Efficiently Detecting Multiple Structural Breaks in Systems of Linear Regression Equations with Integrated and Stationary Regressors

Karsten Schweikert*

University of Hohenheim

[Latest update: January 13, 2022]

Abstract

In this paper, we propose a two-step procedure based on the group LASSO estimator in combination with a backward elimination algorithm to efficiently detect multiple structural breaks in linear regressions with multivariate responses. Applying the two-step estimator, we jointly detect the number and location of change points, and provide consistent estimates of the coefficients. Our framework is flexible enough to allow for a mix of integrated and stationary regressors, as well as deterministic terms. Using simulation experiments, we show that the proposed two-step estimator performs competitively against the likelihood-based approach (Qu and Perron, 2007; Li and Perron, 2017; Oka and Perron, 2018) when trying to detect common breaks in finite samples. However, the two-step estimator is computationally much more efficient. An economic application to the identification of structural breaks in the term structure of interest rates illustrates this methodology.

Keywords: LASSO, shrinkage, model selection, cointegration, multivariate
JEL Classification: C32, C52

*Address: University of Hohenheim, Core Facility Hohenheim & Institute of Economics, Schloss Hohenheim 1 C, 70593 Stuttgart, Germany, e-mail: *karsten.schweikert@uni-hohenheim.de*

1 Introduction

Accounting for structural breaks is crucial in time series analysis, particularly in settings involving long spans of data, where the models are more likely to be affected by multiple structural breaks. More specifically, we focus on systems of equations with a mix of integrated and stationary regressors. Thus far, the literature on structural breaks has provided only few methods applicable to linear regressions with multiple equations and integrated regressors (see, for example, Bai et al., 1998; Li and Perron, 2017; Oka and Perron, 2018). Without prior knowledge about the structural breaks, methods are needed that precisely determine the number of structural breaks, their timing, and simultaneously estimate the model's coefficients. To deal with large sample sizes, it is further necessary to lower the computational burden compared with the existing likelihood-based approaches using dynamic programming techniques to solve the change-point problem (characterized by computational costs quadratic in the number of observations).

We consider a penalized regression approach based on the group LASSO estimator to account for multiple structural breaks in such systems which, to the best of our knowledge, has not been explored in the literature yet. Although estimators based on the penalized regression principle have become popular in the context of change-point problems, few prior studies apply them to linear regressions with integrated regressors (Schmidt and Schweikert, 2021; Schweikert, 2021) or linear regressions with multivariate responses (Gao et al., 2019; Safikhani and Shojaie, 2020). While existing approaches follow a specific-to-general principle utilizing a likelihood-based approach to sequentially increase the number of breakpoints in a model (Bai et al., 1998; Qu and Perron, 2007; Li and Perron, 2017; Oka and Perron, 2018), we take a general-to-specific approach shrinking down the number of breakpoint candidates to find the best fitting model. While the likelihood-based approach employs dynamic programming techniques and is computationally efficient in rather short samples with (possibly) many structural breaks, the proposed model selection approach is particularly useful for long samples with a moderate number of structural breaks (having computational costs linear in the number of observations). Therefore, it is a well-suited solution to account for structural breaks in the long-run relationships between trending variables.

In this paper, we extend the two-step estimator proposed in Chan et al. (2014) for univariate structural break autoregressive (SBAR) processes. To do so, we modify the group LARS algorithm specifically tailored for univariate change-point problems and extend it to cover multivariate systems. Moreover, we generalize the model specification and allow for a mix of stationary and integrated regressors as well as deterministic trends. Consequently, our approach is flexible enough to model structural breaks in several special cases like, for example, seemingly unrelated regression (SUR) models and dynamically augmented cointegrating regressions.[1]

The idea to perceive the change-point problem in linear regressions as a model selection problem has spawned a diverse literature (see, for example, Harchaoui and Lévy-Leduc, 2010; Bleakley and Vert, 2011; Chan et al., 2014; Safikhani and Shojaie, 2020; Schweikert, 2021). In principle, it is possible to shift and turn the regression hyperplane at every point in time using appropriate indicator variables. Finding the true structural breaks corresponds to selecting relevant indicators and eliminating irrelevant indicators thereby optimizing the fit under sparsity. This leads to a high-dimensional penalized regression model with the total number of parameters of the model close to the number of observations. LASSO-type estimators, introduced by Tibshirani (1996), have attractive properties in high-dimensional settings with a sparse model structure. Their objective function includes a penalty for nonzero parameters and a tuning parameter controls the sparsity of the selected model. However, quite restrictive regularity conditions about the design matrix (restricted eigenvalue condition (Bickel et al., 2009) or strong irrepresentable condition (Zhao and Yu, 2006)) need to be imposed to ensure simultaneous variable selection and parameter estimation consistency. Unfortunately, these conditions are usually violated in change-point settings, where adjacent columns of the design matrix differ only by one entry and the design matrix is highly collinear if the sample size grows large. Consequently, the conventional LASSO-type estimators need to be improved to both estimate and select the true model consistently.

Harchaoui and Lévy-Leduc (2010) are among the first to use penalized regression methods to detect structural breaks. They focus on a piecewise constant white noise process and detect structural breaks using a total variation penalty. Bleakley and Vert (2011) use the group fused LASSO for detection of piecewise constant signals and Chan et al. (2014) develop the aforementioned two-step method. In a related study, Jin et al. (2013) apply the smoothly clipped absolute deviation (SCAD) penalty and the minimax concave penalty (MCP) to estimate the number, location and lag length of piecewise autoregressive segments. Ciuperca (2014), Jin et al. (2016) and Qian and Su (2016b) consider LASSO-type estimators for the detection of multiple structural breaks in linear regressions. Behrendt and Schweikert (2021) propose an alternative strategy to eliminate superfluous breakpoints identified by the group LASSO estimator. They suggest a second step adaptive group LASSO which performs comparably to the backward elimination algorithm suggested in Chan et al. (2014). Schweikert (2021) uses the adaptive group LASSO estimator to estimate structural breaks in single-equation cointegrating regressions.

Work on multiple structural change models in the context of a system of multivariate equations is relatively scarce. Quintos (1995, 1997) considers a general time-varying structure

[1]While the model structure, in principle, includes the possibility to consider piece-wise stationary VAR models, our technical analysis relies on Assumption 2 stated below which is not compatible with VAR models. We refer to Safikhani and Shojaie (2020) who use a slightly different penalty to cover a high-dimensional version of this case.

for the reduced-rank matrix of a vector error correction model (VECM) so that both the cointegrating vector and the adjustment dynamics may change over time. Similarly, Seo (1998) develops a test for changing cointegrating vectors and adjustment coefficients at a single unknown breakpoint. Bai et al. (1998) concentrate on dating and estimating a single structural break in vector autoregressions (VARs) and multiple equation cointegrating regressions. Qu and Perron (2007) consider the restricted quasi-maximum likelihood estimation of and inference for multiple structural changes in a system of equations. A sequential break test can be used to determine the number of structural breaks. In related studies, Eo and Morley (2015), Li and Perron (2017), and Oka and Perron (2018) extend the likelihood-based approach in several directions. Eo and Morley (2015) propose confidence sets for the timing of structural break estimation in multiple equation regression models. Li and Perron (2017) introduces the concept of locally ordered breaks. They model structural breaks in systems of equations with a combination of integrated and stationary regressors, dealing with situations where the breaks cannot be separated by a positive fraction of the sample size. Oka and Perron (2018) highlight that the estimation of common breaks allows for a more precise detection of break dates in multivariate systems. They develop common break tests for this assumption.

Recently, the model selection approach has been applied by Gao et al. (2019) and Safikhani and Shojaie (2020) to estimate change-points in a piece-wise stationary VARs. While the former study estimates the change-points for each equation separately, thereby decomposing the problem into smaller single-equation problems, the latter uses a fused LASSO penalty to deal with high-dimensional VAR systems. Additionally, Qian and Su (2016a) apply the adaptive group fused LASSO estimator to estimate common breaks in panel data models.

In the following, we provide a rigorous analysis of the statistical properties of the proposed two-step estimator and extensive simulation experiments to analyze its finite sample properties. We conduct our technical analysis under relatively mild assumptions about the error term process. Prior studies employing LASSO-type estimators to detect structural breaks assume Gaussian white noise error terms (see, for example, Chan et al., 2014; Gao et al., 2019; Safikhani and Shojaie, 2020), which can be useful to model (V)AR processes. However, this assumption is too restrictive in (multiple equations) linear regressions with integrated regressors often having autocorrelated errors with strong persistence. Naturally, it becomes more difficult to detect structural breaks if the error term process is strongly autocorrelated. Under those assumptions, we show that our estimator is able to consistently estimate the number of structural breaks, their timing, and jointly estimates the model's coefficients.

We use simulation experiments to evaluate our new approach against existing approaches like the likelihood-based approach by, inter alia, Qu and Perron (2007), Li and Perron (2017), and Oka and Perron (2018). It is shown that the two-step estimator has competitive finite sample properties with a slight reduction in precision, but substantially improved computational efficiency. Reducing the computational burden over the likelihood-based approach is an important advantage when large sample sizes are available and a moderate number of structural breaks is expected as is often the case in empirical applications involving trending regressors. Another advantage is the joint estimation of the number of breaks, their timing, and the model's coefficients. In the likelihood-based framework, the number of breaks has to be determined based on the evaluation of two tests with the usual implications regarding size and power.[2] In contrast, the approach taken in this paper does not rely on statistical testing, instead we determine the number of breaks as the number of nonzero groups estimated by the group LASSO estimator which is then further reduced by a second step backward elimination algorithm. Consequently, both approaches are conceptually very different so that one approach can serve as a valuable robustness check for the model specification chosen by the other approach.

Finally, we apply the two-step estimator to a term structure model of US interest rates to demonstrate its properties in a real world setting. Relying on the reduced computational burden of the proposed estimator, we are able to estimate the term structure model with daily data over a 30 year span and detect three important structural breaks. Our results reveal substantial differences in the parametrization of those four term structure regimes. While the coefficient estimates for the first two and the last regime at least have reasonable signs, the third regime from August 2010 to April 2014 is characterized by very unusual coefficients. The expectations hypothesis is only supported in the most recent regime.

The paper is organized as follows. Section 2 outlines the proposed model selection procedure to estimate structural breaks in multivariate systems and presents our main technical results. Section 3 is devoted to the Monte Carlo simulation study. Section 4 reports the results of an empirical application of our methodology to the term structure of US interest rates, and Section 5 concludes. Proofs of all theorems in the paper are provided in the Mathematical Appendix.

2 Methodology

Using penalized regression techniques for structural break detection, we aim to divide a set of breakpoint candidates into two groups of active and inactive breakpoints. Our starting points are Chan et al. (2014) and Schweikert (2021), where a two-step procedure is proposed to detect and estimate multiple structural breaks in autoregressive processes and single equation cointegrating regressions, respectively. Here, the model of interest is a multiple equations system of linear regressions with integrated and stationary regressors, q equations, and T time periods.

[2] First, a double maximum test is conducted to test whether at least one break is present, then the exact number of breaks is determined testing the hypothesis of l breaks versus the alternative of $l+1$ breaks. Naturally, the sequential test procedure requires the specification of a nominal significance level α which implies that also for large samples, the number of breaks is overestimated in roughly $\alpha \cdot 100\%$ of all cases.

2.1 First step estimator

We consider the following potentially cointegrated system in triangular form

$$\begin{aligned} Y_t &= AX_t + \delta t + \mu + Bw_t + u_t, \qquad t = 1, 2, \ldots, \\ X_t &= X_{t-1} + \xi_t, \end{aligned} \tag{1}$$

where Y_t is a $q \times 1$ vector of dependent variables, X_t is a $r \times 1$ vector of integrated regressors, w_t is a $s \times 1$ vector of stationary variables, u_t and ξ_t are I(0) error processes. The coefficient matrices A and B have dimension $q \times r$ and $q \times s$, respectively. We study the asymptotic properties of our estimator under the following assumptions about the involved processes:

Assumption 1. *(i)* $u_t = \sum_{j=0}^{\infty} C_j \epsilon_{t-j} = C(L)\epsilon_t$, $\xi_t = \sum_{j=0}^{\infty} D_j e_{t-j} = D(L)e_t$, $C(1)$ *and* $D(1)$ *are full rank,* $\sum_{j=0}^{\infty} j\|C_j\| < \infty$ *and* $\sum_{j=0}^{\infty} j\|D_j\| < \infty$, (ϵ_t, e_t) *are i.i.d. with finite* $4 + a$ $(a > 0)$ *moment.* w_t *is a mean-zero second order stationary process with uniformly bounded* $4 + a$ *moment.*
(ii) Further, we require that

$$\sup_T E\left|\frac{1}{T}\sum_{i=1}^{t} X_{li}u_i\right|^{4+\epsilon} < \infty, \qquad \text{for } 1 \le l \le r,\ 1 \le t \le T \text{ and some } \epsilon > 0,$$

and

$$\sup_T E\left|\frac{1}{T}\sum_{i=1}^{t} w_{li}u_i\right|^{4+\epsilon} < \infty, \qquad \text{for } 1 \le l \le s,\ 1 \le t \le T \text{ and some } \epsilon > 0.$$

Assumption 2. *The error process* u_t *is independent of the regressors for all leads and lags.*

Assumption 3.

$$E\left(\frac{tX_{l,t}^2}{X_{l,1}^2 + X_{l,2}^2 + \cdots + X_{l,t}^2}\right) \le M_X, \qquad \forall t \ge 1 \text{ and } l = 1, \ldots, r.$$

Assumption 4.

$$E\left(\frac{tw_{l,t}^2}{w_{l,1}^2 + w_{l,2}^2 + \cdots + w_{l,t}^2}\right) \le M_w, \qquad \forall t \ge 1 \text{ and } l = 1, \ldots, s.$$

Except for some additional moment conditions in Assumption 1 (ii), these assumptions are identical to those used in Bai et al. (1998). The error term processes are assumed to be linear processes in Assumption 1, satisfying the required conditions to ensure the validity of the functional central limit theorem for partial sum processes constructed from them (see, for example, Theorem 3.4 in Phillips and Solo (1992) and its multivariate extension in Phillips

(1995)). Further, w_t is given as a stationary process with a sufficiently well-behaved distribution. Assumption 2 is less restrictive as it initially seems considering that w_t might include the leads and lags of changes in X_t (see, for example, Saikkonen (1991), Phillips and Loretan (1991), and Stock and Watson (1993) for a treatment of second-order biased estimators in the context of cointegrating regressions).[3]

The moment conditions in Assumption 3 and Assumption 4 are required to ensure the existence of a uniformly bounded first moment over all t. This is satisfied for unit root processes with i.i.d. normal increments and linear Gaussian processes with absolutely summable coefficients, respectively.

We follow Li and Perron (2017) and Oka and Perron (2018) and apply scaling factors in Equation (1) so that the order of all regressors is the same.[4] To simplify notation, we write the system of equations in its stacked form as

$$Y_t = (Z_t' \otimes I)\theta + u_t, \tag{2}$$

where $Z_t = (T^{-1/2}X_t', T^{-1}t, 1, w_t')'$ and $\theta = \mathrm{Vec}(A, \delta, \mu, B)$ is a $d = q(r+2+s)$ column vector, concatenating the coefficients for each regressor over all equations. The operator $\mathrm{Vec}(\cdot)$ stacks the rows of a matrix into a column vector, $\otimes$ denotes the Kronecker product and I is a $q \times q$ identity matrix. While model (1) allows for very flexible specifications and covers several special cases (e.g., SUR models for $A = 0$ and $\delta = 0$, VAR(s) models for $A = 0$ and $\delta = 0$ and $w_t = (Y_{t-1}, \ldots, Y_{t-s})'$, or pure cointegrating regressions for $\delta = \mu = 0$ and $B = 0$), our main focus is on a full model specification with a mix of integrated and stationary regressors.[5] For example, if w_t contains the leads and lags of changes in X_t, we estimate common breaks in a dynamically augmented cointegrating regression with multivariate responses. If it is known to the researcher that some of the coefficients do not change, it is possible to introduce a selection matrix S to consider only partial structural breaks in the sense that some coefficients are constant over the sampling period. The selection matrix contains elements that are 0 or 1 and hence, specifies which regressors appear in each equation.[6]

We assume that the true data generating process includes m^0 true structural breaks. Multiple (partial) structural breaks in the regression coefficients can be expressed using the

[3]We note that Assumption 2 is stronger than the moment conditions given in Assumption 1 (ii) and replaces them in the corresponding results (Theorem 2 – Theorem 4). However, Assumption 2 is not necessary to proof Theorem 1.

[4]Higher order deterministic trends can be included in the model in the same way as long as the corresponding scaling factors are applied.

[5]Note that Assumption 2 is violated in VAR models. We refer to Safikhani and Shojaie (2020) for the appropriate assumptions to show that LASSO-type estimators can be used to estimate piece-wise stationary VAR models.

[6]Note that $S'S$ is idempotent with non zero elements only on the diagonal. The rank of S is equal to the number of coefficients that are allowed to change.

following model

$$Y_t = (Z_t' \otimes I)\theta_{t_0} + \sum_{k=1}^{m} d(t_k)(Z_t' \otimes I)S'S\theta_{t_k} + u_t, \tag{3}$$

where $d(t_k) = 0$ for $t \leq t_k$ and $d(t_k) = 1$ for $t \geq t_k$. The total number of structural breaks in this model is denoted by m, θ_{t_0} is the baseline coefficient vector, and θ_{t_k}, $k = 1, \ldots, m$ are regime-dependent changes in the regression coefficients. In situations where $m > m^0$, our structural break model in Equation (3) considers more structural breaks than necessary which implies that the true coefficient vector does not change at some t_k, i.e., $\theta^0_{t_k} = 0$. For the breakpoints or break dates t_k, it holds by general convention that $1 = t_0 < t_1 < \cdots < t_m < t_{m+1} = T + 1$. The relative timing of breakpoints is denoted by $\tau_k = t_k/T, k \in \{1, \ldots m\}$. We assume that there is a change in at least one coefficient matrix at each true structural break, so that $\|S\theta^0_{t_k}\| \neq 0$. To simplify the notation, we assume that all coefficients change at all breakpoints for the remainder of the paper.

In case of unknown number and timing of structural breaks, each point in time has to be considered as a potential breakpoint. Therefore, it is helpful to estimate the model in Equation (3) with $m = T$ under the condition that the set $\theta(T) = \{\theta_1, \theta_2, \ldots, \theta_T\}$ exhibits a certain sparse nature so that the total number of distinct vectors in the set equals the true number of breaks $m = m^0$. To use a convenient matrix notation, we define

$$\mathcal{Z} = \begin{pmatrix} Z_1' & 0 & 0 & \ldots & 0 \\ Z_2' & Z_2' & 0 & \ldots & 0 \\ Z_3' & Z_3' & Z_3' & \ldots & 0 \\ \vdots & & & & \\ Z_T' & Z_T' & Z_T' & \ldots & Z_T' \end{pmatrix}, \tag{4}$$

$\mathcal{Y} = (Y_1', \ldots, Y_T')'$, $\mathcal{U} = (u_1', \ldots, u_T')'$ and $\boldsymbol{\theta}(T) = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)'$. Furthermore, we define $\boldsymbol{Y} = \text{Vec}(\mathcal{Y})$, $\boldsymbol{Z} = I \otimes \mathcal{Z}$, and $\boldsymbol{U} = \text{Vec}(\mathcal{U})$. Now, the system for T breakpoint candidates can be rewritten as

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\theta}(T) + \boldsymbol{U}, \tag{5}$$

where $\boldsymbol{Y} \in \mathbb{R}^{Tq \times 1}$, $\boldsymbol{Z} \in \mathbb{R}^{Tq \times Td}$, $\boldsymbol{\theta}(T) \in \mathbb{R}^{Td \times 1}$, and $\boldsymbol{U} \in \mathbb{R}^{Tq \times 1}$. Note that this model specification reorders the groups so that $\boldsymbol{\theta}(T)$ contains the breakpoint candidates for each equation successively, i.e. $\boldsymbol{\theta}_i = \boldsymbol{K}\theta_i$ with commutation matrix $\boldsymbol{K}$. We assume that at least one of the baseline coefficients is nonzero to distinguish between active and inactive breakpoints without making any case-by-case considerations. This implies that the vector of true coefficients $\boldsymbol{\theta}^0(T)$ contains $m_0 + 1$ nonzero groups and $\boldsymbol{\theta}_1 \neq \boldsymbol{0}$. For the remainder of this paper, $\boldsymbol{\theta}_i = \boldsymbol{0}$ means that $\boldsymbol{\theta}_i$ has all entries equaling zero and $\boldsymbol{\theta}_i \neq \boldsymbol{0}$ means that $\boldsymbol{\theta}_i$ has at least one non-zero entry. We define the index sets $\bar{\mathcal{A}} = \{1 \leq i \leq T : \boldsymbol{\theta}^0_i \neq \boldsymbol{0}\}$ denoting the indices of truly non-zero coefficients (including the baseline coefficient) and

$\mathcal{A} = \{i \geq 2 : \boldsymbol{\theta}_i^0 \neq \mathbf{0}\}$ denoting the non-zero parameter changes. The set $\mathcal{A}_T = \{\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_m\}$ denotes the m breakpoints estimated in the first step, i.e., indices of those coefficient changes which are estimated to be non-zero. $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$ and $\mathcal{A}^c$ denotes the complementary set.

We propose to estimate the set of coefficient changes $\boldsymbol{\theta}(T)$ by minimizing the following penalized least squares objective function (Yuan and Lin, 2006):

$$Q^*(\boldsymbol{\theta}(T)) \quad = \quad \frac{1}{T} \|\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\theta}(T)\|^2 + \lambda_T \sum_{i=1}^{T} \|\boldsymbol{\theta}_i\|, \tag{6}$$

where λ_T is a tuning parameter and $\|\cdot\|$ denotes the L_2-norm. Minimizing the objective function in (6) yields the group LASSO estimator which is denoted by $\hat{\boldsymbol{\theta}}(T)$.[7] In this way, we frame the detection of structural breaks as a model selection problem and, in principle, are able to use efficient algorithms from this strand of the literature (Huang et al., 2012; Chan et al., 2014; Yau and Hui, 2017) to eliminate irrelevant breakpoint candidates. Depending on the value of λ_T, a sparse solution is obtained so that the number of nonzero groups corresponds to the number of estimated breakpoints and the coefficient changes at each break are contained within the nonzero groups.[8] In the next subsection, we investigate the asymptotic properties of the first step estimator in this setting.

2.2 Asymptotic properties of the first step estimator

We show in the following that the group LASSO estimator for structural breaks in a system of linear regression equations is consistent in terms of prediction error but inherits the same problems, namely estimation inefficiency and model selection inconsistency, as shown for univariate AR models (Chan et al., 2014), single-equation cointegrating regressions Schweikert (2021), and piecewise-stationary VAR models (Gao et al., 2019). As discussed in Chan et al. (2014), any two adjacent columns of the matrix $\mathcal{Z}$ only differ by one entry. Consequently, the restricted eigenvalue condition (Bickel et al., 2009) does not hold in our setting and we cannot establish our consistency proofs based on this assumption (see, e.g., Chan et al., 2014, for a detailed discussion of this issue).

Further assumptions about the timing of true breakpoints ($\tau_k^0,\ k = 1, \ldots, m_0$) and the magnitude of coefficient changes have to be stated to continue our analysis.

Assumption 5. *(i) The break magnitudes are bounded to satisfy* $m_\theta = \min_{1 \leq j \leq m_0+1} \|\boldsymbol{\theta}^0_{t^0_{j-1}}\| \geq \nu > 0$ *and* $M_\theta = \max_{1 \leq j \leq m_0+1} \|\boldsymbol{\theta}^0_{t^0_{j-1}}\| \leq \mathcal{V} < \infty$.

[7] Using a group penalty for $\boldsymbol{\theta}_i$, we assume common breaks across all equations. Efficient estimation of breaks that affect only some equations could be attempted using bi-level selection estimators as proposed in Li et al. (2015) or by treating the equations separately (Gao et al., 2019).

[8] Practical guidance for the choice of λ_T is given in Section 3.

(ii) $\min_{1\leq j\leq m_0+1}|t_j^0 - t_{j-1}^0|/T\gamma_T \to \infty$ *for some* $\gamma_T \to 0$ *and* $\gamma_T/\lambda_T \to \infty$ *as* $T \to \infty$.

The first inequality of Assumption 5(i) is a necessary condition to ensure that a structural break occurs at t_j^0 and the second part excludes the possibility of infinitely large parameter changes.[9] Assumption 5(ii) requires that the length of the regimes between breaks increases with the sample size albeit slower than T. If γ_T is chosen with a slow enough rate, which depends on the tuning parameter sequence λ_T, it allows us to consistently detect and estimate the true break fractions.

The first result for the first step estimator shows that it is consistent in terms of prediction error.

Theorem 1. Under Assumption 1 and Assumption 5, if $\lambda_T = 2dc_0(\log T/T)^{1/4}$ for some $c_0 > 0$, then there exists some $C > 0$ such that with probability greater than $1 - \frac{C}{c_0^4 \log T}$,

$$\frac{1}{T}\|\boldsymbol{Z}\left(\hat{\boldsymbol{\theta}}(T) - \boldsymbol{\theta}^0(T)\right)\|^2 \leq 4dc_0\left(\frac{\log T}{T}\right)^{\frac{1}{4}}(m_0+1)M_\theta.$$

Remark 1. The corresponding convergence rate for univariate piecewise stationary autoregressive processes given in Chan et al. (2014) is $\sqrt{\log T/T}$. However, they assume a white noise error term whereas Assumption 1 in this paper only requires that the error terms of the system follow linear stationary processes and thus allows for persistent errors. The convergence rate given in Safikhani and Shojaie (2020) is not directly comparable because the VAR system is assumed to be high-dimensional with the number of equations increasing with T and a different penalty is used to handle the growing number of time series. Unsurprisingly, Gao et al. (2019) find the same rate as Chan et al. (2014).

The next result shows that the number of estimated breakpoints is at least as large as the true number of breakpoints. Furthermore, the location of the breakpoints can be estimated within an $T\gamma_T$-neighborhood of their true location. To state the theorem, we have to define the Hausdorff distance between the set of estimated breakpoints and the set of true breakpoints. We follow Boysen et al. (2009) and define $d_H(A,B) = \max_{b\in B}\min_{a\in A}|b-a|$ with $d_H(A,\emptyset) = d_H(\emptyset, B) = 1$, where $\emptyset$ is the empty set.

Theorem 2. If Assumption 1 - 5 hold, then as $T \to \infty$

$$P\left(|\mathcal{A}_T| \geq m_0\right) \to 1,$$

and

$$P\left(d_H(\mathcal{A}_T, \mathcal{A}) \leq T\gamma_T\right) \to 1.$$

[9] Our definitions in Assumption 5(i) include the baseline coefficients which can, in principle, be relaxed but simplifies the technical analysis.

Remark 2. When m_0 is known, we can show that the breakpoints are estimated with the same convergence rate (Lemma 4 in the Mathematical Appendix). Setting $\gamma_T = \log T/T$ and $\lambda_T = O(\log T/(T \log \log T))$, the convergence rate is identical to the one found by Chan et al. (2014), assuming Gaussian white noise errors. Hence, it is also faster than the rate found in Schweikert (2021) for cointegrating regressions without scaling the regressors. Finally, Safikhani and Shojaie (2020) show that the rate could be as low as $(\log \log T \log q)/T$ for high-dimensional piecewise stationary VAR processes, where q denotes the number of variables in the VAR system and $q > T$.

Remark 3. The second part of Theorem 2 implies that the Hausdorff distance from the set of estimated breakpoints to the true breakpoints diverges slower than the sample size. Consequently, the Hausdorff distance of the relative location to their true location of the breakpoints converges to zero at rate $\log T/T$. The optimal rate for estimators of break fractions in regression models is $1/T$ (Bai, 1997, 2000). This provides us with a consistency result for the estimated break fractions in the first step and gives us justification to consider multiple structural breaks at once, since the Hausdorff distance evaluates the joint location of all breakpoints.

2.3 Second step estimator

To obtain a consistent estimator for the number of breaks, their timing and coefficient changes, we need to design a second step refinement reducing the number of superfluous breaks. Immediate candidates are using a backward elimination algorithm (BEA) optimizing some information criterion (Chan et al., 2014; Gao et al., 2019; Safikhani and Shojaie, 2020) or applying the adaptive group LASSO estimator as a second step using the group LASSO estimates as weights (Behrendt and Schweikert, 2021; Schweikert, 2021). In the following, we outline the former approach and discuss the latter approach in the Supplementary Material because the BEA produces more accurate results in this setting.[10]

According to Theorem 2, the group LASSO estimator slightly overselects breaks under the right tuning. To distinguish between active and non-active breakpoints in the set $\mathcal{A}_T$, we employ an information criterion for the second step which consists of a goodness-of-fit measure, here the sum of squared residuals, and a penalty term as a function of the number of breaks. We define $\hat{\hat{\boldsymbol{\theta}}}_j$, $1 \leq j \leq m_0$ as the least squares estimator of $\boldsymbol{\theta}_j^0$, based on breakpoints estimated in the first step. Further, we define the sum of squared residuals over all q equations as

$$S_T(t_1, \ldots, t_m) = \sum_{j=1}^{m+1} \sum_{t=t_{j-1}}^{t_j - 1} \|Y_t - \bar{Z}_t \sum_{s=1}^{j} \boldsymbol{K}' \hat{\hat{\boldsymbol{\theta}}}_s\|^2, \tag{7}$$

[10]The Supplementary Material B can be found here: https://karstenschweikert.github.io/mequ_ci/mequ_ci_suppB_20210930.pdf

where $\bar{Z}_t = (Z_t' \otimes I)$. For m and the breakpoints $\boldsymbol{t} = (t_1, \ldots, t_m)$, we can define the information criterion (IC)

$$IC(m, \boldsymbol{t}) = S_T(t_1, \ldots, t_m) + m\omega_T, \tag{8}$$

where ω_T is the penalty term that is further characterized below in Theorem 3. We estimate the number of breaks and the timing by solving

$$(\widehat{\widehat{m}}, \widehat{\widehat{\boldsymbol{t}}}) = \arg \min_{\substack{m \in \{1, \ldots, |\mathcal{A}_T|\} \\ \boldsymbol{t} = (t_1, \ldots, t_m) \subset \mathcal{A}_T}} IC(m, \boldsymbol{t}). \tag{9}$$

If the maximum number of breaks in the first step algorithm is chosen to be small, the evaluation of the information criterion for each combination of breakpoints can be achieved easily. The following result shows that minimizing the IC gives us a consistent estimator for m_0 and $\mathcal{A}$.

Theorem 3. If Assumption 1 - 5 hold and ω_T satisfies the conditions $\lim_{T\to\infty} T\gamma_T/\omega_T = 0$ and $\lim_{T\to\infty} \omega_T / \min\limits_{1 \le i \le m_0} |t_i^0 - t_{i-1}^0| = 0$, then, as $T \to \infty$, $(\widehat{\widehat{m}}, \widehat{\widehat{\boldsymbol{t}}})$ satisfies

$$P\left(\widehat{\widehat{m}} = m_0\right) \to 1,$$

and it exists a constant $B > 0$ such that

$$P\left(\max_{1 \le i \le m_0} |\widehat{\widehat{t}}_i - t_i^0| \le BT\gamma_T\right) \to 1.$$

Remark 4. The conditions for ω_T given in the theorem, combined with the assumption that $\gamma_T/\lambda_T \to \infty$ as $T \to \infty$ and, for example, the sequence of tuning parameters found in Theorem 1 to be of order $O(\log T/T)^{1/4}$, are satisfied for $\omega_T = CT^{3/4} \log T$ for some $C > 0$. For practical purposes, C can be set in analogy to the construction of the BIC so that the second term in Equation (8) penalizes the total number of nonzero coefficients.

Remark 5. Theorem 3 shows that the second step refinement leads to consistent estimation of the true number of breakpoints. The second property is stronger that the one expressed in Theorem 2 and does not refer to the Hausdorff distance but shows that every breakpoint is identified within a $T\gamma_T$ neighborhood.

If $|\mathcal{A}_T|$ is relatively large, evaluating every combination of breakpoints again becomes computationally intensive. Hence, we follow Chan et al. (2014) and use a backward elimination algorithm to successively remove the most redundant breakpoint that corresponds to the largest reduction of the IC until no further improvement is possible. The details of the algorithm are outlined in the Supplementary Material.[11] We denote the set of estimated

[11] The Supplementary Material A can be found here: `https://karstenschweikert.github.io/mequ_ci/mequ_ci_suppA_20210930.pdf`

breakpoints obtained from the BEA by $\mathcal{A}_T^* = (\hat{t}_1^*, \ldots, \hat{t}_{|\mathcal{A}_T^*|}^*)$. The next theorem shows that the estimator based on the BEA has identical asymptotic properties.

Theorem 4. For the same conditions as in Theorem 3, it holds for $T \to \infty$ that

$$P\left(|\mathcal{A}_T^*| = m_0\right) \to 1,$$

and it exists a constant $B > 0$ such that

$$P\left(\max_{1 \le i \le m_0} |\hat{t}_i^* - t_i^0| \le BT\gamma_T\right) \to 1.$$

Using the BEA, it is also possible to optimize another information criterion, say the BIC, for each regime to eliminate some variables from all equations. This allows us to investigate whether some variables lose importance during parts of the sample period. Depending on the chosen model structure this could even be interpreted as some variables dropping out of the long-run equilibrium relationship for a certain period.

Applying scaling factors to the integrated regressors and the linear trend in our model ensures that all regressors have the same order. In turn, this means that the OLS estimator after the first step, $\hat{\hat{\boldsymbol{\theta}}}_j$, has the same convergence rates for all coefficients in the model. In principle, it is also possible to conduct post-LASSO OLS estimation (after the second step) without scaling factors to benefit from higher convergence rates of the estimator for coefficients of trending variables.

3 Simulation

We conduct simulation experiments to assess the adequacy of our technical results presented in Section 2. Specifically, we investigate the finite sample performance of our estimator with respect to the accuracy in finding the exact number of breaks, their location and the magnitude of parameter changes. For the simulations and the empirical application, we rely on a modified group LARS algorithm for the first step and the BEA for the second step. The choice of the tuning parameter λ_T is translated to pre-specifying the maximum number of breakpoint candidates M, i.e. the maximum number of non-zero groups in $\hat{\boldsymbol{\theta}}(T)$, when the LARS implementation is used. Since the group LASSO estimator overselects breaks in the first step, M should be set large enough to encompass all true breakpoints and some additional falsely selected non-zero groups. The BEA then asymptotically guarantees that the set of change-points is attained in the second step. Further, the minimum distance between breaks needs to be specified based on the number of coefficients in the model to guarantee consistent estimation in each regime. Details about the modified group LARS algorithm, the BEA, and additional simulation results are included in Supplementary Material A.

We consider model specifications with one, two and four breakpoints, respectively. The following DGP is employed to model a multiple equations cointegrating regression with multiple structural breaks,

$$\begin{array}{rclcrcl} Y_t & = & A_t X_t + \delta_t t + \mu + B_t w_t + u_t, & & u_t & \sim & N(0, \Sigma_u), \\ X_t & = & X_{t-1} + \xi_t, & & \xi_t & \sim & N(0, \Sigma_\xi), \\ w_t & = & \Phi w_{t-1} + e_t, & & e_t & \sim & N(0, \Sigma_e), \end{array} \tag{10}$$

where $X_t = (X_{1t}, X_{2t}, \dots, X_{Nt})'$, $\Sigma_u = diag(\sigma_u^2)$, $\Sigma_\xi = diag(\sigma_\xi^2)$ and $\Sigma_e = diag(\sigma_e^2)$, i.e. the innovations of our generated processes have multivariate normal distributions with diagonal covariance matrices. μ is a non-zero intercept vector, A_t and B_t are time-varying coefficient matrices with at least one non-zero entry in the baseline specification and a finite number of breaks. δ_t is a time-varying q-dimensional vector and Φ is a coefficient matrix for the VAR(1) process that fulfills the required stationarity conditions. For the main results, we set $q = 2$ and use the following coefficient matrices:

$$A_0 = B_0 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \qquad A_i = A_{i-1} + c \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \qquad B_i = B_{i-1} + c \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad i = 1, \dots, m, \tag{11}$$

with $c = 1$. Moreover, we set $\delta_0 = (2, 2)'$ and $\delta_i = \delta_{i-1} + c(2, 2)'$ for $i = 1, \dots, m$. Similar to Bai et al. (1998), we also specify $c \in \{0.5, 1.5\}$ to investigate the performance for smaller and larger break magnitudes. The results of those robustness checks are included in Supplementary Material A.

Naturally, the ability of all structural break estimators to detect breaks depends on the overall signal strength. Niu et al. (2015) define signal strength in change-point models by $S_{NHZ} = m_\theta^2 I_{\min}$, where $I_{\min} = \min_{1 \le j \le m_0+1} |t_j - t_{j-1}|$ is the minimum distance between breaks and m_θ is the minimum jump size as defined in Assumption 5. For our main simulations concerned with consistency of the two-step estimator, we follow Schweikert (2021) and use equal jump sizes for multiple breaks as well as locating the breaks with equidistant spacing between them. Hence, overall signal strength is a linear function of the sample size in our simulations. We choose a minimum of 50 observation per regime and double the sample sizes in line with the conventional asymptotics specified in Subsection 2.1. Consequently, the sample sizes chosen differ for an increasing number of breakpoints. Note that 12 coefficients are present in the full model specification which requires a substantial number of observations in each regime to estimate them precisely.

In Table 2, we report our results for $r = 2$ integrated regressors, $s = 2$ stationary regressors, a time trend, and $q = 2$ equations. We specify our model for one break located at $\tau = 0.5$, two breaks at $\tau = (0.33, 0.67)$ and four breaks at $\tau = (0.2, 0.4, 0.6, 0.8)$ to have an equidistant spacing on the unit interval. We first compute the percentages of correct estimation (pce) of the number of breaks m and measure the accuracy of the break date estimation conditional on the correct estimation of m. For this matter, we compute the standard deviations of the estimated relative timing. The estimated coefficients are not reported to conserve space but can be obtained from the author upon request.

Our simulation results reveal that the two-step estimator (panel A) is less precise compared with the likelihood-based approach (panel B) when it comes to the estimation of the individual break locations. This is not surprising considering that the likelihood-based approach according to Qu and Perron (2007) rests on a dynamic programming algorithm and uses repeated OLS regressions to determine the location of the breaks in an almost exact fashion. The trade-off here is then computing time versus precision. As outlined above, the computational complexity of the two-step estimator is much lower ($O(T^2)$ versus $O(M^3 + MT)$) which manifests in reduced computing times at moderate to large samples. For example, while it takes the likelihood-approach several minutes on a modern computer to solve the change-point problem for $T = 2,000$, the two-step estimator solves the same problem in seconds.[12] It is also important to note that conceptually the likelihood-based approach is not able to consistently estimate the exact number of breaks which is the case for the two-step estimator. As expected, the sequential test procedure reaches its specified nominal confidence level at very large samples sizes and in our setting ($\alpha = 0.05$) detects the number of breaks in roughly 95% of all cases.[13] Instead, the two-step estimator already attains close to a 100% detection rate at moderate sample sizes. This also means that the most difficult cases (leading to a non-rejection of the sequential test's hypothesis) are not considered for the evaluation of the precision of the likelihood-based approach because in columns two to six of Table 2 only those cases with the correctly estimated number of breaks can be properly evaluated.

We further study the performance of the two-step estimator in settings with structural breaks of smaller or larger magnitude. For instance, we apply the factor $c = 1.5$ to increase the break magnitude in Equation (11). Although we find a slightly better detection of the number of breaks, we observe no substantial effect on the precision in terms of finding the true location of the breaks. It appears that increasing the magnitude does not further improve the performance in small samples. A reason for this upper bound in precision is the pre-selection of breakpoint candidates in the first step which is only accurate up to a $T\gamma_T$ neighborhood of the true breakpoints. In contrast, reducing the magnitude when applying the factor $c = 0.5$ leads to worse results for small sample sizes. In the case of one active break and $T = 100$ observations, we find that the rate of correctly detecting the number of breaks reduces from

[12] More specifically, we obtain the following computational times for both algorithms: using the DGP in Equation 11, a sample size of $T = 2,000$, $c = 1$, and $m = 4$, it takes the two-step estimator 5.89 seconds, but the likelihood-based approach 16 minutes to solve the change-point problem. Of course, the difference in computing times increases for larger sample sizes. All simulations are computed on a computer with an Intel i5-6500 CPU at 3.20GHz and 16GB RAM.

[13] It seems to be undersized for small sample sizes and a larger number of breaks, reaching a 100% detection rate for $m = 4$ and $T = 250$ but then declining to 95% for larger samples.

98.9% in the baseline specification ($c = 1$) to 67.9%. However, when the sample size is increased to $T = 200$, we almost reach the same rate of detection and a similar precision for the timing. We conclude that the performance of the two-step approach naturally depends on the size of the break magnitude but it is fairly robust in medium to large sample sizes which should be its primary field of application considering its improvements in terms of computational costs.

To investigate whether the results are driven by the integrated regressors or the linear trend in the model, we run the simulation experiments for a reduced SUR model specification. Here, we generate data under the restriction that $A_i = diag(0)$ and $\delta_i = 0$ for $i = 0, \dots m$. The results are reported in Table 3. We find almost identical precision for the SUR model. This shows that the precision of the two-step approach is predominately driven by the magnitude of the breaks in terms of their Euclidean distance for the vector of coefficients and the magnitude is the same for the SUR specification. The reduced number of coefficients in the SUR model does not seem to improve the performance substantially, because the sample size in each regime is already large enough to estimate the full model. Additional simulation experiments (not reported) show that this aspect certainly gains importance for smaller regimes with less observations.

We extend the model to a third equation to investigate if the results either improve because a common break is indicated in another regression equation (similar to results obtained in Bai et al., 1998; Qu and Perron, 2007) or deteriorate because the detection relies on additional coefficient changes that need to be estimated. We consider a special case in which the break magnitudes stay constant after adding the third equation. In practice, we hope that estimating the structural breaks jointly in all equations includes some larger coefficient changes that help to find the common break dates. The results for the $q = 3$ case, reported in Table 4, show that while it becomes more difficult to detect the true number of breaks, we reach approximately the same precision for the timing of the breaks and obtain similar standard deviations for the coefficients. This can again be explained by the fact that more difficult cases (with the wrong number of breaks) are excluded. As the break magnitudes are identical for each coefficient, the slightly worse performance compared with the $q = 2$ case can be explained by the additional number of estimated coefficients in each regime.

4 Empirical Application

In our empirical application, we apply our two-step estimator to US term structure data. Thereby, we revisit the study by Hansen (2003) who finds two significant structural breaks in September 1979 and October 1982. Only after accounting for these structural changes, the long-run implication of the expectations hypothesis cannot be rejected. Following Campbell and Shiller (1987), we expect that the term structure of interest rates is that the expected future spot rates equals the future rate plus a time-invariant term premium. This implies that, independent of the maturity, the yields should be cointegrated with cointegrating vector $(1, -1)$. However, several early studies report that the expectations hypothesis fails in empirical practice (see, for example, Froot, 1989; Campbell and Shiller, 1991). Besides other empirical difficulties, structural breaks are named as one of the important reasons for this failure (Lanne, 1999; Sarno et al., 2007; Bulkley and Giordani, 2011). Several studies investigate whether regime shifts in the term structure of interest rates are related to changes in monetary policy (Tillmann, 2007; Thornton, 2018). The important question for applied researchers and policy-makers is whether these equilibrium relationships are robust over different regimes.

Using daily data from January 1990 to July 2021 on the term structure of US interest rates, we end up with more than 8,000 observations to estimate the term structure model. We again emphasize that almost exact segmentation algorithms like that used for the likelihood-based approach are substantially slower than the two-step procedure based on the group LASSO estimator. Taking into account that several re-estimations of the model must be conducted to find the right specification and perform robustness checks, a reduced computational burden is important to encourage routine checks for structural breaks in multivariate systems. We use fitted yields on zero coupon US bonds with 10-year ($r_{10y,t}$), 5-year ($r_{5y,t}$), and 1-year ($r_{1y,t}$) maturity in the term structure model,

$$\begin{aligned} r_{10y,t} &= \mu_1 + \beta_1 r_{1y,t} + u_{1,t} \\ r_{5y,t} &= \mu_2 + \beta_2 r_{1y,t} + u_{2,t}. \end{aligned} \tag{12}$$

Additional maturities could be analyzed, leading to additional equations in the model, but we try to maintain a simple model structure. The data are produced according to the approach of Kim and Wright (2005) fitting a simple three-factor arbitrage-free term structure model to U.S. Treasury yields since 1990, in order to evaluate the behavior of long-term yields, distant-horizon forward rates, and term premiums.[14] Figure 1 provides a time series plot of the data. We assume that the individual variables follow unit root processes.[15] The results of a Johansen trace test for the full sample suggest that the trivariate system is cointegrated with one cointegrating vector.

We estimate the cointegrated term structure regression in Equation (12) with a dynamic OLS specification adding two leads and lags of $\Delta r_{1y,t}$. First, we estimate the model for the full sample without accounting for any structural breaks. The coefficient estimates are $\hat{\beta}_1 = 0.764(0.116)$ and $\hat{\beta}_2 = 0.894(0.078)$ which lead to a rejection of the expectations hypothesis

[14]The data can be downloaded from the St Louis Fed's database FRED.

[15]See the discussion in Hansen (2003) and the references given therein why this assumption is useful for the empirical modelling of the term structure although some reasons like the non-negativity of interest rates are arguments against it. We also conduct unit root tests which do not provide evidence against this hypothesis in this specific sample period.

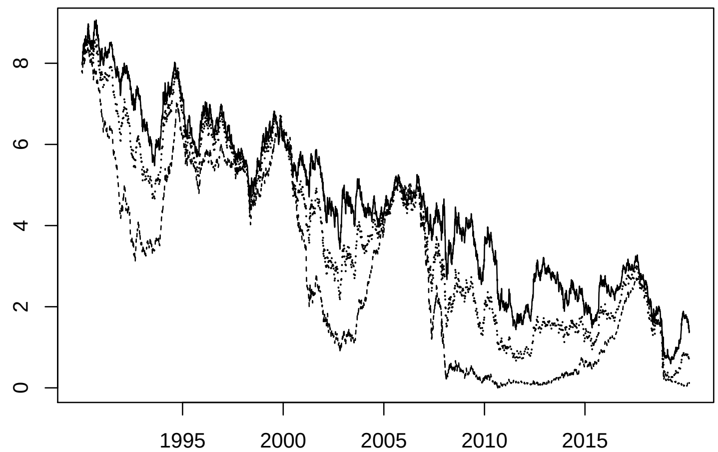

Figure 1: Fitted yields (in %) on 10-year (solid), 5-year (dotted, and 1-year (dashed) zero coupon US bonds.

at the 5% significance level for the 10-year maturity but not for the 5-year maturity. In a second step, we try to capture all relevant structural breaks. Due to the large number of observations, we pre-specify a large maximum number of breaks, $M = 40$, and maintain a minimum break distance of two month (50 daily observations). We consider a specification with a constant, linear trend, and apply the required scaling factors so that each regressor has the same order.[16] Using the two-step group LASSO estimator, we obtain three structural breaks. The first break is located in January 1995, the second break is located in August 2010, and the third break is located in April 2014. After the structural breaks are obtained, we re-estimate the model in each regime without scaling factors for higher precision and report the resulting coefficients in Table 1.

It can be observed that the cointegrating vectors for most regimes substantially differ from $(1, -1)$. A simple t-test of the hypothesis is not rejected for the last regime (t-statistics -0.812 and -1.178, respectively). While the estimates for the first two regimes are reasonable (albeit still being able to reject the expectations hypothesis), the third regime from August 2010 to April 2014 is characterized by very unusual coefficients. However, the third regime can also be associated with an unusually steep yield curve. The first regime ends after the yield spread narrows in 1995 and the second regime ends after the recession following the Global Financial Crisis. The fourth regime is marked by relatively narrow yield spreads that begin to widen during the COVID-19 pandemic. Still, our results suggest that this most recent

[16]The results with and without linear trend do not differ substantially.

Table 1: Regime-specific coefficients of the term structure model

Regimes:	$r_{10y,t}$: $\hat{\mu}_1$	$r_{10y,t}$: $\hat{\beta}_1$	$r_{5y,t}$: $\hat{\mu}_2$	$r_{5y,t}$: $\hat{\beta}_2$
1990 m01 - 1995 m01	6.012 (0.271)	0.364 (0.212)	3.440 (0.208)	0.645 (0.162)
1995 m01 - 2010 m08	5.245 (0.653)	0.220 (0.175)	3.260 (0.491)	0.520 (0.132)
2010 m08 - 2014 m04	2.054 (0.317)	-4.033 (1.462)	0.880 (0.203)	-1.424 (0.954)
2014 m04 - 2021 m07	1.987 (0.328)	0.570 (0.365)	1.133 (0.268)	0.767 (0.287)

Note: The coefficient estimates are obtained from a post-LASSO dynamic OLS estimation. We compute bootstrap standard errors based on 600 replications of the sieve bootstrap method for cointegrating regressions proposed in Chang et al. (2006). Standard errors are given in parentheses.

regime comes closest to satisfying the expectations hypothesis. In summary, accounting for multiple structural breaks in the term structure model reveals some important differences for subsamples of the data but does not solve the term structure puzzle. Instead, it seems that the expectations hypothesis does not hold in the majority of the sampling period.

5 Conclusion

We have proposed a computationally efficient alternative to the existing likelihood-based approach solving the change-point problem in multivariate systems with a mix of integrated and stationary regressors. Our two-step estimator is able to consistently determine the number of breaks, their timing and to jointly estimate the coefficients for each regime. This can be achieved without the need to conduct sequential tests and does not require generating critical values beyond those supplied in the original paper, for example, in situations with larger systems composed of many regressors. The algorithm is much faster than the dynamic programming algorithm used in the likelihood-approach and can solve change-point problems for large multivariate systems and several thousand observations in seconds. In turn, the likelihood-based approach allows for a straightforward construction of valid confidence bands by inverting the likelihood ratio test for a given break date (Eo and Morley, 2015). It remains to be investigated if such confidence bands can be constructed within the model selection approach taken in this paper. Moreover, it is unclear if the two-step estimator can deal with locally ordered breaks as proposed in (Li and Perron, 2017).

The crucial first step estimation is based on the group LASSO estimator and we utilize a group LARS algorithm to solve the change-point problem. Alternative choices of other penalties in the objective function could be used to potentially improve the estimator in some directions. For example, Safikhani and Shojaie (2020) use a fused LASSO penalty to allow the number of equations to grow with the sample size. Such high-dimensional extensions might be useful in panel settings where both the time and the cross-sectional dimension increase asymptotically. In principle, the proposed two-step estimator can be applied to change-point problems in other important model classes. If we use more restrictive assumptions about the error terms, for example, assuming Gaussian white noise errors, we could also deal with structural breaks in VECMs which involve a mix of integrated and stationary variables. Since this application poses additional technical difficulties, we leave this topic for future research.

6 Acknowledgements

The author thanks Konstantin Kuck, Thomas Dimpfl, Markus Mößler, and Robert Jung for valuable comments. Further, he thanks Chun Yip Yau for sharing the programs for the group LARS algorithm, and Zhongjun Qu, Pierre Perron and Tatsushi Oka for sharing the programs for the likelihood-based approach. Funding by the German Research Foundation (Grant SCHW 2062/1-1) is gratefully acknowledged.

A Mathematical Appendix

Lemma 1. *Under Assumption 1, we note that the following properties*

$$
\begin{aligned}
&a) && \frac{1}{T}\sum_{t=k}^{T}\frac{X_{l,t}}{\sqrt{T}}u_{j,t} = O_p(T^{-\frac{1}{2}})\\
&b) && \frac{1}{T}\sum_{t=k}^{T}\frac{t}{T}u_{j,t} = O_p(T^{-\frac{1}{2}})\\
&c) && \frac{1}{T}\sum_{t=k}^{T}u_{j,t} = O_p(T^{-\frac{1}{2}})\\
&d) && \frac{1}{T}\sum_{t=k}^{T}w_{l,t}u_{j,t} = O_p(T^{-\frac{1}{2}}),
\end{aligned}
$$

hold for $1 \leq j \leq q$ *and all* $k = 1, \ldots, T$.

Proof of Lemma 1. Property a) involving integrated regressors follows from Lemma A.4 in Li and Perron (2017). The remaining properties are standard results for (linear) stationary processes.

□

Lemma 2. *Under Assumption 1, for any* $c_0 > 0$*, there exists some constant* $C > 0$ *such that*

$$P\left(\max_{1\leq k\leq T}\max_{1\leq l\leq d}\left|\frac{\boldsymbol{Z}(k,l)'\boldsymbol{U}}{T}\right| \geq c_0\left(\frac{\log T}{T}\right)^{\frac{1}{4}}\right) \leq \frac{C}{c_0^4\log T}. \tag{A.1}$$

Proof of Lemma 2. We first need to define some notation for the proof of this lemma: $\boldsymbol{Z}(k,.)$ is the k-th block column of $\boldsymbol{Z}$ and is of dimension $Tq \times d$. $\boldsymbol{Z}(k,l)$ is the l-th column of the k-th block column. Then, note that the following cases

$$a) \quad \frac{\boldsymbol{Z}(k,l)'\boldsymbol{U}}{T} = \frac{1}{T}\sum_{t=k}^{T}\frac{X_{i,t}}{\sqrt{T}}u_{j,t}, \qquad 1 \leq i \leq r, \quad 1 \leq j \leq q, \tag{A.2}$$

$$b) \quad \frac{\boldsymbol{Z}(k,l)'\boldsymbol{U}}{T} = \frac{1}{T}\sum_{t=k}^{T}\frac{t}{T}u_{j,t}, \qquad 1 \leq j \leq q, \tag{A.3}$$

$$c) \quad \frac{\boldsymbol{Z}(k,l)'\boldsymbol{U}}{T} = \frac{1}{T}\sum_{t=k}^{T}u_{j,t}, \qquad 1 \leq j \leq q, \tag{A.4}$$

$$d) \quad \frac{\boldsymbol{Z}(k,l)'\boldsymbol{U}}{T} = \frac{1}{T}\sum_{t=k}^{T}w_{i,t}u_{j,t}, \qquad 1 \leq i \leq s, \quad 1 \leq j \leq q, \tag{A.5}$$

can appear in the full model depending on the choice of l. In each case, we obtain scalar partial sums. Although $\boldsymbol{Z}(k,l)$ is of dimension $Tq \times 1$, its structure guarantees that not more than T elements are nonzero.

We first turn to case a) and prove the lemma for this case in detail. Under Assumption 1, we have $E\left(\left|\frac{1}{T}\sum_{t=1}^{k}\frac{X_{i,t}}{\sqrt{T}}u_{j,t}\right|^4\right) \leq \frac{C}{32T^2}$ for $1 \leq i \leq r$, $1 \leq j \leq q$, $1 \leq k \leq T$, and all T. It follows that

$$\begin{aligned}
P\left(\max_{1\leq k\leq T}\left|\frac{1}{T}\sum_{t=1}^{k}\frac{X_{i,t}}{\sqrt{T}}u_{j,t}\right| \geq \frac{c_0}{2}\left(\frac{\log T}{T}\right)^{\frac{1}{4}}\right) &\leq \sum_{k=1}^{T} P\left(\left|\frac{1}{T}\sum_{t=1}^{k}\frac{X_{i,t}}{\sqrt{T}}u_{j,t}\right| \geq \frac{c_0}{2}\left(\frac{\log T}{T}\right)^{\frac{1}{4}}\right) \\
&\leq \sum_{k=1}^{T}\frac{16T}{c_0^4\log T}E\left(\left|\frac{1}{T}\sum_{t=1}^{k}\frac{X_{i,t}}{\sqrt{T}}u_{j,t}\right|^4\right) \qquad \text{(A.6)} \\
&\leq \frac{C}{2c_0^4\log T},
\end{aligned}$$

for all i, j, and some $C > 0$. Thus, it holds that

$$\begin{aligned}
& P\left(\max_{1\leq k\leq T}\left|\frac{1}{T}\sum_{t=k}^{T}\frac{X_{i,t}}{\sqrt{T}}u_{j,t}\right| \geq c_0\left(\frac{\log T}{T}\right)^{\frac{1}{4}}\right) \\
= \; & P\left(\max_{1\leq k\leq T}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{X_{i,t}}{\sqrt{T}}u_{j,t} - \frac{1}{T}\sum_{i=1}^{k-1}\frac{X_{l,t}}{\sqrt{T}}u_{j,t}\right| \geq c_0\left(\frac{\log T}{T}\right)^{\frac{1}{4}}\right) \qquad \text{(A.7)} \\
\leq \; & P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\frac{X_{i,t}}{\sqrt{T}}u_{j,t}\right| \geq \frac{c_0}{2}\left(\frac{\log T}{T}\right)^{\frac{1}{4}}\right) + P\left(\max_{1\leq k\leq T}\left|\frac{1}{T}\sum_{t=1}^{k-1}\frac{X_{i,t}}{\sqrt{T}}u_{j,t}\right| \geq \frac{c_0}{2}\left(\frac{\log T}{T}\right)^{\frac{1}{4}}\right) \\
\leq \; & \frac{C}{c_0^4\log T}.
\end{aligned}$$

Now, considering that all components of $d = q(r+2+s)$ are finite, Equation (A.1) follows for case a). The remaining cases are proven similarly, relying on the moment bounds

$$\begin{aligned}
E\left(\left|\frac{1}{T}\sum_{t=1}^{k}\frac{t}{T}u_{j,t}\right|^4\right) &= \frac{1}{T^8}E\left(\left|\sum_{t=1}^{k}tu_{j,t}\right|^4\right) \leq \frac{1}{T^8}E\left(\left|\sum_{t=1}^{k}Tu_{j,t}\right|^4\right) \qquad \text{(A.8)} \\
&= \frac{1}{T^2}E\left(\left|\sum_{t=1}^{k}\frac{u_{j,t}}{\sqrt{T}}\right|^4\right) \leq \frac{C}{32T^2}, \qquad 1 \leq j \leq q,
\end{aligned}$$

and

$$E\left(\left|\frac{1}{T}\sum_{t=1}^{k}w_{i,t}u_{j,t}\right|^4\right) = \frac{1}{T^2}E\left(\left|\sum_{t=1}^{k}\frac{w_{i,t}u_{j,t}}{\sqrt{T}}\right|^4\right) \leq \frac{C}{32T^2}, \qquad 1 \leq i \leq s, \quad 1 \leq j \leq q, \quad \text{(A.9)}$$

respectively. The last inequality in both cases holds, because the processes u_t and w_t are stationary and the partial sums scaled by $1/\sqrt{T}$ have finite fourth moments for each T according to Assumption 1.

$\square$

Lemma 3. *Let $\hat{\boldsymbol{\theta}}(T)$ be the estimator of $\boldsymbol{\theta}(T)$ as defined in Equation (6) and $\bar{Z}_t = (Z_t' \otimes I)$, then it holds under the same conditions as in Theorem 1 that*

$$\sum_{s=\hat{t}_j}^{T} \bar{Z}_s' \left(Y_s - \bar{Z}_s \sum_{i=1}^{s} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_i \right) - \frac{1}{2} T\lambda_T \frac{\boldsymbol{K}'\hat{\boldsymbol{\theta}}_{\hat{t}_j}}{\|\boldsymbol{K}'\hat{\boldsymbol{\theta}}_{\hat{t}_j}\|} = \mathbf{0}, \qquad \forall \hat{\boldsymbol{\theta}}_{\hat{t}_j} \neq \mathbf{0},$$

and

$$\left\| \sum_{s=j}^{T} \bar{Z}_s' \left(Y_s - \bar{Z}_s \sum_{i=1}^{s} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_i \right) \right\| \leq \frac{1}{2} T\lambda_T, \qquad \forall j.$$

Proof of Lemma 3. This lemma is a direct consequence of the Karush-Kuhn-Tucker (KKT) conditions for group LASSO estimators.

$\square$

Lemma 4. *Under Assumption 1 - 5, if m_0 is known and $|\mathcal{A}_T| = m_0$, then*

$$P\left(\max_{1\leq j\leq m_0} |\hat{t}_j - t_j^0| \leq T\gamma_T \right) \to 1, \qquad \text{as } T \to \infty.$$

Proof of Lemma 4. Define $A_{Ti} = \left\{ |\hat{t}_i - t_i^0| > T\gamma_T \right\}$, $i = 1, 2, \ldots, m_0$ such that

$$P\left(\max_{1\leq i\leq m_0} |\hat{t}_i - t_i^0| > T\gamma_T \right) \leq \sum_{i=1}^{m_0} P\left(|\hat{t}_i - t_i^0| > T\gamma_T \right) = \sum_{i=1}^{m_0} P\left(A_{Ti}\right). \tag{A.10}$$

Further define $C_T = \left\{ \max_{1\leq i\leq m_0} |\hat{t}_i - t_i^0| \leq \min_i |t_i^0 - t_{i-1}^0|/2 \right\}$. It suffices to show that

$$\sum_{i=1}^{m_0} P\left(A_{Ti}C_T\right) \to 0 \text{ and } \sum_{i=1}^{m_0} P\left(A_{Ti}C_T^c\right) \to 0. \tag{A.11}$$

The proof follows along the lines of the proof of Theorem 2.2 in Chan et al. (2014) but instead of Lemma A.2 in their paper which bounds the tail probability for β-mixing stationary processes, we rely on the stronger moment conditions stated in Assumptions 1, 3, and 4 and the exogeneity condition in Assumption 2 to do so. In the following, we focus on $\sum_{i=1}^{m_0} P\left(A_{Ti}C_T\right) \to 0$ because the complementary part can be shown using similar arguments.

In the set C_T, it holds that

$$t_{i-1}^0 < \hat{t}_i < t_{i+1}^0, \qquad \forall\, 1 \leq i \leq m_0. \tag{A.12}$$

Next, we split A_{Ti} into two parts (i) $\hat{t}_i < t_i^0$ and (ii) $\hat{t}_i > t_i^0$ to show that $P\left(A_{Ti}C_T\right) \to 0$.

In case of (i), applying Lemma 3 (KKT conditions) yields

$$\|\sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l'(Y_l - \bar{Z}_l \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j)\| \leq T\lambda_T. \tag{A.13}$$

Note that because of $\hat{t}_i < t_i^0$, the true coefficient has not yet changed at $\hat{t}_i$. Hence, plugging in for $Y_l = \bar{Z}_l \sum_{i=j}^{t_i^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 + u_l$ yields

$$\|\sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l' u_l + \sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0\right) + \sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j\right)\| \leq T\lambda_T. \tag{A.14}$$

It follows for $\hat{t}_i < t_i^0$ that,

$$\begin{aligned}
P\left(A_{Ti}C_T\right) \leq\; & P\left(\left\{\tfrac{1}{3}\|\sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0\right)\| \leq T\lambda_T\right\} \cap \left\{|\hat{t}_i - t_i^0| > T\gamma_T\right\}\right) \\
&+P\left(\left\{\|\sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l' u_l\| > \tfrac{1}{3}\|\sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0\right)\|\right\} \cap \left\{|\hat{t}_i - t_i^0| > T\gamma_T\right\}\right) \\
&+P\left(\left\{\|\sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j\right)\|\right.\right. \\
&\left.\left. > \tfrac{1}{3}\|\sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0\right)\|\right\} \cap A_{Ti}C_T\right) \\
&= P\left(A_{Ti1}\right) + P\left(A_{Ti2}\right) + P\left(A_{Ti3}\right).
\end{aligned} \tag{A.15}$$

For the first term, it can be shown under Assumptions 1 - 4 and on the set $\left\{|\hat{t}_i - t_i^0| > T\gamma_T\right\}$ that

$$\begin{aligned}
& \frac{1}{3}\|\sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0\right)\| && \text{(A.16)} \\
\geq\; & \frac{|\hat{t}_i - t_i^0|}{6}\|E(\bar{Z}_{t_i^0-1}'\bar{Z}_{t_i^0-1})\left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0\right)\| && \text{(A.17)} \\
\geq\; & \frac{T\gamma_T}{6}\|E(\bar{Z}_{t_i^0-1}'\bar{Z}_{t_i^0-1})\left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}'\boldsymbol{\theta}_j^0\right)\| = c_0 T\gamma_T, && \text{(A.18)}
\end{aligned}$$

holds with probability going to one. Taking into account that $\gamma_T/\lambda_T \to \infty$, we conclude that

$P(A_{Ti1}) \to 0$ for $T \to \infty$. For the second term, we have

$$\| \sum_{l=\hat{t}_i}^{t_i^0 - 1} \bar{Z}_l' u_l \| = O_p((T\gamma_T)^{\frac{1}{2}}), \tag{A.19}$$

using Lemma 2 but since $\frac{1}{3} \| \sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}' \boldsymbol{\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}' \boldsymbol{\theta}_j^0 \right) \| > c_0 T\gamma_T$ with probability going to one, we conclude that the right hand side of the inequality asymptotically dominates the left hand side and $P(A_{Ti2}) \to 0$ for $T \to \infty$.

Turning to the third term, we apply Lemma 3 to the interval $[t_i^0, (t_i^0 + t_{i+1}^0)/2]$ which yields

$$\| \sum_{l=t_i^0}^{(t_i^0+t_{i+1}^0)/2-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}' \boldsymbol{\theta}_j^0 - \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K}' \hat{\boldsymbol{\theta}}_j \right) \| \le T\lambda_T + \| \sum_{l=t_i^0}^{(t_i^0+t_{i+1}^0)/2-1} \bar{Z}_l' u_l \|. \tag{A.20}$$

Since $|t_{i+1}^0 - t_i^0| \ge 4T\gamma_T$, it follows that for any $x > 0$,

$$\| \sum_{l=t_i^0}^{(t_i^0+t_{i+1}^0)/2-1} \bar{Z}_l' u_l \| \le x |t_{i+1}^0 - t_i^0|. \tag{A.21}$$

However, it also holds that

$$\| \sum_{l=t_i^0}^{(t_i^0+t_{i+1}^0)/2-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K}' \boldsymbol{\theta}_j^0 - \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K}' \hat{\boldsymbol{\theta}}_j \right) \| \tag{A.22}$$

$$\ge \quad \frac{|t_{i+1}^0 - t_i^0|}{4} \| E(\bar{Z}_{t_i^0-1}' \bar{Z}_{t_i^0-1}) \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}' \boldsymbol{\theta}_j^0 - \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K}' \hat{\boldsymbol{\theta}}_j \right) \|, \tag{A.23}$$

with probability approaching one. Combining both results yields

$$\| E(\bar{Z}_{t_i^0-1}' \bar{Z}_{t_i^0-1}) \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K}' \boldsymbol{\theta}_j^0 - \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K}' \hat{\boldsymbol{\theta}}_j \right) \| \le \frac{4T\lambda_T}{|t_{i+1}^0 - t_i^0|} + 4x. \tag{A.24}$$

It follows with probability approaching one that

$$\| \sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K'\hat{\theta}}_j \right) \| \tag{A.25}$$

$$\leq 2(t_i^0 - \hat{t}_i) \| E(\bar{Z}_{t_i^0-1}' \bar{Z}_{t_i^0-1}) \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{\hat{t}_{i+1}-1} \boldsymbol{K'\hat{\theta}}_j \right) \| \tag{A.26}$$

$$\leq \frac{C_1 T \lambda_T (t_i^0 - \hat{t}_i)}{(t_{i+1}^0 - t_i^0)} + C_2 x (t_i^0 - \hat{t}_i), \tag{A.27}$$

but at the same time $\frac{1}{3} \| \sum_{l=\hat{t}_i}^{t_i^0-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K'\theta}_j^0 \right) \| > c_0 (t_i^0 - \hat{t}_i)$ with probability going to one according to Equation (A.18). Since $T\gamma_T/(t_{i+1}^0 - t_i^0) \to 0$ according to Assumption 5, this implies that $P(A_{Ti3}) \to 0$ for $T \to \infty$. It follows that $P\left(A_{Ti} C_T \cap \{\hat{t}_i < t_i^0\} \right) \to 0$.

In case of (ii), we have

$$\| \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' (Y_l - \bar{Z}_l \sum_{j=1}^{\hat{t}_i-1} \boldsymbol{K'\hat{\theta}}_j) \| \leq T\lambda_T, \tag{A.28}$$

Since $\hat{t}_i > t_i^0$, the true coefficient has changed at t_i^0 and we plug in for $Y_l = \bar{Z}_l' \sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K'\theta}_j^0 + u_l$, which yields

$$\| \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' u_l + \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{t_i^0-1} \boldsymbol{K'\theta}_j^0 \right) + \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{\hat{t}_i-1} \boldsymbol{K'\hat{\theta}}_j \right) \| \leq T\lambda_T. \tag{A.29}$$

It follows that,

$$\begin{aligned} P(A_{Ti} C_T) \leq {} & P\left(\left\{ \tfrac{1}{3} \| \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{t_i^0-1} \boldsymbol{K'\theta}_j^0 \right) \| \leq T\lambda_T \right\} \cap \left\{ |\hat{t}_i - t_i^0| > T\gamma_T \right\} \right) \\ & + P\left(\left\{ \| \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' u_l \| > \tfrac{1}{3} \| \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{t_i^0-1} \boldsymbol{K'\theta}_j^0 \right) \| \right\} \cap \left\{ |\hat{t}_i - t_i^0| > T\gamma_T \right\} \right) \\ & + P\left(\left\{ \| \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_i^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{\hat{t}_i-1} \boldsymbol{K'\hat{\theta}}_j \right) \| \right. \right. \\ & \left. \left. > \tfrac{1}{3} \| \sum_{l=t_i^0}^{\hat{t}_i-1} \bar{Z}_l' \bar{Z}_l \left(\sum_{j=1}^{t_{i+1}^0-1} \boldsymbol{K'\theta}_j^0 - \sum_{j=1}^{t_i^0-1} \boldsymbol{K'\theta}_j^0 \right) \| \right\} \cap A_{Ti} C_T \right) \\ = {} & P(A_{Ti1}) + P(A_{Ti2}) + P(A_{Ti3}). \end{aligned} \tag{A.30}$$

The same arguments as for case (i) can be used to show that $P\left(A_{Ti}C_T \cap \{\hat{t}_i > t_i^0\}\right) \to 0$. Combining (i) and (ii) completes the proof of $P\left(A_{Ti}C_T\right) \to 0$.

$\square$

Proof of Theorem 1. By the definition of the group LASSO estimator, we obtain

$$\frac{1}{T}\left\|\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{\theta}}(T)\right\|^2 + \lambda_T \sum_{i=1}^{T} \|\hat{\boldsymbol{\theta}}_i\| \leq \frac{1}{T}\left\|\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\theta}^0(T)\right\|^2 + \lambda_T \sum_{i=1}^{T} \|\boldsymbol{\theta}_i^0\|. \tag{A.31}$$

Denoting $\bar{\mathcal{A}} = \{t_0, t_1, t_2, \ldots, t_{m_0}\}$ and inserting $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\theta}^0(T) + \boldsymbol{U}$ into Inequality (A.31), we have

$$\begin{aligned}
\frac{1}{T}\|\boldsymbol{Z}(\boldsymbol{\theta}^0(T) - \hat{\boldsymbol{\theta}}(T))\|^2 &\leq \frac{2}{T}(\boldsymbol{\theta}^0(T) - \hat{\boldsymbol{\theta}}(T))'\boldsymbol{Z}'\boldsymbol{U} + \lambda_T \sum_{i=1}^{T} \|\boldsymbol{\theta}_i^0\| - \lambda_T \sum_{i=1}^{T} \|\hat{\boldsymbol{\theta}}_i\| \qquad \text{(A.32)} \\
&\leq 2d\left\|\frac{\boldsymbol{Z}'\boldsymbol{U}}{T}\right\|_\infty \left(\sum_{i=1}^{T} \|\boldsymbol{\theta}_i^0 - \hat{\boldsymbol{\theta}}_i\|\right) + \lambda_T \sum_{i=1}^{T} \|\boldsymbol{\theta}_i^0\| - \lambda_T \sum_{i=1}^{T} \|\hat{\boldsymbol{\theta}}_i\| \\
&\leq 2dc_0\left(\frac{\log T}{T}\right)^{\frac{1}{4}} \left(\sum_{i=1}^{T} \|\boldsymbol{\theta}_i^0 - \hat{\boldsymbol{\theta}}_i\|\right) + \lambda_T \sum_{i\in\bar{\mathcal{A}}} (\|\boldsymbol{\theta}_i^0\| - \|\hat{\boldsymbol{\theta}}_i\|) - \lambda_T \sum_{i\in\bar{\mathcal{A}}^c} \|\hat{\boldsymbol{\theta}}_i\| \\
&\leq \lambda_T \sum_{i\in\bar{\mathcal{A}}} \|\boldsymbol{\theta}_i^0 - \hat{\boldsymbol{\theta}}_i\| + \lambda_T \sum_{i\in\bar{\mathcal{A}}} (\|\boldsymbol{\theta}_i^0\| - \|\hat{\boldsymbol{\theta}}_i\|) \\
&\leq 2\lambda_T \sum_{i\in\bar{\mathcal{A}}} \|\boldsymbol{\theta}_i^0\| \leq 2\lambda_T(m_0+1) \max_{1\leq j\leq m_0+1} \|\boldsymbol{\theta}_{t_{j-1}^0}^0\| \\
&= 4dc_0\left(\frac{\log T}{T}\right)^{\frac{1}{4}} (m_0+1)M_\theta.
\end{aligned}$$

with probability greater than $1 - \frac{C}{c_0^4 \log T}$ according to Lemma 2.

$\square$

Proof of Theorem 2. We begin to prove the first part. Suppose that $|\mathcal{A}_T| < m_0$, then there exists some $t_{i_0}^0$, $i_0 = 1, 2, \ldots$ and $\hat{t}_{l_0} \in \mathcal{A}_T \cup \{0, \infty\}$, $l_0 = 0, 1, \ldots, |\mathcal{A}_T| + 1$ with $t_{i_0+1}^0 - t_{i_0}^0 \vee \hat{t}_{l_0} \geq T\gamma_T/3$ and $t_{i_0+2}^0 \wedge \hat{t}_{l_0+1} - t_{i_0+1}^0 \geq T\gamma_T/3$ where $\hat{t}_0 = 0$ and $\hat{t}_{|\mathcal{A}_T|+1} = \infty$.

First, applying Lemma 3 to the interval $[t_{i_0}^0 \vee \hat{t}_{l_0}, t_{i_0+1}^0 - 1]$ yields

$$\|\sum_{l=t_{i_0}^0 \vee \hat{t}_{l_0}}^{t_{i_0+1}^0 - 1} \bar{Z}_l'(Y_l - \bar{Z}_l \sum_{j=1}^{\hat{t}_{l_0+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j)\| \leq T\lambda_T. \tag{A.33}$$

Note that the true coefficient has changed at $t_{i_0}^0$ but does not change until $t_{i_0+1}^0$. Hence, plugging in for $Y_l = \bar{Z}_l \sum_{j=1}^{t^0_{i_0+1}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j + u_l$ yields

$$\begin{aligned}\| \sum_{l=t^0_{i_0} \vee \hat{t}_{l_0}}^{t^0_{i_0+1}-1} \bar{Z}_l' u_l + \sum_{l=t^0_{i_0} \vee \hat{t}_{l_0}}^{t^0_{i_0+1}-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t^0_{i_0+1}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j - \sum_{j=1}^{t^0_{i_0}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j \right) \\ + \sum_{l=t^0_{i_0} \vee \hat{t}_{l_0}}^{t^0_{i_0+1}-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t^0_{i_0}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j - \sum_{j=1}^{\hat{t}_{l_0+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j \right) \| \leq T\lambda_T,\end{aligned} \tag{A.34}$$

and

$$\| \sum_{l=t^0_{i_0} \vee \hat{t}_{l_0}}^{t^0_{i_0+1}-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t^0_{i_0}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j - \sum_{j=1}^{\hat{t}_{l_0+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j \right) \| \leq T\lambda_T + \| \sum_{l=t^0_{i_0} \vee \hat{t}_{l_0}}^{t^0_{i_0+1}-1} \bar{Z}_l' u_l \|. \tag{A.35}$$

Second, applying Lemma 3 to the interval $[t^0_{i_0+1}, t^0_{i_0+2} \wedge \hat{t}_{l_0+1} - 1]$ yields

$$\| \sum_{l=t^0_{i_0+1}}^{t^0_{i_0+2} \wedge \hat{t}_{l_0+1}-1} \bar{Z}_l'(Y_l - \bar{Z}_l \sum_{j=1}^{\hat{t}_{l_0+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j) \| \leq T\lambda_T. \tag{A.36}$$

Since the true coefficient has changed at $t^0_{i_0+1}$ but does not change again until $t^0_{i_0+2}$, we plug in $Y_l = \bar{Z}_l \sum_{j=1}^{t^0_{i_0+2}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j + u_l$ which yields

$$\begin{aligned}\| \sum_{l=t^0_{i_0+1}}^{t^0_{i_0+2} \wedge \hat{t}_{l_0+1}-1} \bar{Z}_l' u_l + \sum_{l=t^0_{i_0+1}}^{t^0_{i_0+2} \wedge \hat{t}_{l_0+1}-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t^0_{i_0+2}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j - \sum_{j=1}^{t^0_{i_0+1}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j \right) \\ + \sum_{l=t^0_{i_0+1}}^{t^0_{i_0+2} \wedge \hat{t}_{l_0+1}-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t^0_{i_0+1}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j - \sum_{j=1}^{\hat{t}_{l_0+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j \right) \| \leq T\lambda_T,\end{aligned} \tag{A.37}$$

and

$$\| \sum_{l=t^0_{i_0+1}}^{t^0_{i_0+2} \wedge \hat{t}_{l_0+1}-1} \bar{Z}_l'\bar{Z}_l \left(\sum_{j=1}^{t^0_{i_0+1}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j - \sum_{j=1}^{\hat{t}_{l_0+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j \right) \| \leq T\lambda_T + \| \sum_{l=t^0_{i_0+1}}^{t^0_{i_0+2} \wedge \hat{t}_{l_0+1}-1} \bar{Z}_l' u_l \|. \tag{A.38}$$

As in the proof of Lemma 4, it can be shown that

$$\| E(\bar{Z}'_{t^0_{i_0+1}} \bar{Z}_{t^0_{i_0+1}}) \left(\sum_{j=1}^{t^0_{i_0}-1} \boldsymbol{K}'\boldsymbol{\theta}^0_j - \sum_{j=1}^{\hat{t}_{l_0+1}-1} \boldsymbol{K}'\hat{\boldsymbol{\theta}}_j \right) \| \leq 2T\lambda_T/(t^0_{i_0+1} - t^0_{i_0} \vee \hat{t}_{l_0}) + 2x, \tag{A.39}$$

and

$$\|E(\bar{Z}'_{t^0_{i_0+2}}\bar{Z}_{t^0_{i_0+2}})\left(\sum_{j=1}^{t^0_{i_0+1}-1}\boldsymbol{K}'\boldsymbol{\theta}^0_j-\sum_{j=1}^{\hat{t}_{l_0+1}-1}\boldsymbol{K}'\hat{\boldsymbol{\theta}}_j\right)\|\leq 2T\lambda_T/(t^0_{i_0+2}\wedge\hat{t}_{l_0+1}-t^0_{i_0+1})+2x,\tag{A.40}$$

holds for arbitrary $x>0$ with probability approaching one. Since $t^0_{i_0+1}-t^0_{i_0}\vee\hat{t}_{l_0}\geq T\gamma_T/3$, $t^0_{i_0+2}\wedge\hat{t}_{l_0+1}-t^0_{i_0+1}\geq T\gamma_T/3$, and $\gamma_T/\lambda_T\to\infty$, it follows that $\|\sum_{j=1}^{t^0_{i_0}-1}\boldsymbol{K}'\boldsymbol{\theta}^0_j-\sum_{j=1}^{\hat{t}_{l_0+1}-1}\boldsymbol{K}'\hat{\boldsymbol{\theta}}_j\|\xrightarrow{p}0$ and $\|\sum_{j=1}^{t^0_{i_0+1}-1}\boldsymbol{K}'\boldsymbol{\theta}^0_j-\sum_{j=1}^{\hat{t}_{l_0+1}-1}\boldsymbol{K}'\hat{\boldsymbol{\theta}}_j\|\xrightarrow{p}0$, which contradicts the condition $\|\sum_{j=1}^{t^0_{i_0}-1}\boldsymbol{K}'\boldsymbol{\theta}^0_j-\sum_{j=1}^{t^0_{i_0+1}-1}\boldsymbol{K}'\boldsymbol{\theta}^0_j\|>\nu>0$ stated in Assumption 5. This implies $|\mathcal{A}_T|\geq m_0$ in probability and concludes the proof of the first part.

Turning to the second part, we define $\hat{T}_k=\{\hat{t}_1,\hat{t}_2,\ldots,\hat{t}_k\}$. Then, it is enough to show that

$$\begin{aligned}&P\left(\{d_H(\mathcal{A}_T,\mathcal{A})>T\gamma_T,m_0\leq|\mathcal{A}_T|\leq T\}\right)\\&=\sum_{k=m_0}^{T}P\left(\{d_H(\hat{T}_k,\mathcal{A})>T\gamma_T\}\right)P\left(|\mathcal{A}_T|=k\right)\to 0,\end{aligned}\tag{A.41}$$

as $T\to\infty$. By Lemma 4, we have already shown that $P\left(d_H(\hat{T}_{m_0},\mathcal{A})>T\gamma_T\right)\to 0$ so that it suffices to show

$$\max_{k>m_0}P\left(d_H(\hat{T}_k,\mathcal{A})>T\gamma_T\right)\to 0.\tag{A.42}$$

Given t^0_i, we define

$$\begin{aligned}B_{T,k,i,1}&=\{\forall 1\leq l\leq k,|\hat{t}_l-t^0_i|\geq T\gamma_T\text{ and }\hat{t}_l<t^0_i\}\\B_{T,k,i,2}&=\{\forall 1\leq l\leq k,|\hat{t}_l-t^0_i|\geq T\gamma_T\text{ and }\hat{t}_l>t^0_i\}\\B_{T,k,i,3}&=\{\exists 1\leq l\leq k-1,\text{ such that }|\hat{t}_l-t^0_i|\geq T\gamma_T,\\&\qquad|\hat{t}_{l+1}-t^0_i|\geq T\gamma_T\text{ and }\hat{t}_l<t^0_i<\hat{t}_{l+1}\}.\end{aligned}\tag{A.43}$$

Then,

$$\max_{k>m_0}P\left(d_H(\hat{T}_k,\mathcal{A})>T\gamma_T\right)=\max_{k>m_0}P\left(\bigcup_{i=1}^{m_0}\bigcup_{j=1}^{3}B_{T,k,i,j}\right).\tag{A.44}$$

Using similar arguments as in the proof of Lemma 4, it can be shown that $\max_{k>m_0}P\left(\bigcup_{i=1}^{m_0}B_{T,k,i,j}\right)\to 0$ for $1\leq j\leq 3$. This completes the proof of Theorem 2.

□

Proof of Theorem 3. We prove the first part of Theorem 3 by showing that (a) $P\left(\widehat{\widehat{m}}<m_0\right)\to 0$ and (b) $P\left(\widehat{\widehat{m}}>m_0\right)\to 0$.

For part (a), Theorem 2 implies that points $t_{Ti} \in \mathcal{A}_T$, $i = 1, \ldots, m_0$ exist such that $\max_{1 \le i \le m_0} |\widehat{t}_{Ti} - t_i^0| \le T\gamma_T$. This implies it is enough to show that if $m < m_0$, then

$$IC(\widehat{\widehat{m}}, \widehat{\widehat{\boldsymbol{t}}}) \ge S_T(\widehat{t}_{T1}, \ldots, \widehat{t}_{Tm_0}) + m_0\omega_T, \tag{A.45}$$

in probability. Let $R_T(m_0) = \{(t_1, \ldots, t_{m_0}) : |t_i - t_i^0| \le T\gamma_T, i = 1, \ldots, m_0\}$. For any $\boldsymbol{t} \in R_T(m_0)$, we can write

$$\begin{aligned}
& S_T(t_1, \ldots, t_{m_0}) \\
= & \sum_{i=1}^{t_1^0 - T\gamma_T - 1} \|Y_i - \bar{Z}_i \boldsymbol{K}' \widehat{\widehat{\boldsymbol{\theta}}}_1\|^2 + \sum_{j=2}^{m_0} \sum_{i=t_{j-1}^0 + T\gamma_T}^{t_j^0 - T\gamma_T - 1} \|Y_i - \bar{Z}_i \sum_{s=1}^{j} \boldsymbol{K}' \widehat{\widehat{\boldsymbol{\theta}}}_s\|^2 \\
+ & \sum_{i=t_m^0 + T\gamma_T}^{T} \|Y_i - \bar{Z}_i \sum_{s=1}^{m_0+1} \boldsymbol{K}' \widehat{\widehat{\boldsymbol{\theta}}}_s\|^2 \\
+ & \sum_{j=1}^{m_0} \sum_{i=t_j^0 - T\gamma_T}^{t_j^0 - 1} \|Y_i - \bar{Z}_i \sum_{s=1}^{j} \boldsymbol{K}' \widehat{\widehat{\boldsymbol{\theta}}}_s\|^2 + \sum_{j=1}^{m_0} \sum_{i=t_j^0}^{t_j^0 + T\gamma_T - 1} \|Y_i - \bar{Z}_i \sum_{s=1}^{j+1} \boldsymbol{K}' \widehat{\widehat{\boldsymbol{\theta}}}_s\|^2 \\
= & L_1 + L_2 + L_3 + L_4 + L_5.
\end{aligned} \tag{A.46}$$

Since $\widehat{\widehat{\boldsymbol{\theta}}}_j$, $1 \le j \le m_0$ is the least squares estimator of $\boldsymbol{\theta}_j^0$ on the intervals $[1, t_1^0 - T\gamma_T - 1]$, $[t_{j-1}^0 + T\gamma_T, t_j^0 - T\gamma_T - 1]$, and $[t_m^0 + T\gamma_T, T]$, it holds that

$$L_1 + L_2 + L_3 \le \left(\sum_{i=1}^{t_1^0 - T\gamma_T - 1} + \sum_{j=2}^{m_0} \sum_{i=t_{j-1}^0 + T\gamma_T}^{t_j^0 - T\gamma_T - 1} + \sum_{i=t_m^0 + T\gamma_T}^{T} \right) \|u_i\|^2. \tag{A.47}$$

Further, it can be shown that

$$L_4 + L_5 \le \left(\sum_{j=1}^{m_0} \sum_{i=t_j^0 - T\gamma_T}^{t_j^0 - 1} + \sum_{j=1}^{m_0} \sum_{i=t_j^0}^{t_j^0 + T\gamma_T - 1} \right) \|u_i\|^2 + A_0 m_0 T\gamma_T, \tag{A.48}$$

with probability approaching 1. Thus, with probability approaching one,

$$S_T(t_1, \ldots, t_{m_0}) \le \sum_{i=1}^{T} \|u_i\|^2 + A_0 m_0 T\gamma_T, \tag{A.49}$$

holds uniformly for all $\boldsymbol{t} \in R_T(m_0)$. This implies that

$$S_T(\widehat{t}_{T1}, \ldots, \widehat{t}_{Tm_0}) \le \sum_{i=1}^{T} \|u_i\|^2 + A_0 m_0 T\gamma_T, \tag{A.50}$$

holds with probability approaching one. However, it can be shown using similar arguments as in Lemma A.4 in Chan et al. (2014) that if $m < m_0$, then

$$S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m) \geq \sum_{i=1}^{T} \|u_i\|^2 + \nu(\min_{1\leq i\leq m_0} |t_i^0 - t_{i-1}^0|), \tag{A.51}$$

with probability approaching one. Combining results (A.50) and (A.51), we obtain with probability approaching one that

$$\begin{aligned}
IC(\widehat{\widehat{m}}, \widehat{\widehat{\boldsymbol{t}}}) &= S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m) + m\omega_T && \text{(A.52)}\\
&\geq \sum_{i=1}^{T} \|u_i\|^2 + \nu(\min_{1\leq i\leq m_0} |t_i^0 - t_{i-1}^0|) + m\omega_T \\
&\geq S_T(\widehat{t}_{T1}, \ldots, \widehat{t}_{Tm_0}) + m_0\omega_T + \nu(\min_{1\leq i\leq m_0} |t_i^0 - t_{i-1}^0|) - A_0 m_0 T\gamma_T - (m_0 - m)\omega_T \\
&\geq S_T(\widehat{t}_{T1}, \ldots, \widehat{t}_{Tm_0}) + m_0\omega_T,
\end{aligned}$$

where the last inequality follows from the conditions $\omega_T / \min_{1\leq i\leq m_0} |t_i^0 - t_{i-1}^0| \to 0$ and $\lim_{T\to\infty} T\gamma_T/\omega_T \leq 1$. This result implies that $P\left(\widehat{\widehat{m}} < m_0\right) \to 0$ for $T \to \infty$ and concludes the proof of part (a).

Next, we turn to part (b). Here, it is enough to show that if $m > m_0$, then $IC(m, \widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m) > IC(m_0, \widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_{m_0})$. We note that, when $m > m_0$, it holds that

$$\begin{aligned}
S_T(\widehat{t}_{T1}, \ldots, \widehat{t}_{Tm_0}) &\geq S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_{m_0}) \geq S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m) && \text{(A.53)}\\
&\geq S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m, t_1^0, \ldots, t_{m_0}^0). && \text{(A.54)}
\end{aligned}$$

It can be shown that

$$S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m, t_1^0, \ldots, t_{m_0}^0) \geq \sum_{i=1}^{T} \|u_i\|^2 - (m + m_0)T\gamma_T. \tag{A.55}$$

From Equation A.54, it follows that

$$S_T(\widehat{t}_{T1}, \ldots, \widehat{t}_{Tm_0}) - S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m) \geq S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_{m_0}) - S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m) \geq 0, \tag{A.56}$$

and, since we have established a upper bound for $S_T(\widehat{t}_{T1}, \ldots, \widehat{t}_{Tm_0})$ in Equation A.55 and a lower bound for $S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m)$ in Equation A.50, it holds with probability approaching one that

$$S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_{m_0}) - S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m) \geq (m + m_0(1 + A_0))T\gamma_T. \tag{A.57}$$

This yields

$$IC(m, \widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_m) - IC(m_0, \widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_{m_0}) \geq (m - m_0)\omega_T - (m + m_0(1 + A_0))T\gamma_T > 0, \quad \text{(A.58)}$$

in probability because of the condition $\lim_{T\to\infty} T\gamma_T/\omega_T = 0$. This implies $P\left(\widehat{\widehat{m}} > m_0\right) \to 0$ for $T \to \infty$ and concludes the proof of part (b).

For the second property, if there exists one $1 \leq i \leq m_0$ such that $|\widehat{\widehat{t}}_i - t_i^0| \geq 2A_0 m_0 T\gamma_T/\nu = BT\gamma_T$, then there must exist a t_l^0 such that $|\widehat{\widehat{t}}_i - t_l^0| \geq BT\gamma_T$ for all $1 \leq i \leq m_0$ and it can be shown that

$$S_T(\widehat{\widehat{t}}_1, \ldots, \widehat{\widehat{t}}_{m_0}) \geq \sum_{i=1}^{T} \|u_i\|^2 + 2A_0 m_0 T\gamma_T. \quad \text{(A.59)}$$

However, this contradicts Equation (A.50) and thus $P\left(\max_{1\leq i\leq m_0} |\widehat{\widehat{t}}_i - t_i^0| \leq BT\gamma_T\right) \to 1$.

□

Proof of Theorem 4. Again, we can decompose the proof of the first property into two parts, (a) $P(|\mathcal{A}_T^*| < m_0) \to 0$ and (b) $P(|\mathcal{A}_T^*| > m_0) \to 0$. Using similar arguments as in the proof of Theorem 3, it follows that $P(|\mathcal{A}_T^*| < m_0) \to 0$ for $T \to \infty$. Turning to part (b), Suppose that $\mathcal{A}_T^* = m > m_0$, then

$$S(\mathcal{A}_T^*) = \min_{(t_1,\ldots,t_m)\subseteq \mathcal{A}_T} S_T(t_1, \ldots, t_m), \quad \text{(A.60)}$$

which implies

$$S_T(\widehat{t}_1^*, \ldots, \widehat{t}_m^*) \leq S_T(\widehat{t}_1, \ldots, \widehat{t}_m) \leq S_T(\widehat{t}_1, \ldots, \widehat{t}_{m_0}) \leq \sum_{i=1}^{T} \|u_i\|^2 + m_0 \mathcal{V} T\gamma_T, \quad \text{(A.61)}$$

with probability approaching one. However, we also have

$$S_T(\widehat{t}_1^*, \ldots, \widehat{t}_m^*) \geq S_T(\widehat{t}_1^*, \ldots, \widehat{t}_m^*, t_1^0, \ldots, t_{m_0}^0) \geq \sum_{i=1}^{T} \|u_i\|^2 - (m + m_0)T\gamma_T, \quad \text{(A.62)}$$

with probability approaching one. Using the previous inequalities as in Theorem 3, it follows that $P(|\mathcal{A}_T^*| > m_0) \to 0$ which concludes the proof of the first property. The proof of the second property follows as in Theorem 3.

□

B Tables

Table 2: Estimation of (multiple) structural breaks in the full model

Panel A: Group LASSO with BEA

SB1: ($\tau = 0.5$)

T	*pce*	τ
100	98.9	0.501 (0.014)
200	100	0.500 (0.007)
400	100	0.500 (0.003)
800	100	0.500 (0.001)

SB2: ($\tau_1 = 0.33$, $\tau_2 = 0.67$)

T	*pce*	τ_1	τ_2
150	91.7	0.337 (0.030)	0.660 (0.024)
300	98.4	0.334 (0.017)	0.667 (0.014)
600	99.8	0.332 (0.009)	0.668 (0.007)
1200	100	0.331 (0.004)	0.669 (0.003)

SB4: ($\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, $\tau_4 = 0.8$)

T	*pce*	τ_1	τ_2	τ_3	τ_4
250	89.0	0.217 (0.030)	0.404 (0.022)	0.596 (0.019)	0.788 (0.028)
500	98.2	0.203 (0.017)	0.402 (0.012)	0.598 (0.009)	0.803 (0.012)
1000	99.9	0.199 (0.008)	0.401 (0.006)	0.599 (0.005)	0.800 (0.008)
2000	100	0.200 (0.003)	0.401 (0.003)	0.599 (0.002)	0.800 (0.003)

Panel B: Likelihood-based approach

SB1: ($\tau = 0.5$)

T	*pce*	τ
100	91.3	0.499 (0.041)
200	93.0	0.500 (0.010)
400	94.5	0.500 (0.005)
800	94.7	0.500 (0.003)

SB2: ($\tau_1 = 0.33$, $\tau_2 = 0.67$)

T	*pce*	τ_1	τ_2
150	94.0	0.327 (0.005)	0.667 (0.004)
300	95.0	0.330 (0.002)	0.670 (0.002)
600	96.1	0.330 (0.001)	0.670 (0.001)
1200	95.5	0.330 (0.001)	0.670 (0.001)

SB4: ($\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, $\tau_4 = 0.8$)

T	*pce*	τ_1	τ_2	τ_3	τ_4
250	99.8	0.200 (0.004)	0.400 (0.004)	0.600 (0.004)	0.800 (0.004)
500	96.7	0.200 (0.002)	0.400 (0.001)	0.600 (0.001)	0.800 (0.001)
1000	95.6	0.200 (0.001)	0.400 (0.001)	0.600 (0.001)	0.800 (0.001)
2000	95.1	0.200 (0.001)	0.400 (0.000)	0.600 (0.000)	0.800 (0.000)

Note: We use 1,000 replications of the data-generating process given in Equation (10). The variance of the error terms is $\sigma_\xi^2 = \sigma_e^2 = \sigma_u^2 = 1$. The first panel reports the results for one active breakpoint at $\tau = 0.5$, the second panel considers two active breakpoints at $\tau_1 = 0.33$ and $\tau_2 = 0.67$ and the third panel has four active breakpoints at $\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, and $\tau_4 = 0.8$. Standard deviations are given in parentheses. We conduct the $\sup(l+1|l)$ test at the 5% level to determine the number of breaks.

Table 3: Estimation of (multiple) structural breaks in the SUR model specification using the group LASSO with BEA

	SB1: ($\tau = 0.5$)				
T	*pce*	τ			
100	100	0.500 (0.010)			
200	100	0.500 (0.005)			
400	100	0.500 (0.002)			
800	100	0.500 (0.001)			
	SB2: ($\tau_1 = 0.33$, $\tau_2 = 0.67$)				
T	*pce*	τ_1	τ_2		
150	98.9	0.337 (0.034)	0.656 (0.033)		
300	100	0.333 (0.019)	0.667 (0.019)		
600	100	0.333 (0.009)	0.668 (0.009)		
1200	100	0.331 (0.005)	0.669 (0.005)		
	SB4: ($\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, $\tau_4 = 0.8$)				
T	*pce*	τ_1	τ_2	τ_3	τ_4
250	87.2	0.215 (0.031)	0.406 (0.021)	0.594 (0.020)	0.786 (0.031)
500	99.9	0.201 (0.013)	0.402 (0.012)	0.597 (0.011)	0.800 (0.013)
1000	100	0.199 (0.008)	0.401 (0.006)	0.599 (0.005)	0.801 (0.008)
2000	100	0.200 (0.003)	0.400 (0.003)	0.599 (0.003)	0.800 (0.003)

Note: We use 1,000 replications of the data-generating process given in Equation (10). The variance of the error terms is $\sigma_\xi^2 = \sigma_e^2 = \sigma_u^2 = 1$. The first panel reports the results for one active breakpoint at $\tau = 0.5$, the second panel considers two active breakpoints at $\tau_1 = 0.33$ and $\tau_2 = 0.67$ and the third panel has four active breakpoints at $\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, and $\tau_4 = 0.8$. Standard deviations are given in parentheses.

Table 4: Estimation of (multiple) structural breaks in the full model ($q = 3$) using the group LASSO with BEA

	SB1: ($\tau = 0.5$)				
T	pce	τ			
100	95.1	0.500 (0.012)			
200	100	0.500 (0.005)			
400	100	0.500 (0.002)			
800	100	0.500 (0.001)			
	SB2: ($\tau_1 = 0.33$, $\tau_2 = 0.67$)				
T	pce	τ_1	τ_2		
150	81.7	0.336 (0.027)	0.662 (0.020)		
300	93.9	0.333 (0.017)	0.667 (0.014)		
600	99.7	0.333 (0.008)	0.668 (0.007)		
1200	100	0.331 (0.004)	0.669 (0.004)		
	SB4: ($\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, $\tau_4 = 0.8$)				
T	pce	τ_1	τ_2	τ_3	τ_4
250	86.5	0.218 (0.031)	0.406 (0.020)	0.597 (0.019)	0.788 (0.029)
500	96.6	0.202 (0.017)	0.402 (0.011)	0.597 (0.010)	0.802 (0.013)
1000	99.2	0.199 (0.008)	0.401 (0.006)	0.599 (0.005)	0.800 (0.008)
2000	99.8	0.200 (0.003)	0.401 (0.003)	0.599 (0.002)	0.800 (0.003)

Note: We use 1,000 replications of a variant of the data-generating process given in Equation (10) with an additional equation but the same break magnitudes. The variance of the error terms is $\sigma_\xi^2 = \sigma_e^2 = \sigma_u^2 = 1$. The first panel reports the results for one active breakpoint at $\tau = 0.5$, the second panel considers two active breakpoints at $\tau_1 = 0.33$ and $\tau_2 = 0.67$ and the third panel has four active breakpoints at $\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, and $\tau_4 = 0.8$. Standard deviations are given in parentheses.

References

Bai, J., 1997. Estimation of a Change Point in Multiple Regression Models. Review of Economics and Statistics 79, 551–563. doi:10.1162/003465397557132.

Bai, J., 2000. Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices. Annals of Economics and Finance 1, 303–339.

Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and Dating Common Breaks in Multivariate Time Series. The Review of Economic Studies 65, 395–432. doi:10.1111/1467-937X.00051.

Behrendt, S., Schweikert, K., 2021. A Note on Adaptive Group Lasso for Structural Break Time Series. Econometrics and Statistics 17, 156–172. doi:10.1016/j.ecosta.2020.04.001.

Bickel, P.J., Ritov, Y., Tsybakov, A.B., 2009. Simultaneous analysis of lasso and dantzig selector. Annals of Statistics 37, 1705–1732. doi:10.1214/08-AOS620.

Bleakley, K., Vert, J., 2011. The group fused lasso for multiple change-point detection , 1–25URL: http://arxiv.org/abs/1106.4199.

Boysen, L., Kempe, A., Liebscher, V., Munk, A., Wittich, O., 2009. Consistencies and rates of convergence of jump-penalized least squares estimators. Annals of Statistics 37, 157–183. doi:10.1214/07-AOS558.

Bulkley, G., Giordani, P., 2011. Structural breaks, parameter uncertainty, and term structure puzzles. Journal of Financial Economics 102, 222–232. doi:10.1016/j.jfineco.2011.05.009.

Campbell, J.Y., Shiller, R.J., 1987. Cointegration and Tests of Present Value Models. Journal of Political Economy 95, 1062–1088. doi:10.1086/261502.

Campbell, J.Y., Shiller, R.J., 1991. Yield spreads and interest rate movements: A bird's eye view. Review of Economic Studies 58, 495–514. doi:10.2307/2298008.

Chan, N.H., Yau, C.Y., Zhang, R.M., 2014. Group LASSO for Structural Break Time Series. Journal of the American Statistical Association 109, 590–599. doi:10.1080/01621459.2013.866566.

Chang, Y., Park, J.Y., Song, K., 2006. Bootstrapping cointegrating regressions. Journal of Econometrics 133, 703–739. doi:10.1016/j.jeconom.2005.06.011.

Ciuperca, G., 2014. Model selection by LASSO methods in a change-point model. Statistical Papers 55, 349–374. doi:10.1007/s00362-012-0482-x.

Eo, Y., Morley, J., 2015. Likelihood-ratio-based confidence sets for the timing of structural breaks. Quantitative Economics 6, 463–497. doi:10.3982/qe186.

Froot, K.A., 1989. New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates. The Journal of Finance 44, 283–305. doi:10.1111/j.1540-6261.1989.tb05058.x.

Gao, W., Yang, H., Yang, L., 2019. Change points detection and parameter estimation for multivariate time series. Soft Computing doi:10.1007/s00500-019-04135-8.

Hansen, P.R., 2003. Structural changes in the cointegrated vector autoregressive model. Journal of Econometrics 114, 261–295. doi:10.1016/S0304-4076(03)00085-X.

Harchaoui, Z., Lévy-Leduc, C., 2010. Multiple Change-Point Estimation With a Total Variation Penalty. Journal of the American Statistical Association 105, 1480–1493. doi:10.1198/jasa.2010.tm09181.

Huang, J., Breheny, P., Ma, S., 2012. A Selective Review of Group Selection in High-Dimensional Models. Statistical Science 27, 481–499. doi:10.1214/12-STS392.

Jin, B., Shi, X., Wu, Y., 2013. A novel and fast methodology for simultaneous multiple structural break estimation and variable selection for nonstationary time series models. Statistics and Computing 23, 221–231. doi:10.1007/s11222-011-9304-6.

Jin, B., Wu, Y., Shi, X., 2016. Consistent two-stage multiple change-point detection in linear models. Canadian Journal of Statistics 44, 161–179. doi:10.1002/cjs.11228.

Kim, D.H., Wright, J.H., 2005. An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates. Finance and Economics Discussion Series - Federal Reserve Board , 1–21doi:10.17016/feds.2005.33.

Lanne, M., 1999. Near Unit Roots and the Predictive Power of Yield Spreads for Changes in Long-Term Interest Rates. The Review of Economics and Statistics 81, 393–398. doi:10.1162/003465399558328.

Li, Y., Nan, B., Zhu, J., 2015. Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure. Biometrics 71, 354–363. doi:10.1111/biom.12292.

Li, Y., Perron, P., 2017. Inference on locally ordered breaks in multiple regressions. Econometric Reviews 36, 289–353. doi:10.1080/07474938.2015.1114552.

Niu, Y.S., Hao, N., Zhang, H., 2015. Multiple Change-point Detection: A Selective Overview. Statistical Science 31, 611–623. doi:10.1214/16-STS587.

Oka, T., Perron, P., 2018. Testing for common breaks in a multiple equations system. Journal of Econometrics 204, 66–85. doi:10.1016/j.jeconom.2018.01.003.

Phillips, P.C.B., 1995. Fully Modified Least Squares and Vector Autoregression. Econometrica 63, 1023–1078. doi:10.2307/2171721.

Phillips, P.C.B., Loretan, M., 1991. Estimating Long-run Economic Equilibria. Review of Economic Studies 58, 407–436.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. The Annals of Statistics 20, 971–1001. doi:10.1214/aos/1176348666.

Qian, J., Su, L., 2016a. Shrinkage estimation of common breaks in panel data models via adaptive group fused Lasso. Journal of Econometrics 191, 86–109. doi:10.1016/j.jeconom.2015.09.004.

Qian, J., Su, L., 2016b. Shrinkage Estimation of Regression Models With Multiple Structural Changes. Econometric Theory 32, 1376–1433. doi:10.1017/S0266466615000237.

Qu, Z., Perron, P., 2007. Estimating and testing structural change in multivariate regressions. Econometrica 75, 459–502. doi:10.1111/j.1468-0262.2006.00754.x.

Quintos, C.E., 1995. Sustainability of the deficit process with structural shifts. Journal of Business and Economic Statistics 13, 409–417. doi:10.1080/07350015.1995.10524615.

Quintos, C.E., 1997. Stability tests in error correction models. Journal of Econometrics 82, 289–315. doi:10.1016/S0304-4076(97)00059-6.

Safikhani, A., Shojaie, A., 2020. Joint Structural Break Detection and Parameter Estimation in High-Dimensional Nonstationary VAR Models. Journal of the American Statistical Association , 1–14doi:10.1080/01621459.2020.1770097.

Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7, 1–21. doi:10.1017/S0266466600004217.

Sarno, L., Thornton, D.L., Valente, G., 2007. The empirical failure of the expectations hypothesis of the term structure of bond yields. Journal of Financial and Quantitative Analysis 42, 81–100. doi:10.1017/s0022109000002192.

Schmidt, A., Schweikert, K., 2021. Multiple Structural Breaks in Cointegrating Regressions: A Model Selection Approach. Studies in Nonlinear Dynamics & Econometrics , forthcoming. doi:10.1515/snde-2020-0063.

Schweikert, K., 2021. Oracle Efficient Estimation of Structural Breaks in Cointegrating Regressions. Journal of Time Series Analysis , forthcoming. doi:10.1111/jtsa.12593.

Seo, B., 1998. Tests for Structural Change in Cointegrated Systems. Econometric Theory 14, 222–259. doi:10.1017/S0266466698142044.

Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820. doi:10.2307/2951763.

Thornton, D.L., 2018. Greenspan's Conundrum and the Fed's Ability to Affect Long-Term Yields. Journal of Money, Credit and Banking 50, 513–543. doi:10.1111/jmcb.12469.

Tibshirani, R., 1996. Regression Selection and Shrinkage via the Lasso. Journal of the Royal Statistical Society B (Methodological) 58, 267–288. doi:10.2307/41262671.

Tillmann, P., 2007. Inflation regimes in the US term structure of interest rates. Economic Modelling 24, 203–223. doi:10.1016/j.econmod.2006.07.001.

Yau, C.Y., Hui, T.S., 2017. LARS-type algorithm for group lasso. Statistics and Computing 27, 1041–1048. doi:10.1007/s11222-016-9669-7.

Yuan, M., 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. doi:10.1111/j.1467-9868.2005.00532.x.

Zhao, P., Yu, B., 2006. On Model Selection Consistency of Lasso. The Journal of Machine Learning Research 7, 2541–2563. doi:10.1109/TIT.2006.883611.