Schweikert, Karsten

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

Oxford Bulletin of Economics and Statistics

Oxford Bulletin of Economics and Statistics

ORIGINAL ARTICLE OPEN ACCESS

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

Karsten Schweikert

Core Facility Hohenheim & Institute of Economics, University of Hohenheim, Stuttgart, Germany

Correspondence: Karsten Schweikert (karsten.schweikert@uni-hohenheim.de)

Received: 26 April 2024 | **Revised:** 14 January 2025 | **Accepted:** 28 January 2025

Funding: This work was supported by German Research Foundation (Grant SCHW 2062/1-1).

Keywords: cointegration | LASSO | model selection | multivariate | shrinkage

ABSTRACT

In this paper, we propose a two-step procedure based on the group LASSO estimator in combination with a backward elimination algorithm to detect multiple structural breaks in linear regressions with multivariate responses. Applying the two-step estimator, we jointly detect the number and location of structural breaks 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 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.

JEL Classification: C32, C52

MSC2020 Classification: 62E20, 62M10, 91B84

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 a few methods applicable to linear regressions with multiple equations and integrated regressors see, for example, [1–3]. Without prior knowledge about the structural breaks, methods are needed that precisely determine the number of structural breaks and their timing, and simultaneously estimate the model's coefficients.

Considering the increasingly larger sample sizes in empirical studies, [4] cautions against the use of algorithms that are $O(T^\eta)$ for $\eta \gg 1$ which become impractical for sufficiently large T. The currently available methods to solve change-point problems in models that we consider in this paper consist of likelihood-based approaches using dynamic programming techniques [5, 6] and are characterised by quadratic computational costs ($\eta = 2$). While these approaches are generally very precise, they are not computationally efficient in situations where T is large. Consequently, we propose to lower the computational burden and design a structural break detection algorithm that is feasible for large T systems.

We consider a penalised regression approach based on the group LASSO estimator to account for multiple structural breaks in such systems, which has not been explored in the literature yet. Although estimators based on the penalised regression principle have become popular in the context of change-point problems, few prior studies apply them to linear regressions with integrated regressors [7, 8] or linear regressions with multivariate responses [9, 10]. While existing approaches follow a specific-to-general principle utilising a likelihood-based approach to sequentially increase the number of breaks in a model [1–3, 11], 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. It can be shown that it has linear computational costs ($\eta = 1$). 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 [12] 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 generalise 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, e.g., 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, [8, 10, 12–14]. 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 optimising the fit under sparsity. This leads to a high-dimensional penalised regression model with the total number of parameters of the model close to the number of observations. LASSO-type estimators, introduced by [15], have attractive properties in high-dimensional settings with a sparse model structure. Their objective function includes a penalty for non-zero parameters and a tuning parameter controls the sparsity of the selected model. However, quite restrictive regularity conditions about the design matrix (restricted eigenvalue condition [16] or strong irrepresentable condition [17]) 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.

[13] are among the first to use penalised regression methods to detect structural breaks. They focus on a piecewise constant white noise process and detect structural breaks using a total variation penalty [14] use the group fused LASSO for detection of piecewise constant signals and [12] develop the aforementioned two-step method [18], [19] and [20] consider LASSO-type estimators for the detection of multiple structural breaks in linear regressions [21] 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 [12] [7] consider cointegration tests in the presence of structural breaks in the long-run relationship and estimate those breaks with an adaptive LASSO estimator [8] uses the adaptive group LASSO estimator to estimate structural breaks in single-equation cointegrating regressions. The estimator developed in this paper can be understood as an extension of the [8] approach to multiple equation cointegrated systems, so that structural change of more than one equilibrium relationship can be modelled.

Another general-to-specific approach to detect structural breaks is the indicator saturation approach described, inter alia, in [22] and [23]. Similar to a penalised regression approach, the starting point is a saturated model with break indicators for every observation, but instead of estimating the regression with an objective function that induces sparsity, a backward elimination path is constructed, removing, one-by-one, non-significant regressors. Outliers can be identified jointly with structural breaks if the search path is designed accordingly. Moreover, this approach to break detection imposes no minimum break length and reliably controls for the false discovery rate. It thus constitutes a very general break detection algorithm but it is considerably slower than the LARS algorithm [24]. Recently, this approach has also been used to model structural breaks in cointegrated vector autoregressions [25].

Work on multiple structural change models in the context of a system of multivariate equations is relatively scarce [26, 27] considers a general time-varying structure 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, [28] develops a test for changing cointegrating vectors and adjustment coefficients at a single unknown breakpoint [3] concentrate on dating and estimating a single structural break in vector autoregressions (VARs) and multiple equation cointegrating regressions [11] 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, [29], [1] and [2] extend the likelihood-based approach in several directions.[2]

Recently, the model selection approach has been applied by [9] and [10] to estimate change-points in 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.

In the following, we provide a rigorous analysis of the statistical properties of the proposed two-step estimator and extensive simulation experiments to analyse 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, [9, 10, 12], 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 serially correlated errors. Naturally, it becomes more difficult to detect structural breaks if the error term process is serially correlated. Under those assumptions, we show that our estimator is able to consistently estimate the number of structural breaks and their timing and jointly estimate the model's coefficients.

We use simulation experiments to evaluate our new approach against existing approaches like the likelihood-based approach by, inter alia, [11], [1] and [2]. 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.[3] 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 non-zero groups estimated by the group LASSO estimator, which is then further reduced by a second step backward elimination algorithm. While we also need to ensure that each identified regime has a sufficient number of observations to estimate the coefficient changes, we need a much smaller trimming parameter than commonly applied in the literature e.g., 15% of the sample between breakpoints as suggested in [2, 11]. Hence, the two-step estimator is able to detect breaks near the boundary much more reliably than the likelihood-based approach. Furthermore, since we detect structural breaks by the Euclidean norm of the group of coefficient changes, the precision of the group LASSO estimator is mostly determined by the total magnitude of each break. This implies that we do not rely on a distinction between common and partial breaks, which is important for the properties of hypothesis tests conducted in the likelihood-based approach to determine the number of breaks. In total, both approaches are conceptually very different, so 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 period of 30 years and detect four important structural breaks. Our results reveal substantial differences in the parametrisation of the resulting five term structure regimes. The estimates of the proportionality coefficient are smaller than one in every regime and the long-run implications of the expectations hypothesis are rejected in three out of five regimes.

The paper is organised 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 penalised 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 [12] and [8], 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.

2.1 | First Step Estimator

We consider the following potentially cointegrated system in triangular form

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

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 dimensions $q \times r$ and $q \times s$, respectively. We study the asymptotic properties of our estimator under the following assumptions about the involved processes:

Assumption 1.

a. $u_t = \sum_{j=0}^{\infty} C_j \epsilon_{t-j} = C(L)\epsilon_t$, $\xi_t = \sum_{j=0}^{\infty} D_j e_{t-j} = D(L)e_t$, $C(1)$ 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. The long-run covariance matrix $\Omega_\xi = \sum_{j=-\infty}^{\infty} E\xi_t \xi'_{t-j}$ is positive definite.

b. Further, we require that

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

and

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

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

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, e.g., Theorem 3.4 in [30] and its multivariate extension in [31]). Further, w_t is given as a stationary process with a sufficiently well-behaved distribution. Note that these assumptions could be replaced by other sufficient conditions that conform with a strong invariance principle or functional central limit theorem. We assume that Ω_ξ is positive definite, which implies that X_t is non-cointegrated. This ensures that the coefficients of X_t have the standard T rate of convergence. Assumption 2 is chosen for technical reasons but it is less restrictive than it initially seems considering that w_t might include the leads and lags of changes in X_t (see, e.g., [32], [33] and [34] for treatment of second-order biased estimators in the context of cointegrating regressions).[4]

Assumption 1 allows for serially correlated error terms but this conflicts with endogenous regressors like lagged dependent variables. The results presented in the following remain valid for endogenous regressors as long as the errors are not permitted to be serially correlated [5, discuss this issue on page 51].[5]

We follow [1] and [2] and apply scaling factors in Equation (1) so that the order of all regressors is the same.[6] 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. For example, if w_t contains the leads and lags of changes in X_t, we estimate structural breaks in a dynamically augmented cointegrating regression with multivariate responses. Assumption 2 is violated in VAR models, which invalidates our technical analysis for this specification. Hence, we refer to [10] for the appropriate assumptions to show that LASSO-type estimators can be used to estimate piece-wise stationary VAR models. 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 entire sampling period. This allows the researcher to estimate the respective coefficients with full efficiency. The selection matrix contains elements that are either 0 or 1 and hence, specifies which regressors appear in each equation.[7] For example, we can use the selection matrix to focus only on breaks in the matrix A, i.e., detecting breaks in the long-run coefficients of a cointegrating regression but leaving the matrix B constant over the sample period.

We assume that the system includes m_0 true structural breaks. Multiple (partial) structural breaks in the regression coefficients can be expressed using the 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 $t \le t_k$ and $d(t_k) = 1$ for $t \ge t_k$. The total number of potential structural breaks or breakpoint candidates 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 an 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_0. To use a convenient matrix notation, we define

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

$\mathcal{Y} = (Y_1', \ldots, Y_T')'$, $\mathcal{U} = (u_1', \ldots, u_T')'$ and $\boldsymbol{\theta}(T) = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_T')'$. Furthermore, we define $\boldsymbol{Y} = \mathrm{Vec}(\mathcal{Y})$, $\boldsymbol{Z} = I \otimes \mathcal{Z}$, and $\boldsymbol{U} = \mathrm{Vec}(\mathcal{U})$. Now, the system for $m = T$ 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 non-zero 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 $m_0 + 1$ non-zero 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 $\mathcal{A} = \{1 \le i \le T : \boldsymbol{\theta}^0_i \neq \boldsymbol{0}\}$ denoting the indices of truly non-zero coefficients (including the baseline coefficient) and $\mathcal{A} = \{i \ge 2 : \boldsymbol{\theta}^0_i \neq \boldsymbol{0}\}$ denoting the non-zero parameter changes. The set $\mathcal{A}_T = \{\hat{\imath}_1, \hat{\imath}_2, \ldots, \hat{\imath}_{\hat{m}}\}$ denotes the $\hat{m}$ 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 $\theta(T)$ by minimising the following penalised least squares objective function [35]:

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

where λ_T is a tuning parameter and $\|\cdot\|$ denotes the L_2-norm. Minimising the objective function in (6) yields the group LASSO estimator, which is denoted by $\hat{\theta}(T)$. Using a group penalty for θ_i, in principle, we assume common breaks across all equations as defined by [11] and [2]. However, it should be noted that the issue of distinguishing between common or partial breaks is crucial for the specification of hypothesis tests required for the likelihood-based approach to detect the true number of breaks, but it is not as important for the group LASSO estimator. Since we detect structural breaks by the Euclidean norm of the group of coefficient changes (vectorising all coefficient matrices), only the total magnitude of each break enters our objective function. Consequently, prior knowledge about partial breaks would not improve our sensitivity in detecting those breaks as much as it would in the likelihood-based approach.

Using these definitions, we frame the detection of structural breaks as a model selection problem and are able to use efficient algorithms from this strand of the literature [12, 36, 37] to eliminate irrelevant breakpoint candidates. Depending on the value of λ_T, a sparse solution is obtained so that the number of non-zero groups corresponds to the number of estimated breaks and the coefficient changes at each breakpoint are contained within the non-zero groups.[8]

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 [12], single-equation cointegrating regressions [8], and piecewise-stationary VAR models [9]. As discussed in [12], any two adjacent columns of the matrix $\mathcal{Z}$ only differ by one entry. Consequently, the restricted eigenvalue condition [16] does not hold in our setting and we cannot establish our consistency proofs based on this assumption.

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

Assumption 3.

a. The break magnitudes are bounded to satisfy $m_\theta = \min_{1\le j\le m_0+1} \|\theta^0_{t^0_{j-1}}\| \ge \nu > 0$ and $M_\theta = \max_{1\le j\le m_0+1} \|\theta^0_{t^0_{j-1}}\| \le \mathcal{V} < \infty$.

b. $\min_{1\le j\le m_0+1} |t^0_j - t^0_{j-1}|/T\gamma_T \to \infty$ for some $\gamma_T \to 0$ and $\gamma_T/\lambda_T \to \infty$ as $T \to \infty$.

The first inequality of Assumption 3 (i) is a necessary condition to ensure that a structural break occurs at t^0_j and the second part excludes the possibility of infinitely large parameter changes.[9] We do not consider breaks with local-to-zero behaviour in this setting (see [3] for assumptions used in this context). This assumption is not believed to be restrictive for the intended empirical applications. In the case of a full specification with integrated regressors, applied researchers aim to estimate structural long-run relationships that are often needed for their follow-up analysis, e.g., in a cointegrated VAR as in [38]. Essentially, they need optimal in-sample forecasts in terms of mean squared error of the cointegrating regressions under structural instability to consistently estimate deviations from the long-run equilibrium [39] show that in-sample forecasts are largely unaffected by local-to-zero breaks. Assumption 3 (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 if the tuning parameter λ_T grows at the right rate.

Theorem 1. *Under Assumptions 1 and 3, 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* $1 - \frac{C}{c_0^4 \log T}$,

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

Remark 1. The corresponding convergence rate for univariate piecewise stationary autoregressive processes given in [12] is $\sqrt{\log T/T}$. However, they assume β-mixing stationary processes with a white noise error term so that the necessary tail bounds can be derived. Instead, Assumption 1 in this paper only requires that the error terms of the system follow stationary processes with some additional moment conditions, i.e., allows for serially correlated errors. The convergence rate given in [10] 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. Expectedly, [9] find the same rate as [12].

The next result shows that the number of estimated breaks is at least as large as the true number of breaks. Furthermore, the location of the breakpoints can be estimated within an $T\gamma_T$-neighbourhood 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 [40] 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 Assumptions 1–3 hold, then as* $T \to \infty$

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

and

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

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 [12], assuming Gaussian white noise errors. Note that the tuning parameter can be set differently than in Theorem 1 as the consistency in prediction error is not important for the proof. For the optimal convergence rate, the tuning parameter must be set conforming to a rate that has a higher order than the one in Theorem 1 and thus penalises the coefficient changes more strongly. Finally, [10] show that the rate could be as low as $(\log\log T \log q)/T$ for high-dimensional piecewise stationary VAR processes with Gaussian white noise error terms, 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 $\log T/T$ and therefore is only slightly above the optimal rate $1/T$ for estimators of break fractions in regression models [41, 42]. 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) optimising some information criterion [9, 10, 12] or applying the adaptive group LASSO estimator as a second step using the group LASSO estimates as weights [8, 21]. In the following, we outline the former approach and discuss the latter approach in Supporting Information B because the BEA produces more accurate results in our simulation experiments.[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 $\widehat{\widehat{\theta}}_j, 1 \le j \le m$ as the least squares estimator of θ_j^0, based on breakpoints estimated in the first step. Further, we define the sum of squared residuals over all q equations as

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

where $\overline{Z}_t = (Z_t' \otimes I)$. 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 characterised below in Theorem 3. We estimate the number of breaks and the timing by solving

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

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 minimising the IC gives us a consistent estimator for m_0 and $\mathcal{A}$.

Theorem 3. *If Assumptions 1–3 hold and ω_T satisfies the conditions $\lim_{T\to\infty} T\gamma_T/\omega_T = 0$ and $\lim_{T\to\infty} \omega_T / \min_{1\le i\le m_0} |t_i^0 - t_{i-1}^0| = 0$, 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 $T \to \infty$, determine the penalty term of the IC. Through γ_T, the penalty term is linked to the tuning parameter of the group LASSO estimator. For example, using the sequence of tuning parameters found in Theorem 1 and being of order $O(\log T/T)^{1/4}$, the conditions 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) penalises the total number of non-zero coefficients.

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

If $|\mathcal{A}_T|$ is relatively large, evaluating every combination of breakpoints again becomes computationally intensive. Hence, we follow [12] 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. This means that including this breakpoint improves the fit sufficiently to outweigh the costs of estimating the coefficients for an additional regime. The details of the algorithm are outlined in Supporting Information A.[11] We denote the set of estimated breakpoints obtained from the BEA by $\mathcal{A}_T^* = (\widehat{t}_1^*, \ldots, \widehat{t}_{|\mathcal{A}_T^*|}^*)$. 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

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

Using the BEA, it is also possible to optimise 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 in the second step, $\widehat{\widehat{\theta}}_j$, 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 finite sample performance of the two-step estimator with respect to the speed and accuracy in finding the exact number of breaks and their location. In principle, we can optimise (6) directly using a coordinate descent algorithm like the one proposed in [43] for a grid of λ_T values. However, such an algorithm is not computationally efficient for change-point problems. Thus, for the simulations and the empirical application, we rely on a modified group LARS algorithm that approximates the solution for the first step and the BEA for the second step. The choice of the tuning parameter λ_T is translated into pre-specifying the maximum number of breakpoint candidates M, i.e., the maximum number of non-zero groups in $\hat{\theta}(T)$ that is returned when the LARS implementation is used. The modified group LARS algorithm evaluates each point in time as a breakpoint candidate but returns only the M most relevant breaks. According to Theorem 2, we know that the group LASSO estimator overselects breaks in the first step. Hence, 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 that these coefficients can be estimated accurately in each regime. Details about the modified group LARS algorithm, the BEA, and additional simulation results are included in Supporting Information 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,

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

where $X_t = (X_{1t}, X_{2t}, \ldots, X_{Nt})'$, $\Sigma_{u,ii} = \sigma_u^2$, $\Sigma_{\xi,ii} = \sigma_\xi^2$ and $\Sigma_{e,ii} = \sigma_e^2$, for $i = 1, \ldots, q$, i.e., the innovations of our generated processes have multivariate normal distributions with identical variances. We choose the variances of these processes so that each column of $Z_t = (T^{-1/2}X_t', T^{-1}t, w_t')'$ has variances less than or equal to the error term variances. To achieve this, we set the variances to $\sigma_\xi^2 = \sigma_e^2 = \sigma_u^2 = 1$. 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 simplicity, we choose a diagonal matrix with each diagonal entry equal to 0.5. For the main results, we set $q = 2$ and use the following coefficient matrices:

$$\begin{aligned} A_{t_0} = B_{t_0} &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \qquad A_{t_j} = A_{t_{j-1}} + c \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \\ B_{t_j} &= B_{t_{j-1}} + c \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad j = 1, \ldots, m_0 \end{aligned} \tag{11}$$

with $c = 1$ and the subscript t_0 denoting the initial coefficient matrix before the first breakpoint. Moreover, we set $\mu = (2, 2)'$, $\delta_{t_0} = (2, 2)'$ and $\delta_{t_j} = \delta_{t_{j-1}} + c(2, 2)'$ for $j = 1, \ldots, m_0$. In our baseline specification, the coefficient changes amount to two standard deviations of the error terms. Similar to [3], we also specify $c \in \{0.25, 0.5, 1.5\}$ to investigate the performance for smaller and larger break magnitudes. Moreover, we investigate the effect of cross-correlated errors, serially correlated errors, and endogenous regressors. The results of those robustness checks are included in Supporting Information A to conserve space.

Naturally, the ability of all structural break estimators to detect breaks depends on the overall signal strength [44] 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 3. For our main simulations concerned with showing the consistency of the two-step estimator, we follow [8] 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 observations per regime and double the sample sizes in line with the conventional asymptotics specified in Section 2.1.

In Table 1, 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. Since the choice of a change-point detection algorithm is often reflective of a trade-off between speed and accuracy, we first compute the average computing time in seconds (Time (s)) for each sample size.[12] Next, we compute the percentages of correct estimation (pce) of the true number of breaks m_0 and measure the accuracy of the break date estimation conditional on the correct estimation of m_0. 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

TABLE 1 | Estimation of (multiple) structural breaks in the full model.

	Panel A: Group LASSO with BEA					
	SB1: ($\tau = 0.5$)					
T	**Time (s)**	pce	τ			
100	0.04	98.9	0.501 (0.014)			
200	0.08	100	0.500 (0.007)			
400	0.26	100	0.500 (0.003)			
800	1.64	100	0.500 (0.001)			
	SB2: ($\tau_1 = 0.33, \tau_2 = 0.67$)					
T	**Time (s)**	pce	τ_1	τ_2		
150	0.08	97.6	0.335 (0.030)	0.659 (0.026)		
300	0.25	100	0.333 (0.018)	0.666 (0.014)		
600	1.53	100	0.331 (0.009)	0.668 (0.007)		
1200	10.65	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	**Time (s)**	pce	τ_1	τ_2	τ_3	τ_4
250	0.17	89.0	0.217 (0.030)	0.404 (0.022)	0.596 (0.019)	0.788 (0.028)
500	0.74	98.2	0.203 (0.017)	0.402 (0.012)	0.598 (0.009)	0.803 (0.012)
1000	4.68	99.9	0.199 (0.008)	0.401 (0.006)	0.599 (0.005)	0.800 (0.008)
2000	36.25	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	**Time (s)**	pce	τ			
100	1.22	91.3	0.499 (0.041)			
200	4.99	93.0	0.500 (0.010)			
400	18.71	94.5	0.500 (0.005)			
800	89.07	94.7	0.500 (0.003)			
	SB2: ($\tau_1 = 0.33, \tau_2 = 0.67$)					
T	**Time (s)**	pce	τ_1	τ_2		
150	4.61	94.0	0.327 (0.005)	0.667 (0.004)		
300	15.28	95.0	0.330 (0.002)	0.670 (0.002)		
600	60.16	96.1	0.330 (0.001)	0.670 (0.001)		
1200	317.78	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	**Time (s)**	pce	τ_1	τ_2	τ_3	τ_4
250	16.59	99.8	0.200 (0.004)	0.400 (0.004)	0.600 (0.004)	0.800 (0.004)
500	64.93	96.7	0.200 (0.002)	0.400 (0.001)	0.600 (0.001)	0.800 (0.001)
1000	322.58	95.6	0.200 (0.001)	0.400 (0.001)	0.600 (0.001)	0.800 (0.001)
2000	2013.69	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 $\sigma_\xi^2 = \sigma_c^2 = \sigma_u^2 = 1$. The first subpanel reports the results for one active breakpoint at $\tau = 0.5$, the second subpanel considers two active breakpoints at $\tau_1 = 0.33$ and $\tau_2 = 0.67$ and the third subpanel has four active breakpoints at $\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, and $\tau_4 = 0.8$. Time (s) denotes the average computing time for one replication in seconds. Standard deviations are given in parentheses. We conduct the $\sup(l+1|l)$ test at the 5% level to determine the number of breaks. Critical values are provided in [11].

break locations. This is not surprising considering that the likelihood-based approach, according to [11] 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 vs. precision. As outlined above, the computational efficiency of the two-step estimator is much higher ($O(M^3 + MT)$ vs. $O(T^2)$), which manifests in reduced computing times at moderate to large samples. For example, while it takes the likelihood-approach several minutes (on average 2,013 s) on a modern computer to solve the change-point problem for $T = 2,000$, the two-step estimator solves the same problem in seconds (on average 36 s). It is also important to note that, in contrast to the two-step estimator, the likelihood-based approach is conceptually not able to consistently estimate the exact number of breaks. As expected, the sequential test procedure reaches its specified nominal confidence level at large sample 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 1 only those cases with the correctly estimated number of breaks can be properly evaluated [5] suggest to let the size of those sequential tests go to zero asymptotically to avoid misspecification. However, implementing this in practice is difficult without any guidance on the exact rate for an adjusted α. Alternatively, an information criterion might be used to compare specifications with a different number of breaks.

We further study the performance of the two-step estimator in settings with structural breaks of smaller or larger magnitude.[14] For instance, we apply the factor $c = 1.5$ to increase the break magnitude in Equation (11) to three standard deviations of the error term. 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 $T\gamma_T$ neighbourhood of the true breakpoints. In contrast, reducing the magnitude when applying the factor $c = 0.5$, i.e., to one standard deviation, leads to worse results for small sample sizes. For example, in the case of four active breaks and $T = 250$ observations, we find that the rate of correctly detecting the number of breaks drops from 89.0% in the baseline specification to 68.3%. However, when the sample size increases to $T = 500$, we almost reach the same rate of detection and a similar precision for the break's location. The likelihood-based approach is still very precise when the break magnitudes are reduced by the factor $c = 0.5$. However, when we reduce them further by setting $c = 0.25$, the detection rate breaks down to, e.g., 6.7% for four breaks and $T = 250$ observations while the two-step approach still reaches a 40.1% detection rate. 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 $i = 0, \ldots, m_0$. The results are reported in Table 2. We consider two variants of the simulation experiment: First without cross-correlation between the error terms for both equations as in Equation (10) and second with cross-correlation coefficient $\rho = 0.5$. We find almost identical precision for the SUR model and no substantial difference in the presence of moderate cross-correlation. This shows that the precision of the two-step approach is predominately driven by the magnitude of 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 with 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 [3, 11] 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 3, show that we reach similar detection rates and approximately the same precision for the timing of the breaks. As the break magnitudes are identical for each coefficient, the $q = 3$ setting is a straightforward extension of the $q = 2$ setting. To investigate further whether breaks in a subset of the coefficients can be reliably detected in a larger system, we consider a partial break specification in the $q = 2$ case. Here, only the coefficients of the first equation change. The results are reported in Table S5 in Supporting Information A and again show that the two-step estimator's performance is mostly determined by the total break magnitude which is slightly reduced in this case.

Finally, we turn to the one-break case of our main specification again and move the break to the right boundary of the unit interval. We fix the number of observations in the second regime to 25 and increase the overall sample size. Consequently the breakpoint moves further to the boundary each time we double the sample size. The true break fractions for the sample size $T \in \{100, 200, 400, 800\}$ are 0.75, 0.875, 0.9375, and 0.96875, respectively. We also consider a two-break setting where the middle regime shrinks relative to the total sample size. We set the first breakpoint at $\tau_1 = 0.5$ and the second one at $\tau_2 = 0.5 + 25/T$. The results for both experiments are reported in Table 4. These exercises show that the two-step estimator is robust to setting a small minimum break distance while the sequential tests to detect breaks in the likelihood-based approach require a rather large trimming parameter to ensure the right size and sufficient power. Of course, if the trimming parameter is set to 0.15 as it is suggested for empirical data [11], it is infeasible to detect breaks that are too close to each other or near the boundary

TABLE 2 | Estimation of (multiple) structural breaks in the SUR model specification using the group LASSO with BEA.

	Panel A: No cross-correlation					
	SB1: ($\tau = 0.5$)					
T	**Time (s)**	pce	τ			
100	0.03	100	0.500 (0.010)			
200	0.06	100	0.500 (0.005)			
400	0.20	100	0.500 (0.002)			
800	1.27	100	0.500 (0.001)			
	SB2: ($\tau_1 = 0.33, \tau_2 = 0.67$)					
T	**Time (s)**	pce	τ_1	τ_2		
150	0.04	98.9	0.337 (0.034)	0.656 (0.033)		
300	0.23	100	0.333 (0.019)	0.667 (0.019)		
600	1.39	100	0.333 (0.009)	0.668 (0.009)		
1200	9.70	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	**Time (s)**	pce	τ_1	τ_2	τ_3	τ_4
250	0.15	87.2	0.215 (0.031)	0.406 (0.021)	0.594 (0.020)	0.786 (0.031)
500	0.63	99.9	0.201 (0.013)	0.402 (0.012)	0.597 (0.011)	0.800 (0.013)
1000	10.01	100	0.199 (0.008)	0.401 (0.006)	0.599 (0.005)	0.801 (0.008)
2000	72.30	100	0.200 (0.003)	0.400 (0.003)	0.599 (0.003)	0.800 (0.003)
	Panel B: Cross-correlated errors ($\rho = 0.5$)					
	SB1: ($\tau = 0.5$)					
T	**Time (s)**	pce	τ			
100	0.03	100	0.500 (0.010)			
200	0.06	100	0.500 (0.005)			
400	0.22	100	0.500 (0.002)			
800	1.36	100	0.500 (0.001)			
	SB2: ($\tau_1 = 0.33, \tau_2 = 0.67$)					
T	**Time (s)**	pce	τ_1	τ_2		
150	0.05	97.8	0.337 (0.034)	0.656 (0.033)		
300	0.22	99.8	0.334 (0.019)	0.667 (0.018)		
600	1.68	100	0.332 (0.009)	0.668 (0.009)		
1200	9.99	100	0.332 (0.005)	0.669 (0.005)		
	SB4: ($\tau_1 = 0.2, \tau_2 = 0.4, \tau_3 = 0.6, \tau_4 = 0.8$)					
T	**Time (s)**	pce	τ_1	τ_2	τ_3	τ_4
250	0.17	93.1	0.216 (0.031)	0.406 (0.020)	0.594 (0.020)	0.785 (0.031)
500	0.58	99.8	0.200 (0.013)	0.402 (0.011)	0.597 (0.011)	0.800 (0.013)
1000	12.26	100	0.199 (0.008)	0.401 (0.006)	0.599 (0.005)	0.801 (0.008)
2000	78.36	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 subpanel reports the results for one active breakpoint at $\tau = 0.5$, the second subpanel considers two active breakpoints at $\tau_1 = 0.33$ and $\tau_2 = 0.67$ and the third subpanel has four active breakpoints at $\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, and $\tau_4 = 0.8$. Time (s) denotes the average computing time for one replication in seconds. Standard deviations are given in parentheses.

TABLE 3 | Estimation of (multiple) structural breaks in the full model ($q = 3$) using the group LASSO with BEA.

			SB1: ($\tau = 0.5$)			
T	**Time (s)**	pce	τ			
100	0.04	99.8	0.500 (0.014)			
200	0.08	100	0.500 (0.005)			
400	0.35	100	0.500 (0.002)			
800	2.15	100	0.500 (0.001)			
			SB2: ($\tau_1 = 0.33, \tau_2 = 0.67$)			
T	**Time (s)**	pce	τ_1	τ_2		
150	0.08	98.1	0.333 (0.032)	0.663 (0.025)		
300	0.28	100	0.333 (0.017)	0.667 (0.014)		
600	1.72	100	0.333 (0.008)	0.668 (0.007)		
1200	14.49	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	**Time (s)**	pce	τ_1	τ_2	τ_3	τ_4
250	0.18	89.3	0.217 (0.031)	0.405 (0.021)	0.597 (0.019)	0.789 (0.029)
500	0.85	97.7	0.202 (0.016)	0.402 (0.011)	0.597 (0.009)	0.802 (0.013)
1000	12.25	99.2	0.199 (0.008)	0.401 (0.006)	0.599 (0.005)	0.800 (0.008)
2000	103.64	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 subpanel reports the results for one active breakpoint at $\tau = 0.5$, the second subpanel considers two active breakpoints at $\tau_1 = 0.33$ and $\tau_2 = 0.67$ and the third subpanel has four active breakpoints at $\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, and $\tau_4 = 0.8$. Time (s) denotes the average computing time for one replication in seconds. Standard deviations are given in parentheses.

of the unit interval. In case of one break at the boundary, the likelihood-based estimator in our simulations (almost) always falsely indicates a break at 0.85.

4 | Empirical Application

In our empirical application, we apply the two-step estimator to US term structure data. Thereby, we revisit the study by [38] who proposes a framework to test for structural change in cointegrated vector autoregressions. The author tests two potential structural breaks in September 1979 and October 1982 that coincide with large changes in the Fed's policy. Only after accounting for these structural changes, the long-run implication of the expectations hypothesis (EHT) cannot be rejected. In the subsequent analysis, we study more recent data on the US term structure and detect structural breaks without assuming any prior knowledge about their location.

Following [45], we expect that the term structure of interest rates is determined by the expected future spot rates being equal to the future rate plus a time-invariant term premium in the long-run. This implies that independent of the maturity, the yields should be cointegrated with pairwise cointegrating vector $(1, -1)$. However, several early studies report that the EHT fails in empirical practice see, for example, [46, 47]. Besides other empirical difficulties, structural breaks are named as one of the important reasons for this failure [48–50]. Several studies investigate whether regime shifts in the term structure of interest rates are related to changes in monetary policy [51, 52]. 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 emphasise that almost exact segmentation algorithms like the one used for the likelihood-based approach are substantially slower than the two-step procedure based on the group LASSO estimator (26,380 s vs. 25 s for solving the change point problem). 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,

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

We choose a model specification matching the long-run component in the cointegrated VAR used in [38]. Each equation models the pairwise relationship between the longer-term maturity and the short-term maturity (a one-to-one relationship under the EHT), while the constant accounts for a term premium. Additional maturities could be analysed, leading to additional equations in the model, but we try to maintain a simple model

TABLE 4 | Estimation of (multiple) structural breaks in the full model with shrinking regimes.

	Panel A: Group LASSO with BEA			
	SB1: $(\tau = 1 - 25/T)$			
T	Time (s)	pce	τ	
100	0.03	100	0.747 (0.019)	
200	0.07	99.5	0.873 (0.010)	
400	0.29	99.7	0.936 (0.006)	
800	1.69	99.8	0.967 (0.004)	
	SB2: $(\tau_1 = 0.5, \tau_2 = 0.5 + 25/T)$			
T	Time (s)	pce	τ_1	τ_2
150	0.09	84.2	0.505 (0.019)	0.661 (0.022)
300	0.26	91.9	0.500 (0.017)	0.584 (0.016)
600	1.59	94.1	0.499 (0.009)	0.542 (0.008)
	Panel B: Likelihood-based approach			
	SB1: $(\tau = 1 - 25/T)$			
T	Time (s)	pce	τ	
100	1.40	96.7	0.750 (0.006)	
200	3.65	97.3	0.854 (0.004)	
400	19.72	98.5	0.851 (0.005)	
800	103.17	98.7	0.850 (0.003)	
	SB2: $(\tau_1 = 0.5, \tau_2 = 0.5 + 25/T)$			
T	Time (s)	pce	τ_1	τ_2
150	3.75	94.3	0.505 (0.019)	0.661 (0.022)
300	14.88	95.5	0.463 (0.034)	0.615 (0.034)
600	72.04	96.3	0.443 (0.056)	0.596 (0.057)

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 subpanel reports the results for the two-step estimator with one active breakpoint at $\tau = 1 - 25/T$; the second subpanel reports the results for two active breakpoints at $\tau_1 = 0.5$, $\tau_2 = 0.5 + 25/T$. The likelihood-based approach is applied with a trimming parameter of 0.15. Time (s) denotes the average computing time for one replication in seconds. Standard deviations are given in parentheses. We conduct the $\sup(l+1|l)$ test at the 5% level to determine the number of breaks. Critical values are provided in [11].

structure. The data are produced according to the approach of [53] fitting a simple three-factor arbitrage-free term structure model to U.S. Treasury yields since 1990, in order to evaluate the behaviour of long-term yields, distant-horizon forward rates, and term premiums.[15] Figure 1 provides a time series plot of the data. We assume that the individual variables follow unit root processes.[16] The results of a Johansen trace test for the full sample suggest that the trivariate system is cointegrated with cointegration rank one or two depending on the specification of the deterministic terms. According to the long-run implications of the EHT, we would expect a cointegration rank of two but it is well-known that the Johansen trace tests are not robust to structural breaks in the cointegrating vectors or the deterministic terms see, for example, [54, 55].

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 EHT at the 5% significance level for the 10-year maturity but not for the 5-year maturity.[17] In the 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 months (50 daily observations) to obtain accurate coefficient estimates in each regime.[18] We consider a specification with a constant and apply the required scaling factors so that each regressor has the same order.[19] Using the two-step group LASSO estimator, we obtain four structural breaks. All breakpoint estimates can be related to important monetary policy events. To show that, we depict the trajectory of the effective federal funds rate (EFFR) in Figure 2 and indicate the regimes. The first break is located in November 1994 after the EFFR sharply increases and the yield spread narrows, the second break is located in April 2003 during the recession. Here, the EFFR falls dramatically and we observe wider spreads. The third break is located in August 2010 after the Global Financial Crisis, and the fourth break is located in March 2015 at the end of the zero target rate regime. After the structural breaks are obtained, we re-estimate the model

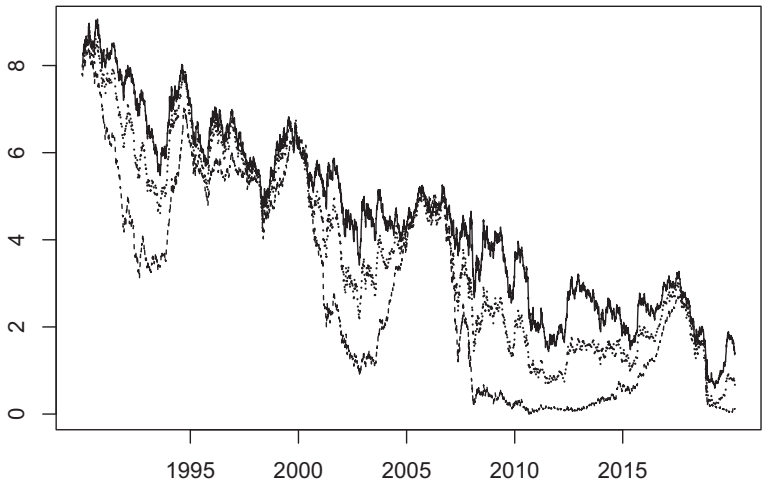

FIGURE 1 | Fitted yields (in %) on 10-year (solid), 5-year (dotted, and 1-year (dashed) zero coupon US bonds.

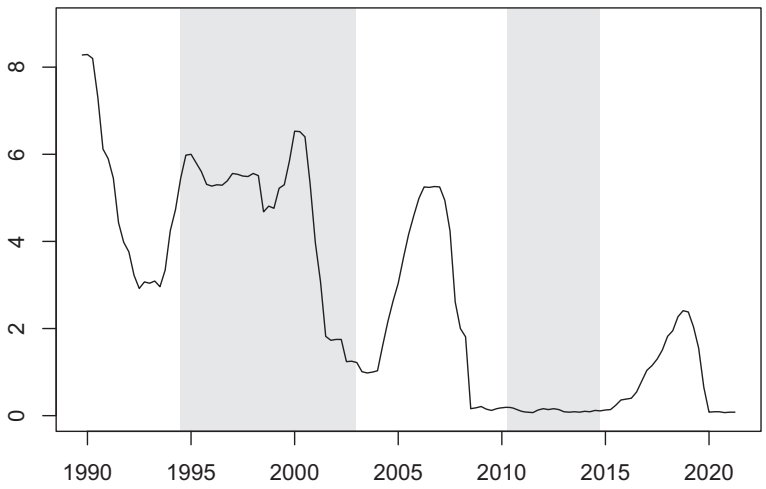

FIGURE 2 | Effective federal funds rate at a quarterly frequency. White and grey-shaded spaces mark the identified term structure regimes.

TABLE 5 | Regime-specific coefficients of the term structure model.

	$r_{10y,t}$		$r_{5y,t}$	
Regimes	$\hat{\mu}_1$	$\hat{\beta}_1$	$\hat{\mu}_2$	$\hat{\beta}_2$
1990 m01–1994 m11	5.013 (0.261)	0.464 (0.082)	3.141 (0.154)	0.678 (0.055)
1994 m11–2003 m04	4.000 (0.211)	0.392 (0.125)	2.297 (0.163)	0.653 (0.093)
2003 m04–2010 m08	3.569 (0.329)	0.267 (0.109)	2.020 (0.220)	0.556 (0.077)
2010 m08–2015 m03	2.311 (0.259)	−0.158 (0.812)	1.103 (0.162)	0.736 (0.532)
2015 m03–2021 m07	1.220 (0.235)	0.669 (0.168)	0.525 (0.191)	0.842 (0.131)

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 [56]. Standard errors are computed for each identified regime separately and therefore do not take the estimation uncertainty regarding the structural break detection into account. Standard errors are given in parentheses.

in each regime without scaling factors for higher precision and report the resulting coefficients in Table 5. For comparison, the estimated breakpoints from the likelihood-based approach are located in January 1995, December 2003, July 2008 and July 2013, corresponding to roughly the same monetary policy events. The algorithm also detects four breakpoints with the sequential test, clearly rejecting the hypothesis of three breaks in favour of a four-break model. More than four breaks cannot be allocated if the default minimum regime length should be maintained (setting the trimming parameter to 0.15).

It can be observed that the pairwise cointegrating vectors for most regimes substantially differ from $(1, -1)$. A simple t-test of the hypothesis does not lead to a rejection of the EHT for the last two regimes.[20] While the estimates of the proportionality coefficient for the first three regimes in case of the 10-year maturity are reasonable (albeit still being able to reject the EHT), the estimated coefficient for the fourth regime from August 2010 to March 2015 is negative with a much larger standard error. The non-rejection of the EHT in this regime can therefore be attributed to the higher error term variance in this period. This regime can also be associated with a zero target rate and an unusually steep yield curve which might explain the unusually small proportionality coefficient for the 10-year maturity. The final 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 regime comes closest to satisfying the EHT. 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 EHT does not hold for most of the sampling period.

5 | Conclusion

We have proposed a computationally efficient alternative to the existing likelihood-based approach to 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 and their timing and to jointly estimate the coefficients for each regime. This can be achieved without the need to conduct sequential tests. 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 [29]. It remains to be investigated if such confidence bands can be constructed within the model selection approach taken in this paper.

The crucial first step estimation is based on the group LASSO estimator and we utilise 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, [10] 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 also be applied to change-point problems in other important model classes. If we relaxed the strict exogeneity assumptions and instead used 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.

Acknowledgements

The author thanks Konstantin Kuck, Thomas Dimpfl, Markus Mößler, Domenic Franjic, Robert Jung, the editor, Anindya Banerjee, two anonymous referees, and participants of the Research Seminar in Economics at the University of Hohenheim, the Asian Meeting of the Econometric Society in China 2022, the Econometric Society 2022 Australasia Meeting, the AMES 2022 Tokyo, and the EEA-ESEM in Milan 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. Open Access funding enabled and organized by Projekt DEAL.

Data Availability Statement

Data can be obtained from the author upon reasonable request.

Endnotes

[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 [10] who use a slightly different penalty to cover a high-dimensional version of this case.

[2] [29] propose confidence sets for the timing of structural break estimation in multiple equation regression models [1] introduce 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 [2] highlight that the estimation of common breaks allows for more precise detection of break dates in multivariate systems. They develop common break tests for this assumption. A common break is defined as a point in time at which at least one coefficient from each equation is not restricted to be the same across two adjacent segments.

[3] 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 vs. 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.

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

[5] The simulation results presented in Table S6 in Supporting Information A show that the structural break detection remains consistent for endogenous regressors and white noise errors.

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

[7] 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.

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

[9] Our definitions in Assumption 3 (i) include the baseline coefficients which can, in principle, be relaxed but simplify the technical analysis.

[10] Supporting Information B can be found here: https://karstenschweikert.github.io/mequ_ci/mequ_ci_suppB.pdf.

[11] Supporting Information A can be found here: https://karstenschweikert.github.io/mequ_ci/mequ_ci_suppA.pdf.

[12] All simulations experiments are conducted on a computer with an Intel i5-6500 CPU at 3.20 GHz and 16 GB 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.

[14] The results can be found in Tables S1–S3 in Supporting Information A.

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

[16] See the discussion in [38] and the references given therein why this assumption is useful for the empirical modelling of the term structure although some features 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.

[17] Bootstrap standard errors are computed based on 600 replications of the sieve bootstrap method for cointegrating regressions proposed in [56].

[18] The results are robust for different choices of M as long as $M > 4$. Since a larger value of M allows the modified group LARS algorithm to take more steps, a larger value of M can, in principle, positively affect the precision of the estimated break locations. However, the estimates do not change for $M > 40$.

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

[20] Note that the reported standard errors do not take the break estimation uncertainty into account.

References

1. Y. Li and P. Perron, "Inference on Locally Ordered Breaks in Multiple Regressions," *Econometric Reviews* 36 (2017): 289–353, https://doi.org/10.1080/07474938.2015.1114552.

2. T. Oka and P. Perron, "Testing for Common Breaks in a Multiple Equations System," *Journal of Econometrics* 204 (2018): 66–85, https://doi.org/10.1016/j.jeconom.2018.01.003.

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

4. J. G. MacKinnon, "Using Large Samples in Econometrics," *Journal of Econometrics* 235 (2023): 922–926, https://doi.org/10.1016/j.jeconom.2022.05.005.

5. J. Bai and P. Perron, "Estimating and Testing Linear Models With Multiple Structural Changes," *Econometrica* 66 (1998): 47–78, https://doi.org/10.2307/2998540.

6. J. Bai and P. Perron, "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics* 18 (2003): 1–22, https://doi.org/10.1002/jae.659.

7. A. Schmidt and K. Schweikert, "Multiple Structural Breaks in Cointegrating Regressions: A Model Selection Approach," *Studies in Nonlinear Dynamics and Econometrics* 26 (2022): 219–254, https://doi.org/10.1515/snde-2020-0063.

8. K. Schweikert, "Oracle Efficient Estimation of Structural Breaks in Cointegrating Regressions," *Journal of Time Series Analysis* 43 (2022): 83–104, https://doi.org/10.1111/jtsa.12593.

9. W. Gao, H. Yang, and L. Yang, "Change Points Detection and Parameter Estimation for Multivariate Time Series," *Soft Computing* 24 (2020): 6395–6407, https://doi.org/10.1007/s00500-019-04135-8.

10. A. Safikhani and A. Shojaie, "Joint Structural Break Detection and Parameter Estimation in High-Dimensional Nonstationary VAR Models," *Journal of the American Statistical Association* 117 (2022): 251–264, https://doi.org/10.1080/01621459.2020.1770097.

11. Z. Qu and P. Perron, "Estimating and Testing Structural Change in Multivariate Regressions," *Econometrica* 75 (2007): 459–502, https://doi.org/10.1111/j.1468-0262.2006.00754.x.

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

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

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

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

16. P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous Analysis of Lasso and Dantzig Selector," *Annals of Statistics* 37 (2009): 1705–1732, https://doi.org/10.1214/08-AOS620.

17. P. Zhao and B. Yu, "On Model Selection Consistency of Lasso," *Journal of Machine Learning Research* 7 (2006): 2541–2563, https://doi.org/10.1109/TIT.2006.883611.

18. G. Ciuperca, "Model Selection by LASSO Methods in a Change-Point Model," *Statistical Papers* 55 (2014): 349–374, https://doi.org/10.1007/s00362-012-0482-x.

19. B. Jin, Y. Wu, and X. Shi, "Consistent Two-Stage Multiple Change-Point Detection in Linear Models," *Canadian Journal of Statistics* 44 (2016): 161–179, https://doi.org/10.1002/cjs.11228.

20. J. Qian and L. Su, "Shrinkage Estimation of Regression Models With Multiple Structural Changes," *Econometric Theory* 32 (2016): 1376–1433, https://doi.org/10.1017/S0266466615000237.

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

22. D. F. Hendry, S. Johansen, and C. Santos, "Automatic Selection of Indicators in a Fully Saturated Regression," *Computational Statistics* 23 (2008): 337–339, https://doi.org/10.1007/s00180-008-0112-1.

23. J. L. Castle, J. A. Doornik, and D. F. Hendry, "Model Selection When There Are Multiple Breaks," *Journal of Econometrics* 169 (2012): 239–246, https://doi.org/10.1016/j.jeconom.2012.01.026.

24. F. Pretis, J. J. Reade, and G. Sucarrat, "Automated General-To-Specific (Gets) Regression Modeling and indicator Saturation for Outliers and Structural Breaks," *Journal of Statistical Software* 86 (2018): 1–44, https://doi.org/10.18637/jss.v086.i03.

25. F. Pretis, "Econometric Modelling of Climate Systems: The Equivalence of Energy Balance Models and Cointegrated Vector Autoregressions," *Journal of Econometrics* 214 (2020): 256–273, https://doi.org/10.1016/j.jeconom.2019.05.013.

26. C. E. Quintos, "Sustainability of the Deficit Process With Structural Shifts," *Journal of Business and Economic Statistics* 13 (1995): 409–417, https://doi.org/10.1080/07350015.1995.10524615.

27. C. E. Quintos, "Stability Tests in Error Correction Models," *Journal of Econometrics* 82 (1997): 289–315, https://doi.org/10.1016/S0304-4076(97)00059-6.

28. B. Seo, "Tests for Structural Change in Cointegrated Systems," *Econometric Theory* 14 (1998): 222–259, https://doi.org/10.1017/S0266466698142044.

29. Y. Eo and J. Morley, "Likelihood-Ratio-Based Confidence Sets for the Timing of Structural Breaks," *Quantitative Economics* 6 (2015): 463–497, https://doi.org/10.3982/qe186.

30. P. C. B. Phillips and V. Solo, "Asymptotics for Linear Processes," *Annals of Statistics* 20 (1992): 971–1001, https://doi.org/10.1214/aos/1176348666.

31. P. C. B. Phillips, "Fully Modified Least Squares and Vector Autoregression," *Econometrica* 63 (1995): 1023–1078, https://doi.org/10.2307/2171721.

32. P. Saikkonen, "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory* 7 (1991): 1–21, https://doi.org/10.1017/S0266466600004217.

33. P. C. B. Phillips and M. Loretan, "Estimating Long-Run Economic Equilibria," *Review of Economic Studies* 58 (1991): 407–436, https://doi.org/10.2307/2298004.

34. J. H. Stock and M. W. Watson, "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica* 61 (1993): 783–820, https://doi.org/10.2307/2951763.

35. M. Yuan and Y. Lin, "Model Selection and Estimation in Regression With Grouped Variables," *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 68 (2006): 49–67, https://doi.org/10.1111/j.1467-9868.2005.00532.x.

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

37. C. Y. Yau and T. S. Hui, "LARS-Type Algorithm for Group Lasso," *Statistics and Computing* 27 (2017): 1041–1048, https://doi.org/10.1007/s11222-016-9669-7.

38. P. R. Hansen, "Structural Changes in the Cointegrated Vector Autoregressive Model," *Journal of Econometrics* 114 (2003): 261–295, https://doi.org/10.1016/S0304-4076(03)00085-X.

39. T. Boot and A. Pick, "Does Modeling a Structural Break Improve Forecast Accuracy?," *Journal of Econometrics* 215, no. 1 (2020): 35–59, https://doi.org/10.1016/j.jeconom.2019.07.007.

40. L. Boysen, A. Kempe, V. Liebscher, A. Munk, and O. Wittich, "Consistencies and Rates of Convergence of Jump-Penalized Least Squares Estimators," *Annals of Statistics* 37 (2009): 157–183, https://doi.org/10.1214/07-AOS558.

41. J. Bai, "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics* 79 (1997): 551–563, https://doi.org/10.1162/003465397557132.

42. J. Bai, "Vector Autoregressive Models With Structural Changes in Regression Coefficients and in Variance-Covariance Matrices," *Annals of Economics and Finance* 1 (2000): 303–339.

43. P. Breheny and J. Huang, "Group Descent Algorithms for Nonconvex Penalized Linear and Logistic Regression Models With Grouped Predictors," *Statistics and Computing* 25 (2015): 173–187, https://doi.org/10.1007/s11222-013-9424-2.

44. Y. S. Niu, N. Hao, and H. Zhang, "Multiple Change-Point Detection: A Selective Overview," *Statistical Science* 31 (2015): 611–623, https://doi.org/10.1214/16-STS587.

45. J. Y. Campbell and R. J. Shiller, "Cointegration and Tests of Present Value Models," *Journal of Political Economy* 95 (1987): 1062–1088, https://doi.org/10.1086/261502.

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

47. J. Y. Campbell and R. J. Shiller, "Yield Spreads and Interest Rate Movements: A bird's Eye View," *Review of Economic Studies* 58 (1991): 495–514, https://doi.org/10.2307/2298008.

48. G. Bulkley and P. Giordani, "Structural Breaks, Parameter Uncertainty, and Term Structure Puzzles," *Journal of Financial Economics* 102 (2011): 222–232, https://doi.org/10.1016/j.jfineco.2011.05.009.

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

50. L. Sarno, D. L. Thornton, and G. Valente, "The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields," *Journal of Financial and Quantitative Analysis* 42 (2007): 81–100, https://doi.org/10.1017/s0022109000002192.

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

52. P. Tillmann, "Inflation Regimes in the US Term Structure of Interest Rates," *Economic Modelling* 24 (2007): 203–223, https://doi.org/10.1016/j.econmod.2006.07.001.

53. D. H. Kim and J. H. Wright, "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* (2005): 1–21, https://doi.org/10.17016/feds.2005.33.

54. H. Lütkepohl, P. Saikkonen, and C. Trenkler, "Testing for the Cointegration Rank of a VAR Process With Level Shift at Unknown Time," *Econometrica* 72 (2004): 647–662, https://doi.org/10.1111/j.1468-0262.2004.00505.x.

55. P. Saikkonen, H. Lütkepohl, and C. Trenkler, "Break Date Estimation for VAR Processes With Level Shift With an Application to Cointegration Testing," *Econometric Theory* 22 (2006): 15–68, https://doi.org/10.1017/S0266466606060026.

56. Y. Chang, J. Y. Park, and K. Song, "Bootstrapping Cointegrating Regressions," *Journal of Econometrics* 133 (2006): 703–739, https://doi.org/10.1016/j.jeconom.2005.06.011.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.