Multiple Structural Breaks in Cointegrating Regressions: A Model Selection Approach

Alexander Schmidt* Karsten Schweikert†

August 16, 2018

Abstract

In this paper, we propose a new comprehensive treatment of structural change in cointegrating regressions. First, we consider a setting with fixed breakpoint candidates and show that a modified adaptive lasso estimator can consistently estimate structural breaks in the intercept and slope coefficient of a cointegrating regression. Second, we extend our approach to a diverging amount of breakpoint candidates and provide simulation evidence that timing and magnitude of structural breaks are estimated consistently. Third, we use the adaptive lasso estimation to design new tests for cointegration in the presence of multiple structural breaks, derive the asymptotic distribution of our test statistics and show that the proposed tests have power against the null of no cointegration. Finally, we use our new methodology to study the effects of structural breaks on the long-run PPP relationship.

Keywords: Cointegration, Structural breaks, Penalized estimation, Adaptive lasso, Purchasing power parity
JEL Classification: C12, C22, C52
MSC Classification: 62E20, 62J07, 91B84

*Address: University of Hohenheim, Department of Econometrics and Statistics, Schloss Hohenheim 1 C, 70593 Stuttgart, Germany, telephone: (0711) 459-24712, e-mail: *schmidt.alexander@uni-hohenheim.de*

†Address: University of Hohenheim, Core Facility Hohenheim, Schloss Hohenheim 1 C, 70593 Stuttgart, Germany, telephone: (0711) 459-24713, e-mail: *karsten.schweikert@uni-hohenheim.de*

1 Introduction

Reliably detecting structural change in multivariate time series models has increasingly gained importance over the last decade. A diverse literature has emerged which is concerned with estimation and testing of unknown structural breaks (see, for example, Perron (2006), Qu and Perron (2007), Kejriwal and Perron (2010), Aue and Horváth (2013), Perron and Yamamoto (2016), Qian and Su (2016) and Kurozumi and Skrobotov (2017)). It is well-known that the coefficient estimates of time series regressions are potentially biased if structural breaks are not accounted for. Further, statistical inference in these situations is unreliable as the size and power properties of statistical tests are distorted. This holds particularly for cointegration models in the spirit of Engle and Granger (1987), for which a long-run equilibrium relationship is estimated under the assumption of parameter constancy. Ignoring break dates at which the cointegrated system attains a new equilibrium might severely confound the cointegration analysis. Thus, accounting for unknown structural breaks in cointegration models and consistently estimating them if they occur during the sample period is of utmost importance for applied economic research.

For this matter, we propose a new approach to detect structural breaks in a potentially cointegrated regression using penalized estimation techniques. Based on this approach, we develop residual-based tests for cointegration which are valid in the presence of multiple structural breaks.

A structural break in a cointegrating regression, as it will be dealt with in this paper, occurs when either the intercept or the slope coefficient (or both) change substantially at one point in time. Early on, Gregory and Hansen (1996) have recognized the need for cointegration tests which account for the presence of structural breaks. They allow the cointegrating vector to change at an unknown point in time during the sample period and use a grid search approach to determine the break date. Their test performs well in situations, where commonly applied residual-based tests fail to detect a cointegration relationship (Gregory et al., 1996). However, their test is limited to only one such change in the long-run equilibrium and performs poorly in the presence of more than one break. Also, they do not consider whether the magnitude of the parameter changes is consistently estimated. Hatemi-J (2008) developed a test for similar models with two structural breaks. The problem with grid search algorithms for structural breaks lies in the exponential increase in computing time with an increasing amount of breaks and the crucial assumption that the exact amount of breaks has to be known a priori.

Bai and Perron (1998) proposed a procedure which allows to detect multiple structural breaks sequentially. This approach was further developed in Kejriwal and Perron (2010) and Maki (2012), adapting it to cointegrating regressions. In contrast to grid search algorithms, their sequential approach does not require the user to know the exact number of breaks, but instead only a maximum number of breaks needs to be specified. After estimating a baseline model without structural breaks, each added breakpoint is tested as to whether it improves the fit of the linear regression model. Westerlund and Edgerton (2006) design LM-based test statistics invariant to structural breaks to test the null of no cointegration and Davidson and Monticini (2010) use subsample procedures to account for structural breaks in their cointegration tests. Carrion-i Silvestre and Sanso (2006) and Arai and Kurozumi (2007) propose a CUSUM-based approach to test the null hypothesis of cointegration with a structural break against the alternative hypothesis of no cointegration.

One major disadvantage of most approaches to model structural change in the cointegration literature is that they solely focus on improving the cointegration test. However, if we plan to further analyze cointegrated data after having tested for cointegration, e.g. specifying error correction models, we are interested in obtaining unbiased cointegration residuals. This necessitates to find the exact break dates and to estimate the magnitude of the breaks consistently. Hence, with our new approach we pursue three objectives, namely (i) detecting multiple structural breaks in cointegrating regressions, (ii) consistently estimating the magnitude of those breaks, and (iii) testing for cointegration in the presence of multiple structural breaks.

We achieve these three objectives by perceiving the task of detecting and estimating structural breaks as a model selection problem. We could potentially shift and turn the regression hyperplane at every point in time using the appropriate indicator functions. Finding the true breakpoints corresponds to selecting the right indicators and eliminating irrelevant indicators. This leads to a high-dimensional setting with the total amount of parameters of the model close to the number of observations. The lasso estimator, introduced by Tibshirani (1996), in principle has attractive properties in these settings. However, quite restrictive regularity conditions are needed for simultaneous variable selection and consistent parameter estimation. Knight and Fu (2000) discuss the asymptotic behavior of the lasso estimator under different regularity conditions. They show that the lasso estimator is not consistent in general and might produce asymptotically biased estimates of the non-zero coefficients. Subsequent studies built on their results and propose slightly different extensions such as the fused lasso

(Tibshirani et al., 2005), the adaptive lasso (Zou, 2006) and grouped lasso (Yuan and Lin, 2006), among others. Particularly, the adaptive lasso is shown to have the oracle property under a broad set of assumptions which means it performs consistent variable selection and parameter estimation.

The use of lasso estimators in cointegrating regressions has been discussed in a few recent studies: To begin with, Mendes (2011) investigates the asymptotic properties of the adaptive lasso estimator in cointegrating regressions with additional stationary components. He shows that also in this context the adaptive lasso estimator has the oracle property under some regularity conditions. The adaptive lasso estimator has further been used by Liao and Phillips (2015) to simultaneously estimate the cointegrating rank and autoregressive order of a VECM. An extension to the conventional Johansen model in high-dimensional settings with a short sampling period can be found in Wilms and Croux (2016). Their sparse cointegration model using a lasso approach is shown to outperform the Johansen method in terms of forecasting precision if some elements of the cointegrating vectors are exactly zero. Koo et al. (2017) apply the lasso to predictive regressions involving highly persistent and potentially cointegrated time series. Their proposed lasso approach leads to superior forecasting performance relative to the OLS estimator based on the full model.

For our purposes, we rely on a modified version of the adaptive lasso procedure, similar to the one presented by Kock (2016), and proceed as follows. First, we operate in a setting with a fixed set of breakpoint candidates and provide a technical proof that our estimator has the oracle property in cointegrating regressions. While being quite restrictive, such a setting, nevertheless, appears to be of practical relevance. One could, for example, be confronted with a situation in which a fixed amount of crises occurred at well-known points during the sample period, all of which could potentially have influenced the cointegration relationship. An important question then would be which of these crises actually changed the long-run equilibrium, i.e. which of these crises led to breaks in either slope or intercept (or both) and which did not. In this context, we allow the breakpoints of intercept and slope coefficient to occur at different points in time (i.e. at different crises).

Second, we extend the procedure to more general situations where we do not have any prior information about potential breakpoint candidates but the breaks can occur at any point in time. This corresponds to a diverging number of parameters in the full model. Despite the increased complexity of the setting, we can provide simulation evidence that our procedure still estimates the breakpoints consistently. Lastly, we discuss how our modified adaptive lasso procedure can be used for residual-based tests for cointegration in the presence of multiple structural breaks.

The remainder of this paper is organized as follows. In Section 2, we describe the models considered, discuss the asymptotic properties of the adaptive lasso estimator and propose suitable cointegration tests under the presence of multiple structural breaks. Section 3 is devoted to the Monte Carlo simulation study and Section 4 presents an empirical application. Section 5 summarizes our results and states objectives for future research.

2 Methodology

To present the main idea, we restrict our analysis to a bivariate cointegration system with structural breaks in the intercept and slope coefficient. Possible extensions to multivariate cointegration systems are discussed in Section 5. Let $\{y_t\}_1^\infty$ denote a scalar process generated by

$$y_t = \mu_t + \beta_t x_t + u_t, \qquad t = 1, 2, \ldots, \tag{1}$$

where μ_t and β_t are time-varying coefficients and $\{x_t\}_1^\infty$ follows a random walk process

$$x_t = x_{t-1} + v_t, \qquad t = 1, 2, \ldots, \tag{2}$$

while $\{u_t\}_1^\infty$ and $\{v_t\}_1^\infty$ are mean-zero weakly stationary error processes. We make the following assumptions about the vector process $w_t = (u_t, v_t)'$:

Assumption 1. *The vector process $\{w_t\}_1^\infty$ satisfies the following conditions*

1. $Ew_t = 0$ *for* $t = 1, 2, \ldots$
2. $\{w_t\}_1^\infty$ *is weakly stationary.*
3. $\{w_t\}_1^\infty$ *is strong mixing with mixing coefficients of size* $-\delta\beta/(\delta - \beta)$ *and* $E|w_t|^\delta < \infty$ *for some* $\delta > \beta > 5/2$.

We further make some assumptions about the coefficients μ_t and β_t concerning the number of total changes in a given sample. We treat structural breaks as rare events and assume that parameter changes persist for some time. This assumption is easily justified by the intended application on economic long-run equilibrium relationships which, in order to be meaningful, have to hold over long periods of time. For true random coefficient models without such strict sparsity assumptions, we refer to Quintos and Phillips (1993), Kuo (1998), Park and Hahn (1999), Xiao (2009a), Xiao (2009b) and Bierens and Martins (2010), among others.

Assumption 2. *The total number of distinct values in any set* $\{\mu_1, \ldots, \mu_T\}$ *is* $p+1$, *where* $0 \leq p \leq p^* \ll T$ *and the total number of distinct values in any set* $\{\beta_1, \ldots, \beta_T\}$ *is* $m+1$, *where* $0 \leq m \leq m^* \ll T$.

We assume that the maximum number of breaks in the intercept, p^*, and slope, m^*, is known beforehand and thereby follow Bai and Perron (1998). The true number of p breaks in the intercept and m breaks in the cointegrating vector is unknown and can be determined from the data. For the moment and in contrast to Bai and Perron (1998), we allow that the intercept and the slope coefficient have a different number of breaks at different points in time. We denote the distinct coefficients in samples of length T as

$$\mu_t = \tilde{\mu}_i, \quad \text{for } t = T_{1,i-1}, T_{1,i-1}+1, \ldots, T_{1,i}-1, \quad i = 1, \ldots, p+1, \tag{3}$$

and

$$\beta_t = \tilde{\beta}_j, \quad \text{for } t = T_{2,j-1}, T_{2,j-1}+1, \ldots, T_{2,j}-1, \quad j = 1, \ldots, m+1, \tag{4}$$

where $T_{1,0} = T_{2,0} = 1$ and $T_{1,p+1} = T_{2,m+1} = T+1$. The relative timing of breakpoints is denoted by $\tau_{1,i} = T_{1,i}/T, i \in \{1, \ldots p\}$ and $\tau_{2,j} = T_{2,j}/T, j \in \{1, \ldots, m\}$, respectively. To study the consistency of our estimator, we need some additional assumptions about the magnitude of the breaks and the distance between them.

Assumption 3. *(i) The minimum break intervals are defined as* $I_{1,\min} = \min_{1 \leq i \leq p^*+1} |T_{1,i} - T_{1,i-1}|$ *and* $I_{2,\min} = \min_{1 \leq j \leq m^*+1} |T_{2,j} - T_{2,j-1}|$. *Both increase with the sample size as* $I_{1,\min} = I_{2,\min} = O(T)$.
(ii) The break magnitudes are bounded by $J_{1,\min} < |\tilde{\mu}_i - \tilde{\mu}_{i-1}| < J_{1,\max}$ *for* $2 \leq i \leq p+1$ *and* $J_{2,\min} < |\tilde{\beta}_j - \tilde{\beta}_{j-1}| < J_{2,\max}$ *for* $2 \leq j \leq m+1$. *The lower bounds are larger than* $O(T^{-1/2})$ *as* $T \to \infty$. *The upper bounds are constants so that* $J_{1,\max} = J_{2,\max} = O(1)$.

Assumption A3(i) requires that the length of the regimes between breaks increases with the sample size and in the same proportions to each other. This allows us to consistently detect and estimate the true break fractions as it makes the break dates asymptotically distinct (Perron, 2006). A3(ii) excludes the possibility of infinite shifts in the parameters and requires parameter changes to be of a substantial magnitude to distinguish active breaks from inactive breaks.

2.1 Fixed breakpoint candidates

First, we consider a special setting, where we have prior information about the timing of potential breakpoint candidates in our sample. Hence, the values of p^*, m^* and $(\tau_{1,1}, \ldots, \tau_{1,p^*}, \tau_{2,1}, \ldots, \tau_{2,m^*})$ are known. The total amount of coefficients in the full model, i.e. of baseline regressors and all breakpoint candidates, is then given by the fixed scalar $d^* = p^* + m^* + 2$. In this case, we can express the long-run equilibrium equation in a regime-specific form such that

$$y_t = \sum_{i=1}^{p^*+1} \mu_i^* \varphi_{t,\tau_{1,i-1}} + \sum_{j=1}^{m^*+1} \beta_j^* x_t \varphi_{t,\tau_{2,j-1}} + u_t, \tag{5}$$

where the indicator variable $\varphi_{t,\tau_{k,l}}$ is defined as

$$\varphi_{t,\tau_{k,l}} = \begin{cases} 0 & \text{if } t < [T\tau_{k,l}] \\ 1 & \text{if } t \geq [T\tau_{k,l}] \end{cases}, \qquad k \in \{1,2\}, t = 1, 2, \ldots, \tag{6}$$

and $\tau_{k,0} = 0$. The coefficients in regime-specific form are $\mu_1^* = \tilde{\mu}_1$, $\mu_i^* = \tilde{\mu}_i - \tilde{\mu}_{i-1}$ for $i = 2, \ldots, p^* + 1$ and $\beta_1^* = \tilde{\beta}_1$, $\beta_i^* = \tilde{\beta}_j - \tilde{\beta}_{j-1}$ for $j = 2, \ldots, m^* + 1$. Hence, μ_1^* and β_1^* denote the parameter values until the first breakpoint (baseline model), while μ_i^*, $i = 2, \ldots, p^* + 1$ and β_j^*, $j = 2, \ldots m^* + 1$ denote the parameter changes at all breakpoint candidates. We are primarily interested in a procedure to detect the true number of breakpoints and to consistently estimate the magnitude of the parameter change. Relevant breakpoints should be indicated by nonzero coefficients while irrelevant breakpoints should be eliminated from the model. For that matter, we estimate the cointegrating regression with potentially multiple breaks using an objective function which shrinks irrelevant breakpoints to zero.

A natural choice for such an objective function would be the lasso of Tibshirani (1996). It allows to select relevant coefficients, i.e. those corresponding to active breakpoints, and shrinks the coefficients of irrelevant coefficients, i.e. those corresponding to the other potential but non-active breakpoints, to zero. However, it is well-known that the least squares estimator of μ has convergence rate $\sqrt{T}$ while the least squares estimator of β is superconsistent at rate T. Since we want to recover breaks in both μ and β, we should not shrink both types of coefficients with the same tuning parameter. Further, it is well-known that the plain lasso estimator is not oracle efficient. One way to deal with these issues is to assign individually chosen weights to each coefficient, as in the adaptive lasso of Zou (2006). With these weights the coefficients will experience different degrees of shrinkage even though there is still only one global tuning parameter in the model. The objective function that we will use in the following is a variant of the adaptive lasso objective function and similar to the objective function used in Kock (2016) who investigates model selection in stationary and nonstationary autoregressions. He modifies the objective function such that a different exponent of the weights is added (either γ_1 or γ_2) which depends on the convergence rate of the least squares estimator. It allows us to shrink all elements of the sets $\{\mu_1, \ldots, \mu_T\}$ and $\{\beta_1, \ldots, \beta_T\}$ to zero where no structural change occurs. Subsequently, we detect structural breaks using the index of all nonzero coefficients left after optimization. The objective function can be written as

$$\begin{aligned} V_T(\{\mu_t, \beta_t\}) &= \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{p^*+1} \mu_i^* \varphi_{t,\tau_{1,i-1}} - \sum_{j=1}^{m^*+1} \beta_j^* x_t \varphi_{t,\tau_{2,j-1}} \right)^2 \\ &+ \lambda_T \sum_{i=2}^{p^*+1} w_{1i}^{\gamma_1} |\mu_i^*| + \lambda_T \sum_{j=2}^{m^*+1} w_{2j}^{\gamma_2} |\beta_j^*|, \qquad \gamma_1, \gamma_2 > 0, \end{aligned} \tag{7}$$

where $w_{1i} = 1/|\hat{\mu}_{I,i}^*|$ for $i = 2, \ldots, p^* + 1$ and $w_{2j} = 1/|\hat{\beta}_{I,j}^*|$ for $j = 2, \ldots, m^* + 1$ are coefficient-specific weights based on initial estimates of the coefficients. Note that we do not apply any shrinkage to the baseline model.

The value of the global tuning parameter, λ_T, is generally unknown and has to be estimated from the data. Cross-validation approaches are commonly used for this matter. However, since we later also consider cointegration tests where the null hypothesis is no cointegration, we cannot meaningfully apply these approaches. Instead, we follow Kock (2016) and select the tuning parameter by minimizing the BIC,

$$BIC(\lambda_T) = \log(\widehat{MSE}) + \log(T)/T \cdot df, \tag{8}$$

where df are the respective degrees of freedom of the model.

In the case of fixed breakpoint candidates, the initial estimates $\hat{\mu}_I^*$ and $\hat{\beta}_I^*$ can be obtained from least squares estimation of the long-run equilibrium equation with multiple structural breaks indicators. The least squares estimator is consistent and yields appropriate weights. However, if the total number of coefficients d^* exceeds the number of observations T, ordinary least squares estimation is not an option and alternatives, e.g. ridge regression or the dimension-reduction procedure outlined in Subsection 2.2, have to be considered.

In the following, we establish the first of our main results. We prove that the adaptive lasso estimator tuned to perform consistent model selection has the oracle property in bivariate cointegrating regressions with multiple structural beaks. We show that the adaptive lasso performs correct model selection which requires that the probability of including all truly nonzero coefficients in the model tends to one while the probability of keeping irrelevant variables tends to zero. This satisfies the first property an oracle procedure should possess. Further, the estimator should have an asymptotic normal distribution (Fan and Li, 2001). We show that our estimator has the same asymptotic distribution as the least squares estimator. However, since our regression involves nonstationary components, the asymptotic distribution of the least squares estimator is naturally given as a functional of Brownian motions. Hence, we say that our estimator satisfies a nonstandard oracle property to distinguish it from its stationary counterpart.

We use the following notation to present our main results: a vector of T observations for the variable y_t is denoted by $y = (y_1, \dots, y_T)'$. Similarly, we denote $x = (x_1, \dots, x_T)'$ and $u = (u_1, \dots, u_T)'$. Further, we define

$$\varphi_{\tau_{k,l}} = (\underbrace{0 \dots 0}_{T_{k,l}-1} \quad \underbrace{1 \dots 1}_{T-T_{k,l}+1}), \qquad k \in \{1, 2\},$$

to denote break indicators in vector form. We define the matrix $\boldsymbol{X} = (1, \varphi_{\tau_{1,1}}, \dots, \varphi_{\tau_{1,p^*}}, x, x\varphi_{\tau_{2,1}}, \dots, x\varphi_{\tau_{2,m^*}})$ with column rank d^* containing the baseline regressors and all potential break indicator variables. We decompose the matrix $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ into one part containing the constant and the intercept break indicators, $\boldsymbol{X}_1$, and one part containing the regressor x and the slope break indicators, $\boldsymbol{X}_2$. $\Sigma = E(\boldsymbol{X}_1 \boldsymbol{X}_1')$ is the covariance matrix of $\boldsymbol{X}_1$. We define the index set $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$, where $\mathcal{A}_1 = \{1 \leq i \leq p^* + 1 : \mu_i^* \neq 0\}$ denotes the set of active intercept break indicators and $\mathcal{A}_2 = \{1 \leq j \leq m^* + 1 : \beta_j^* \neq 0\}$ denotes the set of active slope break indicators. $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$ and $\mathcal{A}^c$ denotes the complementary set. $\Sigma_{\mathcal{A}_1}$ indexes the rows and columns of the covariance matrix belonging to the active variables. $B(s)$ denotes Brownian motion with variance σ^2 and $U(s)$ denotes the weak limit of u_t. For notational convenience we use '$\Rightarrow$' to signify weak convergence of the associated probability measures and $\xrightarrow{p}$ to denote convergence in probability. Continuous stochastic processes such as $B(s)$ on [0,1] are simply written as B if no confusion is caused. We also write integrals with respect to the Lebesgue measure such as $\int_0^1 B(s)ds$

simply as $\int_0^1 B$. We use $x\varphi_{\tau_{2,j}} = x \odot \varphi_{\tau_{2,j}}$ as short-hand notation for the Hadamard product involving indicator terms where no confusion arises.

Theorem 1. *Suppose that the scalar processes $\{y_t\}_1^\infty$ and $\{x_t\}_1^\infty$ are cointegrated as described by Equation (1), Assumptions A1, A2, and A3 hold, $\frac{\lambda_T}{\sqrt{T}} \to 0$, $\frac{\lambda_T}{T^{1/2-\gamma_1/2}} \to \infty$ and $\frac{\lambda_T}{T^{1-\gamma_2}} \to \infty$. Then, the adaptive lasso estimator has nonstandard oracle properties:*

(a)
$$P(\hat{\mu}_{T,\mathcal{A}_1^c} = 0) \to 1$$
$$P(\hat{\beta}_{T,\mathcal{A}_2^c} = 0) \to 1$$

(b)
$$\begin{bmatrix} \sqrt{T}(\hat{\mu}_{T,\mathcal{A}_1} - \mu_{\mathcal{A}_1}) \\ T(\hat{\beta}_{T,\mathcal{A}_2} - \beta_{\mathcal{A}_2}) \end{bmatrix} \Rightarrow \begin{bmatrix} \Sigma_{\mathcal{A}_1} & \mathbf{0} \\ \mathbf{0} & \int_0^1 B_{\tau,\mathcal{A}_2} B'_{\tau,\mathcal{A}_2} \end{bmatrix}^{-1} \times \begin{bmatrix} N(0, \Upsilon_{\mathcal{A}_1}\sigma^2) \\ \int_0^1 B_{\tau,\mathcal{A}_2} dU + C^*_{\mathcal{A}_2} \end{bmatrix}$$

$$C^* = [\Lambda, (1-\tau_{2,1})\Lambda, \ldots, (1-\tau_{2,m^*})\Lambda]', \qquad \Lambda = \lim_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} Ex_{t-1}u_t,$$

$$\Upsilon = \begin{bmatrix} 1 & \tau_{1,1} & \ldots & \tau_{1,p^*} \\ \tau_{1,1} & \tau_{1,1} & & \\ \vdots & & & \vdots \\ \tau_{1,p^*} & \ldots & & \tau_{1,p^*} \end{bmatrix}.$$

Statement (a) of the theorem establishes the convergence to zero of truly zero coefficients with probability approaching one. In statement (b), we derive the limit distribution for truly nonzero coefficients. It follows from statement (b) that truly nonzero coefficients are consistently estimated and converge with the same rate as the least squares estimators. As in Theorem 1 of Kock (2016), the exponents of the weights have to satisfy the restrictions $\gamma_1 > 0, \gamma_2 > 1/2$ to guarantee the nonstandard oracle property.

Remark 1. The proof of Theorem 1, which can be found in Subsection 1.1 of the Supplementary Material, is given under the assumption that initial least squares estimators are available. The idea of the proof is similar to the proof of Theorem 2 in Zou (2006) and Theorem 1 in Kock (2016). We extend some results to match our specific cointegration model involving multiple structural breaks.

Remark 2. It is shown in Theorem 2 of Kock (2016) that selecting the tuning parameter via the BIC results in consistent variable selection. His results for nonstationary autoregressions extend straightforwardly to cointegrating regressions under our assumptions.

Remark 3. The results presented in Theorem 1 describe the pointwise asymptotic distribution of our estimator. As Pötscher and Schneider (2009) show using more general asymptotic theory, the oracle property does not hold uniformly over the parameter space and the rate of convergence can be substantially slower than $(\sqrt{T}, T)'$.

2.2 Diverging amount of breakpoint candidates

If we begin our analysis without any prior information about the timing of the structural breaks, we could consider each $0 < t < T$ to be a potential breakpoint for both μ and β. This would result in a high-dimensional estimation problem with a diverging amount of parameters of the full model as $T \to \infty$. Zhang and Huang (2008) investigate the properties of the lasso estimator in a similar, high-dimensional setting under a general sparsity condition. They find that the lasso tends to include variables with large coefficients but also selects some irrelevant variables. Still, the lasso substantially reduces the dimensionality of the estimation problem and provides coefficient estimates which can be used to construct weights for the adaptive lasso method.

We follow Horowitz and Huang (2013) and use the adaptive lasso again to distinguish between active and non-active structural breaks when the number of potential breaks diverges with the sample size, $d_T^* = O(T)$.[1] Since we want to reduce the dimensionality of our estimation problem a priori as much as possible, we impose two restrictions: first, we apply some lateral trimming to exclude the possibility of selecting structural breaks at the beginning or the end of the sample. The degree of trimming is denoted with the parameter ξ. This is again motivated by our intended empirical applications, where each regime relates to a newly attained long-run equilibrium and should persist for some time. Second, we do not consider breaks in the intercept for our initial estimation. In a later estimation step, we relax the second restriction and instead require the intercept breaks to have the same timing as the slope breaks which is a common restriction in most studies on structural breaks in multiple regression models (Bai and Perron, 1998).

[1] The subscript T for the scalar d_T^* is added to emphasize the dependence of model complexity on the sample size.

A lateral trimming of 5% leaves us with $d_T^* = 0.9 \cdot T + 2$ parameters of the full model and we satisfy the condition $d_T^* < T$ when $T > 20$.

To further reduce the dimensionality of the problem and to obtain useful weights, we apply the plain lasso estimator to our full model,

$$y_t = \mu + \beta_1^* x_t + \sum_{j=2}^{T_\xi} \beta_j^* x_t \varphi_{t,\tau_{2,j-1}} + u_t, \tag{9}$$

where $T_\xi = [(1-2\xi)T]$ and $\tau_{2,j} \in (\xi, 1-\xi)$ on an equidistant grid. While the lasso estimator does not shrink all irrelevant breaks exactly to zero, we nevertheless obtain consistent estimates of all coefficients. Particularly, we do not eliminate any relevant variables asymptotically if the coefficients exceed a certain threshold. Since the detection of breaks becomes more complicated with an increasing number of breakpoint candidates, we slightly modify Assumption A3 and let the threshold value for active breaks depend on the parameter d_T^*:

Assumption 4. *The coefficients of active breaks have large coefficients such that* $|\beta_j^*| \gg \sqrt{\ln d_T^*/T}$ *for* $j \in \mathcal{A}_2$ *and non-active breaks have zero coefficients, hence* $|\beta_j^*| = 0$ *for* $j \in \mathcal{A}_2^c$.

When a variable is not selected by the first stage lasso estimation, we do not include the variable in the subsequent adaptive lasso stage. If the coefficient is nonzero, we take the reciprocal absolute value as its weight (w_j) and optimize the objective function

$$V_T(\{\mu, \beta_t\}) = \sum_{t=1}^{T} \left(y_t - \mu - \beta_1^* x_t - \sum_{j=2}^{T_\xi} \beta_j^* x_t \varphi_{t,\tau_{2,j-1}} \right)^2 + \lambda_T \sum_{j=2}^{T_\xi} w_j |\beta_j^*|,$$

where the exponent of the weights is unity (see Equation (7)). The tuning parameter is selected by minimizing the modified BIC proposed in Wang et al. (2009),

$$BIC^*(\lambda_T) = \log(\widehat{MSE}) + \log(T)/T \cdot df \cdot \log\log d_T^*. \tag{10}$$

This generalization of the BIC can still identify the true model consistently with a diverging number of parameters as long as $d_T^* < T$ holds. The second stage estimation might still indicate more structural breaks in the slope coefficient than assumed a priori. This could be caused by noisy parameter estimates which are close to zero but not exactly zero. We eliminate the additional irrelevant breaks by using a post-lasso

OLS estimation proposed by Belloni and Chernozhukov (2013) as our third and final estimation stage. We select only m^* break indicators corresponding to the m^*-th largest parameter changes for the OLS model. This also allows us to relax the assumption about a constant intercept which might be unrealistic in practice. At the third stage, we can easily add intercept break indicators with the same timing as the slope breaks obtained from the adaptive lasso estimation.[2] However, a crucial aspect for our procedure is the performance of the adaptive lasso estimator if the model is misspecified with respect to the intercept. Taking into consideration the different convergence rates of the least squares estimators and the much higher variation of the slope break indicators, we should be able to detect the slope and indicator breaks sequentially. We analyze this aspect again in our simulation experiments in Section 3.

2.3 Testing for cointegration

The previous sections have revealed that the (modified) adaptive lasso estimator can be used to detect multiple structural breaks in cointegrating regressions. These results hinge on Assumption A1 which specifies u_t as a stationary error term. In most practical applications, we do not know with certainty whether a particular set of non-stationary variables hold a long-run equilibrium relationship. Therefore, we consider residual-based tests for cointegration which allow for the possibility of multiple structural breaks. The regression in Equation (1) becomes spurious under the null hypothesis of no cointegration. In this case, the error term u_t is a cumulative sum of innovations and hence integrated of order one. If we apply our adaptive lasso estimator in such situations, we not only obtain meaningless least squares components but also have to deal with penalization terms applied to non-existing structural breaks. However, the degree of shrinkage naturally depends on the value of the tuning parameter λ_T. Further, we still assume that we know the maximum number of breaks if the variables were indeed cointegrated and thereby limit the number of location shifts of the error term in situations when the variables are not cointegrated.

We conduct our cointegration test in three steps. First, we apply the adaptive lasso estimator to a potentially cointegrating regression with pre-specified maximum number of slope breaks m^*. Without prior knowledge of any break date, we begin to

[2] As Belloni and Chernozhukov (2013) show, the post-lasso estimation only performs well if all components of the true model are included as a subset of the selected model and the selected model is sufficiently sparse. To avoid multicollinearity in our design matrix, we treat estimated breakpoints with adjacent break dates as a single breakpoint in the post-lasso estimation.

shrink the number of breakpoints from T_ξ and select the m^*-th largest breaks. Second, we re-estimate the long-run equilibrium equation with the selected slope breaks and corresponding intercept breaks using post-lasso OLS. Finally, we test the residuals for stationarity using ADF-type and bias-corrected test statistics (Phillips, 1987).

We make the following assumptions to present the asymptotic distributions of our test statistics. The observed data $S_t = (y_t, x_t)$ is generated as a random walk under the null hypothesis with $\Delta S_t = u_t$, where u_t is a vector process similar to the one specified in Assumption A1.

Assumption 5. *The vector process $\{u_t\}_1^\infty$ satisfies the following conditions*

1. $Eu_t = 0$ *for* $t = 1, 2, \ldots$
2. $\{u_t\}_1^\infty$ *is weakly stationary.*
3. $\{u_t\}_1^\infty$ *is strong mixing with mixing coefficients of size* $-\delta\beta/(\delta-\beta)$ *and* $E|w_t|^\delta < \infty$ *for some* $\delta > \beta > 5/2$.
4. *The long-run variance of* $S_t = \sum_{j=1}^{t} u_j$,

$$\Omega = \lim_{T\to\infty} T^{-1} E S_t S_t',$$

is positive definite.

We consider the following auxiliary regression for the post-lasso cointegrating residuals, $\hat{e}_{t\tau}$, which depend on the breakpoint selection indicated by the subscript τ,

$$\hat{e}_{t\tau} = \rho_\tau \hat{e}_{t-1\tau} + v_{t\tau}, \tag{11}$$

and estimate the bias-corrected first-order serial correlation coefficient as suggested in Phillips (1987),

$$\hat{\rho}^*_\tau = \sum_{t=2}^{T} (\hat{e}_{t\tau}\hat{e}_{t-1\tau} - \hat{\psi}_\tau) \Big/ \sum_{t=2}^{T} \hat{e}^2_{t-1\tau}, \tag{12}$$

where the bias-correction term, $\hat{\psi}_\tau$, is an estimate of the weighted sum of autocovariances,

$$\hat{\psi}_\tau = \sum_{j=1}^{M} w(j/M)\hat{\gamma}_\tau(j), \qquad \hat{\gamma}_\tau(j) = \frac{1}{T}\sum_{t=j+1}^{T} \hat{v}_{t-j\tau}\hat{v}_{t\tau}. \tag{13}$$

The kernel weights $w(\cdot)$ satisfy standard conditions and the bandwidth is a function of the sample size so that $M \to \infty$ and $M/T^5 = O(1)$. The estimate of the long-run variance of $\hat{v}_{t\tau}$ is then given by

$$\hat{\sigma}_{\tau}^2 = \hat{\gamma}_{\tau}(0) + 2\hat{\psi}_{\tau}. \tag{14}$$

We employ the standardized bias-corrected test statistic to evaluate the null hypothesis of no cointegration. The test statistic is given by

$$Z_{\tau} = (\hat{\rho}_{\tau}^{*} - 1)/\hat{s}_{\tau}, \qquad \hat{s}_{\tau}^2 = \hat{\sigma}_{\tau}^2 \Big/ \sum_{t=2}^{T} \hat{e}_{t-1\tau}^2. \tag{15}$$

Alternatively, we regress $\Delta\hat{e}_{t\tau}$ upon $\hat{e}_{t-1\tau}$ and K lagged differences $\Delta\hat{e}_{t-1\tau}, \ldots, \Delta\hat{e}_{t-K\tau}$. In practice, we use order selection rules such as AIC, BIC or a general-to-specific pretesting procedure to determine the lag truncation parameter. We follow Chang and Park (2002) and require that K increases with the sample size.

Assumption 6. *K increases with T in such a way that $K = o(T^{1/2})$.*

The ADF test statistic is the t-ratio for the regressor $\hat{e}_{t-1\tau}$. We express the asymptotic distribution of our primary test statistic, Z_{τ}, as functionals of Brownian motion.

Theorem 2. *If the scalar processes $\{y_t\}_1^{\infty}$ and $\{x_t\}_1^{\infty}$ are generated under the null hypothesis of no cointegration and Assumption A5 holds, then*

$$Z_{\tau(m^*,\lambda)} \Rightarrow \pi_0 Z_0 + \pi_1 Z_1 + \cdots + \pi_{m^*} Z_{m^*},$$

where

$$Z_i \sim \int_0^1 W_{\tau_i} dW_{\tau_i} \Big/ \left(\int_0^1 W_{\tau_i}^2 \right)^{1/2} \left(1 + \kappa_{\tau_i}' D_{\tau_i} \kappa_{\tau_i}\right)^{1/2},$$

$$W_{\tau_i} = W_1(s) - \int_0^1 W_1 W_{2\tau_i} \left[\int_0^1 W_{2\tau_i} W_{2\tau_i}' \right]^{-1} W_{2\tau_i}(s),$$

$$\kappa_{\tau_i} = \left[\int_0^1 W_{2\tau_i} W_{2\tau_i}' \right]^{-1} \int_0^1 W_{2\tau_i} W_1,$$

$$W_{2\tau_i} = \left[1, \varphi_{\tau_{2,1}}(s), \ldots, \varphi_{\tau_{2,i}}(s), W_2(s), W_2(s)\varphi_{\tau_{2,1}}(s), \ldots, W_2(s)\varphi_{\tau_{2,i}}(s)\right],$$

D_{τ_i} is a quadratic matrix with rank equal to the column rank of $W_{2\tau_i}$ and the vector of probabilities $\pi = [\pi_0, \pi_1, \ldots, \pi_{m^*}]'$ is a function of the maximum number of breaks specified and the tuning parameter $\lambda \in (0, \infty)$.

Remark 4. The distribution of our test statistic is a weighted average over all possible model selection outcomes in the post-lasso estimation. For a maximum amount of m^* breaks in the slope coefficient, we have m^*+1 models which can be selected by the lasso estimation procedure depending on the value of the tuning parameter. The timing of each breakpoint is assumed to be unknown a priori and is determined by the estimation procedure.

Following Phillips and Ouliaris (1990) and Gregory and Hansen (1996), we can assume that the asymptotic distributions of the ADF-type and bias-corrected test statistics are the same.[3]

3 Simulation results

Monte Carlo experiments are utilized to evaluate the finite sample performance of the (modified) adaptive lasso estimator. First, we consider cointegrated systems with several structural break specifications to investigate the theoretical claims developed in Subsection 2.1. Particularly, we want to find out whether the timing of the breaks is accurately indicated and whether the estimated change in the parameters is becoming more accurate if we increase the sample size successively. Second, we evaluate the performance of our estimator in models with a diverging amount of breakpoint candidates described in Subsection 2.2. Here, we are primarily interested in the precision of our estimator when breaks occur in the intercept and slope coefficients simulaneously but only the slope coefficients are specified correctly. Third, we study the size and power of our residual-based tests proposed in Subsection 2.3 for different configurations of the DGP and finite sample sizes. Approximate critical values of the Z_τ and ADF test statistics are reported in Table 1.

The following DGP is employed to model a bivariate cointegrated system with mul-

[3]Since in the proof of Theorem 2, which can be found in Subsection 1.2 of the Supplementary Material, we only refer to results for a fixed number of break indicator regressors, we follow Gregory and Hansen (1996) and expect Theorem 4.2 of Phillips and Ouliaris (1990) to hold. Our simulation results in Section 3 seem to support this claim.

tiple structural breaks,

$$\begin{aligned} y_t &= \mu_t + \beta_t x_t + e_t && \\ x_t &= x_{t-1} + \omega_t && \omega_t \sim N(0, \sigma^2_\omega) \\ \Delta e_t &= \rho\, e_{t-1} + \vartheta_t && \vartheta_t \sim N(0, \sigma^2_\vartheta). \end{aligned} \tag{16}$$

Using this framework, we study the performance of our estimation procedure and residual-based tests under different breakpoint specifications. In order to evaluate the performance of the modified adaptive lasso estimator in a fixed breakpoint setting, we consider seven potential breakpoint candidates located at the break fractions $\tau = (0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875)$. The first specification has one active break at 0.5 where the baseline coefficients are $\mu_1^* = 2$ and $\beta_1^* = 2$ before the break and $\mu_2^* = 4$ and $\beta_2^* = 4$ after the break. The second specification has a first break of both coefficients at 0.25 and a second break at 0.75. The final specification involves three breaks at 0.25, 0.5 and 0.75. The parameter values are increased by 2 at each breakpoint. If the adaptive lasso performs as described by Theorem 1, we should see that the estimates of truly zero coefficients tend to zero very quickly. In our setting, all indicator terms except for the true breakpoints should have zero coefficients. Further, we should observe different convergence rates for the truly nonzero intercept and slope coefficients, where the convergence rate for the intercept coefficients should be slower.

The results for increasing sample sizes are summarized in Table 2. To begin with, we observe quite accurate estimates of the breakpoints for the slope coefficients. The coefficient estimates approach the true parameter values with increasing sample size, while the variance for truly zero coefficients is already very small at sample sizes of $T = 800$. It appears more difficult to obtain precise estimates of breaks in the intercept than for the slope coefficients. The estimates for nonzero intercept changes on average underestimate the true change which can be attributed to a non-detection of true breakpoints in some replications. Also, they have variances which are some magnitudes larger than the variances of slope changes. Since parameter changes are estimated from one regime to the next, it appears that estimation errors made in previous regimes accumulate and influence the values of parameter changes at later break dates. Consequently, the variance of parameter changes at later break dates is generally larger than the variance at earlier break dates. Thus, these results mostly fulfill our expectations with respect to the different convergence rates of intercept and slope coefficients.

The results for a diverging amount of breakpoint candidates are reported in Table 3.

We consider 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 m and measure the accuracy of the break date estimation conditional on the correct estimation of m. For this matter, we compute the average Hausdorff distance[4] and divide it by T (hd/T) to compare the values across different sample sizes. We find that the number of breaks is detected with increasing precision and the distance between estimated breakpoints and true breakpoints is becoming smaller for increasing sample sizes. The estimated break fractions are obtained from the second estimation stage involving a misspecified intercept which is falsely assumed to be constant over the sample period. However, the estimates are already very accurate at small sample sizes. Using these second stage results, we are able to specify the post-lasso estimation and obtain consistent estimates for the intercept and slope changes. As expected, the parameter changes of models with fewer breakpoints can be estimated more precisely than those of models with a higher amount of breakpoints, as indicated by larger variances of the latter models at all sample sizes.

Turning to the cointegration tests next, we compare the performance of our test against the performance of well-known cointegration tests with structural breaks described in Gregory and Hansen (1996), Hatemi-J (2008) and Maki (2012), henceforth GH-test, HJ-test and Maki-test. Further, we use the benchmark cointegration test for constant coefficients by Engle and Granger (1987), henceforth EG-test, to evaluate the performance for a DGP without any structural breaks. The results for $T = 100$ and $T = 200$ are reported in Table 4 and Table 5, respectively. For the power simulations, we do not set a fixed amount of breakpoint candidates in our adaptive lasso framework, but let every observation along T, excluding the lateral trimming, be a potential breakpoint. The Maki-test with one break at most (M1) is conceptually identical to the GH-test. The only reason to explain the small differences observed in the results is the trimming parameter which is $\xi = 0.05$ for the Maki-test and $\xi = 0.15$ for the GH-test.

Our proposed cointegration test based on the adaptive lasso and the biased-corrected test statistic appears to be increasingly oversized for an increasing number of breaks. However, this can also be observed for the Maki-test and the HJ-test. A reason for this pattern might be the small number of observations per regime invalidating asymptotic

[4]We define $\Delta(A, B) = \sup_{b \in B} \inf_{a \in A} |a - b|$ for any two sets A and B, then $\max\{\Delta(A, B), \Delta(B, A)\}$ is the Hausdorff distance between A and B.

approximations. The size-adjusted power of our test increases with the sample size and faster speed of adjustment. It is generally low for small adjustment coefficients and always in favor of the ADF test statistic over the bias-corrected test statistic for larger values of the adjustment coefficients. Low power against the cointegration alternative with slow adjustment can be attributed to the inaccuracy of detecting the correct breakpoints if the cointegration residuals are near unit root processes. Choosing the correct maximum number of breaks exerts substantial influence on the power curves, as the power of the proposed test is always some magnitudes higher for the correct model choice. The results thereby show that, in terms of power, it is generally better to select irrelevant breaks than to restrict the number of breaks and to miss a crucial parameter change, which should also be a guiding principle in applied work with our proposed general-to-specific modelling approach.

As expected, our test outperforms the EG-test for specifications with structural breaks. Further, it also performs better than the GH-test for more than one break and better than the HJ-test for more than two breaks if the adjustment is moderate. The Maki-test appears to perform better for a high amount of breakpoints and slow adjustment. However, it does not provide a comprehensive modelling framework as it does not attempt to estimate all breakpoints consistently.

4 Empirical application

This section considers an application of our proposed framework to the long-run purchasing power parity (PPP) between the US and the UK with more than a century of data and several potential regime shifts. PPP is a well-known theory in macroeconomics which postulates that the nominal exchange rate between two currencies should be equal to the ratio of the domestic to the foreign price level. This can be tested empirically if the nominal exchange rate and the price ratio are cointegrated. Extensive studies have been conducted to investigate whether the proposition of PPP holds over the long-run (see Taylor and Taylor (2004) for a comprehensive review). However, empirical evidence on the subject still remains contradictory. Particularly the examined data span seems to have a substantial influence on the outcomes of empirical studies (Taylor, 2006; Karoglou and Morley, 2012). The diverging findings could on one hand be rooted in the generally slow adjustment of the real exchange rate which necessitates long samples for robust results. On the other hand these long samples might be composed of regimes shaped by very different macroeconomic environments which have to be accounted for. For example, the US and the UK moved between several fixed and floating exchange rate regimes in the last century.

We apply our methodology to monthly data from 1885 to 2015 and study the effects of structural breaks on the long-run PPP relationship. The dataset is an extended version of the data collected in Grilli and Kaminsky (1991) and Engel and Kim (1999). It is unique in the sense that such a comparatively high sampling frequency is not available for other country pairs over such a long sampling period. The logarithm of the nominal USD/GBP exchange rate is denoted by ex_t while p_t is the log US price level and p_t^* is the log UK price level, respectively. We estimate the regression,

$$ex_t = \alpha + \beta(p_t - p_t^*) + u_t, \tag{17}$$

where ex_t, p_t and p_t^* are generated by integrated processes of order one. Under PPP, we would expect to observe a mean-zero stationary error term u_t so that nominal exchange rate and relative prices are cointegrated. The strong from of PPP, assuming strict proportionality, would be given by the restriction $\alpha = 0$ and $\beta = 1$. Since we can only speculate which macroeconomic events might have changed the long-run equilibrium relationship, we apply our adaptive lasso procedure without the pre-specification of breakpoint candidates. The maximum number of breaks is chosen to be six, which corresponds to the number of changes from fixed to floating exchange rates and vice versa in our sample. A detailed description of US/UK exchange rate regimes is provided in Craighead (2010). The minimum length of a regime was chosen to be one year.

First, we assume constancy of the parameters and ignore potential structural breaks. Estimation of the long-run equilibrium coefficients yields OLS estimates $\hat{\alpha} = 0.46$ and $\hat{\beta} = 0.77$. The Engle-Granger test based on an ADF regression rejects the null hypothesis at the 1% level. Similar results can be obtained for the Phillips-Ouliaris test. The adjustment is slow ($\hat{\rho} = -0.016$) with a half life period of disequilibrium states of more than 3.5 years. Next, we compare several previously mentioned structural break models with our model selection approach. The GH-test finds evidence for cointegration with a breakpoint at 1949 m02, while the HJ-test indicates breakpoints at 1949 m02 and 1982 m01. The Maki-test selects the breakpoints 1919 m07, 1949 m08, 1967 m11, 1978 m07, 1987 m03 and rejects the null hypothesis of no cointegration as well. Our new general-to-specific procedure yields break dates 1919 m12, 1946 m07, 1949 m09, 1967 m12, 1982 m11, 2005 m09 and rejects the null hypothesis of no cointegration at the 1%

(5%) level for the ADF (bias-corrected) test statistic.

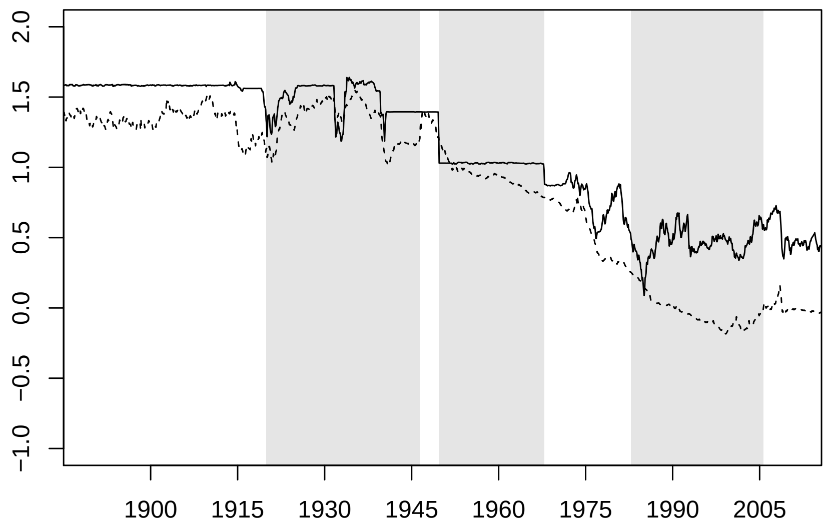

Figure 1: Nominal exchange rate (solid) and relative price levels (dashed). The shaded areas correspond to the regimes identified by the adaptive lasso procedure.

The indicated break dates (see Figure 1) can all be related to economic events affecting the real exchange rate. The first regime extends to the end of the first world war and spans the classical gold standard period with a fixed price of a pound sterling at USD 4.87. The next regime extends to the end of the second world war and comprises of fixed and floating exchange rate regimes in the inter-war period. The third breakpoint is found in September 1949 which coincides with a devaluation of pound sterling by roughly 30% and is followed by another breakpoint after Britain devalued the pound in November 1967. After the Bretton Woods system ended, we find two more breakpoints where the first one can be associated with a deep recession in the UK and the second one slightly pre-dates the financial crisis. We have to emphasize that the break dates might be affected by the usual lead and lag effects, since the parameter changes are representative for the following regime.

We can clearly see from Figure 2 that the post-lasso residuals are characterized by high and low volatility periods. The high volatility periods match with the floating exchange rate periods which supports the hypothesis formulated by Mussa (1986) who

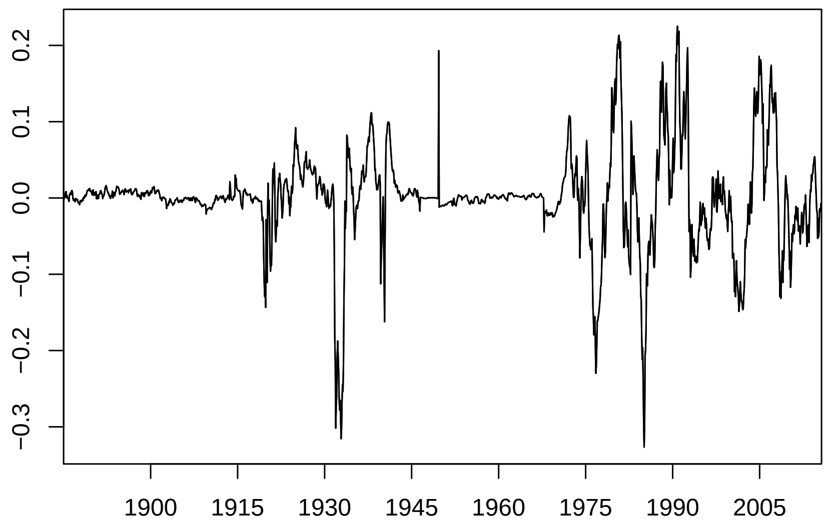

Figure 2: Post-lasso residual series. Fixed exchange rate regimes are marked by white areas, floating exchange rate regimes are marked by light-blue areas and both world wars are marked by orange-red areas.

argues that the nominal exchange rate system is a major determinant of the real exchange rate volatility. Accounting for the different exchange rate regimes yields a much faster adjustment ($\hat{\rho} = -0.062$) and a half life of disequilibrium states of less than a year. Our findings reveal that long-run PPP can only be properly assessed if we model structural instabilities induced by policy decisions which lead to substantial moves in the exchange rate.

5 Conclusion

In this article, we propose a new general-to-specific method for the accurate detection of an unknown number of structural breaks in cointegrating regressions. Furthermore, we design a new test for cointegration under the presence of multiple structural breaks based on our proposed adaptive lasso estimator. Our main goal is to build a comprehensive modeling framework which does not only focus on improving the performance of a specific test for cointegration under parameter instability but also provides consistent estimates of the corresponding parameter changes. Our procedure for a diverging amount of breakpoint candidates imposes very few a priori restrictions on the timing of the breakpoints, only requiring to specify a maximum number of breaks. This promises a high degree of flexibility for applied economists in finding the right model specification.

Although some authors have already used lasso estimators in a cointegration framework, our specific application of the adaptive lasso to cointegrating regressions with multiple structural breaks has not been discussed in the literature before. Thus, we first prove the suitability of the adaptive lasso estimator in our context by showing its nonstandard oracle property for a fixed amount of breakpoint candidates. Subsequently, we provide extensive Monte Carlo evidence that our framework can be extended to a diverging number of breakpoint candidates. Our results show an accurate and consistent estimation of the parameter changes for different choices for the maximum number of possible breakpoints.

The analysis of this paper is confined to a bivariate cointegration model. A generalization to multivariate cointegration models is straightforward in situations where fixed breakpoint candidates are available and $d^* \ll T$. However, the same does not hold for a diverging amount of breakpoint candidates which would require a different penalty for nonzero parameters reducing the overall number of penalized parameters. A suitable solution could be adapting the group-fused lasso estimator proposed in Qian and Su (2016) to nonstationary regressions. Their estimator forces all slope coefficients to have breaks at a common break date. We are currently investigating this possibility in ongoing research. Further, we could use our framework to construct confidence bands around estimated break fractions if the general difficulties with inference about lasso estimators are resolved.

Our residual-based cointegration tests appear to perform well in terms of power, especially for moderate adjustment. They outperform standard tests and are only inferior to the Maki-test if the adjustment is very slow. It seems that the adaptive lasso estimator needs a strong signal to find true breakpoints which directly determines the power of the cointegration tests. It would be interesting to develop an analogous testing framework which reverses the null hypothesis and alternative similar to the tests described in Carrion-i Silvestre and Sanso (2006) and Arai and Kurozumi (2007).

These findings show a promising new direction for cointegration model specification and cointegration testing in the presence of multiple structural breaks. While the currently proposed estimation and testing procedure provide the possibility of several extensions, it also constitutes a solid benchmark for other general-to-specific approaches dealing with structural change in cointegration models.

6 Acknowledgements

We thank Robert Jung, Andrew Tremayne and Jana Mareckova for valuable comments and suggestions. Further, we thank the organizers and participants of the Junior Research Seminar in Econometrics in Obermarchtal and the Economics Brown Bag seminar at the University of Hohenheim.

7 Appendix

Table 1: Approximate critical values

	ADF			Z_t		
$m^* = p^* = 1$						
T	10%	5%	1%	10%	5%	1%
100	−3.91	−4.28	−4.93	−4.09	−4.45	−5.13
200	−3.89	−4.24	−4.88	−4.03	−4.37	−5.04
400	−3.88	−4.23	−4.86	−3.96	−4.31	−4.92
∞	−3.86	−4.19	−4.83	−3.92	−4.26	−4.85
$m^* = p^* = 2$						
T	10%	5%	1%	10%	5%	1%
100	−4.51	−4.90	−5.59	−4.79	−5.18	−5.87
200	−4.51	−4.88	−5.53	−4.70	−5.07	−5.73
400	−4.48	−4.85	−5.47	−4.63	−4.99	−5.63
∞	−4.48	−4.84	−5.47	−4.58	−4.94	−5.57
$m^* = p^* = 3$						
T	10%	5%	1%	10%	5%	1%
100	−4.96	−5.35	−6.07	−5.34	−5.74	−6.45
200	−4.98	−5.36	−6.01	−5.25	−5.62	−6.30
400	−4.99	−5.34	−6.00	−5.19	−5.54	−6.20
∞	−4.98	−5.34	−6.02	−5.10	−5.48	−6.15
$m^* = p^* = 4$						
T	10%	5%	1%	10%	5%	1%
100	−5.29	−5.70	−6.43	−5.79	−6.21	−6.94
200	−5.38	−5.78	−6.46	−5.71	−6.11	−6.84
400	−5.42	−5.80	−6.44	−5.66	−6.03	−6.70
∞	−5.40	−5.78	−6.45	−5.57	−5.95	−6.60
$m^* = p^* = 5$						
T	10%	5%	1%	10%	5%	1%
100	−5.58	−6.00	−6.70	−6.18	−6.64	−7.46
200	−5.70	−6.10	−6.79	−6.11	−6.52	−7.29
400	−5.74	−6.16	−6.82	−6.03	−6.43	−7.15
∞	−5.77	−6.16	−6.86	−5.97	−6.35	−7.06
$m^* = p^* = 6$						
T	10%	5%	1%	10%	5%	1%
100	−5.77	−6.21	−7.02	−6.46	−6.98	−7.84
200	−5.97	−6.39	−7.08	−6.45	−6.88	−7.67
400	−6.05	−6.47	−7.22	−6.39	−6.80	−7.56
∞	−6.10	−6.49	−7.17	−6.33	−6.72	−7.41

Note: The lag truncation parameter in the ADF regression is determined using the AIC. Critical values for other order selection rules are not reported but can be obtained from the author upon request. We use 25,000 replications to compute the finite sample critical values.

Table 2: Estimation of multiple structural breaks in the intercept and slope coefficient (fixed breakpoint candidates)

SB1: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, 2\}$, $(\tau = 0.5)$

T	μ_1^*	μ_2^*	μ_3^*	μ_4^*	μ_5^*	μ_6^*	μ_7^*	μ_8^*
100	2.143 (7.383)	0.004 (0.734)	0.017 (0.763)	0.032 (24.372)	1.744 (27.585)	0.020 (0.733)	0.003 (0.569)	0.007 (0.548)
200	2.136 (2.530)	0.007 (0.469)	0.002 (0.486)	−0.127 (12.809)	1.928 (15.603)	0.019 (0.376)	0.001 (0.390)	−0.005 (0.389)
400	2.101 (1.260)	0.004 (0.391)	0.003 (0.389)	−0.161 (8.530)	1.963 (10.476)	0.008 (0.293)	−0.001 (0.287)	−0.002 (0.297)
800	2.063 (0.904)	0.003 (0.349)	0.000 (0.336)	0.064 (4.385)	1.795 (5.870)	0.006 (0.273)	−0.002 (0.255)	−0.004 (0.255)

T	β_1^*	β_2^*	β_3^*	β_4^*	β_5^*	β_6^*	β_7^*	β_8^*
100	2.198 (0.393)	0.000 (0.023)	0.003 (0.054)	0.285 (1.255)	1.364 (1.376)	0.005 (0.073)	0.001 (0.027)	0.001 (0.037)
200	2.078 (0.137)	0.000 (0.002)	0.001 (0.015)	0.211 (0.619)	1.645 (0.688)	0.000 (0.010)	0.000 (0.009)	0.000 (0.011)
400	2.050 (0.076)	0.000 (0.001)	0.000 (0.004)	0.075 (0.307)	1.843 (0.345)	0.000 (0.005)	0.000 (0.001)	0.000 (0.003)
800	2.031 (0.046)	0.000 (0.000)	0.000 (0.001)	0.012 (0.101)	1.938 (0.131)	0.000 (0.002)	0.000 (0.000)	0.000 (0.000)

SB2: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, 2, 3\}$, $(\tau_1 = 0.25,\ \tau_2 = 0.75)$

T	μ_1^*	μ_2^*	μ_3^*	μ_4^*	μ_5^*	μ_6^*	μ_7^*	μ_8^*
100	2.049 (16.263)	0.313 (13.717)	1.616 (20.403)	0.010 (0.613)	0.009 (0.622)	−0.009 (21.770)	2.000 (26.376)	0.009 (1.225)
200	1.997 (8.298)	−0.069 (8.413)	2.070 (13.372)	0.011 (0.440)	0.004 (0.455)	−0.295 (13.982)	2.335 (19.690)	−0.010 (1.025)
400	2.126 (3.550)	0.001 (5.305)	1.869 (7.313)	0.007 (0.428)	0.007 (0.441)	−0.151 (9.491)	1.959 (16.354)	0.002 (0.867)
800	2.091 (2.273)	0.032 (1.834)	1.868 (3.702)	−0.001 (0.413)	0.003 (0.438)	0.026 (4.531)	1.877 (12.696)	−0.003 (0.871)

T	β_1^*	β_2^*	β_3^*	β_4^*	β_5^*	β_6^*	β_7^*	β_8^*
100	2.641 (0.938)	0.111 (0.779)	1.273 (1.193)	0.005 (0.069)	0.004 (0.061)	0.333 (1.132)	1.243 (1.255)	0.005 (0.071)
200	2.281 (0.453)	0.115 (0.447)	1.604 (0.685)	0.000 (0.009)	0.001 (0.013)	0.225 (0.629)	1.589 (0.764)	0.001 (0.027)
400	2.153 (0.211)	0.034 (0.212)	1.830 (0.325)	0.000 (0.001)	0.000 (0.002)	0.081 (0.301)	1.799 (0.430)	0.000 (0.004)
800	2.100 (0.131)	0.004 (0.048)	1.911 (0.154)	0.000 (0.000)	0.000 (0.001)	0.015 (0.096)	1.902 (0.219)	0.000 (0.001)

SB3: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, \ldots, 4\}$, $(\tau_1 = 0.25,\ \tau_2 = 0.5,\ \tau_3 = 0.75)$

T	μ_1^*	μ_2^*	μ_3^*	μ_4^*	μ_5^*	μ_6^*	μ_7^*	μ_8^*
100	1.979 (20.169)	0.286 (11.902)	1.661 (18.975)	0.092 (15.905)	2.029 (21.521)	0.227 (17.692)	2.009 (22.915)	0.024 (1.388)
200	1.956 (9.361)	−0.104 (7.719)	2.181 (14.867)	−0.097 (13.264)	2.142 (18.240)	−0.381 (14.153)	2.550 (20.722)	−0.013 (1.190)
400	2.098 (4.266)	−0.015 (5.490)	1.907 (9.705)	−0.107 (7.940)	2.199 (15.172)	−0.085 (9.055)	2.005 (19.680)	0.006 (1.190)
800	2.080 (2.804)	0.021 (2.011)	1.898 (5.948)	0.041 (3.465)	1.863 (11.065)	0.019 (3.208)	2.159 (18.747)	0.002 (1.337)

T	β_1^*	β_2^*	β_3^*	β_4^*	β_5^*	β_6^*	β_7^*	β_8^*
100	2.710 (1.111)	0.138 (0.704)	1.293 (1.232)	0.349 (1.004)	1.408 (1.279)	0.386 (1.020)	1.289 (1.181)	0.005 (0.068)
200	2.325 (0.503)	0.101 (0.415)	1.620 (0.798)	0.195 (0.615)	1.721 (0.877)	0.215 (0.614)	1.603 (0.832)	0.001 (0.020)
400	2.180 (0.248)	0.032 (0.212)	1.851 (0.429)	0.047 (0.263)	1.912 (0.524)	0.054 (0.266)	1.784 (0.529)	0.000 (0.004)
800	2.123 (0.162)	0.004 (0.053)	1.912 (0.224)	0.005 (0.074)	1.980 (0.262)	0.006 (0.064)	1.870 (0.337)	0.000 (0.001)

Note: We use 10,000 replications of the data-generating process given in Equation (16) with seven breakpoint candidates using equidistant spacing $\tau = (0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875)$. The adjustment coefficient is $\rho = -1$ (iid error terms), the variance of the error terms is $\sigma_\omega^2 = 1$ and $\sigma_\vartheta^2 = 1$, respectively. The first panel reports the results for one active breakpoint at $\tau = 0.5$, the second panel considers two active breakpoints at $\tau_1 = 0.25$, $\tau_2 = 0.75$ and the third panel has three active breakpoints at $\tau_1 = 0.25$, $\tau_2 = 0.5$, $\tau_3 = 0.75$. The baseline coefficients and parameter changes at all breakpoints take the value 2. We use initial least squares estimates to compute the adaptive lasso weights. Standard deviations are given in parentheses.

Table 3: Estimation of multiple structural breaks in the intercept and slope coefficient (diverging amount of breakpoint candidates)

SB1: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, 2\}$, $(\tau = 0.5)$

T	pce	hd/T	τ	μ_1^*	μ_2^*	β_1^*	β_2^*
100	99.8	0.54	0.502 (0.041)	2.04 (1.510)	1.98 (2.270)	2.01 (0.156)	1.99 (0.282)
200	100	0.32	0.500 (0.019)	2.02 (0.774)	1.96 (1.247)	2.00 (0.074)	2.00 (0.106)
400	100	0.19	0.499 (0.008)	2.00 (0.405)	1.98 (0.739)	2.00 (0.035)	2.00 (0.048)
800	100	0.15	0.499 (0.006)	2.00 (0.244)	1.99 (0.503)	2.00 (0.017)	2.00 (0.025)

SB2: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, 2, 3\}$, $(\tau_1 = 0.33,\ \tau_2 = 0.67)$

T	pce	hd/T	τ_1	τ_2	μ_1^*	μ_2^*	μ_3^*	β_1^*	β_2^*	β_3^*
100	98.7	1.26	0.335 (0.054)	0.677 (0.044)	2.07 (2.230)	1.96 (3.470)	1.93 (3.820)	2.03 (0.292)	2.01 (0.500)	1.96 (0.569)
200	99.8	0.80	0.331 (0.038)	0.673 (0.033)	2.03 (1.130)	2.00 (2.250)	1.95 (2.340)	2.01 (0.171)	2.00 (0.201)	1.98 (0.247)
400	99.9	0.48	0.329 (0.020)	0.671 (0.020)	2.02 (0.602)	2.00 (1.009)	2.02 (1.296)	2.00 (0.084)	2.00 (0.121)	1.99 (0.147)
800	99.9	0.37	0.328 (0.009)	0.670 (0.007)	2.01 (0.362)	1.98 (0.733)	2.01 (0.961)	2.00 (0.030)	2.00 (0.044)	2.00 (0.043)

SB4: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, \ldots, 5\}$, $(\tau_1 = 0.2,\ \tau_2 = 0.4,\ \tau_3 = 0.6,\ \tau_4 = 0.8)$

T	pce	hd/T	τ_1	τ_2	τ_3	τ_4
100	94.4	2.61	0.208 (0.050)	0.410 (0.060)	0.612 (0.057)	0.809 (0.039)
200	97.2	2.49	0.204 (0.044)	0.407 (0.053)	0.610 (0.053)	0.808 (0.035)
400	98.7	1.67	0.202 (0.040)	0.405 (0.045)	0.609 (0.049)	0.806 (0.033)
800	99.2	1.16	0.202 (0.037)	0.403 (0.036)	0.604 (0.034)	0.803 (0.024)

T	μ_1^*	μ_2^*	μ_3^*	μ_4^*	μ_5^*
100	2.11 (3.72)	2.19 (8.47)	1.81 (8.92)	1.85 (6.40)	2.03 (6.44)
200	2.09 (1.92)	2.03 (2.95)	1.92 (3.51)	1.92 (3.81)	1.96 (3.86)
400	2.09 (1.10)	1.94 (2.80)	1.99 (3.40)	1.95 (3.08)	1.97 (2.66)
800	2.07 (0.91)	2.00 (2.41)	1.92 (2.60)	1.99 (1.74)	1.99 (1.68)

T	β_1^*	β_2^*	β_3^*	β_4^*	β_5^*
100	2.09 (0.507)	2.01 (1.436)	2.03 (1.428)	1.98 (0.733)	1.89 (0.831)
200	2.06 (0.344)	2.02 (0.427)	2.01 (0.581)	1.99 (0.575)	1.90 (0.564)
400	2.04 (0.266)	2.01 (0.312)	2.02 (0.298)	2.01 (0.362)	1.90 (0.526)
800	2.04 (0.233)	2.00 (0.321)	2.00 (0.246)	2.00 (0.184)	1.95 (0.308)

Note: We use 5,000 replications of the data-generating process given in Equation (16). The adjustment coefficient is $\rho = -1$ (iid error terms), the variance of the error terms is $\sigma_\omega^2 = 1$ and $\sigma_\vartheta^2 = 2$, respectively. Models with a better signal-to-noise ratio yield more precise estimates for all sample sizes. 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 $\tau_4 = 0.8$. The baseline coefficients and parameter changes at all breakpoints take the value 2. We use the procedure detailed in Subsection 2.2 to compute the adaptive lasso weights and apply post-lasso estimation to obtain the estimates for the intercept breaks. Standard deviations are given in parentheses.

Table 4: Size and size-adjusted power of cointegration tests in the presence of multiple regime shifts ($T = 100$)

ρ	$ADF_{\tau,1}$	$ADF_{\tau,2}$	$ADF_{\tau,4}$	$Z_{\tau,1}$	$Z_{\tau,2}$	$Z_{\tau,4}$	GH	HJ	$M1$	$M2$	$M4$	EG
	SB0: $\mu = 2$, $\beta = 2$											
Size:	0.061	0.059	0.053	0.076	0.074	0.083	0.076	0.099	0.054	0.068	0.097	0.041
−0.05	0.102	0.115	0.120	0.080	0.083	0.060	0.069	0.056	0.068	0.072	0.062	0.099
−0.10	0.174	0.165	0.134	0.136	0.114	0.066	0.106	0.078	0.106	0.106	0.080	0.226
−0.25	0.604	0.388	0.173	0.533	0.304	0.101	0.496	0.274	0.492	0.370	0.218	0.885
−0.50	1.000	0.983	0.740	0.998	0.943	0.496	0.997	0.933	0.998	0.968	0.760	1.000
	SB1: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, 2\}$, $(\tau = 0.5)$											
−0.05	0.095	0.102	0.133	0.073	0.066	0.071	0.663	0.695	0.683	0.571	0.376	0.042
−0.10	0.157	0.176	0.160	0.123	0.106	0.083	0.753	0.746	0.766	0.644	0.450	0.048
−0.25	0.642	0.545	0.353	0.564	0.427	0.197	0.934	0.879	0.935	0.829	0.638	0.066
−0.50	0.995	0.950	0.829	0.992	0.941	0.707	1.000	0.994	1.000	0.995	0.924	0.086
	SB2: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, 2, 3\}$, $(\tau_1 = 0.33,\ \tau_2 = 0.67)$											
−0.05	0.056	0.107	0.143	0.044	0.065	0.061	0.225	0.815	0.256	0.606	0.370	0.043
−0.10	0.057	0.154	0.181	0.044	0.097	0.083	0.237	0.852	0.272	0.668	0.420	0.045
−0.25	0.080	0.505	0.424	0.064	0.396	0.239	0.283	0.950	0.320	0.827	0.580	0.047
−0.50	0.122	0.966	0.884	0.097	0.955	0.784	0.334	0.999	0.372	0.976	0.888	0.054
	SB4: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, \ldots, 5\}$, $(\tau_1 = 0.2,\ \tau_2 = 0.4,\ \tau_3 = 0.6\ \tau_4 = 0.8)$											
−0.05	0.028	0.044	0.154	0.021	0.027	0.056	0.089	0.257	0.100	0.108	0.403	0.039
−0.10	0.033	0.047	0.190	0.022	0.032	0.087	0.085	0.263	0.102	0.118	0.432	0.038
−0.25	0.031	0.054	0.431	0.019	0.037	0.249	0.091	0.284	0.107	0.122	0.542	0.038
−0.50	0.032	0.064	0.902	0.022	0.046	0.814	0.100	0.317	0.113	0.148	0.773	0.039

Note: $ADF_{\tau,1}$, $ADF_{\tau,2}$, and $ADF_{\tau,4}$ denote the cointegration test based on the ADF regression with one, two and four possible breakpoints, respectively. Correspondingly, $Z_{\tau,1}$, $Z_{\tau,2}$, and $Z_{\tau,4}$ denote the cointegration test based on the bias-corrected test statistic. GH denote the cointegration test with one structural break (model C/S) by Gregory and Hansen (1996) and HJ denotes the corresponding test with two structural breaks by Hatemi-J (2008). $M1$, $M2$, and $M4$ denote the cointegration test by Maki (2012) allowing for one, two and four structural breaks, respectively. EG denotes the Engle-Granger cointegration test with constant coefficients. The results are obtained from 2,500 replications. All tests are conducted at the 5% significance level.

Table 5: Size and size-adjusted power of cointegration tests in the presence of multiple regime shifts ($T = 200$)

ρ	$ADF_{\tau,1}$	$ADF_{\tau,2}$	$ADF_{\tau,4}$	$Z_{\tau,1}$	$Z_{\tau,2}$	$Z_{\tau,4}$	GH	HJ	$M1$	$M2$	$M4$	EG
	SB0: $\mu = 2$, $\beta = 2$											
Size:	0.062	0.054	0.051	0.066	0.067	0.074	0.062	0.080	0.045	0.056	0.060	0.040
-0.05	0.162	0.152	0.131	0.140	0.104	0.079	0.102	0.064	0.090	0.092	0.083	0.235
-0.10	0.406	0.296	0.180	0.365	0.231	0.124	0.312	0.154	0.285	0.222	0.167	0.705
-0.25	1.000	0.952	0.557	0.998	0.909	0.386	0.993	0.872	0.990	0.930	0.730	1.000
-0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	SB1: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, 2\}$, ($\tau = 0.5$)											
-0.05	0.148	0.157	0.158	0.128	0.100	0.090	0.707	0.736	0.722	0.598	0.450	0.069
-0.10	0.416	0.358	0.247	0.374	0.278	0.159	0.874	0.848	0.870	0.738	0.588	0.087
-0.25	0.996	0.972	0.784	0.995	0.956	0.682	1.000	0.997	1.000	0.990	0.914	0.121
-0.50	0.998	0.974	0.930	0.998	0.973	0.930	1.000	1.000	1.000	1.000	1.000	0.138
	SB2: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, 2, 3\}$, ($\tau_1 = 0.33$, $\tau_2 = 0.67$)											
-0.05	0.070	0.138	0.156	0.060	0.092	0.084	0.296	0.830	0.339	0.643	0.436	0.055
-0.10	0.090	0.320	0.291	0.077	0.240	0.187	0.352	0.916	0.388	0.761	0.547	0.058
-0.25	0.134	0.966	0.865	0.114	0.946	0.780	0.418	0.998	0.460	0.973	0.873	0.066
-0.50	0.168	0.987	0.937	0.138	0.986	0.936	0.456	1.000	0.505	0.992	1.000	0.069
	SB4: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1, \ldots, 5\}$, ($\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$ $\tau_4 = 0.8$)											
-0.05	0.038	0.042	0.166	0.030	0.028	0.098	0.104	0.284	0.112	0.143	0.462	0.048
-0.10	0.038	0.049	0.297	0.030	0.035	0.189	0.109	0.302	0.119	0.152	0.545	0.050
-0.25	0.043	0.061	0.866	0.030	0.043	0.769	0.119	0.346	0.131	0.173	0.796	0.052
-0.50	0.044	0.075	0.960	0.033	0.049	0.948	0.132	0.382	0.149	0.214	0.948	0.055

Note: $ADF_{\tau,1}$, $ADF_{\tau,2}$, and $ADF_{\tau,4}$ denote the cointegration test based on the ADF regression with one, two and four possible breakpoints, respectively. Correspondingly, $Z_{\tau,1}$, $Z_{\tau,2}$, and $Z_{\tau,4}$ denote the cointegration test based on the bias-corrected test statistic. GH denote the cointegration test with one structural break (model C/S) by Gregory and Hansen (1996) and HJ denotes the corresponding test with two structural breaks by Hatemi-J (2008). $M1$, $M2$,and $M4$ denote the cointegration test by Maki (2012) allowing for one, two and four structural breaks, respectively. EG denotes the Engle-Granger cointegration test with constant coefficients. The results are obtained from 2,500 replications. All tests are conducted at the 5% significance level.

Supplementary Material

1.1 Proof of Theorem 1

Suppose, the scalar partial sum process in Equation (2) of the manuscript satisfies the functional central limit theorem (FCLT) for Reyni-mixing processes, described in Hall and Heyde (1980). For $s \in [0,1]$ and as $T \to \infty$, it holds that

$$x_{[Ts]} = T^{-1/2} \sum_{t=1}^{[Ts]} v_t \Rightarrow B(s), \tag{18}$$

where $B(s)$ is a Brownian motion process with variance σ^2. Next, we define the objective function $V_T(\boldsymbol{b})$ by

$$\begin{aligned} V_T(\boldsymbol{b}) &= \sum_{t=1}^{T} \Big[(u_t - \boldsymbol{b}'\boldsymbol{X}_t \delta_T^{-1})^2 - u_t^2 \Big] && (19) \\ &+ \lambda_T \sum_{i=2}^{p^*+1} w_{1i}^{\gamma_1} |\mu_i^* + b_{1i}/\sqrt{T}| + \lambda_T \sum_{j=2}^{m^*+1} w_{2j}^{\gamma_2} |\beta_j^* + b_{2j}/T|, && (20) \end{aligned}$$

where $\boldsymbol{b} = (\boldsymbol{b}_1', \boldsymbol{b}_2')'$, $\delta_T = diag(T^{1/2} I_{p^*+1}, T I_{m^*+1})$ and

$$\hat{\boldsymbol{b}} = (\hat{\boldsymbol{b}}_1', \hat{\boldsymbol{b}}_2')' = \arg\min V_T(\boldsymbol{b}) \tag{21}$$

is the minimizer of V_T with $\hat{b}_{1i} = \sqrt{T}(\hat{\mu}_{T,i} - \mu_i^*)$ and $\hat{b}_{2j} = T(\hat{\beta}_{T,j} - \beta_j^*)$.

First, we consider the asymptotic counterparts to the least squares terms

$$-2 \sum_{t=1}^{T} u_t \boldsymbol{b}'\boldsymbol{X}_t \delta_T^{-1} + \sum_{t=1}^{T} \boldsymbol{b}'\boldsymbol{X}_t \delta_T^{-1} \delta_T^{-1} \boldsymbol{X}_t' \boldsymbol{b}.$$

We use the decomposition $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ to express the weak convergence result

$$T^{-1/2} I_{m^*+1} \boldsymbol{X}_{2,[Ts]} \Rightarrow (B(s), B(s)\varphi_{\tau_{2,1}}(s), \ldots, B(s)\varphi_{\tau_{2,m^*}}(s)) = B_\tau(s), \tag{22}$$

where

$$\varphi_{\tau_{k,l}}(s) \begin{cases} 0 & \text{if } s < \tau_{k,l} \\ 1 & \text{if } s \geq \tau_{k,l} \end{cases}, \qquad k \in \{1,2\}, s \in [0,1]. \tag{23}$$

Using (A.4) in Gregory and Hansen (1996) and the continuous mapping theorem (CMT, see Billingsley (1999), Theorem 2.7), we observe that

$$\sum_{t=1}^{T} \boldsymbol{b}'\boldsymbol{X}_t\delta_T^{-1}\delta_T^{-1}\boldsymbol{X}_t'\boldsymbol{b} \Rightarrow \boldsymbol{b}'\begin{bmatrix} \Upsilon & 0 \\ 0 & \int_0^1 B_\tau(s)B_\tau(s)'ds \end{bmatrix}\boldsymbol{b},$$

where the weak convergence is uniform over the vector $(\tau_{1,1},\ldots,\tau_{1,p^*},\tau_{2,1},\ldots,\tau_{1,m^*}) \in \mathcal{T}$. Further, using (A.3) in Gregory and Hansen (1996) and Theorem 3.1 in Hansen (1992), we have the weak convergence to a stochastic integral

$$\sum_{t=1}^{T} u_t\boldsymbol{b}'\boldsymbol{X}_t\delta_T^{-1} \Rightarrow \boldsymbol{b}'\begin{bmatrix} U \\ \int_0^1 B_\tau(s)dU(s) \end{bmatrix} + \boldsymbol{b}'\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Lambda \\ (1-\tau_{2,1})\Lambda \\ \vdots \\ (1-\tau_{2,m^*})\Lambda \end{bmatrix},$$

where $U \sim N(0,\Upsilon\sigma^2)$, $\Lambda = \lim_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} Ex_{t-1}u_t$ and $U(s)$ is the weak limit of u_t.

We implicitly assume $d^* \ll T$ and require that initial least squares estimates are available for the weights of the adaptive lasso estimator. We investigate the consistency of the individual coefficients and distinguish between the true coefficients being zero or nonzero:

(a) If $\mu_i^* \neq 0$, we have

$$\begin{aligned} & \lambda_T w_{1i}^{\gamma_1}\left[|\mu_i^* + b_{1i}/\sqrt{T}| - |\mu_i^*|\right] \\ = \; & \frac{\lambda_T}{\sqrt{T}}\left|\frac{1}{\hat{\mu}_{I,i}}\right|^{\gamma_1}\sqrt{T}\left[|\mu_i^* + b_{1i}/\sqrt{T}| - |\mu_i^*|\right] \\ \xrightarrow{p} \; & 0, \end{aligned} \tag{24}$$

since (i) $\frac{\lambda_T}{\sqrt{T}} \to 0$, (ii) $\left|\frac{1}{\hat{\mu}_{I,i}}\right|^{\gamma_1} \xrightarrow{p} \left|\frac{1}{\mu_i^*}\right|^{\gamma_1}$ if the initial estimator is consistent and (iii) $\sqrt{T}\left[|\mu_i^* + b_{1i}/\sqrt{T}| - |\mu_i^*|\right] \xrightarrow{p} b_{1i}\,\mathrm{sgn}(\mu_i^*)$ as in Zou (2006).

(b) If $\mu_i^* = 0$, we have

$$\begin{aligned}
\lambda_T w_{1i}^{\gamma_1}\left[|\mu_i^* + b_{1i}/\sqrt{T}| - |\mu_i^*|\right] &= \frac{\lambda_T}{\sqrt{T}}\left|\frac{1}{\hat{\mu}_{I,i}}\right|^{\gamma_1}|b_{1i}| \\
&= \frac{\lambda_T}{T^{1/2-\gamma_1/2}}\left|\frac{1}{\sqrt{T}\hat{\mu}_{I,i}}\right|^{\gamma_1}|b_{1i}| \qquad (25)\\
&\Rightarrow \begin{cases} \infty & \text{if} \quad b_{1i} \neq 0 \\ 0 & \text{if} \quad b_{1i} = 0 \end{cases},
\end{aligned}$$

since (i) $\frac{\lambda_T}{T^{1/2-\gamma_1/2}} \to \infty$ and (ii) the initial least squares estimator is tight and converges to a normal distribution, $\sqrt{T}\hat{\mu}_{I,i} \Rightarrow W_{1i} \sim N(0, \frac{\sigma^2}{\tau_{1,i}(1-\tau_{1,i})})$.

(c) If $\beta_j^* \neq 0$, we have

$$\begin{aligned}
&\lambda_T w_{2j}^{\gamma_2}\left[|\beta_j^* + b_{2j}/T| - |\beta_j^*|\right] \\
&= \frac{\lambda_T}{T}\left|\frac{1}{\hat{\beta}_{I,j}}\right|^{\gamma_2} T\left[|\beta_j^* + b_{2j}/T| - |\beta_j^*|\right] \qquad (26)\\
&\xrightarrow{p} 0,
\end{aligned}$$

since (i) $\frac{\lambda_T}{T} \to 0$, (ii) $\left|\frac{1}{\hat{\beta}_{I,j}}\right|^{\gamma_2} \xrightarrow{p} \left|\frac{1}{\beta_j^*}\right|^{\gamma_2}$ if the initial estimator is consistent and (iii) $T\left[|\beta_j^* + b_{2j}/T| - |\beta_j^*|\right] \xrightarrow{p} b_{2j}\,\mathrm{sgn}(\beta_j^*)$.

(d) If $\beta_j^* = 0$, we have

$$\begin{aligned}
\lambda_T w_{2j}^{\gamma_2}\left[|\beta_j^* + b_{2j}/T| - |\beta_j^*|\right] &= \frac{\lambda_T}{T}\left|\frac{1}{\hat{\beta}_{I,j}}\right|^{\gamma_2}|b_{2j}| \qquad (27)\\
&= \frac{\lambda_T}{T^{1-\gamma_2}}\left|\frac{1}{T\hat{\beta}_{I,j}}\right|^{\gamma_2}|b_{2j}| \qquad (28)\\
&\Rightarrow \begin{cases} \infty & \text{if} \quad b_{2j} \neq 0 \\ 0 & \text{if} \quad b_{2j} = 0 \end{cases}, \qquad (29)
\end{aligned}$$

since (i) $\frac{\lambda_T}{T^{1-\gamma_2}} \to \infty$, (ii) the least squares estimator is tight and has the following nonstandard distribution

$$T\hat{\beta}_{I,j} \Rightarrow W_{2j} = \frac{\left(\int_0^1 B_{\tau_{2,j}}(s)dU(s) + (1-\tau_{2,j})\Lambda\right)}{\int_0^1 B_{\tau_{2,j}}^2(s)ds}, \qquad (30)$$

and (iii) $P(W_{2j} = 0) \xrightarrow{a.s.} 0$.

Thus, $V_T(\boldsymbol{b}) \Rightarrow V(\boldsymbol{b})$, where

$$V(\boldsymbol{b}) = \begin{cases} \boldsymbol{b}' A \boldsymbol{b} - 2\boldsymbol{b}' B - 2\boldsymbol{b}' C & \text{if} \quad b_k = 0 \text{ for all } k \in \mathcal{A}^c \\ \infty & \text{if} \quad b_k \neq 0 \text{ for some } k \in \mathcal{A}^c \end{cases} \tag{31}$$

with

$$A = \begin{bmatrix} \Upsilon & 0 \\ 0 & \int_0^1 B_\tau(s) B_\tau(s)' ds \end{bmatrix},$$

$$B = \begin{bmatrix} U \\ \int_0^1 B_\tau(s) dU(s) \end{bmatrix}, \qquad U \sim N(0, \Upsilon \sigma^2),$$

$$C = [0, \ldots, 0, \Lambda, (1 - \tau_{2,1})\Lambda, \ldots, (1 - \tau_{2,m^*})\Lambda]'.$$

Since V_T is a convex function and V has a unique minimum, it follows from Knight and Fu (2000) that

$$\arg\min V_T(\boldsymbol{b}) = \hat{\boldsymbol{b}} = \begin{bmatrix} \sqrt{T}(\hat{\mu}_T - \mu^*) \\ T(\hat{\beta}_T - \beta^*) \end{bmatrix} \Rightarrow \arg\min V(\boldsymbol{b}). \tag{32}$$

From these results, we can deduce that

$$\begin{aligned} \sqrt{T}(\hat{\mu}_{T,\mathcal{A}_1^c} - \mu^*_{\mathcal{A}_1^c}) &\Rightarrow \delta_0^{|\mathcal{A}_1^c|} \\ \sqrt{T}(\hat{\mu}_{T,\mathcal{A}_1} - \mu^*_{\mathcal{A}_1}) &\Rightarrow N(0, [\Sigma_{\mathcal{A}_1}]^{-1} \Upsilon_{\mathcal{A}_1} \sigma^2), \end{aligned} \tag{33}$$

where δ_0 denotes the Dirac measure at 0. Correspondingly, we have

$$\begin{aligned} T(\hat{\beta}_{T,\mathcal{A}_2^c} - \beta^*_{\mathcal{A}_2^c}) &\Rightarrow \delta_0^{|\mathcal{A}_2^c|} \\ T(\hat{\beta}_{T,\mathcal{A}_2} - \beta^*_{\mathcal{A}_2}) &\Rightarrow \left[\int_0^1 B_{\tau,\mathcal{A}_2} B'_{\tau,\mathcal{A}_2} \right]^{-1} \left[\int_0^1 B_{\tau,\mathcal{A}_2} dU + C^*_{\mathcal{A}_2} \right]. \end{aligned} \tag{34}$$

It remains to show that coefficients of inactive variables are set to zero with probability approaching one. We begin with a proof of $P(\hat{\mu}_{T,\mathcal{A}_1^c} = 0) \to 1$. Consider the event that $\hat{\mu}_{T,i} \neq 0$ although $i \in \mathcal{A}_1^c$. We know from the Karush-Kuhn-Tucker (KKT) optimality conditions that the first order condition for a minimum is given by

$$\frac{2\varphi'_{\tau_{1,i}}(y - x(\hat{\mu}'_T, \hat{\beta}'_T)')}{\sqrt{T}} = \frac{\lambda_T w_{1i}^{\gamma_1} \operatorname{sgn}(\hat{\mu}_{T,i})}{\sqrt{T}}. \tag{35}$$

Note that

$$\left| \frac{\lambda_T w_{1i}^{\gamma_1} \operatorname{sgn}(\hat{\mu}_{T,i})}{\sqrt{T}} \right| = \frac{\lambda_T}{T^{1/2-\gamma_1/2}} \left| \frac{1}{T^{1/2}\hat{\mu}_{I,i}} \right|^{\gamma_1} \to \infty \tag{36}$$

since (i) $\frac{\lambda_T}{T^{1/2-\gamma_1/2}} \to \infty$ and (ii) $\sqrt{T}\hat{\mu}_{I,i}$ is tight. The left hand side of the equation is equivalent to

$$\frac{2\varphi'_{\tau_{1,i}}(u - x\delta_T^{-1}\delta_T(\hat{\mu}'_T - \mu^{*\prime}, \hat{\beta}'_T - \beta^{*\prime})')}{\sqrt{T}} = \frac{2\varphi'_{\tau_{1,i}}u}{\sqrt{T}} - \frac{2\varphi'_{\tau_{1,i}}x\delta_T^{-1}\delta_T(\hat{\mu}'_T - \mu^{*\prime}, \hat{\beta}'_T - \beta^{*\prime})'}{\sqrt{T}}. \tag{37}$$

For the first term, we have the weak convergence

$$\frac{\varphi'_{\tau_{1,i}}u}{\sqrt{T}} \Rightarrow N(0, \sigma^2\tau_{1,i}) \tag{38}$$

and for the second term, we have the weak convergence of $\frac{\varphi'_{\tau_{1,i}}x\delta_T^{-1}}{\sqrt{T}}$ which depends on the timing of the break fraction $\tau_{1,i}$ relative to all other possible break fractions. Say $\tau_{1,i} = \tau_{1,p^*} > \tau_{2,m^*}$ holds, then

$$\frac{\varphi'_{\tau_{1,i}}x\delta_T^{-1}}{\sqrt{T}} \Rightarrow \left(0, \ldots, 0, \int_0^1 B_{\tau_{2,1}}(s)ds, \ldots, \int_0^1 B_{\tau_{2,m^*}}(s)ds\right). \tag{39}$$

Further, we have already shown the weak convergence of $\delta_T(\hat{\mu}'_T - \mu^{*\prime}, \hat{\beta}'_T - \beta^{*\prime})'$. Hence, the distribution of the first term is tight and

$$P(\hat{\mu}_{T,i} \neq 0) \leq P\left(\frac{2\varphi'_{\tau_{1,i}}(y - x(\hat{\mu}'_T, \hat{\beta}'_T)')}{\sqrt{T}} - \frac{\lambda_T w_{1i}^{\gamma_1} \operatorname{sgn}(\hat{\mu}_{T,i})}{\sqrt{T}} = 0\right) \to 0. \tag{40}$$

Next, we show that $P(\hat{\beta}_{T,\mathcal{A}_2^c} = 0) \to 1$. Again, we consider the event that $\hat{\beta}_{T,j} \neq 0$ although $j \in \mathcal{A}_2^c$. The KKT optimality condition in this case is given by

$$\frac{2(x\varphi_{\tau_{2,j}})'(y - x(\hat{\mu}'_T, \hat{\beta}'_T)')}{T} = \frac{\lambda_T w_{2j}^{\gamma_2} \operatorname{sgn}(\hat{\beta}_{T,j})}{T}, \tag{41}$$

where the factor T substitutes the factor $\sqrt{T}$. For the right hand side of the equation, we observe that

$$\left|\frac{\lambda_T w_{2j}^{\gamma_2}\,\mathrm{sgn}(\hat{\beta}_{T,j})}{T}\right| = \frac{\lambda_T}{T^{1-\gamma_2}}\left|\frac{1}{T\hat{\beta}_{I,j}}\right|^{\gamma_2} \to \infty \tag{42}$$

since $T\hat{\beta}_{I,j}$ is tight. For the left hand side,

$$\frac{2(x\varphi_{\tau_{2,j}})'u}{T} - \frac{2(x\varphi_{\tau_{2,j}})'x\delta_T^{-1}\delta_T(\hat{\mu}_T' - \mu^{*\prime}, \hat{\beta}_T' - \beta^{*\prime})'}{T}, \tag{43}$$

we have the weak convergence of the first term using

$$\frac{(x\varphi_{\tau_{2,j}})'u}{\sqrt{T}} \Rightarrow \int_0^1 B_{\tau_{2,j}}(s)dU(s) + (1-\tau_{2,j})\Lambda. \tag{44}$$

The expression of the weak convergence result for the second term depends on the timing of the break fraction $\tau_{2,j}$. Say $\tau_{2,j} = \tau_{1,m^*} > \tau_{1,p^*}$ holds, we have

$$\frac{(x\varphi_{\tau_{2,j}})'x\delta_T^{-1}}{T} \Rightarrow \Bigg(\int_0^1 B_{\tau_{2,j}}(s)ds, \int_0^1 B_{\tau_{1,1}}(s)ds, \ldots, \int_0^1 B_{\tau_{2,p^*}}(s)ds, \tag{45}$$

$$\int_0^1 B^2_{\tau_{2,j}}(s)ds, \int_0^1 B^2_{\tau_{1,1}}(s)ds, \ldots, \int_0^1 B^2_{\tau_{2,m^*}}(s)ds\Bigg), \tag{46}$$

and as before $\delta_T(\hat{\mu}_T' - \mu^{*\prime}, \hat{\beta}_T' - \beta^{*\prime})'$ is tight. Finally, we have shown that

$$P(\hat{\beta}_{T,j} \neq 0) \leq P\left(\frac{2(x\varphi_{\tau_{2,j}})'(y - x(\hat{\mu}_T', \hat{\beta}_T')')}{T} - \frac{\lambda_T w_{2j}^{\gamma_2}\,\mathrm{sgn}(\hat{\beta}_{T,j})}{T} = 0\right) \to 0. \tag{47}$$

□

1.2 Proof of Theorem 2

For ease of exposition, we assume that the maximum number of breaks is $m^* = 2$ and the true intercept is known to be $\mu_t = 0$ for all t. In this case, we obtain three possible post-lasso regressions:

1. All coefficients of break indicator regressors are shrunk to zero

$$y_t = \beta_1 x_t + e_{t\tau_0}. \tag{48}$$

2. One structural break is (falsely) detected

$$y_t = \beta_1 x_t + \beta_2 x_t \varphi_{t,\tau_{2,1}} + e_{t\tau_1}. \tag{49}$$

3. Two structural breaks are (falsely) detected

$$y_t = \beta_1 x_t + \beta_2 x_t \varphi_{t,\tau_{2,1}} + \beta_2 x_t \varphi_{t,\tau_{2,2}} + e_{t\tau_2}. \tag{50}$$

The probability of each model being selected depends on the value of the tuning parameter λ_T. We continue the proof for the bias-corrected test statistic, Z_2, corresponding to the case of two (falsely) detected structural breaks. The asymptotic distribution of the test statistics for the two remaining cases can be easily deduced from our derivations. We decompose the cumulative sum into $S_t = (S_{1t}, S_{2t})'$. Further, we define the break fraction vector $\tau = (\tau_{2,1}, \tau_{2,2})'$ as a compact set on $(0,1) \times (0,1)$ and define the matrix $X_{t\tau} = (S_t', S_{2t}\varphi_{t,\tau_{2,1}}, S_{2t}\varphi_{t,\tau_{2,2}})' = (S_{1t}, X_{2t\tau}')'$.

Using the result

$$T^{-1/2} S_{[\tau_{2,i}T]} \Rightarrow B(\tau_{2,i}) \tag{51}$$

and the CMT yields the weak convergence of

$$\frac{1}{T^2} \sum_{t=[\tau_{2,i}T]}^{T} S_t S_t' = \int_{\tau_{2,i}}^{1} BB'. \tag{52}$$

Further, Equation (51) and Theorem 4.1 of Hansen (1992) yield the weak convergence of

$$\frac{1}{T} \sum_{t=[\tau_{2,i}T]}^{T} S_{t-1} u_t' = \int_{\tau_{2,i}}^{1} BdB' + (1 - \tau_{2,i})\Lambda. \tag{53}$$

The result in Equation (52) can straightforwardly be extended to

$$\frac{1}{T^2} \sum_{t=1}^{T} X_{t\tau} X_{t\tau}' = \int_{0}^{1} X_\tau X_\tau', \tag{54}$$

where $X_\tau = (B_1, B_2, B_2\varphi_{\tau_{2,1}}, B_2\varphi_{\tau_{2,2}})' = (B_1, X_{2\tau}')'$ and

$$\varphi_{\tau_{2,i}}(s) = \begin{cases} 0 & \text{if } s < \tau_{2,i} \\ 1 & \text{if } s \geq \tau_{2,i} \end{cases}, \qquad s \in [0,1]. \tag{55}$$

Define $\hat{\beta}_\tau = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)'$ as the post-lasso least squares estimator and set $\hat{\eta}_\tau = (1, -\hat{\beta}_\tau')'$ so that

$$\hat{\eta}_\tau \Rightarrow \begin{bmatrix} 1 \\ -\left(\int_0^1 X_{2\tau}X_{2\tau}'\right)^{-1} \int_0^1 X_{2\tau}B_1 \end{bmatrix} = \eta_\tau. \tag{56}$$

We partition

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \qquad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \tag{57}$$

in conformity with S_t and define $\Lambda_{2\cdot} = (\Lambda_{21}, \Lambda_{22})$ and $\Lambda_{\cdot 2} = (\Lambda_{12}, \Lambda_{22})'$.

For each element of $S_{2t\tau} = (S_{2t}\varphi_{t,\tau_{2,1}}, S_{2t}\varphi_{t,\tau_{2,2}})$ it holds that

$$\Delta S_{2t}\varphi_{t,\tau_{2,i}} = \Delta S_{2t}\varphi_{t,\tau_{2,i}} + S_{2t-1}\Delta\varphi_{t,\tau_{2,i}} \tag{58}$$

and $\Delta\varphi_{t,\tau_{2,i}} = \varphi_{t,\tau_{2,i}} - \varphi_{t-1,\tau_{2,i}} = \mathbb{1}\{t = [\tau_{2,i}T]\}$. Since $\varphi_{t-1,\tau_{2,i}}\Delta\varphi_{t,\tau_{2,i}} = 0$ and

$$d(B_2(s)\varphi_{\tau_{2,i}}(s)) = dB_2(s)\varphi_{\tau_{2,i}}(s) + d\varphi_{\tau_{2,i}}(s)B_2(s) \tag{59}$$

for the asymptotic counterpart, we have the identities

$$\int_0^1 BdB_2(s)\varphi_{\tau_{2,i}}(s) = \int_{\tau_{2,i}}^1 BdB_2(s) + B(\tau_{2,i})B_2(\tau_{2,i}) \tag{60}$$

and

$$\int_0^1 B_2\varphi_{\tau_{2,i}}dB_2(s)\varphi_{\tau_{2,i}}(s) = \int_{\tau_{2,i}}^1 B_2\varphi_{\tau_{2,i}}dB_2(s) = \int_{\tau_{2,i}}^1 B_2dB_2(s). \tag{61}$$

Consequently, we can state the following important weak convergence results

$$\frac{1}{T}\sum_{t=2}^T X_{t-1\tau}\Delta S_{2t\tau}' \;\Rightarrow\; \int_0^1 X_\tau dB_{2\tau} + \begin{bmatrix} (1-\tau_{2,1})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{12} \\ (1-\tau_{2,1})\Lambda_{22} & (1-\tau_{2,2})\Lambda_{22} \\ (1-\tau_{2,1})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{22} \\ (1-\tau_{2,2})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{22}, \end{bmatrix} \tag{62}$$

where $dB_{2\tau} = (d(B_2(s)\varphi_{\tau_{2,1}}(s)), d(B_2(s)\varphi_{\tau_{2,2}}(s)))$ and

$$\frac{1}{T}\sum_{t=2}^T X_{t-1\tau}\Delta X_{t\tau}' \Rightarrow \int_0^1 X_\tau dX_\tau' + \Lambda_\tau, \tag{63}$$

where

$$\Lambda_\tau = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & (1-\tau_{2,1})\Lambda_{12} & (1-\tau_{2,2})\Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} & (1-\tau_{2,1})\Lambda_{22} & (1-\tau_{2,2})\Lambda_{22} \\ (1-\tau_{2,1})\Lambda_{12} & (1-\tau_{2,1})\Lambda_{22} & (1-\tau_{2,1})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{22} \\ (1-\tau_{2,2})\Lambda_{12} & (1-\tau_{2,2})\Lambda_{22} & (1-\tau_{2,2})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{22} \end{bmatrix}. \tag{64}$$

Under the null hypothesis, the cointegration residuals can be written as $\hat{e}_{t\tau_2} = \hat{\eta}_\tau' X_{t\tau}$ and we can show weak convergence of the sample moments. It holds that

$$\frac{1}{T^2}\sum_{t=1}^{T} \hat{e}_{t\tau_2}^2 = \hat{\eta}_\tau' \frac{1}{T^2}\sum_{t=1}^{T} X_{t\tau}X_{t\tau}'\hat{\eta}_\tau \tag{65}$$

$$\Rightarrow \eta_\tau' \int_0^1 X_\tau X_\tau' \eta_\tau = \sigma^2 \int_0^1 W_{\tau_2}^2, \tag{66}$$

where $W_{\tau_2}(s) = W_1(s) - \left(\int_0^1 W_1 W_{2\tau_2}'\right)\left(\int_0^1 W_{2\tau_2} W_{2\tau_2}'\right)^{-1} W_{2\tau_2}$ and

$$\frac{1}{T}\sum_{t=2}^{T} \hat{e}_{t-1\tau_2}\Delta\hat{e}_{t\tau_2} = \hat{\eta}_\tau' \frac{1}{T}\sum_{t=2}^{T} X_{t-1\tau}\Delta X_{t\tau}'\hat{\eta}_\tau \tag{67}$$

$$\Rightarrow \eta_\tau' \left(\int_0^1 X_\tau dX_\tau' + \Lambda_\tau\right)\eta_\tau = \sigma^2 \int_0^1 W_{\tau_2} W_{\tau_2}' + \eta_\tau' \Lambda_\tau \eta_\tau. \tag{68}$$

Next, we consider the bias-correction term for the first-order serial correlation coefficient. We denote the kernel weights as $w(j/M) = w_j$ and can show that

$$\hat{\psi}_\tau = \sum_{j=1}^{M} w_j \frac{1}{T}\sum_t \Delta\hat{e}_{t-j\tau_2}\Delta\hat{e}_{t\tau_2} + o_p(1). \tag{69}$$

Hence, we have the weak convergence result

$$\hat{\psi}_\tau = \hat{\eta}_\tau' \sum_{j=1}^{M} w_j \frac{1}{T}\sum_t \Delta X_{t-j\tau}\Delta X_{t\tau}\hat{\eta}_\tau + o_p(1) \Rightarrow \eta_\tau' \Lambda_\tau \eta_\tau. \tag{70}$$

For the long-run variance, we obtain the result

$$\hat{\sigma}_\tau^2 \Rightarrow \eta_\tau' \Omega_\tau \eta_\tau, \tag{71}$$

where

$$\Omega_\tau = \begin{bmatrix} \sigma^2 & \Omega_{12} & (1-\tau_{2,1})\Omega_{12} & (1-\tau_{2,2})\Omega_{12} \\ \Omega_{21} & \Omega_{22} & (1-\tau_{2,1})\Omega_{22} & (1-\tau_{2,2})\Omega_{22} \\ (1-\tau_{2,1})\Omega_{12} & (1-\tau_{2,1})\Omega_{22} & (1-\tau_{2,1})\Omega_{21} & (1-\tau_{2,2})\Omega_{22} \\ (1-\tau_{2,2})\Omega_{12} & (1-\tau_{2,2})\Omega_{22} & (1-\tau_{2,2})\Omega_{21} & (1-\tau_{2,2})\Omega_{22} \end{bmatrix} \tag{72}$$

$$= \begin{bmatrix} 1, -\kappa'_{\tau_2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D_{\tau_2} \end{bmatrix} \begin{bmatrix} 1 \\ -\kappa_{\tau_2} \end{bmatrix} = \sigma^2(1 + \kappa'_{\tau_2} D_{\tau_2} \kappa_{\tau_2}) \tag{73}$$

and

$$D_{\tau_2} = \begin{bmatrix} 1 & (1-\tau_{2,1}) & (1-\tau_{2,2}) \\ (1-\tau_{2,1}) & (1-\tau_{2,1}) & (1-\tau_{2,2}) \\ (1-\tau_{2,2}) & (1-\tau_{2,2}) & (1-\tau_{2,2}) \end{bmatrix}. \tag{74}$$

Now, we use the CMT to show that

$$Z_2 = \frac{\frac{1}{T^2}\sum_{t=2}^{T} \hat{e}_{t-1\tau_2}\Delta\hat{e}_{t\tau_2} - \hat{\psi}_\tau}{\frac{1}{T^2}\sum_{t=2}^{T} \hat{e}^2_{t-1\tau_2}} \times \left(\frac{1}{\hat{\sigma}^2_\tau T^2}\sum_{t=2}^{T} \hat{e}^2_{t-1\tau_2}\right)^{1/2} \tag{75}$$

$$\Rightarrow \frac{\sigma^2 \int_0^1 W_{\tau_2} dW_{\tau_2} + \eta'_\tau \Lambda_\tau \eta_\tau - \eta'_\tau \Lambda_\tau \eta_\tau}{\sigma^2 \int_0^1 W^2_{\tau_2}} \times \left(\frac{1}{\sigma^2(1 + \kappa'_{\tau_2} D_{\tau_2} \kappa_{\tau_2})}\sigma^2 \int_0^1 W^2_{\tau_2}\right)^{1/2} \tag{76}$$

$$= \frac{\int_0^1 W_{\tau_2} dW_{\tau_2}}{\left(\int_0^1 W^2_{\tau_2}\right)^{1/2} \left(1 + \kappa'_{\tau_2} D_{\tau_2} \kappa_{\tau_2}\right)^{1/2}} \tag{77}$$

Correspondingly, the remaining components of our test statistic have the asymptotic distributions

$$Z_1 \sim \int_0^1 W_{\tau_1} dW_{\tau_1} \Big/ \left(\int_0^1 W^2_{\tau_1}\right)^{1/2} \left(1 + \kappa'_{\tau_1} D_{\tau_1} \kappa_{\tau_1}\right)^{1/2}, \tag{78}$$

$$W_{\tau_1} = W_1(s) - \left[\int_0^1 W_1 W_{2\tau_1}\right] \left[\int_0^1 W_{2\tau_1} W'_{2\tau_1}\right]^{-1} W_{2\tau_1}(s),$$

$$\kappa_{\tau_1} = \left[\int_0^1 W_{2\tau_1} W'_{2\tau_1}\right]^{-1} \left[\int_0^1 W_{2\tau_1} W_1\right],$$

$$W_{2\tau_1} = \Big[W_2(s), W_2(s)\varphi_{\tau_{2,1}}(s) \Big],$$

and

$$Z_0 \sim \int_0^1 W_{\tau_0} dW_{\tau_0} \Bigg/ \left(\int_0^1 W_{\tau_0}^2 \right)^{1/2}, \tag{79}$$

$$W_{\tau_0} = W_1(s) - \left[\int_0^1 W_1 W_2 \right] \left[\int_0^1 W_2^2 \right]^{-1} W_2(s),$$

respectively. Finally, the model selection procedure yields a mixture distribution depending on the tuning parameter λ. □

References

Arai, Y., Kurozumi, E., 2007. Testing for the Null Hypothesis of Cointegration with a Structural Break. Econometric Reviews 26 (6), 705–739.

Aue, A., Horváth, L., 2013. Structural breaks in time series. Journal of Time Series Analysis 34 (1), 1–16.

Bai, J., Perron, P., 1998. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica 66 (1), 47–78.

Belloni, A., Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. Bernoulli 19 (2), 521–547.

Bierens, H. J., Martins, L. F., 2010. Time-Varying Cointegration. Econometric Theory 26 (5), 1453–1490.

Billingsley, P., 1999. Convergence of Probability Measures, 2nd Edition. Wiley, New York.

Carrion-i Silvestre, J. L., Sanso, A., 2006. Testing for the Null Hypothesis of Cointegration with Structural Breaks. Oxford Bulletin of Economics and Statistics 68 (5), 623–646.

Chang, Y., Park, J. Y., 2002. On the Asymptotics of ADF Tests for Unit Roots. Econometric Reviews 21 (4), 431–447.

Craighead, W. D., 2010. Across time and regimes: 212 Years of the US-UK real exchange rate. Economic Inquiry 48 (4), 951–964.

Davidson, J., Monticini, A., 2010. Tests for cointegration with structural breaks based on subsamples. Computational Statistics and Data Analysis 54 (11), 2498–2511.

Engel, C., Kim, C.-J., 1999. The Long-Run U.S./U.K. Real Exchange Rate. Journal of Money, Credit and Banking 31 (3), 335–356.

Engle, R. F., Granger, C. W. J., 1987. Co-Integration and Error Correction: Representation, Estimation and Testing. Econometrica 55 (2), 251–276.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96 (456), 1348–1360.

Gregory, A. W., Hansen, B. E., 1996. Residual-based tests for cointegration in models with regime shifts. Journal of Econometrics 70 (2), 99–126.

Gregory, A. W., Nason, J. M., Watt, D. G., 1996. Testing for structural breaks in cointegrated relationships. Journal of Econometrics 71 (1-2), 321–341.

Grilli, V., Kaminsky, G., 1991. Nominal exchange rate regimes and the real exchange rate - Evidence from the United States and Great Britain, 1885-1986. Journal of Monetary Economics 27 (2), 191–212.

Hall, P., Heyde, C. C., 1980. Martingale Limit Theory and Its Application. Academic Press.

Hansen, B. E., 1992. Convergence to Stochastic Integrals for Dependent Heterogeneous Processes. Econometric Theory 8 (4), 489–500.

Hatemi-J, A., 2008. Tests for cointegration with two unknown regime shifts with an application to financial market integration. Empirical Economics 35 (3), 497–505.

Horowitz, J., Huang, J., 2013. Penalized estimation of high-dimensional models under a generalized sparsity condition. Statistica Sinica 23 (2), 725–748.

Karoglou, M., Morley, B., 2012. Purchasing power parity and structural instability in the US/UK exchange rate. Journal of International Financial Markets, Institutions and Money 22 (4), 958–972.

Kejriwal, M., Perron, P., 2010. Testing for Multiple Structural Changes in Cointegrated Regression Models. Journal of Business & Economic Statistics 28 (4), 503–522.

Knight, K., Fu, W., 2000. Asymptotics for Lasso-Type Estimators. The Annals of Statistics 28 (5), 1356–1378.

Kock, A. B., 2016. Consistent and Conservative Model Selection With the Adaptive Lasso in Stationary and Nonstationary Autoregressions. Econometric Theory 32 (1), 243–259.

Koo, B., Anderson, H., Seo, M. H., Yao, W., 2017. High-dimensional predictive regression in the presence of cointegration. Tech. rep., SSRN Working Paper.

Kuo, B.-S., 1998. Test for partial parameter instability in regressions with I(1) processes. Journal of Economics 86, 337–368.

Kurozumi, E., Skrobotov, A., 2017. Confidence Sets for the Break Date in Cointegrating Regressions. Oxford Bulletin of Economics and Statistics 80 (3), 514–535.

Liao, Z., Phillips, P. C. B., 2015. Automated Estimation of Vector Error Correction Models. Econometric Theory 31 (3), 581–646.

Maki, D., 2012. Tests for cointegration allowing for an unknown number of breaks. Economic Modelling 29 (5), 2011–2015.

Mendes, E. F., 2011. Model Selection Consistency for Cointegrating Regressions. Tech. rep., Northwestern University.

Mussa, M., 1986. Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates: Evidence and Implications. Carnegie-Rochester Conference on Public Policy 25, 117–214.

Park, J. Y., Hahn, S. B., 1999. Cointegrating Regressions with Time Varying Coefficients. Econometric Theory 15 (5), 664–703.

Perron, P., 2006. Dealing with structural breaks. In: Hassani, H., Mills, T., Patterson, K. (Eds.), Palgrave Handbook of Econometrics - Volume 1: Econometric Theory. Palgrave Macmillan UK, pp. 278–352.

Perron, P., Yamamoto, Y., 2016. On the Usefulness or Lack Thereof of Optimality Criteria for Structural Change Tests. Econometric Reviews 35 (5), 782–844.

Phillips, P. C. B., 1987. Time Series Regression with a Unit Root. Econometrica 55 (2), 277–301.

Phillips, P. C. B., Ouliaris, S., 1990. Asymptotic Properties of Residual Based Tests for Cointegration. Econometrica 58 (1), 165–193.

Pötscher, B. M., Schneider, U., 2009. On the distribution of the adaptive LASSO estimator. Journal of Statistical Planning and Inference 139 (8), 2775–2790.

Qian, J., Su, L., 2016. Shrinkage Estimation of Regression Models With Multiple Structural Changes. Econometric Theory 32 (6), 1376–1433.

Qu, Z., Perron, P., 2007. Estimating and testing structural change in multivariate regressions. Econometrica 75 (2), 459–502.

Quintos, C. E., Phillips, P. C. B., 1993. Parameter Constancy in Cointegrating Regressions. Empirical Economics 18 (4), 675–706.

Taylor, A. M., Taylor, M. P., 2004. The Purchasing Power Parity Debate. The Journal of Economic Perspectives 18 (4), 135–158.

Taylor, M. P., 2006. Real exchange rates and Purchasing Power Parity: mean-reversion in economic thought. Applied Financial Economics 16 (1-2), 1–17.

Tibshirani, R., 1996. Regression Selection and Shrinkage via the Lasso. Journal of the Royal Statistical Society B (Methodological) 58 (1), 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K., 2005. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society. Series B: Statistical Methodology 67 (1), 91–108.

Wang, H., Li, B., Leng, C., 2009. Shrinkage tuning parameter selection with a diverging number of parameters. Journal of the Royal Statistical Society. Series B: Statistical Methodology 71 (3), 671–683.

Westerlund, J., Edgerton, D. L., 2006. New Improved Tests for Cointegration with Structural Breaks. Journal of Time Series Analysis 28 (2), 188–224.

Wilms, I., Croux, C., 2016. Forecasting using sparse cointegration. International Journal of Forecasting 32 (4), 1256–1267.

Xiao, Z., 2009a. Functional-coefficient cointegration models. Journal of Econometrics 152 (2), 81–92.

Xiao, Z., 2009b. Quantile cointegrating regression. Journal of Econometrics 150 (2), 248–260.

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 (1), 49–67.

Zhang, C. H., Huang, J., 2008. The sparsity and bias of the lasso selection in high-dimensional linear regression. Annals of Statistics 36 (4), 1567–1594.

Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101 (476), 1418–1429.