Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ubes20

Variable Selection Based Testing for Parameter Changes in Regression with Autoregressive Dependence

Lajos Horváth, Piotr Kokoszka & Shanglin Lu

To cite this article: Lajos Horváth, Piotr Kokoszka & Shanglin Lu (2024) Variable Selection Based Testing for Parameter Changes in Regression with Autoregressive Dependence, Journal of Business & Economic Statistics, 42:4, 1331-1343, DOI: 10.1080/07350015.2024.2310025

To link to this article: https://doi.org/10.1080/07350015.2024.2310025

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2024, VOL. 42, NO. 4, 1331–1343
https://doi.org/10.1080/07350015.2024.2310025

Variable Selection Based Testing for Parameter Changes in Regression with Autoregressive Dependence

Lajos Horváth[a], Piotr Kokoszka[b], and Shanglin Lu[c]

[a]Department of Mathematics, University of Utah, Salt Lake City, UT; [b]Department of Statistics, Colorado State University, Fort Collins, CO; [c]China School of Banking and Finance, University of International Business and Economics, Beijing, China

ABSTRACT

We consider a regression model with autoregressive terms and propose significance tests for the detection of change points in this model. Our tests are applicable to both low- or moderate dimension and to high-dimension with sparse regressors. The dimension may be high from the practical point of view of economic and business applications, but in our theoretical framework it is fixed. To accommodate practically high dimension, variable selection is incorporated as an integral part of our approach. The regressors and the errors can exhibit general nonlinear dependence and the model incorporates autoregressive dependence. We develop asymptotic justification and evaluate the performance of the tests both on simulated and real economic data. We test for and estimate changes in responses to risk factors of a U.S. energy stocks portfolio and the Industrial Production index. We relate our findings to macroeconomic policy changes and global impact events.

ARTICLE HISTORY
Received January 2023
Accepted January 2024

KEYWORDS
Change points; Significance test; Time series regression; Variable selection

1. Introduction

We propose and study change point tests for a regression model with scalar responses y_i and regressors that can be moderate- or high-dimensional. The term high-dimension is used in a practical sense of a large number of potential explanatory variables, not in the sense adopted in asymptotic theory that requires the dimension to tend to infinity with the sample size. The regressors can also include autoregressive terms $y_{i-1}, \ldots, y_{i-d}$. The inclusion of these terms is relevant in econometric applications, as pointed out already by Lo (2004) and further elaborated in Lo (2019). We call such a model a dynamic linear model because of the dependence on lagged values of the responses. It is related to dynamic factor models, discussed below, but we consider the usual statistical regression setting in which the regressors are observed and do not need to be disentangled from the coefficients. Using a shrinkage approach, we focus on identifying the most relevant regressors (real risk factors) rather than finding PCA-like transformations to reduce dimension. The precise model formulation is given in Section 2. The importance of accounting for structural breaks, or change-points, in economic and financial data has long been recognized, see for example Andreou and Ghysels (2009), Baek (2023), and Zhu et al. (2023) are very recent contributions. The importance of high dimensionality in economic and financial time series is also well-known, see for example Fan, Lv, and Qi (2011), and there is a rapidly growing research in this direction, see for example Bodnar, Okhrin, and Parolya (2023) and Trucios et al. (2023), among dozens of papers published every year.

Even though our work is most directly motivated by financial and economic data, it is a contribution related to the growing body of statistics research on change point detection in high-dimensional models. The most directly related recent papers are Wang et al. (2022) and Wang and Zhao (2022). Wang et al. (2022) study change point detection in the mean of a high-dimensional sequence with temporal dependence. The difference between the target models is that we consider changes in the regression coefficients of a dynamic regression model in a sparse setting, while Wang et al. (2022) target dense alternatives and consider changes in mean. The methods and their large sample justifications are consequently quite different. Wang and Zhao (2022) study basically the same problem as this article, but without incorporating the autoregressive terms. Their approach is quite different though. They explicitly incorporate sparsity as well as dimension increasing to infinity, and use randomization to eliminate temporal dependence. Wang and Zhao (2022) establish a number of interesting mini-max optimality results. We note that our approach also applies to low or moderate dimension. Sequential change point monitoring for a low dimensional version of the model considered here is studied in Horváth, Liu, and Lu (2022). The test statistic is based on a different process, residuals are not weighted by observations, and the asymptotic theory is quite different than for the historical sample test we propose here. Wang et al. (2022) provide an excellent review of research on change point detection in high-dimensional settings, so we only note influential papers of Jirak (2015), Wang and Samworth (2018), Enikeeva and Harchaoui (2019), Kaul et al. (2019, 2020), Fang, Li, and Siegmund (2020), as well as more recent work that

CONTACT Piotr Kokoszka Piotr.Kokoszka@colostate.edu Department of Statistics, Colorado State University, Fort Collins, CO 80521-1877.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

includes Wang et al. (2022), Dette, Pan, and Yang (2022), Xu et al. (2022), Cho and Owens (2022) and Zhang, Wang, and Shao (2023).

Our work is related to a large body of econometrics research on factor models that considers the evaluation of the effects of change points, testing for them and estimating them. Chen, Dolado, and Gonzalo (2014) show that unaccounted for "big breaks" defined by the matrix Λ in their decomposition (4), can severely impact the number of estimated factors, and principal component analysis in general, as well as forecasting. Chen, Dolado, and Gonzalo (2014) also propose Lagrange multiplier and Wald change point tests in which the alternative is that there is one "big break". Han and Inoue (2015) also consider Wald and LM change point test whose alternative is that the loading of a specified group of factors changes. Chen, Dolado, and Gonzalo (2016) focus on a different but related problem that can be summarized as follows. Suppose time of a break in the factors structure is known and consider the model $\mathcal{M} = (r_1, r_2, B)$, where r_1 is the count of factors before the break, r_2 after the break and B is a binary variable equal to 0 if there is no change and to 1 otherwise. If $B = 1$ and if $r_1 = r_2$, the change is in factor loadings; if $r_1 \neq r_2$ the change is in the number of factors. Using PCA and shrinkage, estimators of r_1 and r_2 are derived. The behavior of the method if the break time is unknown is studied, but a significance test for the presence of the break is not considered; the article focuses on identifying the factors and loadings and providing information about the structure of the model before and after a break whose timing may be misspecified. A piecewise stationary high-dimensional factor model is considered by Barigozzi, Cho, and Fryzlewicz (2018) who propose a PCA and wavelet based methodology for the estimation of its multiple change-points. All these papers provide numerous references to earlier work going back at least 20 years. Generally speaking, econometric factor models involve some form dimension reduction through a method similar to PCA, see for example Uematsu and Yamagata (2023) for a recent contribution.

Our approach targets detection of change points without any transformation of variables. Our objective is to propose a relatively simple method that can be broadly applied using readily available statistical software. We however take care to clearly specify under what assumptions this method is valid. This requires listing mathematical assumption, which however, in our opinion, will hold for most economic time series after some standard transformation to stationarity. The scope of validity of our method, and its limitations, are additionally investigated through simulation studies. We also validate it by application to economic moderate- and high-dimensional time series. The change points we detect can be related to relatively well-known macroeconomic policy changes and global impact events.

The remainder of the article is organized as follows. In Section 2, we formulate the model and state the assumption we impose on it. Test procedures and their asymptotic justification are explained in Section 3. A simulation study in Section 4 is followed by two data examples in Section 5. Online supplementary material contains the proofs of the results of Section 3, additional graphs and a table, and well as the data availability statement.

2. Model Formulation and Assumptions

We consider the scalar response regression

$$y_i = \sum_{\ell=0}^{R} \tilde{\mathbf{x}}_i^\top I\left\{k_\ell^* < i \le k_{\ell+1}^*\right\}(\tilde{\boldsymbol{\beta}}_0 + \tilde{\boldsymbol{\delta}}_\ell) + \epsilon_i, \qquad 1 \le i \le N, \tag{2.1}$$

where $0 < k_1^* < k_2^* < \cdots < k_R^* < N$ denote the unknown times of potential changes and $\tilde{\boldsymbol{\delta}}_1 \neq \tilde{\boldsymbol{\delta}}_2 \neq \cdots \neq \tilde{\boldsymbol{\delta}}_R$ are fixed. sizes of the changes. We use the convention $k_0^* = 0, k_{R+1}^* = N, \tilde{\boldsymbol{\delta}}_0 = \mathbf{0}$. In model (2.1),

$$\tilde{\mathbf{x}}_i = \left(1, x_{i,2}, \ldots, x_{i,\tilde{r}}, y_{i-1}, \ldots, y_{i-\tilde{d}}\right)^\top, \tag{2.2}$$

so the first $\tilde{r}$ coordinates of $\tilde{\mathbf{x}}_i$ are the auxiliary terms and the last $\tilde{d}$ are the autoregressive terms.

We test the null hypothesis

$$H_0: \quad \text{for each } \ell = 1, 2, \ldots, R, \quad \|\tilde{\boldsymbol{\delta}}_\ell\| = 0$$

against the alternative

$$H_A: \quad \text{there is } \ell = 1, 2, \ldots, R, \quad \text{with } \|\tilde{\boldsymbol{\delta}}_\ell\| > 0.$$

We assume that the $\tilde{\boldsymbol{\delta}}_\ell$ (and $\tilde{\boldsymbol{\beta}}_0$) are unknown parameters that do not depend on the sample size N.

The tilde ˜ in (2.1) and (2.2) indicates that the model involves a potentially large number $\tilde{r}$ of regressors, or even lags $\tilde{d}$, but many of them may be not significant and should not have been in the model. Such a situation arises in econometric and financial problems where often over a hundred factors could have a potential impact on the responses, but only a few of them are sufficient to explain the response variable. In other problems, the dimensions $\tilde{r}$ and $\tilde{d}$ are not large, and it is clear, often from extensive previous research, that all, or most, of them should be incorporated in a factor model. We consider both situations in the case studies in Section 5. Our testing methodology covers both settings, by reducing the dimensions $\tilde{r}$ and $\tilde{d}$ to smaller dimensions r and d, if needed. In contrast to extensive previous research on the selection of significant factors, our focus is on change point testing in both situations. The potential dimension reduction is incorporated into the construction of test statistics. In an informal way, we think of the moderate dimensional case if $\tilde{r} + \tilde{d} \approx 10$, with $\tilde{r} \approx r, \tilde{d} \approx d$, and of the sparse high-dimensional case if $\tilde{r} + \tilde{d} \approx 100$, but $r + d \approx 10$. We use of the term "high-dimensional", in an informal, not asymptotic, sense. In both cases, we denote by

$$\mathbf{x}_i = \left(1, x_{i,2}, \ldots, x_{i,r}, y_{i-1}, \ldots, y_{i-d}\right)^\top \tag{2.3}$$

the vectors of regressors in the true regression valid under the null hypothesis of no change points, that is $y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \epsilon_i$, where $\boldsymbol{\beta}_0$ is the vector of the nonzero coordinates of $\tilde{\boldsymbol{\beta}}_0$. The $\mathbf{x}_i$ are the $\tilde{\mathbf{x}}_i$ without the components that have zero coefficients.

We use the LASSO–type estimator $\hat{\boldsymbol{\beta}}_N$ which is the vector of the nonzero coefficients of the minimizer of

$$\mathcal{S}_N(\tilde{\boldsymbol{\beta}}) = \sum_{i=1}^{N}\left(y_i - \tilde{\mathbf{x}}_i^\top \tilde{\boldsymbol{\beta}}\right)^2 + \lambda_N \sum_{j=2}^{\tilde{r}+\tilde{d}} |\tilde{\beta}_j|^\gamma, \tag{2.4}$$

where $\gamma > 0$. The vector $\hat{\boldsymbol{\beta}}_N$ has $\hat{r} + \hat{d}$ nonzero components. Our test is based on the residuals

$$\hat{\epsilon}_i = y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_N, \quad 1 \le i \le N,$$

where $\hat{\mathbf{x}}_i$ is the part of $\tilde{\mathbf{x}}_i$ that corresponds to the coordinates of $\hat{\boldsymbol{\beta}}_N$. Note that if a component of $\hat{\boldsymbol{\beta}}_N$ vanishes, then the corresponding component of $\tilde{\mathbf{x}}_i$ does not enter into the construction of the residuals.

Frank and Friedman (1993) introduced the estimator $\hat{\boldsymbol{\beta}}_N$ with $\gamma = 2$ as the "bridge" estimator, a generalization of the ridge method. The special case of $\gamma = 1$ is related to the "LASSO" estimator of Tibshirani (1996). Our work incorporates autoregressive dependence and very general temporal dependence of the regressors. The large sample justification applies to any $\gamma > 0$, but we work with the most commonly $\gamma = 1$ in our data examples and simulations.

Before formulating the test procedures and their large sample justifications, we list the required assumptions. We formulate them in terms of the regressors (2.3) and the parameter $\boldsymbol{\beta}_0$ decomposed as

$$\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{1,0}^\top, \boldsymbol{\beta}_{2,0}^\top)^\top, \quad \boldsymbol{\beta}_{1,0} \in R^r, \ \boldsymbol{\beta}_{2,0} \in R^d,$$

where $\boldsymbol{\beta}_{1,0}$ and $\boldsymbol{\beta}_{2,0}$ are, respectively, the true regression and the autoregressive parameter vectors in the absence of any change points.

Our first Assumption ensures that under the null hypothesis there is a stationary and causal solution $\{y_i\}$ to (2.1). This assumption is needed only if $d > 0$, that is if there is an autoregressive part. For the explanation of this condition, we refer for example to sec. 3.1 of Brockwell and Davis (1991) or sec. 5.4 of McElroy and Politis (2020).

Assumption 2.1. If $d > 0$, set $\boldsymbol{\beta}_{2,0} = (\beta_{2,0,1}, \beta_{2,0,1}, \ldots, \beta_{2,0,d})$ and assume then that $\beta_{2,0,d} \neq 0$ and the roots of the polynomial $z^d - \beta_{2,0,1} z^{d-1} - \beta_{2,0,2} z^{d-2} - \cdots - \beta_{2,0,d}$ are inside of the unit circle of the complex plane.

In addition to the autoregressive dependence, the exogenous regressors and the errors can exhibit very general linear or nonlinear dependence, but must be stationary. This is quantified in Assumption 2.2.

Assumption 2.2. Setting

$$\mathbf{z}_i = (x_{i,2}, \ldots, x_{i,r}, \epsilon_i)^\top,$$

we assume that

$$\mathbf{z}_i = \mathbf{g}(\eta_i, \eta_{i-1}, \eta_{i-2}, \ldots), \tag{2.5}$$

where $\mathbf{g}$ is a nonrandom measurable function $\mathcal{S}^\infty \to R^r$, $E\|\mathbf{z}_i\|^\nu < \infty$ with some $\nu > 4$, $\{\eta_i, \ -\infty < i < \infty\}$ are independent and identically distributed random variables with values in a measurable space $\mathcal{S}$,

$$\left(E\|\mathbf{z}_i - \mathbf{z}_{i,j}^*\|^\nu\right)^{1/\nu} \le cj^{-\alpha} \quad \text{with some } \ c > 0 \ \text{ and } \ \alpha > 2, \tag{2.6}$$

$\mathbf{z}_{i,j}^* = \mathbf{g}(\eta_i, \eta_{i-1}, \eta_{i-2}, \ldots, \eta_{i-j+1}, \eta_{i-j,i,j}^*, \eta_{i-j,i,j}^*, \ldots)$, $\{\eta_{i,j,\ell}^*, -\infty < i, j, \ell\}$ are independent copies of η_0, independent of $\{\eta_t, -\infty < t < \infty\}$.

Assumptions involving the Bernoulli shift representation (2.5) have been extensively used in recent time series research because (2.5) is a natural and very flexible generalization of linear moving averages. For a specific time series model, verification of moment conditions similar to (2.6) is often easier than the verification of various mixing conditions, like α- or β-mixing, Bradley (2007). The latter have a long tradition, and are routinely imposed in nonparametric inference, so the choice of the quantification of weak dependence is chiefly dictated by the desired mathematical framework. References to papers using assumptions similar to Assumption 2.2 and models that satisfy them are too numerous to be fully cited, so we merely note Wu (2005), Shao and Wu (2007), Hörmann and Kokoszka (2010), Horváth, Kokoszka, and Rice (2014), Zhou (2015), Zhang (2016), Zhang and Cheng (2018), and Dette, Eckle, and Vetter (2020), among many others. We also note that basically all known time series models, including high-dimensional and functional models, satisfy Assumption 2.2. Long-range dependent models with coefficients or correlations decaying like a power function do not satisfy it.

We show that under Assumption 2.2,

$$\frac{1}{N}\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^\top \xrightarrow{P} E\mathbf{x}_0\mathbf{x}_0^\top =: \boldsymbol{A}$$

and we require

Assumption 2.3. The $(r+d) \times (r+d)$ covariance matrix $\boldsymbol{A}$ is nonsingular.

Assumption 2.4. The regressors and the errors are uncorrelated, that is $E\mathbf{x}_0\epsilon_0 = \mathbf{0}$.

Assumption 2.2 implies that

$$\mathbf{D} = \sum_{\ell=-\infty}^{\infty} E\mathbf{x}_0\mathbf{x}_\ell^\top \epsilon_0\epsilon_\ell$$

exists as an absolutely convergent sum. We require that

Assumption 2.5. The $(r+d) \times (r+d)$ matrix $\mathbf{D}$ is nonsingular.

Our tests use the weight function $w(t) = (t(1-t))^\kappa, 0 < \kappa \le 1/2$, but they are justified for a larger class of weight functions that satisfy the following assumption.

Assumption 2.6. The weight function $w(t), t \in (0,1)$ satisfies:
(i) $\inf_{\delta \le t \le 1-\delta} w(t) > 0$ for all $0 < \delta < 1/2$
(ii) $w(t)$ is nondecreasing in a neighborhood of 0
(iii) $w(t)$ is nonincreasing in a neighborhood of 1.

The existence of the asymptotic distribution of our test statistic is determined by the integral functional

$$I(w,c) = \int_0^1 \frac{1}{t(1-t)} \exp\left(-\frac{cw^2(t)}{t(1-t)}\right) dt. \tag{2.7}$$

We impose only minimal conditions on the smoothing parameter and the parameter space.

Assumption 2.7. The smoothing parameter in (2.4) satisfies $\lambda_N \to 0$ and the true parameter vector is in the interior of a compact $\mathcal{K} \subset R^{r+d}$.

3. Test Procedures

Our test statistic is based on the weighted supremum of the process

$$\mathbf{U}_N(t) = N^{-1/2} \left(\sum_{i=1}^{\lfloor (N+1)t \rfloor} \hat{\mathbf{x}}_i \hat{\epsilon}_i - \frac{\lfloor (N+1)t \rfloor}{N} \sum_{i=1}^{N} \hat{\mathbf{x}}_i \hat{\epsilon}_i \right), \quad t \in [0,1], \quad (3.1)$$

where the $\hat{\mathbf{x}}_i$ retain only the selected coordinates with nonzero coefficients and have dimension $\hat{r}+\hat{d}$. The proof of Theorem 3.1 in Section A of the supplementary material shows that $\hat{r} = r$ and $\hat{d} = d$ with probability approaching 1. The process $\mathbf{U}_N(\cdot)$, and therefore the tests based on its functionals, depend only on the statistically significant part of the regression. This is similar in sprit to the ideas of Candes and Tao (2007) and Gareth and Radchenko (2009), but our context and the asymptotic framework are different.

The residuals $\hat{\epsilon}_i$ are weighted by the regressors $\mathbf{x}_i$. This is of practical advantage because, as we will see, the weighted residuals $\mathbf{x}_i\hat{\epsilon}_i$ give tests that are always consistent as long as Assumption 2.3 holds, which is a very weak assumption. For tests based on the CUSUM process of the $\hat{\epsilon}_i$, additional assumptions on the $\mathbf{x}_i$ are needed, see for example Horváth, Liu, and Lu (2022) and references therein. Even in the case of iid ϵ_i and no autoregressive part, tests based on the residuals $\hat{\epsilon}_i$ require stronger conditions, see chap. 4 of Csörgő and Horváth (1997). Observe that according to Assumption 2.2, even the unobservable vectors $\mathbf{x}_i\epsilon_i$ may exhibit complex nonlinear dependence. This makes the theoretical analysis of the tests based on the process $\mathbf{U}_N(\cdot)$ more complex, but justifies them under weaker assumptions (very general, nonlinear dependence quantified by Assumption 2.2.)

The following result provides a justification for a broad family of tests valid both in the moderate and the sparse high-dimensional cases.

Theorem 3.1. If H_0 and Assumptions 2.1–2.7 are satisfied, and $I(w,c) < \infty$ with some $c > 0$, then

$$\sup_{0<t<1} \frac{1}{w(t)} \left(\mathbf{U}_N^\top(t) \mathbf{D}^{-1} \mathbf{U}_N(t) \right)^{1/2} \xrightarrow{\mathcal{D}} \sup_{0<t<1} \frac{1}{w(t)} \left(\sum_{i=1}^{r+d} B_i^2(t) \right)^{1/2}, \quad (3.2)$$

where $\{B_1(t), 0 \le t \le 1\}, \{B_2(t), 0 \le t \le 1\}, \ldots, \{B_{r+d}(t), 0 \le t \le 1\}$ are independent Brownian bridges.

The critical values of the asymptotic distribution in (3.2) can be simulated with arbitrary precision, once the combined dimension $m = r + d$ has been estimated. In the moderate case, m can be taken as the model dimension. More details on the simulation of the RHS of (3.2) are provided in Section 4.

As will be apparent from the proof of Theorem 3.1, the condition $I(w,c) < \infty$ is optimal since the limit in (3.2) exists if and only if $I(w,c) < \infty$ with some $c > 0$. Hence Theorem 3.1 does not cover the function $w(t) = (t(1-t))^{1/2}$ that defines the "self normalized" statistic, for which the expected value of $\mathbf{U}_N^\top(t)\mathbf{D}^{-1}\mathbf{U}_N(t)/(t(1-t))$ is asymptotically $r+d$, for every fixed $0 < t < 1$. It follows from the Law of the Iterated Logarithm for the Wiener process that for $w(t) = (t(1-t))^{1/2}$ the supremum on the right-hand side of (3.2) is infinity. This necessitates truncation to a compact subinterval of $(0,1)$. Such an approach, made popular following the work of Andrews (1993), can be justified within our framework. Minimally changing our proofs, one can show that for any $\delta \in (0, 1/2)$,

$$\sup_{\delta<t<1-\delta} \frac{1}{\sqrt{t(1-t)}} \left(\mathbf{U}_N^\top(t) \mathbf{D}^{-1} \mathbf{U}_N(t) \right)^{1/2} \xrightarrow{\mathcal{D}} \sup_{\delta<t<1-\delta} \frac{1}{\sqrt{t(1-t)}} \left(\sum_{i=1}^{r+d} B_i^2(t) \right)^{1/2}.$$

If one does not wish to exclude the possibility of a change point before time $N\delta$ or after $(1-\delta)N$, a different approximation for the supremum of the "self normalized" statistic must be used. The convergence to the Gumbel distribution in Theorem 3.2 is known as Darling–Erdős type limit result. In our setting, put

$$a(x) = (2\log x)^{1/2} \quad \text{and} \quad b_{r+d} = 2\log x + \frac{r+d}{2} \log \log x - \log \Gamma((r+d)/2),$$

where $\Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx$, $u > 0$ is the Gamma function, and define the statistic

$$G_N = \sup_{0<t<1} \frac{1}{(t(1-t))^{1/2}} \left(\mathbf{U}_N^\top(t) \mathbf{D}^{-1} \mathbf{U}_N(t) \right)^{1/2}. \quad (3.3)$$

Note that G_N is well defined because $\mathbf{U}_N(t) = 0$ if $t < 1/(N+1)$ or $t > 1 - 1/(N+1)$.

Theorem 3.2. If H_0 and Assumptions 2.1–2.5 are satisfied, then for all x,

$$\lim_{N\to\infty} P\left\{ a(\log N) G_N \le x + b_{r+d}(\log N) \right\} = \exp\left(-2e^{-x}\right).$$

It is known that the rate convergence to an extreme value distribution can be slow, so other type of approximations also have been suggested. In Section A of the supplementary material we derive an effective approximation.

To apply the above tests, the long run covariance matric $\mathbf{D}$ must be estimated. Let

$$\hat{\boldsymbol{\gamma}}_N(\ell) = \begin{cases} \dfrac{1}{N-\ell} \displaystyle\sum_{i=1}^{N-\ell} \mathbf{x}_i \hat{\epsilon}_i \mathbf{x}_{i+\ell}^\top \hat{\epsilon}_{i+\ell}, & \text{if } 0 \le \ell < N, \\ \dfrac{1}{N-|\ell|} \displaystyle\sum_{i=-(\ell-1)}^{N} \mathbf{x}_i \hat{\epsilon}_i \mathbf{x}_{i+\ell}^\top \hat{\epsilon}_{i+\ell}, & \text{if } -N < \ell < 0. \end{cases}$$

The kernel long run covariance estimator is

$$\widehat{\mathbf{D}} = \sum_{\ell=-(N-1)}^{N-1} K\left(\frac{\ell}{h}\right) \hat{\boldsymbol{\gamma}}_N(\ell),$$

where K is the kernel and $h = h(N)$ is the window length. The consistency of $\widehat{\mathbf{D}}$ requires the following standard conditions:

Assumption 3.1. (i) $h = h(N) \to \infty$, (ii) $h/N \to 0$.

Assumption 3.2. (i) $K(0) = 1$,
(ii) $K(u) = K(-u)$,
(iii) there is $c > 0$ such that $K(u) = 0$, if $u \notin [-c, c]$,
(iv) $\sup_{-c<u<c} |K(u)| < \infty$,
(v) $K(u)$ is Lipschitz continuous.

If Assumptions 2.1–2.4, 3.1, and 3.2 hold, then, under the null hypothesis of no change points,

$$\left\|\widehat{\mathbf{D}} - \mathbf{D}\right\| = o_P(1). \tag{3.4}$$

The proof of (3.4) can be based on natural modifications of the arguments used in Liu and Wu (2010). Kernel estimators for variances, long run variances, covariance matrices as well as long run covariance matrices are not consistent under the alternative even in simple cases, like changes in the mean, see for example Horváth, Kokoszka, and Reeder (2013) and Horváth and Rice (2014). However, one can show that under the alternative

$$\left\|\frac{1}{h}\hat{\mathbf{D}} - \mathbf{D}^*\right\| = o_P(1),$$

where $\mathbf{D}^*$ is a positive definite matrix. Hence, the condition for the consistency, that is to have (3.5) is $h/N \to 0$, which is part of Assumption 3.1.

We now list other assumptions needed for the consistency of our test. Our first assumption is that under the alternative the separation between the change points must increase linearly with the length of the time series.

Assumption 3.3. There are $0 < \theta_1 < \theta_2 < \cdots < \theta_R < 1$ such that $k_\ell^* = \lfloor N\theta_\ell \rfloor, 1 \le \ell \le R$, where $\theta_0 = 0$ and $\theta_{R+1} = 1$.

Let $\tilde{\boldsymbol{\beta}}_\ell$ be the value of the parameter between $k_{\ell-1} + 1$ and $k_\ell, 1 \le \ell \le R + 1$. Put

$$\tilde{\boldsymbol{\beta}}_\ell = \left(\tilde{\boldsymbol{\beta}}_{1,\ell}^\top, \tilde{\boldsymbol{\beta}}_{2,\ell}^\top\right)^\top, \quad \tilde{\boldsymbol{\beta}}_{1,\ell} \in R^r \quad \text{and} \quad \tilde{\boldsymbol{\beta}}_{2,\ell} \in R^d$$

The following condition implies that the sequence is stationary between change points:

Assumption 3.4. Set $\tilde{\boldsymbol{\beta}}_{2,\ell} = \left(\tilde{\beta}_{2,\ell,1}, \tilde{\beta}_{2,\ell,2}, \ldots, \tilde{\beta}_{2,\ell,d},\right)^\top$. For each $1 \le \ell \le R + 1$, $\tilde{\beta}_{2,\ell,d} \ne 0$, the roots of the polynomial $z^d - \tilde{\beta}_{2,\ell,1} z^{d-1} - \tilde{\beta}_{2,\ell,2} z^{d-2} - \cdots - \tilde{\beta}_{2,\ell,d}$ are inside of the unit circle of the complex plane.

Since the $\tilde{\mathbf{x}}_i$'s depend on $y_j, j < i$, we have stationarity only on the subintervals $(k_{\ell-1}, k_\ell], 1 \le \ell \le R + 1$, on which then we observe realizations of infinite stationary sequences.

Under our assumptions

$$\frac{1}{r_\ell - r_{\ell-1}} \sum_{i=k_{\ell-1}+1}^{k_\ell} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^\top \xrightarrow{P} \widetilde{\mathbf{A}}_\ell, \qquad 1 \le \ell \le R + 1.$$

Similarly to Assumptions 2.3 and 2.4, we require

Assumption 3.5. For all $1 \le \ell \le R + 1$, $\widetilde{\mathbf{A}}_\ell$ is nonsingular.

and

Assumption 3.6. $E\tilde{\mathbf{x}}_{k_\ell^*} \epsilon_{k_\ell^*} = \mathbf{0}, 1 \le \ell \le R + 1$.

We also need

Assumption 3.7. One of the matrices

$$\lim_{N\to\infty} \frac{1}{N} E \left(\sum_{i=k_{\ell-1}+1}^{k_\ell} \tilde{\mathbf{x}}_i \tilde{\epsilon}_i\right) \left(\sum_{i=k_{\ell-1}+1}^{k_\ell} \tilde{\mathbf{x}}_i \tilde{\epsilon}_i\right)^\top, \quad 1 \le \ell \le R + 1,$$

is nonsingular.

Our last theorem establishes the consistency of the hypothesis testing of no change based on Theorem 3.1.

Theorem 3.3. If H_A and Assumptions 2.2, 2.7, and 3.3–3.7 hold, then

$$\sup_{0<t<1} \frac{1}{w(t)} \left(\mathbf{U}_N^\top(t) \widehat{\mathbf{D}}^{-1} \mathbf{U}_N(t)\right)^{1/2} \xrightarrow{P} \infty. \tag{3.5}$$

4. Finite Sample Performance

In this section, we investigate finite sample performance of the tests introduced in Section 3. We consider tests based on Theorems 3.1 and 3.2 with the matrix $\mathbf{D}$ in (3.2) and (3.3) replaced by an estimator $\widehat{\mathbf{D}}$. We use the kernel long run covariance matrix estimator with the Bartlett kernel and either data driven or fixed bandwidth, as specified in Section 4.1. Focusing on a specific weight function, we consider test statistics

$$\widehat{Z}(\kappa) = \sup_{0<t<1} \frac{1}{[t(1-t)]^\kappa} \left(\mathbf{U}_N^\top(t) \widehat{\mathbf{D}}^{-1} \mathbf{U}_N(t)\right)^{1/2}, \quad 0 < \kappa < 1/2, \tag{4.1}$$

and

$$\widehat{G} = \sup_{0<t<1} \frac{1}{(t(1-t))^{1/2}} \left(\mathbf{U}_N^\top(t) \widehat{\mathbf{D}}^{-1} \mathbf{U}_N(t)\right)^{1/2}. \tag{4.2}$$

As discussed in Section 3, statistic (4.2) can be used only in the context of Theorem 3.2, with critical values obtained via approximation (4.4). The residuals needed to construct the process $\mathbf{U}_N$ are computed using the lasso estimator minimizing (2.4) with $\gamma = 1$ and λ_N selected as specified in Section 4.1. We used the MATLAB `lasso` function.

We work with dimension m, which estimates $r + d$. Asymptotic critical values of $\widehat{Z}(\kappa)$, denoted $c_B(\kappa, m, \alpha)$, are determined by

$$P\left\{\sup_{1/(N+1)\le t\le 1-1/(N+1)} \frac{1}{[t(1-t)]^\kappa} \left(\sum_{i=1}^m B_i^2(t)\right)^{1/2} > c_B(\kappa, m, \alpha)\right\} = \alpha. \tag{4.3}$$

The Brownian bridges are approximated on a grid of 1000 equispaced points in $[0, 1]$ and the $c_B(\kappa, m, \alpha)$ are based on 10,000 replications.

Using Theorem 3.2, specifically approximation (A.2), we obtained critical values $c_V(N, m, \alpha)$ which satisfy

$$P\left\{\sup_{0\le s\le r(N)} V(s) > c_V(N, m, \alpha)\right\} = \alpha, \tag{4.4}$$

with $r(N) = \log(N^2) + C(\log\log N)^g$. We used several values of C and g listed in the following. The test rejects at the nominal level α if $\widehat{G} > c_V(N, m, \alpha)$. The selection of m is discussed in the following.

Table 1. Empirical sizes of the test based on $\widehat{Z}(0.35)$ for DGPs(i)–(vi) with critical values computed using (4.3) with $m = \tilde{r} + \tilde{d}$.

N	250			500			1000		
α	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
DGP(i)	0.076	0.035	0.008	0.088	0.042	0.009	0.098	0.041	0.011
DGP(ii)	0.065	0.030	0.012	0.088	0.042	0.011	0.102	0.052	0.017
DGP(iii)	0.050	0.022	0.005	0.078	0.030	0.008	0.085	0.036	0.008
DGP(iv)	0.061	0.033	0.007	0.082	0.042	0.012	0.108	0.058	0.018
DGP(v)	0.051	0.020	0.002	0.073	0.028	0.007	0.083	0.041	0.008
DGP(vi)	0.061	0.032	0.008	0.090	0.042	0.012	0.102	0.054	0.015

Table 2. Empirical sizes of the test based on $\widehat{Z}(0.35)$ for DGPs(i)–(vi) with critical values computed using (4.3) with $m = \tilde{r} + \tilde{d} - J$.

N	250			500			1000		
α	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
DGP(i)	0.082	0.033	0.006	0.096	0.045	0.007	0.093	0.039	0.005
DGP(ii)	0.080	0.041	0.010	0.093	0.046	0.009	0.082	0.041	0.013
DGP(iii)	0.065	0.027	0.003	0.087	0.036	0.009	0.079	0.033	0.008
DGP(iv)	0.069	0.038	0.012	0.103	0.056	0.019	0.108	0.057	0.018
DGP(v)	0.075	0.042	0.009	0.102	0.052	0.011	0.098	0.054	0.008
DGP(vi)	0.100	0.054	0.014	0.118	0.062	0.017	0.123	0.075	0.021

4.1. Empirical Size

We begin by describing data generating processes (DGPs):

DGP(i) (IIDs, IID) The $x_{i,j}$, $1 \le i \le N$, $2 \le j \le \tilde{r}$ are independent, identically distributed standard normal random variables. Also, the errors ϵ_i, $1 \le i \le N$, are iid standard normal random variables, independent of the $x_{i,j}$. (In this case, the normalizing matrix $\widehat{\mathbf{D}}$ is the sample covariance matrix.)

DGP(ii) (IIDs, GARCH) The $x_{i,j}$ are independent, identically distributed standard normal random variables, the errors ϵ_i form a GARCH(1,1) process defined by

$$\epsilon_i = \sigma_{i,\epsilon} h_i \quad \sigma^2_{i,\epsilon} = .3 + .2\epsilon^2_{i-1} + .5\sigma^2_{i-1,\epsilon}, \tag{4.5}$$

where the h_i's are independent, standard normal random variables, independent of the $x_{i,j}$.

DGP(iii) (AR, IID) The $x_{i,j}$ follow an AR(1) process defined by

$$x_{i,j} = 0.5x_{i-1,j} + \eta_{i,j}, \quad 1 \le i \le N,\ 2 \le j \le \tilde{r}, \tag{4.6}$$

where the $\eta_{i,j}$ are independent, identically distributed standard normal random variables, independent of $\{\epsilon_i,\ 1 \le i \le N\}$. The errors ϵ_i are still standard normal.

DGP(iv) (AR, GARCH) The $x_{i,j}$ follow the AR(1) process of (4.6), the errors ϵ_i the GARCH process (4.5).

DGP(v) (XARs, IID) The $x_{i,j}$ are generated from an autoregressive (AR) process defined by

$$x_{i,j} = \rho_j x_{i-1,j} + \eta_{i,j}, \quad 1 \le i \le N,\ 2 \le j \le \tilde{r}, \tag{4.7}$$

where ρ_j is randomly selected from a uniform distribution $U[-.7, \ .7]$ and the $\eta_{i,j}$ are still standard normal random variables, independent of $\{\epsilon_i,\ 1 \le i \le N\}$, as in DGP(iii). But now $\eta_{i,j}$, $2 \le j \le \lfloor \tilde{r}/2 \rfloor$ are iid, while $\eta_{i,\lfloor \tilde{r}/2 \rfloor + 1} = \cdots = \eta_{i,\tilde{r}}$, that is these errors are the same for each variable j, if $j > \lfloor \tilde{r}/2 \rfloor$. The errors ϵ_i are still independent standard normal.

DGP(vi) (XARs, GARCH) The $x_{i,j}$ satisfy (4.7), the errors ϵ_i satisfy (4.5).

In the case of DGPs (i)–(iv), the coordinates of $(x_{i,2}, \ldots, x_{i,\tilde{r}})$ are independent, whereas they are cross–dependent under DGPs (v) and (vi). The sample sizes, numbers of regressors and the structure of the coefficients are designed to resemble situations encountered in the data examples studied in Section 5. We implement the tests with $m = \tilde{r} + \tilde{d}$ and $m = \tilde{r} + \tilde{d} - J$, where J is the count of coefficients set to zero by the lasso. In other words, $m = \tilde{r} + \tilde{d} - J$ is the count of nonzero coefficients selected by LASSO.

Moderate dimension. We first consider a moderate number of regressors, $\tilde{r} = 10$ and $\tilde{d} = 1$. Under the null hypothesis, the parameters of the linear regressors $(1, x_{i,2}, \ldots, x_{i,10})^\top$, $1 \le i \le N$, are

$$\boldsymbol{\beta}_{1,0} = (1, 1, 0.5, 0.1, 0.01, 0, -0.01, -0.1, -0.5, -1)^\top$$

and the parameter of the autoregressive term y_{i-1} is $\beta_{2,0} = 0.5$. We see that there is only one coefficient equal exactly to 0 and two that are almost zero.

Empirical sizes, based on the automatic bandwidth selection of Andrews (1991) and on 2000 replications, are shown in Tables 1 and 2 for $\kappa = 0.35$. The value $\kappa = 0.35$ produced the best balance of empirical size and power in Horváth, Kokoszka, and Wang (2021), even though the testing problem was different (sequential monitoring for a change in distribution) and in several older papers cited there. We have also produced analogous tables for $\kappa = 0, 0.1, 0.2, 0.3, 0.4$ and observed that $\kappa = 0.3$ and $\kappa = 0.4$ produce similar results as $\kappa = 0.35$. For smaller κs, the test is too conservative. Comparing Tables 1 and 2, we see that using $m = \tilde{r} + \tilde{d} - J$ in place of $m = \tilde{r} + \tilde{d}$ improves the accuracy of the test in some cases, for example DGP(v), but makes it slightly worse in others. The main conclusion is that for $\tilde{r} = 10$ and $\tilde{d} = 1$, the test based $\widehat{Z}(0.35)$ is quite accurate for $N \ge 500$ and reasonably accurate for $N \ge 250$. It does not make much difference if $m = \tilde{r} + \tilde{d} - J$ or $m = \tilde{r} + \tilde{d}$ is used. The test is somewhat deficient for DGP(iii), that is when all coordinates $x_{i,j}$. $2 \le j \le \tilde{r}$ have precisely the same temporal dependence. This is unlikely in practice, but this DPG is included to show that the the test does not perform well in some cases.

We now turn to the test based on the statistic $\widehat{G}$ defined in (4.2). The critical values computed according to (4.4) depend on

Table 3. Empirical sizes of the test based on the statistic $\widehat{G}$ defined in (4.2) with critical values computed with $C = g = 11$.

N	250			500			1000		
α	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
Using all $\tilde{r}+\tilde{d}$ variables									
DGP(i)	0.052	0.045	0.030	0.066	0.058	0.046	0.056	0.049	0.037
DGP(ii)	0.068	0.057	0.045	0.087	0.079	0.065	0.091	0.082	0.064
DGP(iii)	0.058	0.047	0.037	0.067	0.059	0.049	0.071	0.062	0.050
DGP(iv)	0.072	0.065	0.049	0.099	0.092	0.068	0.115	0.104	0.083
DGP(v)	0.008	0.005	0.004	0.021	0.019	0.013	0.025	0.023	0.014
DGP(vi)	0.030	0.024	0.019	0.050	0.042	0.036	0.065	0.057	0.046
Using only $\tilde{r}+\tilde{d}-J$ variables									
DGP(i)	0.051	0.042	0.032	0.067	0.058	0.043	0.051	0.045	0.039
DGP(ii)	0.072	0.061	0.049	0.086	0.079	0.062	0.094	0.081	0.062
DGP(iii)	0.056	0.047	0.034	0.072	0.062	0.046	0.072	0.062	0.052
DGP(iv)	0.078	0.069	0.053	0.104	0.096	0.072	0.109	0.098	0.081
DGP(v)	0.070	0.057	0.031	0.083	0.069	0.043	0.071	0.058	0.039
DGP(vi)	0.084	0.070	0.043	0.106	0.087	0.062	0.114	0.097	0.068

positive constants g and C that do not affect the *asymptotic* size. These constants appear in the term $C(\log\log N)^g$, and since the iterated logarithm changes very slowly with N, and the tail in Vostrikova's formula (A.2) decays basically like $\exp(-x^2)$, one might expect that they will have little impact on finite sample test size. We investigated the size for several choices like $C = \tilde{r}+\tilde{d}$, $g = \tilde{r}+\tilde{d}$, $C = 10^3(\tilde{r}+\tilde{d})$, $g = \tilde{r}+\tilde{d}$, $C = \tilde{r}+\tilde{d}$, $g = 2(\tilde{r}+\tilde{d})$, and indeed did not see much effect. We therefore report empirical sizes for $C = \tilde{r}+\tilde{d}$, $g = \tilde{r}+\tilde{d}$. Table 3 shows empirical sizes for the test based on the statistic $\widehat{G}$. Broadly speaking, the conclusions are the same as for the statistic $\widehat{Z}(0.35)$, but the empirical sizes at the nominal level of one percent are too large. The overall recommendation for moderate sample size is to apply tests based on $\widehat{Z}(0.35)$ and $\widehat{G}$ and $m = \tilde{r}+\tilde{d}$, at the standard 5% level.

High dimension. We now turn to the high-dimensional setting. There are clearly even more options for $\tilde{r}$, $\tilde{d}$ and the DGPs than in the moderate dimensional setting. We opted for settings most relevant to the data example in Section 5.2 to make sure that the conclusions reached in that section are well supported. We thus consider three high-dimensional settings with $\tilde{r} = 125$ and $\tilde{d} = 1$, and the following parameter settings:

H1: The parameter vector $\boldsymbol{\beta}_{1,0}$ is 125×1 with 11 ones and 114 zeros.

H2: The vector $\boldsymbol{\beta}_{1,0}$ has 119 zeros and 6 ones.

H3: The vector $\boldsymbol{\beta}_{1,0}$ has 119 zeros, 4 ones and the remaining two coordinates are 0.1.

In the three cases above, the indexes of the nonzero entries are randomly selected. The first coordinate $\beta_{1,0,1}$ (i.e., the parameter of the model intercept) is always 1. In all scenarios, the autoregressive coefficient is $\beta_{2,0} = 0.5$.

For each high-dimensional setting, we considered the DGPs (i)–(vi) with modified DGPs (v) and (vi). In Section 5.2, we estimate the pairwise correlation matrix Σ of the 124 macroeconomic variables using the data from January 1980 to December 2021 (i.e., $N = 504$). The errors of the cross-dependent AR processes $\eta_{i,j}$, $2 \le j \le 125$ in DGPs (v) and (vi) were generated from the multivariate normal distribution with mean zero and covariance Σ_η that has the same eigenvalues as the estimated Σ in Section 5.2. We used the MATLAB `gallery('randcorr', x)` function to produce the random matrix Σ_η, where x is the vector of eigenvalues of Σ.

The most clear cut finding is that using $m = \tilde{r}+\tilde{d}$ in the high-dimensional setting yields tests that are much too conservative. We performed additional simulations with several large values of $\tilde{r}$ and observed that the empirical size declined monotonically with $\tilde{r}$. This is to be expected because if the vast majority of the coefficients are zero, they do not contribute to the test statistics, yet their number, $\tilde{r}$, increases the critical values. The second finding is that using $m = \tilde{r}+\tilde{d}-J$ with λ determined by 10-fold cross-validation still leads to very conservative tests. The explanation is that such a penalty is too small and produces too many nonzero coefficients. This is illustrated in Figure 1.

The difficulties of applying the lasso to select the correct/optimal coefficients are known, see for example (Zou 2006; Zhang and Huang 2008; Meinshausen, Meier, and Bühlmann 2009; Emmert-Streib and Dehmer 2019), and are not peculiar to our setting. Our objective is to recommend a penalty that yields good sizes in the applications we consider. As shown in Table 4, using $\lambda = 2\lambda_{10}$ gives satisfactory results. Recall that λ_{10} is the penalty selected by the 10-fold cross-validation that worked well for the moderate dimension. Another issue is the selection of the bandwidth h. We are not aware of any data driven-methods derived in the setting of our model (HD regression with LD autoregression), so we report the results for $h = 1$. A sensitivity analysis has shown that the test generally performs worse for larger values of h, it becomes too conservative and its power drops for $N = 500$, see Section C of the supplementary material. Table 4 shows that $\lambda = 2\lambda_{10}$ and $h = 1$ produce satisfactory empirical sizes, comparable to those obtained in the moderate setting. We refer to Basu and Matteson (2022) for an informative survey of estimation in several high-dimensional time series models.

4.2. Empirical Power

The DGPs are the same as in Section 4.1 until a change point k^*. Starting from k^*+1, the regression parameters change from $\boldsymbol{\beta}_{1,0}$ to $\boldsymbol{\beta}_{1,A} = \boldsymbol{\beta}_{1,0} + \delta_1$ while the autoregressive parameter changes from $\boldsymbol{\beta}_{2,0}$ to $\boldsymbol{\beta}_{2,A} = \boldsymbol{\beta}_{2,0} + \delta_2$. We consid-

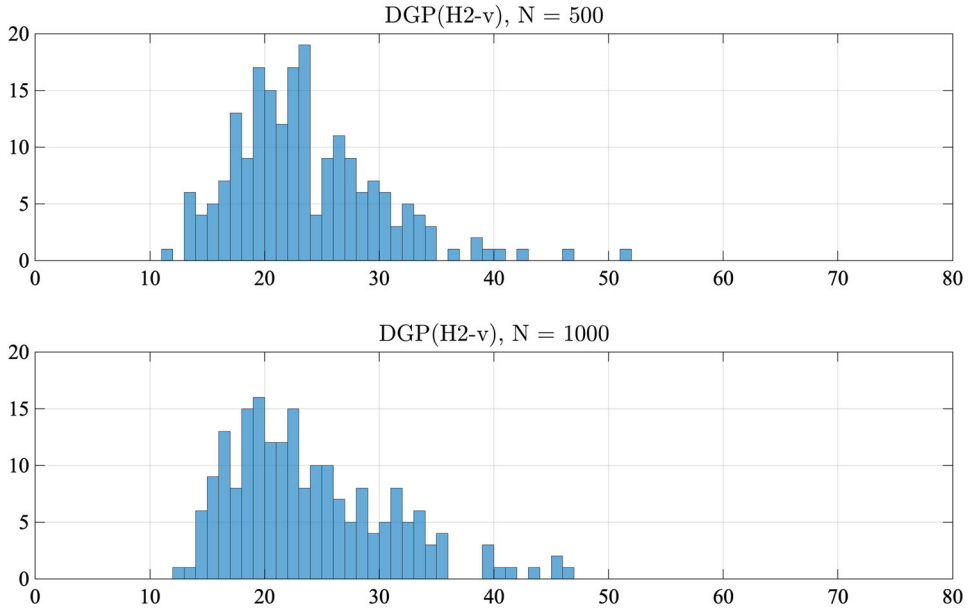

Figure 1. The histogram of the count of nonzero parameters estimated by LASSO with λ determined by 10-fold cross-validation. The actual count is 7.

Table 4. Empirical sizes of the test based on $\widehat{Z}(0.35)$ for DGPs(i)–(vi) under the high-dimensional settings with the penalty $\lambda = 2\lambda_{10}$, the fixed bandwidth $h = 1$, and critical values computed using (4.3) with $m = \tilde{r} + \tilde{d} - J$.

N	250			500			1000		
α	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
The first high-dimensional setting									
DGP(H1-i)	0.085	0.035	0.008	0.095	0.043	0.008	0.105	0.058	0.013
DGP(H1-ii)	0.039	0.017	0.002	0.082	0.040	0.012	0.105	0.058	0.015
DGP(H1-iii)	0.035	0.013	0.002	0.072	0.033	0.005	0.092	0.044	0.008
DGP(H1-iv)	0.044	0.017	0.002	0.094	0.052	0.017	0.107	0.059	0.014
DGP(H1-v)	0.016	0.006	0.001	0.047	0.017	0.003	0.083	0.039	0.004
DGP(H1-vi)	0.017	0.006	0.001	0.055	0.023	0.005	0.099	0.041	0.010
The second high-dimensional setting									
DGP(H2-i)	0.137	0.078	0.025	0.132	0.067	0.011	0.130	0.068	0.017
DGP(H2-ii)	0.089	0.039	0.007	0.111	0.056	0.015	0.119	0.061	0.018
DGP(H2-iii)	0.090	0.043	0.011	0.100	0.045	0.010	0.103	0.049	0.010
DGP(H2-iv)	0.103	0.053	0.010	0.123	0.074	0.017	0.126	0.069	0.022
DGP(H2-v)	0.047	0.019	0.003	0.072	0.030	0.006	0.108	0.051	0.010
DGP(H2-vi)	0.061	0.026	0.003	0.094	0.041	0.008	0.106	0.060	0.012
The third high-dimensional setting									
DGP(H3-i)	0.175	0.108	0.028	0.160	0.089	0.033	0.142	0.081	0.017
DGP(H3-ii)	0.121	0.071	0.015	0.140	0.077	0.025	0.129	0.074	0.019
DGP(H3-iii)	0.114	0.058	0.013	0.121	0.068	0.019	0.116	0.062	0.015
DGP(H3-iv)	0.132	0.074	0.024	0.153	0.080	0.021	0.145	0.079	0.021
DGP(H3-v)	0.077	0.033	0.005	0.093	0.045	0.009	0.110	0.058	0.012
DGP(H3-vi)	0.096	0.048	0.010	0.127	0.068	0.020	0.121	0.066	0.017

NOTE: The results are based on 2000 replications.

ered $\delta_1 \in \{0.0, 0.2, \ldots, 1.0\}$ (increments of 0.2) and $\delta_2 \in \{0, 0.05, 0.10, 0.15, 0.20, 0.25\}$. In the high-dimensional settings, only 15 randomly selected parameters (out of 125) change. This is motivated by the data considered in Section 5.2.

In some high-dimensional scenarios, the sample size of $N = 250$ may be insufficient to detect a change (if only 15 parameters change). This is illustrated in Figure 2. The low power for $N = 250$ occurs if the $x_{i,j}$ are AR(1) with identical autoregressive coefficients $\rho_j = 0.5$, which would be rare in practice. The remaining findings are as expected: the power increases with the sample size, and $N = 500$ is sufficient to detect a change point. The power also increases with the magnitude of the change, and it is easier to detect a change point in the middle of the record than close to the beginning (or end) of the record. More power curves are shown in Section B of the supplementary material.

5. Application to the Detection of Changes in Risk Factor Models

The purpose of this section is to apply our test procedures to data that motivate them. Previous research considered estimation of change points in these data, without checking if they are significant. We use standard binary segmentation to identify significant change points. It is however recommended that change point estimation procedures for which proofs of consistency have been developed are used. We refer to them in

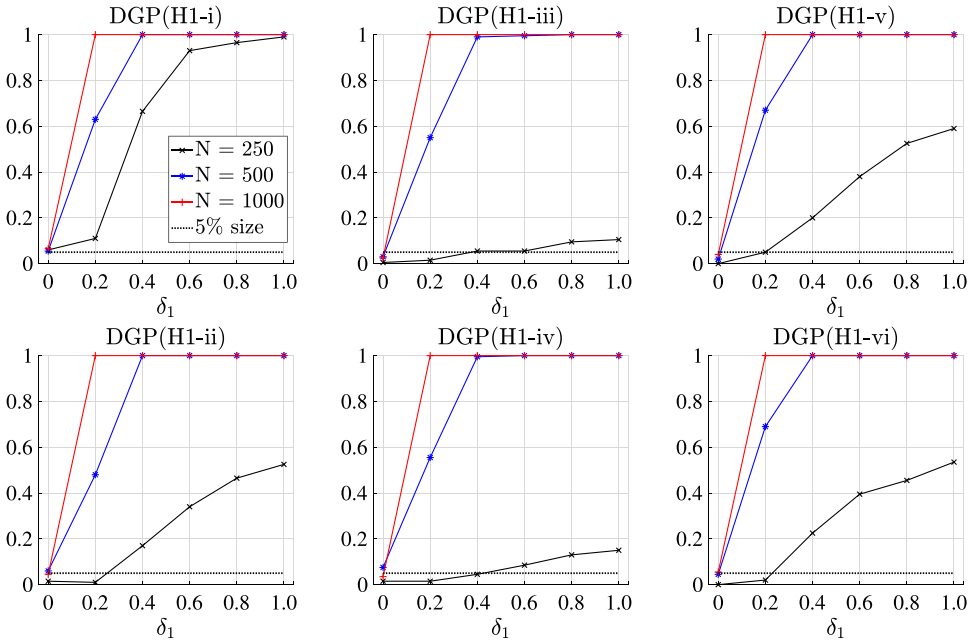

Figure 2. Power of the test based on $\widehat{Z}(0.35)$ as the function of the size of change δ_1 ($\delta_2 = 0$) at the nominal level $\alpha = 5\%$ when $k^* = \lfloor .5N \rfloor$. The power curves are for the high-dimensional scenario H1, and only 15 out of 125 regression parameters change.

the following. It turns out that binary segmentation identifies very similar change points, which lends additional confidence to our analysis.

We consider two datasets to test whether there are structural breaks in the coefficients of risk factors of two important financial and economic variables (the y_i):

1. Monthly returns on U.S. energy stocks portfolio: $\tilde{r} = 23, \tilde{d} = 1, N = 648$, Section 5.1;
2. Monthly log changes of U.S. Industrial Production index: $\tilde{r} = 125, \tilde{d} = 3, N = 504$, Section 5.2.

The portfolio in point 1 is constructed from all stocks traded at the NYSE, AMEX, and NASDAQ exchanges and classified as energy industry stocks by their four-digit SIC code. The U.S. Industrial Production index in point 2 is a key indicator for evaluating macroeconomic growth.

Both datasets are used to shed new light on the perhaps most extensively studied problem of empirical and theoretical finance: determination of risk factors, their impact and significance. This strain of research goes back to early 1970s and is generally associated with the name of E. F. Fama. In the field of empirical asset pricing there is persistent interest in testing for new risk factors that can be used to explain the expected stock returns. The number of proposed factors is reaching two hundred, a big increase from the now classical three factor model of Fama and French (1992). Such a sheer number of factors has lead to what is termed a replication crisis, similar as the credibility crisis in the fields of medicine, psychology, and management (Ioannidis 2005; Nosek, Spies, and Motyl 2012; Bettis 2012). Many researchers have argued that most of the discoveries are likely false due to no internal validity and/or no external validity (Harvey, Liu, and Zhu 2016; Chordia, Goyal, and Saretto 2020; Hou, Xue, and Zhang 2020; Jensen, Kelly, and Pedersen 2021). This means that a sparse, practically high-dimensional regression framework considered in this article may be relevant; the $x_{i,j}$, $1 \le j \le \tilde{r}$, correspond to the series of risk factors. Our objective is however not to contribute to the debate on which factors are optimal, but rather we focus on testing if their impact remains the same or if there is a change in how they affect a response variable y_i. Such research is useful because it may reveal that factors that are significant and important over one time period may not be so over another period, helping form a better understanding of results reported in extensive empirical research. From the perspective of statistics research, testing is complementary to research on change point estimation that assumes that at least one change point exists. In the absence of a prior significance test, an estimated change point may be spurious.

5.1. Changes in Risk Factors of the U.S. Energy Industry Stocks

As the response variable y_i, we consider monthly excess returns on U.S. Energy stocks defined as the raw return less the one–month treasury bill rate. The data were downloaded from the personal website of Professor Kenneth R. French specified in the data availability statement, where one can find more information on the portfolio construction. We collected 22 exogenous risk factors that have been studied, including the FF5 factors, HXZ q factors, HMXZ q^5 factors, liquidity factors, volatility factors, trend factors and trading volume factors, see Table D.1 in the supplementary material for the descriptions and original papers corresponding to them. In addition, a first order autoregressive term is included to account for the adaptive market hypothesis as advocated by Lo (2004, 2019). Our sample covers monthly data from January 1968 to December 2021, so the sample size N is 648. The KPSS test (Kwiatkowski et al. 1992) shows that stationarity cannot be rejected for any of the risk factor series $x_{i,j}$, with the smallest p-value of 0.0835 for FF_{SMP} and most p-values between 0.2 and 0.4. Reflecting the findings of Section 4, we use $\kappa = 0.35, \alpha = 0.05$, the regularization of λ_{10}, and critical

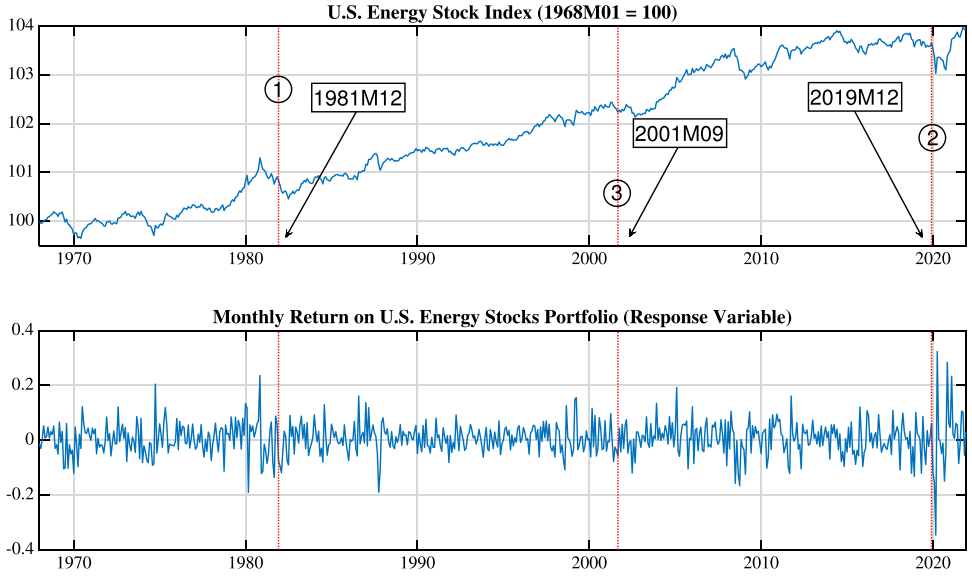

Figure 3. The breaks in the U.S. energy industry stocks.

Table 5. Summary information for the four segments displayed in Figure 3.

	y_i	y_i^2	LRV(y_i)	ϵ_i	ϵ_i^2	LRV(ϵ_i)	# of 0 in $\hat{\boldsymbol{\beta}}_N$	# of changes in $\hat{\boldsymbol{\beta}}_N$
Panel A: The first phase from January 1968 to December 1981								
Autocorrelation	No	Yes	0.0036	No	Yes	0.0008	9	–
p-value	0.2206	0.0262		0.9079	0.0215			
Panel B: The second phase from January 1982 to September 2001								
Autocorrelation	No	Yes	0.0025	Yes	Yes	0.0011	7	22
p-value	0.1950	0.0469		0.0868	0.0009			
Panel C: The second phase from October 2001 to December 2019								
Autocorrelation	No	No	0.0033	No	No	0.0013	8	19
p-value	0.5507	0.7233		0.4032	0.7771			
Panel D: The third phase from January 2020 to December 2021								
Autocorrelation	No	No	0.0195	No	No	0.0037	19	16
p-value	0.9654	0.5314		0.4744	0.1702			

NOTE: The Ljung–Box Q test is conducted by including three lagged terms. The long run variance tabulated is estimated with the Bartlett kernel, and the bandwidth selected through the automatic bandwidth selection method of Andrews (1991).

values computed using (4.3) with $m = \tilde{r} + \tilde{d}$. We note that the change points remain significant at the $\alpha = 0.01$ level. As in the moderate setting in Section 4, the long run covariance matrix estimator is computed with the Bartlett kernel, and the bandwidth is selected through the automatic bandwidth selection method of Andrews (1991).

As shown in Figure 3, our test statistics $\widehat{Z}(0.35)$ detected three structural breaks for the Energy Industry stocks using standard binary segmentation. The first detected break is in December 1981, the second in December 2019 and the third in September 2001. These change points can be validated by association with events that may have impacted the performance of the energy stocks. The trigger of the break in December 1981 is likely the presidential executive order 12,287 of "Decontrol of Crude Oil and Refined Petroleum Products" signed in January 1981.[1] The order aimed to curb oil consumption through higher prices that would result from decontrol and stimulate domestic production. Such a fundamental change in energy policy is expected to impact all companies in the industry. The change point in December 2019 can be clearly associated with the COVID–19 Pandemic. Various public health measures reduced demand for travel and impacted international trade. With this unprecedented shock to not only the U.S. stock market but also the global economy, the Energy stocks returns were driven by only four factors (FF_{MKT}, FF_{MOM}, EP, STDVOL). After the break, the estimated exposures to FF_{MKT}, FF_{MOM}, EP, and STDVOL factors were 0.39, −0.01, 0.85, and −0.58, while exposures to 11 other risk factors, which were significant before the break, changed to zero. The break in September 2001 can be associated with the 9/11 terrorist attacks, which temporarily reduced demand for travel, but also is likely related to the Bush Administration's National Energy Policy (NEP) implementation and two climate change initiatives: the U.S. Climate Change Research Initiative and the National Climate Change Technology Initiative, which had an impact on the whole Energy industry.

In Table 5, we display selected statistical summaries for the four partition periods, including the long run variances of the

[1]For more details, please see the website *https://www.archives.gov/federal-register/codification/executive-order/12287.html* (June 8, 2022).

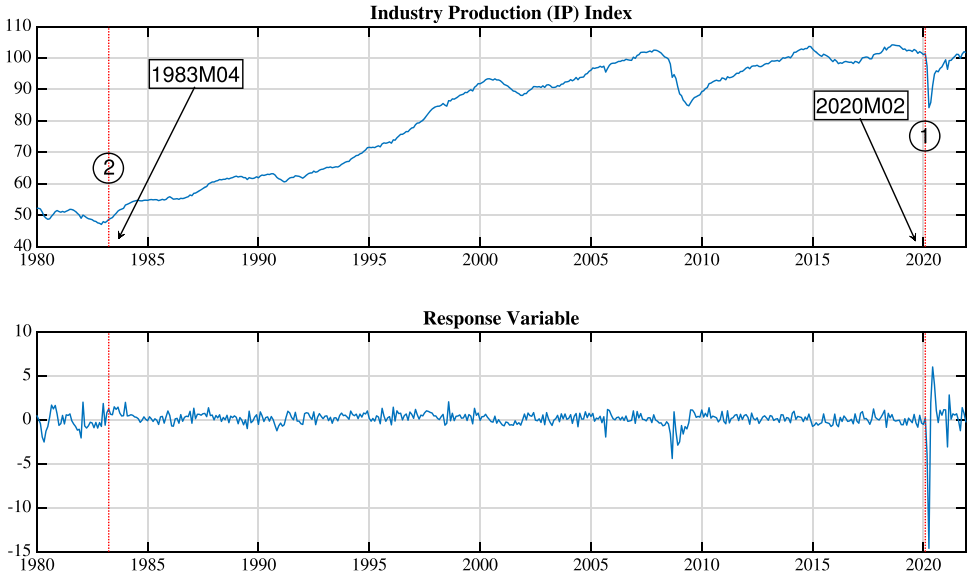

Figure 4. The breaks in the U.S. Industrial Production Index.

Table 6. The LASSO estimates during three phases segmented by the estimated breaks for U.S.

	y_i	y_i^2	LRV(y_i)	ϵ_i	ϵ_i^2	LRV(ϵ_i)	# of 0 in $\hat{\beta}_N$	# of changes in $\hat{\beta}_N$
Panel A: The first phase from January 1980 to April 1983								
Autocorrelation	Yes	Yes	1.7695	No	No	1.0449	125	–
p-value	0.0864	0.0675		0.4088	0.3206			
Panel B: The second phase from May 1983 to February 2020								
Autocorrelation	Yes	Yes	0.8621	No	No	0.2371	94	35
p-value	0.0000	0.0000		0.5287	0.1777			
Panel C: The third phase from March 2020 to December 2021								
Autocorrelation	No	No	16.5100	No	No	0.2344	113	45
p-value	0.2403	0.8849		0.7421	0.3473			

NOTE: Industrial Production Index. The Ljung–Box Q test is conducted by including three lagged terms. The long run variance tabulated is estimated with the Bartlett kernel, and the bandwidth selected through the automatic bandwidth selection method of Andrews (1991).

response variable and the residuals. We also summarize the LASSO estimates before and after the breaks in the last two columns of Table 5. By comparing Panel B to Panel A, one can find that LASSO penalized less risk factors to have zero exposures after the break of December 1981. On top of that, the exposures of 22 risk factors changed, as well as the pricing alpha (or the intercept in the regression). These results demonstrate a significant, not only in the statistical sense, overall structural change in the U.S. Energy stocks because of the dramatic federal policy change.

5.2. Structural Break Detection in the U.S. Industrial Production Index

We consider detection of structural breaks in the impact of 124 macroeconomic variables on the growth rate of U.S. Industrial Production (IP) Index. Our response variable is $y_i = \log(\text{IP}_i/\text{IP}_{i-1}) \times 100$, where IP_t is the U.S. Industrial Production Index in month i. Following He, Jaidee, and Gao (2022), the high-dimensional predictors are the ex-ante 124 variables in month $i-1$ and the lagged response variables y_{i-1}, y_{i-2}, and y_{i-3}. We transformed the raw data to stationary time series using the MATLAB code provided by McCracken and Ng (2016). We used monthly data from January 1980 to December 2021, so the number of the (potential) explanatory variables $\tilde{r} + \tilde{d} = 128$ is roughly a quarter of the sample size $N = 504$. The high-dimensional dataset covering 127 macroeconomic variables is a modified version of the 132 macroeconomic dataset in Stock and Watson (2002, 2005) and Breitung and Eickmeier (2011). We selected the 124 variables out of the 127 due to missing values in the long period we study in the three omitted variables. Chen, Dolado, and Gonzalo (2014) and Han and Inoue (2015) used the 132 macroeconomic variables while Wang and Zhao (2022) used the 127 variables, but for shorter sample periods.

For our detection, we used $\kappa = 0.35$, the regularization with $\lambda_N = 2\lambda_{10}$ and $h = 1$. The critical values are computed using (4.3) with $m = \tilde{r}+\tilde{d}-J$. These settings are based on the findings of Section 4.

As illustrated in Figure 4, the test statistic $\widehat{Z}(0.35)$ found two breaks in the model parameters in April 1983 and February 2020 when $\alpha = 0.05$ (using $\alpha = 0.01$ detects the latter one). The first break, April 1983, likely marks the transition period from the so-called Great Inflation to the Great Moderation. The Great Moderation is known as a period of macroeconomic stability relative to the volatility of the Great Inflation. We refer to the website of Federal Reserve History *https://www.federalreservehistory.*

org/time-period/federal-reserve-history (September 15, 2022) for more information on these two periods. In Table 6, we show that the volatility of the response variable is 0.8621 after the break, much lower than the 1.7695 before the break. The second break in February 2020 almost certainly reflects the shut down the overall economic activities due to the COVID–19 Pandemic. (The method of Xu et al. (2022) identified March 2020 as the change point.) We note that the detected change precedes the emergence of high inflation and the collapse of the stock market by about two years. The volatility of the IP, 16.51, is exceptionally high. We note that, for this particular dataset, we obtained very similar results using the automated bandwidth selection (rather than $h = 1$), and using $\alpha = 0.05$ and $\alpha = 0.01$ leads to the same test decisions. The change points were in August 1982 and April 2020 (estimated Great Moderation longer by 8 months). The analog of Table 6 looks basically the same.

Supplementary Materials

Additional supplementary material may be found online in the supplemental material tab for this article. It contains the proofs of the results stated in Section 3 (Section A), additional power curves graphs (Section B), the definitions of the response variable and the explanatory variables used in Section 5.1 (Section C), and the data availability statement (Section D).

Acknowledgments

We thank the Digital Economy Laboratory at the University of International Business and Economics for providing a high-performance computer, which enabled the execution of our numerical work.

Disclosure Statement

The authors do not declare any conflict of interest.

Funding

This research was partially supported by the United States National Science Foundation grants DMS–1923142, DMS–1914882 and DMS–2123761. Shanglin Lu was supported by the National Natural Science Foundation of China (No. 72303032) and the Fundamental Research Funds for the Central Universities in UIBE (22QD07).

References

Andreou, E., and Ghysels, E. (2009), "Structural Breaks in Financial Time Series," in *Handbook of Financial Time Series*, eds. T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch, pp. 839–870, Berlin: Springer. [1331]

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [1336,1340,1341]

——— (1993), "Test for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856. [1334]

Baek, Y. (2023), "Estimation of a Structural Break Point in Linear Regression Models," *Journal of Business & Economic Statistics*, forthcoming. [1331]

Barigozzi, M., Cho, H., and Fryzlewicz, P. (2018), "Simultaneous Multiple Change-Point and Factor Analysis for High-Dimensional Time Series," *Journal of Econometrics*, 39, 187–225. [1332]

Basu, S., and Matteson, D. (2022), "A Survey of Estimation Methods for Sparse High-Dimensional Time Series Models," Technical Report, Cornell University. [1337]

Bettis, R. A. (2012), "The Search for Asterisks: Compromised Statistical Tests and Flawed Theories," *Strategic Management Journal*, 33, 108–113. [1339]

Bodnar, T., Okhrin, Y., and Parolya, N. (2023), "Optimal Shrinkage-based Portfolio Selection in High Dimensions," *Journal of Business & Economic Statistics*, 41, 140–156. [1331]

Bradley, R. C. (2007), *Introduction to Strong Mixing Conditions* (Vols. 1–3), Heber City, UT: Kendrick Press. [1333]

Breitung, J., and Eickmeier, S. (2011), "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics*, 163, 71–84. [1341]

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer. [1333]

Candes, E., and Tao, T. (2007), "The Dantzig Selector: Statistical Estimation when p is much larger than n," *The Annals of Statistics*, 35, 2313–2351. [1334]

Chen, L., Dolado, J. J., and Gonzalo, J. (2014), "Detecting Big Structural Breaks in Large Factor Models," *Journal of Econometrics*, 180, 30–48. [1332,1341]

Cheng, X., Liao, Z., and Schorfheide, F. (2016), "Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities," *The Review of Economic Studies*, 83, 1511–1543. [1332]

Cho, H., and Owens, D. (2022), "High-Dimensional Data Segmentation in Regression Settings Permitting Heavy Tails and Temporal Dependence," Technical Report, University of Bristol. arXiv:2209.08892. [1332]

Chordia, T., Goyal, A., and Saretto, A. (2020), "Anomalies and False Rejections," *Review of Financial Studies*, 33, 2134–2179. [1339]

Csörgő, M., and Horváth, L. (1997), *Limit Theorems in Change-Point Analysis*, Chichester: Wiley. [1334]

Dette, H., Eckle, T., and Vetter, M. (2020), "Multiscale Change Point Detection for Dependent Data," *Scandinavian Journal of Statistics*, 47, 1243–1274. [1333]

Dette, H., Pan, G., and Yang, Q. (2022), "Estimating a Change Point in a Sequence of Very High-Dimensional Covariance Matrices," *Journal of the American Statistical Association*, 117, 444–454. [1332]

Emmert-Streib, F., and Dehmer, M. (2019), "High-Dimensional LASSO-based Computational Regression Models: Regularization, Shrinkage, and Selection," *Machine Learning and Knowledge Extraction*, 1, 359–383. [1337]

Enikeeva, F., and Harchaoui, Z. (2019), "High-Dimensional Change-Point Detection with Sparse Alternatives," *The Annals of Statistics*, 47, 2051–2079. [1331]

Fama, E. F., and French, K. R. (1992), "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465. [1339]

Fan, J., Lv, J., and Qi, L. (2011), "Sparse High-Dimensional Models in Economics," *Annual Review of Economics*, 3, 291–317. [1331]

Fang, X., Li, J., and Siegmund, D. (2020), "Segmentation and Estimation of Change-Point Models: False Positive Control and Confidence Regions," *The Annals of Statistics*, 48, 1615–1647. [1331]

Frank, I. E., and Friedman, J. H. (1993), "A Statistical View of Some Chemometrics Regression Tools," (with Discussion), *Technometrics*, 35, 109–148. [1333]

Gareth, J., and Radchenko, P. (2009), "A Generalized Danzig Selector with Shrinkage Tuning." *Biometrika*, 96, 323–337. [1334]

Han, X., and Inoue, A. (2015), "Tests for Parameter Instability in Dynamic Factor Models," *Econometric Theory*, 31, 1117–1152. [1332,1341]

Harvey, C. R., Liu, Y., and Zhu, H. (2016), " ... and the Cross-section of Expected Returns," *Review of Financial Studies*, 29, 5–68. [1339]

He, Y., Jaidee, S., and Gao, J. (2022), "Most Powerful Test Against a Sequence of High Dimensional Local Alternatives," *Journal of Econometrics*, 141–159, forthcoming. [1341]

Hörmann, S., and Kokoszka, P. (2010), "Weakly Dependent Functional Data," *The Annals of Statistics*, 38, 1845–1884. [1333]

Horváth, L., Kokoszka, P., and Reeder, R. (2013), "Estimation of the Mean of Functional Time Series and a Two Sample Problem," *Journal of the Royal Statistical Society*, Series B, 75, 103–122. [1335]

Horváth, L., Kokoszka, P., and Rice, G. (2014), "Testing Stationarity of Functional Time Series," *Journal of Econometrics*, 179, 66–82. [1333]

Horváth, L., Kokoszka, P., and Wang, S. (2021), "Monitoring for a Change Point in a Sequence of Distributions," *The Annals of Statistics*, 49, 2271–2291. [1336]

Horváth, L., Liu, Z., and Lu, S. (2022), "Sequential Monitoring of Changes in Dynamic Linear Models Applied to the US Housing Market," *Econometric Theory*, 38, 209–272. [1331,1334]

Horváth, L., and Rice, G. (2014), "Extensions of Some Classical Methods in Change Point Analysis," *Test*, 23, 219–255. [1335]

Hou, K., Xue, C., and Zhang, L. (2020), "Replicating Anomalies," *Review of Financial Studies*, 33, 2019–2133. [1339]

Ioannidis, J. P. (2005), "Why Most Published Research Findings are False," *PLoS Medicine*, 2, e124. [1339]

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2021), "Is There a Replication Crisis in Finance?" Technical Report. [1339]

Jirak, M. (2015), "Uniform Change Point Tests in High Dimension," *The Annals of Statistics*, 46, 2451–2483. [1331]

Kaul, A., Jandhyala, V., and Fotopoulos, S. (2019), "An Efficient Two Step Algorithm for High Dimensional Change Point Regression Models Without Grid Search," *Journal of Machine Learning Research*, 20, 1–40. [1331]

——— (2020), "Detection and Estimation of Parameters in High Dimensional Multiple Change Point Regression Models via ℓ_1/ℓ_0 Regularization and Discrete Optimization," Technical Report, Washington State University, Pullman. arXiv:1906.04396. [1331]

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series have a Unit Root?" *Journal of Econometrics*, 54, 159–178. [1339]

Liu, W., and Wu, W. B. (2010), "Asymptotics of Spectral Density Estimates," *Econometric Theory*, 26, 1218–1245. [1335]

Lo, A. W. (2004), "The Adaptive Markets Hypothesis," *Journal of Portfolio Management*, 30, 15–29. [1331,1339]

——— (2019), *Adaptive Markets*, Princeton: Princeton University Press. [1331,1339]

McCracken, M. W., and Ng, S. (2016), "Fred-md: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics*, 34, 574–589. [1341]

McElroy, T., and Politis, D. (2020), *Time Series*, Boca Raton, FL: CRC Press. [1333]

Meinshausen, N., Meier, L., and Bühlmann, P. (2009), "P-values for High-Dimensional Regression," *Journal of the American Statistical Association*, 104, 1671–1681. [1337]

Nosek, B. A., Spies, J. R., and Motyl, M. (2012), "Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability," *Perspectives on Psychological Science*, 7, 615–631. [1339]

Shao, X., and Wu, W. B. (2007), "Asymptotic Spectral Theory for Nonlinear Time Series," *The Annals of Statistics*, 35, 1773–1801. [1333]

Stock, J. H., and Watson, M. W. (2002), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162. [1341]

Stock, J. H. and Watson, M. W. (2005), "Implications of Dynamic Factor Models for VAR Analysis," National Bureau of Economic Research Working Paper Series No. 11467. DOI:10.3386/w11467. *http://www.nber.org/papers/w11467*. [1341]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [1333]

Trucios, C., Mazzeu, J., Hallin, M., Hotta, L., and Valls Perreira, P. (2023), "Forecasting Conditional Covariance Matrices in High-Dimensional Time Series: A General Dynamic Factor Approach," *Journal of Business & Economic Statistics*, 41, 40–52. [1331]

Uematsu, Y., and Yamagata, T. (2023), "Estimation of Sparsity-Induced Weak Factor Models," *Journal of Business & Economic Statistics*, 41, 213–227. [1332]

Wang, D., and Zhao, Z. (2022), "Optimal Change-Point Testing for High-Dimensional Linear Model with Temporal Dependence," Technical Report, University of Notre Dame. arXiv:2205.03880. [1331,1341]

Wang, F., Madrid Padilla, O., Yu, Y., and Rinaldo, A. (2022), "Denoising and Change Point Localisation in Piecewise-Constant High-Dimensional Regression Coefficients," in *International Conference on Artificial Intelligence and Statistics*, pp. 4309–4338, PMLR. [1332]

Wang, R., Zhu, C., Volgushev, S., and Shao, X. (2022), "Inference for Change Points in High-Dimensional Data via Selfnormalization," *The Annals of Statistics*, 50, 781–806. [1331]

Wang, T., and Samworth, R. (2018), "High Dimensional Change Point Estimation via Sparse Projection," *Journal of the Royal Statistical Society*, Series B, 80, 57–83. [1331]

Wu, W. B. (2005), "Nonlinear System Theory: Another Look at Dependence," *Proceedings of the National Academy of Sciences*, 102, 14150–14154. [1333]

Xu, H., Wang, D., Zhao, Z., and Yu, Y. (2022), "Change Point Inference in High-Dimensional Regression Models under Temporal Dependence," Technical Report, University of Notre Dame. arXiv:2207.12453. [1332,1342]

Zhang, C.-H., and Huang, J. (2008), "The Sparsity and Bias of the Lasso Selection in High-Dimensional Linear Regression," *The Annals of Statistics*, 36, 1567–1594. [1337]

Zhang, X. (2016), "White Noise Testing and Model Diagnostic Checking for Functional Time Series," *Journal of Econometrics*, 194, 76–95. [1333]

Zhang, X., and Cheng, G. (2018), "Gaussian Approximation for High Dimensional Vector Under Physical Dependence," *Bernoulli*, 24, 2640–2675. [1333]

Zhang, Y., Wang, R., and Shao, X. (2023), "Adaptive Inference for Change Points in High-Dimensional Data," *Journal of the American Statistical Association*, forthcoming. [1332]

Zhou, Z. (2015), "Inference for Non-stationary Time Series Regression with or Without Inequality Constraints," *Journal of the Royal Statistical Society*, Series B, 77, 349–371. [1333]

Zhu, F., Liu, M., Ling, S., and Cai, Z. (2023), "Testing for Structural Change of Predictive Regression Model to Threshold Predictive Regression Model," *Journal of Business & Economic Statistics*, 41, 228–240. [1331]

Zou, H. (2006), "The Adaptive Lasso and its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [1337]