Sieve bootstrap inference for linear time-varying coefficient models*

Marina Friedrich[a,b] and Yicong Lin[a,b]

[a]Vrije Universiteit Amsterdam
[b]Tinbergen Institute

Abstract

We propose a sieve bootstrap framework to conduct pointwise and simultaneous inference for time-varying coefficient regression models based on a local linear estimator. The asymptotic validity of the sieve bootstrap in the presence of autocorrelation is established. The bootstrap automatically produces a consistent estimate of nuisance parameters, both at the interior and boundary points. In addition, we develop a bootstrap-based test for parameter constancy and examine its asymptotic properties. An extensive simulation study demonstrates a good finite sample performance of our methods. The proposed methods are applied to assess the price development of CO_2 certificates in the European Emissions Trading System. We find evidence of time variation in the relationship between allowance prices and their fundamental price drivers. The time variation might offer an explanation for previous contradicting findings using linear regression models with constant coefficients.

JEL classifications: C14, C22, Q48, Q56.
Keywords: sieve bootstrap, nonparametric estimation, simultaneous confidence bands, energy economics, emission trading.

*Corresponding author: Marina Friedrich. E-mail: m.friedrich@vu.nl.

1 Introduction

Many climatological and economic series, and their interrelations, are subject to time variation. Attention has been drawn to nonparametric and semiparametric methods in applied studies, mainly due to their flexibility and robustness to model misspecification (e.g., Chang et al., 2016). A common way to capture time-varying behavior, among many others, is to allow coefficients in linear regression models to evolve deterministically and smoothly over time.[1] Econometric methods for these time-varying regression models have been developed in various contexts, see e.g., Phillips et al. (2017) and Li et al. (2020) for nonlinear cointegration, and Yousuf and Ng (2021) for high-dimensional predictive regressions. More work in this field includes Cai (2007), Li et al. (2011), Kristensen (2012), Zhang and Wu (2015), Chen (2015). Notwithstanding this body of work, little attention has been paid to constructing simultaneous confidence bands around coefficient curves. To the best of our knowledge, there are only two papers investigating this issue. Zhou and Wu (2010) and Karmakar et al. (2021) provide asymptotic simultaneous bands based on novel Gaussian approximations. Our paper adds to this literature by proposing a new bootstrap framework for this purpose.

In fact, residual bootstrap methods have recently been applied in the context of time-varying coefficient models although the theoretical justification has been missing from the literature (see e.g., Cai et al., 2018; Li and Zhao, 2019; Churchill et al., 2020; Liddle et al., 2020; Uddin et al., 2020). This underlines the importance of bootstrap methods for empirical work in this direction. At the same time, it emphasizes the need for a thorough investigation of the performance of bootstrap methods in this context. The current paper fills this gap. We establish the asymptotic validity of an autoregressive (AR) sieve bootstrap procedure for time-varying coefficient models. The AR sieve bootstrap handles serial dependence of time series data based on approximating a linear process by a finite autoregressive process of increasing order (with the sample size), and resampling from the centered fitted residuals. A similar idea has been studied by Bühlmann (1998) and Friedrich et al. (2020) for nonparametric trend estimation using local constant kernel smoothing. Our procedure extends this line of research by allowing for random regressors and by considering the local linear kernel smoother which shows superior performance at the boundaries (Cai, 2007). Additionally, we develop a bootstrap-based test for parameter constancy. This test can be used as an initial step in modeling. We show in an extensive simulation study that the proposed methods perform well in finite samples.

Compared to the existing asymptotic confidence bands by Zhou and Wu (2010) and Karmakar et al. (2021), we see two main advantages of our method. First, while asymptotic methods require consistent estimation of nuisance parameters, such as the asymptotic bias and long-run variance (LRV), the bootstrap correctly reproduces these parameters without an extra estimation step. The second-order bias term is not simple to estimate and often requires a careful bandwidth selection (Neumann and Polzehl, 1998). It is also well acknowledged that inference based on consistent LRV estimators can perform very poorly in finite samples with strong dependence (Müller, 2007). Second, the bootstrap confidence bands yield accurate coverage with small sample sizes, as illustrated in

[1] Alternative methods that can handle changing behavior include, for example, indicator saturation techniques which focus on structural breaks in various forms (Castle and Hendry, 2019) and smooth transition models (González and Teräsvirta, 2008). These parametric methods have been applied to a wide range of topics in economics and climate science, see e.g., Pretis et al. (2016), Holt and Teräsvirta (2020), Doornik et al. (2021), and references therein.

our simulation study. In contrast, a relatively large sample size is usually required to achieve the theoretical coverage due to the slow convergence rate for the asymptotic simultaneous confidence bands as discussed in Zhou and Wu (2010) and Karmakar et al. (2021).

One market in which time variation has become a topic of interest is the European Emissions Trading System (EU ETS). Specifically, the relationship between the prices of CO_2 emission allowances and their fundamental price drivers such as coal and gas prices is suspected to be subject to change (Lutz et al., 2013). However, to our best knowledge, none of the related papers has applied a formal, statistical test for parameter constancy. We provide first evidence for time variation with our proposed test. It potentially explains why the coal price is often found to be insignificant in linear time-constant regressions.

The paper is structured as follows. Section 2 introduces our model and the nonparametric estimator. In Section 3, we introduce the sieve bootstrap and establish its validity. We also present the bootstrap-based test for parameter stability. Section 4 discusses the practical implementation. Section 5 presents the simulation study and Section 6 the empirical application. Section 7 concludes.

Finally, some words on notation. For a vector $\boldsymbol{x} = (x_j) \in \mathbb{R}^n$, its p-norm is denoted by $\|\boldsymbol{x}\|_p = (\sum_{j=1}^n |x_j|^p)^{1/p}$. Similarly, $\|\boldsymbol{A}\|_p = \sup_{\boldsymbol{x}\neq\boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}\|_p/\|\boldsymbol{x}\|_p$ stands for the induced p-norm for a matrix $\boldsymbol{A}$. We will omit the subscripts whenever $p = 2$. The Kronecker product is denoted by "$\otimes$". For vectors $\boldsymbol{a} = (a_j)$ and $\boldsymbol{b} = (b_j)$, $\boldsymbol{a} \geq \boldsymbol{b}$ means $a_j \geq b_j$ for all j. The symbols "$\xrightarrow{p}$" and "$\xrightarrow{d}$" denote convergence in probability and in distribution, respectively. Bootstrap quantities are given a superscript *, expressing that they are conditional on the original sample. For instance, "$\xrightarrow[p]{d^*}$" denotes bootstrap weak convergence in probability (cf. Gine and Zinn (1990)). Let $\mathcal{C}^i\mathcal{I}$, $i \in \mathbb{N}$, be the collection of functions that have i_{th}-order continuous derivatives on the interval $\mathcal{I} \subset \mathbb{R}$, and $f^{(i)}(x) = \frac{d^i}{dx^i} f(x)$ be the i-th derivative with respect to x. The generic constant C can change from line to line.

2 The model and nonparametric estimation

Consider the following linear time-varying coefficient model:

$$y_t = \beta_{0,t} + \sum_{j=1}^{d} \beta_{j,t} x_{j,t} + z_t = \boldsymbol{\beta}_t' \boldsymbol{x}_t + z_t, \qquad t = 1 \ldots, n, \tag{2.1}$$

where $\boldsymbol{\beta}_t = (\beta_{0,t}, \beta_{1,t}, \ldots, \beta_{d,t})'$ is a $(d+1)$-dimensional vector of time-varying coefficients, and $\boldsymbol{x}_t = (x_{0,t}, x_{1,t}, \ldots, x_{d,t})'$ is a vector of covariates. We observe the data $\{(y_t, \boldsymbol{x}_t)\}_{t=1}^n$. We shall assume that $\{(z_t, \boldsymbol{x}_t)\}$ is a stationary process (Assumption A1) and $\boldsymbol{\beta}_t := \boldsymbol{\beta}(t/n)$ with $\boldsymbol{\beta}(\cdot) = (\beta_0(\cdot), \ldots, \beta_d(\cdot))' : [0,1] \to \mathbb{R}^{d+1}$ being a vector of functions (Assumption A2). An advantage of this model is that it circumvents the curse of dimensionality commonly arising in unrestricted nonparametric regressions while maintaining the flexibility to capture the observed nonlinearity.

The aim of this paper lies in the inference on the parameter curves $\boldsymbol{\beta}_t$. One can conduct pointwise and simultaneous inference using the asymptotic results in Cai (2007) and Zhou and Wu (2010), respectively. However, asymptotic confidence bands usually require consistent estimation of nuisance parameters which is not simple (Sections 4.1 and 4.3 of Zhou and Wu (2010)). Zhou and Wu (2010) and Karmakar et al. (2021) have further shown that the theoretical simultaneous

confidence bands have slow logarithmic convergence. A large sample size is needed to achieve the desired coverage. To alleviate these difficulties, we propose a sieve bootstrap procedure.

2.1 The local linear estimator and assumptions

Initial estimation of $\boldsymbol{\beta}_t$ is needed for the residual bootstrap. Since local linear estimators suffer less from boundary effects and have a smaller bias than Nadaraya-Watson-type estimators (Cai, 2007), we will consider the former approach here. For $\tau \in (0,1]$, write $\beta_j(t/n) \approx \beta_j(\tau) + \beta_j^{(1)}(\tau)(t/n-\tau)$, $j=0,\ldots,d$. The estimator is based on the following approximation:

$$y_t \approx \boldsymbol{\beta}(\tau)'\boldsymbol{x}_t + \boldsymbol{\beta}^{(1)}(\tau)'\boldsymbol{x}_t(t/n-\tau) + z_t =: \tilde{\boldsymbol{x}}_t(\tau)'\boldsymbol{\theta}(\tau) + z_t, \tag{2.2}$$

where $\boldsymbol{\beta}^{(1)}(\tau) = \left(\beta_0^{(1)}(\tau),\ldots,\beta_d^{(1)}(\tau)\right)'$, $\boldsymbol{\theta}(\tau) = \left(\boldsymbol{\beta}(\tau)', \boldsymbol{\beta}^{(1)}(\tau)'\right)'$, and $\tilde{\boldsymbol{x}}_t(\tau) = (\boldsymbol{x}_t', \boldsymbol{x}_t'(t/n-\tau))'$. The local linear estimator minimizes the following weighted sum of squares:

$$\widehat{\boldsymbol{\theta}}(\tau) = \begin{pmatrix} \widehat{\boldsymbol{\beta}}(\tau) \\ \widehat{\boldsymbol{\beta}}^{(1)}(\tau) \end{pmatrix} = \underset{\boldsymbol{\theta}(\tau)}{\operatorname{argmin}} \sum_{t=1}^{n} \left(y_t - \tilde{\boldsymbol{x}}_t(\tau)'\boldsymbol{\theta}(\tau)\right)^2 K\left(\frac{t/n-\tau}{h}\right), \tag{2.3}$$

where $K(\cdot)$ is a kernel function and $h>0$ is a bandwidth. Let $\tau_t = t/n$. The solution to this minimization problem has a closed-form expression:

$$\widehat{\boldsymbol{\theta}}(\tau) = \begin{pmatrix} \boldsymbol{S}_{n,0}(\tau) & \boldsymbol{S}_{n,1}'(\tau) \\ \boldsymbol{S}_{n,1}(\tau) & \boldsymbol{S}_{n,2}(\tau) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{T}_{n,0}(\tau) \\ \boldsymbol{T}_{n,1}(\tau) \end{pmatrix} =: \boldsymbol{S}_n^{-1}(\tau)\boldsymbol{T}_n(\tau), \qquad \tau \in [0,1], \tag{2.4}$$

where, for $k=0,1,2$,

$$\begin{aligned} \boldsymbol{S}_{n,k}(\tau) &= \frac{1}{nh}\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t'(\tau_t-\tau)^k K\left(\frac{\tau_t-\tau}{h}\right), \\ \boldsymbol{T}_{n,k}(\tau) &= \frac{1}{nh}\sum_{t=1}^{n} \boldsymbol{x}_t(\tau_t-\tau)^k K\left(\frac{\tau_t-\tau}{h}\right) y_t. \end{aligned} \tag{2.5}$$

Our asymptotic analysis hinges on the following regularity conditions.

Assumptions:

A1 *Let $\delta>0$ be some constant. Suppose $\{(z_t,\boldsymbol{x}_t)\}_{t\in\mathbb{Z}}$ is a strictly stationary and mixing process satisfying the following conditions.*

- (a) *$\{\boldsymbol{x}_t\}_{t\in\mathbb{Z}}$ is strictly stationary α-mixing with mixing coefficients $\alpha(m) = O(m^{-\varphi})$, where $\varphi = \max\{(2+\delta)(1+\delta)/\delta, 3(1+\delta)/\delta\}$. Moreover, $\mathbb{E}\|\boldsymbol{x}_t\|^{2(2+\delta)} < \infty$.*
- (b) *Assume $z_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$ with $\psi_0 = 1$, where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence of continuous variables with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}|\varepsilon_t|^{2(2+\delta)} < \infty$.*
- (c) *The density function f_ε of ε_t satisfies $\int_{x\in\mathbb{R}} |f_\varepsilon(x+a) - f_\varepsilon(x)|\,dx \le M|a|$, $M<\infty$, whenever $|a| \le \tau$ for some $\tau>0$.*
- (d) *The lag polynomial function $\Psi : z \mapsto \sum_{j=0}^{\infty}\psi_j z^j$, $z\in\mathbb{C}$, is bounded away from zero for $|z|\le 1$. Moreover, for some $\lambda > (5\varphi+7)/4$, $|\psi_j| \ll C_1 j^{-\lambda}$ as $j\to\infty$, where φ is defined*

in A1(a).

(e) *The next moment conditions hold almost surely: (i)* $\mathbb{E}(z_t|\boldsymbol{x}_t) = 0$*; (ii)* $\mathbb{E}(z_t z_s|\boldsymbol{x}_t, \boldsymbol{x}_s) = \mathbb{E}(z_t z_s)$*,* $s, t \in \mathbb{Z}$*.*

A2 The coefficient function $\boldsymbol{\beta}(\cdot) \in \mathcal{C}^3[0,1]$*, i.e.,* $\beta_j(\cdot) \in \mathcal{C}^3[0,1]$*,* $j = 0, 1, \ldots, d$*.*

A3 The kernel function $K(\cdot)$ *is positive, symmetric, Lipschitz continuous, and has compact support* $[-1,1]$ *with* $\mu_0 \equiv \int_{-1}^{1} K(u)\,du = 1$*.*

A4 The bandwidth $h \equiv h(n)$ *satisfies* $\max\left\{h, \frac{\ln n}{nh}, \frac{1}{nh^2}, nh^7, h^4 \ln n\right\} \overset{n\to\infty}{\to} 0$*.*

Assumption A1(a) reflects a tradeoff between the mixing coefficients and the moments. The condition on the mixing coefficient enables us to apply the well-known uniform convergence results for strong mixing processes established in Theorem 2 by Hansen (2008). Although it is crucial for some empirical applications, as required in Assumption A1(a), we do not allow for nonstationary regressors in model (2.1). It is mainly because of the asymptotic degeneracy issue of the kernel-weighted signal matrix $\boldsymbol{S}_{n,k}(\tau)$ given in (2.5), see Phillips et al. (2017) and Li et al. (2020). A non-standard technique using a path-dependent local coordinate transformation can be applied to resolve this degeneracy. However, this issue dramatically complicates the asymptotic theory and is beyond our current scope.

Assumption A1(b) imposes a linear structure on the error process, which is usually required for the autoregressive sieve bootstrap (SB) procedure (e.g., Park (2002), Palm et al. (2008), Smeekes and Taylor (2012)). We follow a large body of previous literature to impose this linear dependence structure. It is worth noting that Theorem 3.1 in Kreiss et al. (2011) has shown wide applicability of the SB even under nonlinear dependence. Yet, unlike their Eq. (3.1), our core statistics typically rely on the properties of both $\{z_t\}$ and $\{\boldsymbol{x}_t\}$. Showing the validity of the SB in our semiparametric regression models allowing for more general error processes is a nontrivial task.

Not all linear processes are strong mixing, see e.g., Bosq (1998, Chapter 1.1). To apply the results of uniform convergence mentioned above, some degree of smoothness on the distributions of $\{\varepsilon_t\}$ is needed as given in Assumption A1(c). Many common densities fulfill this condition. For instance, if $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, by the mean value theorem, we have $\int_{x\in\mathbb{R}} |f_\varepsilon(x+a) - f_\varepsilon(x)|\,dx = \sqrt{\frac{2}{\pi\sigma_\varepsilon^2}}\,|a|$. If $\varepsilon_t \sim t(\nu)$, $\nu > 0$, we have $\int_{x\in\mathbb{R}} |f_\varepsilon(x+a) - f_\varepsilon(x)|\,dx = \frac{2}{\sqrt{\nu}B(\nu/2,1/2)}|a|$, where $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$ is the Beta function.

The first part of Assumption A1(d) ensures the invertibility of the MA(∞) process. The condition $|\phi_j| \ll C_1 j^{-\lambda}$ implies that $\{z_t\}$ is a short-memory process with $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ and $\sum_{j=0}^{\infty} j|\,\mathbb{E}(z_t z_{t+j})\,| < \infty$. Any causal and invertible ARMA(p,q) model with $0 \le p, q < \infty$ satisfies this condition. Moreover, by Lemma 2.1 of Bühlmann (1995), $\{z_t\}$ admits an AR(∞) representation

$$\sum_{j=0}^{\infty} \phi_j z_{t-j} = \varepsilon_t, \tag{2.6}$$

where $\phi_0 = 1$ and $\sum_{j=0}^{\infty} j|\phi_j| < \infty$. By Theorem 14.9 in Davidson (1994), Assumption A1(b), (c), and (d) jointly imply that $\{z_t\}$ is a strictly stationary α-mixing process with $\alpha(m) \ll Cm^{-\varphi}$,

where φ is given in A1(a). Therefore, the uniform convergence results can be applied to the process $\{(z_t, \boldsymbol{x}_t)\}$. Given z_t satisfies A1(b), and using the uniform results in Li et al. (2012), it is possible to only require $\{(z_t, \boldsymbol{x}_t)\}$ to be a strictly stationary L_p-NED process with $p > 4$ under some high-level assumptions.

It is worthwhile to mention that our Assumption A1(e)(ii) is stronger than Assumption 3 in Cai (2007) in two ways. First, we rule out conditional heteroscedasticity of the form $\mathbb{E}\left(z_t^2|\boldsymbol{x}_t\right) = g(\boldsymbol{x}_t, t/n)$, where $g(\cdot,\cdot)$ is a continuous and bounded function of the stationary regressors and rescaled time. Although $\widehat{\boldsymbol{\theta}}(\tau)$ retains consistency and asymptotic normality in the presence of heteroscedasticity, the sieve bootstrap scheme may not work. In this case, it is natural to consider variants of the wild bootstrap. For instance, when the model has unconditional heteroscedasticity, i.e., $\mathbb{E}\left(z_t^2|\boldsymbol{x}_t\right) = g(t/n)$, the sieve wild bootstrap (SWB) may be a solution, see Remark 5. In the Monte Carlo study (Section 5.1.2), we shall explicitly investigate the robustness of our methods when the homoskedasticity assumption is violated. Second, Assumption A1(e)(ii) requires some "independence" in terms of second-order moments between $\boldsymbol{x}_t$ and z_t. A similar but stronger condition has been imposed in Kapetanios (2008, Assumption 3) for the i.i.d. bootstrap. We will come back to this point in Remark 3.

Assumption A2 is a common smoothness condition, see e.g., Assumption 6 in Zhou and Wu (2010). It can be weakened by assuming that $\boldsymbol{\beta}(\cdot)$ is twice continuously differentiable with minor modification of the proofs. Assumption A3 on the kernel function $K(\cdot)$ is satisfied by many commonly used kernels such as the Epanechnikov kernel. Assumption A4 ensures consistency and asymptotic normality of the local linear estimator. The natural logarithm in $\ln n/nh \overset{n\to\infty}{\to} 0$ is due to our uniform convergence results. It is not stringent as common bandwidth choices have the rate $n^{-\kappa}$ with $\kappa \in (0,1)$, satisfying this condition.

3 Confidence bands with the autoregressive (AR) sieve bootstrap

We now propose an (AR) sieve bootstrap to construct pointwise/simultaneous confidence bands. A similar method was discussed in Bühlmann (1998) for a deterministic trend model.

3.1 The sieve bootstrap (SB)

Recall that the errors admit an AR(∞) process in (2.6). The focus of the SB lies in approximating the dependence structure of the error terms by AR(p) models. We first estimate model (2.1). To the residuals from this estimation we fit an AR(p) model and obtain the residuals from which we draw the bootstrap errors. We can describe the bootstrap algorithm in six steps:

Step 1 Estimate model (2.1) and form a residual series. This means, calculate

$$\hat{z}_t = y_t - \boldsymbol{x}_t'\widetilde{\boldsymbol{\beta}}(t/n), \qquad t = 1, \ldots, n,$$

where the estimate $\widetilde{\boldsymbol{\beta}}(t/n)$ is obtained by bandwidth $\tilde{h} > h$.

Step 2 To the residuals $\hat{z}_t$, for $t = 1, \ldots, n$, fit an AR(p) model and form the new series of residuals

$$\widehat{\varepsilon}_{t,p} = \hat{z}_t - \sum_{j=1}^{p} \widehat{\phi}_j \hat{z}_{t-j}, \qquad t = p+1, ..., n.$$

Recenter the residuals $\tilde{\varepsilon}_{t,p} = \hat{\varepsilon}_{t,p} - \frac{1}{n-p} \sum_{t=p+1}^{n} \widehat{\varepsilon}_{t,p}$.

Step 3 Draw randomly with replacement from $\{\tilde{\varepsilon}_{t,p}\}$ to obtain $\{\varepsilon_t^*\}$.

Step 4 Calculate the bootstrap errors z_t^* as $z_t^* = \sum_{j=1}^{p} \widehat{\phi}_j z_{t-j}^* + \varepsilon_t^*$. Generate bootstrap observations

$$y_t^* = \boldsymbol{x}_t' \widetilde{\boldsymbol{\beta}}(t/n) + z_t^*, \qquad t = 1, \ldots, n,$$

where $\widetilde{\boldsymbol{\beta}}(t/n)$ is the same as in the first step.

Step 5 Obtain the bootstrap estimator $\hat{\boldsymbol{\beta}}^*(\cdot)$ as defined in (2.4) using the bootstrap series $\{y_t^*\}$, with the same bandwidth h as used for the original estimate $\hat{\boldsymbol{\beta}}(\cdot)$.

Step 6 Repeat Steps 3 to 5 B times, and let

$$\hat{q}_{j,\alpha}(\tau) = \inf \left\{ u \in \mathbb{R} : \mathbb{P}^* \left(\hat{\beta}_j^*(\tau) - \tilde{\beta}_j(\tau) \leq u \right) \geq \alpha \right\}, \qquad j = 0, \ldots, d, \tag{3.1}$$

denote for the 100αth percentile of the B centered bootstrap statistics $\hat{\beta}_j^*(\tau) - \tilde{\beta}_j(\tau)$. These bootstrap quantiles are used to construct confidence bands as described in the next section.

The lag length p in Step 2 should satisfy Assumption B2. How to select it in practice is discussed in Section 4. Note that in Step 1, a different bandwidth $\tilde{h}$ is used to perform the nonparametric estimation. Compared to the original bandwidth h, this bandwidth produces an oversmoothed estimate. The reason for this is the presence of the asymptotic bias with local polynomial estimation. The bias contains the second derivatives of the coefficient functions, which can only be consistently estimated using a larger, oversmoothed bandwidth. Alternatively, one may use an undersmoothing bandwidth which attempts to eliminate the bias asymptotically, see e.g., Neumann and Polzehl (1998). We follow Bühlmann (1998) and find the oversmoothing works well in practice. In Remark 4, we provide an intuition of why oversmoothing can consistently estimate the asymptotic bias.

Assumptions:

B1 The oversmoothing bandwidth $\tilde{h} = \tilde{h}(n)$ *satisfies* $\max\left\{\tilde{h}, nh\tilde{h}^4, h \ln n/\tilde{h}\right\} \to 0$ *as* $n \to \infty$.

B2 The lag order $p = p(n) \to \infty$ *with* $p \max\left\{\tilde{h}, \left(\ln n/(n\tilde{h})\right)^{1/4}\right\} \to 0$ *as* $n \to \infty$.

We consider the asymptotic properties of the bootstrap estimators at boundary points (Section 3.3) unlike Theorem 3.1 in Bühlmann (1998) and Theorem 2 in Friedrich et al. (2020). It comes at the cost of requiring a slightly stronger condition $h \ln n/\tilde{h} \overset{n\to\infty}{\to} 0$, compared to Assumption (K) in Bühlmann (1998) and Assumption 8 in Friedrich et al. (2020).

Remark 1. Residuals that correspond to a time point close to the boundary might be problematic as nonparametric estimators exhibit edge effects. The accuracy of estimates for points close to the boundary thus cannot be guaranteed. As a solution, Bühlmann (1998, p. 53) suggests only to resample residuals for points $\tau \in [\delta, 1-\delta]$ for a small $\delta > 0$. Formally, this means that in Step 1, residuals $\hat{z}_t$ are obtained for $t = \lfloor n\delta \rfloor + 1, \ldots, \lfloor n(1-\delta) \rfloor$, where $\lfloor x \rfloor$ is the integer part of $x \in \mathbb{R}$. The remainder of the bootstrap procedure then proceeds with this smaller set of residuals. However, given that we use a local linear estimator and not a local constant approach as in Bühlmann (1998), we expect the effect to be small. Our preliminary simulation study confirmed this conjecture.

3.2 Constructing confidence bands

To construct pointwise confidence intervals for every $\beta_j(\cdot)$, the quantity $\hat{\beta}_j^*(\cdot) - \tilde{\beta}_j(\cdot)$ is needed for $j = 0, \ldots, d$. It is straightforward to determine pointwise two-sided confidence intervals for a confidence level of $1-\alpha$. These are exactly the values, for every τ, between which $1-\alpha$ of the deviations fall. Formally, this can be stated as

$$I_{j,n,\alpha}^*(\tau) = \left[\hat{\beta}_j(\tau) - \hat{q}_{j,1-\alpha/2}(\tau),\ \hat{\beta}_j(\tau) - \hat{q}_{j,\alpha/2}(\tau)\right], \qquad \tau \in (0,1), \tag{3.2}$$

where $\hat{q}_{j,\alpha}(\tau)$ is defined in (3.1). From Eq. (3.2), it can be seen that the confidence intervals are only valid for a fixed point $\tau \in (0,1)$. In general, asymptotic pointwise confidence intervals $I_{j,n,\alpha}(\tau)$ for $\beta_j(\tau)$ are designed to satisfy

$$\liminf_{n\to\infty} \mathbb{P}\big(\beta_j(\tau) \in I_{j,n,\alpha}(\tau)\big) \geq 1-\alpha, \qquad \tau \in (0,1). \tag{3.3}$$

Many interesting research questions, like whether a coefficient remains zero over the whole period or whether there is an upward trend over a certain period of time, cannot be answered with pointwise confidence intervals. Simultaneous confidence bands are needed to answer these questions. That is, for $G \subset [0,1]$, we seek for $I_{j,n,\alpha}^G(\cdot)$ that satisfies

$$\liminf_{n\to\infty} \mathbb{P}\left(\beta_j(\tau) \in I_{j,n,\alpha}^G(\tau),\ \forall \tau \in G\right) \geq 1-\alpha. \tag{3.4}$$

As the bootstrap counterpart we consider a three-step procedure which is similar to the one described in Bühlmann (1998) and Friedrich et al. (2020). This procedure provides confidence bands which are simultaneous within a finite union of neighborhoods $G = \bigcup_{i=1}^m U_i(h) \subset [0,1]$ of the form $U_i(h) = [\tau_i - ah, \tau_i + bh]$, with $\tau_i \in (0,1)$, $i = 1, \ldots, m$, $m \in \mathbb{Z}^+$, and $0 \leq a, b < \infty$. Clearly, the length of these neighborhoods depends on the bandwidth used for the original estimation, and through the bandwidth it depends on the sample size. The first step is to construct pointwise quantiles from the deviations $\hat{\beta}_j^*(\cdot) - \tilde{\beta}_j(\cdot)$:

Step 1 Compute the pointwise quantiles $\hat{q}_{j,\alpha_p/2}(\tau), \hat{q}_{j,1-\alpha_p/2}(\tau)$ by varying $\alpha_p \in [1/B, \alpha]$, for $\tau \in G$, $j = 0, \ldots, d$.

Step 2 Choose $\hat{\alpha}_s = \hat{\alpha}_s(\alpha)$ as

$$\hat{\alpha}_s = \underset{\alpha_p \in [1/B, \alpha]}{\operatorname{argmin}} \left| \mathbb{P}^* \left(\hat{q}_{j,\alpha_p/2}(\tau) \leq \hat{\beta}_j^*(\tau) - \tilde{\beta}_j(\tau) \leq \hat{q}_{j,1-\alpha_p/2}(\tau),\ \forall \tau \in G \right) - (1-\alpha) \right|.$$

Step 3 Given $\hat{\alpha}_s$ from Step 2, construct the simultaneous confidence bands as

$$I^{G*}_{j,n,\hat{\alpha}_s}(\tau) = \left[\hat{\beta}_j(\tau) - \hat{q}_{j,1-\hat{\alpha}_s/2}(\tau),\ \hat{\beta}_j(\tau) - \hat{q}_{j,\hat{\alpha}_s/2}(\tau)\right], \qquad \tau \in G.$$

Note that a pointwise error $\hat{\alpha}_s$ is found for which a fraction of approximately $1-\alpha$ of all centered bootstrap estimates falls within the resulting confidence bands, for all points of the set G. As such, the confidence intervals with pointwise coverage $1-\hat{\alpha}_s$ become simultaneous confidence bands with coverage $1-\alpha$.[2] Asymptotically, this procedure may lead to the so-called balance property as discussed in Beran (1988) and Romano and Wolf (2010).[3] Since this model can have more than one explanatory variable, we have to construct confidence bands for $d+1$ coefficient curves using the above procedure. It is noted that the confidence bands do not need to have equal width, a feature which we find valuable, since in finite samples the fluctuations are likely to vary.

Remark 2. There is a fundamental difference between the SB and the multiplier bootstrap (MB) presented in Section 4.2 of Zhou and Wu (2010). The MB is closely linked to the asymptotic simultaneous confidence bands:

$$\hat{\beta}_j(\tau) - h^2 b_j(\tau) \pm \frac{\sigma_{\beta_j}}{\sqrt{nh}}\left[B_K\left(\frac{1}{h}\right) + \frac{q_{1-\alpha}}{\sqrt{2\log(1/h)}}\right] =: \hat{\beta}_j(\tau) - h^2 b_j(\tau) \pm \frac{\sigma_{\beta_j}}{\sqrt{nh}} c_{1-\alpha,h},$$

where $b_j(\tau)$ and σ_{β_j} are the bias and long-run variance associated with $\hat{\beta}_j(\tau)$, respectively (see Section 3.3 below for more details). The term $c_{1-\alpha,h}$ is a result of approximating the maximum of Gaussian processes, where $B_K(u) = \sqrt{2\log u} + \log(C_K)/\sqrt{2\log u}$ with some constant C_K depending on the kernel $K(\cdot)$, and $q_{1-\alpha} = -\log\left(-0.5\log(1-\alpha)\right)$ is the $100(1-\alpha)$ percentile of the Gumbel distribution $u \mapsto e^{-2e^{-u}}$. It is well-known that this approximation merely has a logarithmic rate of convergence. Based on a strong invariance principle, the MB (only) improves the approximation by generating i.i.d. *Gaussian* draws in simulations. The nuisance parameters have to be estimated or eliminated, e.g., by jackknife, see Steps (b) and (d) on p.520 in Zhou and Wu (2010). As such, the accuracy of nuisance parameter estimates may substantially affect the empirical coverage of bands by the MB. On the other hand, by resampling from the centered fitted residuals, the SB estimates the entire quantity $h^2 b_j(\cdot) \pm \sigma_{\beta_j} c_{1-\alpha,h}/\sqrt{nh}$.[4] It directly mimics the finite sample distributions of $\hat{\beta}_j(\cdot) - \beta_j(\cdot)$ over G. Intuitively speaking, the resampling step in the SB exploits the information of the nuisance parameters, thus producing them automatically.

3.3 Asymptotic theory

An insightful exposition of our results requires further notation.

(*a*) Quantities associated with $K(\cdot)$: $\mu_k = \int_{-1}^{1} u^k K(u)du$, $\nu_k = \int_{-1}^{1} u^k K^2(u)du$; $\mu_{k,c} = \int_{-c}^{1} u^k K(u)du$, $\nu_{k,c} = \int_{-c}^{1} u^k K^2(u)du$, where $c \in (0,1)$. Let $\boldsymbol{\mu}_c = \left(\begin{smallmatrix}\mu_{0,c} & \mu_{1,c}\\ \mu_{1,c} & \mu_{2,c}\end{smallmatrix}\right)$ and $\boldsymbol{\nu}_c = \left(\begin{smallmatrix}\nu_{0,c} & \nu_{1,c}\\ \nu_{1,c} & \nu_{2,c}\end{smallmatrix}\right)$. Moreover, we

[2] We provide a MATLAB package to implement the estimation and bootstrap confidence bands on www.yiconglin.com and www.sites.google.com/view/mfriedrich.

[3] We thank one of the reviewers for pointing out the property of balance.

[4] To be more precise, the term $c_{1-\alpha,h}$ is different in our asymptotic approximation because we form the set G by local neighborhoods.

define $\boldsymbol{\kappa}(\tau_1,\tau_2)=\int_{\mathbb{R}}\boldsymbol{w}(u;\tau_1,\tau_2)K(u-\tau_1)K(u-\tau_2)du$ and $\boldsymbol{\kappa}_+(\tau_1,\tau_2)=\int_{\mathbb{R}^+}\boldsymbol{w}(u;\tau_1,\tau_2)K(u-\tau_1)K(u-\tau_2)du$, where $\boldsymbol{w}(u;\tau_1,\tau_2)=\left(\begin{smallmatrix}1 & u-\tau_2\\ u-\tau_1 & (u-\tau_1)(u-\tau_2)\end{smallmatrix}\right)$.

(*b*) Bias terms: $\boldsymbol{b}(\tau)=\frac{1}{2}\left(\begin{smallmatrix}\mu_2\boldsymbol{\beta}^{(2)}(\tau)\\ \mathbf{0}\end{smallmatrix}\right)$, $\boldsymbol{b}_c(0+)=\frac{1}{2}\boldsymbol{\mu}_c^{-1}\left(\begin{smallmatrix}\mu_{2,c}\boldsymbol{\beta}^{(2)}(0+)\\ \mu_{3,c}\boldsymbol{\beta}^{(2)}(0+)\end{smallmatrix}\right)$, where $\boldsymbol{\beta}^{(2)}(0+)=\lim_{\tau\downarrow 0}\boldsymbol{\beta}^{(2)}(\tau)$.

(*c*) Scaling matrix: $\boldsymbol{H}=\operatorname{diag}(\boldsymbol{I}_{d+1},h\boldsymbol{I}_{d+1})$.

(*d*) Short/long-run covariance matrices: $\boldsymbol{\Omega}_0=\mathbb{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$, $\boldsymbol{\Lambda}=\sum_{j=-\infty}^{\infty}\operatorname{cov}(\boldsymbol{x}_tz_t,\boldsymbol{x}_{t+j}z_{t+j})$.

The following pointwise results illustrate that the nuisance parameters in asymptotic inference are consistently estimated by the sieve bootstrap.

Theorem 1

Under Assumptions A1, A2, A3, A4, B1, and B2, we have

(*i*) *for any fixed* $\tau\in(0,1)$,

$$\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(\tau)-\widetilde{\boldsymbol{\theta}}(\tau)-h^2\boldsymbol{b}(\tau)\right)\xrightarrow{d^*}_p\mathcal{N}\left(\mathbf{0},\left(\begin{smallmatrix}\nu_0 & \\ & \nu_2/\mu_2^2\end{smallmatrix}\right)\otimes\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Omega}_0^{-1}\right)\right),\tag{3.5}$$

and $\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau)-h^2\boldsymbol{b}(\tau)\right)$ *converges in distribution to the same limit;*

(*ii*) *for the left endpoint* $\tau=ch$, $c\in(0,1)$,

$$\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(ch)-\widetilde{\boldsymbol{\theta}}(ch)-h^2\boldsymbol{b}_c(0+)\right)\xrightarrow{d^*}_p\mathcal{N}\left(\mathbf{0},\left(\boldsymbol{\mu}_c^{-1}\boldsymbol{\nu}_c\boldsymbol{\mu}_c^{-1}\right)\otimes\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Omega}_0^{-1}\right)\right),\tag{3.6}$$

and $\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}(ch)-\boldsymbol{\theta}(ch)-h^2\boldsymbol{b}_c(0+)\right)$ *converges in distribution to the same limit.*

Theorem 1 has three implications. First, the theoretical validity in the sense of (3.3) for the pointwise bootstrap intervals follows directly from the theorem. Second, the sieve bootstrap successfully replicates the asymptotic behavior of the local linear estimators at both interior points and the left boundary point. Similar results hold for the right endpoint $\tau=1-ch$ and are omitted here. Third, the bootstrap consistently estimates the bias terms $\boldsymbol{b}(\tau)/\boldsymbol{b}_c(0+)$. As mentioned, undersmoothing conditions of $\tilde{h}$ aim at making the bias vanish asymptotically. Yet, it is not necessary here. The bootstrap automatically mimics the second-order bias terms despite them being negligible or dominating the stochastic variation as also found in Friedrich et al. (2020).

We further investigate the validity of the simultaneous confidence bands as in (3.4) by considering h-neighborhoods around fixed time points. When the distance between two points, τ_1 and τ_2, is of order h, there is non-negligible correlation between $\widehat{\boldsymbol{\theta}}(\tau_1)$ and $\widehat{\boldsymbol{\theta}}(\tau_2)$. Theorem 2 below shows that the sieve bootstrap correctly mimics this neighboring correlation. This property can be used for constructing simultaneous confidence bands over G, that is, a finite union of h-neighborhoods.

Some additional notation is needed. Let $q=2(d+1)$. Denote by $C[-1,1]$ the space of real-valued continuous functions on $[-1,1]$, and by $C[-1,1]^q=C[-1,1]\times\cdots\times C[-1,1]$ the space of continuous vector functions. That is, $C[-1,1]^q$ is the q-fold Cartesian product of the space $C[-1,1]$. Let $d_U(x,y)=\sup_{\tau\in[-1,1]}|x(\tau)-y(\tau)|$, $x,y\in C[-1,1]$, be the uniform metric. The symbol "$\Rightarrow$" signifies weak convergence in $C[-1,1]^q$ endowed with $d_U^q(\boldsymbol{x},\boldsymbol{y})=\max_{1\le i\le q}\{d_U(x_i,y_i)\}$,

where $\boldsymbol{x} = (x_i), \boldsymbol{y} = (y_i) \in C[-1,1]^q$. Note that $d_U^q(\boldsymbol{x}, \boldsymbol{y})$ induces the product topology that makes $C[-1,1]^q$ complete and separable, like $C[-1,1]$, see e.g., Theorem 6.16 of Davidson (1994).

Theorem 2

Suppose the assumptions in Theorem 1 hold.

(*i*) *For any fixed* $\tau_0 \in (0,1)$,

$$\begin{aligned} &\left\{\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}(\tau_0+\tau h) - \boldsymbol{\theta}(\tau_0+\tau h) - h^2\boldsymbol{b}(\tau_0)\right)\right\}_{\tau\in[-1,1]} \Rightarrow \{\boldsymbol{W}(\tau)\}_{\tau\in[-1,1]}, \\ &\left\{\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(\tau_0+\tau h) - \widetilde{\boldsymbol{\theta}}(\tau_0+\tau h) - h^2\boldsymbol{b}(\tau_0)\right)\right\}_{\tau\in[-1,1]} \Rightarrow \{\boldsymbol{W}(\tau)\}_{\tau\in[-1,1]} \text{ in probability,} \end{aligned} \tag{3.7}$$

where $\{\boldsymbol{W}(\tau)\}_{\tau\in[-1,1]}$ *is a multivariate Gaussian process with* $\mathbb{E}\boldsymbol{W}(\tau) = \boldsymbol{0}$ *and*

$$\operatorname{cov}\left(\boldsymbol{W}(\tau_1), \boldsymbol{W}(\tau_2)\right) = \left[\operatorname{diag}\left(1, \mu_2^{-1}\right)\boldsymbol{\kappa}(\tau_1,\tau_2)\operatorname{diag}\left(1,\mu_2^{-1}\right)\right] \otimes \left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Omega}_0^{-1}\right).$$

(*ii*) *Let* $K \subset (0,1)$ *be a compact set.*

$$\begin{aligned} &\left\{\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}(\tau h) - \boldsymbol{\theta}(\tau h) - h^2\boldsymbol{b}_\tau(0+)\right)\right\}_{\tau\in K} \Rightarrow \{\boldsymbol{W}_+(\tau)\}_{\tau\in K}, \\ &\left\{\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(\tau h) - \widetilde{\boldsymbol{\theta}}(\tau h) - h^2\boldsymbol{b}_\tau(0+)\right)\right\}_{\tau\in K} \Rightarrow \{\boldsymbol{W}_+(\tau)\}_{\tau\in K} \text{ in probability,} \end{aligned} \tag{3.8}$$

where $\{\boldsymbol{W}_+(\tau)\}_{\tau\in K}$ *is a multivariate Gaussian process with* $\mathbb{E}\boldsymbol{W}_+(\tau) = \boldsymbol{0}$ *and*

$$\operatorname{cov}\left(\boldsymbol{W}_+(\tau_1), \boldsymbol{W}_+(\tau_2)\right) = \left(\boldsymbol{\mu}_{\tau_1}^{-1}\boldsymbol{\kappa}_+(\tau_1,\tau_2)\boldsymbol{\mu}_{\tau_2}^{-1}\right) \otimes \left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Omega}_0^{-1}\right).$$

Theorem 2 shows the uniform validity of the bootstrap within an h-neighborhood for any interior point as well as the left endpoint. One can similarly consider the right endpoint. Moreover, the results trivially hold for any interval $[\tau_0 - ah, \tau_0 + bh]$ with $a, b > 0$. The interval $[\tau_0 - h, \tau_0 + h]$ is simply chosen out of convenience. As shown in Corollary 3.3 in Bühlmann (1998), the uniform validity over a finite union of h-neighborhoods follows straightforwardly from the theorem. In finite samples, one can always cover the full sample by taking sufficiently many unions in G.

However, this is fundamentally different from the asymptotic construction based on direct approximations of $\sup_{\tau\in(0,1)} \left\|\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)\right\|$, see e.g., Zhou and Wu (2010) and Karmakar et al. (2021). Although our construction may be theoretically less attractive, it has some advantages. First, it provides an alternative to the the method by Zhou and Wu (2010) from a practical point of view. By choosing some representative periods while constructing G, such as the beginning and the end, the simultaneous bands over these local sets allow one to make conclusion about the most important drivers of y_t in these specific periods (see Fig. 4). Second, intuitively speaking, it better captures the local variability of functions at different points in $[0,1]$, particularly in the boundary regions. We observe in extensive simulations that our simultaneous bands are generally wider than the construction by Zhou and Wu (2010) in the boundary regions, yet similar in the middle area. However, this local variability comes with a cost. Our procedure is similar to the Bonferroni correction in spirit, where we adjust the significance level and obtain $\hat{\alpha}_s$ for simultaneous bands. The more local sets we adjust for, the more conservative bands we have. That is, we likely have a larger length

compared to the asymptotic construction based on $\sup_{\tau\in(0,1)} \|\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)\|$. As will be seen in the simulation study (Section 5.1), we have an overall mildly larger empirical length compared to Zhou and Wu (2010). Several remarks are now in place.

Remark 3. Recall $\tau_t = t/n$ and $\mathbb{E}\left(z_t z_s | \boldsymbol{x}_t, \boldsymbol{x}_s\right) = \mathbb{E}\left(z_t z_s\right)$, $s, t \in \mathbb{Z}$, in Assumption A1(e). The reason for this assumption is given as follows. Our bootstrap scheme hinges on an i.i.d. bootstrap of the AR residuals which are considered to be approximations of the i.i.d. errors ε_t. Intuitively speaking, the i.i.d. resampling of these residuals destroys the potential higher-order dependence structure between $\boldsymbol{x}_t$ and z_t. We find that the next terms mimic the asymptotic normality:

$$\boldsymbol{Z}^*_{n,k}(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \boldsymbol{x}_t z_t^* w_t^k(\tau), \qquad k = 0, 1,$$

where $w_t^k(\tau) := \left(\frac{\tau_t - \tau}{h}\right)^k K\left(\frac{\tau_t - \tau}{h}\right)$, with $\mathbb{E}^*\left(\boldsymbol{Z}^*_{n,k}(\tau)\right) = 0$ and the (conditional) variance

$$\begin{aligned}
\mathbb{V}\mathrm{ar}^*\left(\boldsymbol{Z}^*_{n,k}(\tau)\right) &= \frac{1}{nh} \sum_{i=-n+1}^{n-1} \sum_{t=1}^{n-|i|} \boldsymbol{x}_t \boldsymbol{x}'_{t+|i|} \mathbb{E}^*\left(z_t^* z_{t+|i|}^*\right) w_t^k(\tau) w_{t+|i|}^k(\tau). \\
&\approx \sum_{i=-n+1}^{n-1} \mathbb{E}\left(z_1 z_{1+|i|}\right) \left[\frac{1}{nh} \sum_{t=1}^{n-|i|} \boldsymbol{x}_t \boldsymbol{x}'_{t+|i|} \left(\frac{\tau_t - \tau}{h}\right)^{2k} K^2\left(\frac{\tau_t - \tau}{h}\right)\right] \\
&\approx \sum_{i=-n+1}^{n-1} \mathbb{E}\left(z_1 z_{1+|i|}\right) \left[\mathbb{E}\left(\boldsymbol{x}_1 \boldsymbol{x}'_{1+|i|}\right) \frac{1}{nh} \sum_{t=1}^{n-|i|} \left(\frac{\tau_t - \tau}{h}\right)^{2k} K^2\left(\frac{\tau_t - \tau}{h}\right)\right] \\
&\approx \nu_{2k} \sum_{i=-n+1}^{n-1} \mathbb{E}\left(z_1 z_{1+|i|}\right) \mathbb{E}\left(\boldsymbol{x}_1 \boldsymbol{x}'_{1+|i|}\right)
\end{aligned}$$

for any fixed $\tau \in (0, 1)$. Observe that $\sum_{i=-n+1}^{n-1} \mathbb{E}\left(z_1 z_{1+|i|}\right) \mathbb{E}\left(\boldsymbol{x}_1 \boldsymbol{x}'_{1+|i|}\right)$ converges to $\boldsymbol{\Lambda}$ provided Assumption A1(e) holds. Therefore, $\boldsymbol{Z}^*_{n,k}(\tau)$ may not correctly mimic the asymptotic variance of $\widehat{\boldsymbol{\theta}}(\tau)$ without this assumption.

Remark 4. Some intuition for the oversmoothing condition $h \ln n/\tilde{h} \overset{n\to\infty}{\to} 0$ is provided as follows. For $\tau \in (0, 1]$, we can write

$$\begin{aligned}
&\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(\tau) - \widetilde{\boldsymbol{\theta}}(\tau) - h^2 \boldsymbol{b}(\tau)\right) \\
&\qquad = \sqrt{nh}\boldsymbol{H}\left[\widehat{\boldsymbol{\theta}}^*(\tau) - \mathbb{E}^*(\widehat{\boldsymbol{\theta}}^*(\tau))\right] + \sqrt{nh}\boldsymbol{H}\left[\mathbb{E}^*(\widehat{\boldsymbol{\theta}}^*(\tau)) - \widetilde{\boldsymbol{\theta}}(\tau) - h^2 \boldsymbol{b}(\tau)\right].
\end{aligned}$$

The first part appears to mimic the stochastic variation and to capture the asymptotic normality of the bootstrap estimators. As shown in the proof of Lemma 6, the term $\mathbb{E}^*(\widehat{\boldsymbol{\theta}}^*(\tau)) - \widetilde{\boldsymbol{\theta}}(\tau) - h^2 \boldsymbol{b}(\tau)$ converges to zero but at a rate no faster than $\sqrt{\ln n/(n\tilde{h})}$. When multiplying by $\sqrt{nh}$, the second part can only vanish asymptotically provided $h \ln n/\tilde{h} \overset{n\to\infty}{\to} 0$. This is possible by oversmoothing.

Remark 5. One may add a wild component to the SB when the errors are heteroskedastic. For instance, $z_t = \sigma_t u_t$ with u_t satisfying Assumption A1 and $\sigma_t = \sigma\left(t/n\right)$ is a deterministic process that governs the shape of the volatility. More specifically, generate ε_t^* as $\varepsilon_t^* = \xi_t^* \widehat{\varepsilon}_{t,p}$, $\xi_t^* \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, instead of the original Step 3 of the SB. The other steps remain unchanged. This bootstrap is

called sieve wild bootstrap (SWB). It is suggested in, e.g., Smeekes and Taylor (2012) for unit root testing. Since the residuals are not resampled like in Step 3 of the SB, the boundary residuals stay at the boundary and edge effects, if present, are not a problem in this method (Remark 1). Another potential method that can handle serial correlation and heteroskedasticity is the autoregressive wild bootstrap (AWB). We refer the interested reader to Friedrich et al. (2020) and references therein for more details. Both the SWB and AWB are natural solutions when there is unconditional heteroskedasticity.

3.4 Testing for parameter stability

A large body of research has adopted time-constant linear regression models in related climatic applications. This gives the necessity to test the hypothesis that $\boldsymbol{\beta}(\cdot)$ does not evolve with time. That is, we test the following hypothesis:

$$H_0 : \boldsymbol{\beta}(\cdot) = \boldsymbol{c}, \qquad \text{for some unknown } \boldsymbol{c} \in \mathbb{R}^{d+1}, \tag{3.9}$$

against the alternative hypothesis that $\boldsymbol{\beta}(\cdot)$ satisfies Assumption A2 but $\boldsymbol{\beta}(\cdot) \neq \boldsymbol{c}$ for any $\boldsymbol{c} \in \mathbb{R}^{d+1}$. Under the null hypothesis, a consistent estimator of $\boldsymbol{c}$ can be obtained by OLS, i.e., $\widehat{\boldsymbol{c}} = (\hat{c}_j,\, j = 0, \ldots, d) := (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ and $\boldsymbol{y} = (y_1, \ldots, y_n)'$. In light of Theorem 1, for a fixed $\tau \in (0, 1)$, an infeasible Wald-type test statistic can be constructed:

$$\mathcal{W}_n(\tau) = \left(\widehat{\boldsymbol{\beta}}(\tau) - h^2\boldsymbol{b}(\tau) - \widehat{\boldsymbol{c}}\right)' \left(\nu_0 \boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Omega}_0^{-1}\right)^{-1} \left(\widehat{\boldsymbol{\beta}}(\tau) - h^2\boldsymbol{b}(\tau) - \widehat{\boldsymbol{c}}\right).$$

Under the null hypothesis, the OLS estimator $\widehat{\boldsymbol{c}}$ has a faster convergence rate and is, therefore, more efficient than $\widehat{\boldsymbol{\beta}}(\tau)$. Under the alternative hypothesis, $\widehat{\boldsymbol{\beta}}(\tau)$ is consistent, whereas $\widehat{\boldsymbol{c}}$ is not. As such, $\mathcal{W}_n(\tau)$ may be loosely considered as a functional extension of Hausman-type specification tests.

Clearly, $nh\mathcal{W}_n(\tau)$ is pointwise asymptotically χ^2_{d+1}-distributed by Theorem 1. Although the test is asymptotically pivotal, consistent estimators of the LRV and bias are needed to implement this test. We will further exploit the advantages of the sieve bootstrap. A similar idea has been explored by Kapetanios (2008) using an i.i.d. bootstrap procedure. Since the bootstrap automatically takes care of the standardization, we consider the following test statistic

$$\widehat{\mathcal{W}}_n(\tau) = \left(\widehat{\mathcal{W}}_{n,0}(\tau), \ldots, \widehat{\mathcal{W}}_{n,d}(\tau)\right)' := \left(\left(\hat{\beta}_0(\tau) - \hat{c}_0\right)^2, \ldots, \left(\hat{\beta}_d(\tau) - \hat{c}_d\right)^2\right)' \tag{3.10}$$

without studentizing (and correcting bias), where $\widehat{\boldsymbol{\beta}}(\tau)$ for $\tau \in G$, and $G = \cup_{i=1}^m U_i(h)$ is given below (3.4). The bootstrap procedure is given as follows. Note that the first step will be identical to Steps 1 through 4 in the previous bootstrap algorithm.

Step 1 Perform Steps 1 through 4 of the SB procedure in Section 3.1 to obtain sieve bootstrap observations y_t^* for $t = 1, \ldots, n$.

Step 2 Obtain the bootstrap estimates $\widehat{\boldsymbol{\beta}}^*(\tau)$, $\tau \in G$, using the same h as in Step 1. Construct

$$\widehat{\mathcal{W}}_n^*(\tau) = \left(\widehat{\mathcal{W}}_{n,0}^*(\tau), \ldots, \widehat{\mathcal{W}}_{n,d}^*(\tau)\right)' := \left(\left(\hat{\beta}_0^*(\tau) - \tilde{\beta}_0(\tau)\right)^2, \ldots, \left(\hat{\beta}_d^*(\tau) - \tilde{\beta}_d(\tau)\right)^2\right)'.$$

Step 3 Repeat Steps 3 and 4 B times. Compute the $(d+1)$-dimensional vector of pointwise quantiles $\widehat{\boldsymbol{q}}_{\alpha_p}(\tau) = \left(\hat{q}_{0,\alpha_p}(\tau), \ldots, \hat{q}_{d,\alpha_p}(\tau)\right)'$, $\tau \in G$, where

$$\hat{q}_{j,\alpha_p}(\tau) = \inf\left\{u : \ \mathbb{P}^*\left(\widehat{\mathcal{W}}^*_{n,j}(\tau) \leq u\right) \geq \alpha_p\right\}.$$

From these pointwise quantiles, we obtain critical values for a significance level of α using a similar procedure as introduced in Section 3.1 to construct simultaneous confidence bands. Vary the pointwise error α_p until $\mathbb{P}^*\left(\widehat{\mathcal{W}}^*_n(\tau) \geq \widehat{\boldsymbol{q}}_{1-\alpha_p}(\tau), \ \forall \tau \in G\right) \approx \alpha$. This means that the ratio of $\widehat{\mathcal{W}}^*_{n,j}(\tau) \geq \hat{q}_{j,1-\alpha_p}(\tau)$ for all $\tau \in G$ and $j \in \{0, \ldots, d\}$ is approximately α. Denote this α_p by $\hat{\alpha}_s$, and reject the null whenever $\widehat{\mathcal{W}}_n(\tau) \geq \widehat{\boldsymbol{q}}_{1-\hat{\alpha}_s}(\tau)$ for some $\tau \in G$.

Note that we generate the bootstrap observations from the nonparametric fit rather than the parametric fit. This is because $\widetilde{\boldsymbol{\beta}}(\cdot)$ is consistent under both the null and alternative hypotheses. Therefore, the dynamics of $\{y_t\}$ can be replicated in the bootstrap world even when H_0 is not true. Moreover, we construct the bootstrap test statistics $\widehat{\mathcal{W}}^*_n(\cdot)$ using $\left((\hat{\beta}^*_j(\cdot) - \tilde{\beta}_j(\cdot))^2, \ 0 \leq j \leq d\right)'$ in Step 2, instead of the natural construction $\left((\hat{\beta}^*_j(\cdot) - \hat{c}^*_j)^2, \ 0 \leq j \leq d\right)'$, where $\hat{c}^*_j$ are the bootstrap counterparts of $\hat{c}_j$. In principle, to obtain a good empirical null rejection probability, the bootstrap distribution should reflect the finite sample null distribution of the test statistic. Moreover, for good empirical power, the bootstrap distribution should be bounded under the alternative hypothesis. If not, the resulting critical values would be too large to reject the null. Our construction ensures that the bootstrap distribution mimics the test under the null asymptotically (Proposition 1 below). More importantly, $nh\widehat{\mathcal{W}}^*_n(\cdot)$ is bounded in probability under the alternative, leading to a nontrivial power. The power property is generally important for Hausman-type tests because a false negative would lead to restrictive model specifications and potentially, misleading empirical results. The following results shed some light on the asymptotic properties which are obtained based on Theorems 1 and 2.

Proposition 1

Suppose Assumptions A1, A3, A4, B1, and B2, hold.

(*i*) *For any fixed $\tau_0 \in (0,1)$, under the null hypothesis H_0 in* (3.9), $\left\{nh\widehat{\mathcal{W}}_n(\tau_0 + \tau h)\right\}_{\tau \in [-1,1]}$ *weakly converges, and* $\left\{nh\widehat{\mathcal{W}}^*_n(\tau_0 + \tau h)\right\}_{\tau \in [-1,1]}$ *weakly converges in probability, to the same limiting distribution.*

(*ii*) *Consider the alternative hypothesis $H_1 : \boldsymbol{\beta}(\cdot) \neq \boldsymbol{c}$ for any $\boldsymbol{c} \in \mathbb{R}^{d+1}$. Under H_1 and Assumption A2, for some $\tau_0 \in (0,1)$, there exists a neighborhood $U_\delta(\tau_0) = (\tau_0 - \delta, \tau_0 + \delta)$, $\delta > 0$, such that $nh\widehat{\mathcal{W}}_n(\tau) \xrightarrow{p} \infty$ for every $\tau \in U_\delta(\tau_0)$.*

Proposition 1(*i*) demonstrates that $\widehat{\mathcal{W}}^*_n(\cdot)$ asymptotically reflects the distributions of $\widehat{\mathcal{W}}_n(\cdot)$ on the finite union G of h-neighborhoods using the arguments of Corollary 3.3 in Bühlmann (1998), as also discussed below Theorem 2. Under the null, we have $\boldsymbol{\beta}(\cdot) = \boldsymbol{c}$ over $G \subset [0,1]$. Hence, our test has correct null rejection probability asymptotically. Nevertheless, we may have lower power compared to tests such as $\sup_{\tau \in [0,1]} \widehat{\mathcal{W}}_n(\tau)$, depending on the actual shape of $\boldsymbol{\beta}(\cdot)$. More specifically, we note that $nh\widehat{\mathcal{W}}^*_n(\tau) = O^*_p(1)$ uniformly over G under H_1 by Theorem 2. Proposition 1(*ii*) shows

that the test has a non-trivial power whenever $G \ni \tau_0$. If by chance $\boldsymbol{\beta}(\cdot)$ is relatively flat over G, the null may be rejected by $\sup_{\tau\in[0,1]} \widehat{\mathcal{W}}_n(\tau)$ but not our test. Hence, the former may have a better power. From a practical point of view, one can always choose multiple h-neighborhoods centered around some representative time points such as $\{1/5, \ldots, 4/5\}$ to form the set G. It is generally informative enough to tell if $\boldsymbol{\beta}(\cdot)$ is time-varying in practice.

Remark 6. Since there is no prior information about which coefficients are time-varying in our empirical study, we focus on jointly testing if $\beta_j(\cdot)$, $j = 0, \ldots d$, are constants. As pointed out by an anonymous reviewer, it is also relevant to test whether some coefficients admit certain parametric forms while others are time-invariant, i.e., $H_0 : \beta_j(\cdot) = f_j(\cdot, \boldsymbol{\theta})$, $j \in S \subsetneq \{0, 1, \ldots, d\}$, with $f_j(\cdot, \boldsymbol{\theta})$ belonging to some parametric family and $\boldsymbol{\theta}$ being a vector of unknown parameters, provided $\beta_k(\cdot) = c_k$ for some unknown constants c_k, $k \in S^{\mathrm{C}}$. Under the null, a root-n consistent estimate $f_j(\cdot, \hat{\boldsymbol{\theta}})$ may be obtained. One can further employ profile least squares to estimate the curves $\beta_j(\cdot)$, $j \in S$, consistently under the alternative, see Li et al. (2011). Then, $\widehat{\mathcal{W}}_n(\tau)$ in (3.10) can be modified to $\left(\left(\hat{\beta}_j(\tau) - f_j(\tau, \hat{\boldsymbol{\theta}})\right)^2,\ j \in S\right)'$; the bootstrap counterpart can be similarly constructed by using $\tilde{h}$.

4 Practical implementation

We discuss some issues that arise during implementation: selecting (i) the bandwidth h; (ii) the oversmoothing parameter $\tilde{h}$; and (iii) the lag order p in the bootstrap.

4.1 Bandwidth selection

Assumption A4 gives some guidance on choosing h, however, it does not provide a practical choice. We next discuss some potential ways to select the bandwidth in practice. Although an optimal bandwidth that minimizes asymptotic mean integrated squared error can be obtained, it depends on unknown quantities that are hard to estimate (e.g., the second derivative of $\boldsymbol{\beta}(\cdot)$). Here, we focus on the practical aspect.

We adopt two conventional approaches based on the Akaike information criterion (AIC, Cai (2007)) and the generalized cross-validation (GCV, Craven and Wahba (1978)) in our simulation study and application.[5] Both approaches use the $n \times n$ hat-matrix $\boldsymbol{Q}_h$ that gives $\hat{\boldsymbol{y}}(h) = \boldsymbol{Q}_h \boldsymbol{y}$, where $\hat{\boldsymbol{y}}(h) = (\hat{y}_1(h), \ldots, \hat{y}_n(h))'$. Then, the AIC and GCV criteria are defined as

$$\mathrm{AIC}(h) = \log \hat{\sigma}^2(h) + 2\frac{\mathrm{tr}\left(\boldsymbol{Q}_h\right) + 1}{n - \mathrm{tr}\left(\boldsymbol{Q}_h\right) - 2}, \qquad \mathrm{GCV}(h) = \frac{\hat{\sigma}^2(h)}{\left[1 - \mathrm{tr}\left(\boldsymbol{Q}_h\right)/n\right]^2}, \tag{4.1}$$

where $\hat{\sigma}^2(h) = n^{-1}\sum_{t=1}^{n}(y_t - \hat{y}_t(h))^2$. The resulting optimal bandwidths minimize these criteria, respectively.

Our preliminary investigation shows that AIC/GCV cannot capture the local peaks and troughs over function paths. It leads to large bandwidths and thus produces a local bias if the local peaks and troughs are too pronounced. Intuitively, these bandwidths are the "global average" of the optimal bandwidths for both wiggly and flat areas. They balance the bias-variance tradeoff globally

[5] In Section S3.6 in the online supplement, we additionally investigate the performance of another conventional approach (modified cross-validation, MCV). We find the performance of MCV does not show a qualitative difference from AIC/GCV.

and may work well for estimating functions over the complete unit interval. However, it can be problematic for the simultaneous coverage of confidence bands that are constructed for largely fluctuating functions because intersections between the confidence bands and the coefficient curves can occur when the chosen bandwidth is too large.

To better capture large variations, we further consider the local cross-validation (LCV) approach by Vieu (1991). For time series data, it is natural to combine this approach with the modified cross-validation (MCV) proposed by Chu and Marron (1991). This results in an adapted version of the LCV which we call the local modified-cross-validation (LMCV). For each $\tau \in (0,1)$, the bandwidth $\hat{h}_\tau$ is given by

$$\hat{h}_\tau = \underset{h}{\operatorname{argmin}} \operatorname{LMCV}_\tau(h), \qquad \operatorname{LMCV}_\tau(h) = \frac{1}{n}\sum_{t=1}^{n}\left[y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}^{l,h}(t/n)\right]^2 w_\tau\left(t/n\right), \tag{4.2}$$

where $w_\tau(\cdot)$ is a weight function and $\widehat{\boldsymbol{\beta}}^{l,h}(\tau)$ is the leave-$(2l+1)$-out estimator constructed by omitting the observations at $\lfloor \tau n \rfloor + i$, $-l \leq i \leq l$. That is, $\boldsymbol{S}_{n,k}(\tau)$ and $\boldsymbol{T}_{n,k}(\tau)$ in Eq. (2.4) are replaced by the following leave-$(2l+1)$-out counterparts, for $k = 0, 1, 2$,

$$\boldsymbol{S}^{l,h}_{n,k}(\tau) = \frac{1}{(n-2l-1)h}\sum_{t:|t-\tau n|>l} \boldsymbol{x}_t\boldsymbol{x}_t'(\tau_t-\tau)^k K\left(\frac{\tau_t-\tau}{h}\right),$$
$$\boldsymbol{T}^{l,h}_{n,k}(\tau) = \frac{1}{(n-2l-1)h}\sum_{t:|t-\tau n|>l} \boldsymbol{x}_t(\tau_t-\tau)^k K\left(\frac{\tau_t-\tau}{h}\right) y_t.$$

If $w_\tau(\cdot) \equiv w \in \mathbb{R}$, the LMCV coincides with the MCV by Chu and Marron (1991). Moreover, if $l = 0$, the LMCV reduces to the LCV in Vieu (1991). Note that $\hat{h}_\tau$ varies with τ. For simplicity, we use $\hat{h} = \min_{\tau\in(0,1)} \hat{h}_\tau$ as our optimal bandwidth. Although $\hat{h}_\tau$ might produce a better local fit in flat areas of functions, $\hat{h}$ leads to reasonably good coverage and length in our simulations due to the lower sensitivity to the bandwidth in these areas. In our simulated and empirical examples, we follow Vieu (1991) and take $w_\tau(\cdot)$ to be the density function of $\mathcal{N}(\tau, 0.025)$, i.e., $w_\tau(u) = (2\pi)^{-1/2}\sigma_w^{-1}\exp\left(-(u-\tau)^2/2\sigma_w^2\right)$ with $\sigma_w^2 = 0.025$. In addition to the proposed methods, we also investigate taking the average of the resulting bandwidths from AIC, GCV, and LMCV. Our simulation study shows that using LMCV with $l = 4, 6$ as well as the average rule works well in practice (see Section 5.1.3 for further details).

4.2 Bootstrap implementation

In Step 1 of the sieve bootstrap algorithms (Sections 3.1 and 3.4), a larger bandwidth $\tilde{h}$ is used to perform the nonparametric estimation. Compared to the original bandwidth h, this bandwidth is larger and produces an oversmoothed estimate. We follow the recommendation of Bühlmann (1998) to use $\tilde{h} = Ch^{5/9}$ with $C = 2$. The interested reader is referred to Section S3.3 in the supplement in which we investigate the sensitivity to the parameter C. Furthermore, we use the AIC to select the lag length $p \in [0, p_{max}]$ in Step 2. Throughout our analysis, we use $p_{max} = 10\log_{10}(n)$.

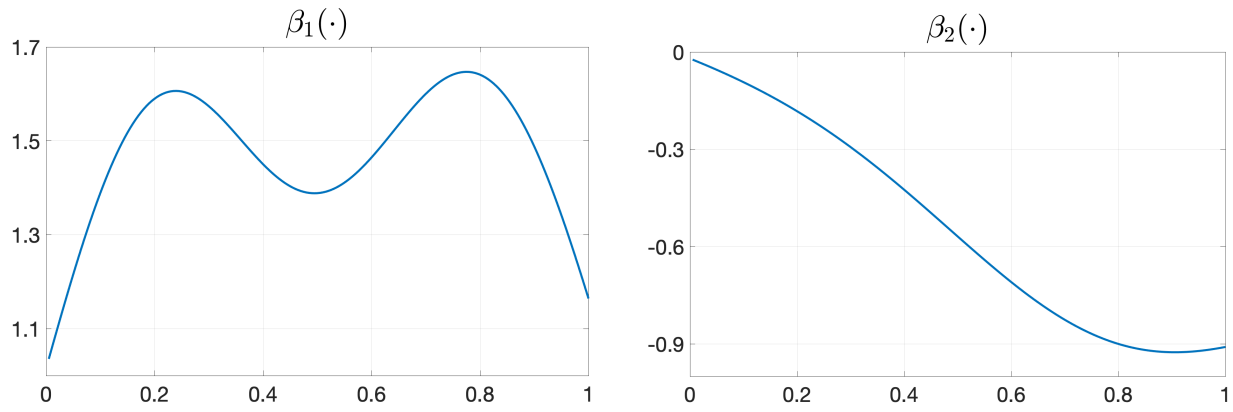


Figure 1: Plot of the coefficient curves $\beta_1(\cdot)$ and $\beta_2(\cdot)$ used in the simulation study

5 Simulation study

We now carry out a Monte Carlo study to investigate the small-sample properties of the proposed bootstrap methods. There are two main themes: (i) empirical coverage and length of confidence bands; (ii) empirical null rejection probability and power of parameter stability tests. Throughout the section, we use the Epanechnikov kernel $K(x) = \frac{3}{4}(1-x^2)\mathbb{1}_{\{|x|\leq 1\}}$, and report results based on 2,000 Monte Carlo replications and $B = 1,299$ bootstrap draws. When generating the bootstrap errors in Step 4 (Section 3.1), we take a presample size of 20 and start the recursion with a starting value of 0. Our simulated example is motivated by the empirical application in Section 6 – the price development of allowance prices in the EU Emissions Trading System. In particular, we try to mimic the relationship between allowance prices and the two main drivers, coal and gas prices.

$$y_t = \beta_1(t/n)x_{1,t} + \beta_2(t/n)x_{2,t} + u_t, \qquad t = 1, \ldots, n, \tag{5.1}$$

where $n \in \{200, 400\}$, and the two coefficient functions are given by

$$\begin{aligned} \beta_1(\tau) &= 1.5\exp(-10(\tau-0.2)^2) + 1.6\exp(-8(\tau-0.8)^2), \\ \beta_2(\tau) &= -0.5\tau - 0.5\exp(-5(\tau-0.8)^2). \end{aligned} \tag{5.2}$$

We plot the coefficient curves in Fig. 1. It shows that $\beta_1(\cdot)$ has two peaks, and $\beta_2(\cdot)$ starts at zero and then turns negative. The explanatory variables are generated based on a VAR(1) model:

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \boldsymbol{A}\begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \xi_{1,t} \\ \xi_{2,t} \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.2 \end{pmatrix},$$

where $(\xi_{1,t}, \xi_{2,t})'$ are i.i.d. bivariate standard normal. The coefficient matrix $\boldsymbol{A}$ is inspired by the empirical example as it is obtained by estimating a bivariate VAR(1) from coal and gas (log) prices in first differences. In the next sections, various specifications of the error process $\{u_t\}$ will be adopted to examine the bootstrap methods.

5.1 Finite sample properties of confidence bands

We are particularly interested in answering two questions. First, since the error dynamics are crucial to the bootstrap, how robust are our methods to different specifications of the error process? We consider two scenarios for $\{u_t\}$ in the DGP (5.1), see (i) and (ii) below, while fixing the bandwidth $h \in \{0.06, 0.09, \ldots, 0.21\}$. Second, due to the essential role of the bandwidth h in nonparametric methods, it is natural to investigate the influence of data-driven bandwidth selection on the performance of our approaches. Therefore, we shall compare the confidence bands using the approaches AIC, GCV, and LMCV (with $l = 0, 2, 4, 6$), as discussed in Section 4.1. We additionally consider the average bandwidth of these methods (AVG).

(i) Assumption A1 is satisfied. The error process is linear and has homoskedastic innovations. We assume $\{u_t\}$ follows an ARMA(1,1) process

$$u_t = \phi u_{t-1} + \epsilon_t + \psi \epsilon_{t-1}, \qquad \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \frac{1-\phi^2}{2(1+\psi^2+2\phi\psi)}\right), \tag{HM}$$

where the variance of ϵ_t is such that the signal-to-noise ratio does not change when varying (ϕ, ψ). We take $(\phi, \psi) \in \{(0,0), (0.3,0), (-0.3,0), (0,0.3), (0.3,0.3)\}$.

(ii) Assumption A1 is violated. (a) We assume unconditionally heteroskedastic errors $u_t = \sigma_t \nu_t$, where $\{\nu_t\}$ follows the ARMA(1,1) process in (HM) with $(\phi, \psi) = (0.3, 0.3)$. The volatility process, σ_t, is varied as follows:

$$\sigma_t = 1 + \kappa \cdot t/n, \tag{HT1}$$
$$\sigma_t = 1 + \kappa |\sin(t/4)|. \tag{HT2}$$

We take $\kappa = 0.5$ in the simulation. Case (HT1) allows for an increase in volatility and (HT2) for a smooth fluctuation. (b) We also consider conditional heteroskedasticity with GARCH(1,1) errors. That is, $u_t = \sigma_t \nu_t$, where ν_t are i.i.d. standard normal, and

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2. \tag{HT3}$$

We let $\omega = 1 - \alpha_1 - \alpha_2$ and $(\alpha_1, \alpha_2) \in \{(0.2, 0.7), (0.3, 0.6)\}$ as in Ling et al. (2003).

As mentioned, we study the empirical coverage and length of the confidence bands. For the empirical coverage, we consider both pointwise and simultaneous coverage. The pointwise coverage represents the fraction of the time points for which the corresponding point on the true parameter curve lies within the confidence intervals. For simultaneous coverage over a subset of $[0, 1]$, the entire coefficient curve must lie within the confidence bands over this subset (counted as a success). The (empirical) simultaneous coverage is then computed as the success rate over the Monte Carlo replications. In light of Theorem 2 and out of practical interest, we consider two subsets of $[0, 1]$, G_{sub} and G, where $G_{sub} = U_1(h) \cup U_4(h)$, $G = \bigcup_{i=1}^4 U_i(h)$, with $U_i(h) = \{(i/5) - h + j/100,\ j = 0, \ldots, \lfloor 200h \rfloor\} \bigcap [0, 1]$. We further investigate simultaneous coverage over the full sample $\{i/n,\, i = 1, \ldots, n\}$. Moreover, we compute the empirical length by first taking the median length over the grid $\{i/n,\, i = 1, \ldots, n\}$

for each pair of confidence bands, then averaging over the Monte Carlo replications.[6]

Finally, the existing multiplier bootstrap (MB) based on i.i.d. Gaussian draws, as given in Section 4.2 of Zhou and Wu (2010), will be compared as a benchmark.[7] The MB approach is originally for constructing simultaneous confidence tubes and is therefore comparable for both pointwise and simultaneous coverage. Since it only has one simultaneous construction, the empirical length does not vary from one subset of $[0, 1]$ to another.

5.1.1 Homoskedastic errors

We report the results of 95%-level confidence bands in Tables 1 and 2. The outputs for $n = 400$ are provided in Sections S3.1 and S3.2 in the supplement. The first column of each table shows which AR (ϕ) and MA (ψ) coefficients were used while the second column refers to the (fixed) bandwidth. We make three main observations.

(i) The proposed sieve bootstrap (SB) confidence bands have a substantial improvement in terms of empirical coverage compared to the MB approach, in particular for the simultaneous coverage. The difference increases when the empirical coverage is evaluated at more grid points, for instance, from G_{sub} to $\{i/n,\, i = 1, \ldots, n\}$. The improvement remains large when the sample size increases to $n = 400$ (Section S3.1). We expect the reason for this to be the crucial difference in the construction of quantiles/bounds between the SB and MB methods. As mentioned in Remark 2, the coverage of the MB depends on the accuracy of the nuisance parameter estimates. It may underestimate the "randomness" in small samples when estimating the long-run covariance matrix, leading to small empirical lengths (Table S5) and relatively low empirical (simultaneous) coverage. Admittedly, the SB exploits the AR dependence structure of the errors under more restrictive assumptions. As such, the current comparison gives an advantage to the SB. In the next section, we shall investigate the performance of both methods when Assumption A1 is violated.

(ii) For both methods, the empirical coverage is closer to the nominal level of 95% for $\beta_2(\cdot)$ than it is for $\beta_1(\cdot)$. Moreover, for the SB, the empirical coverage deteriorates quickly when h increases from 0.15 to 0.21 for $\beta_1(\cdot)$ while remaining similar for $\beta_2(\cdot)$. This sensitivity with respect to h for $\beta_1(\cdot)$ is not unexpected given the more complex shape compared to $\beta_2(\cdot)$.

(iii) For both methods, the pointwise coverage is less affected by the bandwidth choices compared to the simultaneous coverage, particularly for $\beta_1(\cdot)$. This is an expected result because the requirement of the simultaneous coverage is much stricter compared to the pointwise coverage. Simultaneous coverage fails whenever there is an intersection point between the bands and the true paths. Since the shapes of the confidence bands also rely on the ones of estimated curves, the simultaneous coverage depends on h more heavily than the pointwise counterpart. Accordingly, there exists a preferred range of bandwidth choices if one focuses on simultaneous coverage. In our simulation study, this range is $[0.09, 0.15]$ for both the SB and MB. In general,

[6] Taking median over Monte Carlo replications does not make a substantial difference.

[7] Limiting distributions are simulated based on 3,000 draws as in Zhou and Wu (2010). The tuning parameters for estimating the long-run covariance matrix are selected using the suggested simple rule on their p.521. We also find the results are not sensitive to the choice of tuning parameters as mentioned on p.522 in Zhou and Wu (2010).

a smaller h is preferred for the simultaneous bands if the function has more features to be captured, for instance, in the case with more fluctuations. In practice, this can be checked by visualizing how the function estimates change with different values of h.

5.1.2 Heteroskedastic errors

Table 3 reports the empirical coverage of $\beta_1(\cdot)$ for $n = 200$. The remaining outputs can be found in the supplement (Tables S14 - S18). We find similar patterns as in Section 5.1.1. An additional finding is as follows.

(iv) The SB shows some robustness when Assumption A1 is violated. Since it is not deliberately designed for heteroskedastic data, the lower performance of the SB is an expected result. The SB still outperforms the MB for both pointwise and simultaneous coverage, even though the latter is theoretically valid in these cases. The advantage of the SB is smaller but still substantial when n increases to 400 (Tables S15 and S17).

5.1.3 The impacts of data-driven bandwidth selection

Given that the previous results show a dependence of the performance on the bandwidth, we look at the data-driven bandwidth selection methods in Section 4.1, by fixing the error process (HM) with $(\phi, \psi) = (0.3, 0.3)$. Since the previous tables show that the empirical coverage of both the SB and MB is optimal when h is no larger than 0.18, we first display the results by performing a grid search, in each Monte Carlo replication, over a range from 0.06 to 0.2 in steps of 0.005. However, it is generally infeasible to determine such an "optimal" search range from a practical point of view. Therefore, we shall also discuss the influence of the chosen search range. The resulting empirical coverage with $n = 200$ is summarized in Table 4.[8] We make the following additional observations.

(v) Using the data-driven methods, the pointwise coverage of both $\beta_1(\cdot)$ and $\beta_2(\cdot)$ is relatively close to the nominal level of 95%. Nevertheless, in terms of the simultaneous coverage, the SB performs less satisfactorily for $\beta_1(\cdot)$. These results remind us of the sensitivity of the empirical coverage to the shapes of the functions, as discussed in (ii) and (iii) in Section 5.1.1.

(vi) As mentioned, a relatively small bandwidth is preferable for functions that fluctuate more. The left boxplot in Fig. 5 shows that AIC/GCV tends to select large boundary values in our simulation study. By extending the upper bound of our search range to 0.28, values lying close to the upper bound of 0.28 are still frequently chosen by AIC/GCV as seen in the right boxplot. As a result, the simultaneous coverage of $\beta_1(\cdot)$ using AIC/GCV is only about 88% when the upper bound is 0.2 and quickly drops to about 79% when the upper bound increases to 0.28 (Table S19 in the supplement). In general, LMCV4, LMCV6 and AVG perform reasonably well. As shown in Fig. 5, these rules are overall more robust when the search range increases (Table S19).

Despite the reasonably good performance of LMCV4, LMCV6 and AVG, blindly relying on data-driven methods may still be problematic for constructing simultaneous confidence bands. In

[8] We make similar observations about empirical coverage when $n = 400$. Yet, the empirical length of the SB and MB is 20% - 25% shorter when $n = 400$ than $n = 200$, see Table S20 in the supplement.

practice, we recommend to perform a robustness analysis using different fixed bandwidths in addition to such selection methods. A more sophisticated bandwidth selection approach for simultaneous inference, with theoretical justification, remains an open question for future research.

5.2 Parameter stability test

We investigate the empirical null rejection probability and power of the bootstrap test proposed in Section 3.4.[9] Here, we again adopt model (5.1) with the error process (HM). The same sets of G as in Section 5.1 and the full sample are used for the bootstrap test. Moreover, we rely on the same data-driven bandwidth selection methods as before. Due to the influence of the search range as discussed in (vi) above, we consider two upper bounds (0.2 and 0.28). The empirical null rejection probability and power of the tests over the full sample are reported in Table 5.[10] The corresponding results over G are similar (Table S24 in the supplement).

(vii) The empirical null rejection probability is slightly lower than the nominal level of 5% using data-driven bandwidth selection when $n = 200$, except LMCV0. It overall improves when $n = 400$ (Table S23 in the supplement). Moreover, the results show evidence of the nontrivial empirical power of the test, as expected by Proposition 1(ii). Since a false negative may result in model misspecifications, a nontrivial empirical power is crucial in applications.

$(viii)$ Unlike the simultaneous confidence bands, the parameter stability test is overall less affected by the bandwidth selection as well as the search ranges.

6 The EU Emissions Trading System

6.1 Background

The EU ETS – which is short for European Emissions Trading System – is one of the largest and oldest cap-and-trade programs. It operates in the EU, Iceland, Norway and Liechtenstein. Regulated firms and other participants can buy and sell permits that can be used to match emissions of one metric ton of CO_2 or CO_2 equivalent. These permits are called emission allowances (EUAs). The EU ETS covers around 40% of all greenhouse gas emissions of the EU. The main sector which is regulated by the scheme is the power sector. Energy intensive industries and the aviation sector are also part of the current version of the scheme. Since its implementation in 2005, the EU ETS went through three compliance phases and is currently in its fourth phase which began in 2021 (ICAP, 2020). An interesting area of research within the EU ETS is to study the impact of fundamental price drivers on the allowance price. According to economic theory of cap-and-trade markets, the price is determined by future marginal abatement costs and the cap which can be called demand- and supply-side market fundamentals, respectively (Rubin, 1996). The cap is the total amount of allowances in the market and it is determined by regulators. In the EU ETS, the cap is linearly decreasing by a given factor which recently increased from 1.74% in Phase III to its current level

[9] Next to testing for parameter stability it is of interest to test for abrupt changes against smooth variation in practice. We refer the interested reader to, e.g., Chen and Hong (2012) and Chen (2015) for asymptotic tests. To save space, we discuss the potential of adapting our bootstrap test to these situations in the supplement, Section S3.8.

[10] The results are based on 5,000 Monte Carlo replications.

of 2.2% (ICAP, 2020). Regarding the demand-side fundamentals, future abatement costs are not observable. They are replaced in the literature by observable information variables among which the coal and the gas price are the most important. The power sector is the main sector in the EU ETS and therefore, it is responsible for the identification of the main price determinants. In this sector, firms can switch from coal to gas for electricity generation. In theory, due to this fuel switching relationship, the coal price is expected to have a negative effect on allowances prices. In contrast, gas should have a positive effect. Using coal in electricity generation produces much more emissions than using gas. Thus, an increase in the price of coal, ceteris paribus, leads to a switch from coal to gas which reduces emissions and simultaneously, it decreases the demand for allowances (Christiansen et al., 2005).[11] Another price driving factor is economic activity which should have a positive effect on allowance prices. Electricity consumption and therefore emissions will increase with an increase in economic activity. In contrast, electricity generation from renewable sources should have a negative effect due to cleaner technologies, resulting in less demand for allowances. The effect of temperature on allowance prices is expected to be mainly due to increased use of heating during winter and therefore, theory predicts a negative effect.

Next to these price drivers, the expectations of market participants will play a role in price formation. This effect will, however, be hard to capture using econometric models. Given that the EU ETS has been subject to changing rules and a recent major reform, changing market participants' beliefs over the stringency of the scheme are a likely outcome which will result in changes in demand for allowances (Lutz et al., 2013). This explains the potential time variation in the relationship observed in the price data.

6.2 Data

We consider weekly data for the period from January 2008 to December 2021 resulting in $T = 734$ observations. We use price data for EUA December futures traded on the Intercontinental Exchange (ICE). Most related papers rely on the December futures prices (Aatola et al., 2013; Lutz et al., 2013; Koch et al., 2014) since they are frequently traded. As our main set of explanatory variables, we include natural gas, coal and oil prices as month-ahead futures. We further use a stock index as supplementary indicator of current and expected economic activity. As a final explanatory variable we use European average temperature data.

The gas price is the settlement price of month-ahead Dutch TTF futures, denoted in EUR/MWh. TTF stands for Title Transfer Facility and is a virtual trading point for natural gas in the Netherlands. Similarly, we consider the settlement price of month-ahead coal futures based on the API2 index of the ARA region (Amsterdam-Rotterdam-Antwerp). The contract size is 1000 metric tonnes of thermal coal.[12] For oil we rely on the historical futures prices (continuous contract) of Brent crude oil based on raw data from the ICE. The contract size is 1,000 barrels. The stock index is the STOXX Europe 50. The EUA and gas prices as well as the stock index data are retrieved from Nasdaq Data Link. The coal and oil prices are obtained from FactSet. They need to be converted

[11] Conversely, the allowance price is not expected to have a significant effect on coal and gas prices. This assumption is also made in related literature that analyzes the relationship between the energy sector and carbon markets where it is used as an identification restriction in a SVAR analysis (Lovcha et al., 2022).

[12] To convert the coal price data into EUR/MWh, one simply has to divide the series by the conversion factor of 8.14. Since the conversion factor is constant and we consider first differences, this would not change our results.

into EUR, as they are denoted in USD. This is done using the daily USD/EUR exchange rate data from Tullett Prebon. We use temperature data constructed as an average over seven European cities from the European Climate Assessment & Dataset which provides surface air temperature for 199 measurement stations in Europe. It is provided by The Royal Netherlands Meteorological Institute. We aggregate the data to weekly means in order to match our sample frequency. In addition, we remove seasonality by fitting a Fourier regression and subsequently, working with the residual series from this regression (see Section S2 in the supplement).

We plot the allowance price together with the coal and gas price in Fig. 2. Panel (a) plots the EUA price and the coal price. We observe that the two series often move in opposing directions while there is some joint increase visible towards the end of the sample. In Panel (b), the EUA price is jointly plotted with the gas price. The two time series move together over quite some periods of our sample, but they also diverge at the end of 2011 and the beginning of 2014. This gives a first indication of a potentially unstable relation over time.

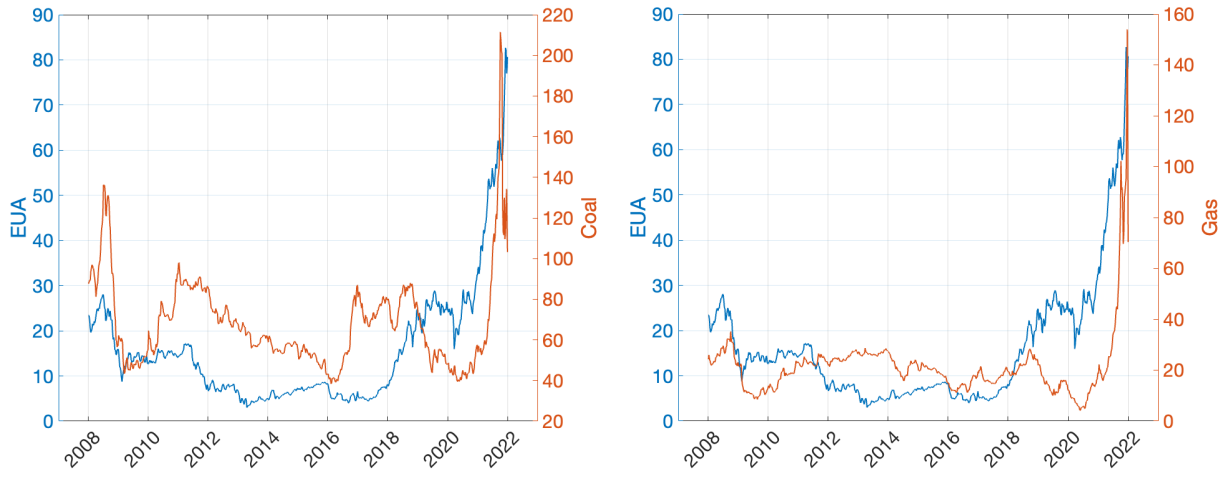
6.3 Related literature

Although market fundamentals should have a major effect on allowance prices, a study of the related literature shows that empirical evidence is mixed. There seems to be a contradiction between theory and empirical applications. This is confirmed by a simple linear regression exercise using on our data. Unit root tests performed on the data give evidence that all series except for the temperature data contain a unit root and we therefore run the regressions using log returns for all but the temperature data. The results of the unit root tests are shown in the supplement. Table 6 displays the linear regression results. The two significant factors are the gas and the oil price in model (a). The coal price, as an important driver, does not show a significant effect on the allowance price in this initial linear regression. The results are robust regarding the choice of indicator of economic activity (model (b)).[13] This does not come as a surprise given the results of previous studies.

The effect of the coal price on allowance prices causes disagreement in findings. Rickels et al. (2014) find a positive effect of the coal price on the allowance price. Aatola et al. (2013) find a negative coefficient of coal, while Hintermann (2010) and Koch et al. (2014) find it to be insignificant. In the latter study, the explicitly calculated fuel switching price is found to have a significant effect. It is obtained from gas and coal prices as well as the efficiency and emission rates of coal and gas plants in the EU ETS. The significant effect could be due to the gas price rather than the coal price since there is no ambiguity in the empirical literature regarding the effect of gas prices. All studies find a positive and significant coefficient of the gas price independent of which approach is used. In particular, in Hintermann (2010) it is the only explanatory variable with a significant effect throughout all considered specifications.

These findings raise the question of whether it might be more appropriate to account for potential time variation when modeling the relationship between allowance prices. This has been discussed by Lutz et al. (2013) who consider potential non-linearities using a regime-switching model. They distinguish two different pricing regimes – one applies during periods of high volatility and the other during periods of low volatility. By construction, the impact of explanatory variables on the allowance price can differ among the two regimes. In both regimes, they find the same set of

[13] The coal price is insignificant by including the oil price and the stock index STOXX 50 together.



(a) EUA and coal

(b) EUA and gas

Figure 2: Joint plot of (a) EUA and coal prices and (b) EUA and gas prices

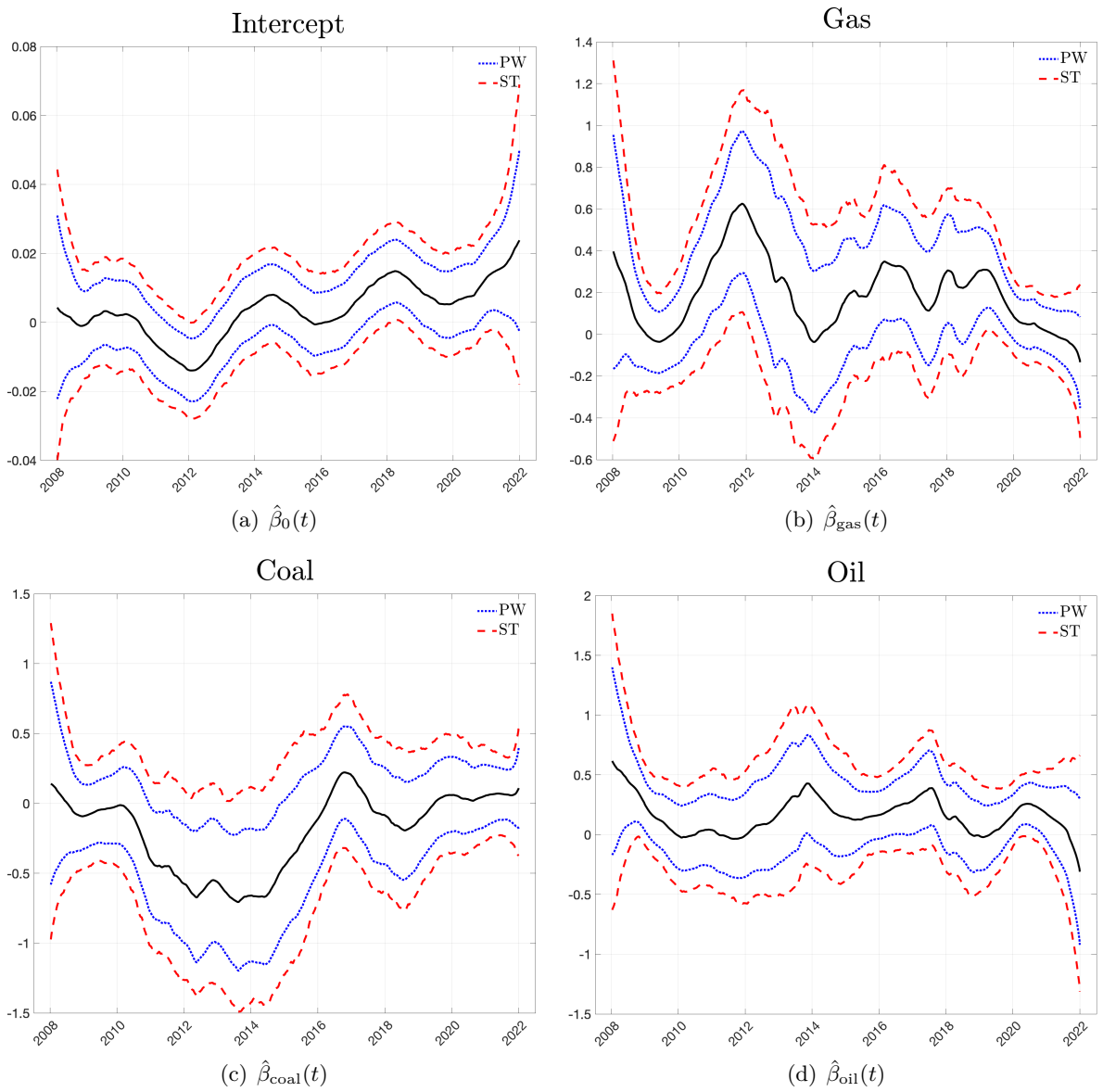
relevant price drivers. Coal and gas prices, oil prices and the stock index are statistically significant determinants of the EUA price. In Regime 2, which is characterized by low and constant volatility, all significant price drivers show the anticipated sign. Regime 1, however, shows a positive impact of the coal price. This goes against economic considerations that predict, as in the second regime, a negative effect of the coal price on allowance prices. These results give further evidence that the relationship between the allowance price and its fundamentals might not be constant over time but subject to (structural) change. This observation is also supported by our parameter stability test. For the implementation, we select the bandwidth by the AVG rule over $[0.06, 0.28]$, resulting in 0.0975 for model (a), and 0.0875 for model (b). A larger upper bound of 0.4 for the search range does not change the results. We find that the null of parameter constancy is rejected in both models (a) and (b) at a significance level of $\alpha = 1\%$. Together with the former studies, a more robust model specification shall be adopted.

6.4 Empirical results

We first check for outliers in allowance price returns as our method is not designed to explain sudden jumps in the dependent variable. We apply the impulse indicator saturation (IIS) approach proposed in Hendry et al. (2008). This approach retains 11 outliers which we detail in the online supplement. The results we present here are the results after outliers have been removed.[14] We employ the Epanechnikov kernel and the SB procedure with $B = 9999$ to construct 95%-level confidence bands. The bands are simultaneous over the full sample $\{i/n,\, i = 1, \ldots, n\}$.

Analogous to the time-constant model (a) (or (b)), we estimate model (2.1) where y_t is the log return on EUAs and $\boldsymbol{x}_t$ contains an intercept as well as four exogenous regressors which consist of log returns obtained from coal, gas and oil (or stock) prices, in addition to temperature data. The same bandwidths as for the parameter constancy tests are used. Namely, based on whether oil or stock prices are included in the model, we use either 0.0975 or 0.0875. Fig. 3 plots the results; additional plots can be found in Section S2 in the supplement. We see significant overall time

[14] We repeat the analysis using the original data, and there are no major differences.

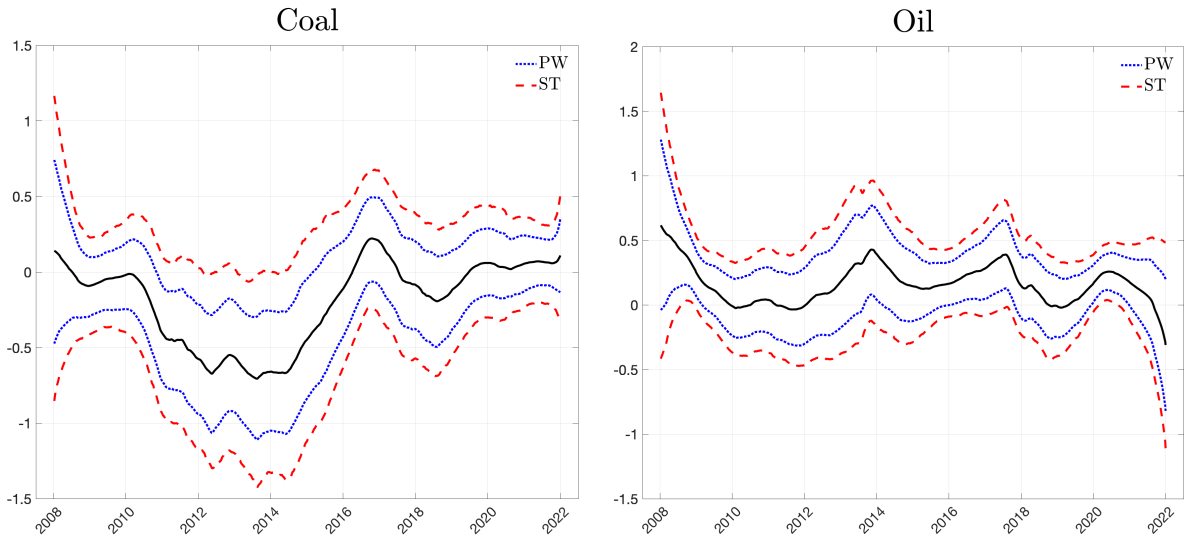


(a) $\hat{\beta}_0(t)$ (b) $\hat{\beta}_{\text{gas}}(t)$

(c) $\hat{\beta}_{\text{coal}}(t)$ (d) $\hat{\beta}_{\text{oil}}(t)$

Figure 3: The estimated coefficient curves and 95%-level pointwise intervals (PW) and simultaneous bands (ST) with the simultaneity over the full sample $\{i/n,\, i = 1\ldots, n\}$

variation. In Fig. 3(a), the nonparametric trend fluctuates around zero for most of the considered time span. However, it turns significantly positive around 2018. This shows that the onset of the drastic price increase, visible in Fig. 2, seems to be picked up by the trend component. In Fig. 3(b), we see that the gas coefficient has the expected positive sign. From the simultaneous bands we get strong evidence for an overall significant effect of this variable because the zero line does not entirely fit within the bands. Specifically, this is the case for two periods, one around 2011/2012 and one in 2019. The statistical evidence for the coal price is mixed in Fig. 3(c). The simultaneous bands show no overall effect given $\alpha = 5\%$. However, at some specific time points around 2014, the pointwise intervals display locally negative effects. Similarly, the path of the oil price in Fig. 3(d) is seemingly positive at the beginning and the end of the sample. It is then interesting to construct simultaneous bands over G and G_{sub} for the coal and oil prices, respectively, to further examine the periods that showed some evidence of a significant effect. The results are plotted in Fig. 4.



(a) simultaneity over G (b) simultaneity over G_{sub}

Figure 4: The estimated coefficient curves for the coal and oil prices and 90%-level pointwise intervals (PW) and simultaneous bands (ST) with the simultaneity over different G sets

The 90%-level simultaneous bands over the sub-periods confirm that the coal price starts out with an insignificant effect that turns significantly negative between mid-2013 and 2014, as observed in Fig. 2(a), and subsequently becomes insignificant again around 2015. In contrast, the effect of oil is positive and significant for the two subsets over which the displayed bands are simultaneous: around 2009 and mid-2017. Zooming in at these two time points, the pointwise intervals show clear evidence of local effects.[15]

The overall time variation found in the relationship does not come as a surprise given the large number of regulatory changes in the market such as the most recent reform. The fact that the relationship is not strengthened despite the higher stringency of the market induced by the reform shows that there might be other factors at play. A potential explanation could be a growing presence of financial actors in the market that are trading allowances for speculation and not for compliance. As a result, it could be beneficial to introduce a monitoring system into the market as mentioned in Pahle and Quemin (2022).

6.5 An illustrative forecasting exercise

With the evidence of time variation, it is natural to ask if the forecasting performance improves using a time-varying (TV) coefficient model compared to a time-constant (TC) one. Since the main scope of this paper is not about improving the forecasting ability of models in the presence of instabilities, here we only consider a simple exercise for illustration. The interested reader is referred to Rossi (2021) for a more extensive discussion.

More specifically, we consider the direct k-step ahead forecasting model: $y_{t+k} = \boldsymbol{\beta}_t' \boldsymbol{x}_t + z_{t+k}$, $k \geq 1$, where $\boldsymbol{x}_t$ includes the same set of exogeneous variables and possibly the dependent variable y_t. The k-step ahead forecast at time t is $\hat{y}_{t+k|t} = \widehat{\boldsymbol{\beta}}(1)' \boldsymbol{x}_t$, where $\widehat{\boldsymbol{\beta}}(1)$ is the local linear estimate at

[15] Although the simultaneous bands for the oil price show a significant effect in early 2020, G_{sub} cannot fully cover this period.

the right endpoint using the observations up to time t. Since $\widehat{\boldsymbol{\beta}}(\cdot)$ is subject to large bias near the boundaries, it is natural to use the average of the last J point estimates for a more robust forecast. That is, $\hat{y}_{t+k|t} = \widehat{\boldsymbol{\beta}}_J(1)'\boldsymbol{x}_t$, where $\widehat{\boldsymbol{\beta}}_J(1) = J^{-1}\sum_{j=1}^{J}\widehat{\boldsymbol{\beta}}\left((n-j+1)/n\right)$. Clearly, $\widehat{\boldsymbol{\beta}}_J(1) = \widehat{\boldsymbol{\beta}}(1)$ when $J = 1$. Moreover, the TC coefficients are also estimated using the data from 1 to t. The forecasting performance is evaluated by the pseudo out-of-sample mean squared forecast error, MSFE, $(n-n_0)^{-1}\sum_{t=n_0-k+1}^{n-k}\left(y_{t+k}-\hat{y}_{t+k|t}\right)^2$. We take the year 2021 as the out-of-sample forecast period, and summarize the results in Table 7. All MSFEs are expressed relative to the benchmark of TC models. Values below 1 indicate a better performance of TV models. We observe from this simple exercise that TV models indeed improve forecasting performance, at least for the one-step ahead forecast. This supports the conclusion in Inoue et al. (2017) and Karmakar et al. (2021). Nevertheless, one should note that the construction of $\widehat{\boldsymbol{\beta}}(\cdot)$ for forecasting and the inclusion of a lagged term of the response in our TV model require more careful theoretical consideration. These issues are left for future research.

7 Conclusion

Recent empirical studies have frequently applied bootstrap methods to construct confidence bands in trending time-varying coefficient models without any theoretical justification. A thorough evaluation of bootstrap methods in finite samples has also been missing. We filled this gap by proposing an autoregressive sieve bootstrap (SB) framework to conduct inference. The SB consistently estimates nuisance parameters both at the interior and at boundary points. Pointwise and simultaneous confidence bands are then simple to obtain. We further suggested a bootstrap-based test for parameter stability. An extensive simulation study has shown that our methods perform well in finite samples. In the empirical application, we studied the price development of allowances in the largest cap-and-trade market for CO_2 emissions – the European Emissions Trading System (EU ETS). We provided first evidence of time variation in the relationship between allowance prices in the EU ETS and their abatement-related fundamental price drivers using our test for parameter stability. It supports previous research that has concluded that the relationship might be unstable. The time variation offers a potential explanation for insignificant coefficients found with linear time-constant regressions used in some of the previous studies.

Acknowledgements

We would like to thank the Editor, Serena Ng, a Guest Editor, and two anonymous reviewers for constructive comments, which have substantially improved the earlier version of the paper. We also thank Eric Beutner, Siem Jan Koopman, Michael Pahle, Franz Palm, Stephan Smeekes, and Bernhard van der Sluis for helpful discussions. We additionally thank all participants at the EC^2 2021, NESG 2022, EcoSta 2022, and at seminars at Tilburg University and the EcoDep seminary series. We dedicate this paper to the memory of Jean-Pierre Urbain without whom we would not be where we are today. All remaining errors are our own.

Table 1: Empirical coverage of confidence bands for $\beta_1(\cdot)$ with $n = 200$ and homoskedastic errors (HM)

		Pointwise		G_{sub}		G		Full Sample	
(ϕ, ψ)	h	SB	MB	SB	MB	SB	MB	SB	MB
(0,0)	0.06	0.947	0.883	0.941	0.409	0.930	0.176	0.861	0.017
	0.09	0.947	0.925	0.927	0.643	0.924	0.448	0.895	0.121
	0.12	0.945	0.933	0.927	0.704	0.931	0.606	0.922	0.180
	0.15	0.939	0.931	0.916	0.560	0.919	0.517	0.911	0.224
	0.18	0.927	0.918	0.898	0.291	0.896	0.278	0.899	0.241
	0.21	0.908	0.894	0.860	0.226	0.854	0.220	0.854	0.221
(0.3,0)	0.06	0.944	0.871	0.944	0.370	0.919	0.147	0.846	0.015
	0.09	0.945	0.920	0.932	0.625	0.923	0.425	0.902	0.117
	0.12	0.943	0.929	0.921	0.679	0.919	0.578	0.910	0.182
	0.15	0.936	0.926	0.918	0.538	0.914	0.499	0.906	0.218
	0.18	0.925	0.914	0.901	0.287	0.901	0.271	0.896	0.225
	0.21	0.909	0.890	0.862	0.220	0.852	0.211	0.853	0.212
(-0.3,0)	0.06	0.948	0.893	0.935	0.417	0.929	0.199	0.858	0.021
	0.09	0.948	0.929	0.937	0.656	0.924	0.471	0.905	0.133
	0.12	0.945	0.934	0.932	0.692	0.928	0.600	0.921	0.192
	0.15	0.938	0.930	0.919	0.541	0.915	0.494	0.910	0.214
	0.18	0.924	0.914	0.894	0.277	0.889	0.263	0.888	0.211
	0.21	0.904	0.884	0.845	0.193	0.839	0.185	0.838	0.183
(0,0.3)	0.06	0.946	0.876	0.933	0.380	0.920	0.152	0.848	0.018
	0.09	0.945	0.921	0.924	0.613	0.917	0.418	0.887	0.121
	0.12	0.944	0.931	0.920	0.676	0.926	0.573	0.913	0.191
	0.15	0.938	0.929	0.913	0.545	0.915	0.501	0.911	0.236
	0.18	0.928	0.915	0.889	0.287	0.891	0.277	0.894	0.234
	0.21	0.913	0.891	0.856	0.215	0.855	0.207	0.855	0.209
(0.3,0.3)	0.06	0.945	0.870	0.928	0.374	0.919	0.142	0.841	0.017
	0.09	0.945	0.916	0.927	0.596	0.915	0.398	0.893	0.110
	0.12	0.944	0.926	0.922	0.650	0.915	0.534	0.908	0.170
	0.15	0.939	0.924	0.917	0.521	0.915	0.467	0.912	0.210
	0.18	0.928	0.912	0.894	0.271	0.888	0.256	0.890	0.216
	0.21	0.913	0.890	0.866	0.200	0.855	0.190	0.859	0.194

Table 2: Empirical coverage of confidence bands for $\beta_2(\cdot)$ with $n = 200$ and homoskedastic errors (HM)

		Pointwise		G_{sub}		G		Full Sample	
(ϕ, ψ)	h	SB	MB	SB	MB	SB	MB	SB	MB
(0,0)	0.06	0.946	0.886	0.925	0.408	0.917	0.175	0.853	0.021
	0.09	0.946	0.927	0.933	0.647	0.931	0.461	0.909	0.126
	0.12	0.946	0.935	0.945	0.705	0.939	0.593	0.930	0.193
	0.15	0.946	0.933	0.938	0.566	0.940	0.520	0.930	0.212
	0.18	0.946	0.929	0.936	0.315	0.937	0.303	0.937	0.250
	0.21	0.946	0.922	0.931	0.263	0.932	0.258	0.932	0.262
(0.3,0)	0.06	0.944	0.876	0.933	0.384	0.917	0.167	0.845	0.020
	0.09	0.946	0.922	0.929	0.628	0.914	0.436	0.889	0.117
	0.12	0.947	0.931	0.929	0.677	0.923	0.576	0.914	0.185
	0.15	0.947	0.931	0.931	0.543	0.932	0.504	0.924	0.220
	0.18	0.944	0.926	0.931	0.312	0.933	0.300	0.933	0.242
	0.21	0.942	0.921	0.923	0.253	0.929	0.250	0.929	0.254
(-0.3,0)	0.06	0.947	0.891	0.934	0.436	0.929	0.207	0.860	0.024
	0.09	0.948	0.929	0.930	0.659	0.921	0.483	0.902	0.148
	0.12	0.949	0.936	0.936	0.711	0.937	0.622	0.930	0.203
	0.15	0.948	0.936	0.935	0.576	0.944	0.534	0.935	0.248
	0.18	0.947	0.932	0.934	0.325	0.938	0.317	0.939	0.260
	0.21	0.946	0.926	0.930	0.274	0.932	0.269	0.933	0.270
(0,0.3)	0.06	0.945	0.882	0.934	0.408	0.924	0.173	0.858	0.020
	0.09	0.946	0.924	0.937	0.643	0.921	0.452	0.889	0.127
	0.12	0.946	0.932	0.939	0.673	0.934	0.578	0.922	0.204
	0.15	0.945	0.932	0.931	0.548	0.929	0.504	0.920	0.236
	0.18	0.944	0.928	0.923	0.302	0.925	0.290	0.927	0.246
	0.21	0.942	0.920	0.927	0.254	0.928	0.251	0.930	0.253
(0.3,0.3)	0.06	0.946	0.875	0.938	0.393	0.924	0.154	0.860	0.016
	0.09	0.947	0.920	0.938	0.629	0.924	0.429	0.902	0.119
	0.12	0.948	0.930	0.938	0.686	0.928	0.581	0.923	0.202
	0.15	0.948	0.931	0.935	0.567	0.937	0.522	0.931	0.226
	0.18	0.946	0.926	0.942	0.306	0.940	0.293	0.942	0.237
	0.21	0.946	0.921	0.936	0.255	0.937	0.250	0.937	0.247

Table 3: Empirical coverage of confidence bands for $\beta_1(\cdot)$ with $n = 200$ and heteroskedastic errors (HT1) - (HT3)

		Pointwise		G_{sub}		G		Full Sample	
Volatility	h	SB	MB	SB	MB	SB	MB	SB	MB
(HT1) smooth increase	0.06	0.944	0.868	0.918	0.381	0.911	0.142	0.824	0.011
	0.09	0.943	0.911	0.918	0.582	0.912	0.374	0.880	0.103
	0.12	0.941	0.923	0.916	0.649	0.913	0.525	0.901	0.168
	0.15	0.938	0.922	0.910	0.524	0.913	0.482	0.903	0.227
	0.18	0.932	0.914	0.897	0.299	0.897	0.288	0.898	0.238
	0.21	0.923	0.899	0.882	0.223	0.881	0.218	0.880	0.219
(HT2) smooth fluctuation	0.06	0.943	0.867	0.933	0.359	0.926	0.134	0.850	0.016
	0.09	0.942	0.913	0.930	0.576	0.923	0.394	0.897	0.101
	0.12	0.942	0.925	0.922	0.646	0.923	0.553	0.911	0.176
	0.15	0.938	0.924	0.920	0.516	0.920	0.473	0.911	0.217
	0.18	0.933	0.915	0.909	0.288	0.906	0.278	0.905	0.240
	0.21	0.924	0.898	0.889	0.231	0.887	0.227	0.888	0.227
(HT3) $(\alpha_1, \alpha_2) = (0.2, 0.7)$	0.06	0.946	0.853	0.901	0.337	0.878	0.105	0.775	0.017
	0.09	0.947	0.902	0.899	0.540	0.888	0.336	0.842	0.095
	0.12	0.947	0.916	0.901	0.597	0.905	0.486	0.881	0.178
	0.15	0.945	0.917	0.902	0.520	0.906	0.469	0.893	0.224
	0.18	0.938	0.914	0.893	0.306	0.893	0.291	0.891	0.245
	0.21	0.928	0.904	0.879	0.257	0.879	0.252	0.882	0.253
(HT3) $(\alpha_1, \alpha_2) = (0.3, 0.6)$	0.06	0.944	0.863	0.876	0.374	0.841	0.110	0.736	0.020
	0.09	0.943	0.902	0.882	0.529	0.862	0.308	0.819	0.111
	0.12	0.943	0.914	0.883	0.577	0.882	0.455	0.860	0.200
	0.15	0.940	0.918	0.886	0.519	0.891	0.467	0.883	0.252
	0.18	0.936	0.916	0.885	0.354	0.887	0.331	0.889	0.289
	0.21	0.930	0.907	0.880	0.288	0.880	0.282	0.878	0.285

Table 4: Empirical coverage of confidence bands when the bandwidth $h \in [0.06, 0.2]$ is selected by AIC, GCV, and LMCV ($l = 0, 2, 4, 6$), as described in Section 4.1, with homoskedastic errors (HM) and $(n, \phi, \psi) = (200, 0.3, 0.3)$

		Pointwise	G_{sub}	G	Full Sample
β_1	AIC	0.922	0.884	0.880	0.877
	GCV	0.924	0.884	0.883	0.872
	LMCV0	0.937	0.916	0.909	0.867
	LMCV2	0.936	0.915	0.905	0.886
	LMCV4	0.935	0.909	0.904	0.892
	LMCV6	0.934	0.906	0.906	0.898
	AVG	0.934	0.913	0.911	0.900
β_2	AIC	0.943	0.935	0.932	0.922
	GCV	0.941	0.929	0.927	0.912
	LMCV0	0.942	0.934	0.924	0.873
	LMCV2	0.944	0.938	0.924	0.894
	LMCV4	0.945	0.943	0.932	0.910
	LMCV6	0.946	0.940	0.935	0.925
	AVG	0.945	0.935	0.928	0.916

Note: The rows labeled "LMCVl", $l = 0, 2, 4, 6$, represent the leave-$(2l+1)$-out estimators used in the LMCV, and "AVG" is the average bandwidth of the other six rules.

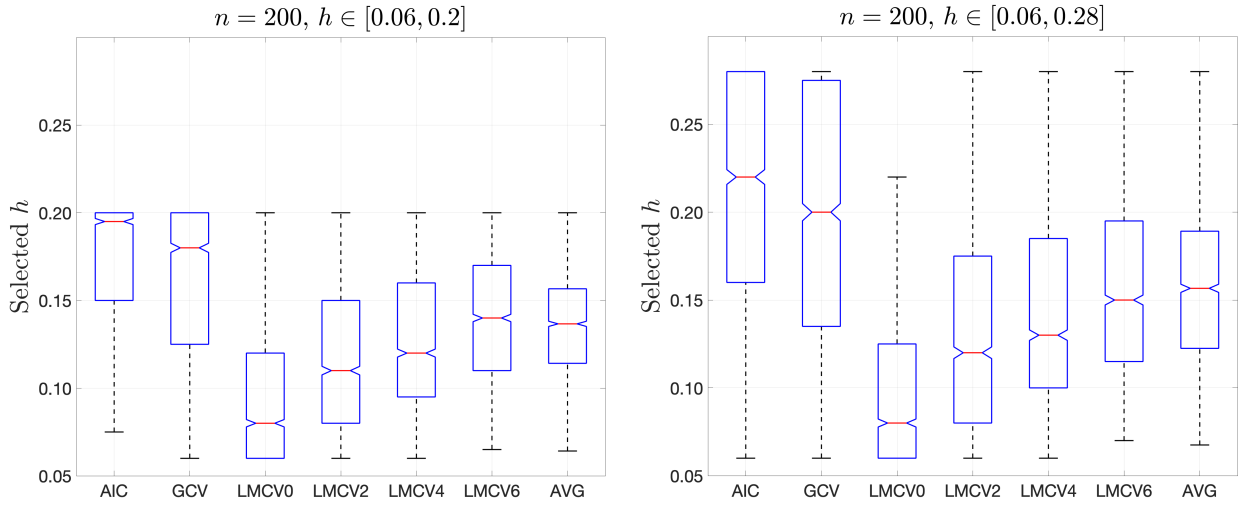


Figure 5: Boxplot of data-driven bandwidths searched over $[0.06, 0.2]$ (left) and $[0.06, 0.28]$ (right) (see Table 4 for further details)

Table 5: Empirical null rejection probability and power of bootstrap tests over the fulll sample for parameter stability with homoskedastic errors (HM) and $n = 200$ (see Table 4 for further details)

	$h \in [0.06, 0.2]$			$h \in [0.06, 0.28]$		
(ϕ, ψ)	(0,0)	(0.3,0)	(0.3,0.3)	(0,0)	(0.3,0)	(0.3,0.3)
Empirical null rejection probability						
AIC	0.027	0.026	0.036	0.023	0.024	0.031
GCV	0.030	0.032	0.041	0.027	0.032	0.039
LMCV0	0.061	0.057	0.064	0.059	0.057	0.062
LMCV2	0.050	0.043	0.049	0.048	0.041	0.048
LMCV4	0.040	0.038	0.039	0.038	0.034	0.038
LMCV6	0.034	0.031	0.036	0.034	0.030	0.032
AVG	0.033	0.032	0.038	0.029	0.030	0.034
Empirical power						
AIC	0.986	0.973	0.959	0.989	0.981	0.968
GCV	0.985	0.973	0.956	0.989	0.980	0.963
LMCV0	0.965	0.942	0.926	0.969	0.946	0.928
LMCV2	0.966	0.949	0.939	0.970	0.955	0.944
LMCV4	0.966	0.955	0.938	0.969	0.960	0.945
LMCV6	0.972	0.959	0.946	0.976	0.966	0.954
AVG	0.976	0.962	0.950	0.985	0.972	0.960

Table 6: The results of linear time-constant models by OLS regressions from allowance price on different explanatory variables: (a) includes the oil price; (b) the stock index STOXX 50. The estimated coefficients ($\hat{\beta}$), HAC-robust standard error estimates se_{HAC} (with Bartlett kernel and the bandwidth selected by Andrews (1991) using AR(1) model, see p.835), and the corresponding p-values are reported. Given $\alpha = 1\%$, both models are rejected for the time-constancy using the bootstrap test (Section 3.4) over the full sample with the bandwidth selected by AVG (see Table 4)

	Model (a)			Model (b)		
	$\hat{\beta}$	se_{HAC}	p-value	$\hat{\beta}$	se_{HAC}	p-value
Intercept	0.004	0.002	0.041	0.004	0.002	0.041
Coal	−0.072	0.047	0.126	−0.038	0.047	0.412
Gas	0.113	0.036	0.002	0.115	0.035	0.001
Oil	0.176	0.049	0.000			
STOXX 50				0.277	0.079	0.000
Temperature	0.000	0.001	0.687	0.000	0.001	0.830
Constancy test	reject***			reject***		

The data-driven bandwidths are 0.0975 and 0.0875 for models (a) and (b), respectively, by the rule of AVG. The search range is $[0.06, 0.28]$. The bandwidth selection here is not sensitive to the upper bound.

Table 7: Pseudo out-of-sample MSFE of time-varying (TV) models. All MSFEs are expressed relative to time-constant models. Values below 1 indicate a better performance of TV models. Model (a) includes the oil price without the stock index STOXX 50, and Model (b) is the opposite. The k-step ahead forecast at time t is given by $\hat{y}_{t+k|t} = \widehat{\boldsymbol{\beta}}_J(1)'\boldsymbol{x}_t$ with $\widehat{\boldsymbol{\beta}}_J(1) = J^{-1}\sum_{j=1}^{J}\widehat{\boldsymbol{\beta}}\left((n-j+1)/n\right)$. Note that $\widehat{\boldsymbol{\beta}}_J(1) = \widehat{\boldsymbol{\beta}}(1)$ when $J = 1$

	Model (a)			Model (b)		
(h, J)	$k=1$	$k=4$	$k=12$	$k=1$	$k=4$	$k=12$
$\boldsymbol{x}_t$: exogeneous variables only						
(AIC, 1)	0.839	1.064	0.889	0.792	0.926	0.867
(AIC, 15)	0.841	1.025	0.888	0.795	0.881	0.867
(AIC, 30)	0.846	0.986	0.888	0.799	0.848	0.867
(LMCV6, 1)	0.835	1.190	0.902	0.808	1.061	0.946
(LMCV6, 15)	0.837	1.124	0.900	0.808	0.945	0.934
(LMCV6, 30)	0.840	1.064	0.897	0.810	0.878	0.922
(AVG, 1)	0.838	1.144	0.910	0.826	1.074	0.930
(AVG, 15)	0.838	1.102	0.905	0.812	0.955	0.913
(AVG, 30)	0.842	1.054	0.901	0.808	0.881	0.901
$\boldsymbol{x}_t$: exogeneous variables and y_t						
(AIC, 1)	0.857	0.945	0.895	0.821	1.053	0.875
(AIC, 15)	0.858	0.932	0.894	0.824	0.983	0.874
(AIC, 30)	0.861	0.920	0.894	0.829	0.923	0.874
(LMCV6, 1)	0.922	1.168	0.909	1.241	1.170	0.881
(LMCV6, 15)	0.907	1.092	0.907	1.094	1.038	0.878
(LMCV6, 30)	0.896	1.030	0.905	0.983	0.944	0.876
(AVG, 1)	0.875	1.035	0.959	0.900	1.100	0.878
(AVG, 15)	0.872	1.001	0.942	0.869	0.996	0.876
(AVG, 30)	0.871	0.971	0.929	0.851	0.920	0.875

References

Aatola, P., Ollikainen, M., and Toppinen, A. (2013). Price determination in the EU ETS market: Theory and econometric analysis with market fundamentals. *Energy Economics*, 36:380–395.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Beran, R. (1988). Balanced simultaneous confidence sets. *Journal of the American Statistical Association*, 83(403):679–686.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. Springer Verlag, New York.

Bühlmann, P. (1995). Moving-average representation of autoregressive approximations. *Stochastic Processes and their Applications*, 60(2):331–342.

Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *The Annals of Statistics*, 26(1):48 – 83.

Cai, B., Cheng, T., and Yan, C. (2018). Time-varying skills (versus luck) in U.S. active mutual funds and hedge funds. *Journal of Empirical Finance*, 49:81–106.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136:163–188.

Castle, J. and Hendry, D. (2019). *Modelling Our Changing World*. Palgrave Texts in Econometrics. Palgrave Macmillan, Open Access.

Chang, Y., Choi, Y., Kim, C. S., Miller, J. I., and Park, J. Y. (2016). Disentangling temporal patterns in elasticities: A functional coefficient panel analysis of electricity demand. *Energy Economics*, 60:232–243.

Chen, B. (2015). Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics*, 185(1):196–215.

Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183.

Christiansen, A. C., Arvanitakis A., and Hasselknippe, H. (2005). Price determinants in the EU emissions trading scheme. *Climate Policy*, 5(1):15–30.

Chu, C.-K. and Marron, J. S. (1991). Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics*, 19(4):1906–1918.

Churchill, S. A., Inekwe, J., Ivanovski, K., and Smyth, R. (2020). The environmental Kuznets curve across Australian states and territories. *Energy Economics*, 90:104869.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Doornik, J. A., Castle, J. L., and Hendry, D. F. (2021). Modeling and forecasting the COVID-19 pandemic time-series data. *Social Science Quarterly*, 102:2070–2087.

Friedrich, M., Smeekes, S., and Urbain, J.-P. (2020). Autoregressive wild bootstrap inference for nonparametric trends. *Journal of Econometrics*, 214:81–109.

Gine, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *The Annals of Probability*, 18(2):851–869.

González, A. and Teräsvirta, T. (2008). Modelling autoregressive processes with a shifting mean. *Studies in Nonlinear Dynamics & Econometrics*, 12(1).

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

Hansen, B. E. (1991). Strong laws for dependent heterogeneous processes. *Econometric Theory*, 7(2):213–221.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, 23:317–335.

Hintermann, B. (2010). Allowance price drivers in the first phase of the EU ETS. *Journal of Environmental Economics and Management*, 59(1):43–56.

Holt, M. and Teräsvirta, T. (2020). Global hemispheric temperatures and co-shifting: A vector shifting-mean autoregressive analysis. *Journal of Econometrics*, 214:198–215.

ICAP (2020). Emissions Trading Worldwide: Status Report 2020. Technical report, Berlin: ICAP.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1):55–67.

Kapetanios, G. (2008). Bootstrap-based tests for deterministic time-varying coefficients in regression models. *Computational Statistics & Data Analysis*, 53(2):534–545.

Karmakar, S., Richter, S., and Wu, W. B. (2021). Simultaneous inference for time-varying models. *Journal of Econometrics*, 227(2):408–428.

Koch, N., Fuss, S., Grosjean, G., and Edenhofer, O. (2014). Causes of the EU ETS price drop: Recession, CDM, renewable policies or a bit of everything?–New evidence. *Energy Policy*, 73:676–685.

Kreiss, J.-P., Paparoditis, E., and Politis, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *The Annals of Statistics*, 39(4):2103 – 2130.

Kristensen, D. (2012). Non-parametric detection and estimation of structural change. *Econonmetrics Journal*, 15:420–461.

Li, D., Chen, J., and Lin, Z. (2011). Statistical inference in partially time-varying coefficient models. *Journal of Statistical Planning and Inference*, 141(2):995–1013.

Li, D., Lu, Z., and Linton, O. (2012). Local linear fitting under near epoch dependence: Uniform consistency with convergence rates. *Econometric Theory*, 28(5):935–958.

Li, D., Phillips, P. C. B., and Gao, J. (2020). Kernel-based inference in time-varying coefficient cointegrating regression. *Journal of Econometrics*, 215(2):607–632.

Li, X. and Zhao, Z. (2019). A time varying approach to the stock return-inflation puzzle. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(5):1509–1528.

Liddle, B., Smyth, R., and Zhang, X. (2020). Time-varying income and price elasticities for energy demand: Evidence from a middle-income panel. *Energy Economics*, 86:104681.

Ling, S., Li, W. K., and McAleer, M. (2003). Estimation and testing for unit root processes with GARCH (1, 1) errors: Theory and Monte Carlo evidence. *Econometric Reviews*, 22(2):179–202.

Lovcha, Y., Perez-Laborda, A., and Sikora, I. (2022). The determinants of CO_2 prices in the EU emission trading system. *Applied Energy*, 305:117903.

Lutz, B. J., Pigorsch, U., and Rotfuß, W. (2013). Nonlinearity in cap-and-trade systems: The EUA price and its fundamentals. *Energy Economics*, 40:222–232.

Müller, U. K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics*, 141(2):1331–1352.

Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, 9(4):307–333.

Pahle, M. and Quemin, S. (2022). Financials threaten to undermine the functioning of emissions markets. *SSRN Working paper*, pages 1–26.

Palm, F. C., Smeekes, S., and Urbain, J.-P. (2008). Bootstrap unit-root tests: Comparison and extensions. *Journal of Time Series Analysis*, 29(2):371–401.

Park, J. Y. (2002). An invariance principle for sieve bootstrap in time series. *Econometric Theory*, 18(2):469–490.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495.

Phillips, P. C. B., Li, D., and Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, 196(1):180–195.

Pretis, F., Schneider, L., Smerdon, J. E., and Hendry, D. F. (2016). Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation. *Journal of Economic Surveys*, 30(3):403–429.

Rickels, W., Görlich, D., and Peterson, S. (2014). Explaining European emission allowance price dynamics: Evidence from phase II. *German Economic Review*, 16(2):181–202.

Romano, J. P. and Wolf, M. (2010). Balanced control of generalized error rates. *The Annals of Statistics*, 38(1):598 – 633.

Rossi, B. (2021). Forecasting in the presence of instabilities: How we know whether models predict well and how to improve them. *Journal of Economic Literature*, 59(4):1135–90.

Rubin, J. D. (1996). A model of intertemporal emission trading, banking, and borrowing. *Journal of Environmental Economics and Management*, 31(3):269–286.

Smeekes, S. and Taylor, A. M. R. (2012). Bootstrap union tests for unit roots in the presence of nonstationary volatility. *Econometric Theory*, 28(2):422–456.

Uddin, M. M., Mishra, V., and Smyth, R. (2020). Income inequality and CO2 emissions in the G7, 1870-2014: Evidence from non-parametric modelling. *Energy Economics*, 88:104780.

Vieu, P. (1991). Nonparametric regression: Optimal local bandwidth choice. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):453–464.

Yousuf, K. and Ng, S. (2021). Boosting high dimensional predictive regressions with time varying parameters. *Journal of Econometrics*, 224(1):60–87. Annals Issue: PI Day.

Zhang, T. and Wu, W. B. (2015). Time-varying nonlinear regression models: Nonparametric estimation and model selection. *The Annals of Statistics*, 43(2):741–768.

Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 72(4):513–531.

Appendix A Technical results

In Section A.1, we first list some auxiliary results that will be needed in the proofs of our main results in Section A.2. The proofs of these auxiliary results can be found in the Supplementary Appendix. Let $w_t^k(\tau) = \left(\frac{\tau_t - \tau}{h}\right)^k K\left(\frac{\tau_t - \tau}{h}\right)$ for convenience.

A.1 Auxiliary lemmas

The first three lemmas do not require $\{z_t\}$ to have an autoregressive structure.

Assumption:

C1 $\{(z_t, \boldsymbol{x}_t), t \geq 1\}$ *is a strictly* α*-mixing stationary process with mixing coefficients* $\alpha(m) = O(m^{-\varphi})$*, where* $\varphi = \max\{(2+\delta)(1+\delta)/\delta, 3(1+\delta)/\delta\}$ *for some* $\delta > 0$*. Further, we assume the moment conditions: (i)* $\mathbb{E}(z_t|\boldsymbol{x}_t) = 0$*; (ii)* $\mathbb{E}(z_t^2|\boldsymbol{x}_t) = \sigma_z^2$*; (iii)* $\mathbb{E}\|\boldsymbol{x}_t\|^{2(2+\delta)} < \infty$*; (iv)* $\mathbb{E}\|z_t\boldsymbol{x}_t\|^{2+\delta} < \infty$.

Lemma 1

For $k = 0, 1, 2, 3$ and $\ell = 1, 2$, consider averages of the form

$$\widehat{\Psi}_{n,k}(\tau) = \frac{1}{nh} \sum_{t=1}^{n} Y_t \left(\frac{\tau_t - \tau}{h} \right)^k K^\ell \left(\frac{\tau_t - \tau}{h} \right), \tag{A.1}$$

where $Y_t \in \mathbb{R}$ is a strictly α-mixing stationary process with the mixing coefficient $\alpha(m) = O(m^{-\beta})$ and $\mathbb{E}|Y_0|^s < \infty$, $\beta = \varphi$ with φ given in Assumption C1 and $s = 2 + \delta$ for some $\delta > 0$. The kernel function $K(\cdot)$ and the bandwidth h satisfy Assumptions A3 and A4, respectively. Then

$$\sup_{\tau \in [0,1]} \left| \widehat{\Psi}_{n,k}(\tau) - \mathbb{E}\left[\widehat{\Psi}_{n,k}(\tau) \right] \right| = O_p\left(\sqrt{\frac{\ln n}{nh}} \right). \tag{A.2}$$

Lemma 2

Under Assumptions C1, A2, A3, and $\max\left\{h, \frac{\ln n}{nh}, \frac{1}{nh^2}\right\} \overset{n\to\infty}{\to} 0$, we have

$$\sup_{\tau \in [0,1]} \left\| \boldsymbol{H} \left(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau) \right) \right\| = O_p\left(h^2 + \sqrt{\frac{\ln n}{nh}} \right). \tag{A.3}$$

Lemma 3

Under Assumption C1, $\sum_{j=-\infty}^{\infty} \left\| \operatorname{cov}\left(\boldsymbol{x}_t z_t, \boldsymbol{x}_{t+j} z_{t+j} \right) \right\| < \infty$ so that $\boldsymbol{\Lambda}$ exists.

Lemma 4

Suppose $\max\left\{h, \frac{\ln n}{nh}, \frac{1}{nh^2}\right\} \overset{n\to\infty}{\to} 0$, and Assumptions A1, A2, A3, B1, B2 hold. It is then possible to represent

$$z_t^* = \sum_{j=0}^{\infty} \widehat{\psi}_{j,n} \, \varepsilon_{t-j}^* \tag{A.4}$$

where $\widehat{\Psi}(z) = \sum_{j=0}^{\infty} \widehat{\psi}_{j,n} z^j = \widehat{\Phi}(z)^{-1}$ (for a large n) with $\widehat{\psi}_{0,n} = 1$ and $\widehat{\Phi}(z) = \sum_{j=0}^{p} \widehat{\phi}_{j,n} z^j$, see Lemmas 6.1 and 6.2 of Bühlmann (1998), or similarly Eq. (31) of Park (2002). We have the following results.

(*i*) $\sup_{n \geq n_1} \sum_{j=0}^{\infty} j \left| \widehat{\psi}_{j,n} \right| < \infty$ *in probability, where n_1 is a random variable.*

(*ii*) $\sup_{j \in \mathbb{N}} \left| \widehat{\psi}_{j,n} - \psi_j \right| = O_p\left(\sqrt{\frac{\ln n}{n}} \right) + O_p\left(p^{-1} \right) + O_p\left(\tilde{h}^2 + \sqrt{\frac{\ln n}{n\tilde{h}}} \right)$.

(*iii*) *For $r \in \{1, 2\}$, $\mathbb{E}^*|\varepsilon_t^*|^{2r} = \mathbb{E}|\varepsilon_t|^{2r} + o_p(1)$.*

(*iv*) *For $j \geq 0$, $\mathbb{E}^*\left(z_t^* z_{t+j}^* \right) = \mathbb{E}(z_t z_{t+j}) + o_p(1)$.*

(*v*) $\sum_{j=0}^{\infty} j \left| \mathbb{E}^*\left(z_t^* z_{t+j}^* \right) \right| = O_p(1)$.

(*vi*) *Let $\boldsymbol{Z}_{n,k}^*(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} \boldsymbol{x}_t z_t^* \left(\frac{\tau_t - \tau}{h} \right)^k K\left(\frac{\tau_t - \tau}{h} \right)$, $k = 0, 1$. Take $k_1, k_2 \in \{0, 1\}$. (i) For any fixed $\tau \in (0, 1)$, $\operatorname{cov}^*\left(\boldsymbol{Z}_{n,k_1}^*(\tau), \boldsymbol{Z}_{n,k_2}^*(\tau) \right) = \nu_{k_1+k_2} \boldsymbol{\Lambda} + o_p(1)$. (ii) For the left endpoint $\tau = ch$, $c \in (0, 1)$, $\operatorname{cov}^*\left(\boldsymbol{Z}_{n,k_1}^*(\tau), \boldsymbol{Z}_{n,k_2}^*(\tau) \right) = \nu_{k_1+k_2,c} \boldsymbol{\Lambda} + o_p(1)$.*

Lemma 5
Recall $w_t^k(\tau) = \left(\frac{\tau_t-\tau}{h}\right)^k K\left(\frac{\tau_t-\tau}{h}\right)$, $k = 0, 1$. *Suppose Assumptions A1, A3, and* $h + (nh)^{-1} \overset{n\to\infty}{\to} 0$ *hold. Suppose* C *is a generic constant that do not depend on* τ_1, τ_2.

(*i*) *For any fixed* $\tau_0 \in (0,1)$, $\tau_1, \tau_2 \in [-1,1]$,

$$\mathbb{E}\left|\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{i,t}z_t\left(w_t^k(\tau_0+\tau_1 h) - w_t^k(\tau_0+\tau_2 h)\right)\right|^2 \le C|\tau_1-\tau_2|^2, \tag{A.5}$$

$$\mathbb{E}^*\left|\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{i,t}z_t^*\left(w_t^k(\tau_0+\tau_1 h) - w_t^k(\tau_0+\tau_2 h)\right)\right|^2 \le O_p(1)|\tau_1-\tau_2|^2. \tag{A.6}$$

(*ii*) *Let* $K \subset (0,1)$ *be a compact set. For* $\tau_1, \tau_2 \in K$,

$$\mathbb{E}\left|\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{i,t}z_t\left(w_t^k(\tau_1 h) - w_t^k(\tau_2 h)\right)\right|^2 \le C|\tau_1-\tau_2|^2, \tag{A.7}$$

$$\mathbb{E}^*\left|\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{i,t}z_t^*\left(w_t^k(\tau_1 h) - w_t^k(\tau_2 h)\right)\right|^2 \le O_p(1)|\tau_1-\tau_2|^2. \tag{A.8}$$

The O_p*-terms in* (A.6) *and* (A.8) *are uniform in* τ_1, τ_2, *and* $i = 0, 1, \ldots, d$.

A.2 Proofs of main results

Lemma 6
Let $\boldsymbol{Z}_n^*(\tau) = \left(\boldsymbol{Z}_{n,0}^*(\tau)', \boldsymbol{Z}_{n,1}^*(\tau)'\right)'$, *where* $\boldsymbol{Z}_{n,k}^*(\tau) = \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} \boldsymbol{x}_t z_t^* \left(\frac{\tau_t-\tau}{h}\right)^k K\left(\frac{\tau_t-\tau}{h}\right)$, $k = 0, 1$, *as defined in Lemma 4. Suppose the assumptions in Lemma 4 hold.*

(*i*) *For any fixed* $\tau \in (0,1)$, *we have*

$$\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(\tau) - \widetilde{\boldsymbol{\theta}}(\tau) - h^2\boldsymbol{b}(\tau)\right) = \boldsymbol{S}^{-1}\boldsymbol{Z}_n^*(\tau) + \boldsymbol{R}_n^*(\tau), \tag{A.9}$$

where $\boldsymbol{S} = \operatorname{diag}(\boldsymbol{\Omega}_0, \mu_2\boldsymbol{\Omega}_0)$, $\mathbb{V}\mathrm{ar}^*\left(\boldsymbol{Z}_n^*(\tau)\right) = \operatorname{diag}(\nu_0\boldsymbol{\Lambda}, \nu_2\boldsymbol{\Lambda}) + o_p(1)$.

(*ii*) *For the left endpoint* $\tau = ch$, $c \in (0,1)$, *we have*

$$\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(ch) - \widetilde{\boldsymbol{\theta}}(ch) - h^2\boldsymbol{b}_c(0+)\right) = \boldsymbol{S}_c^{-1}\boldsymbol{Z}_n^*(ch) + \boldsymbol{R}_n^*(\tau), \tag{A.10}$$

where $\boldsymbol{S}_c = \boldsymbol{\mu}_c \otimes \boldsymbol{\Omega}_0$, $\mathbb{V}\mathrm{ar}^*\left(\boldsymbol{Z}_n^*(ch)\right) = \boldsymbol{\nu}_c \otimes \boldsymbol{\Lambda} + o_p(1)$.

For both cases, the remainder term $\boldsymbol{R}_n^*(\tau)$ *is bounded by*

$$\sup_{\tau\in[0,1]} \|\boldsymbol{R}_n^*(\tau)\| = O_p^*\left(h^2 + \sqrt{\ln n/(nh)} + \sqrt{nh\tilde{h}^4} + \sqrt{h\ln n/\tilde{h}}\right).$$

Proof of Lemma 6 For any $\tau \in [0,1]$, write $\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(\tau) - \widetilde{\boldsymbol{\theta}}(\tau) - h^2\boldsymbol{b}(\tau)\right) = \boldsymbol{M}_n^*(\tau) + \boldsymbol{R}_n^*(\tau)$, where

$$\boldsymbol{M}_n^*(\tau) = \sqrt{nh}\boldsymbol{H}\left[\widehat{\boldsymbol{\theta}}^*(\tau) - \mathbb{E}^*\left(\widehat{\boldsymbol{\theta}}^*(\tau)\right)\right], \quad \widetilde{\boldsymbol{R}}_n^*(\tau) = \sqrt{nh}\boldsymbol{H}\left[\mathbb{E}^*\left(\widehat{\boldsymbol{\theta}}^*(\tau)\right) - \widetilde{\boldsymbol{\theta}}(\tau) - h^2\boldsymbol{b}(\tau)\right].$$

We first look at the properties of $\boldsymbol{H}^{-1}\boldsymbol{S}_n(\tau)\boldsymbol{H}^{-1}$ and $\boldsymbol{Z}_n^*(\tau)$, respectively. Firstly, $\boldsymbol{H}^{-1}\boldsymbol{S}_n(\tau)\boldsymbol{H}^{-1}$ contains the blocks that can be approximated by $\boldsymbol{\Omega}_0\int_{-\tau/h}^{(1-\tau)/h}u^kK(u)du$ according to (S.4). Since $K(\cdot)$ has the support $[-1,1]$, we have $\int_{-\tau/h}^{(1-\tau)/h}u^kK(u)du = \mu_k\mathbb{1}_{\{\tau\in[\tau_L,\tau_U]\}} + \mu_{k,c}\mathbb{1}_{\{\tau=ch\}}$ for any $\tau_L,\tau_U\in(0,1)$, whenever h is sufficiently small. Therefore, by (S.4), we have

$$\boldsymbol{H}^{-1}\boldsymbol{S}_n(\tau)\boldsymbol{H}^{-1} = \boldsymbol{S}\mathbb{1}_{\{\tau\in[\tau_L,\tau_U]\}} + \boldsymbol{S}_c\mathbb{1}_{\{\tau=ch\}} + O_p\left(h^2+\sqrt{\frac{\ln n}{nh}}\right), \tag{A.11}$$

where the O_p-terms are uniform in $\tau\in[0,1]$. Secondly, by Lemma 4(vi), it is immediate to have, $\forall\tau_L,\tau_U\in(0,1)$,

$$\mathbb{V}\mathrm{ar}^*\left(\boldsymbol{Z}_n^*(\tau)\right) = \left(\begin{smallmatrix}\nu_0 & \nu_2\end{smallmatrix}\right)\otimes\boldsymbol{\Lambda}\mathbb{1}_{\{\tau\in[\tau_L,\tau_U]\}} + \left(\begin{smallmatrix}\nu_{0,c} & \nu_{1,c}\\ \nu_{1,c} & \nu_{2,c}\end{smallmatrix}\right)\otimes\boldsymbol{\Lambda}\mathbb{1}_{\{\tau=ch\}} + o_p(1). \tag{A.12}$$

Now we show Part (i). For $\boldsymbol{M}_n^*(\tau)$, because $\boldsymbol{Z}_n^*(\tau) = O_p^*(1)$ uniformly in τ by (A.12),

$$\boldsymbol{M}_n^*(\tau) = \left(\boldsymbol{H}^{-1}\boldsymbol{S}_n(\tau)\boldsymbol{H}^{-1}\right)^{-1}\boldsymbol{Z}_n^*(\tau) = \boldsymbol{S}^{-1}\boldsymbol{Z}_n^*(\tau) + O_p^*\left(h^2+\sqrt{\frac{\ln n}{nh}}\right). \tag{A.13}$$

For $\widetilde{\boldsymbol{R}}_n^*(\tau)$, we note that, if $|\tau_t-\tau|\le h$,

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}(\tau_t) &= \widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\boldsymbol{\beta}}^{(1)}(\tau)(\tau_t-\tau) + \widetilde{\boldsymbol{\beta}}(\tau_t) - \widetilde{\boldsymbol{\beta}}(\tau) - \widetilde{\boldsymbol{\beta}}^{(1)}(\tau)(\tau_t-\tau)\\ &= \widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\boldsymbol{\beta}}^{(1)}(\tau)(\tau_t-\tau) + \boldsymbol{\beta}(\tau_t) - \boldsymbol{\beta}(\tau) - \boldsymbol{\beta}^{(1)}(\tau)(\tau_t-\tau) + O_p\left(\tilde{h}^2+\sqrt{\frac{\ln n}{n\tilde{h}}}\right)\\ &= \widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\boldsymbol{\beta}}^{(1)}(\tau)(\tau_t-\tau) + \left(\boldsymbol{\beta}^{(2)}(\tau)+O(h)\right)\frac{(\tau_t-\tau)^2}{2} + O_p\left(\tilde{h}^2+\sqrt{\frac{\ln n}{n\tilde{h}}}\right),\end{aligned} \tag{A.14}$$

where the final equality is due to (S.1), and the O_p-terms are uniform in $\tau\in[0,1]$ by Lemma 2. Using (S.4), (A.11) and (A.14), we have

$$\begin{aligned}\boldsymbol{H}\left[\mathbb{E}^*\left(\widehat{\boldsymbol{\theta}}^*(\tau)\right) - \widetilde{\boldsymbol{\theta}}(\tau)\right] &= \left(\boldsymbol{H}^{-1}\boldsymbol{S}_n(\tau)\boldsymbol{H}^{-1}\right)^{-1}\left\{\frac{h^2}{2}\begin{pmatrix}h^{-2}\boldsymbol{S}_{n,2}(\tau)\left(\boldsymbol{\beta}^{(2)}(\tau)+O(h)\right)\\ h^{-3}\boldsymbol{S}_{n,3}(\tau)\left(\boldsymbol{\beta}^{(2)}(\tau)+O(h)\right)\end{pmatrix} + O_p\left(\tilde{h}^2+\sqrt{\frac{\ln n}{n\tilde{h}}}\right)\right\}\\ &= \left(\boldsymbol{S}^{-1}+o_p(1)\right)\left\{\frac{h^2}{2}\begin{pmatrix}\mu_2\boldsymbol{\Omega}_0\boldsymbol{\beta}^{(2)}(\tau)\\ \boldsymbol{0}\end{pmatrix} + o_p(h^2) + O_p\left(\tilde{h}^2+\sqrt{\frac{\ln n}{n\tilde{h}}}\right)\right\}\\ &= h^2\boldsymbol{b}(\tau) + o_p(h^2) + O_p\left(\tilde{h}^2+\sqrt{\frac{\ln n}{n\tilde{h}}}\right),\end{aligned}$$

where the O_p-terms are uniform in $\tau\in[0,1]$. Finally, we arrive at

$$\sup_{\tau\in[0,1]}\left\|\widetilde{\boldsymbol{R}}_n^*(\tau)\right\| = o_p\left(\sqrt{nh^5}\right) + O_p\left(\sqrt{nh\tilde{h}^4}+\sqrt{h\ln n/\tilde{h}}\right) = O_p\left(\sqrt{nh\tilde{h}^4}+\sqrt{h\ln n/\tilde{h}}\right). \tag{A.15}$$

Putting $\widetilde{\boldsymbol{R}}_n^*(\tau)$ together with the O_p^*-term in (A.13), denoted by $\boldsymbol{R}_n^*(\tau)$, we have the property of $\boldsymbol{R}_n^*(\tau)$ as given in the lemma. Part (ii) is similar. ■

Proof of Theorem 1 We first consider the bootstrap quantities. By virtue of Lemma 6, we only have to establish the asymptotic normality of $\boldsymbol{Z}_n^*(\tau)$. As such, we use the Cramér-Wold device. For

convenience, we define $Q_t(\tau) = \boldsymbol{a}' \left(\boldsymbol{x}_t', \boldsymbol{x}_t' \left(\frac{\tau_t - \tau}{h}\right)\right)' K\left(\frac{\tau_t - \tau}{h}\right)$, where $\boldsymbol{a} \in \mathbb{R}^{2(d+1)}$ is any unit vector. Define the truncated bootstrap errors $z^*_{t,M_n} = \sum_{j=0}^{M_n} \widehat{\psi}_{j,n}\, \varepsilon^*_{t-j}$ with $M_n \overset{n\to\infty}{\to} \infty$. For $\tau \in [0,1]$,

$$\begin{aligned}
\boldsymbol{a}' \boldsymbol{Z}^*_n(\tau) &= \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} Q_t(\tau) z^*_{t,M_n} + \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} Q_t(\tau) \left(\sum_{j=M_n+1}^{\infty} \widehat{\psi}_{j,n}\, \varepsilon^*_{t-j} \right) \\
&= \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} Q_t(\tau) z^*_{t,M_n} + O^*_p\left(M_n^{-1}\right),
\end{aligned}$$

where the O^*_p-term is uniform in $\tau \in [0,1]$. We obtain it by, for $k \geq 0$, $\left|\sum_{j=M_n+1}^{\infty} \widehat{\psi}_{j,n} \widehat{\psi}_{j+k,n}\right| \leq \left(\sum_{j=M_n+1}^{\infty} \left|\widehat{\psi}_{j,n}\right|\right)^2 = O_p(M_n^{-2})$ and $(nh)^{-1}\mathbb{E}^*\left[\sum_{t=1}^{n} Q_t(\tau)\left(\sum_{j=M_n+1}^{\infty} \widehat{\psi}_{j,n}\, \varepsilon^*_{t-j}\right)\right]^2 = O_p(M_n^{-2})$ by similar arguments in (S.9) and (S.13), together with the Markov's inequality.

It remains to consider $(nh)^{-1/2} \sum_{t=1}^{n} Q_t(\tau) z^*_{t,M_n}$. We now use the common blocking technique to partition $\{1, \ldots, n\} = \bigcup_{j=1}^{k_n} B_j$, where $B_j = \{b_j + 1, \ldots, b_j + \ell_n\} \cup \{b_j + \ell_n + 1, \ldots, b_{j+1}\}$, $k_n = \lceil n/(\ell_n + s_n) \rceil$, $b_j = (j-1)(\ell_n + s_n)$. We truncate the final block B_{k_n} to have n observations in total. It is noted that $k_n \sim n/\ell_n$. We further require $1/\ell_n + \ell_n/(nh) + M_n/\ell_n \to 0$ and $1/s_n + s_n/\ell_n + M_n/s_n$ as $n \to \infty$. For instance, one can take $\ell_n = \lfloor nh^2 \rfloor$, $s_n = \lfloor (nh^2)^{1/2} \rfloor$, and $M_n = \lfloor (nh^2)^{1/4} \rfloor$. Clearly, each block B_j is further separated into two subsets, with either of the relatively large (ℓ_n) and small (s_n) lengths. We have

$$\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} Q_t(\tau) z^*_{t,M_n} = \sum_{j=1}^{k_n} V^*_{n,j}(\tau) + \sum_{j=1}^{k_n} W^*_{n,j}(\tau),$$

where $V^*_{n,j}(\tau) = (nh)^{-1/2} \sum_{t=b_j+1}^{b_j+\ell_n} Q_t(\tau) z^*_{t,M_n}$ and $W^*_{n,j}(\tau) = (nh)^{-1/2} \sum_{t=b_j+\ell_n+1}^{b_{j+1}} Q_t(\tau) z^*_{t,M_n}$

We first show that the small blocks $W^*_{n,j}(\tau)$ are negligible asymptotically. Recall $w^k_t(\tau) = \left(\frac{\tau_t - \tau}{h}\right)^k K\left(\frac{\tau_t - \tau}{h}\right)$. Note that z^*_{t,M_n} are M_n-dependent with respect to the bootstrap probability $\mathbb{P}^*$ conditional on the realization of the original sample. Therefore, $\{W^*_{n,j}(\tau)\}_{j=1}^{k_n}$ are conditionally independent for a sufficiently large n. By the identity $\left(\sum_{t=L}^{U} q_t\right)^2 = \sum_{i=-(U-L)}^{U-L} \sum_{t=L}^{U-|i|} q_t q_{t+|i|}$,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}^*\left(\sum_{j=1}^{k_n} W^*_{n,j}(\tau)\right) &= \frac{1}{nh} \sum_{j=1}^{k_n} \sum_{i=-s_n+1}^{s_n-1} \sum_{t=b_j+\ell_n+1}^{b_{j+1}-|i|} Q_t(\tau) Q_{t+|i|}(\tau) \mathbb{E}^*\left(z^*_{t,M_n} z^*_{t+|i|,M_n}\right) \\
&\leq O_p(1) \frac{1}{nh} \sum_{i=-s_n+1}^{s_n-1} \left|\mathbb{E}^*\left(z^*_{t,M_n} z^*_{t+|i|,M_n}\right)\right| \sum_{j=1}^{k_n} \sum_{t=b_j+\ell_n+1}^{b_{j+1}-|i|} \left\| \begin{pmatrix} w^0_t(\tau) w^0_{t+|i|}(\tau) & w^0_t(\tau) w^1_{t+|i|}(\tau) \\ w^1_t(\tau) w^0_{t+|i|}(\tau) & w^1_t(\tau) w^1_{t+|i|}(\tau) \end{pmatrix} \right\| \\
&\leq O_p(1) \frac{k_n s_n h}{nh} \sum_{i=-s_n+1}^{s_n-1} \left|\mathbb{E}^*\left(z^*_{t,M_n} z^*_{t+|i|,M_n}\right)\right| = O_p\left(\frac{s_n}{\ell_n}\right) = o_p(1),
\end{aligned}$$

where we use that $\sum_{j=0}^{\infty} j \left|\mathbb{E}^*\left(z^*_{t,M_n} z^*_{t+|i|,M_n}\right)\right| = O_p(1)$ by modifying the proof of Lemma 4(v), $\|\boldsymbol{A} \otimes \boldsymbol{B}\| = \|\boldsymbol{A}\| \, \|\boldsymbol{B}\|$ for any matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\sum_{j=1}^{k_n} \sum_{t=b_j+\ell_n+1}^{b_{j+1}-|i|} \left|w^{k_1}_t(\tau) w^{k_2}_{t+|i|}(\tau)\right| \leq C k_n s_n h$, for $|i| \leq s_n$ and $k_1, k_2 \in \{0,1\}$, using the Lipschitz continuity of $x \mapsto x^{k_2} K(x)$ and the Riemann sum approximation (S.6) (replacing n by s_n). Since $\mathbb{E}^*\left(\sum_{j=1}^{k_n} W_{n,j}(\tau)\right) = 0$, we have $\sum_{j=1}^{k_n} W_{n,j}(\tau) = o^*_p(1)$.

Now we establish the asymptotic normality of $\sum_{j=1}^{k_n} V^*_{n,j}(\tau)$ using the Lindeberg central limit

theorem (CLT), see e.g., Theorem 23.6 in Davidson (1994). Again $\mathbb{E}^*\left(\sum_{j=1}^{k_n} V_{n,j}^*(\tau)\right) = 0$, and by the conditional independence of $\{V_{n,j}^*(\tau)\}_{j=1}^{k_n}$, we have $\mathbb{V}\text{ar}^*\left(\sum_{j=1}^{k_n} V_{n,j}^*(\tau)\right) = A_{V,n}^*(\tau) + B_{V,n}^*(\tau)$, where

$$A_{V,n}^*(\tau) = \frac{1}{nh}\sum_{j=1}^{k_n}\sum_{i=-\ell_n+1}^{\ell_n-1}\sum_{t=b_j+1}^{b_j+\ell_n-|i|}\mathbb{E}\left(Q_t(\tau)Q_{t+|i|}(\tau)\right)\mathbb{E}^*\left(z_{t,M_n}^* z_{t+|i|,M_n}^*\right),$$

$$B_{V,n}^*(\tau) = \frac{1}{nh}\sum_{j=1}^{k_n}\sum_{i=-\ell_n+1}^{\ell_n-1}\sum_{t=b_j+1}^{b_j+\ell_n-|i|}\left[Q_t(\tau)Q_{t+|i|}(\tau) - \mathbb{E}\left(Q_t(\tau)Q_{t+|i|}(\tau)\right)\right]\mathbb{E}^*\left(z_{t,M_n}^* z_{t+|i|,M_n}^*\right).$$

Using $\sum_{t=b_j+1}^{b_j+\ell_n-|i|} = \sum_{t=b_j+1}^{b_{j+1}} - \sum_{t=b_j+\ell_n-|i|+1}^{b_{j+1}}$, $\mathbb{E}^*\left(z_{t,M_n}^* z_{t+|i|,M_n}^*\right) = \left(\sum_{k=0}^{M_n}\widehat{\psi}_{k,n}\widehat{\psi}_{k+|i|,n}\right)\mathbb{E}^*|\varepsilon_t^*|^2 = \mathbb{E}(z_t z_{t+|i|}) + o_p(1)$, and by an adaption the proof of (vi) (taking $H_n = s_n$ instead), we have

$$\begin{aligned}
A_{V,n}^*(\tau) &= \boldsymbol{a}'\left\{\frac{1}{nh}\sum_{t=1}^{n}\begin{pmatrix} w_t^0(\tau)w_t^0(\tau) & w_t^0(\tau)w_t^1(\tau) \\ w_t^1(\tau)w_t^0(\tau) & w_t^1(\tau)w_t^1(\tau)\end{pmatrix}\otimes\sum_{i=-\ell_n+1}^{\ell_n-1}\mathbb{E}\left(\boldsymbol{x}_t\boldsymbol{x}_{t+|i|}'\right)\mathbb{E}^*\left(z_{t,M_n}^* z_{t+|i|,M_n}^*\right)\right\}\boldsymbol{a} \\
&\quad + O_p\left(\frac{k_n s_n h}{nh}\right) + o_p(1) + O_p\left(\frac{1}{nh}\right) \\
&\xrightarrow{p} \boldsymbol{a}'\left\{\text{diag}\left(\nu_0,\nu_2\right)\otimes\boldsymbol{\Lambda}\mathbb{1}_{\{\tau\in[\tau_L,\tau_U]\}} + \boldsymbol{\nu}_c\otimes\boldsymbol{\Lambda}\mathbb{1}_{\{\tau=ch\}}\right\}\boldsymbol{a}. \qquad \text{(A.16)}
\end{aligned}$$

Moreover, by defining a deterministic sequence $q_{i,t} = \mathbb{1}_{\{b_j+1\le t\le b_j+\ell_n-|i|,\,1\le j\le k_n\}}$ and then applying Lemma 1, we have $\left\|B_{V,n}^*(\tau)\right\| \le B_{V,n,1}^*(\tau) + O_p\left(\frac{\ell_n}{nh}\sqrt{\frac{\ln n}{nh}}\right)$, where

$$\begin{aligned}
B_{V,n,1}^*(\tau) &= \sum_{i=-\ell_n+1}^{\ell_n-1}\left|\mathbb{E}^*\left(z_{t,M_n}^* z_{t+|i|,M_n}^*\right)\right| \\
&\quad\times\left\|\frac{1}{nh}\sum_{t=1}^{n-|i|} q_{i,t}\begin{pmatrix} w_t^0(\tau)w_t^0(\tau) & w_t^0(\tau)w_t^1(\tau) \\ w_t^1(\tau)w_t^0(\tau) & w_t^1(\tau)w_t^1(\tau)\end{pmatrix}\otimes\left[\boldsymbol{x}_t\boldsymbol{x}_{t+|i|}' - \mathbb{E}\left(\boldsymbol{x}_t\boldsymbol{x}_{t+|i|}'\right)\right]\right\| = O_p\left(\sqrt{\frac{\ln n}{nh}}\right).
\end{aligned}$$

Summing up, we obtain $\text{plim}_{n\to\infty}\mathbb{V}\text{ar}^*\left(\sum_{j=1}^{k_n} V_{n,j}^*(\tau)\right)$ as given in (A.16). The final step is to verify the Lindeberg condition. That is, for every $\kappa > 0$, we shall show

$$\sum_{j=1}^{k_n}\mathbb{E}^*\left[\frac{V_{n,j}^{*2}(\tau)}{\omega_n^{*2}}\mathbb{1}\left\{\left|\frac{V_{n,j}^*(\tau)}{\omega_n^*}\right|>\kappa\right\}\right] = o_p(1), \qquad \text{(A.17)}$$

where $\omega_n^{*2} = \mathbb{V}\text{ar}^*\left(\sum_{j=1}^{k_n} V_{n,j}^*(\tau)\right)$. Note that $Q_t(\tau)z_{t,M_n}^*$ is an L_4-mixingale (provided Lemma 4(i), a straightforward extension of Lemma 4(iii) together with $\mathbb{E}|\varepsilon_t|^{2r} < \infty$, $r > 2$, in Assumption A1) conditionally on the original data. Using Lemma 2 in Hansen (1991) (taking $c_i \le C\left|Q_i(\tau)\right|$) and

the c_r-inequality, the LHS in (A.17) is bounded by

$$\begin{aligned}
\frac{1}{\kappa^2}\frac{1}{\omega_n^{*4}}\sum_{j=1}^{k_n}\mathbb{E}^*\left[V_{n,j}^*(\tau)\right]^4 &\leq C\frac{1}{\kappa^2}\frac{1}{\omega_n^{*4}}\frac{1}{(nh)^2}\sum_{j=1}^{k_n}\left(\sum_{t=b_j+1}^{b_j+\ell_n}Q_t^2(\tau)\right)^2 \\
&\leq C\frac{1}{\kappa^2}\frac{1}{\omega_n^{*4}}\frac{\ell_n}{(nh)^2}\sum_{j=1}^{k_n}\sum_{t=b_j+1}^{b_j+\ell_n}Q_t^4(\tau) \\
&\leq C\frac{1}{\kappa^2}\frac{1}{\omega_n^{*4}}\frac{\ell_n}{(nh)^2}\sum_{t=1}^{n}\left\|\begin{pmatrix} w_t^0(\tau)w_t^0(\tau) & w_t^0(\tau)w_t^1(\tau) \\ w_t^1(\tau)w_t^0(\tau) & w_t^1(\tau)w_t^1(\tau)\end{pmatrix}\right\|^2\|\boldsymbol{x}_t\|^4 \\
&= O_p\left(\frac{\ell_n}{nh}\right),
\end{aligned}$$

where we use that $\omega_n^{*2}=O_p(1)$, and the O_p-term is uniform in $\tau\in[0,1]$. We then obtain (A.17) and thus Equations (3.5) and (3.6) follow given the assumption $\ell_n/(nh)\to 0$.

We just outline the proof for $\widehat{\boldsymbol{\theta}}(\tau)$ because they are similar with the bootstrap counterparts. By (S.2), we have

$$\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau)-h^2\boldsymbol{b}(\tau)\right)=\boldsymbol{S}^{-1}\boldsymbol{Z}_n(\tau)+\boldsymbol{R}_n(\tau),\qquad \forall\tau\in[0,1], \tag{A.18}$$

where $\boldsymbol{Z}_n(\tau)=(\boldsymbol{Z}_{n,0}(\tau)',\boldsymbol{Z}_{n,1}(\tau)')'$ with $\boldsymbol{Z}_{n,k}(\tau)=\frac{1}{\sqrt{nh}}\sum_{t=1}^n\boldsymbol{x}_t z_t\left(\frac{\tau_t-\tau}{h}\right)^k K\left(\frac{\tau_t-\tau}{h}\right)$, $k=0,1$, and $\sup_{\tau\in[0,1]}\|\boldsymbol{R}_n(\tau)\|=O_p\left(\sqrt{nh^7}\right)+O_p\left(h^2\sqrt{\ln n}\right)+o_p(1)$. Replace z_t^* by z_t in the proof above and define similarly the corresponding quantities with respect to z_t similarly, one can find $\boldsymbol{a}'\boldsymbol{Z}_n(\tau)=\frac{1}{\sqrt{nh}}\sum_{t=1}^n Q_t(\tau)z_{t,M_n}+O_p\left(M_n^{-1}\right)=\sum_{j=1}^{k_n}V_{n,j}(\tau)+\sum_{j=1}^{k_n}W_{n,j}(\tau)+O_p\left(M_n^{-1}\right)$, where $\mathbb{V}\mathrm{ar}\left(\sum_{j=1}^{k_n}W_{n,j}(\tau)\right)=O_p\left(\frac{k_n s_n h}{nh}\right)=o_p(1)$ and

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\sum_{j=1}^{k_n}V_{n,j}(\tau)\right) &= \frac{1}{nh}\sum_{j=1}^{k_n}\sum_{i=-\ell_n+1}^{\ell_n-1}\sum_{t=b_j+1}^{b_j+\ell_n-|i|}\mathbb{E}\left(Q_t(\tau)Q_{t+|i|}(\tau)z_{t,M_n}z_{t+|i|,M_n}\right) \\
&= \boldsymbol{a}'\left\{\frac{1}{nh}\sum_{t=1}^{n}\begin{pmatrix} w_t^0(\tau)w_t^0(\tau) & w_t^0(\tau)w_t^1(\tau) \\ w_t^1(\tau)w_t^0(\tau) & w_t^1(\tau)w_t^1(\tau)\end{pmatrix}\otimes\sum_{i=-\ell_n+1}^{\ell_n-1}\mathbb{E}\left(\boldsymbol{x}_t\boldsymbol{x}_{t+|i|}'z_{t,M_n}z_{t+|i|,M_n}\right)\right\}\boldsymbol{a}+o_p(1) \\
&\xrightarrow{p}\boldsymbol{a}'\left\{\mathrm{diag}\left(\nu_0,\nu_2\right)\otimes\boldsymbol{\Lambda}\mathbb{1}_{\{\tau\in[\tau_L,\tau_U]\}}+\boldsymbol{\nu}_c\otimes\boldsymbol{\Lambda}\mathbb{1}_{\{\tau=ch\}}\right\}\boldsymbol{a},
\end{aligned}$$

where we use that $\sum_{j=-\infty}^{\infty}\left\|\mathbb{E}\left(\boldsymbol{x}_t\boldsymbol{x}_{t+|j|}'z_{t,M_n}z_{t+|j|,M_n}\right)\right\|\leq C$. The Lindeberg condition holds similarly by noting that $Q_t(\tau)z_{t,M_n}$ is L_p-mixingale with $p=2+\delta$, where δ is given in Assumption A1, and subsequently applying Lemma 2 in Hansen (1991). ■

Proof of Theorem 2 Consider Part (i) first. By (A.18) and the Lipschitz property of $\boldsymbol{\beta}^{(2)}(\cdot)$ (Assumption A2), we have

$$\begin{aligned}
&\sup_{\tau\in[-1,1]}\left\|\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}(\tau_0+\tau h)-\boldsymbol{\theta}(\tau_0+\tau h)-h^2\boldsymbol{b}(\tau_0)\right)-\boldsymbol{S}^{-1}\boldsymbol{Z}_n(\tau_0+\tau h)\right\| \\
&\qquad\leq\sup_{\tau\in[-1,1]}\|\boldsymbol{R}_n(\tau_0+\tau h)\|+\sup_{\tau\in[-1,1]}\left\|\sqrt{nh^5}\left(\boldsymbol{b}(\tau_0+\tau h)-\boldsymbol{b}(\tau_0)\right)\right\|=o_p(1).
\end{aligned}$$

Similarly, $\sup_{\tau\in[-1,1]}\left\|\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(\tau_0+\tau h)-\widetilde{\boldsymbol{\theta}}(\tau_0+\tau h)-h^2\boldsymbol{b}(\tau_0)\right)-\boldsymbol{S}^{-1}\boldsymbol{Z}_n^*(\tau_0+\tau h)\right\|=o_p^*(1)$ follows from (A.9) in Lemma 6. Therefore, we only have to consider $\boldsymbol{S}^{-1}\boldsymbol{Z}_n(\tau_0+\tau h)$ and $\boldsymbol{S}^{-1}\boldsymbol{Z}_n^*(\tau_0+\tau h)$ in the following. It suffices to show the weak convergence of the $\mathbb{R}^q$-valued processes $\boldsymbol{W}_{\tau_0,n}(\tau):=\boldsymbol{Z}_n(\tau_0+\tau h)$ and $\boldsymbol{W}^*_{\tau_0,n}(\tau):=\boldsymbol{Z}^*_n(\tau_0+\tau h)$ in $C[-1,1]^q$, respectively. The proof involves the same steps as for univariate cases, namely convergence of the finite-dimensional distributions and tightness, see e.g., Theorem 2.1 of Phillips and Durlauf (1986), Chapter 27.7 and Theorem 29.16 of Davidson (1994).

We first establish the asymptotic covariance matrices. For any sequences of vectors $\{\boldsymbol{a}_t\}$ and $\{\boldsymbol{b}_t\}$, we have the identity $\left(\sum_{t=1}^n\boldsymbol{a}_t\right)\left(\sum_{t=1}^n\boldsymbol{b}_t'\right)=\sum_{t=1}^n\boldsymbol{a}_t\boldsymbol{b}_t'+\sum_{i=1}^{n-1}\sum_{t=1}^{n-i}\left(\boldsymbol{a}_t\boldsymbol{b}_{t+i}'+\boldsymbol{a}_{t+i}\boldsymbol{b}_t'\right)$. By this identity and $\mathbb{E}\left(\boldsymbol{W}_{\tau_0,n}(\tau)\right)=0$, $\tau\in[-1,1]$, one can write

$$\operatorname{cov}\left(\boldsymbol{W}_{\tau_0,n}(\tau_1),\boldsymbol{W}_{\tau_0,n}(\tau_2)\right)=:\boldsymbol{\Gamma}_{0,n}(\tau_1,\tau_2;\tau_0)+\boldsymbol{\Gamma}_{1,n}(\tau_1,\tau_2;\tau_0)+\boldsymbol{\Gamma}_{2,n}(\tau_1,\tau_2;\tau_0), \tag{A.19}$$

where $\boldsymbol{\Gamma}_{0,n}(\tau_1,\tau_2;\tau_0)=(nh)^{-1}\sum_{t=1}^n\boldsymbol{Q}_{t,t}(\tau_1,\tau_2;\tau_0)\otimes\mathbb{E}\left(\boldsymbol{x}_t\boldsymbol{x}_t'z_t^2\right)$,

$$\boldsymbol{\Gamma}_{1,n}(\tau_1,\tau_2;\tau_0)=\frac{1}{nh}\sum_{i=1}^{n-1}\sum_{t=1}^{n-i}\boldsymbol{Q}_{t,t+i}(\tau_1,\tau_2;\tau_0)\otimes\mathbb{E}\left(\boldsymbol{x}_t\boldsymbol{x}_{t+i}'z_tz_{t+i}\right),$$

and $\boldsymbol{\Gamma}_{2,n}(\tau_1,\tau_2;\tau_0)=(nh)^{-1}\sum_{i=1}^{n-1}\sum_{t=1}^{n-i}\boldsymbol{Q}_{t+i,t}(\tau_1,\tau_2;\tau_0)\otimes\mathbb{E}\left(\boldsymbol{x}_{t+i}\boldsymbol{x}_t'z_{t+i}z_t\right)$, with

$$\boldsymbol{Q}_{s,t}(\tau_1,\tau_2;\tau_0)=\begin{pmatrix}w_s^0(\tau_0+\tau_1h)w_t^0(\tau_0+\tau_2h) & w_s^0(\tau_0+\tau_1h)w_t^1(\tau_0+\tau_2h)\\ w_s^1(\tau_0+\tau_1h)w_t^0(\tau_0+\tau_2h) & w_s^1(\tau_0+\tau_1h)w_t^1(\tau_0+\tau_2h)\end{pmatrix}. \tag{A.20}$$

Consider $\boldsymbol{\Gamma}_{0,n}(\tau_1,\tau_2;\tau_0)$ first. For $k_1,k_2\in\{0,1\}$, by the Lipschitz property of $x\mapsto(x+\kappa_0)^{k_1}x^{k_2}K(x+\kappa_0)K(x)$ on $[-1,1]$, where κ_0 is a constant, and the Riemann sum approximation (S.6) and a change of variables with $u=(z-\tau_0)/h$,

$$\begin{aligned}\frac{1}{nh}\sum_{t=1}^n w_t^{k_1}(\tau_0+\tau_1h)w_t^{k_2}(\tau_0+\tau_2h)&=\int_{-\tau_0/h}^{(1-\tau_0)/h}(u-\tau_1)^{k_1}(u-\tau_2)^{k_2}K(u-\tau_1)K(u-\tau_2)du+O\left(\frac{1}{nh^2}\right)\\&=\int_{\mathbb{R}}(u-\tau_1)^{k_1}(u-\tau_2)^{k_2}K(u-\tau_1)K(u-\tau_2)du+O\left(\frac{1}{nh^2}\right),\end{aligned} \tag{A.21}$$

when n is sufficiently large. Using (A.21), we obtain $\lim_{n\to\infty}\boldsymbol{\Gamma}_{0,n}(\tau_1,\tau_2;\tau_0)=\boldsymbol{\kappa}(\tau_1,\tau_2)\otimes\mathbb{V}\mathrm{ar}\left(\boldsymbol{x}_tz_t\right)$. For $\boldsymbol{\Gamma}_{i,n}(\tau_1,\tau_2;\tau_0)$, $i=1,2$, we use the splitting technique as in the proof for Lemma 4(vi). That is, we split the summation $\sum_{i=1}^{n-1}=\sum_{i=1}^{H_n-1}+\sum_{i=H_n}^{n-1}$, where $\frac{1}{H_n}+\frac{H_n}{nh}\to0$ as $n\to\infty$, and obtain

$$\begin{aligned}\boldsymbol{\Gamma}_{1,n}(\tau_1,\tau_2;\tau_0)=&\frac{1}{nh}\sum_{t=1}^n\boldsymbol{Q}_{t,t}(\tau_1,\tau_2;\tau_0)\otimes\sum_{i=1}^{H_n-1}\mathbb{E}\left(\boldsymbol{x}_t\boldsymbol{x}_{t+i}'z_tz_{t+i}\right)\\&+O\left(\frac{H_n}{nh}\right)+o(1)+O\left(\frac{1}{nh^2}\right)\to\boldsymbol{\kappa}(\tau_1,\tau_2)\otimes\sum_{i=1}^{\infty}\operatorname{cov}\left(\boldsymbol{x}_tz_t,\boldsymbol{x}_{t+i}z_{t+i}\right).\end{aligned} \tag{A.22}$$

Similarly, $\lim_{n\to\infty}\boldsymbol{\Gamma}_{2,n}(\tau_1,\tau_2;\tau_0)=\boldsymbol{\kappa}(\tau_1,\tau_2)\otimes\sum_{i=-\infty}^{-1}\operatorname{cov}\left(\boldsymbol{x}_tz_t,\boldsymbol{x}_{t+i}z_{t+i}\right)$. Overall,

$$\lim_{n\to\infty}\operatorname{cov}\left(\boldsymbol{Z}_n(\tau_0+\tau_1h),\boldsymbol{Z}_n(\tau_0+\tau_2h)\right)=\boldsymbol{\kappa}(\tau_1,\tau_2)\otimes\boldsymbol{\Lambda}. \tag{A.23}$$

By the Lipschitz property of $x \mapsto x^k K(x)$, $k \in \{0,1\}$, Lemma 1, Lemma 4(v), and Assumption A1(e), one can similarly obtain

$$\begin{aligned}&\operatorname{cov}^*\left(\boldsymbol{W}^*_{\tau_0,n}(\tau_1), \boldsymbol{W}^*_{\tau_0,n}(\tau_2)\right)\\ &\qquad = \boldsymbol{\kappa}(\tau_1,\tau_2) \otimes \sum_{i=-\infty}^{\infty} \mathbb{E}\left(\boldsymbol{x}_t \boldsymbol{x}'_{t+i}\right) \mathbb{E}\left(z_t z_{t+i}\right) + o_p(1) = \boldsymbol{\kappa}(\tau_1,\tau_2) \otimes \boldsymbol{\Lambda} + o_p(1). \quad \text{(A.24)}\end{aligned}$$

For finite $k \in \mathbb{N}$, fix $\tau_1, \ldots, \tau_k \in [-1,1]$. Finite-dimensional distributions of the kq-dimensional vectors $(\boldsymbol{W}_{\tau_0,n}(\tau_1)', \ldots, \boldsymbol{W}_{\tau_0,n}(\tau_k)')'$ weakly converges, and $(\boldsymbol{W}^*_{\tau_0,n}(\tau_1)', \ldots, \boldsymbol{W}^*_{\tau_0,n}(\tau_k)')'$ weakly converges in probability, to the same multivariate normal limiting distribution as in the proof of Theorem 1 using the Cramér-Wold device, (A.23) and (A.24). It remains to consider tightness. It is well-known that probability measures on a product space are tight iff all the marginal probability measures are tight, see e.g., Lemma A.3 in Phillips and Durlauf (1986) and Theorem 26.23 in Davidson (1994). By Lemma 5(i) and the Markov's inequality, the components of $\boldsymbol{W}_{\tau_0,n}(\cdot)$ and $\boldsymbol{W}^*_{\tau_0,n}(\cdot)$ fulfill the moment condition of the tightness criterion in Theorem 12.3 by Billingsley (1968). We have therefore shown the tightness and Part (i) is proved by noting that $\boldsymbol{S}^{-1} = \operatorname{diag}\left(1, \mu_2^{-1}\right) \otimes \boldsymbol{\Omega}_0^{-1}$.

For Part (ii), by (A.10) in Lemma 6 and an adaption of (A.18), we have

$$\begin{aligned}&\sup_{\tau\in K}\left\|\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}^*(\tau h) - \widetilde{\boldsymbol{\theta}}(\tau h) - h^2\boldsymbol{b}_\tau(0+)\right) - \boldsymbol{S}_\tau^{-1}\boldsymbol{Z}^*_n(\tau h)\right\| = o^*_p(1),\\ &\sup_{\tau\in K}\left\|\sqrt{nh}\boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}(\tau h) - \boldsymbol{\theta}(\tau h) - h^2\boldsymbol{b}_\tau(0+)\right) - \boldsymbol{S}_\tau^{-1}\boldsymbol{Z}_n(\tau h)\right\| = o_p(1).\end{aligned}$$

It suffices to consider the weak convergence of $\boldsymbol{S}_\tau^{-1}\boldsymbol{Z}_n(\tau h)$ and $\boldsymbol{S}_\tau^{-1}\boldsymbol{Z}^*_n(\tau h)$ (in probability). We only show their asymptotic covariance matrices here. The convergence of the finite-dimensional distributions and tightness follow similarly from the proof of Part (i) by using Lemma 5(ii) and the Markov's inequality. For $\tau_1, \tau_2 \in K$, one can similarly write $\operatorname{cov}\left(\boldsymbol{S}_{\tau_1}^{-1}\boldsymbol{Z}_n(\tau_1 h), \boldsymbol{S}_{\tau_2}^{-1}\boldsymbol{Z}_n(\tau_2 h)\right) = \boldsymbol{S}_{\tau_1}^{-1}\left(\operatorname{cov}\left(\boldsymbol{Z}_n(\tau_1 h), \boldsymbol{Z}_n(\tau_2 h)\right)\right)\boldsymbol{S}_{\tau_2}^{-1}$ with $\operatorname{cov}\left(\boldsymbol{Z}_n(\tau_1 h), \boldsymbol{Z}_n(\tau_2 h)\right) = \sum_{k=0}^{2}\boldsymbol{\Gamma}_{k,n}(\tau_1,\tau_2;\tau_0 = 0)$, where $\boldsymbol{\Gamma}_{k,n}(\tau_1,\tau_2;\tau_0)$ are defined between (A.19) and (A.20) by plugging $\tau_0 = 0$. Using

$$\frac{1}{nh}\sum_{t=1}^{n} w_t^{k_1}(\tau_1 h) w_t^{k_2}(\tau_2 h) = \int_{\mathbb{R}^+}(u-\tau_1)^{k_1}(u-\tau_2)^{k_2}K(u-\tau_1)K(u-\tau_2)du + O\left(\frac{1}{nh^2}\right), \quad \text{(A.25)}$$

it is not hard to find $\lim_{n\to\infty}\operatorname{cov}\left(\boldsymbol{Z}_n(\tau_1 h), \boldsymbol{Z}_n(\tau_2 h)\right) = \boldsymbol{\kappa}_+(\tau_1,\tau_2) \otimes \boldsymbol{\Lambda}$. Similarly, we have the bootstrap counterpart $\operatorname{cov}^*\left(\boldsymbol{Z}^*_n(\tau_1 h), \boldsymbol{Z}^*_n(\tau_2 h)\right) = \boldsymbol{\kappa}_+(\tau_1,\tau_2) \otimes \boldsymbol{\Lambda} + o_p(1)$. Using $\boldsymbol{S}_\tau^{-1} = \boldsymbol{\mu}_\tau^{-1} \otimes \boldsymbol{\Omega}_0^{-1}$, we obtain Part (ii). ■

Proof of Proposition 1 Under H_0, we have $\sqrt{n}\left(\widehat{\boldsymbol{c}} - \boldsymbol{\beta}(\tau_0 + \tau h)\right) = \sqrt{n}\left(\widehat{\boldsymbol{c}} - \boldsymbol{c}\right) = O_p(1)$ uniformly in $\tau \in [-1,1]$ using a CLT for strictly stationary α-mixing processes, see e.g., Corollary 5.1 in Hall

and Heyde (1980). Then by Theorem 2,

$$\left\{\sqrt{nh}\left(\widehat{\boldsymbol{\beta}}(\tau_0+\tau h)-\widehat{\boldsymbol{c}}-\frac{h^2}{2}\mu_2\boldsymbol{\beta}^{(2)}(\tau_0)\right)\right\}_{\tau\in[-1,1]} \Rightarrow \{\boldsymbol{W}_\beta(\tau)\}_{\tau\in[-1,1]},$$

$$\left\{\sqrt{nh}\left(\widehat{\boldsymbol{\beta}}^*(\tau_0+\tau h)-\widetilde{\boldsymbol{\beta}}(\tau_0+\tau h)-\frac{h^2}{2}\mu_2\boldsymbol{\beta}^{(2)}(\tau_0)\right)\right\}_{\tau\in[-1,1]} \Rightarrow \{\boldsymbol{W}_\beta(\tau)\}_{\tau\in[-1,1]} \text{ in probability,}$$

where $\boldsymbol{W}_\beta(\cdot)$ contains the first $d+1$ components of $\boldsymbol{W}(\cdot)$ corresponding to $\boldsymbol{\beta}(\cdot)$, and $\boldsymbol{W}(\cdot)$ is defined in Theorem 2. Part (i) follows immediately from the continuous mapping theorem for a functional (Theorem A.3, Hall and Heyde (1980)).

Under H_1, there exists a $\tau_0 \in (0,1)$ such that $\boldsymbol{\beta}(\tau_0) \neq \boldsymbol{c}$ for any $\boldsymbol{c} \in \mathbb{R}^{d+1}$. Therefore, for any $\boldsymbol{c} \in \mathbb{R}^{d+1}$ and $\epsilon > 0$, by the continuity of $\boldsymbol{\beta}(\cdot)$, there exists a $\delta = \delta(\epsilon) > 0$ such that $\|\boldsymbol{\beta}(\tau) - \boldsymbol{c}\| \geq \epsilon$, $\forall \tau \in U_\delta(\tau_0)$. As a result, $\sqrt{nh}\,(\boldsymbol{\beta}(\tau) - \widehat{\boldsymbol{c}}) \xrightarrow{p} \infty$ and thus, by Theorem 1,

$$\sqrt{nh}\left(\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{c}}\right) = \sqrt{nh}\,(\boldsymbol{\beta}(\tau) - \widehat{\boldsymbol{c}}) + O_p(1) \xrightarrow{p} \infty, \qquad \forall \tau \in U_\delta(\tau_0).$$

The proof is completed. ■