

Friedrich, Marina; Lin, Yicong

**Working Paper**

## Sieve bootstrap inference for time-varying coefficient models

Tinbergen Institute Discussion Paper, No. TI 2021-107/III

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

*Suggested Citation:* Friedrich, Marina; Lin, Yicong (2021) : Sieve bootstrap inference for time-varying coefficient models, Tinbergen Institute Discussion Paper, No. TI 2021-107/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/248789>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

TI 2021-107/III  
Tinbergen Institute Discussion Paper

# Sieve Bootstrap Inference for Time-Varying Coefficient Models

*Marina Friedrich*<sup>1,2</sup>  
*Yicong Lin*<sup>1</sup>

<sup>1</sup> Vrije Universiteit Amsterdam

<sup>2</sup> Tinbergen Institute

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: [discussionpapers@tinbergen.nl](mailto:discussionpapers@tinbergen.nl)

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900

# SIEVE BOOTSTRAP INFERENCE FOR TIME-VARYING COEFFICIENT MODELS <sup>\*</sup>

Marina Friedrich<sup>a,b</sup> and Yicong Lin<sup>a</sup>

<sup>a</sup>Vrije Universiteit Amsterdam

<sup>b</sup>Tinbergen Institute

## Abstract

We propose a sieve bootstrap framework to conduct pointwise and simultaneous inference for time-varying coefficient regression models based on a nonparametric local linear estimator. The asymptotic validity of the sieve bootstrap in the presence of autocorrelation is established. We find that it automatically produces a consistent estimation of nuisance parameters, both at the interior and boundary points. In addition, we develop a bootstrap test for parameter constancy and show that it is asymptotically correctly sized. An extensive simulation study supports our findings. The proposed methods are applied to assess the price development of CO<sub>2</sub> certificates in the European Emissions Trading System (EU ETS). We find evidence of time variation in the relationship between allowance prices and their fundamental price drivers.

*JEL classifications:* C14, C22, Q48, Q56.

*Keywords:* sieve bootstrap, nonparametric estimation, simultaneous confidence bands, energy economics, emission trading.

---

<sup>\*</sup>We would like to thank Eric Beutner, Stephan Smeekes, Siem Jan Koopman, Michael Pahle and Franz Palm for helpful discussions. We dedicate this paper to the memory of Jean-Pierre Urbain without whom we would not be where we are today. Corresponding author: Marina Friedrich. E-mail: [m.friedrich@vu.nl](mailto: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 behaviors, among many others, is to allow coefficients in linear regression models to evolve deterministically and smoothly over time.<sup>1</sup> There is a considerable amount of papers that develop econometric methodology for these models in various contexts, see e.g. Phillips et al. (2017) and Li et al. (2020) for nonlinear cointegration, Hu et al. (2019) for locally stationary processes, Yousuf and Ng (2021) for high-dimensional predictive regressions. More work in this field includes Cai (2007), Li et al. (2011), Kristensen (2012), Liang and Li (2012), Zhang and Wu (2012, 2015), Chen (2015). Notwithstanding this body of work, little attention has been paid to constructing simultaneous confidence bands around coefficient curves. To 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 the growing literature by proposing a new bootstrap framework for this purpose.

In fact, bootstrap methods have recently been applied in the context of time-varying coefficient models although the theoretical justification has, to the best of our knowledge, 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 a sieve bootstrap procedure for time-varying coefficient models. We show in an extensive simulation study that the proposed method performs well in finite samples. A similar bootstrap approach has been adopted 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 test for parameter constancy. This test can be used as an initial step in modeling.

Compared to the existing asymptotic constructions of 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 the consistent estimation of nuisance parameters such as the asymptotic bias and long-run variance (LRV), the bootstrap correctly reflects these terms without any extra estimation steps. 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 pointwise and simultaneous confidence bands yield accurate coverage with small sample sizes, as illustrated in our simulation studies. In contrast, a relatively large sample size is usually required to achieve the theoretical coverage due to the slow convergence rate for

---

<sup>1</sup>Alternative methods which 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.

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<sub>2</sub> 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 which can potentially explain why the coal price is often found to be insignificant in linear regressions with constant parameters.

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 practical implementation. Section 5 presents the simulation study and Section 6 the empirical application. Section 7 concludes.

Finally, a word on notation. For a vector  $\mathbf{x} = (x_j) \in \mathbb{R}^n$ , its  $p$ -norm is denoted by  $\|\mathbf{x}\|_p = (\sum_{j=1}^n |x_j|^p)^{1/p}$ . And  $\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq \mathbf{0}} \|\mathbf{A}\mathbf{x}\|_p / \|\mathbf{x}\|_p$  stands for the induced  $p$ -norm for a matrix  $\mathbf{A}$ . We will omit the subscripts whenever  $p = 2$ . The Kronecker product is denoted by “ $\otimes$ ”. For vectors  $\mathbf{a} = (a_j)$  and  $\mathbf{b} = (b_j)$ ,  $\mathbf{a} \geq \mathbf{b}$  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 it is conditional on the original sample. For instance, “ $\xrightarrow{d^*}_p$ ” 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:

$$y_t = \boldsymbol{\beta}'_t \mathbf{x}_t + z_t = \beta_{0,t} + \sum_{j=1}^d \beta_{j,t} x_{j,t} + z_t, \quad t = 1 \dots, n, \quad (2.1)$$

where  $\boldsymbol{\beta}_t = (\beta_{0,t}, \beta_{1,t}, \dots, \beta_{d,t})'$  is a  $(d+1) \times 1$  vector of time-varying coefficients, and  $\mathbf{x}_t = (x_{0,t}, x_{1,t}, \dots, x_{d,t})'$  is a  $(d+1)$ -dimensional vector of covariates. We observe the data  $\{(y_t, \mathbf{x}_t)\}_{t=1}^n$ . We shall assume that  $\{z_t, \mathbf{x}_t\}$  is a stationary process (Assumption A1) and  $\boldsymbol{\beta}_t := \boldsymbol{\beta}(t/n)$  with  $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \dots, \beta_d(\cdot))' : [0, 1] \rightarrow \mathbb{R}^{d+1}$  being a vector of functions (Assumption A2). The formulation (2.1) is firstly proposed by Robinson (1989, 1991), Orbe et al. (2005, 2006), and has been considered by many authors as mentioned. 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 on 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 the estimations of nuisance parameters which are generally 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 confi-

dence band has slow logarithmic convergence. Therefore, large sample size is needed to achieve the desired coverage. To bypass these difficulties, at least partially, we will propose a sieve bootstrap procedure in Section 3.1.

## 2.1 The local linear estimator and assumptions

In general, a first-step estimation of  $\beta_t$  is needed for (nonparametric) bootstrap. It is well-known that local linear estimators suffer less from boundary effects and have a smaller bias than Nadaraya-Watson-type estimators (Fan and Gijbels (1996), Cai (2007)). As such, we will consider the former approach here. For  $\tau \in (0, 1)$ , write  $\beta_j(t/n) \approx \beta_j(\tau) + \beta_j^{(1)}(\tau)(t/n - \tau)$ ,  $j = 0, \dots, d$ . The estimator is based on the following approximation:

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

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

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

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:

$$\hat{\boldsymbol{\theta}}(\tau) = \begin{pmatrix} \mathbf{S}_{n,0}(\tau) & \mathbf{S}'_{n,1}(\tau) \\ \mathbf{S}_{n,1}(\tau) & \mathbf{S}_{n,2}(\tau) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{T}_{n,0}(\tau) \\ \mathbf{T}_{n,1}(\tau) \end{pmatrix} =: \mathbf{S}_n^{-1}(\tau) \mathbf{T}_n(\tau), \quad \tau \in (0, 1), \quad (2.4)$$

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

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

Our asymptotic analysis hinges on the following regularity conditions.

### Assumptions:

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

- (a)  $\{\mathbf{x}_t\}_{t \in \mathbb{Z}}$  is a strictly stationary  $\alpha$ -mixing process with the mixing coefficient  $\alpha(m) = O(m^{-\varphi})$ , where  $\varphi = \max\{(2 + \delta)(1 + \delta)/\delta, 3(1 + \delta)/\delta\}$ . Moreover,  $\mathbb{E}\|\mathbf{x}_t\|^{2(2+\delta)} < \infty$ .
- (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  $f_\varepsilon$  of  $\varepsilon_t$  satisfies  $\int_{x \in \mathbb{R}} |f_\varepsilon(x + a) - f_\varepsilon(x)| dx \leq M|a|$ ,  $M < \infty$ , whenever  $|a| \leq \tau$  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  $|z| \leq 1$ . Moreover, for some  $\lambda > (5\varphi + 7)/4$ ,  $|\psi_j| \ll C_1 j^{-\lambda}$  as  $j \rightarrow \infty$ , where  $\varphi$  is defined in A1(a).

(e) The following moment conditions hold almost surely: (i)  $\mathbb{E}(z_t | \mathbf{x}_t) = 0$ ; (ii)  $\mathbb{E}(z_t z_s | \mathbf{x}_t \mathbf{x}'_s) = \mathbb{E}(z_t z_s)$ ,  $s, t \in \mathbb{Z}$ .

A2 The coefficient function  $\beta(\cdot) \in \mathcal{C}^3[0, 1]$ , namely  $\beta_j(\cdot) \in \mathcal{C}^3[0, 1]$ ,  $j = 0, 1, \dots, d$ .

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  $\max \left\{ h, \frac{\ln n}{nh}, \frac{1}{nh^2}, nh^7, h^4 \ln n \right\} \xrightarrow{n \rightarrow \infty} 0$ .

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). Assumption A1(b) imposes a linear structure on the error process, which is usually required for the (AR) sieve bootstrap procedure (e.g. Palm et al. (2008), Smeekes and Taylor (2012)). It is well known that not all linear processes are strong mixing, see e.g. Bosq (1998, Chapter 1.1). To apply the results of uniform convergence as 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  $\int_{x \in \mathbb{R}} |f_\varepsilon(x+a) - f_\varepsilon(x)| dx = \sqrt{\frac{2}{\pi \sigma_\varepsilon^2}} |a|$ . Similarly, 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( $\infty$ ) 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  $0 \leq p, q < \infty$  satisfies this condition. Moreover, by Lemma 2.1 of Bühlmann (1995),  $\{z_t\}$  admits an AR( $\infty$ ) 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  $\alpha$ -mixing process with  $\alpha(m) \ll C m^{-\varphi}$ , where  $\varphi$  is given in Assumption (a). Therefore, the uniform convergence results can be applied to the process  $\{(z_t, \mathbf{x}_t)\}$ . Given  $z_t$  satisfies Assumption A1(b), and using the uniform results in Li et al. (2012), it is possible to only require  $\{(z_t, \mathbf{x}_t)\}$  to be a strictly stationary  $L_p$ -NED process with  $p > 4$  with some high-level assumptions. We do not go in this direction because these assumptions are generally difficult to verify in practice.

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  $\mathbb{E}(z_t^2 | \mathbf{x}_t) = g(\mathbf{x}_t, t/n)$ , where  $g(\cdot, \cdot)$  is some continuous and bounded function of the stationary regressors and trends. Although  $\hat{\theta}(\tau)$  retains consistency and asymptotic normality with such heteroscedasticity, the sieve bootstrap scheme will 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.  $\mathbb{E}(z_t^2|\mathbf{x}_t) = g(t/n)$ , the sieve wild bootstrap (SWB) may be a solution, see Remark 4. Second, the assumption requires some “independence” in terms of second-order moments between  $\mathbf{x}_t$  and  $z_t$ . A similar condition has been imposed in Kapetanios (2008, Assumption 3) for *i.i.d.* bootstrap. We will come back to this point in Remark 2.

Assumption A2 is a common smoothness condition, see e.g. Assumption 6 in Zhou and Wu (2010). It can be weakened to assume that  $\beta(\cdot)$  is twice continuously differentiable with minor modification of the proofs. The kernel function  $K(\cdot)$  in Assumption A3 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 appeared in  $\frac{\ln n}{nh} \xrightarrow{n \rightarrow \infty} 0$  is due to our uniform convergence results. It is not stringent because common bandwidth choices have the rate  $n^{-\kappa}$  with  $\kappa \in (0, 1)$ , satisfying this condition.

### 3 Confidence bands with the sieve bootstrap

We now propose a sieve bootstrap procedure 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  $\text{AR}(\infty)$  process in (2.6). The focus of the sieve bootstrap lies on the dependence structure of the error terms. It attempts to approximate the process by  $\text{AR}(p)$  models. This means that we first estimate model (2.1). To the residuals from this estimation we fit an  $\text{AR}(p)$  model and, again, 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

$$\hat{z}_t = y_t - \mathbf{x}'_t \tilde{\beta}(t/n), \quad t = 1, \dots, n,$$

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

Step 2 To the residuals  $\hat{z}_t$ , for  $t = 1, \dots, n$ , fit an autoregressive model of order  $p$  and form the new series of residuals

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

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

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

Step 4 Calculate the bootstrap errors  $z_t^*$  as  $z_t^* = \sum_{j=1}^p \hat{\phi}_j z_{t-j}^* + \varepsilon_t^*$  and generate the bootstrap observations by

$$y_t^* = \mathbf{x}'_t \tilde{\beta}(t/n) + z_t^*, \quad t = 1, \dots, n,$$

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

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

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

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

denote, for  $j = 0, \dots, d$ , the  $\alpha$ -quantile of the  $B$  centered bootstrap statistics  $\hat{\beta}_j^*(\tau) - \tilde{\beta}_j(\tau)$ . These bootstrap quantiles are then used to construct confidence bands as described below.

When generating the bootstrap error process in Step 4, we start the recursion with an arbitrary starting value and wait until stationarity is reached. The lag length  $p$  in Step 2 should satisfy Assumption B2 below. How to select it in practice is discussed in Section 4. Note that in Step 1 of the above algorithm, a different bandwidth is used to perform the nonparametric estimation. Compared to the original bandwidth  $h$ , this bandwidth should be larger to produce an oversmoothed estimate as starting point for the bootstrap procedure. The reason for this is the presence of the asymptotic bias whenever local polynomial estimation is applied. The bias contains the second derivatives of the coefficient functions, which can only be consistently estimated using a larger, oversmoothed bandwidth  $\tilde{h}$  (Assumption B1). 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 3, we provide an intuition of why oversmoothing can consistently estimate the asymptotic bias. Now we list the extra assumptions needed for the bootstrap.

**Assumptions:**

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

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

A few more words on the assumptions. We will also 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  $h \ln n / \tilde{h} \xrightarrow{n \rightarrow \infty} 0$ , compared to Assumption (K) in Bühlmann (1998) and Assumption 8 in Friedrich et al. (2020).

*Remark 1.* Residuals which lie close to the boundary might be problematic as the nonparametric estimator exhibits edge effects. The quality of estimates for points close to the boundary ( $\tau$  close to 0 or 1) 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 only obtained for  $t = [n\delta] + 1, \dots, [n(1 - \delta)]$ . 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 since the boundary effects are reduced. We investigate the impact of the edge effects in the simulation study.

### 3.2 Constructing confidence intervals

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

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

where  $1 - \alpha$  is the confidence level and

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

The quantiles  $\hat{q}_{j,\alpha}(\tau)$  are obtained in the last step of one of our bootstrap procedure. From Equation (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

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

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

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

As the bootstrap counterparts, 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  $G = \bigcup_{i=1}^m U_i(h)$ , where the neighborhoods  $U_i(h)$  are of the form  $U_i(h) = [\tau_i - ah, \tau_i + bh]$ , with  $0 \leq a, b < \infty$  and  $m < \infty$ . 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  $\tau \in G$ ,  $j = 0, \dots, d$ .

Step 2 Choose  $\alpha_s$  as

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

Step 3 Given  $\alpha_s$  from Step 2, construct the simultaneous confidence bands as

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

Note that a pointwise error  $\alpha_s$  is found for which a fraction of approximately  $(1 - \alpha)$  of all

centered bootstrap estimates falls within the resulting confidence intervals, for all points of the set  $G$ . As such, the confidence intervals with pointwise coverage  $(1 - \alpha_s)$  become simultaneous confidence bands with coverage  $(1 - \alpha)$ . Since this model can have more than one explanatory variable, we have to construct a confidence band for  $d$  coefficient curves and the trend function using the above procedure. It is noted that the confidence bands do not have to have equal width, a feature which we find valuable, since in finite samples the fluctuations are likely to vary over the time span.

### 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  $\boldsymbol{\mu}_c = \begin{pmatrix} \mu_{0,c} & \mu_{1,c} \\ \mu_{1,c} & \mu_{2,c} \end{pmatrix}$  and  $\boldsymbol{\nu}_c = \begin{pmatrix} \nu_{0,c} & \nu_{1,c} \\ \nu_{1,c} & \nu_{2,c} \end{pmatrix}$ . Moreover, we define  $\boldsymbol{\kappa}(\tau_1, \tau_2) = \int_{\mathbb{R}} \mathbf{w}(u; \tau_1, \tau_2) K(u - \tau_1) K(u - \tau_2) du$  and  $\boldsymbol{\kappa}_+(\tau_1, \tau_2) = \int_{\mathbb{R}^+} \mathbf{w}(u; \tau_1, \tau_2) K(u - \tau_1) K(u - \tau_2) du$ , where  $\mathbf{w}(u; \tau_1, \tau_2) = \begin{pmatrix} 1 & u - \tau_2 \\ u - \tau_1 & (u - \tau_1)(u - \tau_2) \end{pmatrix}$ .
- (b) Bias terms:  $\mathbf{b}(\tau) = \frac{1}{2} \begin{pmatrix} \mu_2 \boldsymbol{\beta}^{(2)}(\tau) \\ \mathbf{0} \end{pmatrix}$ ,  $\mathbf{b}_c(0+) = \frac{1}{2} \boldsymbol{\mu}_c^{-1} \begin{pmatrix} \mu_{2,c} \boldsymbol{\beta}^{(2)}(0+) \\ \mu_{3,c} \boldsymbol{\beta}^{(2)}(0+) \end{pmatrix}$ , where  $\boldsymbol{\beta}^{(2)}(0+) = \lim_{\tau \downarrow 0} \boldsymbol{\beta}^{(2)}(\tau)$ .
- (c) Scaling matrix:  $\mathbf{H} = \text{diag}(\mathbf{I}_{d+1}, h\mathbf{I}_{d+1})$ .
- (d) Short/long-run covariance matrices:  $\boldsymbol{\Omega}_0 = \mathbb{E}(\mathbf{x}_t \mathbf{x}_t')$ ,  $\boldsymbol{\Lambda} = \sum_{j=-\infty}^{\infty} \text{cov}(\mathbf{x}_t z_t, \mathbf{x}_{t+j} z_{t+j})$ .

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  $\tau \in (0, 1)$ ,

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

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

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

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

and  $\sqrt{nh} \mathbf{H} \left( \hat{\boldsymbol{\theta}}(ch) - \boldsymbol{\theta}(ch) - h^2 \mathbf{b}_c(0+) \right)$  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 behaviors 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 thus are

omitted here. Third, the bootstrap consistently estimates the bias terms  $\mathbf{b}(\tau)/\mathbf{b}_c(0+)$ . 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 they are negligible or dominate 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 a fixed time point. When the distance between two points,  $\tau_1$  and  $\tau_2$ , is of order  $h$ , there is non-negligible correlation between  $\hat{\boldsymbol{\theta}}(\tau_1)$  and  $\hat{\boldsymbol{\theta}}(\tau_2)$ . Theorem 2 below shows that the sieve bootstrap correctly mimic this neighborhood correlation. This property can be used for constructing simultaneous confidence intervals over  $G$ , that is, a finite union of  $h$ -neighborhoods.

Some additional notation is needed. Throughout, 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  $\mathbf{w} = (w_1, \dots, w_q) : [-1, 1] \rightarrow \mathbb{R}$ . 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 notation “ $\Rightarrow$ ” signifies weak convergence in  $C[-1, 1]^q$  endowed with  $d_U^q(\mathbf{x}, \mathbf{y}) = \max_{1 \leq i \leq q} \{d_U(x_i, y_i)\}$ , where  $\mathbf{x} = (x_i)$ ,  $\mathbf{y} = (y_i) \in C[-1, 1]^q$ . Note that  $d_U^q(\mathbf{x}, \mathbf{y})$  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  $\tau_0 \in (0, 1)$ ,

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

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

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

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

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

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

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

Theorem 2 shows the uniform validity of the bootstrap within an  $h$ -neighborhood for any interior points 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 of the bootstrap over a finite union of  $h$ -neighborhoods follow straightforwardly from the theorem. In finite samples, one can always cover the full sample by taking sufficiently many unions in  $G$ . Nevertheless, this is different from the asymptotic construction which is based on direct approximations to  $\sup_{\tau \in (0,1)} (\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau))$ , see e.g. Zhou and Wu (2010) and Karmakar et al. (2021). Although it is theoretically less attractive, our construction may be practically more relevant. 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 conclude about, for instance, what are the most important drivers of  $y_t$  in these specific periods. Several remarks are now in place.

*Remark 2.* Recall  $\tau_t = t/n$  and  $\mathbb{E}(z_t z_s | \mathbf{x}_t \mathbf{x}'_s) = \mathbb{E}(z_t z_s)$ ,  $s, t \in \mathbb{Z}$ , in Assumption A1(e). The reason for this assumption is given as follows. Our sieve bootstrap (SB) scheme hinges on an *i.i.d.* bootstrap of the AR residuals which are considered to be approximations of the *i.i.d.* errors  $\varepsilon_t$ . Intuitively speaking, the *i.i.d.* resampling of these residuals destroys the potential higher-order dependence structure between  $\mathbf{x}_t$  and  $z_t$ . We find that the next terms mimic the asymptotic normality:

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

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

$$\begin{aligned} \text{Var}^*\left(\mathbf{Z}_{n,k}^*(\tau)\right) &= \frac{1}{nh} \sum_{i=-n+1}^{n-1} \sum_{t=1}^{n-|i|} \mathbf{x}_t \mathbf{x}'_{t+|i|} \mathbb{E}^*\left(z_t^* z_{t+|i|}^*\right) w_t^k(\tau) w_{t+|i|}^k(\tau). \quad (3.9) \\ &\approx \sum_{i=-n+1}^{n-1} \mathbb{E}\left(z_1 z_{1+|i|}\right) \left[ \frac{1}{nh} \sum_{t=1}^{n-|i|} \mathbf{x}_t \mathbf{x}'_{t+|i|} \left(\frac{\tau_t - \tau}{h}\right)^{2k} K^2\left(\frac{\tau_t - \tau}{h}\right) \right] \\ &\approx \sum_{i=-n+1}^{n-1} \mathbb{E}\left(z_1 z_{1+|i|}\right) \left[ \mathbb{E}\left(\mathbf{x}_1 \mathbf{x}'_{1+|i|}\right) \frac{1}{nh} \sum_{t=1}^{n-|i|} \left(\frac{\tau_t - \tau}{h}\right)^{2k} K^2\left(\frac{\tau_t - \tau}{h}\right) \right] \\ &\approx \nu_{2k} \sum_{i=-n+1}^{n-1} \mathbb{E}\left(z_1 z_{1+|i|}\right) \mathbb{E}\left(\mathbf{x}_1 \mathbf{x}'_{1+|i|}\right) \end{aligned}$$

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

*Remark 3.* Some intuition for the oversmoothing condition  $h \ln n / \tilde{h} \xrightarrow{n \rightarrow \infty} 0$  is provided as follows. For  $\tau \in [0, 1]$ , we can write

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

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

*Remark 4.* The SB method, while correctly mimicking the dependence structure, is not valid when the errors are heteroskedastic, for instance,  $z_t = \sigma_t u_t$  with  $u_t$  satisfying Assumption A1 and  $\sigma_t = \sigma(t/n)$  is a deterministic function. In Step 3 of the SB algorithm, we draw randomly with replacement from the  $\text{AR}(p)$  residuals. It leads to bootstrap errors that do not display the same variance pattern as the original residuals in the presence of heteroskedasticity. To overcome this, one may add a wild component to the SB. That is, generate  $\varepsilon_t^*$  as  $\varepsilon_t^* = \xi_t^* \widehat{\varepsilon}_{t,p}$  with  $\xi_t^* \stackrel{iid}{\sim} \mathcal{N}(0, 1)$  instead of the original Step 3. The other steps remain. This bootstrap is called sieve wild bootstrap (SWB). It is suggested in Cavaliere and Taylor (2008, 2009) and 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). The SWB is a natural solution when there is unconditional heteroskedasticity. Interestingly, as also reported in Friedrich et al. (2020), we find in our simulations that SWB has poor coverage and is even outperformed by SB in the presence of unconditional/conditional heteroskedasticity. The issues merit further research.

### 3.4 Testing for parameter stability

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

$$H_0 : \beta(\cdot) = \mathbf{c}, \quad \text{for some unknown } \mathbf{c} \in \mathbb{R}^{d+1}, \quad (3.10)$$

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

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

Under the null, the parametric estimator  $\widehat{\mathbf{c}}$  has faster convergence rate than  $\widehat{\beta}(\tau)$  ( $\sqrt{n}$ -consistent),  $n h \mathcal{W}_n(\tau)$  is therefore pointwise asymptotically  $\chi_{d+1}^2$ -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. Similar idea has been explored by Kapetanios (2008) using *i.i.d.* bootstrap. Since bootstrap automatically takes care of the standardization, we consider the following test statistic

$$\widehat{\mathcal{W}}_n(\tau) = \left( \widehat{\mathcal{W}}_{n,0}(\tau), \dots, \widehat{\mathcal{W}}_{n,d}(\tau) \right)' := \left( \left( \widehat{\beta}_0(\tau) - \widehat{c}_0 \right)^2, \dots, \left( \widehat{\beta}_d(\tau) - \widehat{c}_d \right)^2 \right)' \quad (3.11)$$

without studentizing (and correcting bias), where  $\widehat{\beta}(\tau)$  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  $t = 1, \dots, n$ .

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

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

Step 3 Repeat Steps 3 and 4  $B$  times. Compute the  $(d+1)$ -dimensional vector of pointwise quantiles

$$\widehat{\mathbf{q}}_{\alpha_p}(\tau) = \left( \widehat{q}_{0,\alpha_p}(\tau), \dots, \widehat{q}_{d,\alpha_p}(\tau) \right)', \tau \in G, \text{ where}$$

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

From these pointwise quantiles, we obtain critical values for a significance level of  $\alpha$  using a similar procedure as we introduced in Section 3.1 to construct simultaneous confidence bands. Vary the pointwise error  $\alpha_p$  until

$$\mathbb{P}^* \left( \widehat{\mathcal{W}}_n^*(\tau) \geq \widehat{\mathbf{q}}_{1-\alpha_p}(\tau), \forall \tau \in G \right) \approx \alpha.$$

This means that the ratio of  $\widehat{\mathcal{W}}_{n,j}^*(\tau) \geq \widehat{q}_{j,1-\alpha_p}(\tau)$  for all  $\tau \in G$  and  $j \in \{0, \dots, d\}$  is approximately  $\alpha$ . Denote this  $\alpha_p$  by  $\alpha_s$  (simultaneous error). Reject the null if  $\widehat{\mathcal{W}}_n^*(\tau) \geq \widehat{\mathbf{q}}_{1-\alpha_s}(\tau)$  for some  $\tau \in G$ .

Note that we generate the bootstrap observations from the nonparametric fit rather than the parametric fit. This is because  $\widetilde{\beta}(\tau)$  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  $\widehat{\mathcal{W}}_n^*(\tau)$  using  $\left( \widehat{\beta}_j^*(\tau) - \widetilde{\beta}_j(\tau) \right)^2$ ,  $0 \leq j \leq d$  in Step 2, instead of the natural construction  $\left( \widehat{\beta}_j^*(\tau) - \widehat{c}_j^* \right)^2$ ,  $0 \leq j \leq d$ , where  $\widehat{c}_j^*$  are the bootstrap counterparts of  $\widehat{c}_j$ . Our approach ensures that  $nh\widehat{\mathcal{W}}_n^*(\tau)$  is bounded in probability under the alternative, leading to a nontrivial power. Finally, since we do not possess prior information about which coefficients are time-varying in our empirical study, we focus on jointly testing if  $\beta_j(\cdot)$ ,  $j = 0, \dots, d$ , are constants. It would be interesting to additionally test whether coefficients are partially time-invariant. In this case, one may construct a test statistic based on profile least squares estimators with local linear fitting in the first stage, see e.g. Li et al. (2011) and Gao et al. (2021). We leave this problem to future research. 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 (3.10),  $\left\{ nh\widehat{\mathcal{W}}_n(\tau_0 + \tau h) \right\}_{\tau \in [-1, 1]}$  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  $H_1 : \beta(\cdot) \neq \mathbf{c}$  for any  $\mathbf{c} \in \mathbb{R}^{d+1}$ . 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 can imply  $\beta(\cdot) = \mathbf{c}$  over  $G \subset [0, 1]$ . Hence, our test has correct size 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  $\beta(\cdot)$ . 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  $\beta(\cdot)$  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 to form the set  $G$ , such as  $\{1/5, \dots, 4/5\}$ . It is generally informative enough to tell if  $\beta(\cdot)$  is time-varying in practice.

## 4 Practical implementation

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

### 4.1 Bandwidth selection

Although Assumption A4 gives some guidance on choosing  $h$ , it does not provide a practical choice. We next discuss some potential ways to select bandwidths in practice. Although a theoretically optimal bandwidth can be obtained, it often depends on unobservable quantities that are hard to estimate (e.g. the second derivative of  $\beta(\cdot)$ ). We refer the interested reader to Fan and Gijbels (1996) for more details. We focus on the practical aspect here.

The first approach is based on the leave-one-out cross validation (CV). The first step is to construct the leave-one-out estimator for every time point by leaving out the observation  $t$  that receives the highest weight in the local estimation given an  $h$ , denoted by  $\hat{\beta}_{h,-t}(\cdot)$ . The second step in the least-squares CV approach is to look at the weighted average of the leave-one-out squared residuals

$$CV(h) = \frac{1}{n} \sum_{t=1}^n \left[ y_t - \mathbf{x}'_t \hat{\beta}_{h,-t}(t/n) \right]^2 \quad (4.1)$$

and minimize this criterion with respect to  $h$ .

Cross validation was originally designed for independent data and can therefore cause problems in time series applications. Chu and Marron (1991) propose a modified cross validation (MCV) and show that it works well for time series data. It follows the same general principle as CV, but it is based on a leave- $(2l + 1)$ -out estimator. That is,  $\mathbf{S}_{n,k}(\tau)$  and  $\mathbf{T}_{n,k}(\tau)$  in equation (2.4) are replaced by the following leave- $(2l + 1)$ -out counterparts

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

for  $k = 0, 1, 2$ . For  $l = 0$ , the original cross validation leave-one-out estimator is obtained.

In addition to cross validation methods, we also consider two additional bandwidth selection 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. Both approaches use the  $n \times n$  hat-matrix  $\mathbf{Q}_h$  which gives  $\hat{\mathbf{y}} = \mathbf{Q}_h \mathbf{y}$ , where  $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)'$ . Then, the AIC criterion and GCV are defined as

$$AIC(h) = \log \hat{\sigma}^2 + 2 \frac{\text{tr}(\mathbf{Q}_h) + 1}{n - \text{tr}(\mathbf{Q}_h) - 2}, \quad GCV(h) = \frac{n^{-1} \|\mathbf{y} - \hat{\mathbf{y}}\|^2}{[1 - \text{tr}(\mathbf{Q}_h)/n]^2}, \quad (4.2)$$

where  $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^n (y_t - \hat{y}_t)^2$ .

## 4.2 Bootstrap implementation

In Step 1 of the sieve bootstrap algorithms (Sections 3.1 and 3.4), a larger bandwidth is used to perform the nonparametric estimation. We follow Bühlmann (1998) and suggest using  $\tilde{h} = Ch^{5/9}$  for  $C = 0.5, 1, 2$ . Compared to the original bandwidth  $h$ , this bandwidth is larger and produces an oversmoothed estimate. We investigate the sensitivity to the oversmoothing parameter  $C$  in the simulation study. Furthermore, we use the AIC to select the lag length in Step 2. The selection should be done in such a way that the range of possible lag lengths increases with  $n$ . In practice, we recommend, as in Bühlmann (1998), to select  $p$  from a range of possible values in  $[0, p_{max}]$ , for example let  $p_{max} = 10 \log_{10}(n)$ .

## 5 Simulation study

In this section, we investigate the properties of the proposed bootstrap methods with the help of an extensive Monte Carlo study. First, we describe the data generating process (DGP). Second, we present results for a sample size of  $n = 100$ . Third, we increase the sample size to  $n = 300$ . Fourth, we investigate the performance of our proposed test for parameter stability.

### 5.1 The data generating process

In the empirical application in Section 6 we look at the price development of allowance prices in the European Emissions Trading Scheme (EU ETS). In particular, the relationship between allowance prices and their two main drivers is of interest. As explained in Section 6, in theory, one factor should have a negative effect on allowance prices and the other should have a positive effect. In line with the previous literature, we find, however, that the factor with the negative impact is not significant when using linear regression. This is because the effect is insignificant for the first part of our sample and becomes significant after the first few years. Additionally, the positive price determinant displays a two-peak shape in our dataset. We try to mimic both shapes in our simulation study by generating data from the following model:

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

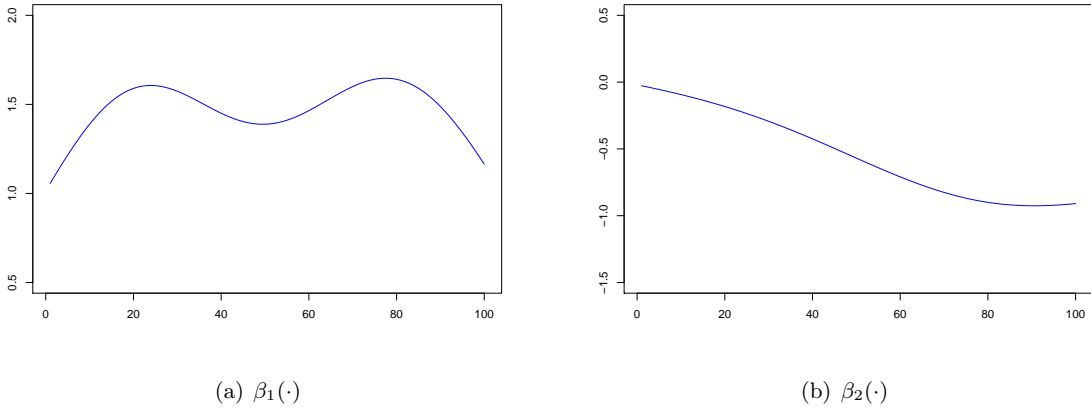


Figure 1: Plot of the coefficient curves  $\beta_1$  and  $\beta_2$  used in the simulation study

where the two coefficient functions are given by

$$\beta_1(t) = 1.5 \exp(-10(t - 0.2)^2) + 1.6 \exp(-8(t - 0.8)^2) \quad (5.2)$$

$$\beta_2(t) = -0.5t - 0.5 \exp(-5(t - 0.8)^2). \quad (5.3)$$

Figure 1 plots the two coefficient curves which are given by (5.2) and (5.3). It shows that  $\beta_1(\cdot)$  has two peaks and  $\beta_2(\cdot)$  starts at zero and then turns negative. As mentioned above, their shape is inspired by the coefficient estimates of the gas and coal price series of the empirical application in Section 6. For the error term  $\{u_t\}$  we consider an ARMA(1,1) process as given by

$$u_t = \phi u_{t-1} + \epsilon_t + \psi \epsilon_{t-1}, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}\left(0, \frac{1 - \phi^2}{2(1 + \psi^2 + 2\phi\psi)}\right), \quad (5.4)$$

where the variance of  $\epsilon_t$  is such that the signal to noise ratio does not change when we vary  $\phi$  and  $\psi$ . We allow for a dynamic specifications between the regressors  $x_{1,t}$  and  $x_{2,t}$ . Consider the following VAR(1) model:

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

where  $(\xi_{1,t}, \xi_{2,t})'$  are bivariate standard normal. The dynamics between  $\{x_{1,t}\}$  and  $\{x_{2,t}\}$  are inspired by the empirical study as the coefficients are obtained by estimating a bivariate VAR(1) from coal and gas prices in first differences.

To obtain the results presented in the next sections, we use 1000 Monte Carlo simulations. The sample size is  $n \in \{100, 300\}$ . We report 95% confidence intervals using  $B = 999$  bootstrap replications. We apply the local linear estimator with the Epanechnikov kernel  $K(x) = \frac{3}{4}(1 - x^2) \mathbb{1}_{\{|x| \leq 1\}}$ . We also report the average median length of the confidence intervals in parenthesis underneath the respective coverage. Next to pointwise confidence intervals, we also look at simultaneous confidence bands. For simultaneous coverage to be a success, we count the num-

ber of successes which means that the trend curve has to lie within the confidence bands for all points of the considered set  $G$ . We study the performance of the method for the two sets  $G_{sub}$  and  $G$  considered by Bühlmann (1998), where  $G_{sub} = U_1(h) \cup U_4(h)$  and  $G = \bigcup_{i=1}^4 U_i(h)$ , with  $U_i(h) = \{(i/5) - h + j/100, j = 0, \dots, [200h]\}$ . We additionally investigate simultaneous coverage over the whole sample.

## 5.2 Results for $n = 100$

The first set of results are presented in Tables 1 through 4. They are obtained with an oversmoothing parameter choice of  $C = 2$  and three different bandwidths  $h = 0.04, 0.06, 0.08$ . Table 1 gives coverage probabilities as well as the average median length for pointwise confidence intervals. The pointwise coverage represents the fraction of the data points for which the corresponding point on the true parameter curve lies within the confidence bands. Tables 2, 3 and 4 look at the three settings for simultaneous confidence bands:  $G$ ,  $G_{sub}$  and the whole sample, respectively. In this case, the coverage probability counts the number of times that the true parameter curve lies entirely within the confidence bands (for every point of the considered set). The left column of each table shows which AR and MA coefficient was used while the top row refers to the bandwidth. For each bandwidth, we present results for  $\beta_1$  and  $\beta_2$ .

Table 1: Pointwise coverage probabilities

$(\phi, \psi)$	$h = 0.04$		$h = 0.06$		$h = 0.08$	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
(0,0)	0.957 (0.307)	0.964 (0.310)	0.941 (0.247)	0.963 (0.250)	0.898 (0.214)	0.960 (0.216)
(0.3,0)	0.958 (0.302)	0.963 (0.301)	0.938 (0.246)	0.962 (0.244)	0.890 (0.214)	0.958 (0.212)
(0.5,0)	0.959 (0.288)	0.966 (0.282)	0.938 (0.236)	0.964 (0.232)	0.887 (0.206)	0.961 (0.202)
(-0.3,0)	0.964 (0.284)	0.967 (0.292)	0.946 (0.228)	0.967 (0.233)	0.891 (0.196)	0.963 (0.201)
(0.0.3)	0.956 (0.306)	0.965 (0.303)	0.940 (0.249)	0.964 (0.247)	0.898 (0.217)	0.960 (0.215)
(0.3,0.3)	0.950 (0.367)	0.956 (0.361)	0.935 (0.300)	0.955 (0.296)	0.901 (0.262)	0.951 (0.258)

Table 2: Coverage probabilities for  $G_{sub}$ 

$(\phi, \psi)$	$h = 0.04$		$h = 0.06$		$h = 0.08$	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
(0,0)	0.947 (0.401)	0.953 (0.404)	0.897 (0.324)	0.947 (0.328)	0.763 (0.283)	0.943 (0.286)
(0.3,0)	0.946 (0.394)	0.955 (0.393)	0.876 (0.322)	0.959 (0.321)	0.753 (0.282)	0.951 (0.280)
(0.5,0)	0.947 (0.376)	0.968 (0.369)	0.872 (0.309)	0.960 (0.305)	0.737 (0.272)	0.940 (0.268)
(-0.3,0)	0.956 (0.372)	0.969 (0.381)	0.902 (0.299)	0.965 (0.307)	0.737 (0.260)	0.970 (0.266)
(0.0,3)	0.947 (0.399)	0.950 (0.396)	0.895 (0.326)	0.949 (0.324)	0.740 (0.286)	0.944 (0.284)
(0.3,0.3)	0.946 (0.479)	0.944 (0.472)	0.878 (0.394)	0.941 (0.388)	0.774 (0.346)	0.938 (0.340)

Table 3: Coverage probabilities for  $G$ 

$(\phi, \psi)$	$h = 0.04$		$h = 0.06$		$h = 0.08$	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
(0,0)	0.946 (0.434)	0.954 (0.437)	0.892 (0.350)	0.951 (0.354)	0.772 (0.302)	0.947 (0.304)
(0.3,0)	0.950 (0.427)	0.955 (0.425)	0.888 (0.348)	0.957 (0.347)	0.747 (0.301)	0.956 (0.299)
(0.5,0)	0.941 (0.407)	0.955 (0.400)	0.880 (0.335)	0.953 (0.329)	0.743 (0.290)	0.937 (0.285)
(-0.3,0)	0.943 (0.403)	0.960 (0.413)	0.903 (0.323)	0.963 (0.331)	0.752 (0.277)	0.961 (0.284)
(0.0,3)	0.949 (0.433)	0.952 (0.429)	0.891 (0.352)	0.953 (0.350)	0.755 (0.305)	0.944 (0.302)
(0.3,0.3)	0.931 (0.519)	0.934 (0.511)	0.880 (0.425)	0.936 (0.419)	0.776 (0.368)	0.941 (0.363)

The coverage of the pointwise confidence intervals (Table 1) obtained by the sieve bootstrap is close to the nominal level of 95% in almost all cases, independent of the chosen bandwidth and the degree of autocorrelation. This holds for both coefficient functions. In Tables 2 and 3 we observe that this is still true for  $\beta_2$  when we move to simultaneous coverage over the sets  $G$  and  $G_{sub}$ . Looking at the same values for  $\beta_1$ , the coverage drops as the bandwidth gets larger. The coverage drops to 74% for  $G_{sub}$  and  $G$  in the case of AR(1) errors with  $\phi = 0.5$  for the largest considered bandwidth. For the smallest bandwidth, the coverage for  $\beta_1$  is close to the nominal level. This sensitivity with respect to the bandwidth observed for  $\beta_1$  is not unexpected given the more complex shape of the parameter curve compared to  $\beta_2$ . In Table 4, the same pattern arises when we consider simultaneity over the whole sample, while the confidence intervals become slightly

wider compared to the previous tables to account for the substantially larger set of values considered for the simultaneity.

Table 4: Coverage probabilities for the whole sample

$(\phi, \psi)$	$h = 0.04$		$h = 0.06$		$h = 0.08$	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
(0,0)	0.925 (0.461)	0.931 (0.464)	0.896 (0.367)	0.948 (0.371)	0.798 (0.312)	0.944 (0.315)
(0.3,0)	0.917 (0.453)	0.940 (0.451)	0.882 (0.365)	0.946 (0.363)	0.771 (0.311)	0.955 (0.309)
(0.5,0)	0.905 (0.432)	0.937 (0.425)	0.880 (0.350)	0.941 (0.345)	0.757 (0.300)	0.938 (0.295)
(-0.3,0)	0.936 (0.427)	0.932 (0.438)	0.907 (0.338)	0.949 (0.347)	0.780 (0.286)	0.952 (0.293)
(0.0,3)	0.919 (0.459)	0.947 (0.455)	0.902 (0.369)	0.951 (0.366)	0.779 (0.315)	0.943 (0.312)
(0.3,0.3)	0.894 (0.551)	0.896 (0.542)	0.873 (0.446)	0.933 (0.439)	0.787 (0.382)	0.936 (0.375)

The next set of results investigates the role of the oversmoothing parameter  $C$ . Here, we only consider a limited number of AR and MA specifications  $((0, 0), (0.3, 0), (0, 0.3))$  and we restrict ourselves to one bandwidth ( $h = 0.04$ ). The results are presented in Table 5. The top part shows pointwise coverage and the bottom part simultaneous coverage over  $G_{sub}$  for an increasing value of  $C = 0.5, 1, 2$  (Bühlmann, 1998). For ease of comparison, the final two columns repeat the values from Tables 1 and 2, respectively. With increasing  $C$  the confidence intervals become wider and thus, the coverage increases. Values of 1 and 2 both produce accurate coverage when it comes to pointwise confidence intervals. In the case of simultaneous confidence bands, the method works best with an oversmoothing parameter of  $C = 2$ .

We additionally investigated a setting that allowed for conditional heteroskedasticity by letting the errors follow a GARCH(1,1) specification. In this scenario, the sieve bootstrap becomes invalid and we use a sieve wild extension as well as an autoregressive wild bootstrap. The results are presented in Tables 9 to 11 in Appendix B. The wild bootstrap methods produced confidence bands with surprisingly low coverage. This might be a small sample problem because the performance improved when considering  $n = 300$ . To further investigate this is a promising avenue for future research.

### 5.3 Increasing the sample size

In Table 6 we see how the coverage and length of the confidence intervals change when we increase the sample size from 100 to 300. We consider pointwise and simultaneous inference for  $G$ ,  $G_{sub}$  and the whole sample. We restrict ourselves to three error specifications from Table 5 and one bandwidth ( $h = 0.04$ ). The method performs well in almost all cases. Coverage probabilities are even slightly too high in some cases. Merely the simultaneity over the whole sample remains at a

Pointwise coverage						
$(\phi, \psi)$	$C = 0.5$		$C = 1$		$C = 2$	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
(0,0)	0.909 (0.262)	0.918 (0.265)	0.942 (0.287)	0.950 (0.290)	0.957 (0.307)	0.964 (0.310)
(0.3,0)	0.90977 (0.258)	0.91164 (0.259)	0.944 (0.283)	0.947 (0.282)	0.958 (0.302)	0.963 (0.301)
(0,0.3)	0.90579 (0.259)	0.9161 (0.257)	0.941 (0.285)	0.950 (0.283)	0.956 (0.306)	0.965 (0.303)
Simultaneous coverage ( $G_{sub}$ )						
(0,0)	0.846 (0.343)	0.849 (0.347)	0.920 (0.375)	0.926 (0.379)	0.947 (0.401)	0.953 (0.404)
(0.3,0)	0.869 (0.337)	0.875 (0.339)	0.929 (0.369)	0.939 (0.368)	0.947 (0.394)	0.955 (0.393)
(0,0.3)	0.836 (0.338)	0.861 (0.336)	0.919 (0.372)	0.929 (0.370)	0.947 (0.399)	0.950 (0.396)

Table 5: Coverage probabilities for varying  $C$ ,  $h = 0.04$

small loss of coverage compared to the nominal level. The most observable change is that for the increased sample size, the intervals become more narrow while the coverage remains similar.

$(\phi, \psi)$	pointwise		$G_{sub}$		$G$		whole sample		
	$n = 100$	$n = 300$	$n = 100$	$n = 300$	$n = 100$	$n = 300$	$n = 100$	$n = 300$	
$\beta_1$	(0,0)	0.957 (0.307)	0.962 (0.169)	0.947 (0.401)	0.932 (0.222)	0.946 (0.434)	0.951 (0.241)	0.925 (0.461)	0.936 (0.258)
	(0.3,0)	0.958 (0.302)	0.963 (0.172)	0.946 (0.394)	0.953 (0.227)	0.950 (0.427)	0.963 (0.245)	0.917 (0.453)	0.942 (0.263)
	(0,0.3)	0.956 (0.306)	0.961 (0.171)	0.947 (0.399)	0.937 (0.225)	0.949 (0.433)	0.941 (0.244)	0.919 (0.459)	0.929 (0.262)
$\beta_2$	(0,0)	0.964 (0.310)	0.970 (0.171)	0.953 (0.404)	0.970 (0.225)	0.954 (0.437)	0.971 (0.244)	0.931 (0.464)	0.957 (0.262)
	(0.3,0)	0.963 (0.301)	0.972 (0.171)	0.955 (0.393)	0.979 (0.225)	0.955 (0.425)	0.971 (0.244)	0.940 (0.451)	0.960 (0.262)
	(0,0.3)	0.965 (0.303)	0.968 (0.171)	0.950 (0.396)	0.971 (0.224)	0.952 (0.429)	0.959 (0.243)	0.947 (0.455)	0.946 (0.261)

Table 6: Coverage probabilities for  $n = 300$ ,  $h = 0.04$

#### 5.4 Parameter stability test

The empirical size and power of the bootstrap test for parameter stability are reported in Table 7. Overall, the test has the best performance in terms of both empirical size and power when  $h = 0.08$ . We further make three observations. First, a larger  $h$ , and therefore a larger  $\tilde{h}$ , yields a better size control due to our construction of the bootstrap test. More specifically, we use  $(\hat{\beta}_j^*(\tau) - \tilde{\beta}_j(\tau))^2$  to

$\tilde{h} = 2h^{5/9}$	$n = 100$			$n = 300$		
	$h = 0.04$	$h = 0.06$	$h = 0.08$	$h = 0.04$	$h = 0.06$	$h = 0.08$
Size						
(0, 0)	9.6	9.0	7.1	5.1	5.7	4.7
(0.3, 0)	7.3	7.5	7.3	6.2	4.5	4.0
(0.5, 0)	9.4	8.4	6.7	5.6	4.5	4.1
(−0.3, 0)	7.6	6.9	6.4	6.0	6.6	4.3
(0, 0.3)	10.8	9.7	5.2	7.5	4.8	3.9
(0.3, 0.3)	9.5	7.3	7.5	5.9	4.5	4.3
Power						
(0, 0)	27.9	42.4	54.5	72.2	92.5	97.2
(0.3, 0)	29.1	41.7	53.4	67.3	88.0	96.3
(0.5, 0)	29.0	44.0	57.3	70.3	87.4	95.4
(−0.3, 0)	30.9	48.3	55.6	76.4	95.7	98.2
(0, 0.3)	28.7	41.2	53.7	71.4	89.6	97.4
(0.3, 0.3)	31.1	45.4	55.0	68.3	87.7	96.2

Table 7: The empirical size and power (in %) of bootstrap tests for parameter stability.

gain approximations to the test statistics  $(\hat{\beta}_j(\tau) - \hat{c}_j)^2$ ,  $j = 0, 1, \dots, d$ . The first-step smoothers  $\tilde{\beta}_j(\tau)$  fluctuate less with a larger  $\tilde{h}$ , mimicking the data better under the null hypothesis. Second, the test is mildly oversized when  $n = 100$ . It yields a more accurate empirical size as  $n$  increases to 300. Finally, the empirical power also increases with the sample size in line with Proposition 1(ii). Since a false negative may result in model misspecifications, a nontrivial empirical power is crucial in applications.<sup>2</sup>

## 6 The EU Emissions Trading System

### 6.1 Background and related literature

The EU ETS – which is short for European Emissions Trading System – is the largest and oldest cap-and-trade program. Regulated firms and other participants can buy and sell permits that can be used to match emissions of one metric ton of CO<sub>2</sub> or CO<sub>2</sub> equivalent. These permits are called emission allowances (EUAs). The EU ETS covers around 45% 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. Economic theory identifies clear price drivers among which the most important ones are the coal and the gas price. 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

<sup>2</sup>A false positive may suffer less from this issue because one would instead use a more flexible semi-parametric model.



electricity generation. In theory, due to the 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). Other price driving factors are e.g. 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.

We plot the allowance price together with the coal and gas price in Figure 2. Panel (a) plots the EUA price (black) and the coal price (red). We indeed see some opposite movements in the two time series. We observe that the two series indeed move in opposing directions most of the time 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 (blue). 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.

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 exercises performed on our data. Table 8 displays the results. The two significant factors are the gas and the oil price. The coal price, as an important driver, does not show a significant effect on the allowance price in these initial linear regressions. The results are robust regarding the choice of indicator of economic activity; the estimates are very similar, if we include the alternative stock index, or if we include the oil price.<sup>3</sup> 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 in Friedrich et al. (2019) and Lutz et al. (2013). The latter paper considers potential non-linearities using a regime-switching model. They distinguish two different pricing regimes - one applies during

---

<sup>3</sup>Inclusion of the stock indices in addition the oil price left the other coefficients nearly unchanged and the index's coefficient is not significant. The results can be obtained on request.

OLS regression results									
	(a)			(b)			(c)		
	$\beta$	$se_{NW}$	$p$ -value	$\beta$	$se_{NW}$	$p$ -value	$\beta$	$se_{NW}$	$p$ -value
Intercept	0.000	0.003	0.774	0.000	0.003	0.928	0.000	0.003	0.769
Coal	-0.119	0.094	0.206	-0.061	0.097	0.528	-0.07	0.097	0.425
Gas	0.190	0.075	0.012	0.198	0.074	0.007	0.198	0.074	0.008
Oil	0.214	0.069	0.002	–	–	–	–	–	–
Temp	-0.001	0.001	0.572	-0.001	0.001	0.450	-0.001	0.001	0.451
Stoxx 50	–	–	–	0.139	0.103	0.031	–	–	–
Stoxx 600	–	–	–	–	–	–	0.296	0.110	0.007
Rejection Rate	77%			64%			63%		

Table 8: Linear regression results from a regression of the allowance price on different sets of explanatory variables: (a) includes the oil price (b) the stock index STOXX 50 and (c) the index STOXX 600. The estimated coefficients ( $\beta$ ) as well as Newey-West standard errors  $se_{NW}$  and the corresponding  $p$ -values are reported. The row labeled “Rejection Rate” indicates the ratio of rejecting the null of parameter constancy, given  $(\alpha, h) = (10\%, 0.08)$  and  $\tilde{h} = 2h^{5/9}$ , by repeating the testing procedure 100 times developed in Section 3.4.

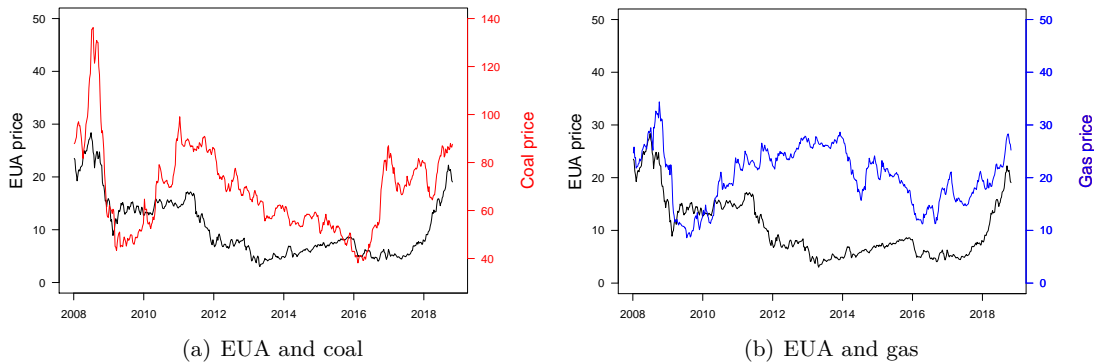


Figure 2: Joint plot of (a) EUA and coal prices and (b) EUA and gas prices.

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 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 be subject to (structural) changes. Such observation is also supported by our parameter stability test given in Section 3.4. Since, for each simulation, our testing procedure gives a binary outcome (reject or not reject) for a significance level  $\alpha$ , say  $\alpha = 10\%$ , we repeat 100 simulations to obtain the rejection rates (Table 8). The results show that there is a more than 60% chance of rejecting parameter constancy. Together with the former studies, a more robust model

specification shall be adopted.

## 6.2 The data and empirical results

We consider weekly data for the period from January 2008 to October 2018 resulting in  $T = 538$  observations. We use EUA December futures contract traded on the European Energy Exchange (EEX). 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 price series. As our main set of explanatory variables, we include natural gas, coal and oil prices as month-ahead futures. We further use two stock indices as supplementary indicators 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.<sup>4</sup> Both are obtained from the EEX. For oil we rely on the historical futures prices (continuous contract) of Brent crude oil based on raw data from the Intercontinental Exchange (ICE), retrieved from Quandl. The contract size is 1,000 barrels. The coal and the oil prices need to be converted into EUR, as they are denoted in USD. This is done using the daily USD/EUR exchange rate data from Tullett Prebon. The first stock index is the STOXX Europe 50. As an alternative, we use data on a comparable index, which is sometimes used in this context, the STOXX Europe 600 index. We use temperature data constructed as an average over seven European cities from the European Climate Assessment & Dataset (ECA&D) which provides surface air temperature for 199 measurement stations in Europe. It is provided by The Royal Netherlands Meteorological Institute (KNMI). We refer to Klein Tank et al. (2002) for more details on the temperature series and measurement stations. 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 Appendix C for details).<sup>5</sup> Figure 2 plots the various data series. All price and index series experienced a sharp drop at the beginning of our sample owing to the global financial crisis. Additionally, panels (d) and (e) show that the two stock indices show an almost identical development over our sample period. A unit root test performed on the data shows that all series except for the temperature data contain a unit root and we work with log returns. The results for the test are shown in the Supplementary Appendix.

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 Santos et al. (2008). This approach retains 7 outliers which we detail in Appendix C. The results we present here are the results after outliers have been removed.<sup>6</sup> We apply the

---

<sup>4</sup>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.

<sup>5</sup>We considered data on hydro power in Norway from the Norwegian Water Resources and Energy Directorate as well as wind and solar production data for Germany obtained from the database of the European Network of Transmission System Operators for Electricity (ENTSO-E). However, none of the time series showed a significant effect in our regression analysis.

<sup>6</sup>In the Supplementary Appendix we repeat the analysis using the original data and there are no major differences.

local linear estimator with the Epanechnikov kernel and a sieve bootstrap procedure with  $B = 999$  to generate 95% confidence bands which are simultaneous over the whole sample. To determine a suitable bandwidth, we perform data-driven bandwidth selection with the methods introduced in Section 4.1. We use  $h = 0.09$  which is the rounded average of the selected values.<sup>7</sup>

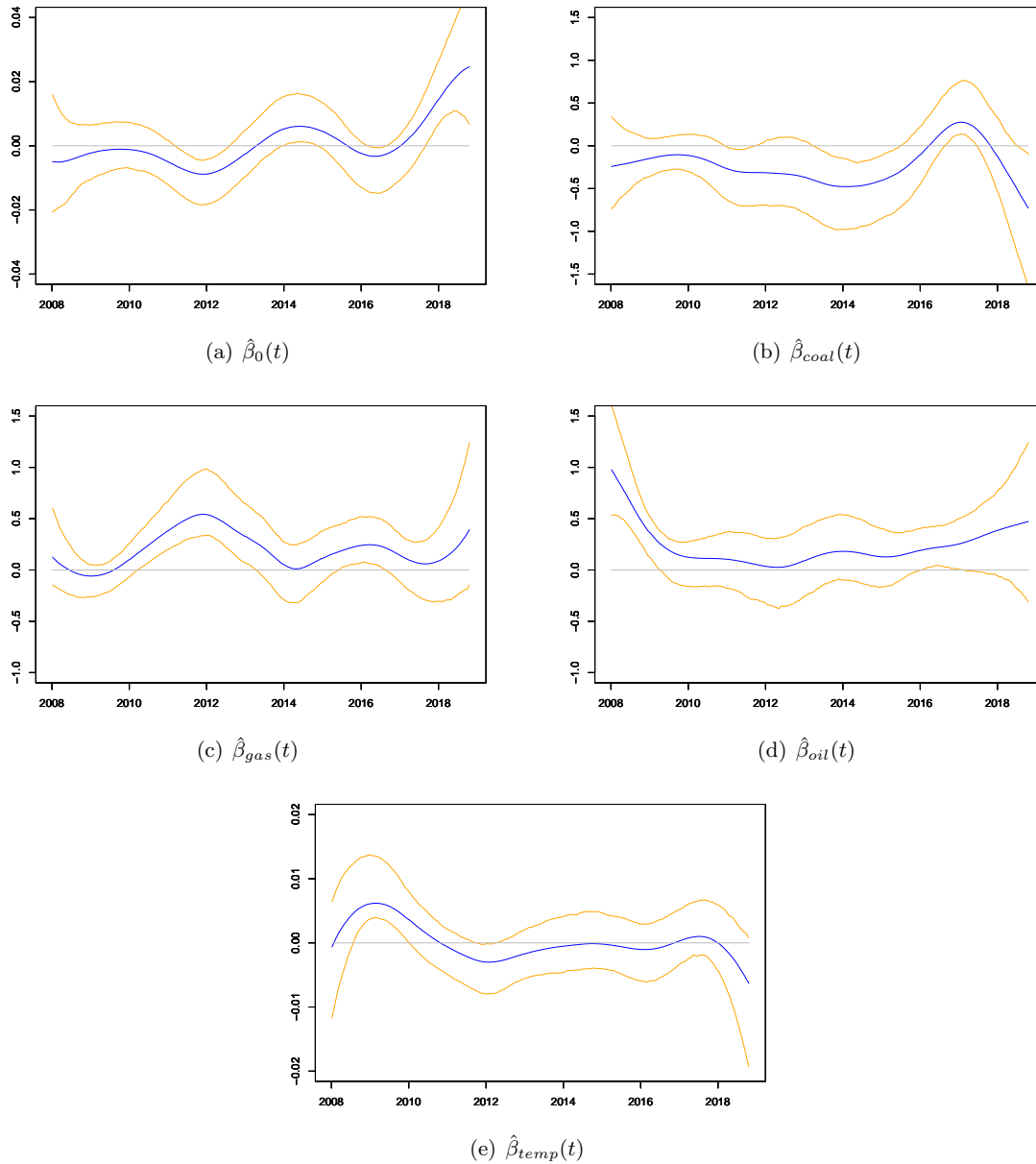


Figure 3: Nonparametrically estimated coefficient curves and 95% confidence intervals

Figure 3 plots the results. We see significant time variation. In Panel (a), the nonparametric trend fluctuates around zero for most of the considered time span. However, at the end of the sample, it turns significantly positive. This shows that the drastic price increase, visible in Figure 2, seems to be picked up by the trend component. Panel (b) shows that the coal price has a

<sup>7</sup>Given our simulation results, we repeat the analysis with a smaller value ( $h = 0.05$ ) for robustness and observe that the main conclusions stay unchanged. The results can be obtained from the authors upon request.

significant negative effect for two periods: one ranging from 2010 to mid 2011 and one from 2013 to mid 2015. Subsequently, the coefficient becomes positive and significant for a short period in 2016/2017. This is a very interesting finding given that its coefficient was found to be insignificant in the linear regression analysis suggesting that time variation might have caused this result. In Panel (c), we see that the gas coefficient has the expected positive sign and is significant over long periods. The coefficient of the oil price, as displayed in Panel (d) is positive over two periods – until 2009 and from 2015 onward. Finally, in Panel (e) we see that the temperature series shows only two very short periods of significance and the magnitude of the coefficient estimate is small.

## 7 Conclusion

In this paper, we have considered trending time-varying coefficient models that allow for serial dependence. These models can capture common relationships that gradually vary over time. Recent empirical studies have frequently applied bootstrap methods to construct simultaneous confidence bands in these models due to the easy implementation and good performance in small samples. However, a theoretical justification is currently unavailable.

We fill this gap by proposing a sieve bootstrap framework to conduct inference. We find that it can consistently estimate nuisance parameters in one go, both at the interior and boundary points. Therefore, pointwise and simultaneous confidence bands are simple to obtain. An extensive simulation study shows that our confidence bands provide accurate coverage in small samples, in line with the theoretical results. Moreover, since current related climatic analyses often adopt linear models with constant parameters, we then suggest a bootstrap-based test for parameter stability. We show that the test has asymptotically correct size and performs well in simulations.

In the empirical application, we study the price development of allowances in the largest cap and trade market for CO<sub>2</sub> emissions – the European Emissions Trading System. We provide 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 which has concluded that the relationship might be unstable. The time variation offers a potential explanation for insignificant coefficients found with linear regression techniques used in some of the previous work.

## 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(October 2007):380–395.
- 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. *Annals of Statistics*, 26:48–83.
- 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 Pivot, Cham., New York, NY.
- Cavaliere, G. and Taylor, A. M. R. (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory*, 24(1):43–71.
- Cavaliere, G. and Taylor, A. M. R. (2009). Heteroskedastic time series with a unit root. *Econometric Theory*, 25(5):1228–1276.
- 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.
- 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*, n/a(n/a).
- Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications*. Chapman and Hall, London.
- Friedrich, M., Mauer, E.-M., Pahle, M., and Tietjen, O. (2019). From fundamentals to financial assets: the evolution of understanding price formation in the EU ETS. Technical report, Kiel, Hamburg.

- Friedrich, M., Smeekes, S., and Urbain, J.-P. (2020). Autoregressive wild bootstrap inference for nonparametric trends. *Journal of Econometrics*, 214:81–109.
- Gao, J., Peng, B., and Yan, Y. (2021). Parameter stability testing for multivariate dynamic time-varying models. *Monash Business School, Working Paper 11/21*.
- Gine, E. and Zinn, J. (1990). Bootstrapping General Empirical Measures. *The Annals of Probability*, 18(2):851–869.
- Gonzalés, 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.
- 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.
- Hu, L., Huang, T., and You, J. (2019). Estimation and identification of a varying-coefficient additive model for locally stationary processes. *Journal of the American Statistical Association*, 114(527):1191–1204.
- ICAP (2020). Emissions Trading Worldwide: Status Report 2020. Technical report, Berlin: ICAP.
- Kapetanios, G. (2008). Bootstrap-based tests for deterministic time-varying coefficients in regression models. *Computational Statistics & Data Analysis*, 53:534–545.
- Karmakar, S., Richter, S., and Wu, W. B. (2021). Simultaneous inference for time-varying models. *Journal of Econometrics*.
- Klein Tank, A. M. G., Wijngaard, J. B., and et al. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, 22(12):1441–1453.
- 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.
- Kristensen, D. (2012). Non-parametric detection and estimation of structural change. *Econometrics 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., 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.
- Liang, Z. and Li, Q. (2012). Functional coefficient regression models with time trend. *Journal of Econometrics* 170, pages 15–31.
- 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.
- 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.
- Orbe, S., Ferreira, E., and Rodriguez-Poo, J. (2005). Nonparametric estimation of time varying parameters under shape restrictions. *Journal of Econometrics*, 126(1):53–77.
- Orbe, S., Ferreira, E., and Rodriguez-Poo, J. (2006). On the estimation and testing of time varying constraints in econometric models. *Statistica Sinica*, 16(4):1313–1333.
- 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., Li, D., and Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, 196(1):180–195.
- 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.
- 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.
- Robinson, P. M. (1989). *Nonparametric estimation of time varying parameters* In: Hackl, P. (Ed.), *Statistics, Analysis and Forecasting of Economic Structural Change*. Springer, Berlin.
- Robinson, P. M. (1991). *Time varying nonlinear regression* In: Hackl, P. (Ed.), *Statistics, Analysis and Forecasting of Economic Structural Change*. Springer, Berlin.
- Santos, C., Hendry, D., and Johansen, S. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, 23:317–335.
- Smeeke, 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.
- 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. (2012). Inference of time-varying regression models. *The Annals of Statistics*, 40(3):1376 – 1402.
- 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*, 72:513–531.

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

**Lemma 1**

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

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

where  $Y_t \in \mathbb{R}$  is a strictly  $\alpha$ -mixing stationary process with the mixing coefficient  $\alpha(m) = O(m^{-\beta})$  and  $\mathbb{E}|Y_0|^s < \infty$ , where  $\beta = \varphi$  with  $\varphi$  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

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

**Lemma 2**

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

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

**Lemma 3**

Under Assumption C1,  $\sum_{j=-\infty}^{\infty} \|\text{cov}(\mathbf{x}_t z_t, \mathbf{x}_{t+j} z_{t+j})\| < \infty$  so that  $\mathbf{A}$  exists.

**Lemma 4**

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

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

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 Equation (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)  $\sup_{j \in \mathbb{N}} \left| \widehat{\psi}_{j,n} - \psi_j \right| = O_p \left( \sqrt{\frac{\ln n}{n}} \right) + O_p(p^{-1}) + O_p \left( \tilde{h}^2 + \sqrt{\frac{\ln n}{nh}} \right)$ .
- (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  $\mathbf{Z}_{n,k}^*(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^n \mathbf{x}_t z_t^* \left( \frac{\tau_t - \tau}{h} \right)^k K \left( \frac{\tau_t - \tau}{h} \right)$ ,  $k = 0, 1$ . Take  $k_1, k_2 \in \{0, 1\}$ . (i) For any fixed  $\tau \in (0, 1)$ ,  $\text{cov}^* \left( \mathbf{Z}_{n,k_1}^*(\tau), \mathbf{Z}_{n,k_2}^*(\tau) \right) = \nu_{k_1+k_2} \mathbf{A} + o_p(1)$ . (ii) For the left endpoint  $\tau = ch$ ,  $c \in (0, 1)$ ,  $\text{cov}^* \left( \mathbf{Z}_{n,k_1}^*(\tau), \mathbf{Z}_{n,k_2}^*(\tau) \right) = \nu_{k_1+k_2,c} \mathbf{A} + o_p(1)$ .

**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  $h + \frac{1}{nh} \xrightarrow{n \rightarrow \infty} 0$  hold. Let  $C$  be uniform over  $\tau_1, \tau_2$  and  $i = 0, 1, \dots, d$ .

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

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

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

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

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

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

The  $O_p(1)$  terms in (A.6) and (A.8) are uniformly over  $\tau_1, \tau_2$ .

**A.2 Proofs of main results**

**Lemma 6**

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

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

$$\sqrt{nh} \mathbf{H} \left( \hat{\boldsymbol{\theta}}^*(\tau) - \tilde{\boldsymbol{\theta}}(\tau) - h^2 \mathbf{b}(\tau) \right) = \mathbf{S}^{-1} \mathbf{Z}_n^*(\tau) + \mathbf{R}_n^*(\tau), \quad (\text{A.9})$$

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

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

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

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

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

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

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

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

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

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

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

$$\text{Var}^*(\mathbf{Z}_n^*(\tau)) = (\nu_0 \nu_2) \otimes \mathbf{A} \mathbb{1}_{\{\tau \in [\tau_L, \tau_U]\}} + (\nu_{1,c} \nu_{2,c}) \otimes \mathbf{A} \mathbb{1}_{\{\tau = ch\}} + o_p(1). \quad (\text{A.12})$$

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

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

For  $\tilde{\mathbf{R}}_n^*(\tau)$ , we note that, if  $|\tau_t - \tau| \leq h$ ,

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

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

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

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

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

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

*Proof of Theorem 1* We first consider the bootstrap quantities. By virtue of Lemma 6, we only

have to establish the asymptotic normality of  $\mathbf{Z}_n^*(\tau)$ . As such, we use the Cramér-Wold device. For convenience, we define  $Q_t(\tau) = \mathbf{a}'(\mathbf{x}'_t, \mathbf{x}'_t(\frac{\tau_t - \tau}{h}))' K(\frac{\tau_t - \tau}{h})$ , where  $\mathbf{a} \in \mathbb{R}^{2(d+1)}$  be any unit vector, and the truncated bootstrap errors  $z_{t,M_n}^* = \sum_{j=0}^{M_n} \widehat{\psi}_{j,n} \varepsilon_{t-j}^*$  with  $M_n \xrightarrow{n \rightarrow \infty} \infty$ . For  $\tau \in [0, 1]$ , then we can write

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

where the  $O_p^*$ -term is uniform over  $\tau \in [0, 1]$ . We obtain it by, for  $k \geq 0$ ,  $\left| \sum_{j=M_n+1}^{\infty} \widehat{\psi}_{j,n} \widehat{\psi}_{j+k,n} \right| \leq \left( \sum_{j=M_n+1}^{\infty} \widehat{\psi}_{j,n} \right)^2 = O_p(M_n^{-2})$  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  $\{1, \dots, n\} = \cup_{j=1}^{k_n} B_j$ , where  $B_j = \{b_j + 1, \dots, b_j + \ell_n\} \cup \{b_j + \ell_n + 1, \dots, b_{j+1}\}$ ,  $k_n = \lceil n/(\ell_n + s_n) \rceil$ ,  $b_j = (j-1)(\ell_n + s_n)$ . 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  $1/\ell_n + \ell_n/(nh) + M_n/\ell_n \rightarrow 0$  and  $1/s_n + s_n/\ell_n + M_n/s_n$  as  $n \rightarrow \infty$ . 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 ( $\ell_n$ ) 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  $w_t^k(\tau) = (\frac{\tau_t - \tau}{h})^k K(\frac{\tau_t - \tau}{h})$ . 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  $\left( \sum_{t=L}^U q_t \right)^2 = \sum_{i=-(U-L)}^{U-L} \sum_{t=L}^{U-|i|} q_t q_{t+|i|}$ ,

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

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),  $\|\mathbf{A} \otimes \mathbf{B}\| = \|\mathbf{A}\| \|\mathbf{B}\|$  for any matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\sum_{j=1}^{k_n} \sum_{t=b_j+\ell_n+1}^{b_{j+1}-|i|} |w_t^{k_1}(\tau) w_{t+|i|}^{k_2}(\tau)| \leq C k_n s_n h$ , 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

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

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

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

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

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

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

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

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  $\mathbb{E} |\varepsilon_t|^2 r < \infty$ ,  $r > 2$ , in Assumption A1) conditionally on the original data. Using Lemma 2 in Hansen (1991) (taking  $c_i \leq C |Q_i(\tau)|$ ) and

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

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

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  $\ell_n/(nh) \rightarrow 0$ .

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

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

where  $\mathbf{Z}_n(\tau) = (\mathbf{Z}_{n,0}(\tau)', \mathbf{Z}_{n,1}(\tau)')$  with  $\mathbf{Z}_{n,k}(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^n \mathbf{x}_t z_t \left( \frac{\tau_t - \tau}{h} \right)^k K \left( \frac{\tau_t - \tau}{h} \right)$ ,  $k = 0, 1$ , and  $\sup_{\tau \in [0,1]} \|\mathbf{R}_n(\tau)\| = O_p(\sqrt{nh^7}) + O_p(h^2 \sqrt{\ln n}) + o_p(1)$ . 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  $\mathbf{a}' \mathbf{Z}_n(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^n Q_t(\tau) z_{t,M_n} + O_p(M_n^{-1}) = \sum_{j=1}^{k_n} V_{n,j}(\tau) + \sum_{j=1}^{k_n} W_{n,j}(\tau) + O_p(M_n^{-1})$ , where  $\mathbb{V}\text{ar} \left( \sum_{j=1}^{k_n} W_{n,j}(\tau) \right) = O_p \left( \frac{k_n s_n h}{nh} \right) = o_p(1)$  and

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

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

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

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

Similarly,  $\sup_{\tau \in [-1, 1]} \left\| \sqrt{nh} \mathbf{H} \left( \widehat{\boldsymbol{\theta}}^*(\tau_0 + \tau h) - \widetilde{\boldsymbol{\theta}}(\tau_0 + \tau h) - h^2 \mathbf{b}(\tau_0) \right) - \mathbf{S}^{-1} \mathbf{Z}_n^*(\tau_0 + \tau h) \right\| = o_p^*(1)$  follows from (A.9) in Lemma 6. Therefore, we only have to consider  $\mathbf{S}^{-1} \mathbf{Z}_n(\tau_0 + \tau h)$  and  $\mathbf{S}^{-1} \mathbf{Z}_n^*(\tau_0 + \tau h)$  in the following. It suffices to show the weak convergence of the  $\mathbb{R}^q$ -valued processes  $\mathbf{W}_{\tau_0, n}(\tau) := \mathbf{Z}_n(\tau_0 + \tau h)$  and  $\mathbf{W}_{\tau_0, n}^*(\tau) := \mathbf{Z}_n^*(\tau_0 + \tau h)$  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  $\{\mathbf{a}_t\}$  and  $\{\mathbf{b}_t\}$ , we have the identity  $(\sum_{t=1}^n \mathbf{a}_t)(\sum_{t=1}^n \mathbf{b}_t) = \sum_{t=1}^n \mathbf{a}_t \mathbf{b}_t' + \sum_{i=1}^{n-1} \sum_{t=1}^{n-i} (\mathbf{a}_t \mathbf{b}_{t+i}' + \mathbf{a}_{t+i} \mathbf{b}_t')$ . By this identity and  $\mathbb{E}(\mathbf{W}_{\tau_0, n}(\tau)) = 0$ ,  $\tau \in [-1, 1]$ , one can write

$$\text{cov}(\mathbf{W}_{\tau_0, n}(\tau_1), \mathbf{W}_{\tau_0, n}(\tau_2)) =: \mathbf{I}_{0, n}(\tau_1, \tau_2; \tau_0) + \mathbf{I}_{1, n}(\tau_1, \tau_2; \tau_0) + \mathbf{I}_{2, n}(\tau_1, \tau_2; \tau_0), \quad (\text{A.19})$$

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

$$\mathbf{I}_{1, n}(\tau_1, \tau_2; \tau_0) = \frac{1}{nh} \sum_{i=1}^{n-1} \sum_{t=1}^{n-i} \mathbf{Q}_{t, t+i}(\tau_1, \tau_2; \tau_0) \otimes \mathbb{E}(\mathbf{x}_t \mathbf{x}_{t+i}' z_t z_{t+i}),$$

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

$$\mathbf{Q}_{s, t}(\tau_1, \tau_2; \tau_0) = \begin{pmatrix} w_s^0(\tau_0 + \tau_1 h) w_t^0(\tau_0 + \tau_2 h) & w_s^0(\tau_0 + \tau_1 h) w_t^1(\tau_0 + \tau_2 h) \\ w_s^1(\tau_0 + \tau_1 h) w_t^0(\tau_0 + \tau_2 h) & w_s^1(\tau_0 + \tau_1 h) w_t^1(\tau_0 + \tau_2 h) \end{pmatrix}. \quad (\text{A.20})$$

Consider  $\mathbf{I}_{0, n}(\tau_1, \tau_2; \tau_0)$  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  $\kappa_0$  is a constant, and the Riemann sum approximation (S.6) and a change of variables with  $u = (z - \tau_0)/h$ ,

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

when  $n$  is sufficiently large. Using (A.21), we obtain  $\lim_{n \rightarrow \infty} \mathbf{I}_{0, n}(\tau_1, \tau_2; \tau_0) = \boldsymbol{\kappa}(\tau_1, \tau_2) \otimes \text{Var}(\mathbf{x}_t z_t)$ . For  $\mathbf{I}_{i, n}(\tau_1, \tau_2; \tau_0)$ ,  $i = 1, 2$ , 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  $\frac{1}{H_n} + \frac{H_n}{nh} \rightarrow 0$  as  $n \rightarrow \infty$ , and obtain

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

Similarly,  $\lim_{n \rightarrow \infty} \mathbf{I}_{2, n}(\tau_1, \tau_2; \tau_0) = \boldsymbol{\kappa}(\tau_1, \tau_2) \otimes \sum_{i=-\infty}^{-1} \text{cov}(\mathbf{x}_t z_t, \mathbf{x}_{t+i} z_{t+i})$ . Overall,

$$\lim_{n \rightarrow \infty} \text{cov}(\mathbf{Z}_n(\tau_0 + \tau_1 h), \mathbf{Z}_n(\tau_0 + \tau_2 h)) = \boldsymbol{\kappa}(\tau_1, \tau_2) \otimes \mathbf{A}. \quad (\text{A.23})$$



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

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

For finite  $k \in \mathbb{N}$ , fix  $\tau_1, \dots, \tau_k \in [-1, 1]$ . Finite-dimensional distributions of the  $kq$ -dimensional vectors  $(\mathbf{W}_{\tau_0, n}(\tau_1)', \dots, \mathbf{W}_{\tau_0, n}(\tau_k)')'$  weakly converges, and  $(\mathbf{W}_{\tau_0, n}^*(\tau_1)', \dots, \mathbf{W}_{\tau_0, n}^*(\tau_k)')'$  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  $\mathbf{W}_{\tau_0, n}(\cdot)$  and  $\mathbf{W}_{\tau_0, n}^*(\cdot)$  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  $\mathbf{S}^{-1} = \text{diag}(1, \mu_2^{-1}) \otimes \boldsymbol{\Omega}_0^{-1}$ .

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

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

It suffices to consider the weak convergence of  $\mathbf{S}_\tau^{-1} \mathbf{Z}_n(\tau h)$  and  $\mathbf{S}_\tau^{-1} \mathbf{Z}_n^*(\tau h)$  (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  $\text{cov} \left( \mathbf{S}_{\tau_1}^{-1} \mathbf{Z}_n(\tau_1 h), \mathbf{S}_{\tau_2}^{-1} \mathbf{Z}_n(\tau_2 h) \right) = \mathbf{S}_{\tau_1}^{-1} (\text{cov}(\mathbf{Z}_n(\tau_1 h), \mathbf{Z}_n(\tau_2 h))) \mathbf{S}_{\tau_2}^{-1}$  with  $\text{cov}(\mathbf{Z}_n(\tau_1 h), \mathbf{Z}_n(\tau_2 h)) = \sum_{k=0}^2 \boldsymbol{\Gamma}_{k, n}(\tau_1, \tau_2; \tau_0 = 0)$ , 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  $\lim_{n \rightarrow \infty} \text{cov}(\mathbf{Z}_n(\tau_1 h), \mathbf{Z}_n(\tau_2 h)) = \boldsymbol{\kappa}_+(\tau_1, \tau_2) \otimes \mathbf{A}$ . Similarly, we have the bootstrap counterpart  $\text{cov}^*(\mathbf{Z}_n^*(\tau_1 h), \mathbf{Z}_n^*(\tau_2 h)) = \boldsymbol{\kappa}_+(\tau_1, \tau_2) \otimes \mathbf{A} + o_p(1)$ . Using  $\mathbf{S}_\tau^{-1} = \boldsymbol{\mu}_\tau^{-1} \otimes \boldsymbol{\Omega}_0^{-1}$ , we obtain Part (ii).  $\blacksquare$

*Proof of Proposition 1* Under  $H_0$ , we note that  $\sqrt{n}(\widehat{\mathbf{c}} - \boldsymbol{\beta}(\tau_0 + \tau h)) = \sqrt{n}(\widehat{\mathbf{c}} - \mathbf{c}) = O_p(1)$ , uniformly over  $\tau \in [-1, 1]$ , by a CLT for strictly stationary  $\alpha$ -mixing processes, see e.g. Corollary 5.1

in Hall and Heyde (1980). Then by Theorem 2, we have

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

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

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

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

$$\sqrt{nh}(\hat{\beta}(\tau) - \hat{c}) = \sqrt{nh}(\beta(\tau) - \hat{c}) + O_p(1) \xrightarrow{p} \infty, \quad \forall \tau \in U_\delta(\tau_0).$$

The proof is completed. ■

## Appendix B Additional simulation results

### B.1 The GARCH(1,1) specification

In order to introduce conditional heteroskedasticity, we consider GARCH(1,1) errors.

$$u_t = \sigma_t \epsilon_t \tag{B.1}$$

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

where  $\epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ . We let  $\omega = 1 - \alpha_1 - \alpha_2$  and consider three different  $(\alpha_1, \alpha_2)$  combinations inspired by the simulation study in Ling et al. (2003): (0.2, 0.7), (0.3, 0.6) and (0.4, 0.5). We additionally compare the sieve and the sieve wild bootstrap with the autoregressive wild bootstrap (AWB) as proposed by Friedrich et al. (2020) with an AR parameter of  $\gamma = 0.2$ . We see in Tables 9 through 11 that the performance of the SB is similar to that of the ARMA specification. In general, the sieve bootstrap outperforms the sieve wild version (SWB) for all considered specifications. It also outperforms the AWB when it comes to simultaneous coverage. The AWB outperforms the sieve wild. We observe that coverage of the wild bootstrap methods, in particular the sieve wild, is surprisingly low for simultaneous confidence bands. The sieve bootstrap intervals are always wider than their wild bootstrap counterparts offering a potential explanation for the lower coverage. The AWB intervals are comparable in length to the SB intervals. In all considered cases, we see only minor differences between the three chosen GARCH parameter specifications in both length and coverage. The lower performance of the SB in this specification compared to the previous one is an expected result given that the method is not designed for heteroskedastic data. The fact that it nevertheless outperforms its wild extension is surprising. The confidence intervals constructed with the latter method are too narrow resulting in coverage probabilities that are too far below the nominal level. To further investigate the validity of the sieve wild and dependent wild bootstrap

		$h = 0.04$			$h = 0.06$			$h = 0.08$		
$(\alpha_1, \alpha_2)$		SB	SWB	AWB	SB	SWB	AWB	SB	SWB	AWB
$\beta_1$	(0.2,0.7)	0.954 (0.302)	0.895 (0.273)	0.936 (0.298)	0.938 (0.244)	0.883 (0.229)	0.927 (0.253)	0.890 (0.211)	0.838 (0.204)	0.887 (0.225)
	(0.3,0.6)	0.952 (0.296)	0.897 (0.261)	0.939 (0.286)	0.937 (0.239)	0.883 (0.220)	0.931 (0.246)	0.885 (0.207)	0.836 (0.196)	0.888 (0.219)
	(0.4,0.5)	0.952 (0.288)	0.900 (0.249)	0.932 (0.310)	0.937 (0.233)	0.885 (0.211)	0.933 (0.237)	0.882 (0.203)	0.834 (0.189)	0.888 (0.227)
$\beta_2$	(0.2,0.7)	0.953 (0.302)	0.899 (0.251)	0.946 (0.271)	0.953 (0.243)	0.902 (0.210)	0.946 (0.228)	0.951 (0.210)	0.900 (0.186)	0.942 (0.201)
	(0.3,0.6)	0.951 (0.295)	0.900 (0.238)	0.949 (0.256)	0.952 (0.238)	0.902 (0.200)	0.950 (0.219)	0.950 (0.206)	0.900 (0.178)	0.946 (0.194)
	(0.4,0.5)	0.951 (0.289)	0.904 (0.226)	0.941 (0.287)	0.953 (0.233)	0.907 (0.190)	0.954 (0.209)	0.951 (0.202)	0.903 (0.169)	0.941 (0.204)

Table 9: Pointwise coverage probabilities, GARCH(1,1)

methods is beyond the scope of this paper but will be left as a promising topic for future research.

## B.2 Bandwidth selection methods

Given that the previous results show a dependence of the bootstrap’s performance on the bandwidth, we look at the data-driven bandwidth selection methods. We generate 1000 samples from each of the following four error specifications: white noise, AR(1) with  $\phi = 0.3$  and  $\phi = -0.3$ . For each draw, we select the optimal bandwidth chosen by the four different methods. We apply Leave- $(2l + 1)$ -out Cross Validation with  $l = 2, 4, 6$ . The set of possible bandwidths to select from in each case ranges from 0.02 to 0.12 in steps of 0.001. The average selected bandwidths as well as the standard deviations are reported in Table 12. The chosen bandwidths often lie close to the largest value we considered in our simulation study. For  $\beta_2$  this value resulted in accurate coverage. For  $\beta_1$  a smaller bandwidth parameter produced better coverage. Although a bandwidth matrix consisting of different bandwidths for each parameter curve is sometimes considered in the theoretical literature, in practice, a data driven bandwidth selection method to determine such a matrix is computationally cumbersome. Combining these results suggests to use a smaller bandwidth in practice than the one selected by the data-driven methods. While a smaller bandwidth might not produce an optimal approximation to coefficient curves with low variations, it nevertheless led to good coverage for such cases in our simulation study. For more difficult to estimate coefficient curves, a small bandwidth is preferable and it produced superior results when looking at coverage. This shows that blindly relying on data-driven bandwidth selection methods can be problematic in practice. We recommend, in addition to such selection methods, to perform a robustness analysis of the main results using different bandwidths.

		$h = 0.04$			$h = 0.06$			$h = 0.08$		
$(\alpha_1, \alpha_2)$		SB	SWB	AWB	SB	SWB	AWB	SB	SWB	AWB
$\beta_1$	(0.2,0.7)	0.912 (0.397)	0.790 (0.354)	0.849 (0.384)	0.882 (0.321)	0.740 (0.299)	0.809 (0.330)	0.745 (0.280)	0.636 (0.267)	0.720 (0.296)
	(0.3,0.6)	0.901 (0.390)	0.787 (0.338)	0.860 (0.372)	0.871 (0.316)	0.747 (0.286)	0.822 (0.309)	0.733 (0.275)	0.624 (0.257)	0.726 (0.289)
	(0.4,0.5)	0.895 (0.383)	0.801 (0.323)	0.826 (0.402)	0.871 (0.310)	0.756 (0.275)	0.821 (0.295)	0.713 (0.270)	0.613 (0.247)	0.736 (0.299)
$\beta_2$	(0.2,0.7)	0.912 (0.396)	0.829 (0.327)	0.917 (0.387)	0.909 (0.320)	0.818 (0.274)	0.900 (0.298)	0.901 (0.279)	0.787 (0.245)	0.896 (0.265)
	(0.3,0.6)	0.911 (0.390)	0.829 (0.310)	0.929 (0.333)	0.904 (0.315)	0.804 (0.261)	0.909 (0.321)	0.895 (0.274)	0.778 (0.234)	0.902 (0.256)
	(0.4,0.5)	0.903 (0.383)	0.842 (0.294)	0.908 (0.374)	0.904 (0.310)	0.811 (0.248)	0.918 (0.309)	0.894 (0.269)	0.785 (0.223)	0.893 (0.269)

Table 10: Coverage probabilities for  $G_{sub}$ , GARCH(1,1)

		$h = 0.04$			$h = 0.06$			$h = 0.08$		
$(\alpha_1, \alpha_2)$		SB	SWB	AWB	SB	SWB	AWB	SB	SWB	AWB
$\beta_1$	(0.2,0.7)	0.827 (0.457)	0.376 (0.409)	0.692 (0.448)	0.848 (0.364)	0.462 (0.339)	0.736 (0.374)	0.739 (0.309)	0.435 (0.295)	0.677 (0.325)
	(0.3,0.6)	0.805 (0.450)	0.387 (0.391)	0.706 (0.430)	0.825 (0.359)	0.464 (0.326)	0.756 (0.363)	0.732 (0.304)	0.426 (0.284)	0.674 (0.317)
	(0.4,0.5)	0.795 (0.442)	0.376 (0.373)	0.655 (0.466)	0.799 (0.353)	0.462 (0.312)	0.769 (0.350)	0.703 (0.299)	0.416 (0.273)	0.692 (0.325)
$\beta_2$	(0.2,0.7)	0.841 (0.457)	0.441 (0.377)	0.797 (0.408)	0.875 (0.363)	0.500 (0.312)	0.842 (0.337)	0.876 (0.308)	0.573 (0.270)	0.857 (0.291)
	(0.3,0.6)	0.811 (0.450)	0.445 (0.358)	0.811 (0.385)	0.864 (0.358)	0.512 (0.297)	0.850 (0.324)	0.870 (0.303)	0.573 (0.258)	0.870 (0.281)
	(0.4,0.5)	0.806 (0.442)	0.440 (0.339)	0.790 (0.432)	0.863 (0.352)	0.512 (0.282)	0.870 (0.308)	0.865 (0.297)	0.575 (0.246)	0.863 (0.296)

Table 11: Coverage probabilities for the whole sample, GARCH(1,1)

Bandwidth selection						
Method	$(\phi, \psi) = (0, 0)$		$\phi = 0.3$		$\phi = -0.3$	
	$\mu$	$(\sigma)$	$\mu$	$(\sigma)$	$\mu$	$(\sigma)$
CV	0.085	(0.024)	0.083	(0.023)	0.084	(0.022)
GCV	0.076	(0.028)	0.074	(0.028)	0.076	(0.027)
AIC	0.095	(0.018)	0.093	(0.017)	0.094	(0.017)
L2O	0.092	(0.021)	0.090	(0.022)	0.091	(0.020)
L4O	0.099	(0.019)	0.096	(0.020)	0.096	(0.019)
L6O	0.103	(0.017)	0.101	(0.018)	0.101	(0.017)

Table 12: Mean ( $\mu$ ) and stand deviation ( $\sigma$ ) of selected bandwidths

SUPPLEMENTAL APPENDIX TO:  
SIEVE BOOTSTRAP INFERENCE FOR TIME-VARYING  
COEFFICIENT MODELS

Marina Friedrich<sup>a,b</sup> and Yicong Lin<sup>b</sup>

<sup>a</sup>Vrije Universiteit Amsterdam

<sup>b</sup>Tinbergen Institute

## S1 Additional Proofs

*Proof of Lemma 1* The lemma is a special case of the well-known uniform result established by Hansen (2008), see Theorem 2. We shall check Assumptions 1-3 of Hansen (2008). We note that  $\bar{K} : x \mapsto x^k K^\ell(x)$  is Lipschitz continuous on  $[-1, 1]$  and has a bounded support, satisfying Assumptions 1 and 3. With the fact that  $\beta > \frac{2s-2}{s-2}$ , and using similar arguments of Kristensen (2009) (p. 1444, Proof of Theorem 2), Assumption 2 is fulfilled. Finally, one can justify Equations (10) and (12) of Hansen (2008) by some straightforward calculations.<sup>1</sup> Overall, we obtain (A.2). ■

*Proof of Lemma 2* By Taylor's expansion and Assumption A2,

$$\beta(\tau_t) = \beta(\tau) + \beta^{(1)}(\tau)(\tau_t - \tau) + \left(\beta^{(2)}(\tau) + O(h)\right)(\tau_t - \tau)^2/2, \quad |\tau_t - \tau| \leq h. \quad (\text{S.1})$$

Since  $K$  has support  $[-1, 1]$ , by (S.1), one can write

$$\begin{aligned} & \left(\mathbf{H}^{-1} \mathbf{S}_n(\tau) \mathbf{H}^{-1}\right) \left[\mathbf{H} \left(\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)\right)\right] \\ &= \frac{h^2}{2} \left( h^{-2} \mathbf{S}_{n,2}(\tau) \left(\beta^{(2)}(\tau) + O(h)\right) + \left( (nh)^{-1} \sum_{t=1}^n \mathbf{x}_t z_t K\left(\frac{\tau_t - \tau}{h}\right) \right) \right) \\ & \quad + \left( h^{-3} \mathbf{S}_{n,3}(\tau) \left(\beta^{(2)}(\tau) + O(h)\right) + \left( (nh)^{-1} \sum_{t=1}^n \mathbf{x}_t z_t \left(\frac{\tau_t - \tau}{h}\right) K\left(\frac{\tau_t - \tau}{h}\right) \right) \right). \end{aligned} \quad (\text{S.2})$$

We first claim that (proofs given below)

$$\sup_{\tau \in [0,1]} \left\| \frac{1}{nh} \sum_{t=1}^n \mathbf{x}_t z_t w_t^k(\tau) \right\| = O_p \left( \sqrt{\frac{\ln n}{nh}} \right), \quad (\text{S.3})$$

$$\sup_{\tau \in [0,1]} \left\| h^{-k} \mathbf{S}_{n,k}(\tau) - \boldsymbol{\Omega}_0 \int_{-\tau/h}^{(1-\tau)/h} w^k K(u) du \right\| = O_p \left( \sqrt{\frac{\ln n}{nh}} \right) + O \left( \frac{1}{nh^2} \right), \quad (\text{S.4})$$

---

<sup>1</sup>Take  $q = \infty$  and  $d = 1$  in the equations in Hansen (2008). Then Equation (10) in Hansen (2008) reduces to  $\beta > \frac{2s-1}{s-2} = \frac{2\delta+3}{\delta}$ . It is satisfied by our Assumption C1. Moreover, Equation (11) in Hansen (2008) reduces to  $\theta = \frac{\beta-2-(1+\beta)/(s-1)}{\beta+2-(1+\beta)/(s-1)} = 1 - \frac{4}{\beta+2-(1+\beta)/(s-1)}$ . Using  $\beta \geq 3(1+\delta)/\delta$  (Assumption C1) and the strict monotonicity of  $\beta \mapsto \beta+2-(1+\beta)/(s-1)$ , we have  $\beta+2-\frac{1+\beta}{s-1} > 4$  and thus  $\theta > 0$ . Therefore Equation (12) of Hansen (2008) is fulfilled.

where  $k = 0, 1, 2, 3$ . By (S.2) - (S.4), it is immediate to obtain (A.3) because, for any  $\tau \in [0, 1]$ ,  $\left| \int_{-\tau/h}^{(1-\tau)/h} u^k K(u) du \right| \leq \int_{-1}^1 |u^k K(u)| du \leq C$ .

It remains to show (S.3) and (S.4). By Lemma 1 with  $Y_t = x_{i,t} z_t$ ,  $i = 0, 1, \dots, d$ , it is immediate to have (S.3) by the norm equivalence. Moreover, by triangle inequality

$$\begin{aligned} & \left\| h^{-k} \mathbf{S}_{n,k}(\tau) - \mathbf{\Omega}_0 \int_{-\tau/h}^{(1-\tau)/h} u^k K(u) du \right\| \\ & \leq \left\| h^{-k} \mathbf{S}_{n,k}(\tau) - \mathbb{E}(h^{-k} \mathbf{S}_{n,k}(\tau)) \right\| + \left\| \mathbb{E}(h^{-k} \mathbf{S}_{n,k}(\tau)) - \mathbf{\Omega}_0 \int_{-\tau/h}^{(1-\tau)/h} u^k K(u) du \right\|. \end{aligned}$$

Similarly, by the norm equivalence, it suffices to consider the individual elements of  $h^{-k} \mathbf{S}_{n,k}(\tau)$ , namely  $(nh)^{-1} \sum_{t=1}^n x_{i,t} x_{k,t} w_t^k(\tau)$ , where  $0 \leq i, k \leq d$ . By Lemma 1 with  $Y_t = x_{i,t} x_{k,t}$ , we have

$$\sup_{\tau \in [0,1]} \left| \frac{1}{nh} \sum_{t=1}^n x_{i,t} x_{k,t} w_t^k(\tau) - \frac{1}{nh} \sum_{t=1}^n \mathbb{E}(x_{i,t} x_{k,t}) w_t^k(\tau) \right| = O_p \left( \sqrt{\frac{\ln n}{nh}} \right) \quad (\text{S.5})$$

It implies  $\sup_{\tau \in [0,1]} \|h^{-k} \mathbf{S}_{n,k}(\tau) - \mathbb{E}(h^{-k} \mathbf{S}_{n,k}(\tau))\| = O_p \left( \sqrt{\frac{\ln n}{nh}} \right)$ . Next, it is well-known that, for instance, see Bühlmann (1998), Equation (6.5),

$$\left| n^{-1} \sum_{t=1}^n g\left(\frac{t}{n}\right) - \int_0^1 g(z) dz \right| \leq \sup_{|x-y| \leq n^{-1}} |g(x) - g(y)|, \quad (\text{S.6})$$

where  $g(\cdot)$  is continuous and Riemann-integrable. Setting  $g : x \mapsto h^{-1} \left(\frac{x-\tau}{h}\right)^k K\left(\frac{x-\tau}{h}\right)$  in (S.6), we obtain

$$\begin{aligned} & \sup_{\tau \in [0,1]} \left\| \mathbb{E}(h^{-k} \mathbf{S}_{n,k}(\tau)) - \mathbf{\Omega}_0 \int_{-\tau/h}^{(1-\tau)/h} u^k K(u) du \right\| \\ & \leq C \sup_{\tau \in [0,1]} \sup_{|x-y| \leq n^{-1}} h^{-1} \left| \left(\frac{x-\tau}{h}\right)^k K\left(\frac{x-\tau}{h}\right) - \left(\frac{y-\tau}{h}\right)^k K\left(\frac{y-\tau}{h}\right) \right| \\ & \leq C \sup_{\tau \in [0,1]} \sup_{|x-y| \leq n^{-1}} h^{-1} \frac{|x-y|}{h} \leq \frac{C}{nh^2}, \end{aligned} \quad (\text{S.7})$$

where (S.7) is due to the Lipschitz continuity of  $x \mapsto x^k K(x)$  on a compact set. Equation (S.4) is now obtained, and thus the lemma follows.  $\blacksquare$

*Proof of Lemma 3* It follows from Corollary A.2 in Hall and Heyde (1980).  $\blacksquare$

*Proof of Lemma 4* Note that Assumption A1 implies Assumption C1. By Lemma 2,

$$n^{-1} \sum_{t=1}^n \left( \mathbf{x}'_t \tilde{\boldsymbol{\beta}}(\tau_t) - \mathbf{x}'_t \boldsymbol{\beta}(\tau_t) \right)^2 \leq \sup_{\tau \in [0,1]} \left\| \tilde{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau) \right\|^2 n^{-1} \sum_{t=1}^n \|\mathbf{x}_t\|^2 = O_p(b(n)),$$

where  $b(n) = \left( \tilde{h}^2 + \sqrt{\frac{\ln n}{nh}} \right)^2$ . Then Parts (i), (ii) and (iii) are direct results of Lemmas 6.3 and 6.4 of Bühlmann (1998) provided  $p(n) = o\left(\min\left\{\left(\frac{n}{\ln n}\right)^{1/4}, \left(\frac{1}{b(n)}\right)^{1/4}\right\}\right)$ . Because  $\frac{1}{b(n)} < \frac{n}{\ln n}$  when  $n$  is

sufficiently large, the condition is clearly satisfied under Assumption B2. Next we consider Part (iv). By the representation in (A.4), the results in Parts (i), (ii) and (iii) lead to

$$\begin{aligned}\mathbb{E}^* \left( z_t^* z_{t+j}^* \right) &= \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \widehat{\psi}_{k,n} \widehat{\psi}_{\ell,n} \mathbb{E}^* \left( \varepsilon_{t-k}^* \varepsilon_{t+j-\ell}^* \right) \\ &= \left( \sum_{k=0}^{\infty} \widehat{\psi}_{j+k,n} \widehat{\psi}_{k,n} \right) \mathbb{E}^* |\varepsilon_t^*|^2 = \left( \sum_{k=0}^{\infty} \psi_{j+k} \psi_k \right) \mathbb{E} |\varepsilon_t|^2 + o_p(1).\end{aligned}$$

Since  $\mathbb{E}(z_t z_{t+j}) = (\sum_{k=0}^{\infty} \psi_{j+k} \psi_k) \mathbb{E} |\varepsilon_t|^2$ , we have shown (iv). Using Parts (i), (iii) and the identity  $\mathbb{E}^* \left( z_t^* z_{t+j}^* \right) = \left( \sum_{k=0}^{\infty} \widehat{\psi}_{j+k,n} \widehat{\psi}_{k,n} \right) \mathbb{E}^* |\varepsilon_t^*|^2$ , it is straightforward to obtain Part (v).

Recall  $w_t^k(\tau) = \left(\frac{\tau_t - \tau}{h}\right)^k K\left(\frac{\tau_t - \tau}{h}\right)$  and write the covariance matrix in Part (vi) as

$$\text{cov}^* \left( \mathbf{Z}_{n,k_1}^*(\tau), \mathbf{Z}_{n,k_2}^*(\tau) \right) = \frac{1}{nh} \sum_{t=1}^n \sum_{s=1}^n \mathbb{E}^* \left( z_t^* z_s^* \right) \mathbf{x}_t \mathbf{x}_s' w_t^{k_1}(\tau) w_s^{k_2}(\tau) =: \mathbf{\Gamma}_{0,n}^*(\tau) + \mathbf{\Gamma}_{1,n}^*(\tau) + \mathbf{\Gamma}_{2,n}^*(\tau), \quad (\text{S.8})$$

where

$$\begin{aligned}\mathbf{\Gamma}_{0,n}^*(\tau) &= \frac{1}{nh} \sum_{t=1}^n \mathbb{E}^* \left( z_t^{*2} \right) \mathbf{x}_t \mathbf{x}_t' \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right), \\ \mathbf{\Gamma}_{1,n}^*(\tau) &= \frac{1}{nh} \sum_{t=1}^{n-1} \sum_{s=t+1}^n \mathbb{E}^* \left( z_t^* z_s^* \right) \mathbf{x}_t \mathbf{x}_s' w_t^{k_1}(\tau) w_s^{k_2}(\tau), \\ \mathbf{\Gamma}_{2,n}^*(\tau) &= \frac{1}{nh} \sum_{t=2}^n \sum_{s=1}^{t-1} \mathbb{E}^* \left( z_t^* z_s^* \right) \mathbf{x}_t \mathbf{x}_s' w_t^{k_1}(\tau) w_s^{k_2}(\tau) = (nh)^{-1} \sum_{t=1}^{n-1} \sum_{s=t+1}^n \mathbb{E}^* \left( z_t^* z_s^* \right) \mathbf{x}_s \mathbf{x}_t' w_s^{k_1}(\tau) w_t^{k_2}(\tau).\end{aligned}$$

Using Part (iv), Lemma 1 (with  $k = k_1 + k_2$ ,  $\ell = 2$ ), the Riemann sum approximation (S.6), and the Lipschitz continuity of  $x \mapsto x^k K^2(x)$ , we have

$$\begin{aligned}\mathbf{\Gamma}_{0,n}^*(\tau) &= \left( \mathbb{E}(z_1^2) + o_p(1) \right) \frac{1}{nh} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \\ &= \left( \mathbb{E}(z_1^2) + o_p(1) \right) \left[ \mathbf{\Omega}_0 \frac{1}{nh} \sum_{t=1}^n \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) + O_p \left( \sqrt{\frac{\ln n}{nh}} \right) \right] \\ &= \left( \mathbb{E}(z_1^2) + o_p(1) \right) \left[ \mathbf{\Omega}_0 \int_{-\tau/h}^{(1-\tau)/h} u^{k_1+k_2} K^2(u) du + O \left( \frac{1}{nh^2} \right) + O_p \left( \sqrt{\frac{\ln n}{nh}} \right) \right] \\ &= \mathbb{E}(z_1^2) \mathbf{\Omega}_0 \left( \nu_{k_1+k_2} \mathbb{1}_{\{\tau \in [\tau_L, \tau_U]\}} + \nu_{k_1+k_2,c} \mathbb{1}_{\{\tau = ch\}} \right) + o_p(1), \quad (\text{S.9})\end{aligned}$$

for any  $\tau_L, \tau_U \in (0, 1)$ . For  $\mathbf{\Gamma}_{1,n}^*(\tau)$ , note that  $\sum_{t=1}^{n-1} \sum_{s=t+1}^n q_{t,s} \mathbf{a}_t \mathbf{a}_s' = \sum_{j=1}^{n-1} \sum_{t=1}^{n-j} q_{t,t+j} \mathbf{a}_t \mathbf{a}_{t+j}'$  for



any vectors  $\mathbf{a}_t$  and  $q_{t,s} \in \mathbb{R}$ . Then,

$$\begin{aligned}
\mathbf{\Gamma}_{1,n}^*(\tau) &= \sum_{j=1}^{n-1} \mathbb{E}^* \left( z_1^* z_{1+j}^* \right) \left[ \frac{1}{nh} \sum_{t=1}^{n-j} \mathbf{x}_t \mathbf{x}'_{t+j} w_t^{k_1}(\tau) w_{t+j}^{k_2}(\tau) \right] \\
&= \sum_{j=1}^{n-1} \mathbb{E}^* \left( z_1^* z_{1+j}^* \right) \left[ \frac{1}{nh} \sum_{t=1}^{n-j} \mathbf{x}_t \mathbf{x}'_{t+j} \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) + O_p \left( \frac{j}{nh^2} \right) \right] \\
&= \sum_{j=1}^{n-1} \mathbb{E} (z_1 z_{1+j}) \left[ \frac{1}{nh} \sum_{t=1}^{n-j} \mathbf{x}_t \mathbf{x}'_{t+j} \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \right] + O_p \left( \frac{1}{nh^2} \right) + o_p(1) \\
&= \sum_{j=1}^{n-1} \mathbb{E} (z_1 z_{1+j}) \mathbb{E} \left( \mathbf{x}_1 \mathbf{x}'_{1+j} \right) \left[ \frac{1}{nh} \sum_{t=1}^{n-j} \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \right] + o_p(1), \quad (\text{S.10})
\end{aligned}$$

where the second equality is due to

$$\left\| \frac{1}{nh} \sum_{t=1}^{n-j} \mathbf{x}_t \mathbf{x}'_{t+j} w_t^{k_1}(\tau) \left( w_{t+j}^{k_2}(\tau) - w_t^{k_2}(\tau) \right) \right\| \leq \frac{1}{nh} \sum_{t=1}^{n-j} \|\mathbf{x}_t \mathbf{x}'_{t+j}\| |w_t^{k_1}(\tau)| |w_{t+j}^{k_2}(\tau) - w_t^{k_2}(\tau)| = O_p \left( \frac{j}{nh^2} \right),$$

the third one is by Parts (iv) and (v), and the final one is by Lemma 1 with the fact that  $\{\mathbf{x}_t \mathbf{x}'_{t+j}\}$  satisfying Assumption C1 for every  $j \geq 0$ . Now we consider the first term in (S.10). As such, we take  $H_n$  that satisfies  $\frac{1}{H_n} + \frac{H_n}{nh} \rightarrow 0$  as  $n \rightarrow \infty$ . Now we split the summation  $\sum_{j=1}^{n-1}$  into  $\sum_{j=1}^{n-1} = \sum_{j=1}^{H_n-1} + \sum_{j=H_n}^{n-1}$ . For  $j \in \{1, 2, \dots, H_n - 1\}$ , we have  $\left| (nh)^{-1} \sum_{t=n-j+1}^n \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \right| \leq C \frac{j}{nh} \leq C \frac{H_n}{nh}$ . Then the Riemann sum approximation implies

$$\begin{aligned}
&\frac{1}{nh} \sum_{t=1}^{n-j} \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \\
&= \frac{1}{nh} \sum_{t=1}^n \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) + \frac{1}{nh} \sum_{t=n-j+1}^n \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \\
&= \frac{1}{nh} \sum_{t=1}^n \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) + O \left( \frac{H_n}{nh} \right) \\
&= \nu_{k_1+k_2} \mathbb{1}_{\{\tau \in [\tau_L, \tau_U]\}} + \nu_{k_1+k_2,c} \mathbb{1}_{\{\tau = ch\}} + O \left( \frac{1}{nh^2} \right) + O \left( \frac{H_n}{nh} \right),
\end{aligned}$$

where the the  $O$ -terms are uniform in  $j$ , and thus

$$\begin{aligned}
&\sum_{j=1}^{H_n-1} \mathbb{E} (z_1 z_{1+j}) \mathbb{E} \left( \mathbf{x}_1 \mathbf{x}'_{1+j} \right) \left[ \frac{1}{nh} \sum_{t=1}^{n-j} \left( \frac{\tau_t - \tau}{h} \right)^{k_1+k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \right] \\
&\quad \xrightarrow{n \rightarrow \infty} \sum_{j=1}^{\infty} \mathbb{E} (z_1 z_{1+j}) \mathbb{E} \left( \mathbf{x}_1 \mathbf{x}'_{1+j} \right) \left( \nu_{k_1+k_2} \mathbb{1}_{\{\tau \in [\tau_L, \tau_U]\}} + \nu_{k_1+k_2,c} \mathbb{1}_{\{\tau = ch\}} \right). \quad (\text{S.11})
\end{aligned}$$

Finally, by Lemma 3 and  $\left| (nh)^{-1} \sum_{t=1}^{n-j} \left( \frac{\tau_t - \tau}{h} \right)^{k_1 + k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \right| \leq C$ ,

$$\begin{aligned} & \left\| \sum_{j=H_n}^{n-1} \mathbb{E}(z_1 z_{1+j}) \mathbb{E}(\mathbf{x}_1 \mathbf{x}'_{1+j}) \frac{1}{nh} \sum_{t=1}^{n-j} \left( \frac{\tau_t - \tau}{h} \right)^{k_1 + k_2} K^2 \left( \frac{\tau_t - \tau}{h} \right) \right\| \\ & \leq C \sum_{j=H_n}^{\infty} |\mathbb{E}(z_1 z_{1+j})| \left\| \mathbb{E}(\mathbf{x}_1 \mathbf{x}'_{1+j}) \right\| = o(1). \quad (\text{S.12}) \end{aligned}$$

Putting (S.11) (S.12) into (S.10), we find, for any  $\tau_L, \tau_U \in (0, 1)$ ,

$$\mathbf{I}_{1,n}^*(\tau) \xrightarrow{n \rightarrow \infty} \mathbf{I}_{1,\infty}^*(\tau) := \sum_{j=1}^{\infty} \mathbb{E}(z_1 z_{1+j}) \mathbb{E}(\mathbf{x}_1 \mathbf{x}'_{1+j}) \left( \nu_{k_1 + k_2} \mathbb{1}_{\{\tau \in [\tau_L, \tau_U]\}} + \nu_{k_1 + k_2, c} \mathbb{1}_{\{\tau = ch\}} \right), \quad (\text{S.13})$$

in probability. Similarly, one can find  $\mathbf{I}_{2,n}^*(\tau) \xrightarrow{n \rightarrow \infty} \mathbf{I}_{1,\infty}^{*'}(\tau)$ . By Assumption A1(e), we have  $\text{cov}(\mathbf{x}_t z_t, \mathbf{x}_{t+j} z_{t+j}) = \mathbb{E}(z_t z_{t+j} \mathbf{x}_t \mathbf{x}'_{t+j}) = \mathbb{E}(z_t z_{t+j}) \mathbb{E}(\mathbf{x}_t \mathbf{x}'_{t+j})$ . Hence, using (S.8), (S.9) and (S.13), we have shown Part (vi).  $\blacksquare$

*Proof of Lemma 5* Consider Part (i) first. Define  $\Delta_t^k(\tau_1, \tau_2) = w_t^k(\tau_0 + \tau_1 h) - w_t^k(\tau_0 + \tau_2 h)$ . We make two observations first: (i) By the Lipschitz continuity of  $x \mapsto x^k K(x)$  on  $[-1, 1]$ , we have  $|\Delta_t^k(\tau_1, \tau_2)| \leq C|\tau_1 - \tau_2|$ ; (ii) Since  $K(\cdot)$  has the support  $[-1, 1]$ , we have  $w_t^k(\tau_0 + \tau h) = 0$  whenever  $|\tau_t - (\tau_0 + \tau h)| > h$  for any  $\tau \in [-1, 1]$  and  $\tau_0 \in (0, 1)$ . Note that  $\{t : |\tau_t - (\tau_0 + \tau h)| \leq h\} \subset \{t : \lfloor n(\tau_0 - 2h) \rfloor \leq t \leq \lceil n(\tau_0 + 2h) \rceil\} =: \mathcal{I}_{n,h,\tau_0}$ . Then  $\Delta_t^k(\tau_1, \tau_2) = 0$  whenever  $t \notin \mathcal{I}_{n,h,\tau_0}$ . Further note that the cardinality of  $\mathcal{I}_{n,h,\tau_0}$  is bounded by  $|\mathcal{I}_{n,h,\tau_0}| \leq Cnh$ . Using these properties, the LHS of (A.5) can be written as

$$\begin{aligned} & \frac{1}{nh} \sum_{\ell=-n+1}^{n-1} \sum_{t=1}^{n-|\ell|} \mathbb{E} \left( x_{i,t} x_{i,t+|\ell|} z_t z_{t+|\ell|} \right) \Delta_t^k(\tau_1, \tau_2) \Delta_{t+|\ell|}^k(\tau_1, \tau_2) \\ & \leq \sum_{\ell=-n+1}^{n-1} \left| \mathbb{E} \left( x_{i,t} x_{i,t+|\ell|} z_t z_{t+|\ell|} \right) \right| \frac{1}{nh} \sum_{t \in \mathcal{I}_{n,h,\tau_0}} \left| \Delta_t^k(\tau_1, \tau_2) \Delta_{t+|\ell|}^k(\tau_1, \tau_2) \right| \leq C|\tau_1 - \tau_2|^2, \end{aligned}$$

where we use that  $\sum_{\ell=-\infty}^{\infty} \left| \mathbb{E} \left( x_{i,t} x_{i,t+|\ell|} z_t z_{t+|\ell|} \right) \right| \leq C$  by Lemma 3. Using Lemma 4(iv), (A.6) is similarly obtained. Part (ii) is similar.  $\blacksquare$

## S2 Additional empirical results

We obtain daily mean temperature series for seven cities, located in seven different countries. They are spread out over Europe: Berlin, Budapest, De Bilt, Dublin, Lyon, Madrid and Stockholm. We take the average over all cities as our temperature series. The resulting series is displayed in Figure 1(a). We remove the seasonal component with the help of sine and cosine functions. More specifically, we fit the following regression

$$\text{Temp}_t = \alpha_1 \cos(2\pi t) + \alpha_2 \sin(2\pi t) + \epsilon_t \quad (\text{S.1})$$

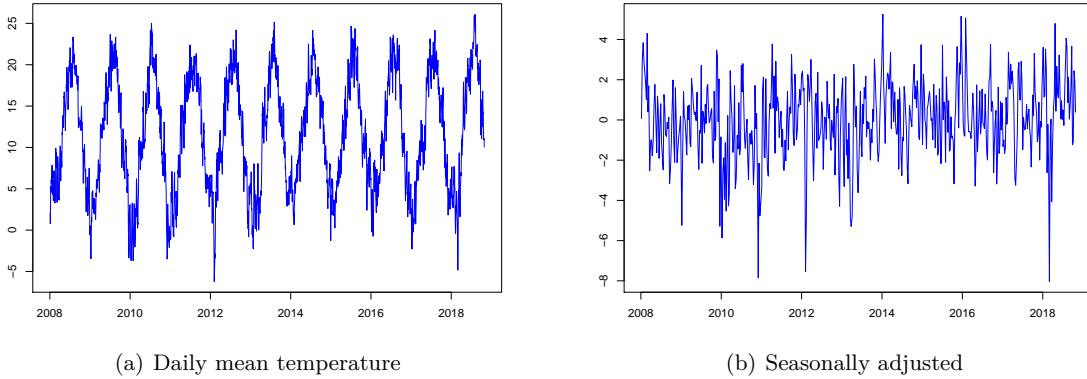


Figure 1: Temperature data before and after the removal of the seasonal component.

and continue to work with the residuals from this regression which are plotted in Figure 1(b).<sup>2</sup>

To determine if our data is stationary, we perform the Augmented Dickey-Fuller (ADF) test (Said and Dickey, 1984) and the tests by Phillips and Perron (1988), Kwiatkowski et al. (1992) and Leybourne et al. (1998). The latter test is similar to the ADF test, but it considers a smoothly varying time trend under the alternative hypothesis. The trend can undergo one transition and the time point as well as speed of the transition is determined endogenously by the test. This modification allows for much more flexibility under the alternative which is an advantage given our complex data series. Table S1 gives the test statistics for all series in levels ( $y_t$ ) as well as the log return data, which is defined as the first difference of the natural logarithm ( $r_{y_t} = \ln(y_t/y_{t-1})$ ). The bottom part of the table shows the critical values for significance levels of 1%, 5% and 10%. Apart from the KPSS test, all tests are left-sided, and hence, we reject the null hypothesis, if the test

<sup>2</sup>As the method requires stationary data, removing the seasonality prior to the regression analysis is important to avoid spurious results. However, in this case part of the link between the temperature series and EUA prices might be lost. Therefore, it could be worthwhile to investigate ways to explicitly incorporate the seasonal component into the model or to construct an indicator for extremely cold periods. This is left for future work.

Unit root tests									
	ADF test		PP test		LNV test		KPSS test		
	$y_t$	$r_{y_t}$	$y_t$	$r_{y_t}$	$y_t$	$r_{y_t}$	$y_t$	$r_{y_t}$	
EUA	-0.562	-14.132	-0.675	-17.641	-1.011	-14.336	1.306	0.072	
Coal	-2.236	-12.879	-2.089	-16.488	-2.336	-13.051	0.553	0.056	
Gas	-1.999	-13.586	-2.108	-18.678	-2.029	13.723	0.574	0.068	
Oil	-1.780	-13.882	-1.761	-18.593	-2.033	-13.958	1.130	0.084	
S <sub>50</sub>	-3.419	-15.965	-4.339	-19.111	-3.705	-16.210	0.739	0.102	
S <sub>600</sub>	-3.452	-15.901	-4.045	-19.868	-3.495	-16.301	0.452	0.071	
Temp	-13.054	-	-15.026	-	-13.147	-	0.070	-	
Critical values: (90%, 95%, 99%)									
(-3.13, -3.42, -3.98)			(-3.13, -3.42, -3.98)			(-4.55, -4.83, -5.42)		(0.12, 0.14, 0.22)	

Table S1: Results from the unit root tests of Said and Dickey (1984), Phillips and Perron (1988), Leybourne et al. (1998) and Kwiatkowski et al. (1992)

statistic is smaller than the critical value. According to the ADF test, the unit root null hypothesis cannot be rejected at a 1% significance level in all cases but the temperature data. This holds true for the data in levels ( $y_t$ ). The only series for which a unit root would be rejected in favor of stationarity are the two stock indices at a 5% and 10% level, respectively. The log return series ( $r_{y_t}$ ) are all stationary. The Phillips-Perron (PP) test results confirm the results from the ADF test with the exception being that the unit root null hypothesis is now rejected for the two stock indices at the 1% significance level. Results of the more flexible test by Leybourne et al. (1998) (LNV) indicate that all series but the temperature data contain a unit root. The KPSS test comes to the same conclusion. Combining the above results with the ADF results, we conclude that allowance prices as well as all fuel prices and stock indices contain a unit root, the temperature data are stationary.

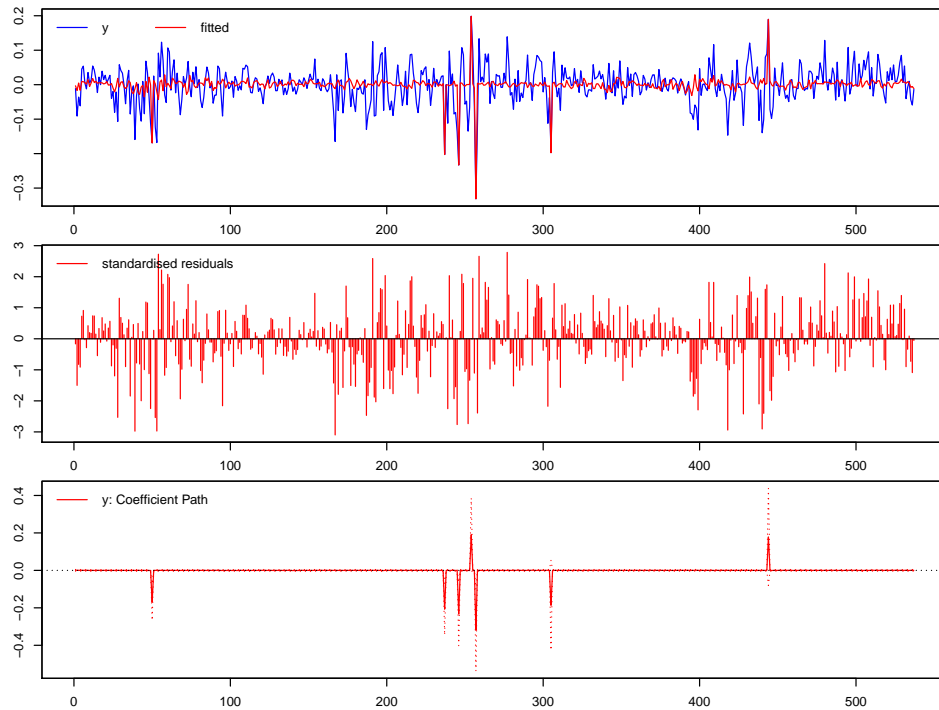


Figure 2: Summary of impulse indicator saturation results. Figure produced with R package *gets*. Top: EUA returns (blue) and fit of dummy regression (red); middle: standardized residuals after regression on dummies; bottom: retained dummy variables

Next, we discuss the results from the IIS method. It retains 10 dummy variables corresponding to 10 outliers located at observations 39, 50, 53, 167, 237, 246, 254, 257, 305 and 444. Corresponding to time points in October 2008, January and February 2009, June 2011, November 2012, January 2013, March and April 2013, March 2014 and December 2016. If we control for the impact of our most important explanatory variables – coal, gas, oil prices and temperature data – three of these outliers can be explained. Figure 2 depicts the results of this IIS application. With this approach we retain 7 outliers which we delete from the EUA series and subsequently perform our analysis. In Figure 3 we repeat the nonparametric regression exercise given in the main text without first removing the outliers from the data. In some cases, the confidence intervals get temporarily wider compared to Figure 3 of the main paper. This holds in particular for the gas and the oil price coefficient around 2013/2014. However, the main conclusions drawn in this paper do not change.

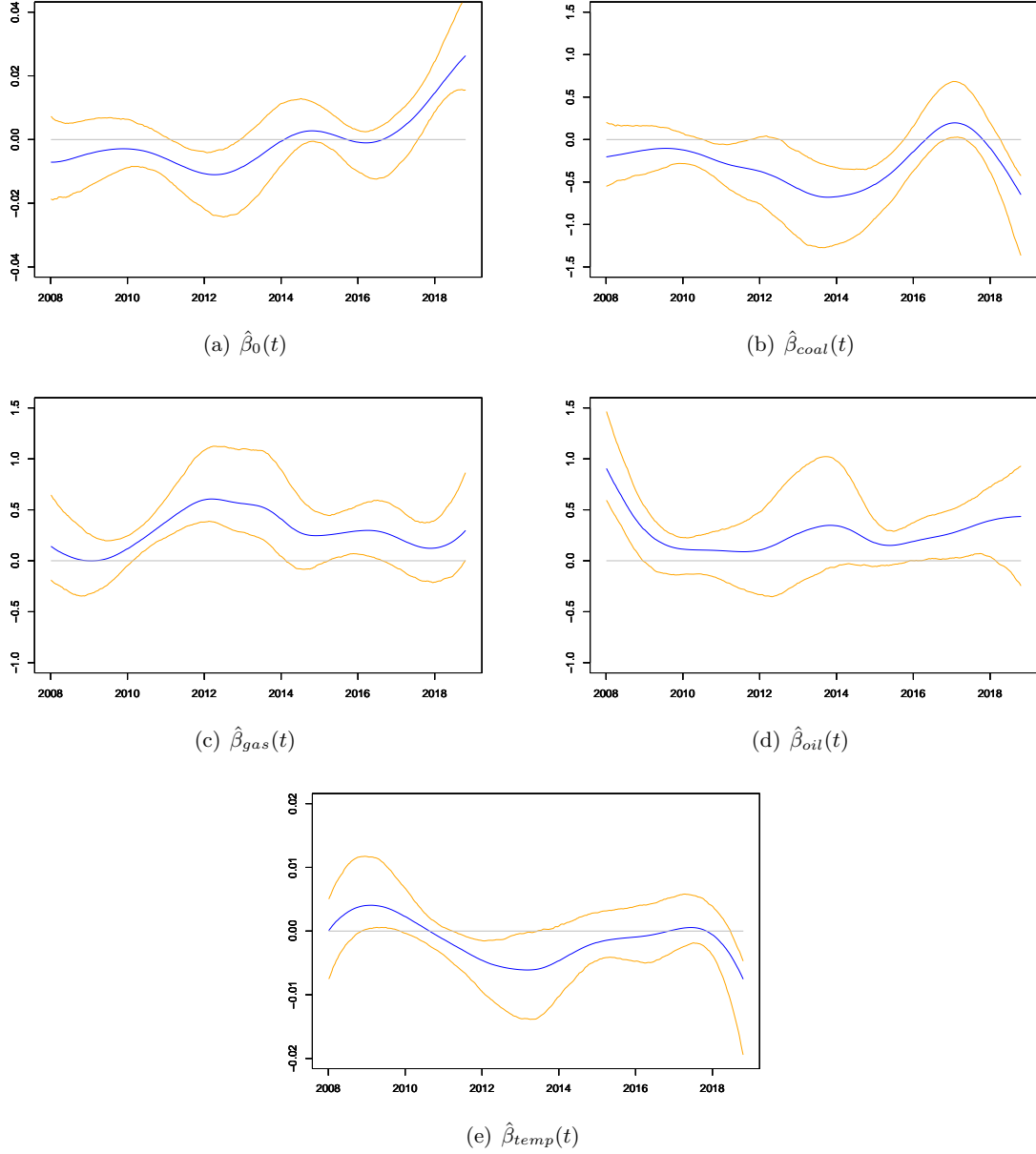


Figure 3: Nonparametrically estimated coefficient curves and 95% confidence intervals before the removal of outliers

We additionally include two explanatory variables to our nonparametric regression: energy supply from hydro power in Norway as well as data on electricity generation from wind for Germany. Both variables should have a negative effect on allowance prices. The hydro power data are weekly data which contain a strong seasonal component which is, as for the temperature data, removed with the help of Fourier terms. The wind generation data is only available until the end of May 2018 which reduces our sample size which is now 517. Added to the nonparametric regression, both new regressors have a coefficient estimate which is extremely low in magnitude for the whole sample. Both estimated coefficient curves are plotted in Figure 4. From Panel (a) we see that hydro power is significant over a very brief period in 2016. Panel (b) shows a period of significance for wind at the

beginning of the sample until 2010. Although this period is quite long, given the small magnitude of the coefficient (in the order of  $10^{-5}$ ), we consider this effect as negligible.

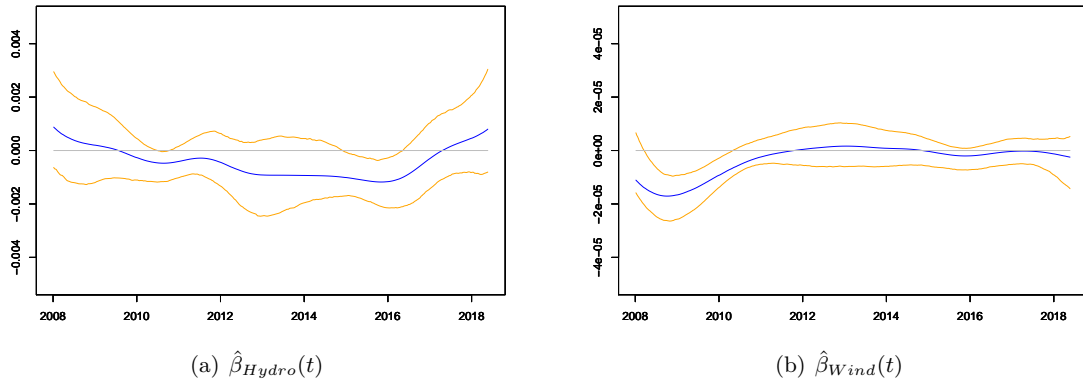


Figure 4: Nonparametrically estimated coefficient curves and 95% confidence intervals for wind and hydro power generation

### S3 Additional simulation results

As noted in Remark 1, the edge effects of the kernel estimator could distort the sieve bootstrap errors if the boundary residuals are resampled. Therefore, we exclude the boundary residuals and resample only from center residuals. We choose  $\delta$  as in Remark 1 to be equal to the bandwidth. We make this choice because it is the truncation of the kernel estimator which is mainly responsible for the boundary effects and  $nh$  observations get truncated. In Table S2 we report results for a small number of ARMA specification from our simulation study. Comparing these new results to the previous results we see at most a small improvement in coverage probabilities. The results do not seem to be greatly affected. The sieve bootstrap performance lies quite close to nominal level independent of whether or not we exclude boundary residuals. The sieve wild bootstrap performance is more problematic. However, as no resampling is involved the boundary residuals stay at the boundary and we do not have to adapt the bootstrap algorithm. We would nevertheless like to investigate the effect of potential edge effects. Therefore, we repeat in Table S3 a limited number of specifications. For the calculations of the coverage probabilities, we simply exclude the first and last  $[nh]$  observations to rule out potential boundary effects – such as truncation of the kernel estimator – as cause for the poor performance of the sieve wild bootstrap in some cases. We look at three specifications for a bandwidth of  $h = 0.04$ : white noise errors, ARMA errors with  $(0.3, 0.3)$  as well as GARCH errors with  $(0.2, 0.7)$ . Since the boundary points are excluded in the calculation of the simultaneous confidence bands over the sets  $G$  and  $G_{sub}$ , we only report results for pointwise coverage and simultaneity over the center part of the sample (i.e. the whole sample minus the boundary points). Comparing the results for pointwise intervals with the previous values from Tables 1 and 9, we see a small improvement in coverage probabilities. Including the boundary regions, the SWB coverage probabilities for  $\beta_1$  and  $\beta_2$  in the white noise case were 90 and 91%, respectively. Excluding the boundary points they are at 92%. For the ARMA errors they were 85 and 86% and increased to 87 and 88%. A larger improvement can be seen for simultaneous coverage

which was as low as 40 and 48% for white noise or 24 and 30% for the ARMA errors (see Table 4). Now, the values are 58 and 66% for white noise as well as 41 and 48% for the ARMA errors. The SB produced better simultaneous results at 92 and 93% for white noise and 89 and 91% for ARMA. Excluding the boundary points, there is a small improvement of around 2 percentage points. In the GARCH case, the results show a similar pattern. Overall, this exercise shows that excluding the boundary points can lead to some improvements but it is not the single explanation for why the sieve wild bootstrap performs so poorly, in particular, at simultaneous coverage.

	white noise		$\phi = 0.1$		$\psi = 0.3$		$(\phi, \psi) = (0.3, 0.3)$	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
pointwise	0.956 (0.307)	0.956 (0.306)	0.954 (0.306)	0.957 (0.305)	0.956 (0.299)	0.957 (0.293)	0.951 (0.360)	0.953 (0.349)
$G_{sub}$	0.951 (0.401)	0.944 (0.399)	0.934 (0.400)	0.937 (0.2398)	0.943 (0.390)	0.948 (0.383)	0.934 (0.470)	0.933 (0.456)
$G$	0.934 (0.434)	0.945 (0.432)	0.938 (0.432)	0.934 (0.431)	0.934 (0.421)	0.943 (0.413)	0.923 (0.509)	0.926 (0.493)
whole sample	0.906 (0.460)	0.916 (0.459)	0.898 (0.459)	0.898 (0.457)	0.911 (0.448)	0.926 (0.439)	0.889 (0.540)	0.887 (0.523)

Table S2: Excluding the first and last  $[nh]$  residuals in Step 1 of the sieve bootstrap, with  $h = 0.04$

	$\beta_1$		$\beta_2$		$\beta_1$		$\beta_2$	
	SB	SWB	SB	SWB	SB	SWB	SB	SWB
white noise	0.960 (0.306)	0.917 (0.288)	0.960 (0.309)	0.921 (0.272)	0.935 (0.459)	0.576 (0.432)	0.921 (0.463)	0.661 (0.409)
ARMA	0.951 (0.356)	0.867 (0.292)	0.959 (0.344)	0.879 (0.277)	0.907 (0.531)	0.405 (0.438)	0.930 (0.515)	0.477 (0.418)
GARCH	0.948 (0.285)	0.910 (0.246)	0.955 (0.283)	0.922 (0.223)	0.811 (0.436)	0.516 (0.368)	0.839 (0.432)	0.612 (0.335)

Table S3: Excluding the first and last  $[nh]$  observations when calculating coverage probabilities, with  $h = 0.04$ . We consider white noise, ARMA(1,1) with (0.3, 0.3) and GARCH(1,1) with (0.2, 0.7).

## References

- Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *Annals of Statistics*, 26:48–83.
- Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and its Application*. Academic Press, New York.
- Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

- Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econometric Theory*, 25(5):1433–1445.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1):159–178.
- Leybourne, S., Newbold, P., and Vougas, D. (1998). Unit roots and smooth transitions. *Journal of Time Series Analysis*, 19(1):83–97.
- Phillips, P. C. B. and Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, 75(2):335–346.
- Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607.