POTSDAM-INSTITUT FÜR KLIMAFOLGENFORSCHUNG

Autoregressive Wild Bootstrap Inference for Nonparametric Trends*

Marina Friedrich[a,b], Stephan Smeekes[a], and Jean-Pierre Urbain[†a]

[a]Maastricht University, Department of Quantitative Economics
[b]Potsdam Institute for Climate Impact Research (PIK)

November 12, 2018

Abstract

In this paper we propose an autoregressive wild bootstrap method to construct confidence bands around a smooth deterministic trend. The bootstrap method is easy to implement and does not require any adjustments in the presence of missing data, which makes it particularly suitable for climatological applications. We establish the asymptotic validity of the bootstrap method for both pointwise and simultaneous confidence bands under general conditions, allowing for general patterns of missing data, serial dependence and heteroskedasticity. The finite sample properties of the method are studied in a simulation study. We use the method to study the evolution of trends in daily measurements of atmospheric ethane obtained from a weather station in the Swiss Alps, where the method can easily deal with the many missing observations due to adverse weather conditions.

JEL classifications: C14, C22.
Keywords: autoregressive wild bootstrap, nonparametric estimation, time series, simultaneous confidence bands, trend estimation.

1 Introduction

The analysis of smoothly evolving trends is of interest in many fields such as economics and climatology. For instance, trend analysis in environmental variables is of major importance, as it can

*Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands and Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 601203, 14412 Potsdam, Germany. E-mail: friedrich@pik-potsdam.de, s.smeekes@maastrichtuniversity.nl. We would like to thank guest editor Tommaso Proietti, two referees, Eric Beutner, David Hendry, Franz Palm and Hanno Reuvers as well as participants at the conference Econometric Models of Climate Change in Aarhus, the Maastricht Workshop on Advances in Quantitative Economics II and a seminar at Oxford University, for helpful discussions and valuable comments. We further thank Whitney Bader, Bruno Franco, Bernard Lejeune and Emmanuel Mahieu for helpful discussions regarding the ethane application. The second author thanks the Netherlands Organisation for Scientific Research (NWO) for financial support.

†Deceased on October 1, 2016

often directly be linked to climate change. Given that trends typically do not evolve in a linear way, fitting linear trends to the data does not uncover actual change accurately. Instead, one would like to use more flexible trend models that avoid making parametric assumptions on the form of the trend. A large body of statistical and econometric research therefore focuses on nonparametric trend modeling and estimation.

In addition, when modeling trends in temperature and emission data, researchers also need to take into account that serial dependence and heteroskedasticity may be present in the data, see for example Franses and Vogelsang (2005) and McKitrick and Vogelsang (2014) who study parametric trend modeling in temperature series in the presence of serial dependence. Bootstrap methods provide an easy and powerful way to account for heteroskedasticity and autocorrelation. Bühlmann (1998) shows the validity of the autoregressive sieve bootstrap for nonparametric trend modeling under general forms of dependence. Neumann (1997) uses a wild bootstrap method to achieve robustness to heteroskedasticity for a similar model.

The wild bootstrap approach is also suitable for dealing with missing data, as advocated for example by Shao (2010b). It does not require any resampling and therefore the missing data points can keep their original date in a bootstrap sample. In particular in climatology, this feature of the wild bootstrap offers an important benefit over other methods, since there is no need of imputing missing data points. Missing data are a prominent feature in many climatological datasets due to instrument failure or adverse weather and measurement conditions; for instance, in our application, data are missing when cloud cover prevents measurements from being taken. The wild bootstrap, however, relies on independence of the error terms, which is a situation rarely encountered in practice. To relax this strong assumption, dependent versions of wild bootstrap methods have been proposed - see Shao (2010b), Leucht and Neumann (2013) and Smeekes and Urbain (2014) - but not in the context of nonparametric trend estimation. Moreover, so far no theory exists on the validity of such wild bootstrap methods in the presence of serial dependence, heteroskedasticty and missing data. In this paper we address this issue and propose an autoregressive wild bootstrap method that provides valid inference under general conditions for nonparametric trend modeling.

Next to the basic pointwise confidence intervals, we also study simultaneous confidence bands, which are often more informative about trend shapes than pointwise confidence intervals. Research questions, like whether upward trends are present over a certain period of time, should be addressed with simultaneous confidence bands as they involve multiple points in time at once. Wu and Zhao (2007) derive such bands for the nonparametric trend model that have asymptotically correct coverage probabilities, but do not consider bootstrap methods. Bühlmann (1998) proposes sieve bootstrap-based simultaneous confidence bands that are not only asymptotically valid but also have good small sample performance. They can, however, not easily be adjusted to be applicable to time series with missing data, as the autoregressive sieve bootstrap requires imputation of the missing values through for instance the Kalman filter. While this is certainly possible, it complicates implmentation. One can also argue about how accurate imputation methods are when a majority of the data are missing, as we face in our climatological application. Instead, we provide a much simpler alternative that requires no adjustments at all in the presence of missing data.

To illustrate our methodology, we study a time series of atmospheric ethane emissions for which

almost 70% of the data points are missing. When weather conditions are unfavorable – in particular due to cloud cover – measurements cannot be taken. The series has previously been investigated by Franco et al. (2015). Atmospheric ethane is an indirect greenhouse gas which can be used as an indicator of atmospheric pollution and transport. It is emitted during shale gas extraction and since shale gas has become more and more important as a source of natural gas, nonparametric trend analysis in ethane data provides geophysicists and climatologists with a tool to link trend changes to shale gas extraction activities, as well as study long-term climatological change.

The paper is organized as follows. In Section 2, our trend model is introduced along with the missing data generating mechanism. Section 3 describes the estimation procedure and the construction of bootstrap confidence bands. Subsequently, Section 4 derives the asymptotic properties of our method. Finite sample performance is analyzed in Section 5 in a simulation study. Trends in atmospheric ethane are studied in Section 6. Section 7 concludes. All technical details including proofs are given in Appendix A, while Supplementary Appendices B to D provide further results.

Finally, a word on notation. We denote by $\xrightarrow{d}$ weak convergence and by $\xrightarrow{p}$ convergence in probability. Whenever a quantity has a subscript *, it denotes a bootstrap quantity, conditional on the original sample. For instance, bootstrap weak convergence in probability is denoted by $\xrightarrow{d^*}_p$ (cf. Giné and Zinn, 1990). $\lfloor x \rfloor$ stands for the largest integer smaller than or equal to x. For any functions $f(x)$ and $g(x)$, defined on the same domain, $f^{(i)}(x) = \frac{d^i}{dx^i} f(x)$ and $[fg]^{(i)}(x) = \frac{d^i}{dx^i} f(x)g(x)$.

2 Trend Model with Missing Data

Consider the following data generating process (DGP):

$$y_t = m\left(\frac{t}{n}\right) + z_t \qquad t = 1, \ldots, n,$$

where $m(\cdot)$ is a smooth deterministic trend function and $z_t = \sigma_t u_t$ is a weakly dependent stochastic component. σ_t captures unconditional heteroskedasticity and $\{u_t\}$ is a linear process

$$u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \qquad \psi_0 = 1, \tag{1}$$

with autocovariance function $R_U(k) = \mathbb{E}\, u_t u_{t+k}$ and long-run variance

$$\Omega_U = \sum_{k=-\infty}^{\infty} \mathbb{E}\, u_t u_{t+k} = \sum_{k=-\infty}^{\infty} R_U(k).$$

Not all observations $y_1, \ldots, y_n$ are observed in practice. For this purpose, define the process $\{D_t\}$ as an indicator for whether the observations at each time are observed:

$$D_t = \begin{cases} 1 & \text{if } y_t \text{ is observed} \\ 0 & \text{if } y_t \text{ is missing} \end{cases} \qquad t = 1, \ldots, n,$$

Assumptions 1 to 4 contain the formal conditions that $\{y_t\}$ and $\{D_t\}$ satisfy.

Assumption 1. $m : [0,1] \to \mathbb{R}$ *is a twice continuously differentiable deterministic function on* $(0,1)$ *with* $\sup_{0<\tau<1} \left|m^{(i)}(\tau)\right| < \infty$ *for* $i = 0,1,2$.

Assumption 2. $\sigma : [0,1] \to \mathbb{R}^+$ *is a Lipschitz continuous deterministic function.*

Assumption 3. $\{u_t\}$ *is generated by* (1)*, where*

(i) $\{\epsilon_t\}$ *is i.i.d. with* $\mathbb{E}(\epsilon_t) = 0$, $\mathbb{E}(\epsilon_t^2) = \sigma_\varepsilon^2 = \left(\sum_{j=0}^\infty \psi_j^2\right)^{-1}$, *and* $\mathbb{E}(\epsilon_t^4) < \infty$.

(ii) $\sum_{j=0}^\infty j|\psi_j| < \infty$ *and the lag polynomial* $\Psi(z) = \sum_{j=0}^\infty \psi_j z^j \neq 0$ *for all* $z \in \mathbb{C}$ *and* $|z| \leq 1$.

Assumption 4. *For all* $t = 1, \ldots, n$, D_t *satisfies the conditions:*

(i) Let $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$ *where* $\mathcal{F}_t = \sigma(\ldots, (y_{t-1}, D_{t-1})', (y_t, D_t)')$. *For all* $s \leq t$ *and* $i \geq 0$, $\mathbb{E}[\mathbb{E}_{t-i} D_s D_t - \mathbb{E}\, D_s D_t]^2 \leq \zeta_i^2$, *where* $\sum_{i=0}^\infty i\zeta_i < \infty$.

(ii) $\mathbb{E}\, D_t = \mathbb{P}(D_t = 1) = p(t/n)$, *where* $p : [0,1] \to [\epsilon, 1]$, *for some* $\epsilon > 0$, *is a twice continuously differentiable function on* $(0,1)$ *with* $\sup_{0<\tau<1} \left|p^{(i)}(\tau)\right| < \infty$ *for* $i = 1,2$.

(iii) $\mathbb{C}\mathrm{ov}(D_t, D_{t+i}) = R_{D,i}\left(\frac{t}{n}, \frac{t+i}{n}\right)$, *where each function* $R_{D,i} : [0,1]^2 \to \mathbb{R}$, $i \geq 0$, *is Lipschitz continuous.*

(iv) For all $s_1, s_2 \in \{1, \ldots, n\}$, $\mathbb{E}(u_{s_1}|D_t) = 0$ *and* $\mathbb{E}(u_{s_1} u_{s_2}|D_t) = \mathbb{E}\, u_{s_1} u_{s_2}$.

Assumption 1 postulates that the trend $m(\cdot)$ is sufficiently smooth, which is the fundamental assumption for the estimation method to work. While it rules out abrupt structural breaks, this does not appear to be particularly restrictive for climatological applications, as many climatological processes tend to be such that change occurs gradually. In particular, as many series are measured daily or even multiple times a day, only instantaneous breaks, which are extremely unlikely in atmospheric processes, would not be covered by the smooth trend model.

Assumption 2 allows for a wide array of unconditional heteroskedasticity. While excluding abrupt breaks, these can be allowed for by generalizing the function $\sigma(\cdot)$ to be piecewise Lipschitz as in Smeekes and Urbain (2014). However, given the limited relevance of abrupt breaks for our climatological focus, we do not pursue this in the current paper.

Assumption 3 is a standard linear process assumption that ensures that sufficient moments of $\{u_t\}$ exist and $\{u_t\}$ is weakly dependent and strictly stationary. These assumptions are satisfied by a large class of processes including, but not limited to, all finite order stationary ARMA models. The assumption also implies that $\Omega_U = \sigma_\varepsilon^2 \sum_{i=-\infty}^\infty \sum_{j=0}^\infty \psi_j \psi_{j+|i|} < \infty$ (cf. Lemma A.1). While our current assumption does not allow for conditional heteroskedasticity, this could be relaxed at the expense of increasing the complexity of the theoretical arguments, by allowing ϵ_t to be a martingale difference sequence. Similarly, alternative dependence concepts such as mixing, which is considered in the same bootstrap context by Smeekes and Urbain (2014), could be used as well. However, as conditional heteroskedasticity is not the focus of our paper, we do not consider these extensions.

Assumption 4 allows the missing data generating mechanism to be weakly dependent and non-stationary. The mixingale assumption (i) along with the summability condition on ζ_i assures weak dependence and summable autocovariances (Lemma A.1). For technical reasons, we need to put the

mixingale assumption on the product $D_s D_t$, but it directly implies that D_t is a mixingale, as well, by setting $s = t$. By (ii) and (iii), the first two moments of the missing data process are allowed to vary smoothly over time, which thereby allows for instance for smooth periodical changing probabilities (e.g. due to seasonal variation), or long-term changes related to climate change. Smoothness is required as our estimator performs an implicit nonparametric estimate of the missing probability, and thus must behave similarly smooth as the trend function. Assumptions (i)-(iii) are met by a large class of generating processes, including many Markov chains with smoothly varying transition probabilities.

Assumption (iv) can be interpreted as an exogeneity assumption on the missing data generating mechanism, which for instance is satisfied if $\{D_t\}$ is independent of $\{u_t\}$. While this assumption could be argued to be restrictive, it does not appear to be problematic for our focus. Though inconclusive, some recent research has found evidence of a relation between greenhouse gases and the occurrence of cloud cover through climate change, see e.g. Norris et al. (2016). As cloud cover may cause missing observations, our assumption might appear restrictive, but this kind of long-run dependence can be accommodated through the slowly varying trend affecting both $\{y_t\}$ and $\{D_t\}$. As such, the exogeneity assumption mostly rules out short-run effects of ethane on cloud cover and vice versa, which we argue is reasonable.

3 Inference on Trends

Our goal is to conduct inference on the trend function $m(\cdot)$ defined in Section 2. We first describe point estimation of $m(\cdot)$, followed by our bootstrap method, and finally treat the construction of the confidence bands.

3.1 Estimation of the Trend Function

We consider local polynomial estimation which is common in the nonparametric regression literature. In particular, we focus on the local constant or Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964), defined as

$$\begin{aligned}\hat{m}(\tau) &= \underset{m(\tau)}{\arg\min} \sum_{t=1}^{n} K\left(\frac{t/n-\tau}{h}\right) D_t \left\{y_t - m(\tau)\right\}^2 \\ &= \left[\sum_{t=1}^{n} K\left(\frac{t/n-\tau}{h}\right) D_t\right]^{-1} \sum_{t=1}^{n} K\left(\frac{t/n-\tau}{h}\right) D_t y_t, \qquad \tau \in (0,1)\,,\end{aligned} \tag{2}$$

where $K(\cdot)$ is a kernel function and $h > 0$ is a bandwidth, which should satisfy Assumptions 5 and 6 given below. Note that by construction of the $\{D_t\}$ series, the formulation in (2) implies that the estimator only depends on the actually observed data.

Assumption 5. *$K(\cdot)$ is a symmetric, Lipschitz continuous function with compact support, where we define $\kappa_k = \int_{\mathbb{R}} K(\omega)^k \, \mathrm{d}\omega$, $\kappa(\tau) = \int_{\mathbb{R}} K(\omega)K(\omega-\tau) \, \mathrm{d}\omega$ and $\mu_k = \int_{\mathbb{R}} \omega^k K(\omega) \, \mathrm{d}\omega$.*

Assumption 6. *The bandwidth $h = h(n)$ satisfies $nh^7 \to 0$ and $nh^2 \to \infty$ as $n \to \infty$.*

Assumption 5 is a standard assumption in the nonparametric kernel smoother literature, and is satisfied by many commonly used kernels. Assumption 6 provides the range in convergence rates allowed for h to ensure consistency and asymptotic normality of the kernel estimator.

The bandwidth, or smoothing parameter h plays an important role. Large bandwidths produce a very smooth estimate, while small bandwidth produce a rough, wiggly trend estimate. Although Assumption 6 gives some guidance, it does not provide us with a practical bandwidth choice. Data-driven bandwidth selection is therefore important for implementation. Leave-one-out cross-validation is the most popular data-based method for bandwidth selection, but it is designed for independent observations and therefore inappropriate for time series data. Chu and Marron (1991) show that in the presence of positive correlation, this criterion systematically selects very small bandwidths, producing estimates which are too wiggly. With negative correlation, bandwidths will be large and the estimate too smooth. Therefore, Chu and Marron (1991) propose to use a time series version of this criterion, called modified cross-validation (MCV). It is based on minimizing the criterion function $\frac{1}{n}\sum_{t=1}^{n} D_t \left(\hat{m}_{k,h}\left(\frac{t}{n}\right) - y_t\right)^2$ with respect to h, where

$$\hat{m}_{k,h}(\tau) = \frac{(n-2k-1)^{-1}\sum_{t:|t-\tau n|>k} K\left(\frac{t/n-\tau}{h}\right) D_t y_t}{(n-2k-1)^{-1}\sum_{t:|t-\tau n|>k} K\left(\frac{t/n-\tau}{h}\right) D_t} \tag{3}$$

is a leave-$(2k+1)$-out version of the leave-one-out estimator of ordinary cross-validation, which leaves out the observation receiving the highest weight. Next to formal selection methods, visual inspection of the estimated trend function for a range of different bandwidths can help determining an appropriate bandwidth.

Remark 1. The assumption that $p(\cdot)$ is bounded away from zero implies that in every subinterval of $(0,1)$, we have enough observed data points as n grows large. This assumption can be relaxed at the expense of more involved notation by restricting our attention to those compact subsets of $(0,1)$, where the probability of observing data is larger than or equal to some $\epsilon > 0$.

In small samples, we may have points τ around which no data are observed in an h-neighborhood. As this is merely a small sample issue, for the theoretical analysis we implicitly assume that n is large enough such that, for every $\tau \in (0,1)$, $\sum_{t=1}^{n} K\left(\frac{t/n-\tau}{h}\right) D_t \geq \epsilon^*$ for some $\epsilon^* > 0$. That is, sufficient data are available around τ. This is not restrictive, as by Assumption 6, h decreases more slowly than n increases. In practice, the points around which insufficient data are available simply have to be excluded from the set of τ values considered for inference.

Remark 2. While the Nadaraya-Watson estimator locally approximates the trend function by a constant function, the local linear estimator locally fits a linear function to the data around a given point $\tau \in (0,1)$:

$$\begin{aligned}\begin{pmatrix}\hat{m}_{ll}(\tau)\\ \hat{m}_{ll}^{(1)}(\tau)\end{pmatrix} &= \underset{(m(\tau),m^{(1)}(\tau))}{\arg\min} \sum_{t=1}^{n} K\left(\frac{t/n-\tau}{h}\right) D_t \left\{y_t - m(\tau) - m^{(1)}(\tau)\left(t/n-\tau\right)\right\}^2 \\ &= \left(\sum_{t=1}^{n} K\left(\frac{t/n-\tau}{h}\right)\mathbf{x}_t(\tau)\mathbf{x}_t(\tau)' D_t\right)^{-1} \sum_{t=1}^{n} K\left(\frac{t/n-\tau}{h}\right)\mathbf{x}_t(\tau) D_t y_t\end{aligned} \tag{4}$$

where $\mathbf{x}_t(\tau) \equiv (1, t/n - \tau)'$. The local linear estimator is more accurate than the local constant estimator at points which are close to the boundaries of the sample. At these points, the local constant estimator suffers from boundary effects which the local linear estimator does not (Cai, 2007; Fan, 1992). Our analysis can be extended to the local linear estimator. However, notation becomes significantly more cumbersome and therefore proofs more complicated, so we focus on the local constant estimator in the theoretical part.

Remark 3. An alternative way to specify the estimators in the presence of missing data is to work with unequally spaced data. Assuming we observe data on times t_i for $i = 1, \ldots, n_1$, where n_1 is the effective number of observations, the local constant estimator can be written as

$$\hat{m}(\tau) = \left[\sum_{i=1}^{n_1} K\left(\frac{(t_i - t_1)/(t_{n_1} - t_1) - \tau}{h}\right)\right]^{-1} \sum_{i=1}^{n_1} K\left(\frac{(t_i - t_1)/(t_{n_1} - t_1) - \tau}{h}\right) y_{t_i}.$$

This formulation has the advantage that it does not require an underlying regular frequency at which the data are observed. However, in many applications it is not hard to define such an underlying frequency (days in our application), and then the two formulations are equivalent.

In the remainder of the paper we will continue to work with the first formulation, which leads to clearer notation and is easier to handle in the proofs, as the randomness in the missing data is modeled through the explicit D_t variable rather than being "hidden" in the (now random) t_i dates.

3.2 Autoregressive Wild Bootstrap

To construct confidence bands around the trend estimate, we modify the wild bootstrap, originally designed to handle heteroskedastic data (Davidson and Flachaire, 2008), to account fo serial dependence. The wild bootstrap generates bootstrap errors as $z_t^* = \xi_t^* \hat{z}_t$, where $\hat{z}_t$ are residuals of the nonparametric trend regression. In the standard wild bootstrap, the random variables $\{\xi_t^*\}$ are i.i.d. and thus, any dependence present in the data gets removed in the bootstrap errors. To overcome this drawback, Shao (2010b) proposed the *dependent wild bootstrap* (DWB) in which $\{\xi_t^*\}$ are generated as ℓ-dependent random variables with $\mathbb{C}\text{ov}(\xi_s^*, \xi_t^*) = K_{DWB}\left(\frac{s-t}{\ell}\right)$, where $K_{DWB}(\cdot)$ is a kernel function. As the tuning parameter ℓ, it has to be selected by the user.

Building on this idea, Smeekes and Urbain (2014) propose the *autoregressive wild bootstrap* (AWB) where $\{\xi_t^*\}$ is generated as an AR(1) process with parameter $\gamma = \gamma(n)$. The AWB has as advantage over the DWB that it is easier to implement and has a more intuitive interpretation. Moreover, as $\{\xi_t^*\}$ is not ℓ-dependent, the AWB has the potential to capture more serial correlation and to be less sensitive to the choice of tuning parameter γ. In the context of unit root testing, Smeekes and Urbain (2014) show that the AWB generally has a superior finite sample performance compared to the DWB. For these reasons, we mainly focus on the AWB in the following, although we consider the DWB in our simulation study as well. The AWB algorithm can be described as follows.

Algorithm 1 (Autoregressive Wild Bootstrap)**.**

1. Let $\tilde{m}(\cdot)$ be defined as in (2), but using bandwidth $\tilde{h}$. Obtain residuals $\hat{z}_t = D_t[y_t - \tilde{m}(t/n)]$

for $t = 1, \ldots, n$.

2. For $0 < \gamma < 1$, generate $\nu_1^*, \ldots, \nu_n^*$ as i.i.d. $\mathcal{N}(0, 1-\gamma^2)$ and let $\xi_t^* = \gamma\xi_{t-1}^* + \nu_t^*$ for $t = 2, \ldots, n$. Take $\xi_1^* \sim \mathcal{N}(0,1)$ to ensure stationarity of $\{\xi_t^*\}$.

3. Calculate the bootstrap errors z_t^* as $z_t^* = D_t\xi_t^*\hat{z}_t$ and generate the bootstrap observations by $y_t^* = D_t[\tilde{m}(t/n) + z_t^*]$ for $t = 1, \ldots, n$, where $\tilde{m}(\cdot)$ is the same estimate as in the first step.

4. Obtain the bootstrap estimator $\hat{m}^*(\cdot)$ as defined in (2) using the bootstrap series $\{y_t^*\}$, with the same bandwidth h as used for the original estimate $\hat{m}(\cdot)$.

5. Repeat Steps 2 to 4 B times, and let

$$\hat{q}_\alpha(\tau) = \inf\{u \in \mathbb{R} : \mathbb{P}^*[\hat{m}^*(\tau) - \tilde{m}(\tau) \leq u] \geq \alpha\} \tag{5}$$

denote the α-quantile of the B centered bootstrap statistics $\hat{m}^*(\tau) - \tilde{m}(\tau)$. These bootstrap quantiles are then used to construct confidence bands as described below.

Note that in Step 3 we only have to draw bootstrap observation for the dates where $D_t = 1$ and we observed the realization y_t; in the description the missing ones are artificially set to zero, but they are actually not used anywhere. In Step 2, we generate $\{\xi_t^*\}$ for all $t = 1, \ldots, n$, although subsequently we only use the subset that corresponds to the actually observed data points. The missing data structure is preserved in the bootstrap sample, while the correlation between consecutive non-missing observations is determined only by their distance, which ensures a coherent bootstrap sample. In this way, the missing data structure is automatically taken into account in the bootstrap without any need for modifications.

Although we suggest to generate $\{\nu_t^*\}$ as a sequence of normally distributed random variables, inspection of the proofs shows normality is not needed; all one needs is a sequence of i.i.d. random variables with $\mathbb{E}^*\,\nu_t^* = 0$, $\mathbb{E}^*\,\nu_t^{*2} = 1 - \gamma^2$ and $\mathbb{E}^{*4}\,\nu_t^{*4} < \infty$. Normality is simply a convenient option that is easy to implement; alternatively one could implement a variant of the Rademacher distribution (corrected for the right variance), which for the independent wild bootstrap has good properties (Davidson and Flachaire, 2008).

For the tuning parameter γ, we follow Smeekes and Urbain (2014) and let $\gamma = \theta^{1/\ell}$ where ℓ is the "block length" parameter also found in the DWB and $0 < \theta < 1$ is a fixed parameter. This specification has the advantage that ℓ can be interpreted in a similar way as the block length parameter in a block bootstrap; its choice constitutes a trade-off between capturing more of the dependence with a large value of the tuning parameter, and allowing for more variation in the bootstrap samples with a smaller value for ℓ. Additionally, it provides a convenient framework for studying the theoretical properties of our method. Specifically, we need $\ell \to \infty$ as $n \to \infty$, such that $\gamma \to 1$. This is analogous to the block (and dependent wild) bootstrap, where the block size must increase to capture more dependence when the sample size increases. Assumption 7 postulates the formal conditions that ℓ needs to satisfy. They imply that $\gamma \to 1$, but not too fast.

Assumption 7. *The bootstrap parameter $\ell = \ell(n)$ satisfies $\ell \to \infty$ and $\ell/\sqrt{nh} \to 0$ as $n \to \infty$.*

Note that we propose to use a different bandwidth $\tilde{h}$ in Step 1 of the algorithm. This is a common feature in the literature on bootstrap methods for nonparametric regression. By either selecting a larger (oversmoothing) or smaller bandwidth (undersmoothing) than used for the estimator, one can account for the asymptotic bias that is present in the local polynomial estimation, see Hall and Horowitz (2013, Section 1.4) for an extensive literature review. While undersmoothing, such as used in the related paper by Neumann and Polzehl (1998), aims at making the bias asymptotically negligible, oversmoothing aims at producing a consistent estimator of the (non-negligible) bias. Both have advantages and disadvantages, see the extensive discussion in Hall and Horowitz (2013). We follow Bühlmann (1998) and consider a solution based on oversmoothing, which we find to work well in practice; also see Härdle and Marron (1991). After presenting our theoretical results in Section 4, Remark 8 provides an intuition of why oversmoothing allows to consistently estimate the asymptotic bias.[1] We now state the formal conditions that $\tilde{h}$ must satisfy in Assumption 8; one is that $h/\tilde{h} \to 0$ as $n \to \infty$, which ensures the oversmoothing.

Assumption 8. *The oversmoothing bandwidth* $\tilde{h} = \tilde{h}(n)$ *satisfies* $\max\left\{\tilde{h}, h/\tilde{h}, nh^5\tilde{h}^4\right\} \to 0$ *and* $\ell \max\left\{\tilde{h}^4, 1/n\tilde{h}\right\} \to 0$ *as* $n \to \infty$.

Remark 4. There are a number of ways to choose the AWB parameter γ in practice. Using the relation $\gamma = \theta^{1/\ell}$, one can fix θ and choose ℓ as a deterministic function of the sample size. Smeekes and Urbain (2014) found that $\theta = 0.01$ paired with $\ell = 1.75n^{1/3}$ performed well in their simulation study on AWB unit root testing; for the local polynomial estimation we might adapt this to let ℓ be a function of nh, see also Remark 5. Alternatively, one may vary γ over a range of reasonable values, choosing a value in the range where the bands are most stable (as a function of γ), akin to the minimum volatility method proposed by Politis et al. (1999). Ideally, one would like to have a data-driven method for choosing an "optimal" γ. However, development of such a method requires a deeper study of higher-order asymptotic properties, which is outside the scope of the paper.

Remark 5. To give some intuition for the interaction between the three tuning parameters h, $\tilde{h}$ and ℓ, consider taking $h = cn^{-p}$, for some $c > 0$ and $\frac{1}{7} < p < \frac{1}{2}$. This satisfies Assumption 6. Now let $\tilde{h} = cn^{-q}$, where $p > q$ as $h/\tilde{h} \to 0$. Furthermore, to satisfy $nh^5\tilde{h}^4 \to 0$, we need that $5p + 4q > 1$. For $p \geq \frac{1}{5}$ this is satisfied for any q; with p approaching $\frac{1}{7}$, $q > \frac{1}{14}$ suffices, which is not very strict.

Finally, take $\ell = cn^r$. From Assumption 7 we know that $r < \frac{1}{2} - p/2$, which for p approaching $\frac{1}{2}$ means $r < \frac{1}{4}$, while for p approaching $\frac{1}{7}$, it implies that $r < \frac{3}{7}$. To satisfy the restrictions from Assumption 8, we need that $r < 4q$ and $r < 1 - q$. As $q < p < \frac{1}{2}$, the second condition is non-binding. The first condition is only restrictive when q is small, say in the vicinity of $\frac{1}{14}$. Then $r < \frac{2}{7}$, which is stricter than the condition implied by Assumption 7. However, for $q \geq \frac{3}{28}$, this restriction is not binding.

As an example, consider the "classical" rates $p = \frac{1}{5}$ and $q = \frac{1}{9}$, cf. Bühlmann (1998). These are allowed under our theory, and additionally imply that $0 < r < \frac{2}{5}$, therefore also allowing for $r = \frac{1}{3}$

[1] Hall and Horowitz (2013) propose an alternative bootstrap approach that requires neither under- nor oversmoothing, however their approach only delivers pointwise intervals, and is therefore not considered in this paper.

as advocated in Smeekes and Urbain (2014).

Remark 6. Instead of the autoregressive wild bootstrap, one could equally imagine a moving-average wild bootstrap (MA-WB), where $\xi_t^* = \sum_{j=0}^{\ell} \psi_{n,j}\nu_t^*$ and ν_t^* are i.i.d. random variables as before. By letting $\psi_{n,j} \to 1$, if $n \to \infty$ and j fixed, and $\psi_{n,j} \to 0$, if n is fixed and $j \to \infty$, one could show validity of such an MA-WB as well. For practical purposes, one could take $\psi_{n,j} = f(j/\ell)$, for instance with $f(x) = 1 - x^r$ to simplify implementation.

Such a moving-average representation is closely related to the DWB, which can be seen as a two-sided MA process, with both bootstrap methods delivering ℓ-dependent bootstrap samples. In his paper, Shao (2010b) shows the asymptotic equivalence of the variance estimator of the DWB with the tapered block bootstrap (Paparoditis and Politis, 2002). Similalry, one can show that the MA-WB has a close link to the extended tapered wild bootstrap of Shao (2010a), and depending on the distribution chosen for $\{\nu_t\}$, the smooth extended tapered block bootstrap of Gregory et al. (2015, 2018). If $\{\nu_t\}$ is drawn from a continuous distribution like the normal, we have an automatic smoothing in the wild bootstrap variants. As Gregory et al. (2015, 2018) show that smoothing helps in the context of the block bootstrap, it may similarly do so for the wild bootstrap and provide a potential reason to prefer the normal distribution over discrete distributions like the Rademacher distribution.

3.3 Bootstrap Confidence Bands

Pointwise bootstrap confidence intervals with a confidence level of $(1-\alpha)$ for $m(\tau)$, are denoted by $I_{n,\alpha}^{(p)}(\tau)$ and constructed with the objective that

$$\liminf_{n\to\infty} \mathbb{P}\left[\left(m(\tau) \in I_{n,\alpha}^{(p)}(\tau)\right)\right] \geq 1-\alpha \qquad \tau \in (0,1). \tag{6}$$

Using our bootstrap algorithm, we can construct such pointwise intervals as

$$I_{n,\alpha}^{(p)}(\tau) = \left[\hat{m}(\tau) - \hat{q}_{1-\alpha/2}(\tau), \hat{m}(\tau) - \hat{q}_{\alpha/2}(\tau)\right].$$

As these intervals are constructed separately for each τ, links over time cannot be established with these intervals. Therefore, we next consider how to construct simultaneous confidence bands. Let $I_{n,\alpha}^{G}(\tau)$, for $\tau \in G$, denote a confidence band that is simultaneous over the set G. Formally, we seek to construct $I_{n,\alpha}^{G}(\tau)$ such that

$$\liminf_{n\to\infty} \left[\mathbb{P}\left(m(\tau) \in I_{n,\alpha}^{G}(\tau) \quad \forall \tau \in G\right)\right] \geq 1-\alpha. \tag{7}$$

Our practical implementation follows the three-step procedure proposed by Bühlmann (1998):

1. For all $\tau \in G$, obtain pointwise quantiles $\hat{q}_{\alpha_p/2}(\tau), \hat{q}_{1-\alpha_p/2}(\tau)$ for varying $\alpha_p \in [1/B, \alpha]$.
2. Choose $\alpha_s = \arg\min_{\alpha_p \in [1/B,\alpha]} \left|\mathbb{P}^*\left[\hat{q}_{\alpha_p/2}(\tau) \leq \hat{m}^*(\tau) - \tilde{m}(\tau) \leq \hat{q}_{1-\alpha_p/2}(\tau) \quad \forall \tau \in G\right] - (1-\alpha)\right|$.

3. Construct the simultaneous confidence bands as

$$I^G_{n,\alpha}(\tau) = \left[\hat{m}(\tau) - \hat{q}_{1-\alpha_s/2}(\tau), \hat{m}(\tau) - \hat{q}_{\alpha_s/2}(\tau)\right] \qquad \tau \in G.$$

In the second step, a pointwise error α_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 $(1-\alpha)$.[2]

Remark 7. As an alternative to the variable-size bands proposed by Bühlmann (1998), one could consider Kolmogorov-Smirnov-type simultaneous confidence bands of fixed size. They would be of the form $I^*_\alpha(\tau) = \left[\hat{m}(\tau) - t^*_{1-\alpha}, \hat{m}(\tau) + t^*_{1-\alpha}\right]$, where the quantile $t^*_{1-\alpha}$ is determined as the $(1-\alpha)$-quantile of the distribution of the quantity $U^*_n = \sup_{\tau\in G}\{|\hat{m}^*(\tau) - \tilde{m}(\tau)|\}$. Neumann and Polzehl (1998) establish the asymptotic validity of such simultaneous bands under serial independence. We do not go in this direction, because we believe confidence bands with variable width to be more informative. They have the feature of becoming wider at points with more variability and more narrow for periods with less variability. To obtain variable width intervals with the Kolmogorov-Smirnov approach, one has to estimate the variance of the estimator at each τ and bootstrap a pivotal quantity, see e.g. Neumann and Polzehl (1998, Section 2.2). This adds additional complications in order to achieve consistent variance estimation.

4 Asymptotic Theory

We first provide the pointwise limiting normal distribution of the local constant estimator $\hat{m}(\cdot)$. Although the result is similar to the non-bootstrap part of Theorem 3.1 in Bühlmann (1998), we extend the asymptotic theory for the local constant estimator to allow for the presence of nonstationary volatility and missing data. As we feel this is a noteworthy result in its own right, we present it in Theorem 1.

Theorem 1. *Under Assumptions 1-6, for any $\tau \in (0,1)$, we have as $n \to \infty$:*

$$\sqrt{nh}\left(\hat{m}(\tau) - m(\tau) - h^2 B_{as}(\tau)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{as}(\tau)\right),$$

where

$$B_{as}(\tau) = \mu_2 p(\tau)^{-1}\left[mp\right]^{(2)}(\tau) \qquad \text{and} \qquad \sigma^2_{as}(\tau) = p(\tau)^{-1}\sigma(\tau)^2\Omega_U\kappa_2, \tag{8}$$

The term $B_{as}(\tau)$ reflects the familiar asymptotic bias generally found in local polynomial estimators, although the exact form is different due the presence of the missing data parameter $p(\tau)$. The asymptotic variance $\sigma^2_{as}(\tau)$ is not only affected by $p(\tau)$, but also by the volatility process $\sigma^2(\tau)$. If one were to use these distributions directly for inference, one would need to plug in consistent

[2] We provide R code to implement the estimator and bootstrap confidence bands on www.stephansmeekes.nl.

estimators of these nuisance parameters. However, as we show next, in the bootstrap these are automatically consistently estimated, and we have consistency of the autoregressive wild bootstrap method for the local constant estimator.

Theorem 2. *Under Assumptions 1-8, for any $\tau \in (0,1)$, we have as $n \to \infty$:*

$$\sqrt{nh}\left(\hat{m}^*(\tau) - \hat{m}(\tau) - h^2 B_{as}(\tau)\right) \xrightarrow{d^*}_p \mathcal{N}\left(0, \sigma^2_{as}(\tau)\right),$$

where $B_{as}(\tau)$ and $\sigma^2_{as}(\tau)$ are defined in (8).

The pointwise validity of the bootstrap confidence intervals in the sense of (6) follows directly from this pointwise convergence result. Note that, as the bias term $B_{as}(\tau)$ is the same in both theorems, it is consistently estimated by the bootstrap. As such, we do not need the bias to disappear, which happens when undersmoothing if $nh^5 \to 0$, or to be $O(1)$, when $nh^5 \to c$. Even if $nh \to \infty$, and the asymptotic bias dominates the stochastic variation, the bootstrap correctly mimics this and can be used for asymptotically valid inference. As such, we can relax the assumption in Bühlmann (1998, p. 55) that $h \sim Cn^{-1/5}$ to allow for a wider range of bandwidths. In practice, this means that the bootstrap provides additional protection against a misspecified bandwidth, by letting the widths of confidence bands automatically adapt.

Next, to study the validity of simultaneous confidence bands as in (7), we consider h-neighborhoods around time points τ. We do so because estimates $\hat{m}(\tau_1)$ and $\hat{m}(\tau_2)$ are asymptotically independent for $\tau_1 \neq \tau_2$ being two fixed distinct time points. When the distance between τ_1 and τ_2 is of order h, the estimators show a non-zero correlation. Therefore a major benefit of "zooming in" on local h-neighborhoods is that we can study how the bootstrap mimics the correlation between close points, a feature which is lost when considering simultaneity globally.

Theorem 3. *For any $\tau_0 \in (0,1)$, let*

$$Z_{\tau_0,n}(\tau) = \sqrt{nh}\left(\hat{m}(\tau_0 + \tau h) - m(\tau_0 + \tau h)\right), \qquad Z^*_{\tau_0,n}(\tau) = \sqrt{nh}\left(\hat{m}^*(\tau_0 + \tau h) - \tilde{m}(\tau_0 + \tau h)\right).$$

Then, under Assumptions 1-8, we have for all $\tau_0 \in (0,1)$

$$\begin{aligned}
&\left\{Z_{\tau_0,n}(\tau) - B_{as}(\tau_0)\right\}_{\tau\in[-1,1]} \Rightarrow \left\{W(\tau)\right\}_{\tau\in[-1,1]}, \\
&\left\{Z^*_{\tau_0,n}(\tau) - B_{as}(\tau_0)\right\}_{\tau\in[-1,1]} \Rightarrow_p \left\{W(\tau)\right\}_{\tau\in[-1,1]},
\end{aligned}$$

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

$$\mathbb{C}\mathrm{ov}(W(\tau_1), W(\tau_2)) = \sigma_{W,\tau_0}(\tau_1,\tau_2) = p(\tau_0)^{-1}\sigma(\tau_0)^2\Omega_U\kappa(\tau_1 - \tau_2).$$

Here, $\Rightarrow$ denotes weak convergence in the space of continuous real-valued functions on $[-1,1]$ endowed with the sup-norm.

Theorem 3 establishes the uniform validity of the bootstrap within an h-neighborhood around any point $0 < \tau_0 < 1$, where, since $h = o(1)$, we assume without loss of generality that $m(\tau_0 + \tau h)$

is always defined. Note that the interval $[-1,1]$ is mainly chosen out of convenience, and the results can trivially be shown to hold over any interval $[\tau_0 - ah, \tau_0 + bh]$ with $0 < a, b < \infty$. Moreover, it follows directly from Theorem 3 that the bootstrap will be valid uniformly on sets that contain a union of any finite number of such h-neighborhoods, see e.g. Bühlmann (1998, Corollary 3.3). While, in finite samples, one can always take h and the intervals such that the full sample is covered in G, this kind of "too large" simultaneity should be considered with caution, as this is not what the asymptotic analysis covers. Although simultaneity over such local sets might appear less attractive than simultaneity over the whole sample, it can nevertheless be of great interest in applications. For example, constructing confidence bands with simultaneous coverage over two time periods - one located early in the sample and the other one at the end - is useful when judging if there was an upward (or downward) movement of the trend at the end of the time period when compared to the beginning. This allows the empirical researcher to draw conclusions about developments spanning time stretches, which is not possible with pointwise confidence intervals.

Remark 8. To provide some intuition for the required oversmoothing with bandwidth $\tilde{h}$ in the bootstrap, note that from Theorem 1 (or more formally Lemma A.6) we can deduce that the estimator used in the first step of the bootstrap algorithm satisfies

$$\tilde{m}(\tau) - m(\tau) = \tilde{h}^2 B_{as}(\tau) + Z_n(\tau)/\sqrt{n\tilde{h}} + o_p\left(1/\sqrt{n\tilde{h}}\right), \tag{9}$$

where $Z_n(\tau) \xrightarrow{d} N(0, \sigma^2_{as}(\tau)$. Furthermore, letting $w_{t,n}(\tau) = K\left(\frac{t/n-\tau}{h}\right) D_t / \left[\sum_{t=1}^n K\left(\frac{t/n-\tau}{h}\right) D_t\right]$, and using that $y_t^* = \tilde{m}(t/n) + z_t^*$, we can write

$$\hat{m}^*(\tau) - \tilde{m}(\tau) = \left[\sum_{t=1}^n w_{t,n}(\tau)\tilde{m}(t/n) - \tilde{m}(\tau)\right] + \sum_{t=1}^n w_{t,n}(\tau) z_t^* + o_p\left(1/\sqrt{n\tilde{h}}\right).$$

While the second term, $\sum_{t=1}^n w_{t,n}(\tau) z_t^*$, mimics the stochastic variation in the trend estimation, and ensures the asymptotic normality of the bootstrap trend estimator, the bias arises from the first term, $\sum_{t=1}^n w_{t,n}(\tau)\tilde{m}(t/n) - \tilde{m}(\tau)$. Using (9), this term can be decomposed as

$$\begin{aligned}\sum_{t=1}^n w_{t,n}(\tau)\tilde{m}(t/n) - \tilde{m}(\tau) = &\left[\sum_{t=1}^n w_{t,n}(\tau) m(t/n) - m(\tau)\right] + \tilde{h}^2\left[\sum_{t=1}^n w_{t,n}(\tau) B_{as}(t/n) - B_{as}(\tau)\right] \\ &+ \left[\sum_{t=1}^n w_{t,n}(\tau) Z_n(t/n) - Z_n(\tau)\right]/\sqrt{n\tilde{h}} + o_p\left(1/\sqrt{n\tilde{h}}\right).\end{aligned}$$

The asymptotic bias arises, as for the original estimator, from the first term of the decomposition. As shown in the proof of Lemma A.7, by the smoothness of $B_{as}(\tau)$, the second term converges to zero, thus canceling out the bias in $\tilde{m}(\tau)$. However, to make the third, stochastic, term vanish when multiplying by $\sqrt{nh}$, it must be that $h/\tilde{h} \to 0$, such that this term is of small enough magnitude. This is achieved by oversmoothing.

Remark 9. Although it is not visible from the theorems – as we only consider pointwise τ (or τ_0 in Theorem 3) – τ needs to be bounded away from the boundaries (0 and 1) to make the results

hold. As the estimator exhibits edge effects, the quality of $\tilde{m}(\tau)$ can only be guaranteed for τ away from 0 or 1. Formally, we need to take a small $\delta > 0$ and then consider $\tau \in [\delta, 1-\delta]$. Consequently, in the bootstrap we then obtain the limit distribution in a slightly smaller set $\tau \in [\delta^*, 1-\delta^*]$, for some $\delta^* > \delta$. However, as we can take δ and δ^* as small as we like, this does not affect the pointwise statements of the theorems.[3]

Because of the above reasons, Bühlmann (1998, p. 53) suggests in his bootstrap algorithm to obtain residuals $\hat{z}_t = D_t[y_t - \tilde{m}(t/n)]$ for $t = [n\delta]+1, \ldots, [n(1-\delta)]$, as the residuals too close to the boundary may contaminate the bootstrap sample when they are sampled. This is not a problem for our method, as the AWB doesn't involve resampling: boundary residuals remain at the boundary in the bootstrap sample, and can therefore not affect results away from the boundaries.

5 Simulation Study

For the simulation exercise, we simulate time series with a trending behavior in both mean and variance, inspired by patterns observed in climatological time series, and allow for similar patterns of missing data. We will first describe the setting and then present and discuss the results.

5.1 Simulation Setup

We consider the following smooth transition model:

$$y_t = m(t/n) + \sigma_t u_t, \qquad m(\tau) = \beta_1 \tau + \beta_2 \tau G(\tau, \lambda, c), \tag{10}$$

where for $\lambda > 0$,

$$G(\tau, \lambda, c) = \left(1 + \exp\left\{-\lambda(\tau - c)\right\}\right)^{-1}. \tag{11}$$

The error term $\{u_t\}$ follows an ARMA$(1,1)$ model

$$u_t = \phi u_{t-1} + \psi \epsilon_{t-1} + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}\left(0, \frac{(1-\phi^2)/4}{1+\psi^2-2\phi\psi}\right), \tag{12}$$

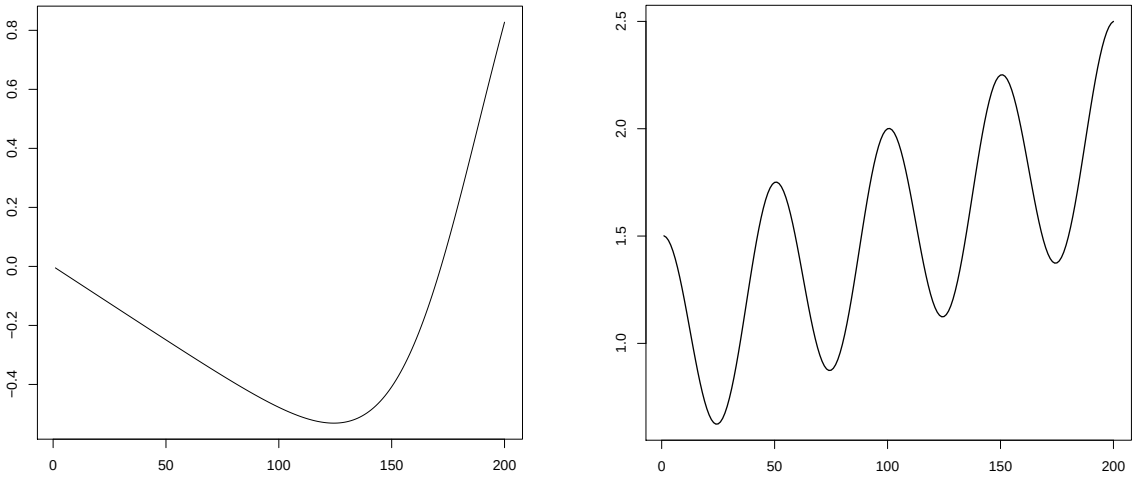
where we vary the parameters ϕ and ψ to investigate the impact of serial correlation on our method. The variance of ϵ_t is normalized such that the signal to noise ratio does not depend on the specific choice of the AR and MA parameter. Furthermore, we introduce heteroskedasticity with the process $\{\sigma_t\}$. We consider two scenarios, where σ_t is constant over time or $\sigma_t = \sigma(t/n)$, with the volatility process $\sigma(\tau)$ given by

$$\sigma(\tau) = \sigma_0 + (\sigma_* - \sigma_0)(\tau) + a\cos\left(2\pi k \tau\right). \tag{13}$$

Equation (10) is a shifting mean model as considered by Gonzalés and Teräsvirta (2008), and can be seen as a smooth transition version of a broken trend model with one break. The function $G(\tau, \lambda, c)$

[3] See Lemmas A.6 and A.7 in Appendix A for statements where these constants do appear.

as given in (11) is the transition function with time as transition variable. Its inputs apart from time are the location of the shift – the parameter c – as well as the smoothness of the shift, determined by λ. For large values of λ the shift happens almost instantaneous, while it is smoother for smaller values of this parameter. In our simulations, we fix $\lambda = 10$. The other parameters of our DGP will be chosen in such a way that the time series experiences a downward trend during the first three quarters which turns into a steeper upward trend in the last quarter.



(a) The trend function $m(\tau)$

(b) The variance process $\sigma(\tau)$ with $k = 4$ and $a = 0.5$

Figure 1: Trends in mean and variance in the simulation DGP

This mimics the general pattern which is expected to occur in atmospheric ethane time series and therefore fits our application well. More specifically, this means we set the location of the shift to occur at $c = 0.9$. The slope of the trend gradually changes from $\beta_1 = -1$ before the shift to $\beta_2 = 2.5$ after the shift. This is illustrated in Figure 1(a). For the variance process, inspired by the series considered in the empirical application, we consider a cyclical component with trend. We have to choose four parameters in (13): the start and end point of the trend – σ_0 and σ_* – as well as the specifics of the cyclical component. The parameter a fixes the amplitude of the cycle, while k determines how many cycles there are. We set $\sigma_0 = 1$, $\sigma_* = 2$ and consider different combinations of values for a and k. We let $a = 0.3, 0.5, 0.7$ and $k = 2, 3, 4$. An example of this process is displayed in Figure 1(b).

We also consider different degrees of dependence by varying the AR and MA parameters. For the AR parameter we take $\phi = 0, 0.2, 0.5, -0.5$, while the MA parameter varies between $\psi = 0$, $\psi = 0.2$ and $\psi = 0.5$. We only look at pure AR or MA processes with these coefficient values. The different specifications will be abbreviated in the tables with self-explanatory names, e.g. we write $AR_{-0.5}$ for $\phi = -0.5$, $\psi = 0$ and use $MA_{0.5}$ when $\phi = 0$ and $\psi = 0.5$.

In addition, we consider cases of missing data for which we generate a missing pattern that is representative for the ethane data. We implement a first-order Markov Chain for D_t with transition

probabilities

$$\begin{array}{c} \\ D_{t-1}=0 \\ D_{t-1}=1 \end{array} \begin{array}{c} \begin{array}{cc} D_t=0 & D_t=1 \end{array} \\ \left(\begin{array}{cc} 0.80 & 0.20 \\ 0.45 & 0.55 \end{array} \right), \end{array} \tag{14}$$

which are estimated from the ethane time series considered in Section 6. This transition matrix results in an average fraction of around 70% missing observations.

In the estimation step, we apply the local constant estimator based on the Epanechnikov kernel which is given by the function $K(x) = \frac{3}{4}(1-x^2)\mathbb{1}_{\{|x|\leq 1\}}$. For the bandwidth parameter h we use $h = 0.02$, $h = 0.04$ and $h = 0.06$. In the first step of the bootstrap procedure we follow the recommendation of Bühlmann (1998) to use $\tilde{h} = Ch^{5/9}$ with $C = 2$. In the second step of the bootstrap, we consider different values for the AR parameter γ; next to $\gamma = 0$, which reduces the AWB to a standard wild bootstrap (WB), we also consider $\gamma = 0.2, 0.4, 0.6$.

For each specification, we run 5000 Monte Carlo simulations. We report average pointwise as well as simultaneous coverage for a sample size of $n = 200$, based on $B = 999$ bootstrap replications. For ease of comparison, we choose the sample size in cases with missing data such that we have approximately 200 data points remaining. Given the large fraction of missings, an expected effective sample size of 200 translates into a original sample size for our Markov Chain of $n = 666$.

The nominal coverage in all cases is 95%. We also report the average median length of the confidence intervals in parenthesis underneath the respective coverage. For simultaneous coverage, the trend curve has to lie within the confidence bands for all points of the considered set G, for which we take 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 $U_i(h) = \{(i/5) - h + j/100;\ j = 0, ..., [200h]\}$.

To compare the performance of our AWB method to related bootstrap methods, we also implement the dependent wild bootstrap (DWB) and the sieve wild bootstrap (SWB), with a standard normal distribution for generation of the wild bootstrap errors. Since the SWB cannot easily be adapted to work with missing data, we provide results for this method only for the cases with no missing data. For the DWB, we convert the γ parameter into the corresponding value for the tuning parameter ℓ, using the formula $\gamma = \theta^{1/\ell}$. Smeekes and Urbain (2014) found in their simulation study that $\theta = 0.01$ provides a sensible conversion between the AWB and DWB, in the sense of yielding comparable performance of the two methods, therefore we use $\theta = 0.01$ as well to convert the AWB parameter into the DWB parameter. This tuning parameter does not exist in the case of the SWB; instead, the lag length has to be selected, for which we use AIC.

5.2 Simulation Results

First, we report results for equally spaced data with no missing observations and a variance process with the same specifications ($k = 4$ and $a = 0.5$) as displayed in Figure 1(b). The results on pointwise coverage are given in Table 1, while Tables 2 and 3 show simultaneous coverage probabilities for the two sets $G.sub$ and G, respectively. The tables consist of three main blocks, one for each bandwidth. Within each block, the individual rows contain results for different combinations of AR

and MA parameters. Results for other choices of the variance parameters k and a, including the homoskedastic case, are available in Supplementary Appendix C. Qualitatively, these settings yield the same conclusions as the ones considered here.

Table 1 shows that the autoregressive wild bootstrap provides confidence intervals with good pointwise coverage in the presence of heteroskedasticity and mild autocorrelation. For the independent case and the cases with negative or small positive correlation, the coverage probabilities are close to the nominal level. The only specifications for which the coverage lies below the nominal level are when $\phi = 0.5$ and $\psi = 0.5$. In these cases, the data deviate from the trend line in clusters due to the strong positive correlation. This causes the nonparametric estimate to go through these clusters and thus, to deviate significantly from the true trend. The confidence bands are in these situations not wide enough to cover the true trend, resulting in too low coverage. Interestingly, all methods/tuning parameters are similarly affected.

Concerning the autoregressive parameter of the wild bootstrap, we can observe that whenever the data are serially correlated, the autoregressive wild bootstrap ($\gamma \neq 0$) provides better coverage than the standard wild bootstrap ($\gamma = 0$). In addition, with stronger correlation, a larger value for γ should be preferred, except that the case $\gamma = 0.6$ provides consistently lower coverage, indicating that simply going for a very large value of γ is not sensible in practice. However, even if we do see these patterns, in general, the coverage probabilities do not vary substantially with the autoregressive parameter and therefore, the sensitivity to this parameter appears to be fairly limited.

When we look at the different blocks of Table 1, the bootstrap shows a similar overall performance regardless of the value we select for the bandwidth parameter. Since the bandwidth plays such an important role in nonparametric estimation, yet there are no fully satisfactory ways to select optimal bandwidth from the data in most applications, robustness to bandwidth "misspecifcation" is an important finding. This implies that the bootstrap can correct for poorly chosen bandwidths.

We observe similar patterns in Tables 2 and 3, while overall coverage is lower for the set G than for G_{sub}. This is not surprising, since the former set covers twice as many points as the latter. Interestingly, the confidence bands are consistently more narrow with G than they are with G_{sub}. This appears counterintuitive at first, as G is twice as large as G_{sub}. However, while G_{sub} is made up entirely of points relatively close to the boundaries, for which estimation is more variable, G additionally contains "stable" regions closer to the center. It may be that the stability of these regions has an offsetting effect compared with the boundary regions, reducing the size of the intervals. As a side effect, overall coverage is also reduced. As such, if one is only interested in coverage near the beginning and the end of the sample, it may be wiser to only attempt to achieve uninformity over these regions, rather than over the full sample.

Similar to the pointwise coverage results, the simultaneous coverage is close to the nominal level for the two cases with negative correlation as well as the independent case. Weak positive correlation can also be handled decently. The cases $\phi = 0.5$ and $\psi = 0.5$ are more problematic, as coverage drops to around 60% for G and 70% for G_{sub}. In these cases, the smallest bandwidth $h = 0.02$ seems to be preferred.

Comparing the AWB to the other two bootstrap methods, we can see that in almost all cases, AWB and DWB show similar results and they outperform the sieve version. Often, the AWB results

in slightly higher coverage with shorter intervals. Only when $\gamma = 0.6$, the DWB displays better coverage. Further increases of this parameter did not lead to improvements. An exception is the DGP with negative correlation, where the DWB displays coverage that is too high, independent of the choice of γ. In such cases, the AWB is often more accurate for $\gamma = 0.4$. In all other cases, we see that for both methods, the best performance is similar in magnitude but obtained at a different value of the tuning parameter. Since the tuning parameters do not have exactly the same meaning in both methods, there is no reason to expect identical variation. We chose $\theta = 0.01$ to link the two methods; changing this value will likely change the relation between the methods as well.

Next, we consider the setting with missing data. Given the previous findings, we restrict ourselves to one bandwidth ($h = 0.06$) but consider all AR and MA models. The results for pointwise as well as simultaneous coverage probabilities are given in Table 4; further results, with similar conclusions, are available in Supplementary Appendix C.

The first block presents pointwise coverage, while the second and third blocks show results for the sets G_{sub} and G. The AWB performs well even if a significant proportion of the data are missing, as both pointwise and simultaneous coverage is close to the nominal level for almost all models. There is a significant increase in pointwise coverage for the cases with strong positive correlation, which is now close to 90%. The same increase is visible for both G_{sub} and G. This phenomenon does not come as a surprise, as the missing data points create space between consecutive observations, thus effectively reducing the serial dependence between observed points. Comparing the AWB with the DWB, we can see that the coverage is slightly closer to 95% for the AWB in many cases. As before, the best performance is obtained at different values of γ for the AWB and the DWB. The general pattern, however, is as in the previous setting. The DWB outperforms the AWB for higher values of γ, while the AWB obtains the most accurate coverage for smaller values of this parameter. A notable exception are again the cases with negative autocorrelation. Results for other bandwidths are similar, and presented in Supplementary Appendix C.

Overall, this simulation study indicates that the autoregressive wild bootstrap performs well in most of our considered scenarios. The pointwise confidence intervals show coverage close to the nominal level in the presence of heteroskedasticity and serial correlation. In addition, the method still performs well in the presence of missing data. The AWB provides simultaneous confidence bands with good coverage as long as the correlation does not become too strong. It outperforms the sieve wild bootstrap whenever we could compare results. In comparison to the dependent wild bootstrap, we saw that both methods provide very similar coverage probabilities, while the DWB produces slightly wider intervals.

6 Trends in Atmospheric Ethane

We use our methodology to investigate the trending behavior of a time series of atmospheric ethane emissions which is derived from observations performed at the Jungfraujoch station in the Swiss Alps. This station can be found on the saddle between the Jungfrau and the Mönch, located at 46.55° N, 7.98° E, 3580 m altitude. Ethane is the most abundant hydrocarbon gas in the atmosphere after methane and it is used as a measure of atmospheric pollution. It contributes

		$\gamma = 0$	$\gamma = 0.2$		$\gamma = 0.4$		$\gamma = 0.6$		$-$
h	DGP	WB	AWB	DWB	AWB	DWB	AWB	DWB	SWB
0.02	0	0.946	0.952	0.946	0.942	0.945	0.905	0.945	0.925
		(0.732)	(0.715)	(0.733)	(0.658)	(0.733)	(0.555)	(0.733)	(0.719)
	$AR_{0.2}$	0.902	0.918	0.902	0.912	0.902	0.871	0.902	0.861
		(0.710)	(0.711)	(0.711)	(0.672)	(0.711)	(0.582)	(0.711)	(0.688)
	$AR_{0.5}$	0.778	0.818	0.776	0.828	0.778	0.791	0.777	0.672
		(0.604)	(0.630)	(0.605)	(0.621)	(0.604)	(0.562)	(0.603)	(0.536)
	$AR_{-0.5}$	0.989	0.988	0.989	0.982	0.989	0.958	0.989	0.960
		(0.623)	(0.575)	(0.624)	(0.499)	(0.624)	(0.397)	(0.624)	(0.552)
	$MA_{0.2}$	0.910	0.925	0.911	0.918	0.909	0.879	0.910	0.872
		(0.712)	(0.712)	(0.712)	(0.671)	(0.712)	(0.580)	(0.712)	(0.690)
	$MA_{0.5}$	0.864	0.891	0.864	0.891	0.864	0.854	0.864	0.773
		(0.638)	(0.657)	(0.639)	(0.636)	(0.638)	(0.564)	(0.639)	(0.571)
0.04	0	0.950	0.952	0.951	0.940	0.951	0.905	0.951	0.922
		(0.567)	(0.561)	(0.568)	(0.527)	(0.568)	(0.460)	(0.568)	(0.548)
	$AR_{0.2}$	0.903	0.911	0.901	0.905	0.901	0.872	0.902	0.862
		(0.553)	(0.563)	(0.553)	(0.544)	(0.553)	(0.487)	(0.554)	(0.525)
	$AR_{0.5}$	0.768	0.803	0.766	0.818	0.767	0.795	0.765	0.667
		(0.478)	(0.511)	(0.479)	(0.519)	(0.478)	(0.489)	(0.477)	(0.409)
	$AR_{-0.5}$	0.996	0.994	0.996	0.989	0.996	0.969	0.996	0.970
		(0.492)	(0.457)	(0.492)	(0.405)	(0.492)	(0.338)	(0.493)	(0.423)
	$MA_{0.2}$	0.912	0.922	0.913	0.913	0.912	0.880	0.912	0.873
		(0.553)	(0.563)	(0.554)	(0.542)	(0.554)	(0.485)	(0.554)	(0.526)
	$MA_{0.5}$	0.867	0.890	0.869	0.890	0.868	0.861	0.867	0.784
		(0.501)	(0.528)	(0.503)	(0.522)	(0.502)	(0.479)	(0.503)	(0.436)
0.06	0	0.955	0.957	0.956	0.946	0.956	0.917	0.956	0.925
		(0.492)	(0.492)	(0.493)	(0.471)	(0.493)	(0.422)	(0.493)	(0.465)
	$AR_{0.2}$	0.910	0.917	0.907	0.913	0.907	0.887	0.909	0.865
		(0.480)	(0.494)	(0.481)	(0.486)	(0.480)	(0.447)	(0.481)	(0.444)
	$AR_{0.5}$	0.781	0.814	0.777	0.831	0.779	0.815	0.776	0.669
		(0.420)	(0.455)	(0.421)	(0.468)	(0.420)	(0.453)	(0.420)	(0.347)
	$AR_{-0.5}$	0.997	0.996	0.997	0.992	0.997	0.977	0.997	0.979
		(0.432)	(0.408)	(0.432)	(0.373)	(0.432)	(0.325)	(0.433)	(0.358)
	$MA_{0.2}$	0.918	0.928	0.920	0.921	0.918	0.894	0.918	0.875
		(0.480)	(0.495)	(0.481)	(0.485)	(0.481)	(0.444)	(0.481)	(0.446)
	$MA_{0.5}$	0.878	0.900	0.879	0.901	0.878	0.878	0.878	0.789
		(0.439)	(0.467)	(0.440)	(0.470)	(0.440)	(0.441)	(0.440)	(0.370)

Table 1: Pointwise coverage probabilities (average median interval length) for $k = 4$ and $a = 0.5$.

		$\gamma = 0$	$\gamma = 0.2$		$\gamma = 0.4$		$\gamma = 0.6$		$-$
h	DGP	WB	AWB	DWB	AWB	DWB	AWB	DWB	SWB
0.02	0	0.941	0.944	0.936	0.934	0.935	0.870	0.937	0.917
		(0.672)	(0.656)	(0.673)	(0.604)	(0.672)	(0.510)	(0.673)	(0.661)
	$AR_{0.2}$	0.884	0.908	0.883	0.888	0.879	0.822	0.884	0.805
		(0.652)	(0.653)	(0.653)	(0.617)	(0.653)	(0.534)	(0.653)	(0.632)
	$AR_{0.5}$	0.700	0.759	0.689	0.769	0.700	0.687	0.700	0.427
		(0.555)	(0.578)	(0.556)	(0.570)	(0.555)	(0.515)	(0.553)	(0.493)
	$AR_{-0.5}$	0.982	0.981	0.983	0.974	0.987	0.920	0.984	0.985
		(0.572)	(0.528)	(0.573)	(0.459)	(0.573)	(0.365)	(0.573)	(0.507)
	$MA_{0.2}$	0.899	0.913	0.891	0.900	0.888	0.835	0.895	0.820
		(0.654)	(0.653)	(0.654)	(0.616)	(0.654)	(0.533)	(0.654)	(0.634)
	$MA_{0.5}$	0.830	0.860	0.821	0.853	0.810	0.792	0.821	0.619
		(0.586)	(0.604)	(0.587)	(0.584)	(0.586)	(0.518)	(0.587)	(0.524)
0.04	0	0.930	0.931	0.929	0.907	0.932	0.818	0.933	0.902
		(0.521)	(0.516)	(0.522)	(0.484)	(0.522)	(0.423)	(0.522)	(0.504)
	$AR_{0.2}$	0.849	0.858	0.836	0.834	0.835	0.759	0.848	0.774
		(0.508)	(0.517)	(0.508)	(0.500)	(0.508)	(0.448)	(0.509)	(0.483)
	$AR_{0.5}$	0.592	0.648	0.564	0.685	0.577	0.604	0.573	0.376
		(0.439)	(0.470)	(0.440)	(0.476)	(0.440)	(0.449)	(0.439)	(0.376)
	$AR_{-0.5}$	0.997	0.991	0.995	0.980	0.994	0.917	0.995	0.993
		(0.452)	(0.420)	(0.453)	(0.373)	(0.452)	(0.310)	(0.453)	(0.390)
	$MA_{0.2}$	0.859	0.877	0.857	0.860	0.864	0.774	0.863	0.790
		(0.508)	(0.517)	(0.509)	(0.498)	(0.509)	(0.446)	(0.510)	(0.483)
	$MA_{0.5}$	0.767	0.814	0.770	0.807	0.771	0.733	0.766	0.594
		(0.461)	(0.485)	(0.462)	(0.480)	(0.461)	(0.440)	(0.462)	(0.401)
0.06	0	0.930	0.929	0.931	0.907	0.939	0.819	0.936	0.906
		(0.452)	(0.453)	(0.453)	(0.433)	(0.453)	(0.388)	(0.453)	(0.428)
	$AR_{0.2}$	0.852	0.845	0.832	0.835	0.840	0.757	0.851	0.783
		(0.441)	(0.455)	(0.442)	(0.447)	(0.442)	(0.411)	(0.442)	(0.409)
	$AR_{0.5}$	0.564	0.624	0.554	0.663	0.565	0.594	0.537	0.366
		(0.386)	(0.418)	(0.387)	(0.431)	(0.387)	(0.416)	(0.386)	(0.319)
	$AR_{-0.5}$	0.998	0.995	0.998	0.987	0.998	0.937	0.998	0.995
		(0.397)	(0.376)	(0.398)	(0.343)	(0.397)	(0.299)	(0.398)	(0.330)
	$MA_{0.2}$	0.862	0.870	0.857	0.856	0.865	0.771	0.862	0.796
		(0.442)	(0.455)	(0.443)	(0.446)	(0.443)	(0.409)	(0.443)	(0.410)
	$MA_{0.5}$	0.768	0.801	0.761	0.811	0.773	0.732	0.762	0.611
		(0.404)	(0.430)	(0.405)	(0.432)	(0.404)	(0.405)	(0.405)	(0.341)

Table 2: Simultaneous coverage probabilities (average median interval length) over G_{sub} for $k = 4$ and $a = 0.5$.

		$\gamma = 0$	$\gamma = 0.2$		$\gamma = 0.4$		$\gamma = 0.6$		$-$
h	DGP	WB	AWB	DWB	AWB	DWB	AWB	DWB	SWB
0.02	0	0.925	0.939	0.925	0.919	0.926	0.845	0.925	0.894
		(0.527)	(0.515)	(0.528)	(0.474)	(0.528)	(0.400)	(0.528)	(0.518)
	$AR_{0.2}$	0.855	0.887	0.855	0.863	0.853	0.768	0.866	0.738
		(0.512)	(0.513)	(0.512)	(0.485)	(0.512)	(0.420)	(0.513)	(0.496)
	$AR_{0.5}$	0.608	0.690	0.604	0.691	0.602	0.590	0.606	0.266
		(0.435)	(0.454)	(0.435)	(0.449)	(0.435)	(0.407)	(0.434)	(0.386)
	$AR_{-0.5}$	0.982	0.979	0.979	0.972	0.986	0.915	0.984	0.984
		(0.448)	(0.413)	(0.449)	(0.359)	(0.449)	(0.285)	(0.449)	(0.398)
	$MA_{0.2}$	0.874	0.896	0.869	0.876	0.865	0.792	0.870	0.770
		(0.513)	(0.513)	(0.513)	(0.484)	(0.513)	(0.419)	(0.513)	(0.497)
	$MA_{0.5}$	0.780	0.833	0.779	0.812	0.764	0.728	0.776	0.498
		(0.459)	(0.474)	(0.460)	(0.459)	(0.460)	(0.408)	(0.460)	(0.411)
0.04	0	0.914	0.915	0.915	0.882	0.917	0.762	0.917	0.875
		(0.401)	(0.397)	(0.402)	(0.373)	(0.402)	(0.326)	(0.402)	(0.388)
	$AR_{0.2}$	0.812	0.822	0.793	0.798	0.799	0.688	0.806	0.697
		(0.391)	(0.398)	(0.391)	(0.385)	(0.391)	(0.345)	(0.392)	(0.372)
	$AR_{0.5}$	0.461	0.544	0.439	0.571	0.435	0.480	0.441	0.206
		(0.338)	(0.362)	(0.339)	(0.368)	(0.338)	(0.347)	(0.338)	(0.289)
	$AR_{-0.5}$	0.996	0.991	0.995	0.981	0.995	0.907	0.995	0.993
		(0.348)	(0.323)	(0.348)	(0.287)	(0.348)	(0.239)	(0.349)	(0.300)
	$MA_{0.2}$	0.824	0.848	0.824	0.818	0.824	0.705	0.832	0.718
		(0.391)	(0.398)	(0.392)	(0.384)	(0.392)	(0.344)	(0.392)	(0.372)
	$MA_{0.5}$	0.694	0.755	0.700	0.743	0.692	0.646	0.697	0.464
		(0.355)	(0.374)	(0.356)	(0.370)	(0.355)	(0.339)	(0.356)	(0.309)
0.06	0	0.913	0.911	0.912	0.881	0.922	0.761	0.922	0.884
		(0.345)	(0.346)	(0.346)	(0.331)	(0.346)	(0.297)	(0.346)	(0.327)
	$AR_{0.2}$	0.794	0.804	0.781	0.776	0.784	0.663	0.790	0.705
		(0.337)	(0.348)	(0.338)	(0.342)	(0.338)	(0.315)	(0.338)	(0.313)
	$AR_{0.5}$	0.415	0.497	0.395	0.529	0.396	0.450	0.390	0.196
		(0.295)	(0.320)	(0.296)	(0.330)	(0.295)	(0.319)	(0.295)	(0.244)
	$AR_{-0.5}$	0.999	0.995	0.997	0.988	0.998	0.932	0.998	0.996
		(0.304)	(0.287)	(0.304)	(0.262)	(0.304)	(0.228)	(0.304)	(0.252)
	$MA_{0.2}$	0.815	0.835	0.817	0.807	0.827	0.682	0.826	0.740
		(0.338)	(0.348)	(0.339)	(0.341)	(0.338)	(0.313)	(0.338)	(0.313)
	$MA_{0.5}$	0.680	0.740	0.686	0.729	0.683	0.632	0.683	0.480
		(0.308)	(0.329)	(0.310)	(0.331)	(0.309)	(0.310)	(0.309)	(0.260)

Table 3: Simultaneous coverage probabilities (average median interval length) over G for $k = 4$ and $a = 0.5$.

		$\gamma = 0$	$\gamma = 0.2$		$\gamma = 0.4$		$\gamma = 0.6$	
	DGP	WB	AWB	DWB	AWB	DWB	AWB	DWB
pw	0	0.960	0.959	0.960	0.949	0.960	0.923	0.961
		(0.303)	(0.303)	(0.303)	(0.290)	(0.303)	(0.264)	(0.304)
	$AR_{0.2}$	0.939	0.940	0.937	0.933	0.939	0.906	0.938
		(0.297)	(0.302)	(0.297)	(0.295)	(0.298)	(0.272)	(0.298)
	$AR_{0.5}$	0.885	0.897	0.884	0.894	0.883	0.874	0.883
		(0.267)	(0.280)	(0.267)	(0.280)	(0.267)	(0.266)	(0.268)
	$AR_{-0.5}$	0.989	0.986	0.989	0.978	0.989	0.957	0.989
		(0.271)	(0.263)	(0.272)	(0.246)	(0.272)	(0.221)	(0.272)
	$MA_{0.2}$	0.942	0.945	0.942	0.936	0.943	0.911	0.943
		(0.298)	(0.302)	(0.298)	(0.294)	(0.297)	(0.271)	(0.298)
	$MA_{0.5}$	0.924	0.931	0.925	0.926	0.924	0.903	0.924
		(0.276)	(0.285)	(0.275)	(0.282)	(0.275)	(0.263)	(0.276)
G	0	0.937	0.936	0.941	0.906	0.937	0.844	0.943
		(0.237)	(0.237)	(0.237)	(0.227)	(0.237)	(0.207)	(0.238)
	$AR_{0.2}$	0.898	0.899	0.903	0.863	0.885	0.797	0.888
		(0.233)	(0.236)	(0.233)	(0.231)	(0.233)	(0.213)	(0.233)
	$AR_{0.5}$	0.769	0.797	0.773	0.782	0.767	0.716	0.766
		(0.209)	(0.219)	(0.209)	(0.219)	(0.209)	(0.208)	(0.209)
	$AR_{-0.5}$	0.988	0.981	0.985	0.968	0.987	0.918	0.985
		(0.212)	(0.206)	(0.213)	(0.192)	(0.213)	(0.173)	(0.213)
	$MA_{0.2}$	0.900	0.903	0.901	0.875	0.899	0.813	0.911
		(0.233)	(0.236)	(0.233)	(0.230)	(0.233)	(0.212)	(0.233)
	$MA_{0.5}$	0.856	0.873	0.862	0.852	0.861	0.793	0.868
		(0.216)	(0.223)	(0.216)	(0.221)	(0.215)	(0.206)	(0.216)
G_{sub}	0	0.945	0.949	0.953	0.923	0.942	0.890	0.951
		(0.273)	(0.273)	(0.273)	(0.262)	(0.273)	(0.238)	(0.274)
	$AR_{0.2}$	0.923	0.919	0.917	0.899	0.911	0.858	0.913
		(0.268)	(0.273)	(0.268)	(0.266)	(0.268)	(0.245)	(0.269)
	$AR_{0.5}$	0.828	0.855	0.843	0.844	0.835	0.809	0.834
		(0.241)	(0.252)	(0.241)	(0.253)	(0.241)	(0.240)	(0.241)
	$AR_{-0.5}$	0.987	0.982	0.985	0.972	0.985	0.946	0.987
		(0.245)	(0.237)	(0.245)	(0.222)	(0.245)	(0.199)	(0.245)
	$MA_{0.2}$	0.918	0.928	0.927	0.903	0.918	0.868	0.926
		(0.268)	(0.272)	(0.268)	(0.265)	(0.268)	(0.244)	(0.268)
	$MA_{0.5}$	0.891	0.910	0.904	0.890	0.893	0.857	0.902
		(0.248)	(0.257)	(0.248)	(0.254)	(0.248)	(0.237)	(0.249)

Table 4: Coverage probabilities (average median interval length) with missing data ($h = 0.06$).

to the formation of ground-level ozone and it influences the lifetime of methane which classifies it as an indirect greenhouse gas. This series, which has been studied by Franco et al. (2015), is available from the Network for the Detection of Atmospheric Composition Change website at ftp://ftp.cpc.ncep.noaa.gov/ndacc/station/jungfrau/hdf/ftir/. It is argued in Franco et al. (2015) that the measurement conditions are very favorable at the Jungfraujoch location due to high dryness and low local pollution. Further details on the ground-based station and on how the measurements are obtained can be found in the aforementioned reference. It is a time series consisting of daily ethane columns (i.e. the number of molecules integrated between the ground and the top of the atmosphere) recorded under clear-sky conditions between September 1994 and August 2014 with a total of 2260 data points. Whenever more than one measurement is taken on one day, a daily mean is considered.

The average number of data points per year is 112.6 - giving an indication of the severity of the missing data problem present in this series. This shows that, in line with the above discussion, it is of major importance to use a bootstrap method which can replicate the missing data pattern correctly. The estimated transition probabilities of a first order Markov Chain reported in (14) already indicated the presence of (weak) serial dependence in the missing data generating mechanism. As a further exploration of this mechanism, note that the local constant estimator implicitly provides an estimator for the smoothly varying proportion of non-missing data $p(\tau)$. We can write (2) as

$$\hat{m}(\tau) = \hat{p}(\tau)^{-1} \frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right) D_t y_t, \qquad \text{where} \quad \hat{p}(\tau) = \frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right) D_t,$$

and $\hat{p}(\tau)$ can be seen as an estimator of $p(\tau)$.[4] In Figure 2, we plot $\hat{p}(\cdot)$ as a diagnostic tool to investigate how data availability evolves over time. It fluctuates around the average proportion of 0.3, with a maximum of almost 0.4 and a minimum of 0.2. Although no overall trend appears to be present, the fluctuations are serious enough to cast doubt on the stationarity of the missing data generating mechanism; however, our method can handle this without problems.

In addition, the data exhibit strong seasonality, as ethane degrades faster in summer than it does in winter, causing the series to displays peaks during winter and troughs during summer. Franco et al. (2015) take care of this seasonality with the help of Fourier terms by fitting the following model to ethane measurements x_t:

$$x_t = \sum_{j=1}^{M} a_j \cos(2j\pi t) + b_j \sin(2j\pi t) + y_t. \tag{15}$$

They continue their analysis with the residuals from this estimation, where $M = 3$ is selected by inspecting the residual variance. To investigate the sensitivity of the choice of M, we perform a frequency domain analysis to give more insight about the form of the periodic pattern present in our data. Due to the missing data, we use the Lomb-Scargle periodogram, which is suitable in this situation (see Lomb (1976); Scargle (1982)). Figure 3 plots the periodogram of the Jungfraujoch series with the frequency, rescaled to years, on the horizontal axis.

[4] Lemma A.4 establishes the consistency of this estimator.

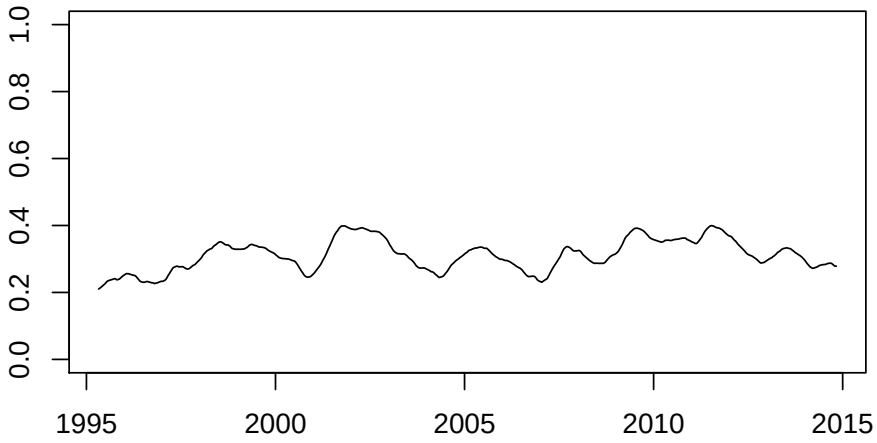
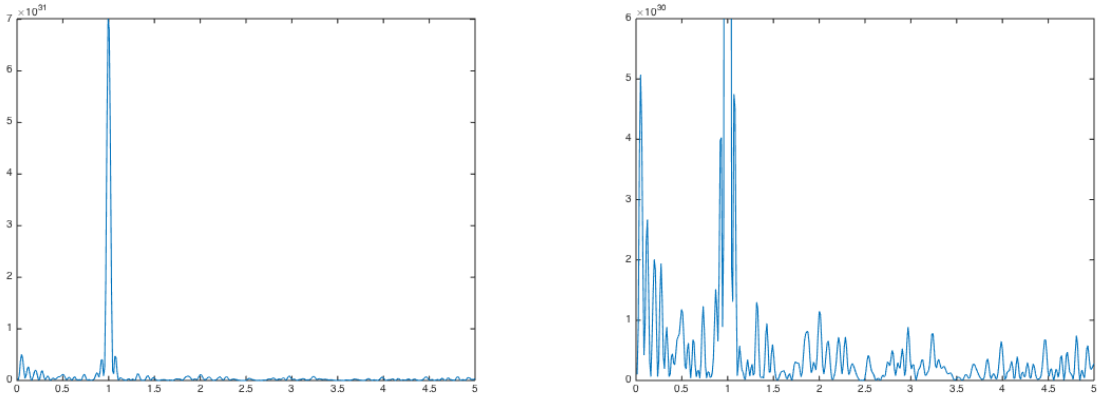


Figure 2: Time-varying proportion of non-missing data $\hat{p}$



(a) Full periodogram

(b) Zoomed version

Figure 3: The Lomb-Scargle periodogram of the ethane series

The peaks around zero are the smooth long-run trend we model with our trend estimator. The present seasonality induces additional peaks at higher frequencies. There is a large peak at 1, representing an annual periodicity, which is so pronounced that it obscures peaks at other frequencies. Therefore, the right panel displays the same spectrum as the left panel, but with a smaller vertical axis such that other peaks are observed more clearly. Moreover, we can observe that there are peaks at 2 and 3. They are, however, not as clear-cut as the peak at 1 and might not contribute as much to the seasonality, yet provide further justification for the choice $M = 3$. In Supplementary Appendix D, we consider the periodograms of the residuals of the regression on 1 up to 4 Fourier terms. These show that while inclusion of one term is clearly needed, including more terms does indeed further reduce periodicity, although for increasing M, the effect becomes less pronounced.

To corroborate our results and to be able to compare our findings to Franco et al. (2015), we additionally look at the Akaike and Bayesian information criteria as well as the residual variance

M	AIC	BIC	MSE
1	158447.5	158464.6	2.79909×10^{30}
2	158420.0	158448.6*	2.76038×10^{30}
3	158413.4	158453.5	2.74750×10^{30}
4	158411.3*	158462.8	2.74010×10^{30}
5	158411.3*	158474.3	2.73521×10^{30}
6	158413.7	158488.1	2.73329×10^{30}
7	158416.8	158502.7	2.73219×10^{30}

Table 5: Fourier term investigation

(MSE in Franco et al. (2015)) from the above regression for different values of M. The results are summarized in Table 5. While the Akaike criterion (AIC) is indifferent between adding 4 or 5 Fourier terms, BIC selects 2. The residual variance (MSE) decreases by only small increments when more than 3 terms are included. Based on our analysis, it is not clear how many Fourier terms we should include; any value of M between 1 and 4 seems reasonable. In the following, we report results for applying the trend estimation on the residuals of the regression with $M = 3$ Fourier terms, as in Franco et al. (2015). In Supplementary Appendix D we perform the same analysis with M varying between 1 and 4, with hardly any difference in the results.

We next investigate bandwidth selection. As suggested in Section 3.1, we determine a possible bandwidth using modified cross-validation. In line with the discussion in Chu and Marron (1991), for our series, the ordinary leave-one-out cross-validation criterion selects a bandwidth which is too small ($h_{cv} = 0.0006$). This value for the bandwidth parameter gives almost no smoothing of the data and the resulting trend curve is too wiggly. Leaving out $k = 5$ observations on each side of any point, the modified criterion yields a value of $h_{mcv} = 0.03$. Albeit a still small bandwidth, this value gives a much more reasonable picture of the trend estimate. The resulting nonparametric estimate as well as 95% simultaneous confidence bands are depicted in Figure 4. The confidence bands are simultaneous over the whole sample. Although the validity has not been established, the algorithm works when we cover the whole sample and the results are easier to interpret. The bands are obtained using $B = 999$ replications of the bootstrap procedure and an autoregressive parameter of $\gamma = 0.5$.

As a robustness check, we also applied the trend estimation to the data without explicitly modeling the seasonality using Fourier terms. The nonparametric kernel estimator can be interpreted as a low pass filter which suppresses high frequency oscillations. A sufficiently large bandwidth should introduce enough smoothing to provide a trend curve which is not driven by the seasonal component. The bandwidth selected by MCV is too small for this effect to appear in our data. When we increase the bandwidth to $h = 0.06$, the resulting trend shows the same pattern as the one in Figure 4. This analysis shows that when strong seasonality is present in the data, the nonparametric kernel estimator can be used to filter out the seasonality if the bandwidth is large enough. In this case, however, bandwidth selection becomes a critical issue and the proposed MCV criterion should be applied with care. Further details can be found in Supplementary Appendix D.

We observe a slight downward trend of the ethane time series until around 2009, with local

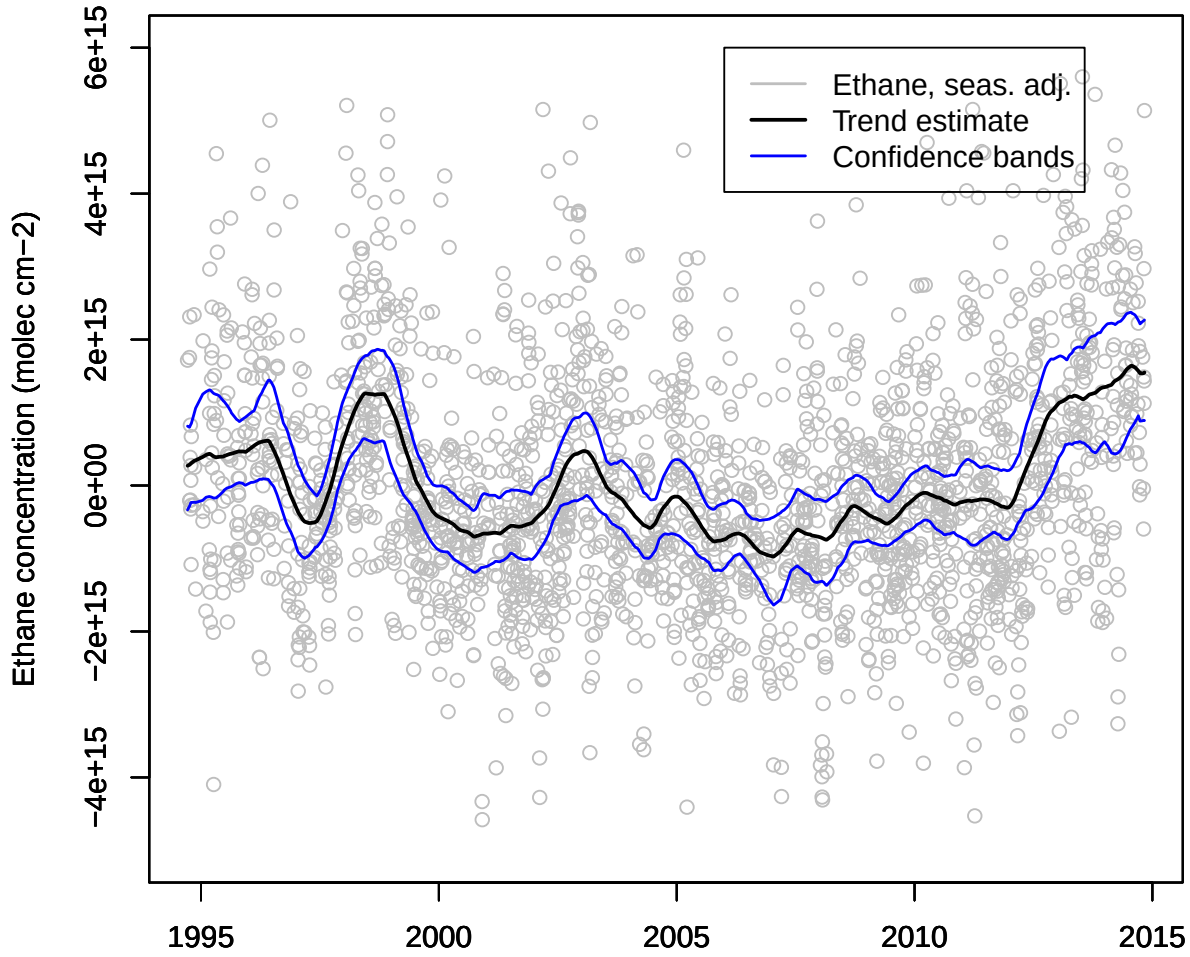


Figure 4: Trend estimate with uniform 95% confidence bands for ethane concentration at the Jungfraujoch (with bandwidth selected by MCV).

peaks in 1998 and 2002-2003, and an upward trend thereafter. This general development of the trend supports the findings in Franco et al. (2015) who estimate a linear trend model with a break at the beginning of 2009. They find a negative slope of the trend line before the break and a positive slope after the break. As mentioned by Franco et al. (2015), the initial downward trend can be explained by a general emission reduction since the mid 1980's, of the fossil fuel sources in the Northern Hemisphere. This has also been reported by Simpson et al. (2012). The upward trend seems to be a more recent phenomenon. Studies attribute it to the recent growth in the exploitation of shale gas and tight oil reservoirs, taking place in North America, see e.g. Vinciguerra (2015) and Franco et al. (2016). Since previous studies have mainly used methods based on linear trends, the two local peaks have to our knowledge not yet been analyzed. They can potentially be explained by boreal forest fires which were taking place mainly in Russia during both periods. Geophysical studies have investigated these events in association with anomalies in carbon monoxide emissions (Yurganov et al., 2004, 2005). In such fires, carbon monoxide is co-emitted with ethane, such that these events are likely explanations for the peaks we observe.

As a final step, we look at the standard deviation of the residuals. When estimating it with a nonparametric kernel smoother, we see a cyclical pattern with upward trend, similar to the process we generate in our simulations. We plot the estimated standard deviation in Figure 5. This clearly shows that the residuals are heteroskedastic which further underlines the importance of a flexible bootstrap method.

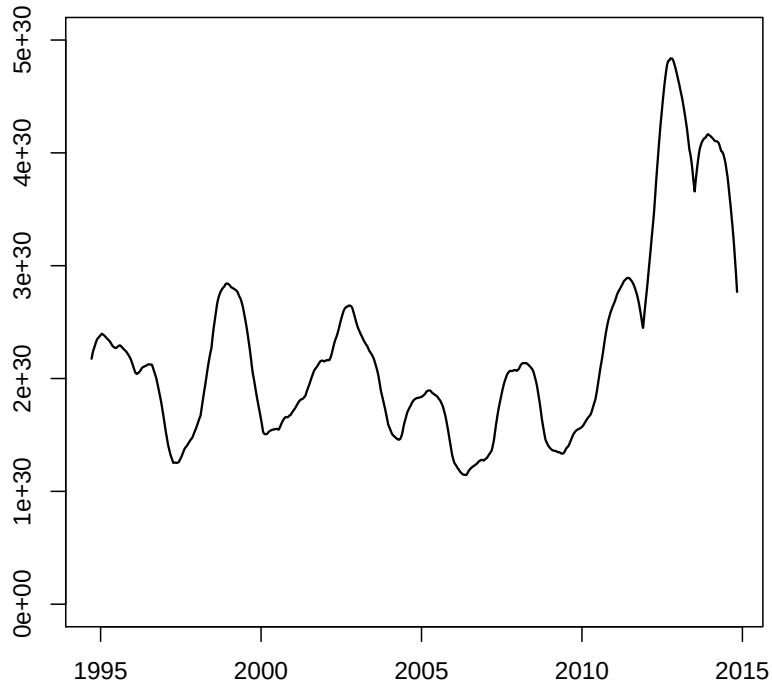


Figure 5: Estimate of the standard deviation of the residuals, obtained using the local constant kernel smoother with Epanechnikov kernel and $h = 0.04$

7 Conclusion

In this paper we have proposed a dependent version of the wild bootstrap, the autoregressive wild bootstrap, to construct confidence intervals around a nonparametrically estimated trend. Consistency of the bootstrap has been established such that it can be used to construct pointwise and simultaneous confidence bands. While the pointwise intervals always show good coverage in finite samples, simulation results for the simultaneous bands indicate that strong positive autocorrelation leads to a drop in coverage whenever simultaneous confidence bands are considered. However, other bootstrap methods such as the dependent wild bootstrap and the sieve bootstrap are equally affected, and overall the autoregressive wild bootstrap performs at least on par with these other methods, and often outperforms them.

One major advantage of the proposed approach is its broad applicability as it can be used under general forms of serial dependence and heteroskedasticity. Furthermore, it can be applied without further adjustments when data points are missing. This feature of the autoregressive wild bootstrap

is particularly relevant in economic and climatological applications where the problem of missing data is often encountered. In addition to simulation results, we provide a rigorous asymptotic analysis where asymptotically pervasive missing data are allowed for. While the missing data generating mechanism affects the asymptotic distribution of the estimator, our bootstrap method correctly mimics this and is therefore valid in the presence of general forms of missing data patterns.

An application to atmospheric ethane measurements from Switzerland demonstrates our methodology. An upward trend in this time series is an indication of increasing atmospheric pollution and it has been visible in the data for the last quarter. This finding is in line with previous studies in geophysics and provides further evidence that an increased activity in shale gas extraction might have caused an increase in the ethane burden measured over the Jungfraujoch. In addition, we find two local peaks in the ethane series, which can be explained by boreal forest fires. Natural limitations of linear trend estimation have prevented these peaks from being discovered in previous research. This underlines the flexibility of our approach compared to parametric methods.

An open end to our analysis is the choice of the autoregressive parameter in the autoregressive wild bootstrap. Although our simulation results suggest that a range of values for this parameter perform adequately, its selection in practice remains an open issue. Theoretical results on the choice of this parameter are not trivial; moreover, such theoretical results do not translate directly into selection methods with good properties in small samples. This issue therefore merits deeper study and is left as an exercise for future research.

References

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

Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. Annals of Statistics 26, 48-83.

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

Chu, C.-K. and J. S. Marron (1991). Comparison of two bandwidths selectors with dependent errors. Annals of Statistics 19, 1906-1918.

Davidson, J. (2002). Stochastic Limit Theory (2nd edition). Oxford: Oxford University Press.

Davidson, R. and E. Flachaire (2008). The wild bootstrap, tamed at last. Journal of Econometrics 146, 162-169.

Fan, J. (1992). Design-adaptive nonparametric regression. Journal of the American Statistical Association 87, 998-1004.

Franco, B., Bader W., Toon G.C., Bray C., Perrin A., Fischer E.V., Sudo K., Boone C.D., Bovya B., Lejeune B., Servais C. and E. Mahieu (2015). Retrieval of ethane from ground-based FTIR solar spectra using improved spectroscopy: Recent burden increase above Jungfraujoch. Journal of Quantitative Spectroscopy and Radiative Transfer 160, 36-49.

Franco, B., Mahieu, E., Emmons, L. K., Tzompa-Sosa, Z. A., Fischer, E. V., Sudo, K., Bovy, B., Conway, S., Griffin, D., Hannigan, J. W., Strong, K. and K. A. Walker (2016). Evaluating ethane and methane emissions associated with the development of oil and natural gas extraction in North America. Environmental Research Letters, 11, 044010.

Franses, P. H. and T. J. Vogelsang (2005). Are winters getting warmer? Environmental Modeling and Software 20, 1449-1455.

Gasser, T. and H. G. Müller (1984). Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics 11, 171-185.

Giné, E. and J. Zinn (1990). Bootstrapping general empirical measures. Annals of Probability 18, 851-869.

Gonzalés, A. and T. Teräsvirta (2008). Modelling autoregressive processes with a shifting mean. Studies in Nonlinear Dynamics & Econometrics 12, 1-26.

Gregory, K. B., Lahiri, S. N., and D. J. Nordman (2015). A Smooth Block Bootstrap for Statistical Functionals and Time Series. Journal of Time Series Analysis 36, 442-461.

Gregory, K. B., Lahiri, S. N., and D. J. Nordman (2015). A smooth block bootstrap for quantile regression with time series. Annals of Statistics 46, 1138-1166.

Hall, P. and J. Horowitz (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. Annals of Statistics 41, 1892-1921.

Härdle, W. and J. S. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression. Annals of Statistics 19, 778-796.

Hansen, B. E. (1991). Strong laws for dependent heterogeneous processes. Econometric Theory 7, 213-221.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. Econometric Theory 18, 1449-1459.

Leucht, A. and M. H. Neumann (2013). Dependent wild bootstrap for degenerate U- and V-statistics. Journal of Multivariate Analysis 117, 257-280.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. Astrophysics and Space Science 39, 447-462.

McKitrick, R. and T. J. Vogelsang (2014). HAC robust trend comparisons among climate series with possible level shifts. Environmetrics 25, 528-547.

McLeish, D. L. (1975). A maximal inequality and dependent strong laws. Annals of Probability 3, 829-839.

Nadaraya, E. A. (1964). On estimating regression. Theory of Probability & Its Applications 9, 141-142.

Neumann, M. H. (1998). Pointwise confidence intervals in nonparametric regression with heteroskedastic error structure. Statistics 29, 1-36.

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

Norris, J. R., Allen, R. J., Evan, A. T., Zelinka, M. D., O'Dell, C. W. and S. A. Klein (2016). Evidence for climate change in the satellite cloud record. Nature 536, 72-75.

Palm, F. C., Smeekes, S. and J. P. Urbain (2011). Cross-sectional dependence robust block bootstrap panel unit root tests. Journal of Econometrics 163, 85-104.

Paparoditis, E. and D. Politis (2002). The tapered block bootstrap for general statistics from stationary sequences. Econometrics Journal 5, 131-148.

Politis, D. N., J. P. Romano, and M. Wolf (1999). Subsampling. New York: Springer.

Scargle, J. D. (1982). Studies in Astronomical Time Series Analysis II. Statistical Aspects of Spectral Analysis of Unevenly Spaced Data. The Astrophysical Journal 263, 835-853.

Shao, X. (2010a). Extended tapered block bootstrap. Statistica Sinica 20, 807-821.

Shao, X. (2010b). The dependent wild bootstrap. Journal of the American Statistical Association 105, 218-235.

Simpson, I. J., Sulbaek Andersen, M. P., Meinardi, S., Bruhwiler, L., Blake, N. J., Helmig, D., Rowland, F. S. and D. R. Blake (2012). Longterm decline of global atmospheric ethane concentrations and implications for methane. Nature 488, 490-494.

Smeekes, S. and J.-P. Urbain (2014). A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing. GSBE Research Memorandum, RM/14/008, Maastricht University.

Vinciguerra, T., Yao, S., Dadzie, J., Chittams, A., Deskins, T., Ehrman, S. and R. R. Dickerson (2015). Regional air quality impacts of hydraulic fracturing and shale natural gas activity: evidence from ambient VOC observations. Atmospheric Environment 110, 144-50.

Watson, G. S. (1964). Smooth regression analysis. Sankhyā Ser. A 26, 359-372.

Wu, W. B. and Z. Zhao (2007). Inference of trends in time series. Journal of the Royal Statistical Society Series B 69, 391-410.

Yurganov, L. N., Blumenstock, T., Grechko, E. I., Hase, F., Hyer, E. J., Kasischke, I. S., Koike, M., Kondo, Y., Kramer, I., Leung, F.-Y., Mahieu, E., Mellqvist, J., Notholt, J., Novelli, P. C., Rinsland, C. P., Scheel, H. E., Schulz, A., Strandberg, A., Sussmann, R., Tanimoto, H., Velazco,

V., Zander, R. and Y. Zhao (2004). A quantitative assessment of the 1998 carbon monoxide emission anomaly in the Northern Hemisphere based on total column and surface concentration measurements. Journal of Geophysical Research 109, D15305.

Yurganov, L. N., Duchatelet, P., Dzhola, A. V., Edwards, D. P., Hase, F., Kramer, I., Mahieu, E., Mellqvist, J., Notholt, J., Novelli, P. C., Rockmann, A., Scheel, H. E., Schneider, M., Schulz, A., Strandberg, A., Sussmann, R., Tanimoto, H., Velazco, V., Drummond, J. R. and J. C. Gille (2005). Increased Northern Hemispheric carbon monoxide burden in the troposphere in 2002 and 2003 detected from the ground and from space. Atmospheric Chemistry and Physics, 5, 563-573.

Appendix A Technical Results

For the remainder of this appendix, we define some further notation that will help lighten the notational load. Define $k_t(\tau) = K\left(\frac{t/n-\tau}{h}\right)$ and $\tilde{k}_t(\tau) = K\left(\frac{t/n-\tau}{\tilde{h}}\right)$. For any random variable X, let $\mathbb{E}_D X = \mathbb{E}(X|\{D_t\}_{t=1}^n)$ and $\mathbb{V}\text{ar}_D X = \mathbb{E}_D[X - (\mathbb{E}_D X)^2]^2$. Furthermore, let $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ denote the L_p-norm of X, where we use $\|X\|$ as short-hand notation for $\|X\|_2$, and define he bootstrap equivalent as $\|X^*\|_p^* = (\mathbb{E}^*|X^*|^p)^{1/p}$. Finally, $C, C_1, C_2, \ldots$ denote arbitrary fixed positive constants not depending on n or h, whose value can change each line, while ϕ_n denotes a generic deterministic sequence with the property that $\lim_{n\to\infty}\phi_n = 0$.

A.1 Auxiliary lemmas

We first establish some auxiliary lemmas with basic results that will be needed in the proofs of our main results. The proofs of these auxiliary lemmas are relegated to Supplementary Appendix B.

Lemma A.1. *Let Assumptions 3-4 hold. Then for all $n \geq 1$,*

$$(i)\ \sum_{i=0}^{n-1} i \max_{1\leq t\leq n-i} \mathbb{C}\text{ov}(D_t, D_{t+i}) \leq C_1;$$
$$(ii)\ \sum_{i=0}^{n-1} i \max_{1\leq t\leq n-i} \mathbb{C}\text{ov}(u_t, u_{t+i}) \leq C_2;$$
$$(iii)\ \sum_{i=0}^{n-1} i \max_{1\leq t\leq n-i} \mathbb{C}\text{ov}(D_t u_t, D_{t+i} u_{t+i}) \leq C_3.$$

Proof of Lemma A.1. See Supplementary Appendix B. □

Lemma A.2. *Under Assumption 5, we have for any $\delta > 0$ that*

$$\sup_{\tau\in[\delta,1\text{-}\delta]} \max_{0\leq i\leq n} \frac{1}{nh} \sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) < \infty.$$

Proof of Lemma A.2. See Supplementary Appendix B. □

Lemma A.3. *Let $f_n(\cdot) : [0,1] \to \mathbb{R}$ satisfy Assumption 1, and let $g(\cdot,\cdot) : [0,1]^2 \to \mathbb{R}^2$ be a Lipschitz continuous function with Lipschitz constant $C_g = \sup_{(\tau_1,\tau_2)\in[0,1]^2} |g_n(\tau_1,\tau_2)|$. Let Assumption 5 be satisfied. Then, there exists some $N > 0$, such that for any $0 < \delta < 1/2$, all $n > N$ and all $i = 0, 1, \ldots, n$, we have that*

$$(i) \sup_{\tau\in[\delta,1\text{-}\delta]} \left| \frac{1}{nh} \sum_{t=1}^{n} f\left(\frac{t}{n}\right) k_t(\tau) - f(\tau)\kappa_1 - \frac{1}{2}h^2 f^{(2)}(\tau)\mu_2 \right| \leq C \max\left\{h^3, \frac{1}{nh}\right\};$$

$$(ii) \sup_{\tau_0\in[\delta,1\text{-}\delta]} \sup_{\tau_1,\tau_2\in[\text{-}1,1]} \left| \frac{1}{nh} \sum_{t=1}^{n-i} g\left(\frac{t}{n}, \frac{t+i}{n}\right) k_t(\tau_0+\tau_1 h) k_{t+i}(\tau_0+\tau_2 h) - g(\tau_0,\tau_0)\kappa(\tau_1-\tau_2) \right|$$
$$\leq C_g \left[\frac{C_1 i}{nh} + C_2 \max\left\{h, \frac{1}{nh^2}\right\} + S_n(i) \right],$$

where, for any sequence $\{\beta_i\}_{i=0}^{\infty}$ with $\sum_{i=1}^{\infty} |\beta_i| < \infty$, we have that $\sum_{i=1}^{\infty} |\beta_i| S_{n,i} \leq \phi_n$, where $\lim_{n\to\infty} \phi_n = 0$.

Proof of Lemma A.3. See Supplementary Appendix B. □

Lemma A.4. *Let Assumptions 2-5 be satisfied. Define*

$$Z_{n,D,f,q}(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} k_t(\tau) \left[f\left(\frac{t}{n}\right) - q_n(\tau) \right] \left[D_t - p\left(\frac{t}{n}\right) \right], \tag{A.1}$$

where $f(\cdot)$ satisfies Assumption 1 and $q_n(\cdot) : [0,1] \to \mathbb{R}$ is a sequence of deterministic functions with $\sup_{\tau\in[0,1]} |q_n(\tau)| < \infty$ for all n. Then, there exists some $N > 0$, such that for any $0 < \delta < 1/2$ and all $n > N$ the following hold:

(i) $\mathbb{E}\, Z_{n,D,f,q}(\tau) = 0$ *and*

$$\sup_{\tau\in[\delta,1\text{-}\delta]} \left| \mathbb{E}\, Z_{n,D,f,l}(\tau)^2 - [f(\tau) - q_n(\tau)]^2 \Omega_D(\tau)\kappa_2 \right| \leq C \max\left\{h, \frac{1}{nh^2}, \phi_n\right\},$$

where $\Omega_D(\tau) = R_{D,0}(\tau,\tau) + 2\sum_{i=1}^{\infty} R_{D,i}(\tau,\tau)$ and $\lim_{n\to\infty} \phi_n = 0$.

(ii) For all $\tau \in (0,1)$,

$$\frac{1}{nh} \sum_{t=1}^{n} f\left(\frac{t}{n}\right) k_t(\tau) D_t - p(\tau) f(\tau) - \frac{1}{2}h^2 [fp]^{(2)}(\tau)\mu_2 = \frac{Z_{n,D,f,0}(\tau)}{\sqrt{nh}} + R_{n,D}(\tau),$$

where $Z_{n,D,f,0}(\tau)$ is defined as in (A.1) *with $q_n(\cdot) = 0$ and $\sup_{\tau\in[\delta,1\text{-}\delta]} |R_{n,D}(\tau)| < C \max\left\{h^3, \frac{1}{nh}\right\}$.*

(iii) Let $\tilde{g}_i(\cdot,\cdot)$ be a Lipschitz continuous function with Lipschitz constant $C_g(i) = \sup_{(\tau_1,\tau_2)\in[0,1]^2} |\tilde{g}_i(\tau_1,\tau_2)|$.

Then for all $i = 0, 1, \ldots, n-1$,

$$\sup_{\tau\in[\delta,1\text{-}\delta]} \sup_{\tau_1,\tau_2\in[\text{-}1,1]} \mathbb{E}\left| \frac{1}{nh}\sum_{t=1}^{n} \tilde{g}_i\left(\frac{t}{n}, \frac{t+i}{n}\right) k_t(\tau_0+\tau_1 h) k_{t+i}(\tau_0+\tau_2 h) D_t D_{t+i} - p(\tau_0)\tilde{g}(\tau_0,\tau_0)\kappa(\tau_1-\tau_2)\right|$$
$$\leq C_g(i)\left[\frac{C_1}{\sqrt{nh}} + C_2 \max\left\{h, \frac{1}{nh^2}\right\} + S_n(i)\right],$$

where $S_n(i)$ *is defined in Lemma A.3.*

Proof of Lemma A.4. See Supplementary Appendix B. □

Lemma A.5. *Let Assumptions 1-6 hold. Then, for all* $0 \leq i \leq n-1$ *and any* $0 < \delta < 1/2$,

$$\sup_{\tau\in[\delta,1\text{-}\delta]} \mathbb{E}\left| \frac{1}{nh}\sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) \sigma_t \sigma_{t+i} D_t D_{t+i} \left[u_t u_{t+i} - \mathbb{E}\, u_t u_{t+i}\right]\right| \leq \beta_i \phi_n + \frac{\eta_n}{\sqrt{nh}},$$

where $\sum_{i=1}^{\infty} \beta_i < \infty$, $\lim_{n\to\infty} \phi_n = 0$ *and* $\limsup_{n\to\infty} \eta_n < \infty$.

Proof of Lemma A.5. See Supplementary Appendix B. □

A.2 Pointwise Results

Lemmas A.6 and A.7 decompose the (bootstrap) estimator in the relevant components, establishing not only consistency but also providing the necessary building blocks towards asymptotic normality. We therefore present these lemmas and their proofs together with the proofs of Theorems 1 and 2.

Lemma A.6. *Let Assumptions 1-6 be satisfied. Then, for some* $N > 0$ *and any* $0 < \delta < 1/2$, *we have for all* $\tau \in (0,1)$ *and* $n > N$ *that*

$$\sqrt{nh}\left[\hat{m}(\tau) - m(\tau) - h^2 B_{as}(\tau)\right] = p(\tau)^{-1} Z_{n,U}(\tau) + R_n(\tau),$$

where $B_{as}(\tau)$ *is defined in* (8),

$$Z_{n,U}(\tau) = \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} k_t(\tau) D_t \sigma_t u_t, \tag{A.2a}$$

$$\sigma_Z^2(\tau) = \operatorname*{plim}_{n\to\infty} \mathbb{V}\mathrm{ar}\left[Z_{n,U}(\tau) \,|\{D_t\}_{t=1}^n\right] = p(\tau)\sigma(\tau)^2 \Omega_U \kappa_2, \tag{A.2b}$$

$$\sup_{\tau\in[\delta,1\text{-}\delta]} \|R_n(\tau)\| \leq C \max\left\{h^2, \sqrt{nh^7}, \frac{1}{\sqrt{nh}}\right\}. \tag{A.2c}$$

Proof of Lemma A.6. Defining

$$\hat{p}(\tau) = \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau) D_t, \qquad \breve{m}(\tau) = \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau) D_t y_t,$$
$$\overline{m}_D(\tau) = \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau) D_t m(t/n) \qquad \overline{m}_p(\tau) = \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau) p(t/n) m(t/n),$$

and realizing that $Z_{n,U}(\tau) = \sqrt{nh}\,[\breve{m}(\tau) - \overline{m}_D(\tau)]$, we can write

$$\begin{aligned}\sqrt{nh}\left[\hat{m}(\tau) - m(\tau) - h^2 B_{as}(\tau)\right] &= p(\tau)^{-1} Z_{n,U}(\tau) + [\hat{p}(\tau)^{-1} - p(\tau)^{-1}] Z_{n,U}(\tau)/\sqrt{nh} \\ &+ \sqrt{nh}\left\{\hat{p}(\tau)^{-1}\overline{m}_D(\tau) - p(\tau)^{-1}\overline{m}_p(\tau)\right\} + \sqrt{nh}\left\{p(\tau)^{-1}\overline{m}_p(\tau) - m(\tau) - h^2 B_{as}(\tau)\right\} \qquad \text{(A.3)} \\ &= p(\tau)^{-1} Z_{n,U}(\tau) + I_n(\tau) + II_n(\tau) + III_n(\tau).\end{aligned}$$

We first derive $\sigma_Z^2(\tau)$. Let $\mathbb{E}_D(\cdot) = \mathbb{E}(\cdot|\{D_t\}_{t=1}^n)$. Then $\mathbb{E}_D\, Z_{n,U}(\tau) = 0$ and

$$\begin{aligned}\mathbb{E}_D\, Z_{n,U}(\tau)^2 &= \frac{1}{nh}\sum_{s=1}^{n}\sum_{t=1}^{n} k_s(\tau) k_t(\tau) D_s D_t \sigma_s \sigma_t\, \mathbb{E}\, u_s u_t \\ &= \frac{1}{nh}\sum_{i=-n+1}^{n-1} R_U(i) \sum_{t=1}^{n-|i|} k_t(\tau) k_{t+|i|}(\tau) \sigma_t \sigma_{t+|i|} D_t D_{t+|i|}.\end{aligned}$$

Using Lemma A.4(iii) with $\tilde{g}_i(\tau_1, \tau_2) = \sigma(\tau_1)\sigma(\tau_2)$, we find that

$$\begin{aligned}&\sup_{\tau\in[\delta,1\text{-}\delta]} \mathbb{E}\left|\sum_{i=-n+1}^{n-1} R_U(i) \frac{1}{nh}\sum_{t=1}^{n-|i|} k_t(\tau) k_{t+|i|}(\tau)\sigma_t\sigma_{t+|i|} D_t D_{t+|i|} - p(\tau)\sigma(\tau)^2\kappa_2\Omega_U\right| \\ &\leq 2\sum_{i=0}^{n-1} |R_U(i)| \left[\frac{C_1}{\sqrt{nh}} + C_2 \max\left\{h, \frac{1}{nh^2}\right\} + S_n(i)\right] + 2\sum_{i=n}^{\infty} |R_U(i)|\,, \\ &\leq \frac{C_3}{\sqrt{nh}} + C_4 \max\left\{h, \frac{1}{nh^2}\right\} + C_5\phi_n.\end{aligned}$$

where $\lim_{n\to\infty}\phi_n = 0$, from which it follows that, for $\tau \in (0,1)$, $\mathbb{E}\left|\mathbb{E}_D\, Z_{n,U}(\tau)^2 - \sigma_Z^2(\tau)\right| \leq C\max\{(nh)^{-1/2}, h, n^{-1}h^2, \phi_n\} = o(1)$, where $\sigma_Z^2(\tau)$ is defined in (A.2b).

Furthermore, by the arguments used above it follows that $\sup_{\tau\in[\delta,1\text{-}\delta]} \|Z_{n,U}(\tau)\| \leq C$, while by Lemma A.4($i$)

$$\|\hat{p}(\tau) - p(\tau)\| \leq \frac{1}{\sqrt{nh}}\|Z_{n,D}(\tau)\| + C_1 h^2 + C_2 \max\left\{h^3, \frac{1}{nh}\right\} \leq C_3 \max\left\{\frac{1}{\sqrt{nh}}, h^2\right\}.$$

By combining these results with the fact that $\|\hat{p}(\tau)\|^{-1} \leq 1/\epsilon^*$ by Assumption 6, we find that

$$\sup_{\tau\in[\delta,1\text{-}\delta]} \|I_n(\tau)\| \leq \sup_{\tau\in[\delta,1\text{-}\delta]} \|\hat{p}(\tau) - p(\tau)\| \left\|\hat{p}(\tau)^{-2}\right\| \left|p(\tau)^{-2}\right| \|Z_{n,U}(\tau)\| \leq C_4 \max\left\{\frac{1}{\sqrt{nh}}, h^2\right\}.$$

For $II_n(\tau)$ we note that

$$\hat{p}(\tau)^{-1} - p(\tau)^{-1} = p(\tau)^{-2}\{p(\tau) - \hat{p}(\tau)\} + \hat{p}(\tau)^{-1}p(\tau)^{-2}\{p(\tau) - \hat{p}(\tau)\}^2, \tag{A.4}$$

such that we can rewrite $II_n(\tau)$ as

$$\begin{aligned} II_n(\tau)/\sqrt{nh} &= p(\tau)^{-1}\{\overline{m}_D(\tau) - \overline{m}_p(\tau)\} + \left\{\hat{p}(\tau)^{-1} - p(\tau)^{-1}\right\}\overline{m}_p(\tau) \\ &\quad + \left\{\hat{p}(\tau)^{-1} - p(\tau)^{-1}\right\}\{\overline{m}_D(\tau) - \overline{m}_p(\tau)\} \\ &= p(\tau)^{-1}\{\overline{m}_D(\tau) - \overline{m}_p(\tau)\} - p(\tau)^{-2}\{\hat{p}(\tau) - p(\tau)\}\overline{m}_p(\tau) \\ &\quad + \hat{p}(\tau)^{-1}p(\tau)^{-2}\{\hat{p}(\tau) - p(\tau)\}^2\overline{m}_p(\tau) + \left\{\hat{p}(\tau)^{-1} - p(\tau)^{-1}\right\}\{\overline{m}_D(\tau) - \overline{m}_p(\tau)\} \\ &= [II_{n,11}(\tau) - II_{n,12}(\tau) + II_{n,2}(\tau) + II_{n,3}(\tau)]/\sqrt{nh}. \end{aligned}$$

As $II_{n,11}(\tau) = p(\tau)^{-1}\frac{1}{\sqrt{nh}}\sum_{t=1}^n k_t(\tau)m(t/n)[D_t - p(t/n)]$ and $II_{n,12}(\tau) = p(\tau)^{-2}\overline{m}_p(\tau)\frac{1}{\sqrt{nh}}\sum_{t=1}^n k_t(\tau)$ $\times [D_t - p(t/n)]$, we can write

$$\begin{aligned} II_{n,1}(\tau) &= II_{n,11}(\tau) - II_{n,12}(\tau) = p(\tau)^{-1}\frac{1}{\sqrt{nh}}\sum_{t=1}^n k_t(\tau)\left[m(t/n) - p(\tau)^{-1}\overline{m}_p(\tau)\right][D_t - p(t/n)] \\ &= p(\tau)^{-1}Z_{n,D}(\tau), \end{aligned}$$

where $Z_{n,D}(\tau)$ is defined as $Z_{n,D,f,q}(\tau)$ in Lemma A.4(i) with $f(\cdot) = m(\cdot)$ and $q_n(\cdot) = p(\cdot)^{-1}\overline{m}_p(\cdot)$. Applying this lemma we find that

$$\sup_{\tau\in[\delta,1\text{-}\delta]}\left|\mathbb{E}\,Z_{n,D}(\tau)^2 - \left[m(\tau) - p(\tau)^{-1}\overline{m}_p(\tau)\right]^2\Omega_D(\tau)\kappa_2\right| \le C\max\{h, n^{-1}h^2\} + \phi_n.$$

As $|x^2 - y^2| \le |x - y|(|x - y| + 2|y|)$, it follows from Lemma A.3(i) that

$$\begin{aligned} \left|\left[m(\tau) - p(\tau)^{-1}\overline{m}_p(\tau)\right]^2 - h^4B_{as}(\tau)^2\right| &\le C\max\{h^3, (nh)^{-1}\}\left[C\max\{h^3, (nh)^{-1}\} + 2h^2B_{as}(\tau)\right] \\ &\le C_1\max\{h^5, n^{-1}h, (nh)^{-2}\}, \end{aligned}$$

and therefore, with $\sigma_D^2(\tau) = h^4B_{as}(\tau)^2\Omega_D(\tau)\kappa_2$,

$$\sup_{\tau\in[\delta,1\text{-}\delta]}\left|\mathbb{E}\,Z_{n,D}(\tau)^2 - \sigma_D^2(\tau)\right| \le C_1\max\{h^5, n^{-1}h, (nh)^{-2}\} + C_2\phi_n.$$

It then follows directly that

$$\sup_{\tau\in[\delta,1\text{-}\delta]}\|II_{n,1}(\tau)\| \le \frac{1}{\epsilon^*}\left(\sup_{\tau\in[\delta,1\text{-}\delta]}\left|\mathbb{E}\,Z_{n,D}(\tau)^2 - \sigma_D^2(\tau)\right|^{1/2} + \sup_{\tau\in[\delta,1\text{-}\delta]}|\sigma_D(\tau)|\right) \le C\max\left\{h^2, \sqrt{\frac{h}{n}}, \frac{1}{nh}\right\}.$$

As $\sup_{\tau\in[\delta,1\text{-}\delta]}\|\hat{p}(\tau) - p(\tau)\| \le C\max\left\{h^2, \frac{1}{\sqrt{nh}}\right\}$, it follows that $\sup_{\tau\in[\delta,1\text{-}\delta]}\|II_{n,2}(\tau)\| \le C\max\left\{\sqrt{nh^9}, \frac{1}{\sqrt{nh}}\right\}$. Furthermore, by Lemma A.4(i), $\|\overline{m}_D(\tau) - \overline{m}_p(\tau)\| \le \frac{1}{\sqrt{nh}}\|Z_{n,D,m,0}(\tau)\| \le \frac{C}{\sqrt{nh}}$, such that it also

follows that

$$\sup_{\tau\in[\delta,1\text{-}\delta]} \|II_{n,3}(\tau)\| \leq C \sup_{\tau\in[\delta,1\text{-}\delta]} \|\hat{p}(\tau) - p(\tau)\| \left\|\overline{m}_{|D}(\tau) - \overline{m}_p(\tau)\right\| \leq C \max\left\{h^2, \frac{1}{\sqrt{nh}}\right\}.$$

Finally, it follows directly from Lemma A.3(i) that $\sup_{\tau\in[\delta,1\text{-}\delta]} |III_n(\tau)| \leq C \max\left\{\sqrt{nh^7}, \frac{1}{\sqrt{nh}}\right\}$. Collecting all remainder terms in $R_n(\tau) = I_{n,1}(\tau) + II_{n,2}(\tau) + II_{n,3}(\tau) + III_n(\tau)$, we find that $\sup_{\tau\in[\delta,1\text{-}\delta]} \|R_n(\tau)\| \leq C \max\left\{h^2, \sqrt{nh^7}, \frac{1}{\sqrt{nh}}\right\}$. □

Proof of Theorem 1. Given Lemma A.6, we only have to prove asymptotic normality of $Z_{n,U}(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^n k_t(\tau) D_t \sigma_t u_t$. To simplify the proofs, we first condition on $\{D_t\}_{t=1}^n$ and thus prove conditional asymptotic normality of $Z_{n,U}(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^n k_t(\tau) D_t \sigma_t u_t$. As before, let $\mathbb{E}_D(\cdot) = \mathbb{E}(\cdot|\{D_t\}_{t=1}^\infty)$. As the limit results do not depend on $\{D_t\}_{t=1}^n$, the results then directly hold unconditionally as well.

Take an M such that $M \to \infty$ as $n \to \infty$, and truncate the MA(∞) representation of u_t at M lags, say $u_{t,M} = \sum_{j=1}^M \psi_j \epsilon_{t-j}$. Then we can write

$$\begin{aligned} Z_n(\tau) &= \frac{1}{\sqrt{nh}} \sum_{t=1}^n k_t(\tau) D_t \sigma_t u_{t,M} + \frac{1}{\sqrt{nh}} \sum_{t=1}^n k_t(\tau) D_t \sigma_t \sum_{j=M+1}^\infty \psi_j \epsilon_{t-j} \\ &= \overline{Z}_{n,M}(\tau) + \overline{W}_{n,M}(\tau). \end{aligned}$$

Let $R_W(k) = \mathbb{E}\left(\sum_{j=M+1}^\infty \psi_j \epsilon_{t-j}\right)\left(\sum_{j=M+1}^\infty \psi_j \epsilon_{t+k-j}\right) = \sigma_\varepsilon^2 \sum_{j=M+1}^\infty \psi_j \psi_{j+|k|}$. Then

$$\begin{aligned} \mathbb{E}_D \overline{W}_{n,M}^2 &= (nh)^{-1} \sum_{t=1}^n \sum_{s=1}^n D_t D_s \sigma_t \sigma_s k_t(\tau) k_s(\tau) R_W(t-s) \\ &\leq 2(nh)^{-1} \sup_{\tau\in[0,1]} \sigma(\tau)^2 \sum_{i=0}^n \sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) |R_W(i)| \\ &\leq 2C \sup_{\tau\in[0,1]} \sigma(\tau)^2 \sum_{i=0}^\infty \left| \sum_{j=M+1}^\infty \psi_j \psi_{j+i} \right| \leq 2C \sup_{\tau\in[0,1]} \sigma(\tau)^2 \left(\sum_{j=M+1}^\infty |\psi_j| \right)^2 = o(M^{-2}), \end{aligned}$$

which follows as $0 \leq D_t D_{t+i} \leq 1$, by Lemma A.2 and the fact that the summability condition in Assumption 3 implies that $\sum_{j=M+1}^\infty |\psi_j| = o(M^{-1})$. As $\mathbb{E}_D \overline{W}_{n,M} = 0$, the Markov inequality implies that the truncation is asymptotically negligible.

We next split $\overline{Z}_{n,M}(\tau)$ into two sequences of blocks: one with small, negligible blocks $Y_{n,j}(\tau)$ and one with dominating blocks $X_{n,j}(\tau)$. Define $B_j = (j-1)(a+b)$, then

$$X_{n,j}(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=B_j+1}^{B_j+a} k_t(\tau) D_t \sigma_t u_{t,M}, \qquad Y_{n,j}(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=B_j+a+1}^{B_j+a+b} k_t(\tau) D_t \sigma_t u_{t,M},$$

such that $\overline{Z}_{n,M}(\tau) = \sum_{j=1}^k X_{n,i}(\tau) + \sum_{i=1}^k Y_{n,i}(\tau)$, where $k = \lceil n/(a+b) \rceil$, and the final block is

truncated to have n observations in total. Now take sequences $a = a(n)$ and $b = b(n) \to \infty$ such that $a/(nh) + M/a \to 0$ and $b/a + M/b \to 0$ as $n \to \infty$.

We first show that the small blocks are asymptotically negligible. First note that $\mathbb{E}_D\left(\sum_{i=1}^k Y_{n,i}(\tau)\right) = 0$. Consider n large enough such that $a > M$ and the blocks $Y_{n,i}$ are mutually independent conditionally on $\{D_t\}_{t=1}^n$. Then, with $R_M(i) = \mathbb{E}\, u_{t,M}u_{t+i,M}$, where $\sum_{i=0}^\infty |R_M(i)| < \infty$ by Assumption 3, we have that

$$\begin{aligned}
&\mathbb{E}_D\left(\sum_{j=1}^k Y_{n,j}(\tau)\right)^2 = \sum_{j=1}^k \mathbb{E}_D\, Y_{n,j}(\tau)^2 = \frac{1}{nh}\sum_{j=1}^k \sum_{s,t=B_j+a+1}^{B_j+a+b} k_s(\tau)k_t(\tau)D_sD_t\sigma_s\sigma_t R_M(s-t) \\
&\leq 2C\frac{1}{nh}\sum_{i=0}^{b-1}|R_M(i)|\sum_{j=1}^k\sum_{t=B_j+a+1}^{B_j+a+b-i} k_t(\tau)k_{t+i}(\tau) \leq C_1\frac{1}{nh}\max_{0\leq i\leq b-1}\sum_{j=1}^k\sum_{t=B_i+a+1}^{B_i+a+b-i} k_t(\tau)k_{t+i}(\tau) \\
&\leq C_2\frac{kbh}{nh} \leq C_3\frac{b}{a} = o(1),
\end{aligned}$$

where we use that $\sum_{j=1}^k\sum_{t=B_i+a+1}^{B_i+a+b-i} k_t(\tau)k_{t+i}(\tau) \leq Ckbh$ and $k \sim an$.

We next employ the Lindeberg central limit theorem (see e.g. Davidson, 2002, Thm 23.6) to show that $\sum_{j=1}^k X_{n,j}(\tau) \xrightarrow{d} \mathcal{N}(0, p(\tau)^2\sigma_{as}^2(\tau))$. Consider again n sufficiently large such that such that $b > M$ and the blocks $X_{n,i}(\tau)$ are conditionally independent. Then $\mathbb{E}_D\sum_{j=1}^k X_{n,j}(\tau) = 0$ and

$$\begin{aligned}
\mathbb{E}_D\left(\sum_{j=1}^k X_{n,j}(\tau)\right)^2 &= \sum_{j=1}^k \mathbb{E}_D\, X_{n,j}(\tau)^2 = \frac{1}{nh}\sum_{i=1}^k\sum_{s,t=B_j+1}^{B_j+a} k_s(\tau)k_t(\tau)D_sD_t\sigma_s\sigma_t R_M(s-t) \\
&= \frac{1}{nh}\sum_{j=1}^k\sum_{i=-a+1}^{a-1} R_M(i)\sum_{t=B_j+1}^{B_j+a-|i|} k_t(\tau)k_{t+|i|}(\tau)\sigma_t\sigma_{t+|i|}D_tD_{t+|i|}.
\end{aligned}$$

As $M \to \infty$, $R_M(k) \to R_U(k)$. A straightforward extension of Lemma A.4 then shows that $\mathbb{E}_D\left(\sum_{j=1}^k X_{n,j}(\tau)\right)^2 \xrightarrow{p} p(\tau)\sigma(\tau)^2\Omega\kappa_2$.

The final step is to verify the Lindeberg condition, that is, we verify that, for every $\kappa > 0$, $\sum_{j=1}^k \mathbb{E}_D\left[\frac{X_{n,j}(\tau)^2}{\omega_n^2}\mathbb{1}\left(\left|\frac{X_{n,i}(\tau)}{\omega_n}\right| > \kappa\right)\right] = o_p(1)$, with $\omega_n^2 = \mathbb{E}_D\left(\sum_{j=1}^k X_{n,j}(\tau)\right)^2$. Note that

$$\sum_{j=1}^k \mathbb{E}_D\left[\frac{X_{n,j}(\tau)^2}{\omega_n^2}\mathbb{1}\left(\left|\frac{X_{n,i}(\tau)}{\omega_n}\right| > \kappa\right)\right] \leq \frac{1}{\kappa^2\omega_n^4}\sum_{j=1}^k \mathbb{E}_D\, X_{n,j}(\tau)^4$$

and define $\tilde{X}_t(\tau) = D_tk_t(\tau)\sigma_t u_{t,M}$, which, due to the M-dependence of $u_{t,M}$, is an L_4-mixingale with $\mathbb{E}_D\,\tilde{X}_t^4 \leq Ck_t(\tau)^4\,\mathbb{E}\, u_{t,M}^4$, where it follows from Minkowski's inequality that

$$\mathbb{E}\, u_{t,M}^4 \leq \left(\sum_{j=0}^M \left(\mathbb{E}\,|\psi_j\varepsilon_{t-j}|^4\right)^{1/4}\right)^4 \leq C\,\mathbb{E}\,\varepsilon_t^4\left(\sum_{j=0}^M \psi_j\right)^4 < \infty, \tag{A.5}$$

by stationarity of $\{\varepsilon_t\}$ and $\mathbb{E}\,\varepsilon_t^4 < \infty$. Lemma 2 of Hansen (1991) then implies that

$$\mathbb{E}_D\, X_{n,j}(\tau)^4 = \mathbb{E}_D \left| \frac{1}{\sqrt{nh}} \sum_{t=B_j+1}^{B_j+a} \tilde{X}_t(\tau) \right|^4 \leq C \frac{1}{(nh)^2} \left(\sum_{t=B_j+1}^{B_j+a} k_t(\tau)^2 \right)^2 .$$

As the blocks are non-overlapping, it follows from the properties of the kernel function and the c_r-inequality that

$$\sum_{j=1}^{k} \left(\sum_{t=B_j+1}^{B_j+a} k_t(\tau)^2 \right)^2 \leq a \sum_{j=1}^{k} \sum_{t=B_j+1}^{B_j+a} k_t(\tau)^4 \leq Cnha.$$

Therefore, with $\omega_n^{-2} = O(1)$ and $a = o(nh)$, $\sum_{i=1}^{k} \mathbb{E}_D \left[\frac{X_{n,i}(\tau)^2}{\omega_n^2} \mathbb{1}\left(\left| \frac{X_{n,i}(\tau)}{\omega_n} \right| > \kappa \right) \right] \leq \frac{Ca}{\omega_n^4 nh} = o(1)$. □

Lemma A.7. *Let Assumptions 1-8 hold. Then, for any $0 < \delta < \delta^* < \frac{1}{2}$ and some $N > 0$, it holds for all $\tau \in [\delta^*, 1-\delta^*]$ and $n > N$ that*

$$\sqrt{nh} \left[\hat{m}^*(\tau) - \tilde{m}(\tau) - h^2 B_{as}(\tau) \right] = p(\tau)^{-1} Z_{n,U}^*(\tau) + R_n^*(\tau),$$

where

$$Z_{n,U}^*(\tau) = \frac{1}{\sqrt{nh}} \sum_{t=1}^{n} k_t(\tau) D_t z_t \xi_t^* \tag{A.6a}$$

$$\operatorname*{plim}_{n\to\infty} \mathbb{V}\mathrm{ar}^*\, Z_{n,U}^*(\tau) = \sigma_Z^2(\tau) \tag{A.6b}$$

$$\sup_{\tau\in[\delta^*,1-\delta^*]} \mathbb{E}\, \|R_n^*(\tau)\|^* \leq C \max \left\{ \frac{1}{\sqrt{nh}}, \sqrt{nh^7}, \sqrt{h\tilde{h}^{-1}}, \sqrt{nh^5\tilde{h}^4}, \tilde{h}^2, \sqrt{\ell\tilde{h}^4}, \sqrt{\frac{\ell}{n\tilde{h}}} \right\}, \tag{A.6c}$$

and $B_{as}(\tau)$ and $\sigma_Z^2(\tau)$ are defined in (8) *and* (A.2b) *respectively.*

Proof of Lemma A.7. Analogously to the proof of Lemma A.6, we define

$$\breve{m}^*(\tau) = \frac{1}{nh} \sum_{t=1}^{n} k_t(\tau) D_t y_t^*, \qquad \overline{m}_D^*(\tau) = \frac{1}{nh} \sum_{t=1}^{n} k_t(\tau) D_t \tilde{m}\left(\frac{t}{n}\right)$$
$$\overline{m}_p^*(\tau) = \frac{1}{nh} \sum_{t=1}^{n} p\left(\frac{t}{n}\right) k_t(\tau) \tilde{m}\left(\frac{t}{n}\right),$$

such that we can write

$$\begin{aligned}
&\sqrt{nh} \left[\hat{m}^*(\tau) - \tilde{m}(\tau) - h^2 B_{as}(\tau) \right] = \sqrt{nh}\hat{p}(\tau)^{-1} \left\{ \breve{m}^*(\tau) - \overline{m}_D^*(\tau) \right\} \\
&\quad + \left\{ \hat{p}(\tau)^{-1} \overline{m}_D^*(\tau) - p(\tau)^{-1} \overline{m}_p^*(\tau) \right\} + \sqrt{nh} \left\{ p(\tau)^{-1} \overline{m}_p^*(\tau) - \tilde{m}(\tau) - h^2 B_{as}(\tau) \right\} \\
&= p(\tau)^{-1} Z_{n,U}^*(\tau) + I_n^*(\tau) + II_n^*(\tau) + III_n^*(\tau)
\end{aligned} \tag{A.7}$$

As a general observation, note that taking the sums in the expressions above from $t = 1$ to n appears

to include the boundary points for $\tilde{m}(\cdot)$, for which the properties of the estimator are not satisfied. However, as we consider $\tau \in [\delta^*, 1-\delta^*]$ and $\tilde{m}\left(\frac{t}{n}\right)$ is always multiplied by $k_t(\tau)$, $k_t(\tau) = 0$ at all points $\frac{t}{n} < \delta$ and $\frac{t}{n} > 1-\delta$ for large enough n, and therefore

$$\frac{1}{nh}\sum_{t=1}^{n} k_t(\tau)\tilde{m}\left(\frac{t}{n}\right) = \frac{1}{nh}\sum_{t=[n\delta]+1}^{[n(1-\delta)]} k_t(\tau)\tilde{m}\left(\frac{t}{n}\right),$$

and analogously for all related sums. Hence, in the following we simply take sums from $t = 1$ to n, under the implicit assumption that n is large enough to do so.

We now first derive (A.6b). With $\mathbb{E}^* Z^*_{n,U}(\tau)^2 = \frac{1}{nh}\sum_{i=-n+1}^{n-1}\sum_{t=1}^{n-|i|} k_t(\tau)k_{t+|i|}(\tau)D_tD_{t+|i|}z_tz_{t+|i|}\gamma^{|i|}$, it follows from Lemma A.5 and the fact that $\sum_{i=0}^{\infty}\gamma^i = \frac{1}{1-\theta^{1/\ell}} = -\ell/\ln\theta + o(\ell)$,

$$\mathbb{E}\left|\mathbb{E}^* Z^*_{n,U}(\tau)^2 - \mathbb{E}_D\,\mathbb{E}^* Z^*_{n,U}(\tau)^2\right| \leq \phi_n\sum_{i=0}^{\infty}\gamma^i\beta_i + \frac{\eta_n}{\sqrt{nh}}\sum_{i=0}^{\infty}\gamma^i \leq \frac{C\ell}{\sqrt{nh}} = o(1).$$

Furthermore

$$\begin{aligned}\mathbb{E}_D\,\mathbb{E}^* Z^*_{n,U}(\tau)^2 &= \frac{1}{nh}\sum_{i=-n+1}^{n-1} R_U(i)\sum_{t=1}^{n-|i|} k_t(\tau)k_{t+|i|}(\tau)D_tD_{t+|i|}\sigma_t\sigma_{t+|i|}\gamma^{|i|}\\ &= \mathbb{E}_D\,Z^2_{n,U} + \frac{1}{nh}\sum_{i=-n+1}^{n-1} R_U(i)\sum_{t=1}^{n-|i|} k_t(\tau)k_{t+|i|}(\tau)\sigma_t\sigma_{t+|i|}D_tD_{t+|i|}(\gamma^{|i|}-1)\end{aligned}$$

where $Z_{n,U}$ is defined in Lemma A.6. For the second term, take $M = M(n)$ such that $1/M + M^2/\ell \to 0$ as $n \to \infty$, then we have that

$$\begin{aligned}&\sup_{\tau\in[\delta^*,1\text{-}\delta^*]}\left|\frac{1}{nh}\sum_{i=-n+1}^{n-1} R_U(i)(\gamma^{|i|}-1)\sum_{t=1}^{n-|i|} k_t(\tau)k_{t+|i|}(\tau)\sigma_t\sigma_{t+|i|}D_tD_{t+|i|}\right|\\ &\leq C\sum_{i=1}^{M} R_U(i)\left|\theta^{i/\ell}-1\right| + C\sum_{i=M+1}^{n-1} R_U(i)\left|\theta^{i/\ell}-1\right|\\ &\leq C_1\sum_{i=1}^{M}(1-\theta^{i/\ell}) + C_2\sum_{i=M+1}^{\infty} R_U(i) = o(1),\end{aligned}$$

as $\sum_{i=1}^{M}(1-\theta^{i/\ell}) \leq C\sum_{i=1}^{M}(-\ln\theta)i/\ell \leq C_1M^2/\ell = o(1)$ and $\sum_{i=M+1}^{\infty} R_U(i) = o(M^{-1})$. It then follows directly from the proof of Lemma A.6 that $\operatorname{plim}_{n\to\infty}\mathbb{V}\mathrm{ar}^* Z^*_{n,U}(\tau) = \sigma^2_Z(\tau)$ for all $\tau \in [\delta^*, 1-\delta^*]$.

Now consider $I^*_n(\tau)$, which we write as

$$\begin{aligned}I^*_n(\tau) &= \left[\hat{p}(\tau)^{-1} - p(\tau)^{-1}\right]Z^*_{n,U}(\tau) + \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} k_t(\tau)\left[m\left(\frac{t}{n}\right) - \tilde{m}\left(\frac{t}{n}\right)\right]D_t\xi^*_t\\ &= I^*_{n,1}(\tau) + I^*_{n,2}(\tau).\end{aligned}$$

It follows directly as in the proof of Lemma A.6 and using Jensen's inequality that $\sup_{\tau\in[\delta^*,1\text{-}\delta^*]}\mathbb{E}\left\|I^*_{n,1}(\tau)\right\|^* \leq C\max\left\{h^2,\frac{1}{\sqrt{nh}}\right\}$. Furthermore, as

$$\mathbb{E}^* I^*_{n,2}(\tau)^2 = \frac{1}{(nh)^2}\sum_{s=1}^{n}\sum_{t=1}^{n} k_s(\tau)k_t(\tau)\left[\tilde{m}\left(\frac{s}{n}\right)-m\left(\frac{t}{n}\right)\right]\left[\tilde{m}\left(\frac{t}{n}\right)-m\left(\frac{t}{n}\right)\right]\gamma^{|s-t|}D_sD_t,$$

it follows from Jensen's and the Cauchy-Schwartz inequality that

$$\begin{aligned}\mathbb{E}\left\|I^*_{n,2}(\tau)\right\|^* &\leq \left[\frac{2}{nh}\sum_{i=0}^{n-1}\gamma^i\sum_{t=1}^{n-i}k_t(\tau)k_{t+i}(\tau)\left\|\tilde{m}\left(\frac{t}{n}\right)-m\left(\frac{t}{n}\right)\right\|\left\|\tilde{m}\left(\frac{t+i}{n}\right)-m\left(\frac{t+i}{n}\right)\right\|\right]^{1/2}\\ &\leq C\left[\sup_{\sup_{\tau\in[\delta,1-\delta]}}\|\tilde{m}(\tau)-m(\tau)\|^2\sum_{i=0}^{\infty}\gamma^i\right]^{1/2} \leq C_1\sqrt{\ell}\max\left\{\tilde{h}^2,\frac{1}{\sqrt{n\tilde{h}}}\right\},\end{aligned}$$

as, using Lemma A.6,

$$\begin{aligned}\sup_{\tau\in[\delta,1-\delta]}\|\tilde{m}(\tau)-m(\tau)\| \leq \tilde{h}^2\sup_{\tau\in[\delta,1-\delta]}|B_{as}(\tau)| + \frac{1}{\sqrt{n\tilde{h}}}\sup_{\tau\in[\delta,1-\delta]}\|Z_{n,U}(\tau)\|\\ + \frac{1}{\sqrt{n\tilde{h}}}\sup_{\tau\in[\delta,1-\delta]}\|R_n(\tau)\| \leq C\max\left\{\tilde{h}^2,\frac{1}{\sqrt{n\tilde{h}}}\right\}.\end{aligned} \tag{A.8}$$

As in the proof of Lemma A.6, write $II^*_n(\tau)$ as

$$\begin{aligned}II^*_n(\tau) &= p(\tau)^{-1}\left\{\overline{m}^*_D(\tau)-p(\tau)^{-1}\hat{p}(\tau)\overline{m}^*_p(\tau)\right\}+\hat{p}(\tau)^{-1}p(\tau)^{-2}\left\{\hat{p}(\tau)-p(\tau)\right\}^2\overline{m}^*_p(\tau)\\ &\quad+\left\{\hat{p}(\tau)^{-1}-p(\tau)^{-1}\right\}\left\{\overline{m}^*_D(\tau)-\overline{m}^*_p(\tau)\right\}\\ &= II^*_{n,1}(\tau)+II^*_{n,2}(\tau)+II^*_{n,3}(\tau),\end{aligned}$$

where $II^*_{n,1}(\tau)=p(\tau)^{-1}Z^*_{n,D}(\tau)$ and $Z^*_{n,D}(\tau)=\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}k_t(\tau)\left[\tilde{m}\left(\frac{t}{n}\right)-p(\tau)^{-1}\overline{m}^*_p(\tau)\right]\left[D_t-p\left(\frac{t}{n}\right)\right]$. Define $Z_{n,D}(\tau)=\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}k_t(\tau)\left[m\left(\frac{t}{n}\right)-p(\tau)^{-1}\overline{m}_p(\tau)\right]\left[D_t-p\left(\frac{t}{n}\right)\right]$. Then

$$\begin{aligned}Z^*_{n,D}(\tau)-Z_{n,D}(\tau) &= \frac{1}{\sqrt{nh}}\sum_{t=1}^{n}k_t(\tau)\left[\tilde{m}\left(\frac{t}{n}\right)-m\left(\frac{t}{n}\right)\right]\left[D_t-p\left(\frac{t}{n}\right)\right]\\ &\quad+\frac{1}{\sqrt{nh}}p(\tau)^{-1}\left[\overline{m}^*_p(\tau)-\overline{m}_p(\tau)\right]\sum_{t=1}^{n}k_t(\tau)\left[D_t-p\left(\frac{t}{n}\right)\right] = II^*_{n,11}(\tau)+II^*_{n,12}(\tau).\end{aligned} \tag{A.9}$$

It follows directly by (A.8) that

$$\begin{aligned}\left\|II^*_{n,11}(\tau)\right\| &\leq \left(\sup_{\tau\in[\delta^*,1-\delta^*]}\|\tilde{m}(\tau)-m(\tau)\|^2\sum_{i=-n+1}^{n-1}\frac{1}{nh}\sum_{t=1}^{n-|i|}k_t(\tau)k_{t+|i|}(\tau)\left|\mathbb{C}\mathrm{ov}(D_t,D_{t+|i|})\right|\right)^{1/2}\\ &\leq C\max\left\{\tilde{h}^2,\frac{1}{\sqrt{n\tilde{h}}}\right\}.\end{aligned}$$

Furthermore, as

$$\sup_{\tau\in[\delta^*,1\text{-}\delta^*]}\left\|\overline{m}_p^*(\tau)-\overline{m}_p(\tau)\right\| \leq \sup_{\tau\in[\delta^*,1\text{-}\delta^*]}\|\tilde{m}_n(\tau)-m(\tau)\| \sup_{\tau\in[\delta^*,1\text{-}\delta^*]}\frac{1}{nh}\sum_{t=1}^{n}p\left(\frac{t}{n}\right)k_t(\tau)$$
$$\leq C\max\left\{\tilde{h}^2,\frac{1}{\sqrt{n\tilde{h}}}\right\}$$

and $\frac{1}{\sqrt{nh}}\left\|\sum_{t=1}^{n}k_t(\tau)\left[D_t-p\left(\frac{t}{n}\right)\right]\right\|\leq C_1$ by Lemma A.4(i), it follows that $\sup_{\tau\in[\delta^*,1\text{-}\delta^*]}\left\|II_{n,12}^*(\tau)\right\|\leq C\max\left\{\tilde{h}^2,\frac{1}{\sqrt{n\tilde{h}}}\right\}$ as well. As it was shown in Lemma A.6 that $\|Z_{n,D}(\tau)\|\leq Ch^2$, it follows that $\left\|II_{n,1}^*(\tau)\right\|\leq C\max\left\{\tilde{h}^2,\frac{1}{\sqrt{n\tilde{h}}}\right\}$.

As $\sup_{\tau\in[\delta,1\text{-}\delta]}\left\|\overline{m}_p^*(\tau)-\overline{m}_p(\tau)\right\|\leq C\max\left\{\tilde{h}^2,\frac{1}{\sqrt{n\tilde{h}}}\right\}$, it follows as in the proof of $II_{n,2}(\tau)$ in Lemma A.6 that $\sup_{\tau\in[\delta^*,1\text{-}\delta^*]}\left\|II_{n,2}^*(\tau)\right\|\leq C\max\left\{\sqrt{nh^9},\frac{1}{\sqrt{nh}}\right\}$, while $\overline{m}_D^*(\tau)-\overline{m}_p^*(\tau)=\overline{m}_D(\tau)-\overline{m}_p(\tau)+II_{n,11}^*(\tau)$, and $II_{n,11}^*(\tau)$ is defined in (A.9). Therefore it follows directly as in the proof of $II_{n,3}(\tau)$ in Lemma A.6 that $\sup_{\tau\in[\delta,1\text{-}\delta]}\left\|II_{n,3}^*(\tau)\right\|\leq C\max\left\{\tilde{h}^2,\frac{1}{\sqrt{nh}}\right\}$.

It follows from Lemma A.6 that $\tilde{m}(\tau)=m(\tau)+\tilde{h}^2B_{as}+\tilde{R}_n(\tau)$, where $\sup_{\tau\in[\delta,1\text{-}\delta]}\left\|\tilde{R}_n(\tau)\right\|\leq\frac{C}{\sqrt{n\tilde{h}}}$. Substituting this into $III_n^*(\tau)$ we find that

$$\begin{aligned}III_n^*(\tau)&=\left[p(\tau)^{-1}\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}p\left(\frac{t}{n}\right)k_t(\tau)m\left(\frac{t}{n}\right)-m(\tau)-h^2B_{as}(\tau)\right]\\&\quad+\tilde{h}^2\left[p(\tau)^{-1}\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}p\left(\frac{t}{n}\right)k_t(\tau)B_{as}\left(\frac{t}{n}\right)-B_{as}(\tau)\right]\\&\quad+\left[p(\tau)^{-1}\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}p\left(\frac{t}{n}\right)k_t(\tau)\tilde{R}_n\left(\frac{t}{n}\right)-\tilde{R}_n(\tau)\right]\\&=III_{n,1}^*(\tau)+III_{n,2}^*(\tau)+III_{n,3}^*(\tau).\end{aligned}$$

Note that $III_{n,1}^*(\tau)$ is equal to $III_n(\tau)$ defined in Lemma A.6, such that $\sup_{\tau\in[\delta^*,1\text{-}\delta^*]}\left|III_{n,1}^*(\tau)\right|\leq C\max\left\{\sqrt{nh^7},\frac{1}{\sqrt{nh}}\right\}$. Furthermore, using the definition of $B_{as}(\tau)$ in (8), we can write

$$\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}p\left(\frac{t}{n}\right)k_t(\tau)B_{as}\left(\frac{t}{n}\right)=\frac{1}{2}\mu_2\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}k_t(\tau)f\left(\frac{t}{n}\right),$$

where $f(\tau)=p(\tau)^{-1}\left[mp\right]^{(2)}(\tau)$ is Lipschitz continuous. Then, by Lemma A.3(i) it follows that

$$\left|III_{n,2}^*(\tau)\right|\leq C\sqrt{nh}\tilde{h}^2\left|\frac{1}{nh}\sum_{t=1}^{n}k_t(\tau)f\left(\frac{t}{n}\right)-f(\tau)\right|\leq C_1\max\left\{\frac{\tilde{h}^2}{\sqrt{nh}},\sqrt{nh^5\tilde{h}^4}\right\}.$$

Finally, for $III^*_{n,3}(\tau)$ we have

$$\|III^*_n(\tau)\| \leq C\sqrt{nh} \sup_{\tau\in[\delta,1\text{-}\delta]} \left\|\tilde{R}_n(\tau)\right\| \frac{1}{nh}\sum_{t=1}^{n} k_t^2(\tau) + \left\|\tilde{R}_n(\tau)\right\| \leq C_1\sqrt{\frac{h}{\tilde{h}}} + \frac{C_2}{\sqrt{n\tilde{h}}}.$$

Collecting all remainder terms, it then follows that

$$\sup_{\tau\in[\delta^*,1\text{-}\delta^*]} \mathbb{E}\left\|R^*_n(\tau)\right\|^* \leq C\max\left\{\frac{1}{\sqrt{nh}}, \sqrt{nh^7}, \sqrt{\frac{h}{\tilde{h}}}, \sqrt{nh^5\tilde{h}^4}, \tilde{h}^2, \sqrt{\ell\tilde{h}^4}, \sqrt{\frac{\ell}{n\tilde{h}}}\right\}. \qquad \square$$

Proof of Theorem 2. Given Lemma A.7, we only have to prove asymptotic normality of $Z^*_{n,U}(\tau)$. As in the proof of Theorem 1, we establish asymptotic normality of the bootstrap process using a blocking technique. By the stationarity of ξ^*_t, we can write $\xi^*_t = \sum_{j=0}^{\infty}\gamma^j\nu^*_{t-j}$ with ν^*_t for $t \leq 1$ defined analogously as for $t > 1$. Take an M such that $M/\ell \to \infty$ as $n \to \infty$, and truncate the MA(∞) representation of ξ^*_t at M lags to define $\xi^*_{t,M} = \sum_{j=0}^{M}\gamma^j\nu^*_{t-j}$. Then we write

$$\begin{aligned} Z^*_{n,U}(\tau) &= \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} k_t(\tau)D_t z_t\xi^*_{t,M} + \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} k_t(\tau)D_t z_t\left(\sum_{j=M+1}^{\infty}\gamma^j\nu^*_{t-j}\right) \\ &= \overline{Z}^*_{n,M}(\tau) + \overline{W}^*_{n,M}(\tau). \end{aligned}$$

Applying Markov's inequality twice, we have that

$$\begin{aligned} \mathbb{E}\,\mathbb{E}^*\,\overline{W}^*_{n,M}(\tau)^2 &\leq \frac{2}{nh}(1-\gamma^2)\sum_{i=0}^{n-1}|R_U(i)|\sum_{t=1}^{n-i} k_t(\tau)k_{t+i}(\tau)\left(\sum_{j=M+1}^{\infty}\gamma^{2j+i}\right) \\ &\leq C\gamma^{2M}\sum_{i=0}^{\infty}\gamma^i\,|R_U(i)| = C\theta^{2M/\ell}\sum_{i=0}^{\infty}\theta^{i/\ell}\,|R_U(i)| \leq C_1\theta^{2M/\ell} = o(1), \end{aligned}$$

as $M/\ell \to \infty$. It then follows that $\overline{W}^*_{n,M}(\tau) = o^*_p(1)$ for all $\tau \in [\delta^*, 1-\delta^*]$.

Now let $\overline{Z}_{t,M}(\tau) = \sum_{j=1}^{k} X^*_{n,j}(\tau) + \sum_{j=1}^{k} Y_{n,j}(\tau)$, where

$$X^*_{n,j}(\tau) = \frac{1}{\sqrt{nh}}\sum_{t=B_j+1}^{B_j+a} k_t(\tau)D_t z_t\xi^*_{t,M}, \qquad Y_{n,j}(\tau) = \frac{1}{\sqrt{nh}}\sum_{t=B_j+a+1}^{B_j+a+b} k_t(\tau)D_t z_t\xi^*_{t,M},$$

with $B_j = (j-1)(a+b)$ and $k = \lceil n/(a+b)\rceil$. Take sequences $a = a(n)$ and $b = b(n)$ such that $a/(nh) + M/a \to 0$ and $b/a + M/b \to 0$ as $n \to \infty$.

We first show that $\sum_{j=1}^{k} Y^*_{n,i}(\tau) = o^*_p(1)$. Consider n large enough such that $a(n) > M$ and the blocks $Y_{n,i}$ are mutually independent conditionally on the original data. Then, with

$$R^*_M(i) = \mathbb{E}^*\,\xi^*_t\xi^*_{t+|i|} = (1-\gamma^2)\sum_{j=0}^{M}\gamma^j\gamma^{j+|i|} = \theta^{|i|/\ell}(1-\theta^{2(M+1)/\ell}) \leq \theta^{|i|/\ell}$$

for $|i| \leq M-1$, we have that

$$\mathbb{E}\,\mathbb{E}^*\left(\sum_{j=1}^{k} Y^*_{n,j}(\tau)\right)^2 = \sum_{j=1}^{k} \mathbb{E}^*\, Y_{n,j}(\tau)^2 \leq \frac{1}{nh}\sum_{i=-b+1}^{b-1} \theta^{i/\ell} R_U(i) \sum_{j=1}^{k} \sum_{t=B_j+a+1}^{B_j+a+b-|i|} k_t(\tau) k_{t+|i|}(\tau)$$
$$\leq C_1 \frac{b}{a} = o(1).$$

As in the proof of Theorem 1, we employ the Lindeberg CLT to establish asymptotic normality of $\sum_{j=1}^{k} X^*_{n,j}(\tau)$, as for n sufficiently large, $b > M$ and the blocks $X^*_{n,j}(\tau)$ are independent. First we show that the asymptotic variance is equal to $p(\tau)^2\sigma^2_{as}(\tau)$. Note that $\mathbb{E}^* \sum_{j=1}^{k} X^*_{n,j}(\tau) = 0$ and

$$\mathbb{E}^*\left(\sum_{j=1}^{k} X^*_{n,j}(\tau)\right)^2 = \frac{1}{nh}\sum_{i=-a+1}^{a-1} R^*_M(i) R_U(i) \sum_{j=1}^{k} \sum_{t=B_j+1}^{B_j+a-|i|} k_t(\tau) k_{t+|i|}(\tau) D_t D_{t+|i|} \sigma_t \sigma_{t+|i|}$$
$$+ \frac{1}{nh}\sum_{i=-a+1}^{a-1} \sum_{j=1}^{k} \sum_{t=B_j+1}^{B_j+a-|i|} k_t(\tau) k_{t+|i|}(\tau) D_t D_{t+|i|} \sigma_t \sigma_{t+|i|} R^*_M(i) \left[u_t u_{t+|i|} - R_U(i)\right]$$
$$= A^*_{X,n}(\tau) + B^*_{X,n}(\tau).$$

By adapting Lemma A.5, we find that

$$\mathbb{E}\left|B^*_{X,n}(\tau)\right| \leq 2\sum_{i=0}^{a-1} \mathbb{E}\left|\sum_{j=1}^{k} \sum_{t=B_j+1}^{B_j+a-|i|} k_t(\tau) k_{t+|i|}(\tau) D_t D_{t+|i|} \sigma_t \sigma_{t+|i|} \left[u_t u_{t+|i|} - R_U(i)\right]\right|$$
$$\leq \phi_n \sum_{i=0}^{a-1} \beta_i + \frac{\eta_n}{\sqrt{nh}} = o(1).$$

Furthermore, the arguments used to prove (A.6b) in Lemma A.7 show that $A^*_{X,n}(\tau) \xrightarrow{p} p(\tau)^2\sigma^2_{as}(\tau)$.

The final step is to verify that, for every $\kappa > 0$, $\sum_{j=1}^{k} \mathbb{E}^*\left[\frac{X^*_{n,j}(\tau)^2}{\omega^{*2}_n} \mathbb{1}\left(\left|\frac{X^*_{n,j}(\tau)^2}{\omega^{*2}_n}\right| > \kappa\right)\right] = o_p(1)$, where $\omega^{*2}_n = \mathbb{E}^*\left(\sum_{j=1}^{k} X^*_{n,j}(\tau)\right)^2$. Similarly as in the proof of Theorem 1, we define $\tilde{X}^*_t(\tau) = D_t k_t(\tau)\sigma_t u_t \xi^*_{t,M}$, as an L_4-mixingale conditionally on the original data, with $\left\|\tilde{X}_t\right\|^*_4 \leq C\,|u_t|\,k_t(\tau)$, such that Lemma 2 of Hansen (1991) then implies that $\mathbb{E}^*\, X_{n,i}(\tau)^4 \leq \frac{C}{(nh)^2}\left(\sum_{t=B_j+1}^{B_j+a} k_t(\tau)^2 u_t^2\right)^2$, and therefore

$$\sum_{j=1}^{k} \mathbb{E}^*\left[\frac{X^*_{n,j}(\tau)^2}{\omega^{*2}_n} \mathbb{1}\left(\left|\frac{X^*_{n,j}(\tau)^2}{\omega^{*2}_n}\right| > \kappa\right)\right] \leq \frac{C}{(nh)^2\omega^{*4}} \sum_{j=1}^{k} \left(\sum_{t=B_j+1}^{B_j+a} k_t(\tau)^2 u_t^2, \right)^2,$$

where $\omega^{*-4}_n = O_p(1)$. By Minkowski's inequality, stationarity of u_t, and $\mathbb{E}\, u_t^4 < \infty$ – let $M \to \infty$ in

(A.5) – we have that

$$\frac{1}{(nh)^2}\sum_{j=1}^{k}\mathbb{E}\left(\sum_{t=B_j+1}^{B_j+a}k_t(\tau)^2u_t^2\right)^2 \leq \frac{1}{(nh)^2}(\mathbb{E}\,u_t^4)\sum_{i=1}^{k}\left(\sum_{t=B_j+1}^{B_j+a}k_t(\tau)^2\right)^2 \leq \frac{C_1 a}{nh} = o(1). \qquad \square$$

A.3 Uniform Results

Before deriving Theorem 3, we propose a few auxiliary lemmas aimed at establishing stochastic equicontinuity, which will be needed to extend the pointwise results to uniformity.

Lemma A.8. *Let $\{X_t\}_{t=1}^n$ be a stochastic process with $\limsup_{n\to\infty}\sum_{i=1}^{n-1}\max_{1\leq t\leq n-i}|\mathbb{E}\,X_tX_{t+i}| < \infty$. Then, for any $\tau_0\in(0,1)$ and $\tau_1,\tau_2\in[-1,1]$, there exists an $N>0$ such that for all $n>N$*

$$\left|\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\left[k_t(\tau_0+\tau_1h)-k_t(\tau_0+\tau_2h)\right]X_t\right| \leq B_n\left|\tau_1-\tau_2\right|,$$

where B_n is a random variable such that $\sup_{n>N,\tau_1,\tau_2\in[-1,1]}\mathbb{E}\,B_n^2<\infty$.

Proof of Lemma A.8. See Supplementary Appendix B. $\square$

Lemma A.9. *Let $\{X_t^*\}_{t=1}^n$ be a bootstrap process defined conditionally on a process $\{X_t\}_{t=1}^n$ with $\limsup_{n\to\infty}\sum_{i=1}^{n-1}\sup_{1\leq t\leq n-i}|\mathbb{E}\,\mathbb{E}^*\,X_t^*X_{t+i}^*| < \infty$. Then, for any $\tau_0\in(0,1)$ and $\tau_1,\tau_2\in[-1,1]$, there exists an $N>0$ such that for all $n>N$*

$$\left|\frac{1}{\sqrt{nh}}\sum_{t=1}^{n}\left[k_t(\tau_0+\tau_1h)-k_t(\tau_0+\tau_2h)\right]X_t^*\right| \leq B_n^*\left|\tau_1-\tau_2\right|,$$

where B_n^* is a random variable such that $\sup_{n>N,\tau_1,\tau_2\in[-1,1]}\mathbb{E}\,\mathbb{E}^*\,B_n^{*2}<\infty$.

Proof of Lemma A.9. See Supplementary Appendix B. $\square$

Lemma A.10. *Let $R_n(\tau)$ and $R_n^*(\tau)$ be defined as in Lemmas A.6 and A.7 respectively. Then we have, for all $\tau_0\in(0,1)$, that $\sup_{\tau\in[\text{-}1,1]}|R_n(\tau_0+\tau_1h)| = o_p(1)$ and $\sup_{\tau\in[\text{-}1,1]}|R_n^*(\tau_0+\tau_1h)| = o_p^*(1)$.*

Proof of Lemma A.10. See Supplementary Appendix B. $\square$

Proof of Theorem 3. It follows directly from Lemma A.10 that

$$\sup_{\tau\in[\text{-}1,1]}\left|Z_{\tau_0,n}(\tau)-p(\tau_0+\tau h)^{-1}Z_{n,U}(\tau_0+\tau h)-B_{as}(\tau_0)\right| \leq \sup_{\tau\in[\text{-}1,1]}|R_n(\tau_0+\tau h| = o_p(1),$$
$$\sup_{\tau\in[\text{-}1,1]}\left|Z_{\tau_0,n}^*(\tau)-p(\tau_0+\tau h)^{-1}Z_{n,U}^*(\tau_0+\tau h)-B_{as}(\tau_0)\right| \leq \sup_{\tau\in[\text{-}1,1]}|R_n^*(\tau_0+\tau h| = o_p^*(1),$$

such that we only have to consider $p(\tau_0+\tau h)^{-1}Z_{n,U}(\tau)$ and $p(\tau_0+\tau h)^{-1}Z^*_{n,U}(\tau)$ in the following.

We next establish the asymptotic covariances. With

$$\mathbb{E}_D\, Z_{n,U}(\tau_0+\tau_1 h)Z_{n,U}(\tau_0+\tau_2 h) = \frac{1}{nh}\sum_{i=-n+1}^{n-1} R_U(i)\sum_{t=1}^{n-|i|} k_t(\tau_0+\tau_1 h)k_{t+|i|}(\tau_0+\tau_2 h)\sigma_t\sigma_{t+|i|}D_tD_{t+|i|},$$

we apply Lemma A.4(*iii*) with $\tilde{g}_i(\tau_1,\tau_2)=\sigma(\tau_1)\sigma(\tau_2)$ to find that

$$\begin{aligned}
&\sup_{\tau_0\in[\delta,1\text{-}\delta]}\sup_{\tau_1,\tau_2\in[\text{-}1,1]}\left\|\mathbb{E}_D\, Z_{n,U}(\tau_0+\tau_1 h)Z_{n,U}(\tau_0+\tau_2 h) - p(\tau_0)\sigma^2(\tau_0)\kappa(\tau_1-\tau_2)\Omega_U\right\| \\
&\quad\le \sum_{i=-n+1}^{n-1}|R_U(i)|\left[\frac{C_1}{\sqrt{nh}}+\frac{C_2}{\sqrt{nh}}\max\left\{h,\frac{1}{nh^2}\right\}+S_n(i)\right]+\sum_{i=-\infty}^{\infty}\mathbb{1}(|i|\ge n)\,|R_U(i)|\,, \\
&\quad\le \frac{C_3}{\sqrt{nh}}+\frac{C_3}{\sqrt{nh}}\max\left\{h,\frac{1}{nh^2}\right\}+\phi_n,
\end{aligned}$$

from which it follows by the law of iterated expectations that

$$\sup_{\tau\in[\delta,1\text{-}\delta]}\sup_{\tau_1,\tau_2\in[\text{-}1,1]}\left|\mathbb{E}\, Z_{n,U}(\tau_0+\tau_1 h)Z_{n,U}(\tau_0+\tau_2 h) - p(\tau_0)\sigma_{W,\tau_0}(\tau_1,\tau_2)\right| = o(1).$$

It then follows that

$$\begin{aligned}
&\sup_{\tau\in[\delta,1\text{-}\delta]}\sup_{\tau_1,\tau_2\in[\text{-}1,1]}\left|\mathbb{E}\, Z_{\tau_0,n}(\tau_1)Z_{\tau_0,n}(\tau_2) - \sigma_{W,\tau_0}(\tau_1,\tau_2)\right| \\
&\le \sup_{\tau\in[\delta,1\text{-}\delta]}\sup_{\tau_1,\tau_2\in[\text{-}1,1]}\left|\mathbb{E}\, Z_{\tau_0,n}(\tau_1)Z_{\tau_0,n}(\tau_2) - p(\tau_0+\tau h)^{-1}p(\tau_0+\tau_2 h)^{-1}p(\tau_0)^2\sigma_{W,\tau_0}(\tau_1,\tau_2)\right| \\
&\quad+\sup_{\tau\in[\delta,1\text{-}\delta]}\sup_{\tau_1,\tau_2\in[\text{-}1,1]}\left|p(\tau_0+\tau_1 h)^{-1}p(\tau_0+\tau_2 h)^{-1}p(\tau_0)^2-1\right|\sigma_{W,\tau_0}(\tau_1,\tau_2)\le o(1)+o(h)=o(1).
\end{aligned}$$

We follow the same steps as in the proof of the asymptotic bootstrap variance in Lemma A.7 for the bootstrap covariances. Note that

$$\mathbb{E}^*\, Z^*_{n,U}(\tau_0+\tau_1 h)Z^*_{n,U}(\tau_0+\tau_2 h) = \frac{1}{nh}\sum_{i=1-n}^{n-1}\sum_{t=1}^{n-|i|} k_t(\tau_0+\tau_1 h)k_{t+i}(\tau_0+\tau_2 h)D_tD_{t+i}z_tz_{t+i}\gamma^i,$$

where it follows from a straightforward adaptation of Lemma A.5 allowing for different τ's that

$$\mathbb{E}\left|\mathbb{E}^*\, Z^*_{n,U}(\tau_0+\tau_1 h)Z^*_{n,U}(\tau_0+\tau_2 h) - \mathbb{E}_D\,\mathbb{E}^*\, Z^*_{n,U}(\tau_0+\tau_1 h)Z^*_{n,U}(\tau_0+\tau_2 h)\right| = o(1).$$

To conclude this part, we can show as in Lemma A.7 that

$$\sup_{\tau_0\in[\delta^*,1\text{-}\delta^*]}\sup_{\tau_1,\tau_2\in[\text{-}1,1]}\left|\mathbb{E}_D\,\mathbb{E}^*\, Z^*_{n,U}(\tau_0+\tau_1 h)Z^*_{n,U}(\tau_0+\tau_2 h) - \mathbb{E}_D\, Z_{n,U}(\tau_0+\tau_1 h)Z_{n,U}(\tau_0+\tau_2 h)\right| = o_p(1).$$

Finite-dimensional convergence of the vectors $(Z_{\tau_0,n}(\tau_1),\ldots,Z_{\tau_0,n}(\tau_m))'$ and $\left(Z^*_{\tau_0,n}(\tau_1),\ldots,Z^*_{\tau_0,n}(\tau_m)\right)'$

for $(\tau_1, \ldots, \tau_m)' \in [-1, 1]^m$ follows from Theorems 1 and 2 and the Cramér-Wold device; it remains to show tightness. By applying Lemma A.8 with $X_t = D_t z_t$ and Lemma A.9 with $X_t^* = D_t z_t \xi_t^*$, it follows directly that

$$\mathbb{E}\left(Z_{\tau_0,n}(\tau_1) - Z_{\tau_0,n}(\tau_2)\right)^2 \leq \mathbb{E}\, B_n^2 \left|\tau_1 - \tau_2\right|^2 \quad \text{and} \quad \mathbb{E}\,\mathbb{E}^*\left(Z_{\tau_0,n}(\tau_1) - Z_{\tau_0,n}(\tau_2)\right)^2 \leq \mathbb{E}\,\mathbb{E}^*\, B_n^{*2}(\tau_1 - \tau_2)^2,$$

where $\mathbb{E}\, B_n^2 \leq C$ and $\mathbb{E}\,\mathbb{E}^*\, B_n^{*2} \leq C$ and tightness follows by Theorem 12.3 of Billingsley (1968). $\square$