Parameter stability and semiparametric inference in time-varying ARCH models

Lionel Truquet [*][†]

Abstract

In this paper, we develop a complete methodology for detecting time-varying/non time-varying parameters in ARCH processes. For this purpose, we estimate and test various semiparametric versions of the time-varying ARCH model (tv-ARCH) which include two well known non stationary ARCH type models introduced in the econometric literature. Using kernel estimation, we show that non time-varying parameters can be estimated at the usual parametric rate of convergence and for a Gaussian noise, we construct estimates that are asymptotically efficient in a semiparametric sense. Then we introduce two statistical tests which can be used for detecting non time-varying parameters or for testing the second order dynamic. An information criterion for selecting the number of lags is also provided. We illustrate our methodology with several real data sets.

1   Introduction

The modeling of financial data using nonstationary time series has recently received considerable attention both in econometrics and in statistics. For classical daily series such as stock market indices or currency exchange rates, the stationarity assumption seems often incompatible with a long history of data and the necessity of using non stationary ARCH models has been pointed out by several authors. See for instance Mikosch and Stărică (2004), Granger and Stărică (2005), Engle and Rangel (2008), Fryzlewicz et al. (2008) and the references therein. However, it is difficult to find in the literature a consensus for representing nonstationary ARCH models. A natural approach is to allow time-varying parameters in the classical ARCH model of Engle (1982). Such an extension has been proposed by Dahlhaus and Subba Rao (2006) with the so-called time-varying ARCH model (tv-ARCH). The tv-ARCH processes are defined by the recursive equations

$$X_t = \xi_t \sigma_t, \quad \sigma_t^2 = a_0\left(\frac{t}{T}\right) + \sum_{j=1}^{p} a_j\left(\frac{t}{T}\right) X_{t-j}^2, \quad p+1 \le t \le T, \tag{1.1}$$

where for $0 \le j \le p$, a_j is a smooth function and ξ a strong white noise with variance 1. Since they can be locally approximated by stationary ARCH processes, the tv-ARCH processes are called locally stationary (the notion of local stationarity is introduced in Dahlhaus (1997) for linear processes but the meaning of local stationarity for the non linear tv-ARCH can be found in Dahlhaus and Subba Rao (2006)). From this important feature, a nice asymptotic theory can be developed for estimation of parameters, in particular local inference methods such as the local Quasi-maximum likelihood estimation studied in Dahlhaus and

[*]UMR 6625 CNRS IRMAR, University of Rennes 1, Campus de Beaulieu, F-35042 Rennes Cedex, France and

[†]ENSAI, Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France. *Email: lionel.truquet@ensai.fr.*

Subba Rao (2006), the local weighted least-squares estimation developed in Fryzlewicz et al. (2008) or the recursive online algorithms considered by Dahlhaus and Subba Rao (2007). In Fryzlewicz et al. (2008), it was shown that tv-ARCH processes provide good fits and accurate forecasts for some financial series. However, statistical inference in model (1.1) is complex, even for large samples, because $p + 1$ functions have to be estimated using nonparametric methods. Thus, in practice, reducing complexity can be interesting to improve model fit or forecasts accuracy. For example, Granger and Stărică (2005) have shown that the simple model

$$X_t = \sigma\left(\frac{t}{T}\right)\xi_t, \quad 1 \leq t \leq T, \tag{1.2}$$

with a smooth deterministic function $\sigma : [0, 1] \to \mathbb{R}_+$ can already produce significantly better forecasts for the returns of the SP&500 index than the classical GARCH$(1, 1)$ model. A process of type (1.2) can be seen as a tv-ARCH process with zero lag coefficients. Note that model (1.2) does not assume autocorrelation for the absolute values or the squares of the process (a correlation property often called second order dynamic in the literature) but only some changes in the unconditional variance. In Granger and Stărică (2005), it is argued that most of dynamic of the S&P index can be explained with a time-varying unconditional variance.

But nonstationarity and second order correlation can also be combined in a very simple way, assuming constant lag coefficients in (1.1):

$$X_t = \xi_t \sqrt{a_0\left(\frac{t}{T}\right) + \sum_{j=1}^{p} a_j X_{t-j}^2}. \tag{1.3}$$

Model (1.3) combine a time-varying unconditional variance compatible with the analysis of Granger and Stărică (2005) and a second order dynamic for the series with a single nonparametric component. Note also that a process $(X_t)_t$ defined by equations (1.3) can be written using the multiplicative form

$$X_t = \sqrt{a_0\left(\frac{t}{T}\right)} \cdot Y_t, \tag{1.4}$$

where

$$Y_t = \xi_t \sqrt{1 + \sum_{j=1}^{p} a_j \frac{a_0\left(\frac{t-j}{T}\right)}{a_0\left(\frac{t}{T}\right)} Y_{t-j}^2} \approx \xi_t \sqrt{1 + \sum_{j=1}^{p} a_j Y_{t-j}^2}$$

if we neglect the ratio $a_0\left(\frac{t-j}{T}\right)/a_0\left(\frac{t}{T}\right)$ which is of order $1+1/T$ when the function a_0 is positive and Lipschitz continuous over $[0, 1]$. One can also notice that writing the model with the latter approximation or not lead to two processes that are both approximated by the same stationary ARCH processes with parameters $a_0(u), a_1, \ldots, a_p$ (see Lemma 1 for this kind of approximation). In the stationary case, we remind that multiplying an ARCH process by a positive constant is equivalent to multiply the initial intercept coefficient. Then for a large sample size T, the process $(Y_t)_t$ behaves as a stationary ARCH process and $a_0(\cdot)$ is (up to a constant) the time-varying unconditional variance of the process $(X_t)_t$. Such a multiplicative form for ARCH models has been first considered by Engle and Rangel (2008) with the so-called Spline-GARCH model which writes as model (1.4) but with a GARCH$(1, 1)$ process $(Y_t)_t$.

Since the previous models satisfy the inclusions $(1.2) \subset (1.3) \subset (1.1)$, a natural question for any real data set is to test some properties of the lag coefficients. Testing the constancy of the lag coefficients can help to decide between model (1.3) and model (1.1) while testing the second order dynamic in model (1.3) is useful to determine if model (1.2) provides a sufficient fit. Statistical tools to help the practitioners to choose

among the three important specifications described above seems not available in the literature except in a recent paper of Patilea and Raïssi (2014) which introduces a test for the second order dynamic in model (1.3).

In this paper, we propose a general approach for estimating an arbitrary subset of non time-varying coefficients in tv-ARCH processes. Our estimators are $\sqrt{T}-$consistent and we will also study the semi-parametric asymptotic efficiency of our method when the noise is Gaussian. Using these results, we construct two statistical tests. The first test can be used to decide whether a given subset of parameters is time-varying or not. This test is based on a $\mathbb{L}^2$ distance between a nonparametric kernel estimator of the coefficients and the semiparametric estimator introduced in this paper. The second test can be used for deciding if the constant parameters are different from zero. Various applications can be considered as a simple particular case of our methodology: testing model (1.3) versus model (1.1), testing model (1.2) versus model (1.3), estimating parameters and selecting lag variables in models (1.3) or (1.1). When some coefficients are assumed to be non time-varying in (1.1), the decomposition $X_t^2 = \sigma_t^2 + \left(\xi_t^2 - 1\right)\sigma_t^2$ leads to semiparametric inference in a time-varying regression model. Detecting and estimating a parametric component in general time-varying regression models has been considered recently by Zhang and Wu (2012). However these authors do not consider the case of tv-ARCH processes with optimal moment condition for the marginal distribution, asymptotic semi-parametric efficiency for the estimation and Lipschitz continuity for the time-varying coefficients. Moreover, our approach for estimating non-time varying coefficients is quite different.

We applied our methodology to three real data sets: the daily exchange rates between the US Dollar and the Euro or between the US Dollar and the Indian Rupee and the FTSE index. For the three series of interest, a non time-varying intercept is clearly rejected over the considered period. The conclusion for the lag coefficients depends on the series. In fitting model (1.3), we also found that incorporating non stationarity reduces the values of lag parameters with respect to the stationary case. Then the time-varying unconditional variance has an important contribution to volatility.

The paper is organized as follows. In Section 2, we introduce our notations and we describe the basis of our method for statistical inference in a tv-ARCH model for which some coefficients are assumed to be non time-varying. In Section 3, we give the asymptotic results for our estimators and we discuss the problem of semiparametric asymptotic efficiency using the LAN theory. Statistical testing and their practical implementation are considered in Section 4 and Section 5 is devoted to real data applications. All the proofs of our results are postponed to the supplementary material which also contains many simulation studies showing the good behavior of our methodology. The Matlab codes and the data sets discussed in Section 5 are available at the URL

https://github.com/time-varying/tests-and-estimation-for-tv-ARCH-

2 Semiparametric volatility and tv-ARCH processes

2.1 Formulation and notations

In this section, we consider semiparametric versions of model (1.1), assuming that some of the ARCH coefficients are not time-varying. For $t \in [\![p + 1, T]\!] = [p + 1, T] \cap \mathbb{N}$, let M_t and N_t be two random vectors of size m and n respectively, with $m + n = p + 1$ and defined as follows. We split the interval $[\![0, p]\!]$ into two parts $\{q_1, q_2, \ldots, q_m\}$ and $\{r_1, r_2, \ldots, r_n\}$ with $q_1 < \cdots < q_m$ and $r_1 < \cdots < r_n$. If $q_1 = 0$, we set $M_t = \left(1, X_{t-q_2}^2, \ldots, X_{t-q_m}^2\right)'$, with the convention $M_t = 1$ if $m = 1$. If $q_1 > 0$, we set $M_t = \left(X_{t-q_1}^2, \ldots, X_{t-q_m}^2\right)'$. The vector N_t is defined similarly, replacing the q_ℓ's with the r_ℓ's. In particular, the coordinates of the random vectors M_t and N_t form a bipartition of the set $\left\{1, X_{t-1}^2, \ldots, X_{t-p}^2\right\}$. Now, we assume that the coefficients

3

vector $\beta = (a_{r_1}, \ldots, a_{r_n})'$ is constant. We also set $\alpha_t = \left(a_{q_1}(t/T), \ldots, a_{q_m}(t/T)\right)'$. Then the model writes

$$X_t = \xi_t \sigma_t, \quad \sigma_t^2 = M_t' \alpha_t + N_t' \beta, \tag{2.1}$$

for $t = p+1, \ldots, T$. Throughout this paper, we will assume that for all realization ω in the probability space, $(X_t(\omega))_{t \le 0}$ is a path of a stationary ARCH process with noise ξ and coefficients $a_j(0)$. From this convention, one can get a local approximation of a tv-ARCH process by stationary ARCH processes with parameters $\left(a_0(u), \ldots, a_p(u)\right)$. See Lemma 1 in the supplementary material for details.

In Section 4, a statistical test will be given for testing H_0: $(a_{r_1}, \ldots, a_{r_n})$ is constant. The sequel of this section is devoted to the statistical inference in model (1.1) under the null hypothesis H_0. The corresponding estimation of parameter β will be necessary to construct the test.

2.2 Estimators of the parametric part β

Considering the square of the process (2.1), statistical inference in model (2.1) can be viewed as a linear regression problem. More precisely, we have for $t \ge p + 1$,

$$X_t^2 = M_t' \alpha_t + N_t' \beta + (\xi_t^2 - 1)\sigma_t^2. \tag{2.2}$$

In the sequel, we consider a sequence of weights $(W_t)_{p+1 \le t \le T}$ such that W_t is a measurable function of t, T and $X_{t-1}^2, \ldots, X_{t-p}^2$. For stating our results, we will only consider sequences of the form

$$W_t = \left(\gamma_0\left(t/T\right) + \sum_{\ell=1}^{p} \gamma_j\left(t/T\right) X_{t-j}^2\right)^{-2}, \tag{2.3}$$

where the γ_j's are positive and Lipschitz continuous functions defined over $[0,1]$. The use of this kind of weights is classical in weighted least squares estimation in order to relax moment conditions on the marginal distribution or to gain in efficiency. The first goal of our procedure is to estimate the parameter β. This is the most difficult part of our methodology since a $\sqrt{T}$-consistent estimate is expected. Once an estimate $\hat{\beta}$ with the classical parametric rate of convergence is available, a pointwise estimate of parameter $\alpha(\cdot)$ can easily be obtained. One can just plug $\hat{\beta}$ in (2.2) and apply standard nonparametric methods (in this paper we will use the local weighted least-squares method studied in Fryzlewicz et al. (2008)). Our aim here is to first eliminate the nonparametric component $M_t' \alpha_t$. Our approach is classical in the setting of partially linear models, for which the regression function involves a parametric component and a nonparametric component (see for example Härdle et al. (2000), Chapter 6 for some results in the case of time series). However, our method is based on nonparametric estimation of linear projections of $\sqrt{W_t}X_t^2$ and $\sqrt{W_t}N_t$ onto the $\mathbb{L}^2$ subspace generated by the components of $\sqrt{W_t}M_t$ instead of a nonparametric estimation of the conditional expectations. Moreover, our two-step approach involving some weights and leading to semiparametric efficient estimates is not common for nonstationary time series and no existing results from the theory of partially linear models can be used here for our purpose. Here, our approach can be also interpreted as a partial regression. For stationary ARCH processes, estimation of the whole set of parameters using a regression model for the squares and least squares estimation has been studied by Bose and Mukherjee (2003) and Horváth and Liese (2004).

Now we introduce our estimator. We first multiply the two members of equation (2.2) by $\sqrt{W_t}$. If $\mathcal{P}_t\left(\sqrt{W_t}X_t^2\right)$ and $\mathcal{P}_t\left(\sqrt{W_t}N_t\right)$ denote the (componentwise) orthogonal projection of $\sqrt{W_t}X_t^2$ and $\sqrt{W_t}N_t$ onto the $\mathbb{L}^2$ linear subspace generated by the coordinates of $\sqrt{W_t}M_t$, it is easily seen that

$$\sqrt{W_t}X_t^2 - \mathcal{P}_t\left(\sqrt{W_t}X_t^2\right) = \left(\sqrt{W_t}N_t - \mathcal{P}_t\left(\sqrt{W_t}N_t\right)\right)' \beta + (\xi_t^2 - 1)\sqrt{W_t}\sigma_t^2. \tag{2.4}$$

The use of these orthogonal projections are natural in order to eliminate the nuisance parameter α_t and to get a partial regression involving parameter β only. Let us introduce some notations for expressing these projections. For $t \geq p + 1$, we set

$$q_{1,t} = s_{3,t}^{-1} s_{1,t}, \quad q_{2,t} = s_{3,t}^{-1} s_{2,t}$$

where

$$s_{3,t} = \mathbb{E}\left(W_t M_t M_t'\right), \quad s_{1,t} = \mathbb{E}\left(W_t M_t X_t^2\right), \quad s_{2,t} = \mathbb{E}\left(W_t M_t N_t'\right),$$

Then setting

$$V_t = X_t^2 - M_t' q_{1,t}, \quad O_t = N_t - q_{2,t}' M_t,$$

we have

$$\sqrt{W_t} X_t^2 - \mathcal{P}_t\left(\sqrt{W_t} X_t^2\right) = \sqrt{W_t} V_t, \quad \sqrt{W_t} N_t - \mathcal{P}_t\left(\sqrt{W_t} N_t\right) = \sqrt{W_t} O_t$$

and equation (2.4) writes

$$\sqrt{W_t} V_t = \sqrt{W_t} O_t \beta + \left(\xi_t^2 - 1\right) \sqrt{W_t} \sigma_t^2.$$

The idea is now to use a least squares estimator for β. Of course, in order to obtain a feasible estimator, it is necessary to first estimate the two quantities $q_{1,t}$ and $q_{2,t}$. To this end, we consider

$$\hat{s}_{3,b,t} = \sum_{i=p+1}^{T} k_{t,i}(b) W_i M_i M_i', \quad \hat{s}_{1,b,t} = \sum_{i=p+1}^{T} k_{t,i}(b) W_i M_i X_i^2,$$

$$\hat{s}_{2,b,t} = \sum_{i=p+1}^{T} k_{t,i}(b) W_i M_i N_i', \quad k_{t,i}(b) = \frac{K\left(\frac{t-i}{Tb}\right)}{\sum_{i=p+1}^{T} K\left(\frac{t-i}{Tb}\right)},$$

where for K is a kernel and $b > 0$ is a bandwidth parameter. Throughout this paper, the kernel K is assumed to be absolutely continuous and with support $[-1, 1]$. Then we set

$$\hat{q}_{1,b,t} = \left(\hat{s}_{3,b,t}\right)^{-1} \hat{s}_{1,b,t}, \quad \hat{q}_{2,b,t} = \left(\hat{s}_{3,b,t}\right)^{-1} \hat{s}_{2,b,t}. \tag{2.5}$$

Now we introduce the following notations. The quantities

$$\hat{V}_t = X_t^2 - M_t' \hat{q}_{1,b,t}, \quad \hat{O}_t = N_t - \hat{q}_{2,b,t}' M_t$$

estimate V_t and O_t respectively. Our estimator of parameter β will be denoted by $\hat{\beta}$ and minimizes the function ℓ_W defined by

$$\ell_W\left(\overline{\beta}\right) = \sum_{t=p+1}^{T} W_t\left(\hat{V}_t - \hat{O}_t' \overline{\beta}\right)^2.$$

We get

$$\hat{\beta} = \left(\sum_{t=p+1}^{T} W_t \hat{O}_t \hat{O}_t'\right)^{-1} \sum_{t=p+1}^{T} W_t \hat{O}_t \hat{V}_t. \tag{2.6}$$

It is now possible to define an estimate of parameter $\alpha(u)$ for $u \in [0, 1]$ by minimizing the function

$$\overline{\alpha} \mapsto \sum_{i=p+1}^{T} K\left(\frac{t-i}{Tb'}\right) W_i\left(X_i^2 - M_i'\overline{\alpha} - N_i'\hat{\beta}\right)^2,$$

5

for an integer $t \in [\![1, T]\!]$ such that $\left| u - \frac{t}{T} \right| \le \frac{1}{T}$ (e.g $t = [Tu]$ where $[x]$ denotes the integer part of a real number x). Since the nonparametric estimation of the function α requires a less restrictive assumption on the bandwidth parameter than for the estimation of parameter β, we introduce a new bandwidth b'. This leads to the estimate

$$\hat{\alpha}_t = \hat{q}_{1,b',t} - \hat{q}_{2,b',t} \cdot \hat{\beta}, \tag{2.7}$$

Note. Expression $\hat{q}_{1,b,t}$ and $\hat{q}_{2,b,t}$ involve the inverse of the matrix $S_{3,b,t}$. One can show that

$$\mathbb{P}\left(\det\left(\hat{s}_{3,t} \right) = 0 \text{ for some } t \le T \right) \to 0.$$

See Lemma 3 and its proof. However invertibility problems can occur when the noise ξ has a mass at point 0, for instance. For simplicity, we always assume that all these matrices are invertible. Studying our estimator on an event with probability tending to one only complicates the statements and proofs of our results by adding some indicator sets but does not change the used approach. One can also show that this distinction is unnecessary for a noise ξ having a density.

Asymptotic normality of the estimates (2.6) and (2.7) will be derived in the next section as well as some plug-in versions to get optimal asymptotic results.

3 Asymptotic results

3.1 Estimation of the parametric component

Our first result shows that the estimator (2.6) is $\sqrt{T}-$consistent under some conditions. Here are our main assumptions.

A1. For $j = 1, \ldots, p$, the function a_j is non-negative and Lipschitz continuous. The function a_0 is positive and Lipschitz continuous. Moreover,

$$c = \sup_{u \in [0,1]} \sum_{j=1}^{p} a_j(u) < 1.$$

A2(h). For the integer $h \ge 2$, there exists a real number $\delta \in (0, 1)$ such that $\mathbb{E}|\xi_1|^{h(1+\delta)} < \infty$.

Assumption **A1** is the classical contraction condition used in Fryzlewicz et al. (2008) to define tv-ARCH processes. Assumption **A2(h)**, used for different values of h in the sequel, implies a restriction on the noise distribution. Let us mention that this condition does not restrict the moment condition for the marginal X_t (in the stationary case, i.e when the a_j's are deterministic, assumption **A1** is the necessary and sufficient condition for the condition $\mathbb{E}\left(X_t^2 \right) < +\infty$).

In the sequel, $(X_t(u))_t$ will denote the stationary ARCH process with coefficients $(\alpha(u), \beta)$. Then we will use the notation $(M_t(u))_t$ (resp. $(N_t(u))_t$, $(W_t(u))_t$) for the stationary approximation of $(M_t)_t$ (resp. $(N_t)_t$, $(W_t)_t$). For example,

$$W_t(u) = \left(\gamma_0(u) + \sum_{\ell=1}^{p} \gamma_\ell(u) X_{t-\ell}^2(u) \right)^{-2}.$$

Theorem 1. *Assume that assumptions* **A1** *and* **A2(4)** *hold,* $b\sqrt{T} \to \infty$ *and* $b^2\sqrt{T} \to 0$. *Then we have the following convergence in distribution:*

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \to_{T\to\infty} \mathcal{N}_n\left(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\right),$$

with

$$\Sigma_1 = \mathbb{E}\int_0^1 W_1(u)\left(N_1(u) - q_2(u)\right)'M_1(u)\right) \cdot \left(N_1(u) - q_2(u)'M_1(u)\right)' du,$$

$$\Sigma_2 = \text{Var}\left(\xi_0^2\right) \cdot \mathbb{E}\int W_1(u)^2\sigma_1(u)^4\left(N_1(u) - q_2(u)'M_1(u)\right) \cdot \left(N_1(u) - q_2(u)'M_1(u)\right)' du,$$

and

$$q_2(u) = \mathbb{E}^{-1}\left(W_1(u)M_1(u)M_1(u)'\right)\mathbb{E}\left(W_1(u)M_1(u)N_1(u)'\right).$$

Notes.

1. The bandwidth conditions used in Theorem 1 are classical for estimating the parametric component in partially linear models. With this restriction, the nonparametric estimation step involved in the expression of $\hat{\beta}$ becomes negligible (i.e for $i = 1, 2$, $\hat{q}_{i,b,t}$ can be replaced with $q_{i,t}$ without changing the asymptotic behavior of $\hat{\beta}$). Let us explain the rule of these conditions. Some nonparametric estimates are introduced to approximate the two ratio $\hat{q}_{1,t}$ and $\hat{q}_{2,t}$. But, up to C/T (C denotes a positive constant), this two ratio can be seen as some Lipschitz functions of t/T (see Lemma 5 in the supplementary material). The mean square error for the kernel estimation of a Lipschitz functional in a regression model with deterministic design is bounded by $b^2 + \frac{1}{Tb}$ (up to a constant). Then our bandwidth conditions entail that this mean square error converges to zero with a faster rate than $\sqrt{T}$.

2. The goal of the proof of Theorem 1 is to show that the asymptotic distribution of $\sqrt{T}\left(\hat{\beta} - \beta\right)$ is the same as if the two quantities $\hat{q}_{1,b,t}$, $\hat{q}_{2,b,t}$ are replaced by $q_{1,t}$, $q_{2,t}$ respectively. Hence, the control of sums involving differences between these quantities are shown to be negligible. To this end, we make Taylor expansions and bound the variance of some multiple weighted sums appearing in this expansion using Lemma 4 given in the supplementary material.

3. The asymptotic variance in Theorem 1 can be estimated consistently using the data. Indeed, the proof of Theorem 1 given in the supplementary material shows that

$$\Sigma_1 = \lim_{T\to+\infty}\frac{1}{T}\sum_{t=p+1}^T W_t\hat{O}_t\hat{O}_t' \text{ a.s.}$$

Moreover, we have

$$\Sigma_2 = \mathbb{E}\int W_1(u)^2\left(X_1(u)^2 - \sigma_1(u)^2\right)^2\left(N_1(u) - q_2(u)'M_1(u)\right) \cdot \left(N_1(u) - q_2(u)'M_1(u)\right)' du. \quad (3.1)$$

Then an estimate of Σ_2 can be obtained if we replace q_2 with $\hat{q}_{2,b,\cdot}$, σ_t^2 with a pointwise estimate $\hat{\sigma}_t^2$ (see Theorem 3) and using the same kind of empirical counterpart as for Σ_1.

The asymptotic variance given in Theorem 1 depends on some weights $W_t(u)$. One can show that its minimal value (in the sense of non-negative definite matrices) is obtained for the choice $W_t^* = \frac{1}{\sigma_t^4}$. Indeed, setting $O_1(u) = N_1(u) - q_2(u)'M_1(u)$ and $O_1^*(u) = N_1(u) - q_2^*(u)'M_1(u)$, where

$$q_2^*(u) = \mathbb{E}^{-1}\left(\frac{M_1(u)M_1(u)'}{\sigma_1(u)^4}\right)\mathbb{E}\left(\frac{M_1(u)N_1(u)'}{\sigma_1(u)^4}\right),$$

we have for all $u \in [0, 1]$,

$$\Sigma_1 = \mathbb{E}\int_0^1 W_1(u)O_1(u)O_1^*(u)'\,du = \mathbb{E}\int_0^1 W_1(u)\sigma_1^2(u)O_1(u)\frac{O_1^*(u)'}{\sigma_1^2(u)}\,du.$$

Then if $x, y \in \mathbb{R}^n$, we have from the Cauchy-Schwarz inequality

$$(x'\Sigma_1 y)^2 \le \frac{x'\Sigma_2 x}{\text{Var}\left(\xi_0^2\right)} \times y'\Sigma y.$$

where $\Sigma = \mathbb{E}\int_0^1 \frac{O_1^*(u)O_1^*(u)'}{\sigma_1^4(u)}\,du$. Now setting $x = \Sigma_1^{-1}z$ and $y = \Sigma^{-1}z$ for $z \in \mathbb{R}^k$, we get

$$\left(z'\Sigma^{-1}z\right)^2 \le \frac{z'\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}z}{\text{Var}\left(\xi_0^2\right)} \times z'\Sigma^{-1}z.$$

Then we have proved the following result.

Proposition 1. *The lower bound for the asymptotic variance given in Theorem 1 is* $\text{Var}\left(\xi_0^2\right)\Sigma^{-1}$, *where*

$$\Sigma = \int_0^1 \mathbb{E}\left(\frac{1}{\sigma_1(u)^4}\left(N_1(u) - q_2^*(u)'M_1(u)\right)\left(N_1(u) - q_2^*(u)'M_1(u)\right)'\right)du,$$

where

$$q_2^*(u) = \mathbb{E}^{-1}\left(\frac{M_1(u)M_1(u)'}{\sigma_1(u)^4}\right)\mathbb{E}\left(\frac{M_1(u)N_1(u)'}{\sigma_1(u)^4}\right).$$

Now, we show that it is possible to construct an estimate of parameter β which has the asymptotic variance given in Proposition 1. A natural candidate is obtained by replacing the weights W_t in (2.6) with an estimation of the optimal weights $W_t^* = \frac{1}{\sigma_t^4}$. We set $\hat{W}_t^* = \frac{1}{\hat{\sigma}_t^4 + v_T}$, where

$$\hat{\sigma}_t^2 = M_t'\left(\hat{q}_{1,b,t} - \hat{q}_{2,b,t}\hat{\beta}\right) + N_t'\hat{\beta}$$

is an estimator of σ_t^2. The sequence $(v_T)_T$ is a sequence of positive real numbers such that $v_T = o\left(\frac{1}{\sqrt{T}}\right)$. The use of this sequence is just technical and avoids possible small values for the fitted volatility $\hat{\sigma}_t^2$ which is not ensured to be bounded away from 0 for finite samples. However, for a large sample size, our simulations show that the choice $v_T = 0$ does not alter the performance of the plug-in estimate. Let us define the quantities

$$\hat{s}_{3,b,t}^* = \sum_{i=p+1}^T k_{t,i}(b)\hat{W}_i^* M_i M_i', \quad \hat{s}_{1,b,t}^* = \sum_{i=p+1}^T k_{t,i}(b)\hat{W}_i^* M_i X_i^2,$$

8

$$\hat{s}^*_{2,b,t} = \sum_{i=p+1}^{T} k_{t,i}(b)\hat{W}^*_i M_i N'_i.$$

We also set

$$\hat{q}^*_{1,b,t} = \left(\hat{s}^*_{3,b,t}\right)^{-1} \hat{s}^*_{1,b,t}, \quad \hat{q}^*_{2,b,t} = \left(\hat{s}^*_{3,b,t}\right)^{-1} \hat{s}^*_{2,b,t}.$$

Now we introduce the following notations in order to simplify the expression of our estimator.

$$\hat{V}^*_t = X^2_t - M'_t \hat{q}^*_{1,b,t}, \quad \hat{O}^*_t = N_t - \left(\hat{q}^*_{2,b,t}\right)' M_t.$$

Our plug-in estimate of parameter β is now defined by

$$\hat{\beta}_* = \left(\sum_{t=p+1}^{T} \hat{W}^*_t \hat{O}^*_t \left(\hat{O}^*_t\right)'\right)^{-1} \sum_{t=p+1}^{T} \hat{W}^*_t \hat{O}^*_t \hat{V}^*_t.$$

With respect to Theorem 1, we impose more restrictive assumptions.

A3. ξ_1 has moments of any order.

A4. For $j = 0, \ldots, p$, the coefficient a_j is a positive function.

Theorem 2. *Assume that the assumptions* **A1**, **A3** *and* **A4** *hold and* $bT^{\frac{1}{2}-\tau} \to \infty$, $b^2\sqrt{T} \to 0$ *for some* $\tau \in (0, 1/4)$. *Then we have*

$$\sqrt{T}\left(\hat{\beta}_* - \beta\right) \to_{T\to\infty} \mathcal{N}_n\left(0, \left(\mathbb{E}\xi^4_0 - 1\right)\Sigma^{-1}\right),$$

The proof of Theorem 2 is similar to that of Theorem 1 but involves more tedious arguments. A detailed proof of Theorem 2 is given in the supplementary material.

3.2 Estimation of the nonparametric component

Now let us investigate the asymptotic properties for time-varying coefficients estimate $\hat{\alpha}$ defined by (2.7). The estimator $\hat{\beta}$ appearing in the expression (2.7) is constructed using the initial bandwidth parameter b which satisfies the assumptions of Theorem 1.

Theorem 3. *Let* $u \in [0, 1]$. *Assume that the assumptions of Theorem 1 hold. Then if* $b' \to 0$, $b'T \to \infty$ *and* $t = t_T$ *satisfies* $|\frac{t}{T} - u| \le \frac{1}{T}$,

$$\sqrt{Tb'}\left(\hat{\alpha}_t - \alpha(u)\right) + \sqrt{Tb'}\mathbb{E}^{-1}\left(W_1(u)M_1(u)M_1(u)'\right)\left(A_t(u) - A^{\#}_t(u)\right) \to \mathcal{N}_m\left(0, \mathcal{V}(u)\right),$$

where

$$\mathcal{V}(u) = Var\left(\xi^2_1\right) \cdot \int K(x)^2 dx$$
$$\cdot \mathbb{E}^{-1}\left(W_1(u)M_1(u)M_1(u)'\right)\mathbb{E}\left(W_1(u)^2\sigma_1(u)^4 M_1(u)M_1(u)'\right)\mathbb{E}^{-1}\left(W_1(u)M_1(u)M_1(u)'\right).$$

$$A_t(u) = \hat{s}_{1,b',t} - \hat{s}_{2,b',t}\beta - \hat{s}_{3,b',t}\alpha(u),$$

$$A^{\#}_t(u) = \sum_{i=p+1}^{T} k_{t,i}(b')W_i(u)M_i(u)\left(X_i(u)^2 - N'_i(u)\beta - M_i(u)'\alpha(u)\right).$$

This result is similar to that obtained in Fryzlewicz et al. (2008), Proposition 3. The part $A_t(u) - A_t^{\#}(u)$ can be interpreted as a term of deviation with respect to stationarity. As pointed out in Fryzlewicz et al. (2008), this term satisfies $A_t(u) - A_t^{\#}(u) = O_{\mathbb{P}}(b)$. One can easily check that the optimal asymptotic variance in Theorem 3 corresponds to the choice $W_t^* = \frac{1}{\sigma_t^2}$ for the weights. Thus, a plug-in approach is natural. We set $\check{W}_{t,i}^* = \frac{1}{\hat{\sigma}_{t,i}^4 + \mu_T}$, where

$$\hat{\sigma}_{t,i}^2 = M_i' \hat{\alpha}_t + N_i' \hat{\beta}$$

and (μ_T) is a sequence of positive real numbers which now plays the rule of the sequence $(\nu_T)_T$ previously used for the optimal estimator $\hat{\beta}_*$. Once again, this choice is only technical and we impose here $\mu_T = O\left(b' + \frac{1}{\sqrt{Tb'}}\right)$. Then we define the following estimator of parameter α_t.

$$\hat{\alpha}_{*,t} = \check{s}_{3,b',t}^{-1}\left(\check{s}_{1,b',t} - \check{s}_{2,b',t}\hat{\beta}\right),$$

where for $j = 1, 2, 3$, $\check{s}_{j,b',t}$ is obtained as $\hat{s}_{j,b',t}$ but replacing W_i with $\check{W}_{t,i}^*$.

Theorem 4. *Assume that assumptions* **A1**, **A2(4)** *and* **A4** *hold*, $b'T \to \infty$ *and* $\sqrt{Tb'}(b')^{\ell_0} \to 0$ *for a given integer* ℓ_0. *Then, if* $|\frac{t}{T} - u| \le \frac{1}{T}$, *we have*

$$\sqrt{Tb'}\left(\hat{\alpha}_{*,t} - \alpha(u)\right) + \sqrt{Tb'}\mathbb{E}^{-1}\left(\frac{M_1(u)M_1(u)'}{\sigma_1(u)^4}\right)\left(B_t(u) - B_t^{\#}(u)\right) \to \mathcal{N}_m\left(0, \mathcal{V}_*(u)\right),$$

where

$\mathcal{V}_*(u) = Var\left(\xi_1^2\right) \cdot \int K(x)^2 dx \cdot \mathbb{E}^{-1}\left(\frac{M_1(u)M_1(u)'}{\sigma_1(u)^4}\right),$

$B_t(u) = \check{s}_{1,b',t} - \check{s}_{2,b',t}\beta - \check{s}_{3,b',t}\alpha(u),$

$B_t^{\#}(u) = \sum_{i=p+1}^{T} k_{t,i}(b')\check{W}_i^*(u)M_i(u)\left(X_i(u)^2 - N_i'(u)\beta - M_i(u)'\alpha(u)\right).$

Moreover, $B_t(u) - B_t^{\#}(u) = O_{\mathbb{P}}(b').$

Notes

1. Compared to Theorem 3, Theorem 4 uses a more restrictive assumption b'. However for powers of the sample size, i.e $b' = CT^{-\ell}$ with constants $C, \ell > 0$, the conditions are equivalent.

2. When all the coefficients of the volatility are time-varying, replacing M_1 by $\left(1, X_0, \ldots, X_{-p+1}\right)'$, we recover the expression of the optimal asymptotic variance given in Fryzlewicz et al. (2008) for the (local) weighted least-squares estimation. This asymptotic variance coincides with that obtained with the local QML estimator studied in Dahlhaus and Subba Rao (2006). However, a crucial assumption in Theorem 4 is the positivity of all the coefficients of the volatility. To avoid this restriction, Fryzlewicz et al. (2008) consider the sequence of weights $\hat{W}_t = \left(\hat{d}_t + \sum_{j=1}^{p} X_{t-j}^2\right)^{-2}$ where $\hat{d}_t$ is a nonparametric estimate of $\mathbb{E}X_t^2$. For the nonparametric estimation of the whole set of coefficients, they show that the corresponding weighted least squares estimator is asymptotically normal, even if the ARCH coefficients are only nonnegative, but at the price of a small loss of efficiency. We claim that the weights $\hat{W}_t$ can also be used in our context to obtain a result similar to Fryzlewicz et al. (2008), Proposition 4. Details are omitted.

3. As usual for this kind of nonparametric estimation, a better finite sample approximation of the distribution of the parameter estimators can often be obtained using bootstrap methods. With straightforward modifications, it is possible to use the bootstrap method studied in Fryzlewicz et al. (2008) (see Section 5 of that paper) to obtain pointwise confidence intervals for the components of α. Since this paper is mainly devoted to testing and estimating some non-time varying coefficients, we will not consider this bootstrap scheme.

4. Detailed proofs of Theorem 3 and Theorem 4 are given in the supplementary material.

3.3 Asymptotic semiparametric efficiency

For Gaussian inputs (i.e, $\xi \sim \mathcal{N}(0, 1)$), it is possible to show that the matrix $2\Sigma^{-1}$ given in Proposition 1 is a lower bound in semiparametric estimation. We refer the reader to Bickel et al. (1998) for a general introduction to semiparametric models and the problem of efficient estimation of a finite dimensional parameter in such models. In our case case, the problem of semiparametric efficiency for estimating the parameter β involves triangular arrays. This is why we will use an abstract result using the classical formalism presented in van der Vaart and Wellner (1996) (see Chapter 3.11). Intuitively, one can see the matrix $2\Sigma^{-1}$ as the smallest asymptotic variance obtained for estimating β in submodels for which the nuisance parameter α_t is projected onto a finite dimensional space of square integrable functions. Formally, the approach consists in writing a LAN expansion of the likelihood ratio and then using a general convolution theorem. In the sequel, we set for $m, n \geq 1$, $\mathbb{H} = \mathbb{L}^2([0, 1])^m \times \mathbb{R}^n$. Then $\mathbb{H}$ is an Hilbert space for the classical scalar product

$$< (g, h); (\bar{g}, \bar{h}) >_1 = \sum_{i=1}^{m} \int_0^1 g_i(u)\bar{g}_i(u)du + \sum_{j=1}^{n} h_j\bar{h}_j.$$

However, in the sequel, the space $\mathbb{H}$ will be endowed with an equivalent scalar product defined by

$$< (g, h)|(\bar{g}, \bar{h}) >_{\mathbb{H}} = \frac{1}{2} \int_0^1 \begin{pmatrix} g(u) \\ h \end{pmatrix}' E(u) \begin{pmatrix} g(u) \\ h \end{pmatrix} du,$$

where

$$E(u) = \begin{pmatrix} \mathbb{E}\left(\frac{M_1(u)M_1(u)'}{\sigma_1(u)^4}\right) & \mathbb{E}\left(\frac{M_1(u)N_1(u)'}{\sigma_1(u)^4}\right) \\ \mathbb{E}\left(\frac{N_1(u)M_1(u)'}{\sigma_1(u)^4}\right) & \mathbb{E}\left(\frac{N_1(u)N_1(u)'}{\sigma_1(u)^4}\right) \end{pmatrix}.$$

Now, we denote by $\mathcal{L}$ the set of Lipschitz functions $f : [0, 1] \rightarrow \mathbb{R}$ and we set $H = \mathcal{L}^m \times \mathbb{R}^n$ where m (resp. n) is the dimension of vector M_t (resp. N_t). Then H is a linear subspace of $\mathbb{H}$. The set $(H, < \cdot, \cdot >_{\mathbb{H}})$ will be referred to the tangent space. For Gaussian inputs, we first derive a LAN expansion for the (conditional) likelihood ratio. We denote by $P_{T,\alpha,\beta}$ the conditional distribution $(X_{p+1}, \ldots, X_T)|X_1 = x_1, \ldots, X_p = x_p$.

Proposition 2. *Assume that $((\alpha, \beta); (g, h)) \in H^2$ where the coordinates of α and β are positive. Then we have*

$$\log \frac{d\mathbb{P}_{T,\alpha+\frac{g}{\sqrt{T}},\beta+\frac{h}{\sqrt{T}}}}{d\mathbb{P}_{T,\alpha,\beta}} \left(X_{p+1}, \ldots, X_T\right) = \Delta_{T,g,h} - \frac{1}{2}\|(g, h)\|_{\mathbb{H}}^2 + o_{\mathbb{P}_{T,\alpha,\beta}}(1),$$

where

$$\Delta_{T,g,h} = \frac{1}{2\sqrt{T}} \sum_{t=p+1}^{T} \frac{X_t^2 - \sigma_t^2}{\sigma_t^4} \left(M_t'g\left(\frac{t}{T}\right) + N_t'h\right).$$

Moreover,

$$\Delta_{T,g,h} \xrightarrow{\mathcal{D}} \mathcal{N}_n\left(0, \|(g, h)\|_{\mathbb{H}}^2\right).$$

From this LAN expansion, we derive a lower bound for the asymptotic variance of regular estimators of β.

For $(g, h) \in H$, we set $\kappa_T(g, h) = \beta + \frac{h}{\sqrt{T}}$. Then, the sequence of parameters $\{\kappa_T(g, h) : (g, h) \in H\}$ is regular: if $\dot{\kappa} : \mathbb{H} \to \mathbb{R}^n$ is the projection operator defined by $\dot{\kappa}(g, h) = h$, then

$$\sqrt{T} \left(\kappa_T(g, h) - \kappa_T(0, 0) \right) = h.$$

Corollary 1. *If the assumptions of Proposition 2 hold, then the adjoint operator $\dot{\kappa}^* : \mathbb{R}^n \to \mathbb{H}$ of $\dot{\kappa}$ is given by*

$$\dot{\kappa}^* v = \begin{pmatrix} -q_2(\cdot) \\ I_n \end{pmatrix} \left(\frac{1}{2} \Sigma \right)^{-1} v, \quad v \in \mathbb{R}^n.$$

Consequently, the limit distribution of a regular estimator of β equals the distribution of a sum $L_1 + L_2$ of independent random vectors of $\mathbb{R}^n$ and such that

$$L_1 \sim \mathcal{N}_n \left(0, 2\Sigma^{-1} \right).$$

4 Statistical testing

4.1 Testing parameter constancy

For a real data set, it is necessary, before applying the methodology given in Section 2, to test if a coefficients vector of the form $\beta = (a_{r_1}, \ldots, a_{r_n})$ is time-varying or not in model (1.1). This is equivalent to test model (2.1) versus model (1.1). When $n = p$ and $\beta = \left(a_1, \ldots, a_p \right)$, such a statistical test is interesting for deciding if model (1.3) is a convenient restriction of the tv-ARCH model. This case is of particular interest for real data applications. In Zhang and Wu (2012), a procedure is proposed for testing if some coefficients are constant in a general time-varying regression model. The null hypothesis is H_0: $\beta(\cdot)$ constant. The test statistic used in Zhang and Wu (2012) is based on a $\mathbb{L}^2$ distance between an estimate under the alternative and an estimate under the null hypothesis. In this part, we derive asymptotic properties of this test for tv-ARCH processes. For simplicity, we will only consider some estimates without plug-in (i.e we fix a sequence of weights $(W_t)_t$ of the form (2.3) and use the corresponding least-squares estimates). Let us first introduce some additional notations.

For a function $f : [-1, 1] \to \mathbb{R}$, we set $\|f\|_2 = \sqrt{\int_{-1}^{1} f(u)^2 du}$ and for $x \in [-1, 1]$,

$$K^*(x) = \int_{-1}^{1-2|x|} K(v) K(v + 2|x|) \, dv.$$

Setting for $u \in [0, 1]$, $e_i(u) = \frac{1}{Tb} K \left(\frac{uT-i}{Tb} \right)$ and for $p + 1 \le t \le T$, $X_t = (M_t', N_t')'$, the kernel estimate of the full vector of ARCH coefficients $a(u) = (\alpha(u)', \beta(u)')'$ is given by (see Fryzlewicz et al. (2008))

$$\widetilde{a}(u) = S_u^{-1} \sum_{i=p+1}^{T} e_i(u) W_i X_i^2 \mathcal{X}_i,$$

where $S_u = \sum_{i=p+1}^{T} e_i(u) W_i \mathcal{X}_i \mathcal{X}_i'$. Then we set $\widetilde{\beta}(u) = A\widetilde{a}(u)$ where A is the matrix of size $(p + 1) \times n$ defined

by $Ax = \begin{pmatrix} x_{m+1} \\ \vdots \\ x_{p+1} \end{pmatrix}$. Note also that $\beta(u) = Aa(u)$. We also set $\kappa_u = \mathbb{E}\left(W_1(u)X_1(u)X_1(u)'\right)$ and

$$
\begin{aligned}
O(u) &= \kappa_u^{-1} \cdot \mathbb{E}\left(W_1(u)^2 \left(X_1(u)^2 - \sigma_1^2(u)\right)^2 X_1(u)X_1(u)'\right) \cdot \kappa_u^{-1} \\
&= \operatorname{Var}\left(\xi_1^2\right) \kappa_u^{-1} \cdot \mathbb{E}\left(W_1(u)^2 \sigma_1(u)^4 X_1(u)X_1(u)'\right) \cdot \kappa_u^{-1}.
\end{aligned}
$$

Let $(\Gamma(u))_{u \in [0,1]}$ be a family of positive definite matrices of size $(p+1) \times (p+1)$ such that $u \mapsto \Gamma(u)$ is a Lipschitz function. Finally, we set for $j = 1, 2$,

$$
\varpi_j = \int_0^1 \operatorname{Tr}\left[\Gamma(u)^{1/2} A O(u) A' \Gamma(u)^{1/2}\right]^j du.
$$

We define our test statistic $\mathcal{S}_T(\widetilde{a}, \beta)$ by

$$
\mathcal{S}_T(\widetilde{a}, \beta) = \int_0^1 \left(\widetilde{\beta}(u) - \beta\right)' \Gamma(u) \left(\widetilde{\beta}(u) - \beta\right) du. \tag{4.1}
$$

The proof of the following theorem is given in the supplementary material.

Theorem 5. *Assume that assumptions* **A1***,* **A2(8)** *hold and β is non time-varying. Then if $Tb^2 \to \infty$ and $Tb^{3.5} \to 0$, we have*

$$
T\sqrt{b}\left\{\mathcal{S}_T(\widetilde{a}, \beta) - \frac{\|K\|_2^2 \varpi_1}{Tb}\right\} \to \mathcal{N}\left(0, 4\|K^*\|_2^2 \varpi_2\right). \tag{4.2}
$$

Moreover (4.2) holds if $\mathcal{S}_T(\widetilde{a}, \beta)$ is replaced with $\mathcal{S}_T\left(\widetilde{a}, \hat{\beta}\right)$ where $\hat{\beta}$ is the estimate of Theorem 1.

Notes

1. As pointed out in Zhang and Wu (2012), if we are interested in prediction, the matrix Γ can be chosen as the asymptotic variance of the kernel estimate $\hat{\beta}(\cdot)$ (which has to be estimated in practice). In our numerical studies, we will use the simple choice $\Gamma(u) = I_n$ where I_n denotes the identity matrix of size n.

2. Quantities ϖ_1 and ϖ_2 involved in the bias and asymptotic variance in (4.2) can be estimated consistently, taking empirical counterpart. Then we obtain a pivotal statistic
   $\mathcal{E}_T = T\sqrt{b}\left\{\mathcal{S}_T\left(\widetilde{a}, \hat{\beta}\right) - \frac{\|K\|_2^2 \widehat{\varpi}_1}{Tb}\right\} / \left(2\|K^*\|_2 \sqrt{\widehat{\varpi}_2}\right)$ and one can reject the null hypothesis for large values of this statistic. However, in practice, such nonparametric tests suffer from the slow convergence in Theorem 5. As in Zhang and Wu (2012), one can use a Monte-Carlo type procedure which can improve the finite-sample performance (a similar Monte Carlo procedure is also used in Patilea and Raïssi (2014)). Note that the result of Theorem 5 is valid for i.i.d series with a standard Gaussian marginal distribution. In particular, if $\mathcal{E}_T^*$ denotes the pivotal statistics computed with an i.i.d sample of standard Gaussian variables, we have $\lim_{T \to \infty} \mathcal{E}_T^* = \lim_{T \to \infty} \mathcal{E}_T = \mathcal{N}(0, 1)$ in distribution. Then one can use the quantiles of the distribution of $\mathcal{E}_T^*$ to compute the critical values for the test (instead of the Gaussian quantiles). Let us give the details of the method proposed in Zhang and Wu (2012). We assume that the bandwidth b has been already selected.

13

- First simulate B samples of size T of i.i.d Gaussian random variables. For each sample, compute the values of the estimators $\widetilde{\beta}(\cdot)$ and $\hat{\beta}$ as well as the realization of the pivotal statistics $\mathcal{P}_T^*$.
- Then, from these B realizations of the random variable $\mathcal{E}_T^*$, compute the empirical quantile $q_{MC}(\alpha)$ of order $1 - \alpha$.
- Reject H_0 if $\mathcal{E}_T$ is greater than $q_{MC}(1 - \alpha)$.

3. In Zhang and Wu (2012), the power of the test under some local alternatives is studied (see Theorem 3.2 of this paper). A similar result can be derived here under the assumptions of Theorem 5. In particular, for some local alternatives of the form $\beta_1(t/T) = \beta + z_T f(t/T)$, with $1/\sqrt{T \sqrt{b}} = o(z_T)$ and f a Lipschitz function defined over $[0, 1]$, the power of the test still converges to 1.

4.2   Testing if a constant parameter is equal to zero

In this part, we consider model (2.1) and our goal is to test whether the vector β is equal to zero. Two approaches are discussed below.

1. One possibility is to use the asymptotic normality of the estimator $\hat{\beta}$ given in Theorem 1. Under the null hypothesis H_0: $\beta = 0$, the asymptotic distribution of $\sqrt{T}\hat{\beta}$ is that of a centered Gaussian vector with covariance matrix $\mathcal{V} = \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$. We have already discussed how to estimate the covariance matrix $\mathcal{V}$. If $\hat{\mathcal{V}}$ denotes such an estimate, the statistics $T\|\hat{\mathcal{V}}^{-\frac{1}{2}}\hat{\beta}\|^2$ is asymptotically distributed as a χ^2 with n degrees of freedom (here $\|\cdot\|$ denotes the euclidean norm on $\mathbb{R}^n$). As for the test given in the previous subsection, one can use a Monte Carlo method instead of using the quantiles of the asymptotic distribution (the convergence in distribution of the previous statistics is quite slow because of the incorporation of nonparametric kernel estimates). If a bandwidth b is selected, one can simulate B samples of Gaussian i.i.d random variables and compute the corresponding values of our statistics. From these values, we can compute the empirical quantile $q_{MC,\alpha}$ of order $1 - \alpha$ and reject the null hypothesis if $T\|\hat{\mathcal{V}}^{-\frac{1}{2}}\hat{\beta}\|^2 > q_{MC,\alpha}$. This test has an asymptotic level α and a power converging to 1 under a fixed alternative. However such approach is not completely natural because the value $\beta = 0$ is on the boundary of the parameter space, our test is similar to the bilateral test for testing the hypothesis $\beta = 0$ in regression models and we ignore the sign of β. This will result in a loss of power and it is more natural to consider a statistics based on the random vector $\left(\sqrt{T}\max\left(\hat{\beta}_i, 0\right)\right)_{1 \leq i \leq n}$. The vector $\left(\max(\hat{\beta}_i, 0)\right)_{1 \leq i \leq n}$ will be called truncated least squares estimator. As discussed in Francq and Zakoïan (2008) for stationary ARCH processes, truncated least squares estimators are natural for testing if some lag coefficients are equal to zero. However, the limiting distribution of this truncated random vector is that of $(\max(Z_i, 0))_{1 \leq i \leq n}$ where $Z = (Z_1, \ldots, Z_n)$ is a Gaussian vector, with dependent entries in general. Then, except if $\text{Var}(Z)$ is diagonal, it is not possible to get a pivotal statistics from truncated least squares estimators. Then the Monte-Carlo method used for testing parameter constancy cannot be applied. Note also that bootstrapping the model is not appropriated here because the bootstrap is generally inconsistent for testing a parameter on the boundary (see for instance Andrews (2000) for this problem). However, when β denotes the full vector of lag coefficients, it is possible to use truncated least squares and the Monte Carlo method, provided $W_t \equiv 1$. This point is discussed below.

2. When β is the full vector of lag coefficients, the problem is to test model (1.2) versus model (1.3). For testing if the lag coefficients are equal to zero in model (1.3), our test is based on the following result. The following notation will be used. If $p + 1 \leq t \leq T$, we set $\hat{d}_t = \sum_{i=p+1}^{T} k_{t,i}(b)X_i^2$. Note that $\hat{d}_t$ is an estimator of $\mathbb{E}X_t^2$.

Proposition 3. *Assume that* **A2(4)** *holds and that* $b \in [cT^{-h}, CT^{-h}]$ *with* $\frac{1}{4} < h < \frac{3}{4} - \frac{1}{2(1+\delta)}$, *where* c, C *are positive constants. Then under* H_0,

$$\left(\hat{a}_1, \ldots, \hat{a}_p\right)' = \arg \min_{a_1,\ldots,a_p} \sum_{t=p+1}^{T} \left(X_t^2 - \hat{d}_t - \sum_{j=1}^{p} a_j \left(X_{t-j}^2 - \hat{d}_{t-j} \right) \right)^2$$

satisfies $T \sum_{j=1}^{p} \max\left(\hat{a}_j, 0\right)^2 \to_{\mathcal{D}} \sigma^2 \sum_{j=1}^{p} \max\left(Z_j, 0\right)^2$, *where* Z *standard Gaussian vector and*

$$\sigma^2 = \frac{\int_0^1 Var\left(X_0(u)^2\right)^2 du}{\left(\int_0^1 Var\left(X_0(u)^2\right) du\right)^2}.$$

Testing the second order dynamic. From this result, we reject H_0 for large values of the statistics $\Psi_T = T \sum_{j=1}^{p} \max\left(\hat{a}_j, 0\right)^2 / \hat{\sigma}^2$ where $\hat{\sigma}^2$ is a consistent estimator of σ^2. Typically one can choose

$$\hat{\sigma}^2 = T \sum_{t=1}^{T} \hat{a}_0(t/T)^4 / \left(\sum_{t=1}^{T} \hat{a}_0(t/T)^2 \right)^2 \quad \text{with } \hat{a}_0(t/T) = \hat{d}_t,$$

which gives a consistent estimate under H_0. A first solution is to reject H_0 if Ψ_T is larger than the quantile of order $1 - \alpha$ of the distribution of the random variable $\sum_{j=1}^{p} \max\left(Z_j, 0\right)^2$. But the statistics Ψ_T is also asymptotically pivotal and its quantiles can be approximated by a Monte Carlo procedure similar to that used for testing parameter constancy.

Notes

1. Note that the estimators of the lag coefficients corresponds to the estimators (2.6) with $W_t \equiv 1$. The optimal rate of convergence $h = \frac{1}{3}$ for kernel estimation of Lipschitz functionals can be used if $\delta > \frac{1}{5}$ in assumption **A2(4)**. A proof of Proposition 3 is given in the supplementary material.

2. In the stationary case, two benchmark tests are usually used for testing the second order dynamic: the Lagrange multiplier test of Engle (1982) and the portmanteau test of McLeod and Li (1983). Patilea and Raïssi (2014) have recently extended these two tests for model (1.3), taking in account of nonstationarity. Here we provide an alternative test based on a direct estimation of the lag coefficients. A comparison of the different approaches is beyond the scope of this paper. Let us observe that in the stationary case, the constant σ^2 in Proposition 3 is equal to 1 whereas in the nonstationary case, this constant σ^2 is a correction factor which can be written

$$\sigma^2 = \int_0^1 a_0(u)^4 du / \left(\int_0^1 a_0(u)^2 du \right)^2 > 1.$$

The correction factor σ^2 also appears in the asymptotic results of Patilea and Raïssi (2014) (see the ratio $\overline{\omega}_4^2 / \overline{\omega}_8$ appearing in the two statistics used in that paper). Ignoring this factor leads to an oversized test and the null hypothesis will be often rejected when the data are independent but not identically distributed. Moreover, let us notice that our moment condition for the noise distribution is less restrictive for applying the test (a moment greater than 8 is assumed in Patilea and Raïssi (2014)).

3. One can study the power of our test under local alternatives of type $\sqrt{T}(a_1, \ldots, a_p)' \to s$ where s is a vector of nonnegative real numbers. One can show that if $\mathbb{E}\xi_0^{8(1+\delta)} < \infty$ and the bandwidth b satisfies $Tb^2 \to \infty$ and $Tb^4 \to 0$, then

$$\mathbb{P}\left(T \sum_{j=1}^{p} \max\left(\hat{a}_j, 0\right)^2 / \hat{\sigma}^2 > q_{1-\alpha}\right) \to \mathbb{P}\left(\sum_{j=1}^{p} \max\left(\frac{s_j}{\sigma} + Z_j, 0\right)^2 > q_{1-\alpha}\right),$$

where $q_{1-\alpha}$ is the quantile of order $1 - \alpha$ of the distribution of $\sum_{j=1}^{p} \max(Z_j, 0)^2$. Details are omitted.

5 Real data applications

Before real data applications, let us provide some recommendations on the choice of tuning parameters.

5.1 Choice of tuning parameters

The practical implementation of our estimation and testing procedures requires the choice of some weights W_t, some bandwidth parameters as well as the number of lags p in the model. In this subsection, we discuss the practical choices of these parameters. In all our studies, the kernel W will be the Epanechnikov kernel. For simplicity, only one bandwidth parameter will be selected for the semiparametric models (this means that we set $b' = b$ in (2.7)). We use a cross-validation method as specified below. The selected bandwidth will be used for the tests. For the tests, the Monte-Carlo procedure will be always applied with $B = 2000$ samples of i.i.d standard Gaussian random variables.

- In practice, a sequence of weights $(W_t)_t$ has to be chosen for applying our method. One possibility is to use the weights $W_t = \left(1 + \sum_{j=1}^{p} X_{t-j}^2\right)^{-2}$ suggested in Horváth and Liese (2004). In our implementation, we use the weights $W_t = \left(\hat{v} + \sum_{j=1}^{p} X_{t-j}^2\right)^{-2}$ where $\hat{v} = \frac{1}{T}\sum_{t=1}^{T} X_t^2$ is an estimate of the average of the variance $v = \int_0^1 \mathbb{E}\left(X_0(u)^2\right) du$. There are several advantages in using these weights. First, the lag estimates obtained in model 1.3 do not depend on the scale of the returns, a property always satisfied for the true lag coefficients (if W_t denotes the price at time t, $X_t = \log(P_t) - \log(P_{t-1})$ or $100 \times X_t$ are two different scales used in practice). Moreover, for stationary Arch processes, this choice is equivalent to the weights used in Fryzlewicz et al. (2008). We also noticed better finite sample performances for our tests and inference procedures with this choice. However, the introduction of the random quantity $\hat{v}$ is not taken in account in our theoretical results. But, inspection of our proofs shows that the conclusions of the theorems remain unchanged if $\hat{v} - v = o_{\mathbb{P}}\left(T^{-1/4}\right)$. The latter condition is satisfied if $\mathbb{E}|X_t|^h < \infty$ for $h > \frac{8}{3}$ (this can be justified using the moment inequality given in Fryzlewicz et al. (2008), see Lemma A2). A sufficient condition for the finiteness of this moment is $\mathbb{E}^{1/h}\left(|\xi_0|^h\right)\sup_{u \in [0,1]}\sum_{j=1}^{p} a_j(u) < 1$ which is more restrictive than the initial condition given in **A3**. Despite this slight restriction, we only consider the aforementioned sequence of weights in the sequel.

- For model (1.1), we use the cross-validation method considered in Fryzlewicz et al. (2008). The bandwidth parameter is selected by minimizing the function

$$b \mapsto \sum_{t=p+1}^{T} W_t \left(X_t^2 - \mathcal{X}_t' \hat{a}_t^{(-t)}(b)\right)^2,$$

where $\hat{a}_t^{-(t)}(b) = \left(\sum_{\substack{k=p+1 \\ k \neq t,\dots,t+p}}^{T} K\left(\frac{t-k}{Tb}\right) W_k \mathcal{X}_k^2 \mathcal{X}_k' \right)^{-1} \sum_{\substack{k=p+1 \\ k \neq t,\dots,t+p}}^{T} K\left(\frac{t-k}{Tb}\right) W_k X_k^2 \mathcal{X}_k.$

- For model (1.3), we choose the bandwidth b by minimizing

$$(\beta, b) \mapsto \sum_{t=p+1}^{T} W_t \left(X_t^2 - \hat{q}_{1,b,t}^{(-t)} - \left(N_t - \hat{q}_{2,b,t}^{(-t)}\right)' \beta \right)^2.$$

Here for $n = 1, 2$, $\hat{q}_{n,b,t}^{(-t)}$ is the version of $\hat{q}_{n,b,t}$ (see (2.5)) defined as $\hat{a}_t^{-(t)}(b)$, $N_t = \left(X_{t-1}^2, \dots, X_{t-p}^2\right)'$ and $\beta = (a_1, \dots, a_p)'$. This type of cross-validation is a weighted version of the method proposed by Hart (1994) for AR models with a time-varying mean. Note that minimizing the latter function with respect to β, b being fixed, leads to an estimate close to the estimate $\hat{\beta}$ defined in (2.6).

- Finally, we discuss the selection of the number of lags p in model 1.1. An information criterion has been studied recently by Zhang and Wu (2012) for time-varying regression models. It is possible to adapt the approach used by these authors to our setting. To this end, we define for $p = 0, 1, \dots, q$,

$$C(p) = \log\left[\sum_{t=p+1}^{T} W_t^{(q)} \left(X_t^2 - \mathcal{X}_i' \hat{a}^{(p)}(t/T)\right)^2 \right] + \zeta_T(p + 1),$$

where $\mathcal{X}_t = \left(1, X_{t-1}^2, \dots, X_{t-p}^2\right)'$, $W_t^{(q)} = \left(\hat{v} + \sum_{j=1}^{q} X_{t-j}^2\right)^{-2}$, $\hat{a}^{(p)}$ are the coefficients estimates obtained for the tv-ARCH with p lags but computed with the weights $W^{(q)}$ and ζ_T is a vanishing sequence of positive numbers. The goal of the selection procedure is to minimize $p \mapsto C(p)$ in the spirit of AIC or BIC criterion used for regression models. The bandwidth b is selected by cross validation for the tv-ARCH model with q lags. Of course, condition on the decrease of ζ_T has to be imposed to get consistency (i.e $\mathbb{P}\left(\arg\min_{0 \leq p \leq q} C(p) = p_0\right) \to 1$ if $p_0 \leq q$ is the true number of lags). Inspecting the proofs of Lemma A6 and Theorem 3.3 in Zhang and Wu (2012), we find that condition $T^{2/3}\zeta_T \to \infty$ guarantees consistency (using the arguments of the proof of Theorem 5, one can show that the quantity $\phi_T(\phi_T + \rho_T)$ in Lemma A6 of Zhang and Wu (2012) can be replaced with $T^{-2/3}$ in our context). For real data applications we choose $\zeta_T = \log(\log(T))/(Tb)$ where b is selected using cross-validation. Our intensive simulation study reported in the supplementary material shows that this choice gives reasonable performances. This choice can be also justified using the argument given in Zhang and Wu (2015): $(p + 1)/b$ can be seen as the effective number of parameters in kernel smoothing. Hence our choice has a similarity with the Hannan-Quinn information, except that we gave up the constant $2c$ with $c > 1$ used in Hannan and Quinn (1979). We found that adding such a factor underestimates the order in our case. A precise justification of our choice using a version of the law of iterated logarithm is not the goal of this paper. However, one can notice that applying cross-validation on an interval of type $\left[cT^{-1/3}, CT^{-1/3}\right]$ is compatible with our consistency condition. In the applications, the maximal number of lags is set to $q = 10$.

Using the approach described above, we conducted an extensive simulation study. Some numerical experiments are reported in the supplementary material and show the good behavior of our method for various simulation setups. This simulation study shows that our estimators and tests have reasonable performances for the sample sizes considered in the sequel. Here we only report the results obtained with real data.

For real data applications, the bandwidth will be always chosen using CV over a grid of the form $[c, C] \times T^{-1/3}$ where c and C are two positive constants (we recall that $T^{-1/3}$ is the optimal rate of convergence for the bandwidth in the nonparametric estimation of a Lipschitzian regression function). We will also use some acronyms for our models. Model tv(p) denotes the tvARCH model with p lags (the case $p = 0$ refers to model (1.2)) while model sptv(p) denotes model (1.3) with $p \geq 1$. In the sequel, we consider two currency exchange rates and one stock market index. The log returns $X_t = \log(P_t) - \log(P_{t-1})$ of the initial series $(P_t)_t$ will be modelized with ARCH processes.

5.2   Exchange rate USD/Euro

In this subsection, we study the exchange rate series between the US Dollar and the Euro. We consider the period from January 03, 2000 to February 13, 2015. The sample size is $T = 3799$. As usual for this type of series, the autocorrelograms suggest correlation for the squares of the transformed series.
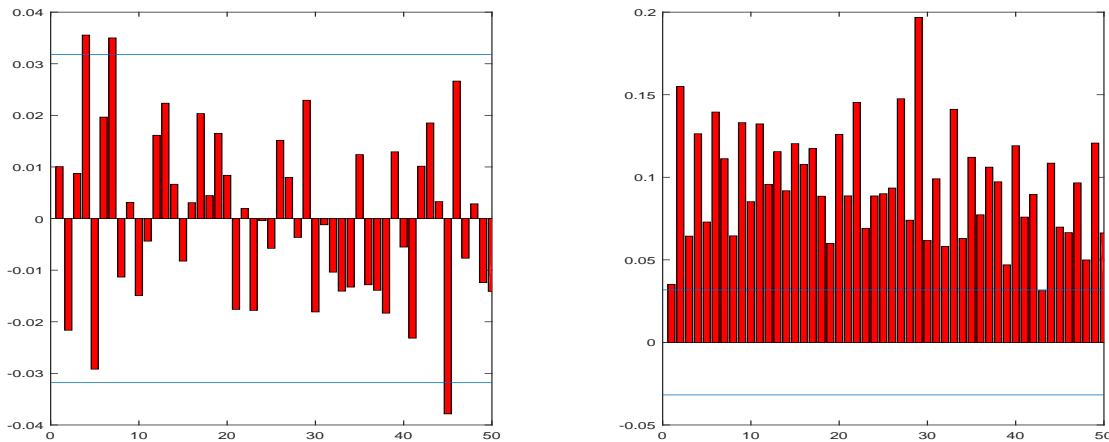

Figure 1: Autocorrelogram and autocorrelogram of the squares for the logged and differenced daily exchange rates USD/Euro

For this data set, the information criterion selects $p = 0$. To confirm this choice, we apply our procedure with $p = 2$. The results for testing the hypothesis of non time-varying coefficients are reported in Table 1. The intercept function seems not constant in contrast to the lag coefficients for which it is not possible to reject the null hypothesis. Fitting a sptv(2) process gives small negative values for the lag coefficients and it is of course not necessary to test if they are equal to zero. This conclusion suggests an absence of second order dynamic for this series. We refer the reader to Granger and Stărică (2005) and Herzel et al. (2006) for other analysis suggesting a similar behavior for some financial time series.

18

Non t-v a_0	Non t-v a_1	Non t-v a_2	Non t-v (a_1, a_2)	$\hat{b}$
0.0005	0.138	0.1645	0.6415	0.028

Table 1: The $p-$values for testing the hypothesis of non time-varying coefficients (the first line gives the null hypothesis)

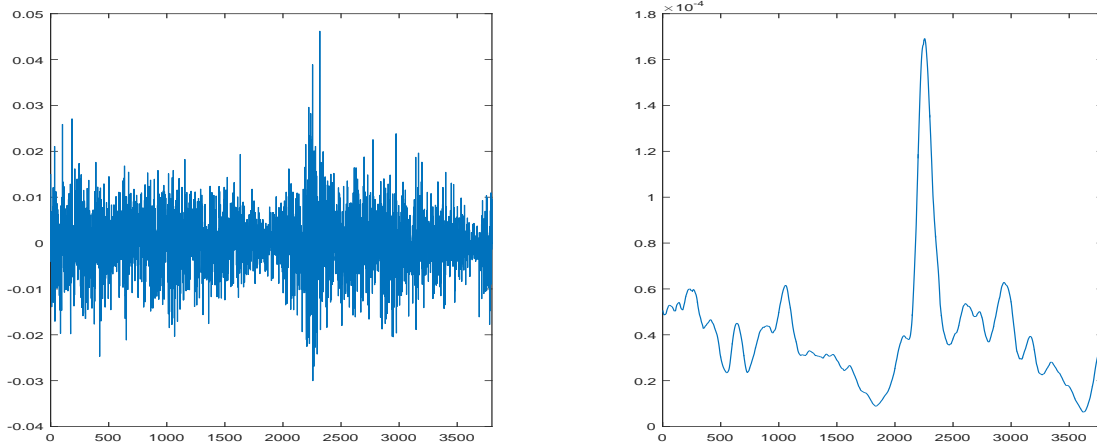A plot of the series with the final estimate of the intercept function is given Figure 2.

Figure 2: Logged and differenced daily exchange rates between USD and Euro and estimation of the unconditional variance

5.3   A second example: the exchange rates between the US Dollar and the Indian Rupee

In this subsection, we analyze the exchange rates series between the US Dollar and the Indian Rupee over the period starting from December 19, 2005 to February 18, 2015. The sample size is $T = 2303$. The information criterion selects $p = 1$ lag for this series. From the $p-$values reported in Table 2, the hypothesis of a constant intercept function is clearly rejected but it is not possible to reject the assumption of a non time-varying first lag coefficient. Fitting a sp(1) process gives a small but significant lag estimates (the $p-$value for testing the second order dynamic is less than 10^{-4}). In contrast, fitting a stationary ARCH process with one lag (one can simply use $b = 1$ and our procedure) leads to $\hat{a}_1 = 0.3041$ (s.e 0.0717) and several significant lag estimates are found for larger values of p.

Non t-v a_0	Non t-v a_1	$\hat{b}_{tv}$	$\hat{a}_1$	$\hat{b}_{sptv}$
$< 10^{-4}$	0.4215	0.035	0.1527 (s.e 0.0688)	0.028

Table 2: Test and estimation for the USD/Rupee series ($p-$values for the test, the selected bandwidths for fitting a tv(1)/sptv(1) process and estimation of the first lag coefficients)
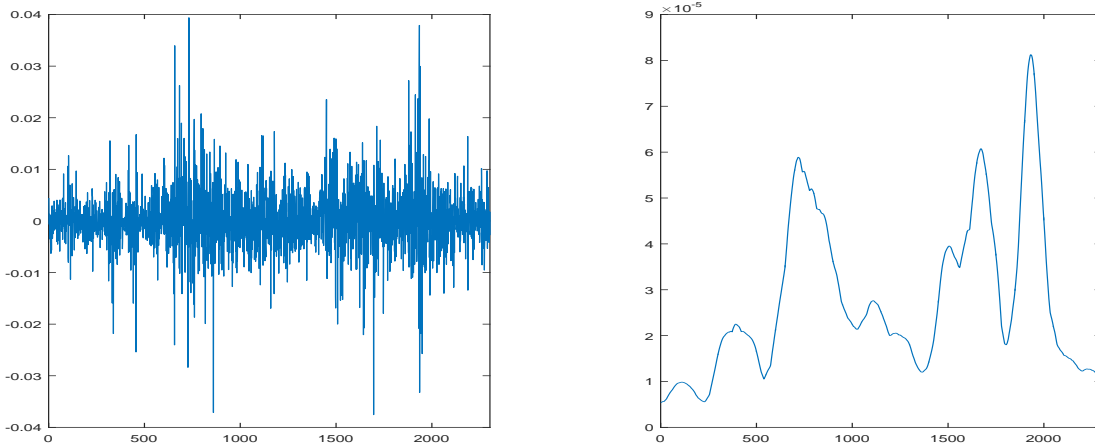
Figure 3: Logged and differenced daily exchange rates between the US Dollar and the Indian Rupee and estimation of the intercept function

5.4 A classical stock market index: the FTSE

Finally, we consider the closing values of the FTSE index from January 04, 2005 to March 04, 2015, taking as usual the logged and differenced daily returns. Our information criterion selects $p = 5$ lags. Testing constancy of the intercept function gives a $p-$value of 3×10^{-3} and the $p-$value for testing constant lag coefficients is 0.066. Hence, the assumption of constant lag coefficients is rejected at level $\alpha = 10\%$ (testing constancy of the third and fourth coefficients gives the p-values 0.084 and 0.067 respectively, the other $p-$values exceed 10%) and considering a tv-ARCH process for this data set could be interesting. We also fit a sptv(5)−process. The estimated lag parameters and their standard errors are reported in Table 3. The $p-$value for testing the absence of second order dynamic is close to zero.

$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$
0.0547 (s.e 0.0321)	0.1155 (s.e 0.0320)	0.1204 (s.e 0.0311)
$\hat{a}_4$	$\hat{a}_5$	$\hat{b}_{SP}$
0.0942 (s.e 0.0367)	0.1201 (s.e 0.0324)	0.063

Table 3: Estimated values of the lag coefficients and selected bandwidth for the FTSE

Here selecting the number of lags is important because fitting a sptv(2)−process for instance does not give significant lag estimates and the selected bandwidth for $p = 2$ is very small. In Fryzlewicz et al. (2008), it is suggested that stationnary GARCH models give better forecasts for stock market indices than tvARCH processes and that this result could be explained by a more stationarity behavior of these series with respect to currency exchange rates. This observation is compatible with our analyze of the FTSE index on this period of time which suggests that adding non linearity has a tendency to take away non stationarity, with larger selected bandwidths. However, Figure 4 shows that incorporating a time-varying unconditional variance significantly reduces the values of ARCH parameters. In Figure 4, two extreme cases are observed. When

20

$b \to 0$, the sum of lag coefficients becomes arbitrary small whereas the value $b = 1$ (which corresponds to the fitting of a stationnary ARCH process) leads to larger lag estimates. Moreover, in fitting a sptv(5) process, the ratio $\sqrt{\widehat{a_0(u)}}/\hat{\sigma}_t$ has an average of 0.75 (s.e 0.14) which means that the contribution of the time-varying intercept has a strong contribution to volatility.
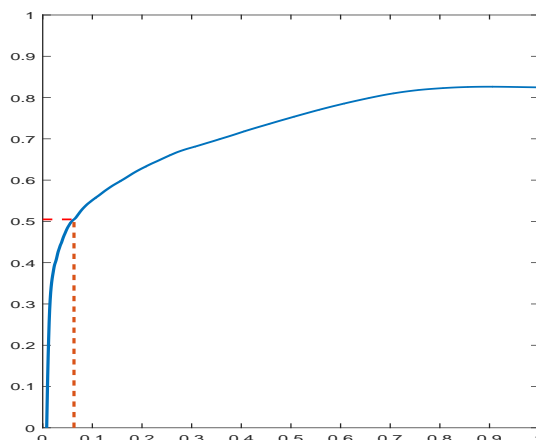

Figure 4: Sum of the first five lag coefficients with respect to the value of the bandwidth b (red dashed lines correspond to the bandwidth selected using our method)

References

D.K.W Andrews. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68:399–405, 2000.

P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag New York, 1998.

A. Bose and K. Mukherjee. Estimating the arch parameters by solving linear equations. *J. Time Ser. Anal.*, 24:127–136, 2003.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25:1–37, 1997.

R. Dahlhaus and S. Subba Rao. Statistical inference for time-varying arch processes. *Ann. Statist.*, 34: 1075–1114, 2006.

R. Dahlhaus and S. Subba Rao. A recursive online algorithm for the estimation of time-varying arch parameters. *Bernoulli*, 13:389–422, 2007.

J. Dedecker, P. Doukhan, J.-R. Leon, S. Louhichi, and C. Prieur. *Weak dependence, with examples and some applications*. Springer Berlin, 2007.

R. Engle. Autoregressive conditional heteroscedastic with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1007, 1982.

R. Engle and J.G. Rangel. The spline-garch model for low-frequency volatility and its global macroeconomic causes. *Rev. Financ. Stud.*, 21:1187–1222, 2008.

C. Francq and J-M. Zakoïan. Estimating arch models when the coefficients are allowed to be equal to zero. *Austrian J. Stat.*, 37:31–40, 2008.

P. Fryzlewicz, T. Sapatinas, and S. Subba Rao. Normalized least-squares estimation in time-varying arch models. *Ann. Statist.*, 36:742–786, 2008.

C. Granger and C. Stărică. Nonstationarities in stock returns. *Review of Economics and Statistics*, 87:503–522, 2005.

E.T. Hannan and B. G. Quinn. The determination of the order of an autoregression. *J. R. Statist. Soc. B*, 41(2):190–195, 1979.

W. Härdle, H. Liang, and J. Gao. *Partially Linear Models*. Springer Physica-Verlag, 2000.

J. D. Hart. Automated kernel smoothing of dependent data by using time series cross-validation. *J. R. Statist. Soc. B*, 56(3):529–542, 1994.

S. Herzel, C. Stărică, and R. Tütüncu. *A non-stationary paradigm for the dynamics of multivariate financial returns*, volume 187 of *Lecture Notes in Statistics*. Springer, 2006.

L. Horváth and F. Liese. l_p- estimators in arch models. *J. Statist. Plann. Inference*, 119:277–309, 2004.

A.I. McLeod and W.K. Li. Diagnostic checking arma time series models using squared-residual autocorrelations. *J. Time Series Anal.*, 4:269–273, 1983.

T. Mikosch and C. Stărică. Non-stationarities in financial time series, the long-range dependence and the igarch effect. *The Review of Economics and Statistics*, 86:378–390, 2004.

V. Patilea and H. Raïssi. Testing second-order dynamics for autoregressive processes in presence of time-varying variance. *J. Amer. Statist. Assoc.*, 109:1099–1111, 2014.

D. Pollard. *Convergence of stochastic processes*. Springer Verlag, 1984.

S. Subba Rao. On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in App. Probab.*, 38:1155–1172, 2006.

A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer Verlag, New-York, 1996.

T. Zhang and W.B. Wu. Inference of time-varying regression models. *Ann. Statist.*, 40:1376–1402, 2012.

T. Zhang and W.B. Wu. Time-varying nonlinear regression models: Nonparametric estimation and model selection. *Ann. Statist.*, 43:741–768, 2015.

Supplementary material

A Auxiliary results for the proofs

In this subsection, we consider a general time-varying ARCH process $(X_t)_{1 \le t \le T}$ defined by

$$X_t = \xi_t \sqrt{a_0\left(\frac{t}{T}\right) + \sum_{j=1}^{p} a_j\left(\frac{t}{T}\right) X_{t-j}^2}, \quad p + 1 \le t \le T.$$

The different coefficients a_j can be time-varying or not and we assume that assumption **A1** is satisfied. Let us introduce additional notations.

- We will always denote by $\|\cdot\|$ the euclidean norm on $\mathbb{R}^g$ for an arbitrary positive integer g. The corresponding operator norm on $\mathcal{M}_g$, the set of matrices of size $g \times g$ and with real coefficients, will be also denoted by $\|\cdot\|$.

- If X is an integrable random variable taking values in $\mathcal{M}_g$, we set $\bar{X} = X - \mathbb{E}(X)$.

- For a sequence $b = b_T \in (0, 1)$ of bandwidths, we recall the notation

$$k_{t,i}(b) = \frac{\frac{1}{Tb} K\left(\frac{t-i}{Tb}\right)}{\frac{1}{Tb} \sum_{j=p+1}^{T} K\left(\frac{t-j}{Tb}\right)}, \quad p + 1 \le i, t \le T.$$

 Note that $\max_{p+1 \le i, t \le T} k_{t,i}(b) = O\left((Tb)^{-1}\right)$. This bound will be extensively used in the sequel.

- For $1 \le t \le T$, we set $\mathcal{F}_t = \sigma\left(\xi_s : s \le t\right)$.

- Finally, we set $Z_t = \xi_t^2 - 1$ for $t \in \mathbb{N}$. Then Z_t is centered.

- **Important notations**: for simplicity of notations, the quantities $\hat{s}_{j,b,t}$ appearing in the statements of Theorems 1 and 3 will be simply denoted by $S_{j,b,t}$ for $j = 1, 2, 3$.

We first give a lemma about the regularity of tv-ARCH processes. The following result is crucial for deriving asymptotic properties of our estimators and it is a direct consequence of Theorem 1 in Dahlhaus and Subba Rao (2006) (see also Subba Rao (2006), Theorem 2.1 and the discussion in Section 5.2).

Lemma 1. *1. There exists a constant $C > 0$ such that for all $(u, v, T) \in [0, 1]^2 \times \mathbb{N}^*$,*

$$\mathbb{E}|X_1(u)^2 - X_1^2(v)| \le C|u - v|, \quad \max_{1 \le t \le T} \mathbb{E}|X_t^2 - X_t^2\left(\frac{t}{T}\right)| \le \frac{C}{T}.$$

From this lemma, we get $\sup_{T \ge p+1} \max_{p+1 \le t \le T} \mathbb{E} X_t^2 < +\infty$.
In the sequel, we will use the following terminology.

Definition 1. *We will say that a sequence of functions $f_T : \{1, \dots, T\} \times \mathbb{R}_+^p \to \mathbb{R}$, $T \ge p + 1$, is in the class $\mathcal{L}$ if there exists two positive real numbers M and L, not depending on T, such that*

$$f_T \text{ is bounded by } M, \quad \max_{1 \le t \le T} |f_T(t, x) - f_T(t, y)| \le L\|x - y\|.$$

23

Now, we will consider two particular classes of processes.

Definition 2. *1. A process $(Y_t)_{p+1 \leq t \leq T}$ is said to be of type I if*

$$Y_t = f_T \left(t, X_{t-1}^2, \ldots, X_{t-p}^2 \right), \quad p + 1 \leq t \leq T,$$

where $(f_T)_{T \geq p+1}$ is in the class $\mathcal{L}$.

2. *A process $(Y_t)_{p+1 \leq t \leq T}$ is said to be of type II if*

$$Y_t = f_T \left(t, X_{t-1}^2, \ldots, X_{t-p}^2 \right) + Z_t g_T \left(t, X_{t-1}^2, \ldots, X_{t-p}^2 \right), \quad p + 1 \leq t \leq T,$$

where $(f_T)_{T \geq p+1}$ and $(g_T)_{T \geq p+1}$ are both in the class $\mathcal{L}$ and $g_T \neq 0$.

3. *A process $\bar{S}$ defined by*

$$\bar{S}_t = \sum_{i=p+1}^{T} k_{t,i}(b) \bar{Y}_i, \quad p + 1 \leq t \leq T,$$

with Y of type I or II will be called a smoothing .

Notes

1. An important example of processes of type II is $Y_t = W_t X_{t-j}^2 X_t^2$, for $j \in [\![1, p]\!]$. Here W_t is given by equation (2.3) in the paper. This is due to the decomposition $Y_t = W_t \sigma_t^2 X_{t-j}^2 + Z_t W_t \sigma_t^2 X_{t-j}^2$ and to the particular form of the weights W_t and of σ_t^2.

2. Some smoothings appear in the expression of $\hat{q}_{j,b,t}$ for $j = 1, 2$. Our method for proving Theorem 1 is to make an asymptotic expansion of the estimator $\hat{\beta}$ and to show that the effect of the smoothings incorporated in $\hat{\beta}$ is negligible by computing some moments. The terminology type I or type II is just used for identifying the number of smoothings which impose a moment restriction.

A.1   Covariance inequalities

Here, we assume that the assumptions of Theorem 1 are fulfilled. Sometimes, assumption **A2(h)** can be used for a general value of the integer h, this will be precised in the statements of our Lemma/Propositions.

Lemma 2. *Let s and t be two natural integers such that $T \geq t \geq s+1 \geq p+1$. Now let $\left(\Upsilon_s, \Upsilon_{s-1}, \ldots, \Upsilon_{s-p+1} \right)$ be a random vector independent from the sequence $(\xi_t)_{1 \leq t \leq T}$ and with the same distribution as $\left(X_s, X_{s-1}, \ldots, X_{s-p+1} \right)$. For $s + 1 \leq k \leq t$, we define recursively*

$$\Upsilon_k = \xi_k \sqrt{a_0 \left(\frac{t}{T} \right) + \sum_{j=1}^{p} a_j \left(\frac{t}{T} \right) \Upsilon_{k-j}^2}.$$

Then

$$\mathbb{E}|X_t^2 - \Upsilon_t^2| \leq 2dc^{\frac{t-s-1}{p}},$$

*where $d = \sup_{T \geq p+1} \max_{1 \leq t \leq T} \mathbb{E} X_t^2$ and c is defined in assumption **A1**.*

Proof of Lemma 2

- Assume first that $s + 1 \leq t \leq s + p$. Then

$$
\begin{aligned}
\mathbb{E}|X_t^2 - \Upsilon_t^2| &= \sum_{j=1}^{p} a_j\left(\frac{t}{T}\right)\mathbb{E}|X_{t-j}^2 - \Upsilon_{t-j}^2| \\
&\leq c \max_{1 \leq j \leq p} \mathbb{E}|X_{t-j}^2 - \Upsilon_{t-j}^2| \\
&\leq 2dc.
\end{aligned}
$$

Since $\frac{t-s-1}{p} \leq 1$, the result follows in this case.

- Suppose the inequality true for any $t \in [\![s+p, N]\!]$, where $N \in [\![s+p, T-1]\!]$. Then

$$
\begin{aligned}
\mathbb{E}|X_{N+1}^2 - \Upsilon_{N+1}^2| &\leq c \times \max_{1 \leq j \leq p} \mathbb{E}|X_{N+1-j}^2 - \Upsilon_{N+1-j}^2| \\
&\leq 2d \max_{1 \leq j \leq p} c^{1+\frac{N+1-j-s-1}{p}} \\
&\leq 2dc^{\frac{N-s}{p}}.
\end{aligned}
$$

Then the result of the lemma follows from a finite induction.□

Lemma 3. *Let h, s, t be three integers such that $p \leq s \leq t \leq T$ and $h \geq 1$. Assume that $\mathbb{E}\xi_0^{2h(1+\delta)} < \infty$ with $0 < \delta < 1$. Let U_s be an integrable random variable $\mathcal{F}_s$-measurable and $G : \mathbb{R}^{p+k+1} \to \mathbb{R}$ a bounded and Lipschitzian function. We set*

$$
U_t = Z_{t+\ell_1} \cdots Z_{t+\ell_o} G\left(X_{t-p}^2, X_{t-p+1}^2, \ldots, X_{t+k}^2\right),
$$

with $0 \leq \ell_1, \ldots, \ell_o \leq k$ and $o \leq h$. Assume that $\mathbb{E}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}|^{1+\delta} < \infty$. Then, we have

$$
|Cov\left(U_s, U_t\right)| \leq (C_1 \vee C_2) c^{\frac{\kappa(t-s)}{p}},
$$

where, setting $\kappa = \frac{\delta}{1+\delta}$,

$$
C_1 = M(G)c^{-\kappa}\left(\mathbb{E}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}| + \mathbb{E}|U_s| \cdot \mathbb{E}|Z_{t+\ell_1} \cdots Z_{t+\ell_o}|\right).
$$

and

$$
C_2 = \frac{d^\kappa c^{-\kappa\frac{p+1}{p}}}{1 - c^{\frac{\kappa}{p}}} L(G)^\kappa M(G)^{1-\kappa} \mathbb{E}^{\frac{1}{1+\delta}}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}|^{1+\delta},
$$

where d is given in Lemma 2 and $M(G)$ (resp. $L(G)$) denotes the supremum (resp. the Lipschitz constant) of the function G.

Proof of Lemma 3.

- Assume first that $t - p \leq s$. Then, it is easy to get the bounds

$$
\begin{aligned}
|\text{Cov}\,(U_s, U_t)| &\leq M(G)\left(\mathbb{E}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}| + \mathbb{E}|U_s| \cdot \mathbb{E}|Z_{t+\ell_1} \cdots Z_{t+\ell_o}|\right) \\
&\leq C_1 c^{\kappa \frac{t-s}{p}}.
\end{aligned}
$$

- Now, assume that $t \geq s + 1 + p$. We have the equality $\text{Cov}\,(U_s, U_t) = \mathbb{E}\,(U_s\,(U_t - U_t'))$, where

$$
U_t' = Z_{t+\ell_1} \cdots Z_{t+\ell_o} G\left(\Upsilon_{t-p}^2, \ldots, \Upsilon_{t+k}^2\right)
$$

is a random variable independent from $\mathcal{F}_s$. Here Υ denotes the process introduced in Lemma 2. Using Lemma 2, the following bounds are valid.

$$
\begin{aligned}
&|\text{Cov}\,(U_s, U_t)| \\
\leq\ & \mathbb{E}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o} \cdot |G\left(X_{t-p}^2, \ldots, X_{t+k}^2\right) - G\left(\Upsilon_{t-p}^2, \ldots, \Upsilon_{t+k}^2\right)| \\
\leq\ & (2M(G))^{1-\kappa} \mathbb{E}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}| \cdot |G\left(X_{t-p}^2, \ldots, X_{t+k}^2\right) - G\left(\Upsilon_{t-p}^2, \ldots, \Upsilon_{t+k}^2\right)|^{\kappa} \\
\leq\ & L(G)^{\kappa}\,(2M(G))^{1-\kappa} \sum_{i=-p}^{k} \mathbb{E}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}| \cdot |X_{t+i}^2 - \Upsilon_{t+i}^2|^{\kappa} \\
\leq\ & L(G)^{\kappa}\,(2M(G))^{1-\kappa} \sum_{i=-p}^{k} \mathbb{E}^{\frac{1}{1+\delta}}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}|^{1+\delta} \times \mathbb{E}^{\kappa}|X_{t+i}^2 - \Upsilon_{t+i}^2| \\
\leq\ & (2d)^{\kappa} L(G)^{\kappa}\,(2M(G))^{1-\kappa} \mathbb{E}^{\frac{1}{1+\delta}}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}|^{1+\delta} \times \sum_{i=-p}^{k} c^{\kappa \frac{t+i-s-1}{p}} \\
\leq\ & \frac{d^{\kappa} c^{-\kappa \frac{p+1}{p}}}{1 - c^{\frac{\kappa}{p}}} L(G)^{\kappa} M(G)^{1-\kappa} \mathbb{E}^{\frac{1}{1+\delta}}|U_s Z_{t+\ell_1} \cdots Z_{t+\ell_o}|^{1+\delta} c^{\kappa \frac{t-s}{p}}.
\end{aligned}
$$

Then the result announced in Proposition 3 is a direct consequence of the two previous points. $\square$

The following corollary is a direct consequence of Lemma 3.

Corollary 2. *Let $h \geq 1$ be an integer. Assume that $\mathbb{E}\xi_0^{2h(1+\delta)} < \infty$ with $0 < \delta < 1$. Let s, t, q, o four non-negative integers such that $p \leq s \leq t$ and $q + o \leq h$. Let U_s and U_t two random variables defined by*

$$
U_s = Z_{h_1} \cdots Z_{h_q} H\left(X_1^2, \ldots, X_s^2\right), \quad U_t = Z_{t+\ell_1} \cdots Z_{t+\ell_o} G\left(X_{t-p}^2, \ldots, X_{t+k}^2\right),
$$

where H and G are two elements of $\mathcal{L}$ and $1 \leq h_1 \leq \cdots \leq h_q \leq s$ and $0 \leq \ell_1 \leq \cdots \leq \ell_o \leq k$. Then, we have the bound

$$
\text{Cov}(U_s, U_t) \leq (C_1 \vee C_2)\, c^{\kappa \frac{t-s}{p}},
$$

where $C_1 = M(G)M(H)c^{-\kappa}\left(\mathbb{E}|Z_1|^{q+o} + \mathbb{E}|Z_1|^o \cdot \mathbb{E}|Z_1|^q\right)$ and

$$
C_2 = \frac{d^{\kappa} c^{-\kappa \frac{p+1}{p}}}{1 - c^{\frac{\kappa}{p}}} L(G)^{\kappa} M(G)^{1-\kappa} M(H) \mathbb{E}^{\frac{1}{1+\delta}}\left(|Z_1|^{(q+o)(1+\delta)}\right).
$$

A.2 Moment bounds

The two next results are crucial for proving the asymptotic normality of our estimators. In particular, Proposition 4 gives conditions under which some partial sums involving smoothings are convergent to zero with a faster rate than $\sqrt{T}$.

Lemma 4. *Assume that $\mathbb{E}\xi_0^{2h(1+\delta)} < \infty$ for a positive integer h. Let $Y^{(1)}, \ldots, Y^{(q)}$ be $q \geq 1$ process of type I or II with at most h processes of type II. Then for a family $\{z_{T,i} : p + 1 \leq i \leq T, T \geq p + 1\}$ of deterministic positive weights, we have*

$$\sum_{p+1 \leq i_1, \ldots, i_q \leq T} z_{T,i_1} \cdots z_{T,i_q} |\mathbb{E}\left(\bar{Y}_{i_1}^{(1)} \cdots \bar{Y}_{i_q}^{(q)}\right)| = O\left((\phi_T s_T)^{\frac{q}{2}}\right),$$

where $s_T = \sum_{i=p+1}^{T} z_{T,i}$ and $\phi_T = \max_{p+1 \leq i \leq T} z_{T,i}$.

Proof of Lemma 4 We set $\beta = c^{\frac{\delta}{p(1+\delta)}}$. The result is clear for $q = 1$. Assume that $q \geq 2$. First, we observe that from Corollary 2, we have for $p + 1 \leq t_1 \leq \cdots \leq t_q \leq T$ and $1 \leq j \leq q - 1$,

$$|\text{Cov}\left(\bar{Y}_{t_1}^{(1)} \cdots \bar{Y}_{t_j}^{(j)}, \bar{Y}_{t_{j+1}}^{(j+1)} \cdots \bar{Y}_{t_q}^{(q)}\right)| \leq C\beta^{t_{j+1}-t_j}, \tag{A.1}$$

where $C > 0$ does not depend on T and on $t_1, \ldots, t_q$. Inequality (A.1) follows from the fact that the covariance given in (A.1) can be decomposed as a sum of covariances of the form $\text{Cov}\left(U_{t_j}, U_{t_{j+1}}\right)$ given in Corollary 2 (replacing s and t with t_j and t_{j+1} respectively). Inequality (A.1) is crucial for the sequel.

We set for $1 \leq j \leq q$,

$$A_T^{(j)}\left(\bar{Y}^{(1)}, \ldots, \bar{Y}^{(j)}\right) = \sum_{p+1 \leq i_1 \leq \cdots \leq i_j \leq T} z_{T,i_1} \cdots z_{T,i_j} |\mathbb{E}\left(\bar{Y}_{i_1}^{(1)} \cdots \bar{Y}_{i_j}^{(j)}\right)|.$$

We use a classical method for bounding sums of cross moments using bounds on covariances (see Dedecker et al. (2007) p. 78). For a q−uplet $\mathbf{i} = (i_1, \ldots, i_q) \in [\![p + 1, T]\!]$ such that $i_1 \leq \cdots \leq i_q$, we define

$$s(\mathbf{i}) = \min\left\{j \leq q : i_{j+1} - i_j = \max_{1 \leq \ell \leq q-1}(i_{\ell+1} - i_\ell)\right\}.$$

Then, using the bound (A.1), we have

$$A_T^{(q)}\left(\bar{Y}^{(1)}, \ldots, \bar{Y}^{(q)}\right)$$

$$\leq \sum_{j=1}^{q-1} A_T^{(j)}\left(\bar{Y}^{(1)}, \ldots, \bar{Y}^{(j)}\right) \cdot A_T^{(q-j)}\left(\bar{Y}^{(j+1)}, \ldots, \bar{Y}^{(q)}\right)$$

$$+ C\sum_{j=1}^{q-1}\sum_{r=0}^{T-1}\beta^r \sum_{\mathbf{i}:s(\mathbf{i})=j, i_{j+1}=i_j+r} z_{T,i_1} \cdots z_{T,i_q}.$$

Since,

$$\sum_{\mathbf{i}:s(\mathbf{i})=j, i_{j+1}=i_j+r} z_{T,i_1} \cdots z_{T,i_q} \leq s_T (\phi_T)^{q-1} (r + 1)^q,$$

27

we conclude that

$$A_T^{(q)}\left(\bar{Y}^{(1)},\ldots,\bar{Y}^{(q)}\right) \le \sum_{j=1}^{q-1} A_T^{(j)}\left(\bar{Y}^{(1)},\ldots,\bar{Y}^{(j)}\right) \cdot A_T^{(q-j)}\left(\bar{Y}^{(j+1)},\ldots,\bar{Y}^{(q)}\right) + O\left(s_T(\phi_T)^{q-1}\right).$$

Since $\phi_T \le s_T$, we have $s_T(\phi_T)^{q-1} \le (s_T\phi_T)^{\frac{q}{2}}$. Then using an induction on q, it easy to prove that

$$A_T^{(q)}\left(\bar{Y}^{(1)},\ldots,\bar{Y}^{(q)}\right) = O\left((s_T\phi_T)^{\frac{q}{2}}\right).$$

This proves Lemma 2.□

Proposition 4. *Assume that $\mathbb{E}\xi_0^{4h(1+\delta)} < \infty$ for a positive integer h and that $b\sqrt{T} \to \infty$. Let $Y^{(1)},\ldots,Y^{(q)}$ be $q \ge 2$ processes of type I or II with at most h processes of type II. We denote by $\bar{S}^{(1)},\ldots,\bar{S}^{(q)}$ the corresponding smoothings.*

1. *If $\{\mu_{t,T} : p + 1 \le t \le T, T \ge p + 1\}$ denotes a family of real numbers such that*

$$\sup_{T \ge p+1} \max_{p+1 \le t \le T} |\mu_{t,T}| < \infty,$$

   we have, using the notations of point 2,

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \mu_{t,T}\bar{S}_t^{(1)}\cdots\bar{S}_t^{(q)} = o_{\mathbb{P}}(1).$$

2. *We have also $\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \bar{Y}_t^{(1)}\bar{S}_t^{(2)}\cdots\bar{S}_t^{(q)} = o_{\mathbb{P}}(1)$.*

3. *If Y is a process of type I (resp. II), we have for all positive integer h' (resp. $h = h'$), $\max_{1 \le t \le T} \mathbb{E}\bar{S}_t^{2h'} = O\left((Tb)^{-h'}\right)$.*

4. *Assume that $Y^{(1)}$ is a process of type I. Then $\frac{1}{T} \sum_{t=p+1}^{T} \overline{Y}_t^{(1)}$ converges to 0 a.s.*

Proof of Proposition 4

1. Assume that $\sup_{t,T} |\mu_{t,T}| \le C$. Taking the second order moment, we get

$$\mathbb{E}|\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \mu_{t,T}\bar{S}_t^{(1)}\cdots\bar{S}_t^{(q)}|^2$$

$$\le \frac{C^2}{T} \sum_{s,t=p+1}^{T} \sum_{i_1,M_1,\ldots,i_q,j_q=p+1}^{T} k_{s,i_1}k_{t,M_1}\cdots k_{s,i_q}k_{t,j_q} \cdot$$

$$\mathbb{E}\left(\bar{Y}_{i_1}^{(1)}\bar{Y}_{M_1}^{(1)}\cdots\bar{Y}_{i_q}^{(q)}\bar{Y}_{j_q}^{(q)}\right)|.$$

Using Lemma 2 (replacing h with $2h$) with $s_T = 1$ and $\phi_T = O\left(\frac{1}{Tb}\right)$, the last bound is $O\left(\frac{1}{T^{q-1}b^q}\right)$. The result follows using the bandwidth assumptions.

2. The second order moment writes

$$
\frac{1}{T} \sum_{s,t=p+1}^{T} \sum_{i_1,M_1,\dots,i_{q-1},j_{q-1}=1}^{T} k_{s,i_1} k_{t,j_1} \cdots k_{s,i_{q-1}} k_{t,j_{q-1}} \cdot
$$
$$
\times \mathbb{E}\left(\bar{Y}_s^{(1)} \bar{Y}_t^{(1)} \bar{Y}_{i_1}^{(2)} \bar{Y}_{j_1}^{(2)} \cdots \bar{Y}_{i_{q-1}}^{(q)} \bar{Y}_{j_{q-1}}^{(q)} \right).
$$

Using the bound $\max_{p+1 \le t,i \le T} k_{t,i} = O\left(\frac{1}{Tb}\right)$ and applying Lemma 2 with $z_{T,i} = 1$, it is easy to show that this second order moment is $O\left(\frac{1}{T^{q-1} b^{2(q-1)}}\right)$. This leads to the result. $\square$

3. This is a consequence of Lemma 2, using the inequality

$$
\mathbb{E}\left(\bar{S}_t^{2h'} \right) \le \sum_{p+1 \le i_1,\dots i_{2h'} \le T} k_{t,i_1} \cdots k_{t,i_{2h'}} \left| \mathbb{E}\left(\bar{Y}_{i_1} \cdots \bar{Y}_{i_{2h'}} \right) \right|.
$$

4. We have for $\epsilon > 0$,

$$
\mathbb{P}\left(\frac{1}{T} | \sum_{t=p+1}^{T} \bar{Y}_t^{(1)} | > \epsilon \right)
$$
$$
\le \frac{1}{\epsilon^4 T^4} \mathbb{E} | \sum_{t=p+1}^{T} \bar{Y}_t^{(1)} |^4
$$
$$
\le \frac{1}{\epsilon^4 T^4} \sum_{p+1 \le i_1,\dots i_4 \le T} |\mathbb{E}\left(\bar{Y}_{i_1}^{(1)} \cdots \bar{Y}_{i_4}^{(1)} \right)|
$$

Using Lemma 2, the last bound is $O\left(\frac{1}{T^2}\right)$. Then the result follows from the Borel-Cantelli Lemma. $\square$

A.3   Control of deterministic quantities using local stationarity

Lemma 5.   *1. For $u \in [0,1]$, we set $s_2(u) = \mathbb{E}\left(W_1(u) M_1(u) N_1(u)' \right)$ and $s_3(u) = \mathbb{E}\left(W_1(u) M_1(u) M_1(u)' \right)$. Then we have*

$$
\max_{p+1 \le t \le T} \left\{ \|s_{3,t} - s_3\left(\frac{t}{T}\right)\| + \|s_{2,t} - s_2\left(\frac{t}{T}\right)\| \right\} = O\left(\frac{1}{T}\right), \tag{A.2}
$$

$$
\inf_{u \in [0,1]} \det\left(s_3(u) \right) > 0, \tag{A.3}
$$

$$
\sup_{u \in [0,1]} \|q_2(u)\| < \infty, \tag{A.4}
$$

$$
\max_{p+1 \le t \le T} \|q_{2,t} - q_2\left(\frac{t}{T}\right)\| = O\left(\frac{1}{T}\right). \tag{A.5}
$$

2. *We have*

$$\sup_{T \geq p+1} \max_{p+1 \leq t \leq T} \left\{ \|\mathbb{E}^{-1}(S_{3,t})\| + \|\mathbb{E}(S_{1,t})\| + \|\mathbb{E}(S_{2,t})\| \right\} < \infty.$$

3. *Setting* $\eta_{j,t} = \mathbb{E}^{-1}(S_{3,t})\mathbb{E}\left(S_{j,t}\right) - q_{j,t}$ *for* $j = 1, 2$, *we have*

$$\max_{p+1 \leq t \leq T} \left\{ \|\eta_{1,t}\| + \|\eta_{2,t}\| \right\} = O(b).$$

4. *We have*

$$\max_{p+1 \leq t \leq T} \left\{ \|S_{3,t}\| + \|S_{3,t}^{-1}\| \right\} = O_{\mathbb{P}}(1). \tag{A.6}$$

Proof of Lemma 5

1. We prove the four assertions successively.

   - Since $W_t M_t M_t' = f\left(t/T, X_{t-1}^2, \ldots, X_{t-p}^2\right)$ where $f : [0,1] \times \mathbb{R}_+^p \to \mathcal{M}_{m,m}$ satisfies

   $$\|f(u, x_1, \ldots, x_p) - f(u, y_1, \ldots, y_p)\| \leq C \sum_{i=1}^{p} |x_i - y_i|$$

   for some positive constant C, Lemma 1 given in the paper yields to $\max_{p+1 \leq t \leq T} \|s_{3,t} - s_3\left(\frac{t}{T}\right)\| = O(1/T)$. The conclusion for $s_{2,t}$ follows in the same way. This shows (A.2).

   - Next, we show (A.3). Let $\lambda(u)$ be the smallest eigenvalue of $\mathbb{E}(W_1(u)M_1(u)M_1(u)')$. From Lemma 1, the application $u \mapsto \mathbb{E}(W_1(u)M_1(u)M_1(u)')$ is Lipschitz continuous. Moreover, it is easily shown that for all $u \in [0,1]$, the matrix $\mathbb{E}(W_1(u)M_1(u)M_1(u)')$ is positive definite. This entails that the application $u \mapsto \lambda(u)$ is continuous and positive. This implies (A.3).

   - Since $\sup_{u \in [0,1]} \|W_1(u)M_1(u)N_1(u)'\|$ is bounded, we deduce from (A.3) that $\sup_{u \in [0,1]} \|q_2(u)\| < \infty$.

   - The assertion (A.5) easily follows from (A.2), (A.3) and (A.4).

2. Since

$$C = \sup_{\omega, T, i} W_i(\omega) \|M_i(\omega)\| < \infty,$$

   we have

$$\|\mathbb{E}(S_{1,t})\| \leq \sum_{i=p+1}^{T} k_{t,i} \|\mathbb{E}\left(W_i M_i X_i^2\right)\| \leq C \sup_{T \geq p+1} \max_{1 \leq t \leq T} \mathbb{E}\left(X_t^2\right).$$

   The same kind of inequality holds for $\|\mathbb{E}(S_{2,t})\|$.
   It remains to prove that

$$\sup_{T \geq p+1} \max_{p+1 \leq t \leq T} \|\mathbb{E}^{-1}(S_{3,t})\| < \infty. \tag{A.7}$$

   If $x \in \mathbb{R}^j$, with $\|x\| = 1$, we have using (A.2),

$$x' \mathbb{E}(S_{3,t}) x \geq \inf_{p+1 \leq t \leq T} x' s_{3,t} x \geq \inf_{u \in [0,1]} x' s_3(u) x - \frac{C}{T},$$

30

for a suitable constant $C > 0$. Then, using (A.3), there exists $\lambda > 0$ such that $x'\mathbb{E}(S_{3,t})x \geq \lambda - \frac{C}{T}$. We deduce that if T is large enough, the smallest eigenvalue of $\mathbb{E}(S_{3,t})$ is bounded from below. This means that there exists an integer $T_0 \geq p + 1$ such that

$$\sup_{T \geq T_0} \max_{p+1 \leq t \leq T} \|\mathbb{E}^{-1}(S_{3,t})\| < \infty.$$

But since each of the matrices $\mathbb{E}(S_{3,t})$ is easily shown to be positive definite for $p + 1 \leq t \leq T$ and $T \geq T_0$, (A.7) easily follows.

3. For a Lipschitzian function f defined over $[0, 1]$, the assumptions made on the kernel K implies that

$$\max_{1 \leq t \leq T} |f(t/T) - \sum_{i=1}^{T} k_{t,i} f(i/T)| = O(b).$$

We only prove that $\max_{p+1 \leq t \leq T} \|\eta_{1,t}\| = O(b)$, the proof for $\eta_{2,t}$ is similar. We use the decomposition

$$\eta_{1,t} = s_{3,t}^{-1}\left(s_{3,t} - \mathbb{E}(S_{3,t})\right)\mathbb{E}^{-1}(S_{3,t})s_{1,t} + \mathbb{E}^{-1}(S_{3,t})\left(\mathbb{E}\left(S_{1,t}\right) - s_{1,t}\right). \tag{A.8}$$

From the proof of the two first points of the present Lemma, it is easily seen that

$$\sup_{T \geq p+1} \max_{p+1 \leq t \leq T} \left\{\|\mathbb{E}^{-1}(S_{3,t})\| + \|s_{1,t}\| + \|s_{3,t}^{-1}\|\right\} < \infty. \tag{A.9}$$

Moreover

$$\max_{p+1 \leq t \leq T} \|s_{3,t} - \mathbb{E}\left[W_1(t/T)M_1(t/T)M_1'(t/T)\right]\| = O(1/T).$$

Since $\mathbb{E}\left(W_t M_t X_t^2\right) = \mathbb{E}\left(W_t M_t \sigma_t^2\right)$, the choice of the weights W_t entails also

$$\max_{p+1 \leq t \leq T} \|s_{1,t} - \mathbb{E}\left[W_1(t/T)M_1(t/T)X_1^2(t/T)\right]\| = O(1/T).$$

Now, since the two applications

$$u \mapsto d(u) = \mathbb{E}\left(W_1(u)M_1(u)X_1^2(u)\right) \quad u \mapsto e(u) = \mathbb{E}\left(W_1(u)M_1(u)M_1(u)'\right)$$

are Lipschitz continuous, we get

$$\max_{p+1 \leq t \leq T} \left\{\|\mathbb{E}\left(S_{1,t}\right) - s_{1,t}\| + \|\mathbb{E}(S_{3,t}) - s_{3,t}\|\right\} = O(b). \tag{A.10}$$

Then, the result announced follows easily from (A.8), (A.9) and (A.10).

4. We use the decomposition $S_{3,t} = \bar{S}_{3,t} + \mathbb{E}(S_{3,t})$. From the previous points, we have

$$\max_{p+1 \leq t \leq T} \left\{\|\mathbb{E}(S_{3,t})\| + \|\mathbb{E}^{-1}(S_{3,t})\|\right\} = O(1).$$

Moreover for $\epsilon > 0$, we have using point 3 of Proposition 3,

$$\mathbb{P}\left(\max_{p+1 \leq t \leq T} \|\bar{S}_{3,t}\| > \epsilon\right) \leq \frac{1}{\epsilon^4}\sum_{t=p+1}^{T}\mathbb{E}\|\bar{S}_{3,t}\|^4$$

$$\leq \frac{C}{\epsilon^4 T b^2}.$$

Then we conclude that $\max_{p+1 \leq t \leq T} \|\bar{S}_{3,t}\| = o_{\mathbb{P}}(1)$. Then (A.6) easily follows.□

B Proof of Theorem 1

Setting

$$\hat{\beta} = D_T^{-1} L_T, \quad L_T = \sum_{t=p+1}^{T} W_t \hat{O}_t \hat{V}_t, \quad D_T = \sum_{t=p+1}^{T} W_t \hat{O}_t \hat{O}_t',$$

we have $\hat{\beta} - \beta = D_T^{-1} \sum_{t=p+1}^{T} W_t \hat{O}_t \left(\hat{V}_t - \hat{O}_t' \beta \right)$. Using the two relations $\hat{V}_t = V_t - M_t' (\hat{q}_{1,t} - q_{1,t})$, $\hat{O}_t = O_t - (\hat{q}_{2,t} - q_{2,t})' M_t$, we obtain

$$\widetilde{\beta} - \beta = D_T^{-1} \sum_{t=p+1}^{T} W_t \left(O_t - (\hat{q}_{2,t} - q_{2,t})' M_t \right) \cdot \left((\xi_t^2 - 1)\sigma_t^2 - M_t'(\hat{q}_{1,t} - q_{1,t}) + M_t'(\hat{q}_{2,t} - q_{2,t})\beta \right).$$

We also set $U_t = W_t O_t M_t'$. Observe that $\mathbb{E}U_t = 0$. This yields to the following decomposition.

$$\hat{\beta} - \beta = D_T^{-1} \left(L_{1,T}\beta - L_{2,T}\beta - L_{3,T} + L_{4,T} + L_{5,T} - L_{6,T} \right),$$

where

$$L_{1,T} = \sum_{t=p+1}^{T} U_t (\hat{q}_{2,t} - q_{2,t}), \quad L_{2,T} = \sum_{t=p+1}^{T} W_t (\hat{q}_{2,t} - q_{2,t})' M_t M_t' (\hat{q}_{2,t} - q_{2,t}),$$

$$L_{3,T} = \sum_{t=p+1}^{T} U_t (\hat{q}_{1,t} - q_{1,t}), \quad L_{4,T} = \sum_{t=p+1}^{T} W_t (\hat{q}_{2,t} - q_{2,t})' M_t M_t' (\hat{q}_{1,t} - q_{1,t}),$$

$$L_{5,T} = \sum_{t=p+1}^{T} W_t O_t Z_t \sigma_t^2, \quad L_{6,T} = \sum_{t=p+1}^{T} W_t (\hat{q}_{2,t} - q_{2,t})' M_t Z_t \sigma_t^2.$$

$L_{5,T}$ is the main term in the asymptotic expansion of the numerator L_T. We will use the formula

$$B^{-1}A - b^{-1}a = b^{-1}(A - a) - b^{-1}(B - b)b^{-1}(A - a) - b^{-1}(B - b)b^{-1}a + B^{-1}(B - b)b^{-1}(B - b)b^{-1}A. \quad \text{(B.1)}$$

Now for $j = 1, 2, 3$, we set $\bar{S}_{j,t} = S_{j,t} - \mathbb{E}\left(S_{j,t}\right)$. Using (B.1), we have for $j = 1, 2$,

$$
\begin{aligned}
\hat{q}_{j,t} - q_{j,t} &= \mathbb{E}^{-1}(S_{3,t})\mathbb{E}\left(S_{j,t}\right) - q_{j,t} + S_{3,t}^{-1}\bar{S}_{3,t}\mathbb{E}^{-1}(S_{3,t})\bar{S}_{3,t}\mathbb{E}^{-1}(S_{3,t})S_{j,t} \\
&+ \mathbb{E}^{-1}(S_{3,t})\bar{S}_{j,t} - \mathbb{E}^{-1}(S_{3,t})\bar{S}_{3,t}\mathbb{E}^{-1}(S_{3,t})\bar{S}_{j,t} - \mathbb{E}^{-1}(S_{3,t})\bar{S}_{3,t}\mathbb{E}^{-1}(S_{3,t})\mathbb{E}\left(S_{j,t}\right)
\end{aligned}
$$

(B.2)

To prove Theorem 1, we will prove that

$$\frac{1}{\sqrt{T}}L_{5,T} \xrightarrow{\mathcal{D}} \mathcal{N}_n(0, \Sigma_2), \tag{B.3}$$

$$\frac{1}{\sqrt{T}}L_{j,T} = o_{\mathbb{P}}(1), \quad j \in \{1, 2, 3, 4, 6\}, \tag{B.4}$$

$$\lim_{T \to \infty} \frac{1}{T}D_T = \Sigma_1, \text{ a.s..} \tag{B.5}$$

The proofs of (B.3), (B.4) and (B.5) are established in the following subsections.

Proof of assertion (B.3) To prove (B.3), we use the central limit theorem for triangular arrays of martingale differences (see Pollard (1984) Chapter VIII.1, Theorem 1). Using the Cramer-Wold device, it is enough to prove that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} W_t \sigma_t^2 x' O_t x Z_t \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, x'\Sigma_2 x\right), \tag{B.6}$$

for each vector $x \in \mathbb{R}^n$. We set

$$A_t = W_t \sigma_t^2 x' O_t x Z_t, \quad \widetilde{A}_t = W_t \sigma_t^2 x' \left(N_t - q_2\left(\frac{t}{T}\right)' M_t\right) x Z_t.$$

Then, if $\mathcal{F}_t = \sigma\left(\xi_s : s \leq t\right)$, the two families $\{(A_t, \mathcal{F}_t) : 1 \leq t \leq T\}$ and $\left\{\left(\widetilde{A}_t, \mathcal{F}_t\right) : 1 \leq t \leq T\right\}$ form a martingale difference. Their corresponding partial sums are asymptotically equivalent because the quantity $q_{2,t}$ is simply replaced by $q_2(t/T)$ in the expression of $\widetilde{A}_t$. Indeed, we have

$$|A_t - \widetilde{A}_t| \leq \|x\|^2 \cdot |Z_t| \cdot \|q_{2,t} - q_2(t/T)\| \cdot \|M_t W_t \sigma_t^2\|.$$

Using the fact that $M_t W_t \sigma_t^2$ is bounded uniformly in t and Lemma 5, 1., we deduce that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T}(A_t - \widetilde{A}_t) \to 0$, in probability. As a consequence, it is sufficient to prove (B.6) for $\widetilde{A}_t$ instead of A_t. Moreover,

$$\mathbb{E}\left(\widetilde{A}_t^2 | \mathcal{F}_{t-1}\right) = \mathbb{E}|\xi_0^2 - 1|^2 \cdot x' W_t^2 \sigma_t^4 \left(N_t - q_2\left(\frac{t}{T}\right)' M_t\right)\left(N_t - q_2\left(\frac{t}{T}\right)' M_t\right)' x.$$

The process G defined by $G_t = W_t^2 \sigma_t^4 \left(N_t - q_2\left(\frac{t}{T}\right)' M_t\right)\left(N_t - q_2\left(\frac{t}{T}\right)' M_t\right)'$ is (coordinatewise) a process of type I and Proposition 4 (point 4) leads to $\lim_{T \to \infty} \frac{1}{T} \sum_{t=p+1}^{T} \overline{G}_t = 0$ in probability. Moreover, using Lemma 1, Lemma 5 (A.3) and some Lipschitz properties, one can show that

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=p+1}^{T} \mathbb{E}|\xi_0^2 - 1|^2 \cdot \mathbb{E}(G_t) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=p+1}^{T} \mathbb{E}|\xi_0^2 - 1|^2 \mathbb{E}\left(G_1(t/T)\right) = x'\Sigma_2 x.$$

Then we get $\frac{1}{T} \sum_{t=p+1}^{T} \mathbb{E}\left(\widetilde{A}_t^2 | \mathcal{F}_{t-1}\right) \to x'\Sigma_2 x$, in probability. Next, we check the Lindberg condition. If $\epsilon > 0$, we have

$$\mathbb{E}\left(\widetilde{A}_t^2 \mathbb{1}_{|\widetilde{A}_t| > \epsilon \sqrt{T}} | \mathcal{F}_{t-1}\right) \leq \frac{1}{\epsilon^\delta T^{\frac{\delta}{2}}} \mathbb{E}|\xi_0^2 - 1|^{2+\delta} W_t^{2+\delta} \sigma_t^{4+2\delta} |x'\left(N_t - q_2\left(\frac{t}{T}\right)' M_t\right)x|^{2+\delta}.$$

We easily deduce that $\frac{1}{T} \sum_{t=p+1}^{T} \mathbb{E}\left(\widetilde{A}_t^2 \mathbb{1}_{|\widetilde{A}_t| > \epsilon} | \mathcal{F}_{t-1}\right) \to 0$ in probability. This proves (B.6) and (B.3) follows.

Proof of $\frac{1}{\sqrt{T}} N_{j,T} = o_{\mathbb{P}}(1)$ for $j \in \{1, 3, 6\}$.

We only prove the result for $j = 3$. The two other cases can be treated in the same way. We set $\eta_{1,t} = \mathbb{E}^{-1}(S_{3,t})\mathbb{E}(S_{1,t}) - q_{1,t} \in \mathcal{M}_{j,1}$. The proof of the result follows from the following points.

1. We first prove that $\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} U_t \eta_{1,t} = o_{\mathbb{P}}(1)$. It is enough to prove that for $(w, y) \in [\![1, n]\!] \times [\![1, m]\!]$,

$$z_T = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} U_t(w, y) \eta_{1,t}(y, 1) = o_{\mathbb{P}}(1).$$

But this assertion follows from Lemma 5 (3.) and Lemma 4, since $\mathbb{E}z_T^2$ can be bounded by b^2 (up to a positive constant).

2. Now we prove that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} U_t \mathbb{E}^{-1}(S_{3,t}) \bar{S}_{1,t} = o_{\mathbb{P}}(1). \tag{B.7}$$

In order to prove (B.7), it is enough to prove that for a given vector $(w, y, z) \in [\![1, n]\!] \times [\![1, m]\!] \times [\![1, m]\!]$,

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} U_t(w, y) \mathbb{E}^{-1}(S_{3,t})(y, z) \bar{S}_{1,t}(z, 1) = o_{\mathbb{P}}(1).$$

But the result is a direct consequence of Lemma 5 (point 2) and Proposition 4 (point 2 applied with $h = 1$).

3. Using the same arguments as for point 2, one can easily show that

$$\sum_{t=p+1}^{T} U_t \mathbb{E}^{-1}(S_{3,t}) \bar{S}_{3,t} \mathbb{E}^{-1}(S_{3,t}) \bar{S}_{1,t} = o_{\mathbb{P}}\left(\sqrt{T}\right), \quad \sum_{t=p+1}^{T} U_t \mathbb{E}^{-1}(S_{3,t}) \bar{S}_{3,t} \mathbb{E}^{-1}(S_{3,t}) \mathbb{E}(S_{1,t}) = o_{\mathbb{P}}\left(\sqrt{T}\right).$$

4. Finally, we prove that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} U_t S_{3,t}^{-1} \bar{S}_{3,t} \mathbb{E}^{-1}(S_{3,t}) \bar{S}_{3,t} \mathbb{E}^{-1}(S_{3,t}) S_{1,t} = o_{\mathbb{P}}(1). \tag{B.8}$$

Since U_t is uniformly bounded in ω, t, T, there exists $C > 0$ such that

$$\|\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} U_t S_{3,t}^{-1} \bar{S}_{3,t} \mathbb{E}^{-1}(S_{3,t}) \bar{S}_{3,t} \mathbb{E}^{-1}(S_{3,t}) S_{1,t}\|$$

$$\leq \frac{C}{\sqrt{T}} \max_{p+1 \leq i \leq T} \|S_{3,i}^{-1}\| \cdot \sum_{t=p+1}^{T} \|\bar{S}_{3,t}\|^2 \cdot \|S_{1,t}\|$$

$$\leq C \max_{p+1 \leq i \leq T} \|S_{3,i}^{-1}\| \cdot \sqrt{\frac{1}{T} \sum_{t=p+1}^{T} \|\bar{S}_{3,t}\|^4} \cdot \sqrt{\sum_{t=p+1}^{T} \|S_{1,t}\|^2}.$$

From Proposition 4, we have the bounds

$$\max_{p+1 \leq t \leq T} \mathbb{E}\|\bar{S}_{3,t}\|^4 = O\left(\frac{1}{T^2 b^2}\right), \quad \max_{p+1 \leq t \leq T} \mathbb{E}\|\bar{S}_{1,t}\|^2 = O\left(\frac{1}{Tb}\right).$$

Then using the point 2 of Lemma 5, we conclude that $\frac{1}{T} \sum_{t=p+1}^{T} \|S_{1,t}\|^2 = O_{\mathbb{P}}(1)$. Finally, using Lemma 5 (4.), we conclude that

$$\|\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} U_t S_{3,t}^{-1} \bar{S}_{3,t} \mathbb{E}^{-1}(S_{3,t}) \bar{S}_{3,t} \mathbb{E}^{-1}(S_{3,t}) S_{1,t}\| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}b}\right).$$

Hence, (B.8) follows using the assumption $b\sqrt{T} \to \infty$.

34

Proof of $\frac{1}{\sqrt{T}} N_{j,T} = o_{\mathbb{P}}(1)$ **for** $j \in \{2, 4\}$.

We only prove the result for $j = 4$, the proof for the case $j = 2$ is similar. Here, we only use the basic decompositions

$$\hat{q}_{k,t} - q_{k,t} = \eta_{k,t} + \mathbb{E}^{-1}(S_{3,t})\bar{S}_{k,t} - S_{3,t}^{-1}\bar{S}_{3,t}\mathbb{E}^{-1}(S_{3,t})S_{k,t}, \tag{B.9}$$

for $k = 1, 2$ and with

$$\eta_{k,t} = \mathbb{E}^{-1}(S_{3,t})\mathbb{E}(S_{k,t}) - q_{k,t}.$$

Then we have

$$
\begin{aligned}
\|N_{4,T}\| &\leq \sum_{t=p+1}^{T} \|\hat{q}_{2,t} - q_{2,t}\| \cdot \|W_t M_t J_t'\| \cdot \|\hat{q}_{1,t} - q_{1,t}\| \\
&\leq \frac{C_1}{2} \sum_{t=p+1}^{T} \left\{ \|\hat{q}_{2,t} - q_{2,t}\|^2 + \|\hat{q}_{1,t} - q_{1,t}\|^2 \right\},
\end{aligned}
$$

where $C_1 = \sup_{\omega,T,t} \|W_t M_t M_t'\|$. It remains to prove that for $k = 1, 2$,

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \|\hat{q}_{k,t} - q_{k,t}\|^2 = o_{\mathbb{P}}(1). \tag{B.10}$$

We only prove (B.10) for $k = 1$, the proof for $k = 2$ being the same. The proof easily follows from the three following points.

1. From Lemma 5 (3.) and our bandwidth condition, we have $\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \frac{\|\eta_{1,t}\|^2}{\sqrt{T}} = o(1)$.

2. Since

$$C_2 = \sup_{\substack{T \geq p+1 \\ p+1 \leq t \leq T}} \|\mathbb{E}^{-1}(S_{3,t})\|$$

   is finite using Lemma 5 (point 2), we use the inequality

   $$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \|\mathbb{E}^{-1}(S_{3,t})\bar{S}_{1,t}\|^2 \leq \frac{C_2^2}{\sqrt{T}} \sum_{t=p+1}^{T} \|\bar{S}_{1,t}\|^2.$$

   But we know from Proposition 4 that $\max_{p+1 \leq t \leq T} \mathbb{E}\left(\|\bar{S}_{1,t}\|^2\right) = O\left(\frac{1}{Tb}\right)$. Then condition $b\sqrt{T} \to \infty$ entails that

   $$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \mathbb{E}\left(\|\bar{S}_{1,t}\|^2\right) = o(1).$$

   This shows that $\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \|\mathbb{E}^{-1}(S_{3,t})\bar{S}_{1,t}\|^2 = o_{\mathbb{P}}(1)$.

3. Finally, we show that

   $$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \|S_{3,t}^{-1}\bar{S}_{3,t}m_t^{-1}S_{1,t}\|^2 = o_{\mathbb{P}}(1). \tag{B.11}$$

35

We have $\|S_{3,t}^{-1}\bar{S}_{3,t}\mathbb{E}^{-1}(S_{3,t})S_{1,t}\|^2 \leq C_2^2 \max_{p+1\leq i\leq T} \|S_{3,i}^{-1}\|^2 \|\bar{S}_{3,t}\|^2 \|S_{1,t}\|^2$, where $C_2 > 0$ is defined in the previous point. We have $\max_{p+1\leq i\leq T} \|S_{3,i}^{-1}\|^2 = O_{\mathbb{P}}(1)$ (see Lemma 5, point 4) and

$$\mathbb{E}\left(\|\bar{S}_{3,t}\|^2 \cdot \|S_{1,t}\|^2\right) \leq \mathbb{E}^{\frac{\delta}{1+\delta}}\left(\|\bar{S}_{3,t}\|^{2\frac{1+\delta}{\delta}}\right) \cdot \mathbb{E}^{\frac{1}{1+\delta}}\left(\|S_{1,t}\|^{2(1+\delta)}\right).$$

Without loss of generality, one can assume that $\frac{1+\delta}{\delta}$ is an integer. Using Proposition 4 (point 3), we have $\max_{p+1\leq t\leq T} \mathbb{E}\left(\|\bar{S}_{3,t}\|^{2\frac{1+\delta}{\delta}}\right) = O\left((Tb)^{-\frac{1+\delta}{\delta}}\right)$. Moreover, using convexity, the moment assumption on the noise and the fact that $W_i M_i \sigma_i^2$ is bounded,

$$\max_{p+1\leq t\leq T} \mathbb{E}\|S_{1,t}\|^{2(1+\delta)} \leq \max_{p+1\leq i\leq T} \mathbb{E}\|W_i M_i X_i^2\|^{2(1+\delta)} = O(1).$$

Then assertion (B.11) easily follows from the condition $\sqrt{T}b \to \infty$.

B.1 Proof of assertion B.5

Recalling that $\hat{O}_t = O_t - (\hat{q}_{2,t} - q_{2,t})' M_t$, we have

$$
\begin{aligned}
D_T &= \frac{1}{T}\sum_{t=p+1}^{T} W_t O_t O_t' + \frac{1}{T}\sum_{t=p+1}^{T} W_t (\hat{q}_{2,t} - q_{2,t})' M_t M_t' (\hat{q}_{2,t} - q_{2,t}) \\
&\quad - \frac{1}{T}\sum_{t=p+1}^{T} W_t (\hat{q}_{2,t} - q_{2,t})' M_t O_t' - \frac{1}{T}\sum_{t=p+1}^{T} W_t O_t M_t' (\hat{q}_{2,t} - q_{2,t}).
\end{aligned}
$$

We have already shown that the three last terms in the previous decomposition are $o_{\mathbb{P}}(1)$. Then, it remains to show that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=p+1}^{T} W_t O_t O_t' = \Sigma_1, \tag{B.12}$$

in probability. One can obtain (B.12) using the same arguments as for deriving the limit of

$$\frac{1}{T}\sum_{t=p+1}^{T} \mathbb{E}\left(\widetilde{A_t^2}|\mathcal{F}_{t-1}\right)$$

in the proof of assertion (B.3), . $\square$

C Auxiliary results for the proof of Theorem 2

Corollary 3. *Let $(Y_t)_{1\leq t\leq T}$ be a process of type II. We denote by $(\bar{S}_t)_{1\leq t\leq T}$ the corresponding smoothing. Then, under the assumptions of Theorem 2, we have*

$$\max_{1\leq t\leq T} |\bar{S}_t| = o_{\mathbb{P}}\left(T^{-\frac{1}{4}}\right).$$

Consequently, if for $i \in \{1,2\}$, $\bar{S}^{(i)}$ is a smoothing then

$$\max_{1\leq t_1,t_2\leq T} |\bar{S}_{t_1}^{(1)} \bar{S}_{t_2}^{(2)}| = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).$$

Proof of Corollary 3 We set $j = 2\left(\left[\frac{1}{\tau}\right] + 1\right)$. Then using the point 4 of Proposition 3, we have

$$\mathbb{P}\left(\max_{p+1 \leq t \leq T} |\bar{S}_t| > \epsilon T^{-\frac{1}{4}}\right)$$

$$\leq \quad \frac{CT^{\frac{j}{4}}}{\epsilon^j} \sum_{t=p+1}^{T} \mathbb{E}|\bar{S}_t|^j$$

$$\leq \quad C\frac{T^{\frac{j}{4}}}{(Tb)^{\frac{j}{2}}}.$$

where $C > 0$ denotes a generic constant and $\epsilon > 0$ is arbitrary. Since, $\frac{T^{\frac{j}{4}+1}}{(Tb)^{\frac{j}{2}}} \leq \frac{1}{\left(T^{\frac{1}{2}-\tau}b\right)^{\frac{j}{2}}}$, the result follows

from the assumptions made on b.□

Finally, we state a Lemma which will be useful for the proof of Theorem 2. For a matrix-valued process $(H_t)_t$, measurable with the sigma field $\mathcal{F}_T$, we will use the equality $H_t = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$ when $\max_{p+1 \leq t \leq T} \|H_t\| = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$. We also introduce additional notations. For $j = 1, 2, 3$, we define $S^*_{j,t}$ as $S_{j,t}$, replacing W_t with $\frac{1}{\sigma_t^4}$.

Lemma 6. *Assume that the assumptions of Theorem 2 hold.*

1. For $j = 1, 2$, we have

$$\hat{q}_{j,t} - q_{j,t} = \eta_{j,t} + \mathbb{E}^{-1}(S_{3,t})\bar{S}_{j,t} - \mathbb{E}^{-1}(S_{3,t})\bar{S}_{3,t}\mathbb{E}^{-1}(S_{3,t})\mathbb{E}(S_{j,t}) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right),$$

and

$$\|\hat{q}_{j,t} - q_{j,t}\|^2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).$$

2. We have

$$\hat{W}^*_t = \frac{1}{\sigma_t^4}\left\{1 - \frac{2}{\sigma_t^2}\left(M'_t L_t + O'_t\left(\hat{\beta} - \beta\right)\right) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)\right\},$$

with

$$L_t = \hat{q}_{1,t} - q_{1,t} - (\hat{q}_{2,t} - q_{2,t})\beta.$$

In particular, we have $\hat{W}^*_t = \frac{1}{\sigma_t^4}(1 + E_t)$ with $E_t^2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$.

3. We have $\max_{p+1 \leq t \leq T} \|\hat{S}^{-1}_{3,t}\| = O_{\mathbb{P}}(1)$.

4. For $j = 1, 2, 3$, there exists two matrices $a_{j,t}$ and $c_{j,t}$ such that

$$\hat{s}^*_{j,t} - \mathbb{E}\left(S^*_{j,t}\right) = \overline{S^*}_{j,t} + R_{j,t} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right), \tag{C.1}$$

where

$$R_{j,t} = a_{j,t}\left(\hat{\beta} - \beta\right) + \sum_{i=p+1}^{T} k_{t,i}c_{j,i}L_i = o_{\mathbb{P}}\left(T^{-1/4}\right).$$

Moreover, $\|\hat{s}^*_{j,t} - \mathbb{E}\left(S^*_{j,t}\right)\|^2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$.

5. *For $j = 1, 2$, we have $\hat{q}^*_{j,t} - q^*_{j,t} = F^*_{j,t} + \Delta_{j,t}$, where*

$$F^*_{j,t} = \eta^*_{j,t} + \mathbb{E}^{-1}(S^*_{3,t})\overline{S}^*_{j,t} - \mathbb{E}^{-1}(S^*_{3,t})\overline{S}^*_{3,t}\mathbb{E}^{-1}(S^*_{3,t})\mathbb{E}\left(S^*_{j,t}\right),$$

$$\Delta_{j,t} = \mathbb{E}^{-1}(S^*_{3,t})R_{j,t} - \mathbb{E}^{-1}(S^*_{3,t})R_{3,t}\mathbb{E}^{-1}(S^*_{3,t})\mathbb{E}\left(S^*_{j,t}\right) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).$$

Moreover, $\|\hat{q}^*_{j,t} - q^*_{j,t}\|^2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).$

Proof of Lemma 6

1. Under the assumptions of Theorem 2, we recall that $\bar{S}_t = o_{\mathbb{P}}\left(T^{-1/4}\right)$ for all univariate smoothing (see Corollary 3). Then, applying this property coordinatewise, we have $\|\overline{S}_{j,t}\| = o_{\mathbb{P}}\left(T^{-1/4}\right)$ for $j = 1, 2, 3$. The announced decomposition is then a consequence of decomposition (6.4) given in the paper and of Lemma 5, point 4. Next, the second assertion follows from the condition $b^2\sqrt{T} \to 0$ and Lemma 5 (point 3).

2. We use the decomposition

$$\hat{W}^*_t = \frac{1}{\sigma^4_t}\left\{1 + \frac{\sigma^4_t - \hat{\sigma}^4_t - \nu_T}{\sigma^4_t} + \frac{\left(\sigma^4_t - \hat{\sigma}^4_t - \nu_T\right)^2}{\sigma^4_t\left(\hat{\sigma}^4_t + \nu_T\right)}\right\}. \tag{C.2}$$

Now, using the fact that $L^2_t + \|\hat{\beta} - \beta\|^2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$, we get

$$\hat{\sigma}^4_t = \sigma^4_t\left(1 + \frac{2}{\sigma^2_t}\left(O'_t\left(\hat{\beta} - \beta\right) + M'_t L_t\right) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)\right).$$

From this decomposition, we deduce that $\frac{1}{\hat{\sigma}^4_t + \nu_T} = O_{\mathbb{P}}(1)$. Then we get

$$\hat{W}^*_t = \frac{1}{\sigma^4_t}\left\{1 - \frac{2}{\sigma^2_t}\left(O'_t\left(\hat{\beta} - \beta\right) + M'_t L_t\right) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)\right\},$$

which also yields to the approximation $\hat{W}^*_t = \frac{1}{\sigma^4_t}\left(1 + E_t\right)$ with $E^2_t = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$.

3. We have

$$\begin{aligned}\hat{s}^*_{3,t} &= S^*_{3,t} + \sum_{i=p+1}^{T} k_{t,i}\frac{M_i M'_i}{\sigma^4_i}E_i \\ &= S^*_{3,t} + o_{\mathbb{P}}(1).\end{aligned}$$

Indeed, $\max_{p+1\leq t\leq T}\frac{\|M_i M'_i\|}{\sigma^4_i}$ is bounded uniformly in (ω, T) and from point 2, $\max_{p+1\leq t\leq T}\|E_t\| = o_{\mathbb{P}}(1)$. Then the result follows from the fact that $\max_{p+1\leq t\leq T}\|(S^*_{3,t})^{-1}\| = O_{\mathbb{P}}(1)$.

4. We only consider the case $j = 1$, the cases $j = 2$ or $j = 3$ being similar. Using the point 2, we have the decomposition

$$\hat{s}_{1,t}^* = S_{1,t}^* - 2 \sum_{i=p+1}^{T} k_{t,i} \frac{O_i' \left(\hat{\beta} - \beta \right) + M_i' L_i}{\sigma_i^6} M_i X_i^2 + o_{\mathbb{P}} \left(\frac{1}{\sqrt{T}} \right).$$

Now we set $a_{1,t} = -2 \sum_{i=p+1}^{T} k_{t,i} \mathbb{E} \left(\frac{M_i O_i' X_i^2}{\sigma_i^6} \right)$ and $c_{1,t} = -2 \sum_{i=p+1}^{T} k_{t,i} \mathbb{E} \left(\frac{M_i M_i' X_i^2}{\sigma_i^6} \right)$. Moreover, we have from Lemma 3

$$-2 \sum_{i=p+1}^{T} k_{t,i} \frac{M_i N_i' X_i^2}{\sigma_i^6} - a_{1,t} = o_{\mathbb{P}} \left(T^{-\frac{1}{4}} \right),$$

which leads to

$$\left(-2 \sum_{i=p+1}^{T} k_{t,i} \frac{M_i N_i' X_i^2}{\sigma_i^6} - a_{1,t} \right) \cdot \left(\hat{\beta} - \beta \right) = o_{\mathbb{P}} \left(\frac{1}{\sqrt{T}} \right).$$

Then it remains to show that

$$\sum_{i=p+1}^{T} k_{t,i} \left(\frac{M_i M_i' X_i^2}{\sigma_i^6} - c_{1,i} \right) L_i = o_{\mathbb{P}} \left(\frac{1}{\sqrt{T}} \right). \tag{C.3}$$

Considering the decomposition given in point 1, assertion (C.3) will follow if we show the two following assertions. For all real-valued sequences (c_t) such that $\max_{p+1 \leq t \leq T} c_t = O(b)$ and all real-valued processes (Y_t), (G_t) of type I or II,

$$\sum_{i=p+1}^{T} k_{t,i} c_i \bar{Y}_i = o_{\mathbb{P}} \left(\frac{1}{\sqrt{T}} \right), \tag{C.4}$$

and

$$\sum_{i=p+1}^{T} k_{t,i} \bar{Y}_i \bar{S}_i = o_{\mathbb{P}} \left(\frac{1}{\sqrt{T}} \right), \tag{C.5}$$

where

$$\bar{S}_t = \sum_{i=p+1}^{T} k_{t,i} \left(G_i - \mathbb{E}(G_i) \right).$$

The assertion (C.5) can be proved as follows. For $\epsilon > 0$ and an even integer $h > 0$, we have

$$\mathbb{P} \left(\max_{p+1 \leq t \leq T} | \sum_{i=p+1}^{T} k_{t,i} \bar{Y}_i \bar{S}_i | > \epsilon \frac{1}{\sqrt{T}} \right) \leq \frac{T^{\frac{h}{2}}}{\epsilon^h} \sum_{t=p+1}^{T} \mathbb{E} | \sum_{i=p+1}^{T} k_{t,i} \bar{Y}_i \bar{S}_i |^h$$

$$\leq \frac{T^{\frac{h}{2}+1}}{\epsilon^h} \sum_{p+1 \leq i_1, j_1, \ldots, i_h, j_h \leq T} p_{i_1} p_{j_1} \cdots p_{i_h} p_{j_h} | \mathbb{E} \left(\bar{Y}_{i_1} \bar{G}_{j_1} \cdots \bar{Y}_{i_h} \bar{G}_{j_h} \right) |,$$

39

where $p_i = \max_{p+1 \le t \le T} k_{t,i}$. Using Lemma 4, we deduce that the right hand side of the previous inequality is $O\left(\frac{1}{T^{\frac{h}{2}-1}b^h}\right)$ But $\frac{h}{2} - 1 = \frac{h}{2}\left(1 - \frac{2}{h}\right) \ge \frac{h}{2}(1 - \tau)$ when $h \ge \frac{2}{\tau}$. Using the bandwidth conditions, we get (C.5). Next, using Corollary 3 and the bandwidth conditions, we get

$$\max_{p+1 \le t \le T} |\sum_{i=p+1}^{T} k_{t,i} c_i \bar{S}_i| \le \max_{p+1 \le t \le T} |c_t| \times \max_{p+1 \le t \le T} |\bar{S}_t| = o_{\mathbb{P}}\left(bT^{-\frac{1}{4}}\right) = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).$$

This proves (C.4). Finally, since $\|L_t\|^2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$ and $\hat{\beta} - \beta = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$, the decomposition (C.1) holds true. Using some arguments discussed before, we easily get $\|\hat{s}_{1,t}^* - \mathbb{E}\left(S_{1,t}^*\right)\|^2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$.

5. Using the equality

$$
\begin{aligned}
\hat{q}_{j,t}^* - q_{j,t}^* &= \eta_{j,t}^* + \mathbb{E}^{-1}(S_{3,t}^*)\left(\hat{s}_{j,t}^* - \mathbb{E}\left(S_{j,t}^*\right)\right) \\
&- \mathbb{E}^{-1}(S_{3,t}^*)\left(\hat{s}_{j,t}^* - \mathbb{E}\left(S_{j,t}^*\right)\right)\mathbb{E}^{-1}(S_{3,t}^*)\mathbb{E}\left(S_{j,t}^*\right) \\
&+ \left(\hat{s}_{3,t}^*\right)^{-1}\left(\hat{s}_{3,t}^* - \mathbb{E}\left(S_{j,t}^*\right)\right)\mathbb{E}^{-1}(S_{3,t}^*)\left(\hat{s}_{3,t}^* - \mathbb{E}\left(S_{3,t}^*\right)\right)\mathbb{E}^{-1}(S_{3,t}^*)\hat{s}_{j,t}^*,
\end{aligned}
$$

the result of point 5 is an easy consequence of the previous points.□

D   Proof of Theorem 2

We recall that for a triangular array $\{H_t = H_{t,T}\}$ of matrices, we will denote $H_t = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$ when

$$\max_{p+1 \le t \le T} \|H_t\| = o_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).$$

Notations.   Let us also recall the following notations.

$$S_{1,t}^* = \sum_{t=p+1}^{T} k_{t,i}\frac{M_i X_i^2}{\sigma_i^4}, \quad S_{2,t}^* = \sum_{t=p+1}^{T} k_{t,i}\frac{M_i N_i'}{\sigma_i^4},$$

$$S_{3,t}^* = \sum_{t=p+1}^{T} k_{t,i}\frac{M_i M_i'}{\sigma_i^4},$$

$$\eta_{1,t}^* = \mathbb{E}^{-1}(S_{3,t}^*)\mathbb{E}\left(S_{1,t}^*\right) - q_{1,t}^*,$$

$$\eta_{2,t}^* = \mathbb{E}^{-1}(S_{3,t}^*)\mathbb{E}\left(S_{2,t}^*\right) - q_{2,t}^*,$$

$$q_{1,t}^* = \mathbb{E}^{-1}\left(\frac{M_t M_t'}{\sigma_t^4}\right)\mathbb{E}\left(\frac{M_t X_t^2}{\sigma_t^4}\right), \quad q_{2,t}^* = \mathbb{E}^{-1}\left(\frac{M_t M_t'}{\sigma_t^4}\right)\mathbb{E}\left(\frac{M_t N_t'}{\sigma_t^4}\right),$$

$$O_t^* = N_t - \left(q_{2,t}^*\right)' M_t.$$

The proof of Theorem 2 uses the same decomposition as for Theorem 1. More precisely, we have

$$\hat{\beta}_* - \beta = \hat{D}_T^{-1}\left(\hat{L}_{1,T}\beta - \hat{L}_{2,T}\beta - \hat{L}_{3,T} + \hat{L}_{4,T} + \hat{L}_{5,T} - \hat{L}_{6,T}\right),$$

where

$$\hat{L}_{1,T} = \sum_{t=p+1}^{T} \hat{W}_t^* O_t^* M_t' \left(\hat{q}_{2,t}^* - q_{2,t}^* \right),$$

$$\hat{L}_{2,T} = \sum_{t=p+1}^{T} \hat{W}_t^* \left(\hat{q}_{2,t}^* - q_{2,t}^* \right)' M_t M_t' \left(\hat{q}_{2,t}^* - q_{2,t}^* \right),$$

$$\hat{L}_{3,T} = \sum_{t=p+1}^{T} \hat{W}_t^* O_t^* M_t' \left(\hat{q}_{1,t}^* - q_{1,t}^* \right),$$

$$\hat{L}_{4,T} = \sum_{t=p+1}^{T} \hat{W}_t^* \left(\hat{q}_{2,t}^* - q_{2,t}^* \right)' M_t M_t' \left(\hat{q}_{1,t}^* - q_{1,t}^* \right),$$

$$\hat{L}_{5,T} = \sum_{t=p+1}^{T} \hat{W}_t^* O_t^* Z_t \sigma_t^2,$$

$$\hat{L}_{6,T} = \sum_{t=p+1}^{T} \hat{W}_t^* \left(\hat{q}_{2,t}^* - q_{2,t}^* \right)' M_t Z_t \sigma_t^2.$$

$$\hat{D}_T = \sum_{t=p+1}^{T} \hat{W}_t^* O_t^* (O_t^*)' - \hat{L}_{1,T} - \hat{L}_{1,T}' + \hat{L}_{2,T}.$$

Using Lemma 6, it is easy to get

$$\|\hat{L}_{2,T}\| + \|\hat{L}_{4,T}\| = o_{\mathbb{P}}\left(\sqrt{T} \right).$$

Then the proof of Theorem 2 will follow from the three following points.

1. We first show that $L_{j,T} = o_{\mathbb{P}}\left(\sqrt{T} \right)$ for $j = 1, 3, 6$. We only consider the case $j = 1$, the proofs for the cases $j = 3$, or $j = 6$ being similar. Using the notations introduced in Lemma 6 (points 2 and 5), we have

$$\hat{L}_{1,T} = \sum_{t=p+1}^{T} (1 + E_t) \cdot \frac{O_t^* M_t'}{\sigma_t^4} \cdot \left(F_{2,t}^* + \Delta_{2,t} \right).$$

The proof will be a consequence of the following points.

- Since $\frac{O_t^* M_t'}{\sigma_t^4}$ is centered, we get $\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \frac{O_t^* M_t'}{\sigma_t^4} F_{2,t}^* = o_{\mathbb{P}}(1)$ The proof is similar to the proof given in Theorem 1 (see the proof of $\frac{1}{\sqrt{T}} N_{j,T} = o_{\mathbb{P}}(1)$ for $j = 1, 3, 6$).

- Next, we prove that $\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \frac{O_t^* M_t'}{\sigma_t^4} \Delta_{2,t} = o_{\mathbb{P}}(1)$. First we will show that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \frac{O_t^* M_t'}{\sigma_t^4} \mathbb{E}^{-1}(S_{3,t}^*) R_{2,t} = o_{\mathbb{P}}(1).$$

Using Lemma 6, we have the expression

$$R_{2,t} = a_{2,t} \left(\hat{\beta} - \beta \right) + \sum_{i=p+1}^{T} k_{t,i} c_{2,i} L_i.$$

41

Since $\frac{O_t^* M_t'}{\sigma_t^4}$ is centered, we have

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \frac{W_t^* M_t'}{\sigma_t^4} \mathbb{E}^{-1}(S_{3,t}^*) a_{2,t} \left(\hat{\beta} - \beta\right) = o_{\mathbb{P}}(1).$$

Then, it remains to show that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \frac{O_t^* M_t'}{\sigma_t^4} \mathbb{E}^{-1}(S_{3,t}^*) \sum_{i=p+1}^{T} k_{t,i} c_{2,i} L_i = o_{\mathbb{P}}(1). \tag{D.1}$$

The assertion (D.1) will follow if we work with each entry of different matrices and if we prove the following property. If (Y_t) is a real-valued centered process of type I or II, $\left(\bar{S}_t\right)$ is a real-valued smoothing of type I or II and (c_t), $(\widetilde{c}_t)$ are deterministic sequences satisfying $\max_{p+1 \leq t \leq T} |c_t| = O(b)$ and $\max_{p+1 \leq t \leq T} |\widetilde{c}_t| = O(1)$, then

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} Y_t c_t = o_{\mathbb{P}}(1), \quad \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} Y_t \sum_{i=p+1}^{T} k_{t,i} \widetilde{c}_i \bar{S}_i = o_{\mathbb{P}}(1). \tag{D.2}$$

The first assertion in (D.1) has been already proved using the bound for the covariance function of the process (Y_t) (see the proof of $\frac{1}{\sqrt{T}} L_{3,T} = o_{\mathbb{P}}(1)$ in the proof of Theorem 1). Then it remains to prove the second assertion in (D.2). Writing $\bar{S}_i = \sum_{j=p+1}^{T} k_{i,j} \overline{Y'}_j$, we have

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} Y_t \sum_{i=p+1}^{T} k_{t,i} \widetilde{c}_i \bar{S}_i = \frac{1}{\sqrt{T}} \sum_{t,j=p+1}^{T} p_{t,j} \bar{Y}_t \overline{Y'}_i,$$

with $p_{t,j} = \sum_{i=p+1}^{T} \widetilde{c}_i k_{t,i} k_{i,j}$ satisfies $\max_{p+1 \leq j, t \leq T} |p_{t,j}| = O\left(\frac{1}{Tb}\right)$. Moreover, using Lemma 2, we have for a generic constant $C > 0$,

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{\sqrt{T}} \sum_{t,j=p+1}^{T} p_{t,j} \bar{Y}_t \overline{Y'}_i\right|^2 &= \frac{1}{T} \sum_{t_1,t_2,j_1,j_2=p+1}^{T} p_{t_1,j_1} p_{t_2,j_2} |\mathbb{E}\left(\bar{Y}_{t_1} \overline{Y'}_{t_2} \bar{Y}_{j_1} \overline{Y'}_{j_2}\right)| \\
&\leq \frac{C}{T^3 b^2} \sum_{t_1,t_2,j_1,j_2=p+1}^{T} |\mathbb{E}\left(\bar{Y}_{t_1} \overline{Y'}_{t_2} \bar{Y}_{j_1} \overline{Y'}_{j_2}\right)| \\
&\leq \frac{C}{T b^2}.
\end{aligned}$$

Using the assumption $Tb^2 \to \infty$, we deduce the second assertion in (D.2). This proves (D.1).

- Finally, there exists a constant $C > 0$ such that

$$\left\|\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} E_t \cdot \frac{O_t^* M_t'}{\sigma_t^4} \cdot \left(\hat{q}_{2,t}^* - q_{2,t}^*\right)\right\| \leq \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \frac{\|O_t^* M_t'\|}{\sigma_t^4} \cdot \max_{p+1 \leq t \leq T} |E_t| \times \max_{p+1 \leq t \leq T} \|\hat{q}_{2,t}^* - q_{2,t}^*\|.$$

Then using Lemma 6, we deduce that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} E_t \cdot \frac{O_t^* M_t'}{\sigma_t^4} \left(\hat{q}_{2,t}^* - q_{2,t}^*\right) = o_{\mathbb{P}}(1).$$

42

2. Next, we prove the following convergence in distribution: $\frac{1}{\sqrt{T}}\hat{L}_{5,T} \to \mathcal{N}_n\left(0, \text{Var}\left(\xi_1^2\right)\Sigma\right)$. We have

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\hat{L}_{5,T} &= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\frac{O_t^*}{\sigma_t^2}Z_t + \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\frac{E_t O_t^*}{\sigma_t^2}Z_t \\
&= A_T + B_T.
\end{aligned}
$$

The convergence of (A_T) is obtained as in the proof of Theorem 1. Moreover, using the expression of E_t given in Lemma 6 and arguments which are now familiar, the convergence $B_T = o_{\mathbb{P}}(1)$ can be obtained.

3. Finally we show that $\frac{\hat{D}_T}{T} \to \Sigma$ a.s. It just remains to prove that

$$
\frac{1}{T}\sum_{t=p+1}^{T}\frac{E_t}{\sigma_t^4}O_t^*\left(O_t^*\right)' = o_{\mathbb{P}}(1), \quad \frac{1}{T}\sum_{t=p+1}^{T}\frac{O_t^*\left(O_t^*\right)'}{\sigma_t^4} \to \Sigma.
$$

Using the fact that $\max_{p+1\leq t\leq T}\|E_t\| = o_{\mathbb{P}}(1)$ and $\frac{O_t^*}{\sigma_t^2}$ is uniformly bounded in t, T, ω, the first assertion follows. The second assertion has been already shown in the proof of Theorem 1 with general weights W_t.□

E   Proof of Theorem 3

The proof uses the arguments given in Fryzlewicz et al. (2008) (see Proposition 3 of that paper). We have the decomposition

$$
\hat{\alpha}_t - \alpha(u) = -\hat{q}_{2,b',t}\left(\hat{\beta} - \beta\right) + S_{3,b',t}^{-1}\left(A_t(u) - A_t^{\#}(u)\right) + S_{3,b',t}^{-1}A_t^{\#}(u). \tag{E.1}
$$

Note that

$$
A_t^{\#}(u) = \sum_{i=p+1}^{T} w_{t,i}(b')W_i(u)M_i(u)\sigma_i^2(u)Z_i
$$

and using the central limit theorem for martingale differences, we obtain as in Fryzlewicz et al. (2008),

$$
\sqrt{Tb'}A_t^{\#}(u) \to \mathcal{N}_j\left(0, \text{Var}\left(\xi_1^2\right) \cdot \int K(x)^2 dx \cdot \mathbb{E}\left(W_1(u)\sigma_1(u)^4 M_1(u)M_1(u)'\right)\right).
$$

From Theorem 1, we have $\hat{\beta} - \beta = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$. Moreover, using Lemma 5 (point 4), and the fact that $S_{2,b',t}$ is uniformly bounded in t, T, ω, we get

$$
\|\hat{q}_{2,b',t}\| \leq \max_{p+1\leq t\leq T}\|S_{3,b',t}^{-1}\| \cdot \max_{p+1\leq t\leq T}\|S_{2,b',t}\| = O_{\mathbb{P}}(1).
$$

This leads to

$$
\hat{q}_{2,b',t}\left(\hat{\beta} - \beta\right) = o_{\mathbb{P}}\left(\frac{1}{\sqrt{Tb'}}\right). \tag{E.2}
$$

Moreover we have

$$\lim_{T \to \infty} S_{3,b',t} = \mathbb{E}\left(W_1(u)M_1(u)M_1(u)'\right) \text{ a.s.} \tag{E.3}$$

To justify (E.3), it is sufficient to show that for every real-valued process $(Y_t)_{p+1 \le t \le T}$ of the form $Y_t = f\left(\frac{t}{T}, X_{t-1}^2, \dots, X_{t-p}^2\right)$ with f bounded and Lipschitz, we have $\sum_{i=p+1}^{T} k_{t,i}(b')Y_i \to \mathbb{E}(Y_1(u))$. Using Lemma 1, the Lipschitz property of f and the properties of the kernel K, it is easy to get

$$\| \sum_{i=p+1}^{T} k_{t,i}(b')\mathbb{E}\left(W_i M_i M_i'\right) - \mathbb{E}\left(W_1(u)M_1(u)M_1(u)'\right) \| = O\left(b' + \frac{1}{T}\right).$$

For the stochastic part, we have, using Proposition 1, $\sum_{i=p+1}^{T} k_{t,i}(b')\bar{Y}_i = O_{\mathbb{P}}\left(\frac{1}{\sqrt{Tb'}}\right)$. This justifies the convergence (E.3).

The convergence in distribution announced in Theorem 3 now easily follows from (E.2), (E.3) and decomposition (E.1). $\square$

F Proof of Theorem 4

In this proof, we will set

$$\hat{a}_t = \begin{pmatrix} \hat{\alpha}_t \\ \hat{\beta} \end{pmatrix}, \quad X_i = \begin{pmatrix} M_i \\ N_i \end{pmatrix}, \quad X_i(u) = \begin{pmatrix} M_i(u) \\ N_i(u) \end{pmatrix}.$$

Then, setting $a_t = \begin{pmatrix} \alpha_t \\ \beta \end{pmatrix}$, we have the expressions

$$\hat{\sigma}_{t,i}^2 = X_i'\hat{a}_t, \quad \hat{\sigma}_{t,i}^2(u) = X_i(u)'\hat{a}_t.$$

As in the proof of Theorem 3, we have the decomposition

$$\hat{\alpha}_{*,t} - \alpha(u) = (\check{s}_{3,b',t})^{-1} \check{s}_{2,b',t}\left(\beta - \hat{\beta}\right) + (\check{s}_{3,b',t})^{-1}\left(\sum_{i=p+1}^{T} k_{t,i}(b')\check{W}_i^*(u)\sigma_i(u)^2 M_i(u)Z_i + B_t(u) - B_t^\#(u)\right).$$

- We first show that $\check{s}_{3,b',t} \to \mathbb{E}\left(\frac{M_1(u)M_1(u)'}{\sigma_1(u)^4}\right)$ in probability. We set $t_- = \max(1, t - Tb')$ and $t^+ = \min(T, t + Tb')$. Define the following event

$$A_{t,T} = \cap_{i=t_-}^{t^+} \left\{\hat{\sigma}_{t,i}^2 > \frac{1}{2}\sigma_i^2\right\}.$$

  Note that we have,

$$\frac{\hat{\sigma}_{t,i}^2}{\sigma_i^2} = 1 + \frac{X_i'}{\sigma_i^2}\left(\hat{a}_t - a_i\right)$$

  and $\frac{\|X_i\|}{\sigma_i^2}$ is a random variable bounded by a constant $C > 0$. Then we have the inclusion

$$\left\{\hat{\sigma}_{t,i}^2 \le \frac{1}{2}\sigma_i^2\right\} \subset \left\{\max_{t_- \le i \le t^+} \|\hat{a}_t - a_i\| \ge \frac{1}{2C}\right\}.$$

44

But

$$\max_{t_-\leq i\leq t^+} \|\hat{a}_t - a_i\| \leq \|\hat{a}_t - a(u)\| + \max_{t_-\leq i\leq t^+} \|a(u) - a_i\| \leq C_1\left(\|\hat{a}_t - a(u)\| + b' + \frac{1}{T}\right)$$

for a suitable constant $C_1 > 0$ and from Theorem 1 and Theorem 3, $\|\hat{a}_t - a(u)\| = O_{\mathbb{P}}\left(b' + \frac{1}{\sqrt{Tb'}}\right)$. This yields to $\max_{t_-\leq i\leq t^+} \|\hat{a}_t - a_i\| = O_{\mathbb{P}}\left(b' + \frac{1}{\sqrt{Tb'}}\right)$. Then we conclude that $\mathbb{P}\left(A_{t,T}^c\right) \to_{T\to\infty} 0$. On the event $A_{t,T}$, we have

$$
\begin{aligned}
|\check{W}_{t,i}^* - \frac{1}{\sigma_i^4}| &= \frac{|\sigma_i^4 - \hat{\sigma}_{t,i}^4 - \mu_T|}{\sigma_i^4\left(\hat{\sigma}_{t,i}^4 + \mu_T\right)} \\
&\leq 4\frac{|\sigma_i^4 - \hat{\sigma}_{t,i}^4 - \mu_T|}{\sigma_i^8}.
\end{aligned}
$$

But since $\frac{\hat{\sigma}_{t,i}^2}{\sigma_i^2} - 1 = \frac{X_i'}{\sigma_i^2}(\hat{a}_t - a_i)$, we conclude that

$$\check{W}_{t,i}^* = \frac{1}{\sigma_i^4}\left(1 + F_{t,i}\right),$$

with $\max_{t_-\leq i\leq t^+} |F_{t,i}| = o_{\mathbb{P}}(1)$. Using the properties of the kernel K, this yields to

$$\check{s}_{3,b',t} = \sum_{i=p+1}^{T} k_{t,i}(b')\frac{M_i M_i'}{\sigma_i^4} + o_{\mathbb{P}}(1).$$

Using the arguments used in the proof of Theorem 3, we have $\sum_{i=p+1}^{T} k_{t,i}(b')\frac{M_i M_i'}{\sigma_i^4} \to \mathbb{E}\left(\frac{M_1(u)M_1(u)'}{\sigma_1(u)^4}\right)$ in probability and the last expectation is also the limit of $\check{s}_{3,b',t}$.

- Next we show the convergence in distribution

$$\sqrt{Tb'}\sum_{i=p+1}^{T} k_{t,i}(b')\check{W}_{t,i}^*(u)M_i(u)\sigma_i(u)^2 Z_i \to \mathcal{N}_m\left(0, \mathcal{V}_*(u)\right).$$

As shown in the proof of Theorem 3, this convergence holds if $\check{W}_{t,i}^*(u)$ is replaced with $\frac{1}{\sigma_i(u)^4}$ in the last expression and it remains to show that

$$\sqrt{Tb'}\sum_{i=p+1}^{T} k_{t,i}(b')\left(\check{W}_{t,i}^*(u) - \frac{1}{\sigma_i(u)^4}\right)M_i(u)\sigma_i(u)^2 Z_i \to 0 \tag{F.1}$$

in probability. We will use the following equality

$$\check{W}_i^*(u) - \frac{1}{\sigma_i(u)^4} = \sum_{\ell=1}^{\ell_0} \frac{\left(\sigma_i(u)^4 - \hat{\sigma}_{t,i}(u)^4 - \mu_T\right)^\ell}{\sigma_i(u)^{4\ell+4}} + \frac{\left(\sigma_i(u)^4 - \hat{\sigma}_{t,i}(u)^4 - \mu_T\right)^{\ell_0+1}}{\sigma_i(u)^{4\ell_0+4}\hat{\sigma}_{t,i}(u)^4}.$$

First, it is easy to show that for $\ell = 1, \ldots, \ell_0$,

$$\sqrt{Tb'}\sum_{i=p+1}^{T} k_{t,i}(b')\frac{\left(\sigma_i(u)^4 - \hat{\sigma}_{t,i}(u)^4 - \mu_T\right)^\ell}{\sigma_i(u)^{4\ell+4}}M_i(u)\sigma_i(u)^2\left(\xi_i^2 - 1\right) = o_{\mathbb{P}}(1). \tag{F.2}$$

45

Indeed, using the equality

$$\frac{\hat{\sigma}^4_{t,i}(u)}{\sigma_i(u)^4} - 1 = (\hat{a}_t - a_i)' \frac{X_i(u)X_i(u)'}{\sigma_i(u)^4} (\hat{a}_t - a_i) + 2\frac{X_i(u)'}{\sigma_i(u)^2} (\hat{a}_t - a_i),$$

developing $\left(\frac{\hat{\sigma}^4_{t,i}(u)}{\sigma_i(u)^4} - 1\right)^\ell$ and using the decomposition $\hat{a}_t - a_i = \hat{a}_t - a(u) + a(u) - a_i$ it is clear that

$$\sum_{i=p+1}^{T} k_{t,i}(b') \frac{\left(\sigma_i(u)^4 - \hat{\sigma}_{t,i}(u)^4\right)^\ell}{\sigma_i(u)^{4\ell+4}} M_i(u)\sigma_i(u)^2 Z_i$$

is composed of terms which write as a product of factors involving some coordinates of the vector $\hat{a}_t - a_t$, μ_T and of one factor of type $\sum_{i=p+1}^{T} k_{t,i}(b')O_{u,t,i}\left(\xi_i^2 - 1\right)$ where $O_{u,t,i}$ is bounded and measurable w.r.t $\sigma\left(\xi_{i-s} : s \geq 1\right)$. Such a term is $o_{\mathbb{P}}\left(\frac{1}{\sqrt{Tb'}}\right)$ because $\hat{a}_t - a(u) = o_{\mathbb{P}}(1)$ and since $(O_{u,t,i}Z_i)_{p+1 \leq i \leq T}$ is a sequence of martingale differences, we have

$$\sum_{i=p+1}^{T} k_{t,i}(b')O_{u,t,i}Z_i = O_{\mathbb{P}}\left(\frac{1}{\sqrt{Tb'}}\right).$$

This proves (F.2).

Now we consider the remainder term. Recall that $\max_{t_- \leq i \leq t^+} \|\hat{a}_t - a_i\| = O_{\mathbb{P}}\left(b' + \frac{1}{\sqrt{Tb'}}\right)$. Using the assumption on b', this entails

$$\max_{t_- \leq i \leq t^+}\left(1 - \frac{\hat{\sigma}_{t,i}(u)^4}{\sigma_i(u)^4} - \frac{\mu_T}{\sigma_i(u)^4}\right)^{\ell_0+1} = O_{\mathbb{P}}\left(\left(b' + \frac{1}{\sqrt{Tb'}}\right)^{\ell_0+1}\right) = o_{\mathbb{P}}\left(\frac{1}{\sqrt{Tb'}}\right). \tag{F.3}$$

If we define

$$A_{t,T}(u) = \cap_{i=t_-}^{t^+}\left\{\hat{\sigma}^2_{t,i}(u) > \frac{1}{2}\sigma_i(u)^2\right\},$$

we have $\mathbb{P}\left(A_{t,T}(u)^c\right) \rightarrow_{T \rightarrow \infty} 0$ (the proof is the same as for $A_{t,T}$). Moreover we have, using (F.3),

$$\mathbb{1}_{A_{t,T}(u)}\left\|\sum_{i=p+1}^{T} k_{t,i}(b')M_i(u)\sigma_i^2(u) \frac{\left(\sigma_i(u)^4 - \hat{\sigma}_{t,i}(u)^4 - \mu_T\right)^{\ell_0+1}}{\sigma_i(u)^{4\ell_0+4}\hat{\sigma}_{t,i}(u)^4}\left(\xi_i^2 - 1\right)\right\|$$

$$\leq 4 \max_{t_- \leq i \leq t^+}\left\{\left|1 - \frac{\hat{\sigma}_{t,i}(u)^4}{\sigma_i(u)^4} - \frac{\mu_T}{\sigma_i(u)^4}\right|^{\ell_0+1} \cdot \frac{\|M_i(u)\|}{\sigma_i(u)^2}\right\} \cdot \sum_{i=p+1}^{T} k_{t,i}(b')|\xi_i^2 - 1|$$

$$= o_{\mathbb{P}}\left(\frac{1}{\sqrt{Tb'}}\right).$$

Then, we conclude that

$$\sum_{i=p+1}^{T} k_{t,i}(b') \frac{\left(\sigma_i^4(u) - \hat{\sigma}^4_{t,i}(u) - \mu_T\right)^{\ell_0+1}}{\sigma_i(u)^{4\ell_0+4}\hat{\sigma}^4_{t,i}(u)} M_i(u)\sigma_i(u)^2 Z_i = o_{\mathbb{P}}\left(\frac{1}{\sqrt{Tb'}}\right).$$

This shows (F.1).

- Finally, we show that $B_t(u) - B_t^\#(u) = O_\mathbb{P}(b')$. We consider the event $\widetilde{A}_{t,T} = A_{t,T} \cap A_{t,T}(u)$. We have of course $\lim_{T\to\infty} \mathbb{P}\left(\left(\widetilde{A}_{t,T}\right)^c\right) = 0$. We will just show that

$$H_t = \check{s}_{1,b',t} - \sum_{i=p+1}^{T} k_{t,i}(b')\check{W}_{t,i}^*(u)M_i(u)X_i(u)^2 = O_\mathbb{P}(b'),$$

the control of the difference being the same for the two other quantities $\check{s}_{2,b',t}$ and $\check{s}_{3,b',t}$. Since $\mathbb{P}\left(\widetilde{A}_{t,T}\right) \to 0$, we have $H_t \mathbb{1}_{\left(\widetilde{A}_{t,T}\right)^c} = O_\mathbb{P}(b')$. Now, on the event $\widetilde{A}_{t,T}$, we have

$$\check{W}_{t,i}^* M_i X_i^2 - \check{W}_{t,i}^*(u)M_i(u)X_i^2(u) = \frac{M_i}{\sqrt{\hat{\sigma}_{t,i}^4 + \mu_T}}F_i + \frac{X_i(u)^2}{\sqrt{\hat{\sigma}_{t,i}^4(u) + \mu_T}}G_i,$$

where

$$F_i = \frac{X_i^2}{\sqrt{\hat{\sigma}_{t,i}^4 + \mu_T}} - \frac{X_i(u)^2}{\sqrt{\hat{\sigma}_{t,i}(u)^4 + \mu_T}},$$

$$G_i = \frac{M_i}{\sqrt{\hat{\sigma}_{t,i}^4 + \mu_T}} - \frac{M_i(u)}{\sqrt{\hat{\sigma}_{t,i}(u)^4 + \mu_T}}.$$

On $\widetilde{A}_{t,T}$ we have for a suitable constant $C > 0$,

$$\begin{aligned}
&|F_i| \\
\leq\ & \frac{|X_i^2 - X_i(u)|}{\sigma_i^2} + \xi_i^2 \cdot \frac{|\hat{\sigma}_{t,i}^2 - \hat{\sigma}_{t,i}^2(u)|}{\sigma_i^2} \cdot \frac{\hat{\sigma}_{t,i}^2 + \hat{\sigma}_{t,i}^2(u)}{\sigma_i^2 + \sigma_i^2(u)} \\
\leq\ & C\left(|X_i^2 - X_i(u)^2| + \|\hat{a}_t\|^2 \xi_i^2 \sum_{\ell=1}^{p} |X_{i-\ell}^2 - X_{i-\ell}^2(u)|\right).
\end{aligned}$$

Moreover, $\mathbb{1}_{\widetilde{A}_{t,T}} \frac{M_i}{\sqrt{\hat{\sigma}_{t,i}^4 + \mu_T}}$ is bounded uniformly in t, i, T, ω. Then, we obtain for a suitably chosen constant $D > 0$,

$$\begin{aligned}
&\mathbb{1}_{\widetilde{A}_{t,T}} \left\| \sum_{i=p+1}^{T} k_{t,i}(b') \frac{M_i}{\sqrt{\hat{\sigma}_{t,i}^4 + \mu_T}} F_i \right\| \\
\leq\ & D\left\{ \sum_{i=p+1}^{T} k_{t,i}(b')|X_i^2 - X_i(u)^2| + \|\hat{a}_t\|^2 \sum_{\ell=1}^{p} \sum_{i=p+1}^{T} k_{t,i}(b')\xi_i^2|X_{i-\ell}^2 - X_{i-\ell}^2(u)| \right\}.
\end{aligned}$$

Using the fact that $\|\hat{a}_t\| = O_\mathbb{P}(1)$, the support condition on the kernel K and Lemma 1, we conclude that

$$\mathbb{1}_{\widetilde{A}_{t,T}} \sum_{i=p+1}^{T} k_{t,i}(b') \frac{M_i}{\sqrt{\hat{\sigma}_{t,i}^4 + \mu_T}} F_i = O_\mathbb{P}(b').$$

47

Similarly, we also get

$$\mathbb{1}_{\widetilde{A}_{t,T}} \sum_{i=p+1}^{T} k_{t,i}(b') \frac{X_i(u)^2}{\sqrt{\hat{\sigma}_{t,i}^4(u) + \mu_T}} G_i = O_{\mathbb{P}}(b').$$

This proves that $H_t \mathbb{1}_{\widetilde{A}_{t,T}} = O_{\mathbb{P}}(b')$ and then $H_t = O_{\mathbb{P}}(b')$. The proof is now complete. □

G Auxiliary results for the proof of Theorem 5

In this section, we assume that the assumptions of Theorem 5 hold true. The quantities studied in the following lemma are defined in (H.1).

Lemma 7. *We set $\mathcal{B} = [0, b] \cup [1 - b, 1]$.*

1. *We have $\max_{u \in [0,1]} \|\mathbb{E}(B_u)\| = O(b)$ and for all positive integer q, $\max_{u \in [0,1]} \mathbb{E}\left(\|\bar{B}_u\|^{2q}\right) = O\left(\frac{b^q}{T^q}\right)$.*

2. *$\max_{u \in [0,1]} \mathbb{E}\left(\|\Delta_u\|^4\right) = O\left((Tb)^{-2}\right)$.*

3. *We have for all positive integer q,*

$$\max_{u \in \mathcal{B}} \mathbb{E}\|S_u - \kappa_u\|^{2q} = O(1), \qquad \max_{u \in [b, 1-b]} \mathbb{E}\|S_u - \kappa_u\|^{2q} = O\left(b^{2q} + (Tb)^{-q}\right).$$

4. *We set $R_u = \widetilde{R}_u S_u^{-1}$. We have $\max_{u \in [0,1]} \|S_u^{-1}\| = O_{\mathbb{P}}(1)$, $\max_{u \in [b, 1-b]} \mathbb{E}\left(\|\widetilde{R}_u\|^{2q}\right) = O\left(b^{4q} + (Tb)^{-2q}\right)$ and $\max_{u \in \mathcal{B}} \mathbb{E}\left(\|\widetilde{R}_u\|^{2q}\right) = O(1)$.*

Proof of Lemma 7

1. The first assertion is obvious. For the second assertion, we use Lemma 2. It is sufficient to show that for some subscripts y, w,

$$\max_{u \in [0,1]} \mathbb{E}|\sum_{i=p+1}^{T} z_i(u) \overline{W_i X_{i,y} X_{i,w}}|^{2q} = O\left(\frac{b^q}{T^q}\right),$$

where $z_i(u) = e_i(u) \left(a_w(i/T) - a_w(u)\right)$ satisfies

$$\max_{p+1 \le i \le T, u \in [0,1]} |q_i(u)| = O\left(\frac{1}{T}\right) \text{ and } \max_{u \in [0,1]} \sum_{i=p+1}^{T} z_i(u) = O(b).$$

Then the result follows from Lemma 2 if we write the moment of order $2q$ as a multiple sum.

2. The result follows from Proposition 3.

3. Using Proposition 3, we have $\max_{u \in [0,1]} \mathbb{E}\|\bar{S}_u\|^{2q} = O\left(\frac{1}{(Tb)^q}\right)$. Moreover for the bias part

$$\mathbb{E}(S_u) - \kappa_u = \sum_{i=p+1}^{T} e_i(u) \left(\mathbb{E}\left(W_i X_i X_i'\right) - \kappa_u\right) + \left(\sum_{i=p+1}^{T} e_i(u) - 1\right) \kappa_u.$$

48

The first term is bounded by $C\left(b + \frac{1}{T}\right)$ where $C > 0$ does not depend on u, T and

$$\max_{u \in I} \Big| \sum_{i=p+1}^{T} e_i(u) - 1 \Big|$$

is $O\left(\frac{1}{Tb}\right)$ or $O(1)$ when $I = [b, 1 - b]$ or $I = [0, 1] \setminus [b, 1 - b]$ respectively.

4. Since $S_{t/T} = \frac{1}{Tb} \sum_{i=p+1}^{T} K\left(\frac{t-i}{Tb}\right) W_i X_i X_i'$, note that if $|u - \frac{t}{T}| \leq \frac{1}{T}$, then $\|S_u - S_{t/T}\| \leq \frac{C}{Tb^2}$ where $C > 0$ does not depend on u, t and T (we recall that $W_i X_i X_i'$ is bounded). Then it is sufficient to show that $\max_{p+1 \leq t \leq T} \|S_{t/T}^{-1}\| = O_{\mathbb{P}}(1)$. This can be proved as in Lemma 5, point 4. Details are omitted. The other assertions are a consequence of the previous point and of the bound $\|\widetilde{R}_u\| \leq C \|S_u - \kappa_u\|^2.$ $\square$

Lemma 8. *Let* $\left(\widetilde{\Gamma}(u)\right)_{u \in [0,1]}$ *be a family of positive definite matrices such that* $u \mapsto \widetilde{\Gamma}(u)$ *is Lipschitz. Set* $\mathcal{V}_T = \int_0^1 \Lambda_u' \widetilde{\Gamma}(u) \Lambda_u du$. *Then*

$$T \sqrt{b} \left(\mathcal{V}_T - \mathbb{E}(\mathcal{V}_T)\right) \to_{T \to \infty} \mathcal{N}\left(0, 4\|K^*\|_2^2 Var^2\left(\xi_1^2\right)\zeta\right),$$

where

$$\zeta = \int_0^1 tr\left(\left(\widetilde{\Gamma}(u)G(u)\right)^2\right)du, \quad G(u) = \mathbb{E}\left(W_1(u)^2 \sigma_1(u)^4 X_1(u) X_1(u)'\right).$$

Moreover

$$\mathbb{E}(\mathcal{V}_T) = \frac{Var\left(\xi_1^2\right)\|K\|_2^2}{Tb} \cdot \int_0^1 tr\left(\widetilde{\Gamma}(u)G(u)\right)du + o\left(\frac{1}{T\sqrt{b}}\right).$$

Proof of Lemma 8 We set for $i, j \in [\![p+1, T]\!]$, $Q_{i,j} = \int_0^1 e_i(u)e_j(u)\widetilde{\Gamma}(u)du$. Moreover, let $\mathcal{Y}_i = W_i \sigma_i^2 X_i$. Then

$$\mathcal{V}_T = \sum_{i,\ell=p+1}^{T} Z_\ell Z_i \mathcal{Y}_i' Q_{i,\ell} \mathcal{Y}_\ell.$$

We use the decomposition $\mathcal{V}_T = 2\mathcal{V}_{1,T} + \mathcal{V}_{2,T}$ with

$$\mathcal{V}_{1,T} = \sum_{p+1 \leq \ell < i \leq T} Z_\ell Z_i \mathcal{Y}_i' Q_{i,\ell} \mathcal{Y}_\ell,$$

and

$$\mathcal{V}_{2,T} = \sum_{i=p+1}^{T} Z_i^2 \mathcal{Y}_i' Q_{i,i} \mathcal{Y}_i.$$

Note that $\mathbb{E}(\mathcal{V}_T) = \mathbb{E}(\mathcal{V}_{2,T})$ and $\max_{i,\ell} \|Q_{i,\ell}\| = O\left(\frac{b}{(Tb)^2}\right)$.

- We first show that

$$T \sqrt{b}\left(\mathcal{V}_{2,T} - \mathbb{E}(\mathcal{V}_{2,T})\right) = o_{\mathbb{P}}(1). \tag{G.1}$$

To show this, we decompose

$$\mathcal{V}_{2,T} - \mathbb{E}(\mathcal{V}_{2,T}) = \mathcal{V}_{2,1,T} + \mathbb{E}(Z_1^2)\mathcal{V}_{2,1,T},$$

49

with

$$\mathcal{V}_{2,1,T} = \sum_{i=p+1}^{T} \left(Z_i^2 - \mathbb{E}Z_i^2\right) \mathcal{Y}_i' Q_{i,i} \mathcal{Y}_i,$$

$$\mathcal{V}_{2,2,T} = \sum_{i=p+1}^{T} \overline{\mathcal{Y}_i' Q_{i,i} \mathcal{Y}_i}.$$

Since $(\mathcal{Y}_i)$ is bounded and $\mathcal{V}_{2,1,T}$ is a sum of uncorrelated random variables, the second moment of $\mathcal{V}_{2,1,T}$ is $O\left(\frac{1}{T^3 b^3}\right)$. Then $T\sqrt{b}\mathcal{V}_{2,1,T} = o_{\mathbb{P}}(1)$. Moreover, one can decompose $\mathcal{V}_{2,2,T}$ as a finite sum of terms of order $O_{\mathbb{P}}\left(\frac{1}{\sqrt{T^3 b^2}}\right)$. This can be easily seen taking the second moment of these terms and then using Lemma 2. This leads to (G.1).

- Next, we prove the assertion on $\mathbb{E}(\mathcal{V}_{2,T})$. We have

$$
\begin{aligned}
\mathbb{E}(\mathcal{V}_{2,T}) &= \operatorname{Var}\left(\xi_1^2\right) \sum_{i=p+1}^{T} \mathbb{E}\left(\mathcal{Y}_i' Q_{i,i} \mathcal{Y}_i\right) \\
&= \operatorname{Var}\left(\xi_1^2\right) \sum_{i=p+1}^{T} \int_0^1 \frac{1}{(Tb)^2} K^2\left(\frac{u - \frac{i}{T}}{b}\right) \mathbb{E}\left(\mathcal{Y}_1\left(i/T\right)' \widetilde{\Gamma}(u) \mathcal{Y}_1\left(i/T\right)\right) du + O\left(\frac{b}{(Tb)^2}\right) \\
&= \operatorname{Var}\left(\xi_1^2\right) \sum_{i=p+1}^{T} \int_0^1 \frac{1}{(Tb)^2} K^2\left(\frac{u - \frac{i}{T}}{b}\right) du \mathbb{E}\left(\mathcal{Y}_1\left(i/T\right)' \widetilde{\Gamma}\left(\frac{i}{T}\right) \mathcal{Y}_1\left(i/T\right)\right) + O\left(\frac{1}{T}\right).
\end{aligned}
$$

Next, note that for a Lipschitz function $f : [0, 1] \to \mathbb{R}$,

$$\frac{1}{Tb} \sum_{i=p+1}^{T} \int_0^1 K^2\left(\frac{u - \frac{i}{T}}{b}\right) f\left(\frac{i}{T}\right) du - \int_0^1 f(u) du \|K\|_2^2 = O\left(b + \frac{1}{Tb}\right).$$

This yields to

$$\mathbb{E}(\mathcal{V}_{2,T}) - \frac{\operatorname{Var}\left(\xi_1^2\right)}{Tb} \int_0^1 \mathbb{E}\left(\mathcal{Y}_1(u)' \widetilde{\Gamma}(u) \mathcal{Y}_1(u)\right) du \|K\|_2^2 = O\left(\frac{1}{T} + \frac{1}{(Tb)^2}\right).$$

But using the fact that $Tb^{3/2} \to \infty$, the last quantity is $O\left(\frac{1}{T\sqrt{b}}\right)$.

- Finally, we study the convergence in distribution of $\mathcal{V}_{1,T}$. The argument is to apply the central limit theorem for martingales differences. First, we show the Lindberg condition. We set $\mathcal{P}_i = \sum_{\ell=p+1}^{i-1} Z_\ell \mathcal{Y}_\ell' Q_{i,\ell} \mathcal{Y}_i$. Then $\mathcal{V}_{1,T} = \sum_{i=p+1}^{T} \mathcal{P}_i Z_i$. It is enough to show that

$$\left(T\sqrt{b}\right)^4 \sum_{i=p+1}^{T} \mathbb{E}\left(\mathcal{P}_i^4 Z_i^4\right) = o(1). \tag{G.2}$$

Using the independence between Z_i and $\mathcal{P}_i$, the fact that $\mathcal{Y}_i$ is bounded and the Burhhölder inequality, we have for a generic constant $C > 0$,

$$
\begin{aligned}
\mathbb{E}\left(\mathcal{P}_i^4 Z_i^4\right) &\leq C\mathbb{E}\left(\|\sum_{\ell=p+1}^{i-1} Z_\ell \mathcal{Y}'_\ell Q_{i,\ell}\|^4\right) \\
&\leq C\left(\sum_{\ell=p+1}^{i-1} \|Q_{i,\ell}\|^2\right)^2 \\
&= O\left(\frac{1}{(T^3 b^2)^2}\right).
\end{aligned}
$$

This leads to (G.2), using the condition $Tb^2 \to \infty$. Now, to obtain the convergence mentioned in Lemma 8, it remains to prove that

$$
T^2 b\mathbb{E}\left(Z_1^2\right) \sum_{i=p+1}^{T} \mathcal{P}_i^2 - \|K^*\|_2^2 \mathbb{E}^2\left(Z_1^2\right) \zeta = o_{\mathbb{P}}(1). \tag{G.3}
$$

We will use the following decomposition

$$
\sum_{i=p+1}^{T} \mathcal{P}_i^2 = \sum_{i=p+1}^{T} A_i + 2B_1 + B_2,
$$

where

$$
\begin{aligned}
A_i &= \sum_{\ell,m=p+1}^{i-1} Z_\ell \mathcal{Y}'_\ell Q_{i,\ell} \overline{\mathcal{Y}_i \mathcal{Y}'_i} Q_{i,m} \mathcal{Y}_m Z_m \\
&= \int_0^1 \int_0^1 \sum_{\ell,m=p+1}^{T} e_i(u)e_i(v)e_\ell(u)e_m(v)B_{i,\ell,m}(u,v)dudv,
\end{aligned}
$$

with

$$
B_{i,\ell,m}(u,v) = Z_\ell \mathcal{Y}'_\ell \widetilde{\Gamma}(u)\overline{\mathcal{Y}_i \mathcal{Y}'_i}\widetilde{\Gamma}(v)\mathcal{Y}_m Z_m
$$

and

$$
B_1 = \sum_{p+1\leq\ell<m\leq T} Z_\ell \mathcal{Y}'_\ell E_{\ell,m}\mathcal{Y}_m Z_m,
$$

$$
B_2 = \sum_{\ell=p+1}^{T} Z_\ell^2 \mathcal{Y}'_\ell E_{\ell,\ell}\mathcal{Y}_\ell,
$$

$$
E_{\ell,m} = \sum_{i=\ell\vee m+1}^{T} Q_{i,\ell}\mathbb{E}\left(\mathcal{Y}_i \mathcal{Y}'_i\right) Q_{i,m}.
$$

We also set $p_{i,\ell,m}(u,v) = e_i(u)e_i(v)e_\ell(u)e_m(v) = O\left(\frac{1}{(Tb)^4}\right)$.

– We first show that $T^2 b \sum_{i=p+1}^{T} A_i = o_{\mathbb{P}}(1)$. We have

$$\mathbb{E}\left[|\sum_{i=p+1}^{T} A_i|^2\right]$$

$$\leq \int_0^1 \int_0^1 \sum_{i,i'=p+1}^{T} \sum_{p+1\leq\ell,m<i} \sum_{p+1\leq\ell',m'<i'} p_{i,\ell,m}(u,v) p_{i',\ell',m'}(u,v) |\mathbb{E}\left(B_{i,\ell,m}(u,v) B_{i',\ell',m'}(u,v)\right)| du dv$$

$$\leq \int_{|u-v|\leq 2b} \sum_{i,\ell,m\in\mathcal{M}_u} \sum_{i',\ell',m'\in\mathcal{M}_u} p_{i,\ell,m}(u,v) p_{i',\ell',m'}(u,v) |\mathbb{E}\left(B_{i,\ell,m}(u,v) B_{i',\ell',m'}(u,v)\right)| du dv,$$

where $\mathcal{M}_u = \left\{s \in [\![p+1,T]\!] : |\frac{s}{T} - u| \leq 3b\right\}$. Then the last quantity is

$$O\left(b\frac{1}{(Tb)^8}(Tb)^3\right) = O\left(\frac{1}{T^5 b^4}\right).$$

Indeed, $p_{i,\ell,m}(u,v)$ can be bounded by $\frac{1}{(Tb)^4}$ (up to a constant) and we can apply Lemma 2 with $z_{T,i} = \mathbb{1}_{i\in\mathcal{M}_u}$. Since $\frac{T^4 b^2}{T^5 b^4} = \frac{1}{Tb^2} \to 0$, we get the result.

– Next, we show that $T^2 b B_1 = o_{\mathbb{P}}(1)$. Using the fact that $(\mathcal{Y}_i)_i$ is bounded and the Burkhölder inequality, we get

$$\mathbb{E}\left|\sum_{p+1\leq\ell<m\leq T} Z_\ell \mathcal{Y}'_\ell E_{\ell,m} \mathcal{Y}_m Z_m\right|^2 \leq C \sum_{m=p+1}^{T} \sum_{\ell=p+1}^{m-1} \|E_{\ell,m}\|^2 \leq C \sum_{|\ell-m|\leq 4Tb} \|E_{\ell,m}\|^2.$$

Since $\max_{\ell,m} \|E_{\ell,m}\| = O\left(\frac{1}{T^3 b}\right)$, we find that

$$\mathbb{E}\left|\sum_{p+1\leq\ell<m\leq T} Z_\ell \mathcal{Y}'_\ell E_{\ell,m} \mathcal{Y}_m Z_m\right|^2 = O\left(\frac{1}{T^4 b}\right).$$

This proves $T^2 b B_1 = o_{\mathbb{P}}(1)$.

– Finally, we study the convergence of $T^2 b B_2$. First, observing that $\max_{p+1\leq\ell\leq T} \|E_{\ell,\ell}\| = O\left(\frac{b^2}{(Tb)^3}\right)$ and using Lemma 2, we get

$$\sum_{\ell=p+1}^{T} \left\{Z_\ell^2 \mathcal{Y}'_\ell E_{\ell,\ell} \mathcal{Y}_\ell - \mathbb{E}\left(Z_\ell^2 \mathcal{Y}'_\ell E_{\ell,\ell} \mathcal{Y}_\ell\right)\right\} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T^5 b^2}}\right).$$

Then it remains to show that

$$T^2 b \sum_{\ell=p+1}^{T} \mathbb{E}\left(\mathcal{Y}'_\ell E_{\ell,\ell} \mathcal{Y}'_\ell\right) - \|K^*\|^2 \zeta = o_{\mathbb{P}}(1). \tag{G.4}$$

We first note that

$$\sum_{\ell=p+1}^{T} \mathbb{E}\left(\mathcal{Y}'_\ell E_{\ell,\ell} \mathcal{Y}_\ell\right) = \sum_{p+1\leq\ell<i\leq T} \text{tr}\left(\mathbb{E}\left(\mathcal{Y}'_\ell \mathcal{Y}_\ell\right) Q_{i,\ell} \mathbb{E}\left(\mathcal{Y}_i \mathcal{Y}'_i\right) Q_{i,\ell}\right).$$

Next, one can verify (we skip the details) that one can replace $\mathbb{E}\left(\mathcal{Y}_i\mathcal{Y}_i'\right)$ with $\mathbb{E}\left(\mathcal{Y}_\ell\mathcal{Y}_\ell'\right)$ and $Q_{i,\ell}$ with $\int_0^1 e_i(u)e_\ell(u)du\widetilde{\Gamma}\left(\frac{\ell}{T}\right)$ without changing the limit in (G.4). Then it remains to study the limit of

$$T^2 b \sum_{\ell=p+1}^T f_\ell s_{T,\ell},$$

where

$$f_\ell = \operatorname{tr}\left(\left[\mathbb{E}\left(\mathcal{Y}_\ell\mathcal{Y}_\ell'\right)\widetilde{\Gamma}\left(\frac{\ell}{T}\right)\right]^2\right),$$

and

$$s_{T,\ell} = \sum_{i=\ell+1}^T \left(\int_0^1 e_i(u)e_\ell(u)du\right)^2 = O\left(\frac{1}{T^3 b}\right).$$

Note first that

$$\sum_{p+1\le\ell\le Tb} f_\ell s_{T,\ell} + \sum_{\ell=T(1-3b)}^T f_\ell s_{T,\ell} = O\left(\frac{1}{T^2}\right).$$

Moreover, if $Tb \le \ell \le T(1 - 3b)$, we have

$$s_{T,\ell} = \sum_{i=\ell+1}^{\ell+2Tb} \frac{1}{T^4 b^2} H\left(\frac{i-\ell}{Tb}\right)^2,$$

where $H(x) = \int_{-1}^1 W(v)W(x+v)dv$. Since

$$\frac{1}{Tb}\sum_{k=1}^{2Tb} H\left(\frac{k}{Tb}\right)^2 - \|K^*\|_2^2 = O\left(\frac{1}{Tb}\right)$$

and

$$\frac{1}{T}\sum_{\ell=Tb}^{T(1-3b)} f_\ell \to \int_0^1 \operatorname{tr}\left(\left[\mathbb{E}\left(\mathcal{Y}_1(u)\mathcal{Y}_1(u)'\right)\widetilde{\Gamma}(u)\right]^2\right)du = \zeta,$$

we get (G.4). $\square$

H   Proof of Theorem 5

The proof of Theorem 5 uses two lemma, Lemma 7 and Lemma 8, which are stated in the next section. We set $\widetilde{\kappa}_u = A\kappa_u^{-1}$ and $\mathcal{Y}_i = W_i\sigma_i^2 X_i$. Under the null assumption (β is non time-varying), we use the following decomposition of the difference $\hat{\beta} - \beta$.

$$\widetilde{\beta}(u) - \beta = \widetilde{\kappa}_u\Lambda(u) + \widetilde{\kappa}_u\bar{B}_u + z_u + \left[\widetilde{\kappa}_u\left(\kappa_u - S_u\right)\kappa_u^{-1} + R_u\right]\cdot\left[B_u + \Lambda_u\right], \tag{H.1}$$

where

$$\Lambda_u = \sum_{i=p+1}^T e_i(u)\mathcal{Y}_iZ_i, \quad B_u = \sum_{i=p+1}^T e_i(u)W_iX_iX_i'\left(a(i/T) - a(u)\right),$$

$$R_u = \widetilde{\kappa}_u\left(\kappa_u - S_u\right)\kappa_u^{-1}\left(\kappa_u - S_u\right)S_u^{-1}, \quad z_u = \widetilde{\kappa}_u\sum_{i=p+1}^T e_i(u)\left[\mathbb{E}\left(W_iX_iX_i'\right) - \kappa_u\right]\cdot\left[a(i/T) - a(u)\right].$$

Note that $\max_{u\in[0,1]} \|z_u\| = O\left(\frac{b}{T} + b^2\right) = O\left(b^2\right)$. The first part of Theorem 5 will follow from Lemma 8, if we show that

$$\int_0^1 \left(\widetilde{\beta}(u) - \beta\right)' \Gamma(u) \left(\widetilde{\beta}(u) - \beta\right) du - \int_0^1 \Lambda_u' \widetilde{\kappa}(u)' \Gamma(u) \widetilde{\kappa}_u \Lambda_u du = o_{\mathbb{P}}\left(\frac{1}{T\sqrt{b}}\right). \tag{H.2}$$

We set $H(u) = \widetilde{\kappa}_u \Gamma(u) \widetilde{\kappa}_u$. To show (H.2), we use the previous decomposition and proceed as follows.

1. We first show that

$$T\sqrt{b} \int_0^1 \bar{B}_u' H(u) \Lambda_u du = o_{\mathbb{P}}(1). \tag{H.3}$$

We have for a suitable constant $C > 0$,

$$\max_{u\in[0,1]} \mathbb{E}\left|\bar{B}_u' H(u) \Lambda_u\right| \le C \max_{u\in[0,1]} \sqrt{\mathbb{E}\left(\|\bar{B}_u\|^2\right) \cdot \mathbb{E}\left(\|\Lambda_u\|^2\right)}.$$

Then using Lemma 7, we get

$$\max_{u\in[0,1]} \mathbb{E}\left|\bar{B}_u' H(u) \Lambda_u\right| = O\left(\frac{1}{T}\right)$$

and (H.3) easily follows.

2. Next we show that

$$T\sqrt{b} \int_0^1 \bar{B}_u' H(u) \bar{B}_u du = o_{\mathbb{P}}(1). \tag{H.4}$$

Using Lemma 7, we have $\max_{u\in[0,1]} \mathbb{E}\|\bar{B}_u\|^2 = O\left(\frac{b}{T}\right)$ and (H.4) easily follows.

3. Next, we prove that

$$\int_0^1 z_u' \Gamma(u) \widetilde{\kappa}_u \Lambda_u du = o_{\mathbb{P}}\left(\frac{1}{T\sqrt{b}}\right). \tag{H.5}$$

We have

$$\mathbb{E}\left|\int_0^1 z_u' \Gamma(u) \widetilde{\kappa}_u \Lambda_u du\right| = O\left(\max_{u\in[0,1]} \|z_u\| \cdot \sqrt{\mathbb{E}\left(\|\Lambda_u\|^2\right)}\right).$$

Then using Lemma 7, we get

$$\mathbb{E}\left|\int_0^1 z_u' \Gamma(u) \widetilde{\kappa}_u \Lambda_u du\right| = O\left(\frac{b^2}{\sqrt{Tb}}\right).$$

This leads to (H.5), using our bandwidth conditions.

4. Under our bandwidth conditions, we have $T\sqrt{b} \int_0^1 z_u' \Gamma(u) z_u du = O\left(Tb^{4.5}\right) = o(1)$.

5. Next we show that

$$\int_0^1 \Lambda_u' H(u) \left(\kappa_u - S_u\right) \kappa_u^{-1} \left(B_u + \Lambda_u\right) du = o_{\mathbb{P}}\left(\frac{1}{T\sqrt{b}}\right). \tag{H.6}$$

54

If I is a subinterval of $[0, 1]$, we have for a positive constant C,

$$\mathbb{E}\Big| \int_I \Lambda'_u H(u) (\kappa_u - S_u) \kappa_u^{-1} \big(\bar{B}_u + \Lambda_u\big) du\Big|$$

$$\leq \quad C \cdot |I| \cdot \max_{u \in I} \mathbb{E}^{1/3}\big(\|\Lambda_u\|^3\big) \cdot \mathbb{E}^{1/3}\big(\|\kappa_u - S_u\|^3\big) \cdot \big(\mathbb{E}^{1/3}\big(\|\bar{B}_u\|^3\big) + \mathbb{E}^{1/3}\big(\|\Lambda_u\|^3\big)\big).$$

When $I = [0, b] \cup [1 - b, 1]$, the last bound is $O\big(\frac{1}{T}\big)$, using Lemma 7. Then,

$$\int_I \Lambda'_u H(u) (\kappa_u - S_u) \kappa_u^{-1} \big(\bar{B}_u + \Lambda_u\big) du = o_{\mathbb{P}}\Big(\frac{1}{T\sqrt{b}}\Big).$$

When $I = [b, 1 - b]$, Lemma 7 leads to

$$\int_I \Lambda'_u H(u) (\kappa_u - S_u) \kappa_u^{-1} \big(\bar{B}_u + \Lambda_u\big) du = O\Big(\frac{1}{Tb} \cdot \big(b + \frac{1}{\sqrt{Tb}}\big)\Big).$$

Using our bandwidth conditions, this yields to

$$\int_0^1 \Lambda'_u H(u) (\kappa_u - S_u) \kappa_u^{-1} \big(\bar{B}_u + \Lambda_u\big) du = o_{\mathbb{P}}\Big(\frac{1}{T\sqrt{b}}\Big).$$

Moreover, if $I = [0, b] \cup [1 - b, 1]$ or $I = [b, 1 - b]$, we have

$$\mathbb{E}\Big| \int_I \Lambda'_u H(u) (\kappa_u - S_u) \kappa_u^{-1} \mathbb{E}(B_u) du\Big|$$

$$\leq \quad Cb|I| \max_{u \in I} \sqrt{\mathbb{E}\|S_u - \kappa_u\|^2} \max_{u \in [0,1]} \sqrt{\mathbb{E}\|\Lambda_u\|^2}$$

$$\leq \quad C \frac{\sqrt{b}}{\sqrt{T}} |I| \max_{u \in I} \sqrt{\mathbb{E}\|S_u - \kappa_u\|^2}.$$

Using Lemma 7 and our bandwidth conditions, we also obtain $\int_0^1 \Lambda'_u H(u) (\kappa_u - S_u) \kappa_u^{-1} \mathbb{E}(B_u) du = o_{\mathbb{P}}\big(\frac{1}{T\sqrt{b}}\big)$ and then (H.6).

6. Next, setting $M_u = \widetilde{\kappa}_u (\kappa_u - S_u) \kappa_u^{-1} (B_u + \Lambda_u)$, we show that

$$T\sqrt{b} \int_0^1 M'_u \Gamma(u) M_u du = o_{\mathbb{P}}(1). \tag{H.7}$$

Using Lemma 7, we have

$$\mathbb{E}\|M_u\|^2 \leq C\sqrt{\mathbb{E}\big(\|\kappa_u - S_u\|^4\big) \cdot \mathbb{E}\big(\|B_u + \Lambda_u\|^4\big)} \leq C\sqrt{\mathbb{E}\big(\|\kappa_u - S_u\|^4\big)}\Big(b^2 + \frac{1}{Tb}\Big).$$

Then, studying $\mathbb{E}\big| \int_I M'_u \Gamma(u) M_u du\big|$ when $I = [0, b] \cup [1 - b, 1]$ or $I = [b, 1 - b]$, (H.7) follows using the previous bound, Lemma 7 and our bandwidth conditions.

7. Next we prove that

$$T \sqrt{b} \int_0^1 \Lambda'_u \Gamma(u) R_u \left(B_u + \Lambda_u \right) du = o_{\mathbb{P}}(1). \tag{H.8}$$

When $I = [b, 1 - b]$ or $I = [0, b] \cup [1 - b, 1]$, we use the bound

$$\left| \int_0^1 \Lambda'_u \Gamma(u) R_u \left(B_u + \Lambda_u \right) du \right|^2$$

$$\leq \quad C \int_I \|\Lambda_u\|^2 \left(\|B_u\|^2 + \|\Lambda_u\|^2 \right) du \cdot \int_I \|R_u\|^2 du.$$

Moreover, using Lemma 7, we have

$$\mathbb{E} \int_I \|\Lambda_u\|^2 \left(\|B_u\|^2 + \|\Lambda_u\|^2 \right) du \leq 2|I| \max_{u \in [0,1]} \sqrt{\mathbb{E}\|\Lambda_u\|^4} \cdot \sqrt{\mathbb{E}\|B_u\|^4 + \mathbb{E}\|\Lambda_u\|^4} = |I| \times O\left(\frac{b}{T} + \frac{1}{(Tb)^2} \right).$$

Now, if $I = [0, b] \cup [1 - b, 1]$, we have $\int_I \|R_u\|^2 du = O_{\mathbb{P}}(b)$ and we obtain from the previous bounds

$$\left| \int_0^1 \Lambda'_u \Gamma(u) R_u \left(B_u + \Lambda_u \right) du \right|^2 = O_{\mathbb{P}}\left(\frac{b^3}{T} + \frac{1}{T^2} \right) = o_{\mathbb{P}}\left(\frac{1}{T^2 b} \right).$$

Now, if $I = [b, 1 - b]$, we have, from Lemma 7, $\int_I \|R_u\|^2 du = O_{\mathbb{P}}\left(b^4 + \frac{1}{(Tb)^2} \right)$ and then

$$\left| \int_0^1 \Lambda'_u \Gamma(u) R_u \left(B_u + \Lambda_u \right) du \right|^2 = O_{\mathbb{P}}\left(\left(\frac{b}{T} + \frac{1}{(Tb)^2} \right) \cdot \left(b^4 + \frac{1}{(Tb)^2} \right) \right).$$

This is clearly $o_{\mathbb{P}}\left(\frac{1}{T^2 b} \right)$ under our bandwidth conditions. Then (H.8) follows.

8. Finally, setting $M_u = R_u \left(B_u + \Lambda_u \right)$, we show that

$$T \sqrt{b} \int_0^1 M'_u \Gamma(u) M_u du = o_{\mathbb{P}}(1). \tag{H.9}$$

We have

$$\int_I M'_u \Gamma(u) M_u du \leq \max_{u \in [0,1]} \|S_u^{-1}\|^2 \cdot \int_I \|B_u + \Lambda_u\|^2 \cdot \|\widetilde{R}_u\|^2 du.$$

Moreover,

$$\mathbb{E} \int_I \|B_u + \Lambda_u\|^2 \cdot \|\widetilde{R}_u\|^2 du$$

$$\leq \quad |I| \cdot \max_{u \in I} \sqrt{\mathbb{E}\|\widetilde{R}_u\|^4} \cdot \left(\sqrt{\mathbb{E}\|B_u\|^4} + \sqrt{\mathbb{E}\|\Lambda_u\|^4} \right)$$

$$\leq \quad |I| \cdot \max_{u \in I} \sqrt{\mathbb{E}\|\widetilde{R}_u\|^4} \cdot O\left(b^2 + \frac{1}{Tb} \right).$$

Considering the two cases for I, (H.9) follows from Lemma 7 and the bandwidth conditions.

The first part of the theorem is now complete. Now we prove that the asymptotic of our statistic remains the same if we replace β with the estimate $\hat{\beta}$ of Theorem 1. Since $\hat{\beta} - \beta = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$, we have

$$S_T\left(\widetilde{a}, \hat{\beta}\right) - S_T\left(\widetilde{a}, \beta\right) = -2I_T + O_{\mathbb{P}}\left(\frac{1}{T}\right),$$

where $I_T = \int_0^1 \left(\hat{\beta} - \beta\right)' \Gamma(u) \left(\overline{\beta}(u) - \beta\right) du$ and it remains to prove that

$$I_T = o_{\mathbb{P}}\left(\frac{1}{T\sqrt{b}}\right). \tag{H.10}$$

Using decomposition (H.1), we have already shown that

$$\int_0^1 \left(\overline{\beta}(u) - \beta - \widetilde{\kappa}_u \Lambda_u\right)' \Gamma(u) \left(\overline{\beta}(u) - \beta - \widetilde{\kappa}_u \Lambda_u\right) du = o_{\mathbb{P}}\left(\frac{1}{T\sqrt{b}}\right).$$

Using Cauchy-Schwarz inequality, it is easily seen that

$$\int_0^1 \left(\hat{\beta} - \beta\right)' \Gamma(u) \left(\overline{\beta}(u) - \beta - \widetilde{\kappa}_u \Lambda_u\right) du = o_{\mathbb{P}}\left(\frac{1}{T\sqrt{b}}\right).$$

Then to show (H.10), it remains to show $\int_0^1 \Gamma(u)\widetilde{\kappa}(u)\Lambda(u)du = o_{\mathbb{P}}\left(\frac{1}{\sqrt{Tb}}\right)$. We have

$$\begin{aligned}
&\int_0^1 \Gamma(u)\widetilde{\kappa}(u)\Lambda(u)du \\
= \; &\sum_{i=p+1}^{T} \int_0^1 \Gamma(u)\widetilde{\kappa}(u)e_i(u)du \cdot \mathcal{Y}_i Z_i \\
= \; &O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

The last equality follows after noticing that $\max_{p+1 \leq i \leq T} \| \int_0^1 \Gamma(u)\widetilde{\kappa}(u)e_i(u)du\| = O\left(\frac{1}{T}\right)$ and applying Lemma 2 componentwise. Then we get (H.10). The proof of Theorem 5 is now complete.□

I   Proof of Proposition 3

Setting $\hat{H}_t = X_t^2 - \hat{d}_t$ and $H_t = X_t^2 - \mathbb{E}X_t^2$ for $p + 1 \leq t \leq T$, we have

$$\hat{\beta} = \left(\hat{a}_1, \ldots, \hat{a}_p\right)' = \left(\sum_{i=p+1}^{T} \hat{\mathcal{H}}_t \hat{\mathcal{H}}_t'\right)^{-1} \sum_{t=p+1}^{T} \hat{H}_t \hat{\mathcal{H}}_t,$$

where $\hat{\mathcal{H}}_t = \left(\hat{H}_{t-1}, \ldots, \hat{H}_{t-p}\right)'$. We use the decomposition

$$\hat{H}_t = H_t + \mathbb{E}X_t - \sum_{i=p+1}^{T} k_{t,i}\mathbb{E}X_i^2 - \sum_{i=p+1}^{T} k_{t,i}H_i.$$

57

To prove the result, we will show that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} H_t \sum_{i=p+1}^{T} k_{t-s,i} H_i = o\mathbb{P}(1), \quad s \le t \tag{I.1}$$

and

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \sum_{i=p+1}^{T} k_{t,i} H_i \cdot \sum_{i=p+1}^{T} k_{t-s,i} H_i = o_{\mathbb{P}}(1), \quad s \le t. \tag{I.2}$$

Using (I.1) and (I.2) and some arguments given in the proof of Theorem 1, one can show that $\sqrt{T}\hat{\beta}$ has the same asymptotic distribution than the same estimator but with $\hat{H}$ replaced by H. Then one can deduce that $\sqrt{T}\hat{\beta}$ converges to a p−dimensional Gaussian vector with distribution $\mathcal{N}_p\left(0, \sigma^2 I_p\right)$ and the result of Proposition 3 easily follows. Let us prove (I.1) and (I.2).

1. For (I.1), we decompose the sum into two terms,

$$\sum_{t=p+1}^{T} H_t \sum_{i=p+1}^{T} k_{t-s,i} H_i = \sum_{t=p+1}^{T} k_{t-s,t} H_t^2 + \sum_{i \ne t} k_{t-s,i} H_i H_t.$$

The expectation of the (positive) first term is smaller to C/b where C is a positive constant. Under the assumptions of Proposition 3, we have $Tb^2 \to 0$. Then the latter expectation is $o\left(\sqrt{T}\right)$. The variance of the second term is

$$\sum_{i \ne t, i' \ne t'} k_{t-s,i} k_{t'-s,i'} \mathbb{E}\left(H_i H_t H_{i'} H_{t'}\right)$$

$$= \sum_{i \ne t} k_{t-s,i}^2 \mathbb{E}\left(H_i^2\right) \mathbb{E}\left(H_t^2\right) + \sum_{i \ne t} k_{t-s,i} k_{i-s,t} \mathbb{E}\left(H_i^2\right) \mathbb{E}\left(H_t^2\right).$$

It is easy to show that this variance is of order $O\left(b^{-1}\right) = o(T)$. This shows (I.1).

2. Next we show (I.2). The $H_i's$ are independent. One can use the results given in Zhang and Wu (2012), Lemma A1 and A3, from which we deduce that

$$\max_{p+1 \le t \le T} \left| \sum_{i=p+1}^{T} k_{t,i} H_i \right| = \frac{O_{\mathbb{P}}\left(T^{\frac{1}{2(1+\delta)}} + \sqrt{Tb \log(T)}\right)}{Tb}.$$

Using our bandwidth conditions, this entails (I.2).

The proof of Proposition 3 is now complete.□

J Asymptotic semiparametric efficiency

J.1 Proof of Proposition 2

We set $g_t = g(t/T)$ and $e_t = \frac{M_t' g_t + L_t' h}{\sigma_t^2}$. Then using the inequalities

$$0 \le 1 - \frac{1}{1+x} - x + x^2 \le x^3, \quad \log(1+x) - x + \frac{x^2}{2} \le x^3, \quad x \ge 0,$$

we have

$$\log \frac{d\mathbb{P}_{T,\alpha+\frac{g}{\sqrt{T}},\beta+\frac{h}{\sqrt{T}}}}{d\mathbb{P}_{T,\alpha,\beta}}\left(X_{p+1},\ldots,X_T\right) = \frac{1}{2}\sum_{t=p+1}^{T}\left\{\frac{X_t^2}{\sigma_t^2}\left[1-\frac{1}{1+\frac{e_t}{\sqrt{T}}}\right]-\log\left(1+\frac{e_t}{\sqrt{T}}\right)\right\}$$

$$= \Delta_{T,g,h} - \frac{1}{2T}\sum_{t=p+1}^{T}\left(\xi_t^2-\frac{1}{2}\right)e_t^2 + r_T,$$

with $|r_T| \le \frac{1}{2T\sqrt{T}}\sum_{t=p+1}^{T}\left(\xi_t^2+1\right)e_t^3$. Since $(e_t)_t$ is bounded, we have $r_T = o_{\mathbb{P}_{T,\alpha,\beta}}(1)$. Next, observe that e_t is Lipschitz in $X_{t-1}^2,\ldots,X_{t-p}^2$. Setting $Y_t = Z_t d_t^2 + \frac{e_t^2}{2}$, we have $\frac{1}{T}\sum_{t=p+1}^{T}\left(\xi_t^2-\frac{1}{2}\right)e_t^2 = \frac{1}{T}\sum_{t=p+1}^{T}Y_t$ and since $(Y_t)_t$ is a processes of type II, we get from Lemma 4,

$$\mathbb{E}|\frac{1}{T}\sum_{t=p+1}^{T}\bar{Y}_t|^2 = O\left(\frac{1}{T}\right).$$

This entails $\frac{1}{T}\sum_{t=p+1}^{T}\bar{Y}_t = o_{\mathbb{P}_{T,\alpha,\beta}}(1)$. Finally, noticing that e_t is a bounded and Lipschitz function in $\left(\frac{t}{T},X_{t-1}^2,\ldots,X_{t-p}^2\right)$, we deduce from Lemma 1,

$$\frac{1}{T}\sum_{t=p+1}^{T}\mathbb{E}\left(Y_t\right) - \frac{1}{2T}\sum_{t=p+1}^{T}\mathbb{E}\left(e_1\left(\frac{t}{T}\right)^2\right) = o(1),$$

where $e_t(u) = \frac{M_t(u)'g(u)+N_t(u)'h}{\sigma_t(u)^2}$. From Lemma 1, $u \mapsto \mathbb{E}\left(e_1(u)^2\right)$ is continuous. This leads to

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=p+1}^{T}\left(\xi_t^2-\frac{1}{2}\right)e_t^2 = \int_0^1 \mathbb{E}\left(e_1(u)^2\right)du = \|(g,h)\|_{\mathbb{H}}^2,$$

where the limit is in $\mathbb{P}_{T,\alpha,\beta}-$probability. This achieves the proof of Proposition 2. $\square$

J.2 Proof of Corollary 1

To prove the first assertion, it is enough to check the equality

$$< \dot{k}^*v, (g,h) >_{\mathbb{H}} = h'v, \tag{J.1}$$

for all $(g,h,v) \in H \times \mathbb{R}^n$. Using the notations $E(u) = \begin{pmatrix} E_1(u) & E_2(u) \\ E_2(u)' & E_3(u) \end{pmatrix}$, with

$$E_1(u) = \mathbb{E}\left(\frac{M_1(u)M_1(u)'}{\sigma_1(u)^4}\right), \quad E_2(u) = \mathbb{E}\left(\frac{M_1(u)N_1(u)'}{\sigma_1(u)^4}\right), \quad E_3(u) = \mathbb{E}\left(\frac{N_1(u)N_1(u)'}{\sigma_1(u)^4}\right),$$

we have $q_2^*(u) = E_1(u)^{-1}E_2(u)$ and it is easy to verify the equality

$$\Sigma = \int_0^1 \left[E_3(u) - E_2(u)'E_1(u)E_2(u)\right]du.$$

Then, it is easy to get (J.1) using the expression of the scalar product on H. For the second assertion, we apply Theorem 3.11.2 of van der Vaart and Wellner (1996), using the equality $\|\dot{k}^*v\|_{\mathbb{H}} = 2v'\Sigma^{-1}v$. $\square$

K   Numerical experiments for inference/testing

K.1   Example of semiparametric estimation.

We first illustrate the methods of parameters inference in the semiparametric model with constant lag co-efficients. We consider the noise distributions $\xi_1 \sim \mathcal{N}(0,1)$, $t(9)$ (Student distribution with 9 degrees of freedom) and $t(5)$. These three distributions satisfy the moment assumption $\mathbb{E}\xi_0^{4(1+\delta)} < \infty$ used in Theorem 1, Theorem 3 and Theorem 4. The number of lags is fixed to $p = 2$ and the intercept function is defined by $a_0(u) = 2 + \sin(2\pi u)$. We compare the estimates obtained using the procedure described in the paper and the plug-in estimates which are asymptotic optimal. Two sample sizes are considered: $T = 500$ and $T = 1500$. Only one bandwidth is used and selected by the CV procedure (the same bandwidth is used for estimating the intercept function and plug-in estimates are also computed using this initial bandwidth). Note that the $t-$distributions do not satisfy the moment assumption for the asymptotic normality of the plug-in estimator of lag coefficients (in Theorem 2, we assumed that ξ has moments of any order but our assumption is probably not optimal). The plug-in estimator seems to have a smaller RMSE (see Table 4), even when $T = 500$. Observe also that our estimates are less accurate when the noise has fatter tails. The RMSE for $\hat{a}_0$ is defined by $\sqrt{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\hat{a}_0(t/T) - a_0(t/T)\right)^2}$.

Table 4: RMSE for parameter estimation (notation $*$ is for the plug-in estimator)

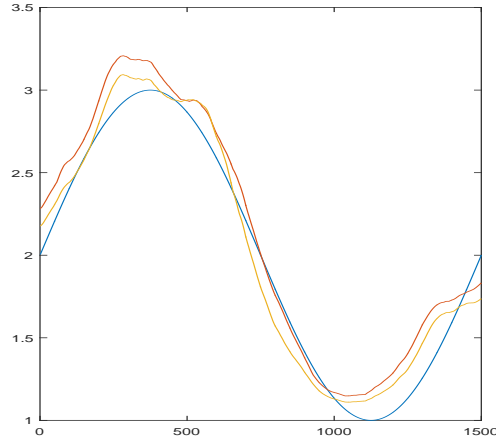|	$\xi_0 \sim \mathcal{N}(0,1)$			$\xi_0 \sim t(9)$			$\xi_0 \sim t(5)$		
	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$
$T = 500$	0.5446	0.0859	0.0769	0.6380	0.1104	0.1000	0.7501	0.1732	0.1430
	$\hat{a}_{0,*}$	$\hat{a_{1,*}}$	$\hat{a}_{2,*}$	$\hat{a}_{0,*}$	$\hat{a_{1,*}}$	$\hat{a}_{2,*}$	$\hat{a}_{0,*}$	$\hat{a_{1,*}}$	$\hat{a}_{2,*}$
	0.5068	0.0750	0.0651	0.5606	0.0949	0.0822	0.6489	0.1619	0.1167
$T = 1500$	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$
	0.3335	0.0473	0.0440	0.3844	0.0615	0.0557	0.5181	0.1012	0.0963
	$\hat{a}_{0,*}$	$\hat{a_{1,*}}$	$\hat{a}_{2,*}$	$\hat{a}_{0,*}$	$\hat{a_{1,*}}$	$\hat{a}_{2,*}$	$\hat{a}_{0,*}$	$\hat{a_{1,*}}$	$\hat{a}_{2,*}$
	0.3192	0.0433	0.0385	0.3571	0.0536	0.0471	0.4365	0.0775	0.0727

Figure 5: Estimation of $a_0(u) = 2 + sin(2\pi u)$ when $\xi_0 \sim \mathcal{N}(0, 1)$, $a_1 = 0.3$, $a_2 = 0.2$ and $T = 1500$ (the red curve is the initial estimate and the yellow curve is the plug-in estimate)

K.2 Testing the constancy of coefficients in a tv(1) process.

We consider here a tv-ARCH model with $p = 1$ and $\xi_0 \sim \mathcal{N}(0, 1)$ or $\xi_0 \sim t(9)$. We consider two setups. In Setup 1, we have $a_0(u) = 2 + sin(2\pi u)$ and $a_1(u) = 0.5$. In Setup 2, we have $a_0(u) = 1$ and $a_1(u) = 0.5 + 0.25 \times \cos(2\pi u)$. Considering two levels $\alpha = 10\%$ and $\alpha = 5\%$, we approximate the probability of rejecting H_0: a_0 constant or H_0: a_1 constant. Results are reported in Table 5. Under H_0, this probability has to be close to the level α of the test. One can observe that using a $t(9)-$ distribution for the noise does not create size distortion. However, under the alternative H_1, the t distribution entails a smaller power than for the standard Gaussian. This suggests that the power of our tests is impacted by a fat tail noise, which is not surprising. Reasonable powers are obtained when $T = 2000$, the order of the sample size used in our real data applications.

Table 5: Approximation of the power for testing parameter constancy in tv(1) processes

		Setup 1				Setup 2			
		$T = 1000$		$T = 2000$		$T = 1000$		$T = 2000$	
		a_0	a_1	a_0	a_1	a_0	a_1	a_0	a_1
$\xi_0 \sim \mathcal{N}(0, 1)$	$\alpha = 5\%$	0.99	0.07	1	0.07	0.08	0.54	0.07	0.91
	$\alpha = 10\%$	0.99	0.12	1	0.13	0.13	0.68	0.12	0.96
$\xi_0 \sim t(9)$	$\alpha = 5\%$	0.97	0.07	1	0.06	0.06	0.34	0.06	0.69
	$\alpha = 10\%$	0.98	0.13	1	0.12	0.13	0.47	0.11	0.8

A comparison with the Gaussian quantiles. For $T = 500$ and $p = 1$, we consider the setup 1. When $\alpha = 10\%$ and $b = 0.01 \times \ell$, $1 \le \ell \le 30$, we compare the coverage probabilities obtained using the Monte Carlo method with the coverage probabilities using the Gaussian quantiles when $\alpha = 10\%$ and for testing

61

the constancy of the first lag coefficient. In Figure 6, one can see that the Monte Carlo method is interesting because the coverage probabilities seem more precise and less sensitive to the bandwidth parameter if we exclude very small bandwidths.
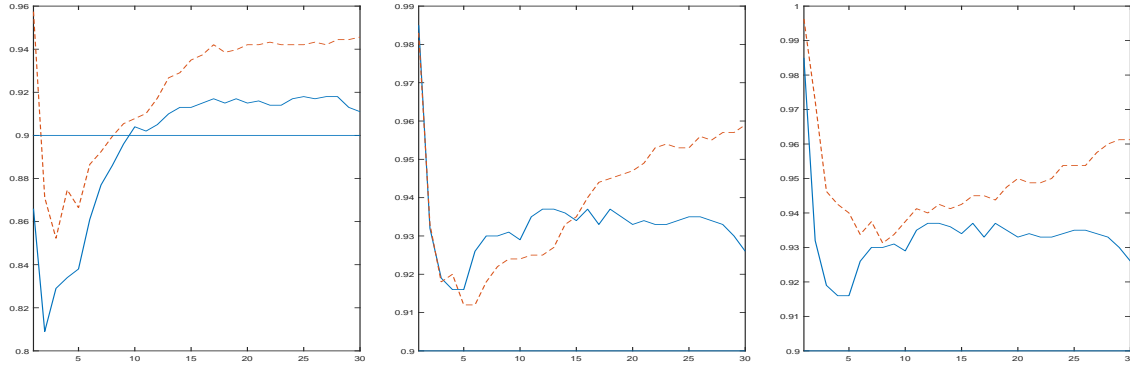

Figure 6: Coverage probabilities for Gaussian inputs (left), $t(9)$ inputs (middle) and the difference of two independent random variables following an exponential distribution with parameter 1 (right). Dashed lines represent the coverage probabilities obtained with the Gaussian quantiles.

K.3 Power curves for testing non time-varying coefficients in a tv(2) process

In this subsection, we simulate approximation of the power for testing $H_0 : a_0$ constant (resp. a_1 constant, a_2 constant, $(a_1, a2)$ constant) when $a_0(u) = 2(1 + \theta \sin(2\pi u))$, $a_1(u) = 0.2 + \frac{\theta}{2} \sin(2\pi u)$, $a_2(u) = 0.2 + \frac{\theta}{2} \cos(2\pi u)$ with $0 \leq \theta \leq 0.45$. The noise distribution will be either Gaussian or a student distribution with 9 degrees of freedom (we remind that Theorem 5 is only valid when $\mathbb{E}\xi_0^{8(1+\delta)} < \infty$). Figure 7 represents an approximation of the power curves when $T = 2500$ and $\alpha = 10\%$. One can observe that a more fat tail for the noise leads to a slightly smaller power for our tests.
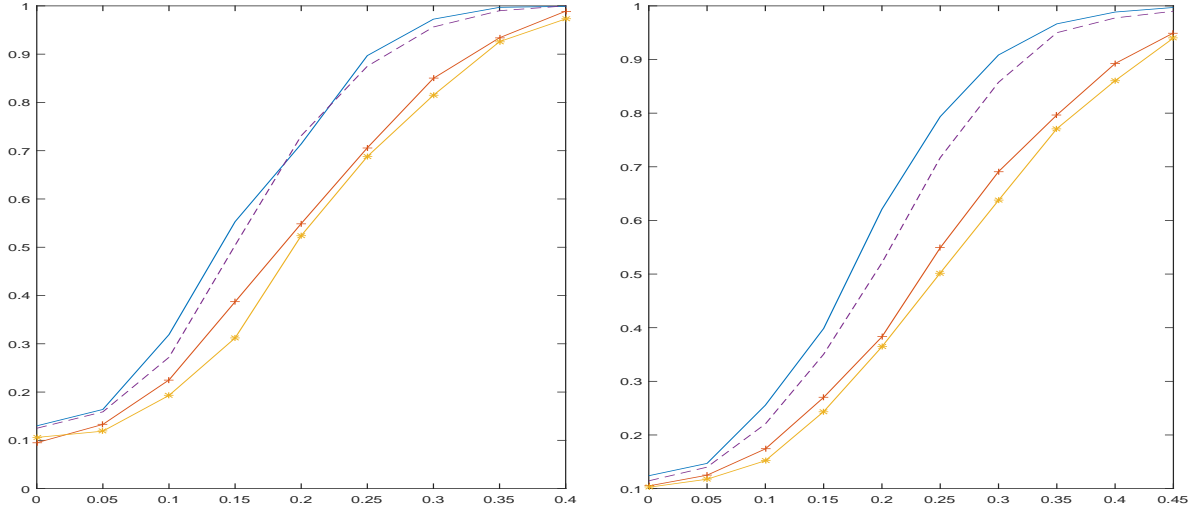
Figure 7: Power curves when the noise is Gaussian (on the left) or follows a $t(9)-$distribution (on the right). The legend for the curves is: $-$ for a_0 constant, $--$ for (a_1, a_2) constant, $+$ for a_1 constant and $*$ for a_2 constant.

K.4 Testing the second order dynamic in the semiparametric model

Here we assume that $\beta = (a_1, \ldots, a_p)'$. The null hypothesis is $\beta = 0$. We use the procedure described in the paper after choosing the bandwidth parameter b by cross-validation. We restrict our study to the case $p = 2$. For the simulation setup, we consider two scenarios. In setup 1, we consider a constant intercept $a_0(u) = 10^{-4}$. In setup 2, a_0 is a piecewise affine function such that $a_0(0) = a_0(0.5) = a_0(1) = 10^{-4}$ and $a_0(0.25) = a_0(0.75) = 4 \cdot 10^{-4}$. The noise distribution will be Gaussian, $t(9)$, or $t(5)$. We also consider two sample sizes: $T = 500$ and $T = 1000$. Table 6 and Table 7 provide approximations of the coverage probabilities. In Figure 8, approximations of some power curves are given under the alternative $a_1 = a_1 = \theta \times 0.02$, with $\theta = 0, \ldots, 6$. The results seem satisfying for the three noise distributions.

Table 6: Approximation of the coverage probabilities when $T = 500$

Setup 1	$\alpha = 10\%$	$\alpha = 5\%$
$\xi_0 \sim \mathcal{N}(0, 1)$	0.92	0.95
$\xi_0 \sim t(9)$	0.92	0.95
$\xi_0 \sim t(5)$	0.91	0.94

Setup 2	$\alpha = 10\%$	$\alpha = 5\%$
$\xi_0 \sim \mathcal{N}(0, 1)$	0.93	0.97
$\xi_0 \sim t(9)$	0.92	0.95
$\xi_0 \sim t(5)$	0.91	0.94

Table 7: Approximation of the coverage probabilities when $T = 1000$

Setup 1	$\alpha = 10\%$	$\alpha = 5\%$
$\xi_0 \sim \mathcal{N}(0, 1)$	0.90	0.96
$\xi_0 \sim t(9)$	0.90	0.94
$\xi_0 \sim t(5)$	0.90	0.93

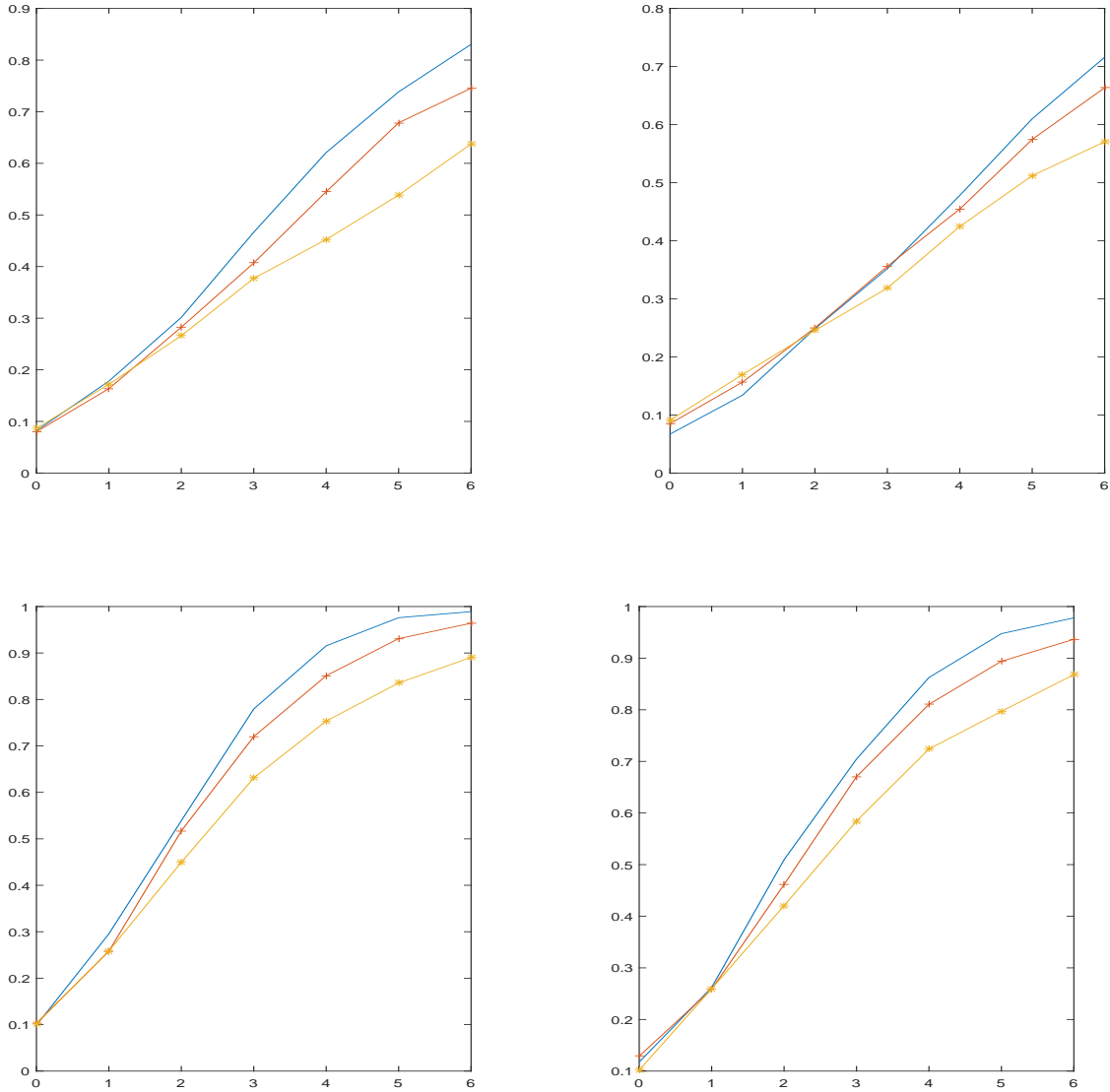| Setup 2	$\alpha = 10\%$	$\alpha = 5\%$
$\xi_0 \sim \mathcal{N}(0, 1)$	0.88	0.94
$\xi_0 \sim t(9)$	0.87	0.93
$\xi_0 \sim t(5)$	0.90	0.93

Figure 8: Power curves for testing the second order dynamic. Setup 1 with $T = 500$ (top left), setup 2 with $T = 500$ (top right), setup 1 with $T = 1000$ (bottom left), setup 2 with $T = 1000$ (bottom right).

K.5 Information criterion for the number of lags in tv-ARCH processes

In this subsection, we study numerically the performance of the information criterion used for selecting the number of lags in time-varying ARCH processes. We first consider the case $p = 1$, with $a_0(u) = 2(1 + 0.4 \sin(2\pi u))$ and $a_1(u) = 0.3$. Two distributions are considered for the noise: the standard Gaussian and the $t(9)$ distribution. In Table 8, we simulate $B = 2000$ models for both noise distributions and three sample sizes: $T = 500$, $T = 1000$ and $T = 2000$. The results are correct for large sample sizes. As for the estimation, one observe that the performance of the criterion is sensitive to the tail of the noise distribution.

Table 8: Percentages of models correctly fitted (CF), underfitted (UF) and overfitted (OF) for Setup 1

		CF	UF	OF
$\xi_0 \sim \mathcal{N}(0,1)$	$T = 500$	88	6	6
	$T = 1000$	97	2	1
	$T = 2000$	99	0	1
$\xi_0 \sim t(9)$	$T = 500$	80	14	6
	$T = 1000$	90	8	2
	$T = 2000$	94	5	1

In a second simulation setup, we consider the case $p = 0, 1, 2$, using the same intercept a_0 and setting $a_1(u) = 0.2 + 0.2 \cdot \sin(2\pi u)$, $a_2(u) = 0.2 + 0.2 \cdot \cos(2\pi u)$. In this case, the lag coefficients can be arbitrary close to zero and the true model more difficult to select. Numerical experiments are reported in Table 9. When $\xi_0 \sim t(9)$, large sample sizes are necessary to obtain good results. Once again, one can explain this behavior by the difficulty of getting accurate estimates with such noise distribution tail when the sample size is not large enough.

Table 9: Percentages of correctly fitted, underfitted and overfitted models for Setup 2

		$p = 0$			$p = 1$			$p = 2$		
		CF	UF	OF	CF	UF	OF	CF	UF	OF
$\xi_0 \sim \mathcal{N}(0,1)$	$T = 500$	93	0	7	78	16	6	74	22	4
	$T = 1000$	93	0	7	92	6	2	91	7	2
	$T = 2000$	96	0	4	99	0	1	99	1	0
$\xi_0 \sim t(9)$	$T = 500$	91	0	9	66	28	6	58	37	5
	$T = 1000$	95	0	5	83	14	3	78	20	2
	$T = 2000$	96	0	4	94	5	1	95	4	1