Les Cahiers du GERAD ISSN: 0711–2440

Validity of the Parametric Bootstrap for Goodness-of-Fit Testing in Dynamic Models

B. Rémillard

G–2011–69

November 2011

Validity of the Parametric Bootstrap for Goodness-of-Fit Testing in Dynamic Models

Bruno Rémillard

GERAD & Department of Management Sciences
HEC Montréal
3000 chemin de la Côte Sainte-Catherine
Montréal (Québec) Canada, H3T 2A7

bruno.remillard@hec.ca

November 2011

Les Cahiers du GERAD

G–2011–69

Abstract

It is shown that parametric bootstrap can be used for computing P-values of goodness-of-fit tests of multivariate time series parametric models. These models include Markovian models, GARCH models with non-Gaussian innovations, regime-switching models, as well as semiparametric models involving copulas of multivariate time series. The methodology is intuitive, easy to implement, and provides an interesting alternative to Khmaladze's transform or other projection methods.

Key Words: Time series; Goodness-of-fit test; Monte Carlo simulation; Parametric bootstrap; P-values; HMM; GARCH; Copulas.

Résumé

On montre dans cet article que le rééchantillonage paramétrique peut être appliqué pour le calcul de probabilités critiques de tests d'adéquation pour des modèles de séries chronologiques multivariées. Ces modèles incluent les modèles markoviens, les modèles GARCH avec des innovations non gaussiennes, les modèles à changement de régimes ainsi que les modèles semi-paramétriques pour des copules de séries chronologiques multidimensionnelles. La méthode proposées est intuitive, facile à implanter, et est une alternative intéressante à la méthode de transformation de Khmaladze ou tout autre méthode de projection.

Acknowledgments: Funding in partial support of this work was provided by the Natural Sciences and Engineering Research Council of Canada, the Fonds québécois de la recherche sur la nature et les technologies, and the Institut de finance mathématique de Montréal. Special thanks to the people of the Department of Finance at Singapore Management University for their hospitality.

1 Introduction

In many financial applications, such as hedging, option pricing, hedge fund replication, risk management and credit risk, it is very important to model correctly multivariate data, including its distribution over time and serial dependence, if any. Sometimes one is interested in modeling the full distribution, hereby referred to as parametric modeling, or one can be interested in modeling the serial dependence and/or interdependence, referred to as semiparametric modeling. In the latter case, only the dependence structure is modeled by a parametric family. In general, tests of goodness-of-fit are based on statistics expressed as functions of empirical processes, and the limiting distribution of these processes always depend on unknown parameters, making it difficult to calculate P-values. This problem can be solved by using projection techniques like Khmaladze's transform, which however, changes completely the original process. Tests based on likelihood ratios are not covered in the paper, since they are not really tests of goodness-of-fit, and they can create problems, specially in Hidden Markov models.

Here it is proposed to use a simpler intuitive technique, called parametric bootstrap, to approximate P-values. That methodology has been shown to work, both in parametric and semiparametric settings, when there is no serial dependence. The aim here is to extend its applicability to dynamic models, i.e., models including serial dependence. Simply stated, if a goodness-of-fit test is based on a statistic S_n of the observations $Y_1, \ldots, Y_n$ with distribution P_θ, for some unknown parameter θ estimated by θ_n, the parametric bootstrap approach consists in generating a large number N of sequences $Y^*_{1,k}, \ldots, Y^*_{n,k}$, $k = 1, \ldots, N$ with distribution P_{θ_n}, evaluating each time the goodness-of-fit statistic $S^*_{n,k}$, and then approximating the p-value by the percentage of values $S^*_{n,k}$ greater than S_n, assuming that the null hypothesis is rejected for large values of S_n. Basically, parametric bootstrap is a valid way to compute P-values if one can show that $(S_n, S^*_{n,1}, \ldots, S^*_{n,N})$ converges in law to $(S, S^*_1, \ldots, S^*_N)$, where $S^*_1, \ldots, S^*_N$ are independent copies of S.

After reviewing the literature on goodness-of-fit tests for dynamic models in Section 2, one proves the main result behind the validity of parametric bootstrap in Section 3, provided the estimators of the parameters involved in the model are regular in some sense. Examples of applications for parametric and semiparametric settings are then given. First, in Section 4, one states general results for the parametric setting, where the null hypothesis to be tested takes the form

H_0: The conditional distribution of Y_t given $\mathcal{F}_{t-1}$ belongs to the parametric family $\{F_{t,\theta}; \theta \in \mathcal{O}\}$.

The asymptotic limit of the empirical processes based on the Rosenblatt transforms is also identified, extending results of Bai (2003) and Bai and Chen (2008). In particular, it is shown that parametric bootstrap can be used in p-Markov, ARMA, GARCH and regime-switching models. Examples of regular estimators are given in each case. Then, in Section 5, one considers goodness-of-fit in semiparametric settings, i.e., for copula models. First, it is shown that parametric bootstrap works for copulas associated with multivariate Markov processes. Here the null hypothesis can be stated as follows:

H_0: The copula C of $(Y_{t-p}, \ldots, Y_t)$ belongs to the parametric family $\mathcal{C} = \{C_\phi; \phi \in \mathcal{P}\}$,

where $(Y_{t-p+1}, \ldots, Y_t)$ is assumed to be a Markov process, which is often referred to as Y being p-Markov, and C is then the associated copula. Here, no structure is imposed of the marginal distributions of the random vectors Y_t. These models have been recently proposed in order to deal with serial dependence and interdependence, i.e., dependence between the components of Y_t (Rémillard et al., 2011).

Another case of semiparametric modeling often considered in applications is the modeling of the joint distribution of innovations $\varepsilon_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,d})$ coming from several univariate time series models that are estimated separately (van den Goorbergh et al., 2005, Patton, 2006, Chen and Fan, 2006a). In that case, the null hypothesis can be stated as H_0: The copula C of ε_t belongs to $\mathcal{C} = \{C_\phi; \phi \in \mathcal{P}\}$. That case has been treated recently (Rémillard, 2010), so it is not covered here. Finally, in Section 6, one gives an example of application for GARCH models with innovations having a generalized error distribution (GED).

2 Review of the literature

From now on, one will concentrate on tests of goodness-of-fit based on the Rosenblatt's transform (Rosenblatt, 1952), which is a mapping (or a series of mappings) so that the output is a sequence of independent and identically distributed (i.i.d.) random vectors $U_1, \ldots, U_n$ with uniform distribution over $[0,1]^d$, denoted by $U_t \sim C_\perp$, where $C_\perp$ is the usual notation for the so-called independence copula defined by $C_\perp(u) = u_1 \times \cdots \times u_d$, $u = (u_1, \ldots, u_d) \in [0,1]^d$. The main reason for ignoring other methods is that, based on recent results of Genest et al. (2009), tests using Rosenblatt's transform seem to be more powerful, at least in the multivariate case. In the univariate case, they are equivalent in general to the tests based on the usual empirical distribution function. Furthermore, these tests are also computationally simpler in general. The idea of using Rosenblatt's transform for testing goodness-of-fit is not new, the main contributions being Durbin (1973) in the i.i.d. case, and Diebold et al. (1998) in time series models, both in univariate cases.

2.1 I.I.D. observations

For the first part of the literature review, suppose that the observations are i.i.d. To test the null hypothesis that the (univariate) observations $Y_1, \ldots, Y_n$ had distribution function F belonging to $\{F_\theta; \theta \in \mathcal{O}\}$, Durbin (1973) proposed to base the test on the empirical distribution function

$$D_n(u) = n^{-1} \sum_{t=1}^{n} \mathbf{1}\{F_{\theta_n}(Y_t) \le u\}, \quad u \in [0,1].$$

His reasoning was that if θ_n is a good estimator of the unknown parameter θ, then under the null hypothesis, the pseudo-observations $u_{n,t} = F_{\theta_n}(Y_t)$ should be close to the (not observable) random variables $U_t = F_\theta(Y_t)$, the latter forming an i.i.d. sequence uniformly distributed over $[0,1]$. Hence, $D_n(u)$ should be close to u, for any $u \in [0,1]$, under the null hypothesis. Assuming additional conditions on θ_n, in particular the convergence in law of $\Theta_n = n^{1/2}(\theta_n - \theta)$ to Θ, together with differentiability conditions on F, he showed that $\mathbb{D}_n(u) = n^{1/2}\{D_n(u) - u\} \rightsquigarrow \mathbb{D}(u) = \mathbb{B}(u) - \Theta^\top \gamma(u)$, where $\rightsquigarrow$ denotes convergence in the Skorohod space $D([0,1])$ of càdlàg functions, and $\mathbb{B}_n \rightsquigarrow \mathbb{B}$, where $\mathbb{B}$ is the usual Brownian bridge process arising as the limiting distribution of the sequence of empirical processes,

$$\mathbb{B}_n(u) = n^{-1/2} \sum_{t=1}^{n} \{\mathbf{1}(U_t \le u) - u\},$$

and $\gamma(u) = \dot{F}_\theta \circ F_\theta^{-1}(u)$, with $\dot{F}_\theta$ being the (column vector) gradient of F_θ with respect to θ. Except for natural location/scale parametric families, like the Gaussian or the exponential families, the distribution of $\Theta^\top \gamma$ depend on the unknown value θ, making it impossible to tabulate statistics based on $\mathbb{D} = \mathbb{B} - \Theta^\top \gamma$.

To overcome the difficulty, there exist some options: Transform the process $\mathbb{D}_n$ so it becomes asymptotically distribution free, or bootstrapp it. One of the first transformation technique is Khmaladze's martingale technique (Khmaladze, 1988, 1993). It is relatively easy to implement for univariate data, but it seems that the level of the test can be imprecise, if the sample size is not quite large. See Bai (2003) for an easy introduction to the technique, even in the context of univariate time series. However, Khmaladze's martingale transform might be difficult to evaluate for semiparametric test statistics and/or imprecise when applied to multivariate data. Some also might not like the idea of working with a transformed process, when the interpretation of the process is not obvious, even if it converges to a Brownian motion. Other examples of the use of that technique is Delgado and Stute (2008), who used it for very special cases of tests on bivariate data. Finally, Li (2009) extended Khmaladze's work to more general projection methods, in view of applications in semiparametric regression settings. So far, no power comparison study has been attempted to answer the question of efficiency between transformation techniques and bootstrapping techniques. This will be done in a forthcoming paper.

By bootstrapping a statistic or more generally a stochastic process $\mathbb{A}_n$, one means a method for generating a process $\tilde{\mathbb{A}}_n$ so that $(\mathbb{A}_n, \tilde{\mathbb{A}}_n) \rightsquigarrow (\mathbb{A}, \tilde{\mathbb{A}})$, where $\tilde{\mathbb{A}}$ is an independent copy of $\mathbb{A}$. Repeating the bootstrapping process N times, one can then approximate P-values of statistics based on $\mathbb{A}_n$. The most known technique of bootstrapping is of course the resampling method (Efron, 1979). Going back to Durbin's process $\mathbb{D}_n$, as a rule of thumb, if one can bootstrap Θ_n, one can also bootstrap $\mathbb{D}_n$. Resampling bootstrap should also work for multivariate data. However, one will not pursue the matter since it might not be applicable for dynamic models. Another form of bootstrapping, called parametric bootstrap, appeared in Efron (1979). The first results about its validity were stated in Beran et al. (1987), Beran and Millar (1987, 1989). Stute et al. (1993) proved its validity for multivariate goodness-of-fit tests based on the empirical distribution function, while Andrews (1997) extended it by incorporating covariates. Finally, Genest and Rémillard (2008) addressed its validity for a wide range of goodness-of-fit tests in semiparametric models, including dependence modeling.

There is another possible bootstrapping technique that could be used, sometimes called weighted bootstrap or multipliers technique. So far it has been used for observations, not pseudo-observations, except in copula contexts (Scaillet, 2005, Rémillard and Scaillet, 2009, Kojadinovic and Yan, 2010). However, its validity and applicability, is beyond the scope of the present paper. For a complete review of that methodology, see, e.g., Rémillard (2011). Finally, note that a technique close to parametric bootstrap, called Maximized Monte Carlo (MMC), has been proposed by Dufour (2006). However, his technique cannot be considered as boostrapping. It seems however that the results proved here could be used to generalized the MMC approach to dynamic models.

2.2 Dynamic models

The idea of using a conditional analog of the Durbin process in a time series context goes back to Diebold et al. (1998). In such settings, one also ends up with limiting processes of the form $\mathbb{D} = \mathbb{B} - \Theta^\top \gamma$, though γ is more complex in general than in the i.i.d. case. For testing purposes, they proposed to divide the sample into two parts, one for estimation and one for testing. However, this sort of approach should be avoided as much as possible, since the samples sizes need to be very large in general and since they are alternative methods than can work much better. As shown in Bai (2003), Khmaladze's transform can also be used in univariate time series. In his paper, he also proved general results on the convergence of empirical process used for parametric goodness-of-fit of dynamic models. In Bai and Chen (2008), the authors studied goodness-of-fit for multivariate GARCH, transforming the pseudo-observations into a univariate series and then using Khmaladze's transform. Although correct, that methodology might lack power. It also shows that Khmaladze's transform is not that easy to implement for multivariate data. Note that even for univariate time series, Khmaladze's approach ca be difficult to implement when one has to estimate parameters for the distribution of the innovations, e.g., for GARCH models with GED or Student distribution for the innovations

For some special cases of dynamic models, Li and Tkacz (2006) based their test on a distance between density estimates and then used parametric bootstrap. Their proof is not general enough to be adapted to general settings considered by Bai (2003), a fortiori to semiparametric settings. Finally, in a univariate time series context similar to the one in Bai (2003), Corradi and Swanson (2006) proposed Kolmogorov-Smirnov type tests for goodness-of-fit and used block bootstrap to approximate P-values. The reason why Corradi and Swanson (2006) did not use parametric bootstrap is that instead of computing conditional expectations with respect to a filtration, they considered smaller sigma-algebras, calling that "dynamic misspecification". Here is an example: For an $AR(1)$ model, one could want to test that Y_t is Gaussian, based on $e_{n,t} = \Phi\{(Y_t - \mu_n)/s_n\}$, where μ_n and s_n are respectively the mean and the standard deviation of the series $Y_1, \ldots, Y_n$. The empirical process based on the pseudo-observations $e_{n,1}, \ldots, e_{n,n}$ would then converge to a quite complex Gaussian process. For the applications mentioned at the beginning of the introduction, no misspecification is allowed, so block bootstrap is not needed.

3 Validity of the parametric bootstrap

The goal of this section is to find sufficient conditions for proving that the parametric bootstrap procedure work. Given a sample $Y_1, \ldots, Y_n$ from a law $P = P_\theta$, with θ unknown and estimated by θ_n, and a statistic $S_n = \psi_n(Y_1, \ldots, Y_n)$ needed to be bootstrapped, the parametric bootstrap procedure based on the estimator θ_n of θ can be described as follows:

> For $k = 1, \ldots, N$, generate a sample $(Y^*_{k,1}, \ldots, Y^*_{k,n})$ from law P_{θ_n} and compute $S^*_{n,k} = \psi_n(Y^*_{k,1}, \ldots, Y^*_{k,n})$.

The parametric bootstrap work if it can be shown that as $n \to \infty$, $(S_n, S^*_{n,1}, \ldots, S^*_{n,N})$ converges jointly to $(S, S^*_1, \ldots, S^*_N)$, where all variables are independent and identically distributed. In other words, $S^*_1, \ldots, S^*_N$ are independent copies of S and, for example, $P(S \geq c)$ can be estimated consistently by $N^{-1}\sum_{k=1}^N \mathbf{1}(S^*_{n,k} > c)$. In particular, assuming that large values of S_n lead to the rejection of H_0, an approximate P-value for the test based on S_n is given by $N^{-1}\sum_{k=1}^N \mathbf{1}(S^*_{n,k} > S_n)$. Note that the size of N has generally an impact on the power of the test.

To describe the conditions for the validity of the parametric bootstrap, set

$$\ell_n(Y_1, \ldots, Y_n, \theta) = \log\left\{ \left.\frac{dP_\theta}{dP}\right|_{\mathcal{F}_n} \right\}.$$

Suppose that uniformly for all a in a compact subset of $\mathbb{R}^s$,

$$\ell_n(Y_1, \ldots, Y_n, \theta + n^{-1/2}a) = a^\top \mathbb{W}_n - \frac{1}{2} a^\top \mathcal{J} a + o_P(1), \quad \text{as } n \to \infty, \tag{1}$$

where $\mathbb{W}_n$ is a statistic of $Y_1, \ldots, Y_n$, and $\mathbb{W}_n \rightsquigarrow \mathbb{W}$, with $\mathbb{W} \sim N(0, \mathcal{J})$. It appears that in goodness-of-fit testing, $\mathbb{W}_n$ plays an essential role, as shown next. The main result for the validity of the parametric bootstrap procedure can now be stated as follows.

Theorem 1 *Suppose that $\mathbb{A}_n$ constructed from $Y_1, \ldots, Y_n$ has values in the Skorohod space $\mathcal{D}(T, \mathbb{R}^m)$, for some closed interval T of $[-\infty, \infty]^p$, and assume that (1) holds true. Further assume that $(\mathbb{W}_n, \Theta_n, \mathbb{A}_n) \rightsquigarrow (\mathbb{W}, \Theta, \mathbb{A})$, the joint law being centered Gaussian, with $a(t) = \mathrm{E}\left\{\mathbb{A}(t)\mathbb{W}^\top\right\}$ for all $t \in T$ and $\Gamma = \mathrm{E}\left(\Theta\mathbb{W}^\top\right)$. Let θ^*_n and $\mathbb{A}^*_n$ be the bootstrap analogs of θ_n and $\mathbb{A}_n$, and set $\Theta^*_n = n^{1/2}(\theta^*_n - \theta)$. Then*

$$(\mathbb{W}_n, \Theta_n, \mathbb{A}_n, \Theta^*_n, \mathbb{A}^*_n) \rightsquigarrow (\mathbb{W}, \Theta, \mathbb{A}, \Theta^*, \mathbb{A}^*),$$

where $\Theta^* = \tilde{\Theta} + \Gamma\Theta$, $\mathbb{A}^* = \tilde{\mathbb{A}} + a\Theta$, and $(\tilde{\Theta}, \tilde{\mathbb{A}})$ is an independent copy of $(\Theta, \mathbb{A})$. In particular, the parametric bootstrap works for $\mathbb{A}_n$ if and only if $\mathbb{A}$ is independent of $\mathbb{W}$.

For the proof, see Appendix A.1.

Example 1 *In many goodness-of-fit tests, the statistics are based on a process $\mathbb{D}_n$ that satisfies $(\mathbb{W}_n, \Theta_n, \mathbb{B}_n, \mathbb{D}_n) \rightsquigarrow (\mathbb{W}, \Theta, \mathbb{B}, \mathbb{D})$, with $\mathbb{D} = \mathbb{B} - \Theta^\top\gamma$, and $\gamma(t) = \mathrm{E}\{\mathbb{B}(t)\mathbb{W}\}$, $t \in T$. Then, according to Theorem 1, the parametric bootstrap will work if $\mathbb{D}$ is independent of $\mathbb{W}$, which in turn is equivalent to $(I - \Gamma)^\top\gamma(t) = 0$ for all $t \in T$. The latter is obviously satisfied if $\Gamma = I$.*

The previous example motivates the following definition, which appeared in Genest and Rémillard (2008) in a serially independent context.

Definition 1 *Suppose that $\theta_n = T_n(Y_1, \ldots, Y_n)$ is an estimator of $\theta \in \mathcal{O} \subset \mathbb{R}^s$ and set $\Theta_n = n^{1/2}(\theta_n - \theta)$. Then θ_n is called regular if $(\mathbb{W}_n, \Theta_n) \rightsquigarrow (\mathbb{W}, \Theta) \sim N(0, \Sigma)$, where $\Sigma = \begin{pmatrix} \mathcal{J} & I \\ I & V \end{pmatrix}$, and I is the identity matrix.*

It follows that the parametric bootstrap procedure works for $\mathbb{D}_n$, as described in Example 1, provided θ_n is regular.

Remark 1 *According to Theorem 1, θ_n is regular if and only if $n^{1/2}(\theta_n^* - \theta_n) \rightsquigarrow \tilde{\Theta}$, an independent copy of Θ. It was shown in Genest and Rémillard (2008) that the usual estimators (MLE, moment matching, minimum distance) calculated from i.i.d. observations were regular. A priori, it might seem strange that in all previous papers proving a version of the validity of the parametric bootstrap, that "regularity" of θ_n was not assumed. Well, it was implicitly assumed. For example, in Andrews (1997), condition E2 states that if θ_n is any non random sequence converging to θ, then one must have $n^{1/2}(\theta_n^* - \theta_n) \rightsquigarrow \Theta$. On the other hand, choosing $\theta_n = \theta + n^{-1/2}\gamma$, it is easy to check that $(\ell_n(\theta_n), \Theta_n) \rightsquigarrow (\gamma^\top \mathbb{W} - \frac{1}{2}\gamma^\top \mathcal{J}\gamma, \Theta)$. As in the proof of Theorem 1, one can use Le Cam's Third Lemma to obtain that $n^{1/2}(\theta_n^* - \theta_n) \rightsquigarrow \Theta + (\Gamma - I)\gamma$. Hence condition E2 of Andrews (1997), also similar to condition (5.4) in Beran et al. (1987), implicitly assumed that $\Gamma = I$, and it is in fact a stronger assumption which is much more difficult to verify than the regularity condition appearing in Definition 1.*

Remark 2 *Instead of studying convergence on the Skorohod space $\mathcal{D}(T;\mathbb{R}^m)$, one may also consider convergence on $\ell_\infty(\mathcal{A})$, over some class of functions $\mathcal{A}$; for more details, one may consult van der Vaart and Wellner (1996). The conclusions of Theorem 1 still hold in that case.*

We are now in a position to study the validity of parametric bootstrap for dynamic models, extending the results of Genest and Rémillard (2008).

4 Goodness-of-fit for dynamic parametric models

The first category of models one considers is the family of parametric models, where the null hypothesis takes the form

H_0: For some $\theta \in \mathcal{O}$, the conditional distribution of Y_t given $\mathcal{F}_{t-1}$ is $F_{t,\theta}$, for all $t \geq 1$.

Throughout this section, one will assume that $F_{t,\theta}$ has a strictly positive density $f_{t,\theta}$ with respect to a reference measure λ_t, not depending on θ. Since each $F_t(\cdot,\theta)$ is a distribution function, one can then associate with it a Rosenblatt transform $R_{t,\theta}$.

Recall that for a given multivariate distribution function H, with $X = (X_1, \ldots, X_d) \sim H$, and continuous marginal distributions $F_1, \ldots, F_d$, the Rosenblatt transform $\mathcal{R}$, studied in Rosenblatt (1952), can be defined as $\mathcal{R}(x) = \left(\mathcal{R}^{(1)}(x_1), \ldots, \mathcal{R}^{(d)}(x_1, \ldots, x_d)\right)$, with $\mathcal{R}^{(1)}(x_1) = F_1(x_1)$, and for $j = 2, \ldots, d$,

$$\mathcal{R}^{(j)}(x_1, \ldots, x_j) = P(X_j \leq x_j | X_1 = x_1, \ldots, X_{j-1} = x_{j-1}).$$

The main property of the Rosenblatt transform is that $X \sim H$ if and only if $U = \mathcal{R}(X) \sim C_\perp$, i.e., U is uniformly distributed in $[0,1]^d$. It also follows that by inverting the mapping, one can generate $X \sim H$ viz. $X = \mathcal{R}^{-1}(U)$, by simulating $U \sim C_\perp$.

Therefore, if $d > 1$, the null hypothesis can be restated as follows:

H_0: For some $\theta \in \mathcal{O}$, the Rosenblatt's transforms of Y_t given $\mathcal{F}_{t-1}$ is $R_{t,\theta}$, for all $t \geq 1$.

It follows that under the null hypothesis, $U_1, \ldots, U_n$ are independent and uniformly distributed in $[0,1]^d$, where $U_t = R_{t,\theta}(Y_t) \sim C_\perp$. Unfortunately, θ is unknown and must be estimated by some statistic $\theta_n = T_n(Y_1, \ldots, Y_n)$. So U_t is not observable and has to to replaced by the pseudo-observation $u_{n,t} = R_{t,\theta_n}(Y_t)$. Therefore it is natural to base a goodness-of-fit test on statistic on the empirical process $\mathbb{D}_n = n^{1/2}(D_n - C_\perp)$, with

$$D_n(u) = \frac{1}{n}\sum_{t=1}^n \mathbf{1}(u_{n,t} \leq u) = \frac{1}{n}\sum_{t=1}^n \prod_{k=1}^d \mathbf{1}(u_{n,t,k} \leq u_k), \quad u = (u_1, \ldots, u_d) \in [0,1]^d.$$

Following the results obtained in many power comparisons, in particular, Genest et al. (2009), it is suggested to use the Cramér-von Mises type statistic

$$\begin{aligned} S_n &= S_n(u_{n,1},\ldots,u_{n,n}) = \int_{[0,1]^d} \mathbb{D}_n^2(u)du \\ &= \frac{1}{n}\sum_{t=1}^n \sum_{i=1}^n \prod_{k=1}^d \{1-\max(u_{n,t,k},u_{n,i,k})\} - \sum_{t=1}^n \prod_{k=1}^d \left(1-u_{n,t,k}^2\right) + \frac{n}{3^d}. \end{aligned} \tag{2}$$

Remark 3 *One could also consider Kolmogorov-Smirnov type statistics but their power seem to be much smaller than those using S_n in addition to be more difficult to compute.*

4.1 Convergence of the empirical process

Suppose that H_0 is true, i.e., assume that for some $\theta \in \mathcal{O}$, the conditional law of Y_t given $\mathcal{F}_{t-1}$ is $F_{t,\theta}$. Furthermore, the following assumptions will be imposed on the parametric family $\{F_{t,\theta};\theta\in\mathcal{O}\}$:

Assumption 1 *For every $t \ge 1$, $F_{t,\theta}$ has density $f_{t,\theta}$ with respect to some reference measure λ_t, not depending on θ. Furthermore,*

- *A1: For every $t \ge 1$, the density $f_{t,\theta}$ admits first and second order continuous derivatives with respect to all components of θ. The gradient (column) vector with respect to θ is denoted $\dot f_{t,\theta}$, and the Hessian matrix is represented by $\ddot f_{t,\theta}$.*
- *A2: For every $t \ge 1$, and for every $\theta_0 \in \mathcal{O}$, there exist a neighborhood $\mathcal{N}$ of θ_0 and a λ_t-integrable function h_t such that $\sup_{\theta\in\mathcal{N}} \left\|\ddot f_{t,\theta}(y)\right\| \le h_t(y)$ and $\sup_{\theta\in\mathcal{N}} \left\|\dot f_{t,\theta}(y)\right\|^2 \le h_t(y)$, λ_t-almost surely.*
- *A3: Setting $\xi_t = \frac{\dot f_{t,\theta_0}(Y_t)}{f_{t,\theta_0}(Y_t)}$, then*
$$\frac{1}{n}\sum_{t=1}^n \xi_t\xi_t^\top \xrightarrow{Pr} \mathcal{J}, \tag{3}$$
with $\mathcal{J}$ invertible, and for every $\epsilon > 0$,
$$\frac{1}{n}\sum_{t=1}^n \mathrm{E}\left\{\|\xi_t\|^2 \mathbf{1}(\|\xi_t\| > \epsilon n^{1/2})|\mathcal{F}_{t-1}\right\} \xrightarrow{Pr} 0. \tag{4}$$
- *A4:*
$$\frac{1}{n}\sum_{t=1}^n \frac{\ddot f_{t,\theta_0}(Y_t)}{f_{t,\theta_0}(Y_t)} \xrightarrow{Pr} 0. \tag{5}$$

In the sequel, θ_0 represents the true (unknown) value of θ and $P = P_{\theta_0}$. Furthermore,

$$f_t = f_{t,\theta_0}, \quad \dot f_t = \dot f_{t,\theta_0}, \quad \ddot f_t = \ddot f_{t,\theta_0}.$$

Remark 4 *Using Assumptions A1-A2, together with Lebesgue's dominated convergence theorem, one may conclude that*

$$\frac{\partial}{\partial\theta}\int f_{t,\theta}(y)g(y)\lambda_t(dy) = \int \dot f_{t,\theta}(y)g(y)\lambda_t(dy), \tag{6}$$

for any bounded measurable function g on $\mathbb{R}^d$ not depending on θ. In particular,

$$\mathrm{E}\left\{ \left.\frac{\dot f_t(Y_t)}{f_t(Y_t)}\right|_{\mathcal{F}_{t-1}} \right\} = \int \dot f_t(y)\lambda_t(dy) = 0. \tag{7}$$

Set

$$\mathbb{W}_n = n^{-1/2}\sum_{t=1}^n \frac{\dot f_t(Y_t)}{f_t(Y_t)}. \tag{8}$$

Note that by Assumption A3 and (7), $\xi_t = \frac{\dot f_t(Y_t)}{f_t(Y_t)}$ are square integrable martingale differences satisfying the conditions of the Lindeberg-Feller Theorem for martingale differences stated in Appendix C. As a result, $\mathbb{W}_n \rightsquigarrow \mathbb{W} \sim N(0,\mathcal{J})$. Furthermore, in most applications, ξ_t is a function of an ergodic Markov process with a unique stationary measure, so assumptions A3-A4 are automatically met.

The following theorem generalizes the results of Bai (2003) and Bai and Chen (2008) and identifies the limiting deterministic function γ appearing in the limit. To state it, set $\mathbb{B}_n = n^{1/2}(B_n - C_\perp)$, where

$$B_n(u) = n^{-1}\sum_{t=1}^n \mathbf{1}(U_t \le u), \qquad u \in [0,1]^d,$$

and where $U_t = R_{t,\theta_0}(Y_t)$, $t \ge 1$.

Theorem 2 *Suppose that Assumptions A1–A4 are met and that $\Theta_n = n^{-1/2}\sum_{t=1}^n \eta_t + o_P(1)$, where the η_t are square integrable martingale differences satisfying the conditions of the Lindeberg-Feller Theorem for martingale differences in Appendix C. If in addition,*

$$\frac{1}{n}\sum_{t=1}^n \mathbf{1}(U_t \le u)\xi_t \xrightarrow{Pr} \gamma(u), \qquad u \in [0,1]^d, \tag{9}$$

$$\frac{1}{n}\sum_{t=1}^n \eta_t \xi_t^\top \xrightarrow{Pr} I, \tag{10}$$

and

$$\frac{1}{n}\sum_{t=1}^n \mathbf{1}(U_t \le u)\eta_t \xrightarrow{Pr} \psi(u), \qquad u \in [0,1]^d, \tag{11}$$

where γ and ψ are continuous, then $(\mathbb{W}_n, \Theta_n, \mathbb{B}_n, \mathbb{D}_n) \rightsquigarrow (\mathbb{W}, \Theta, \mathbb{B}, \mathbb{D})$, with $\mathbb{D}_n \rightsquigarrow \mathbb{D} = \mathbb{B} - \Theta^\top\gamma$, where $\gamma(u) = \mathrm{E}\{\mathbb{B}(u)\mathbb{W}\}$, $u \in [0,1]^d$, and $\mathbb{B}$ is a $C_\perp$-Brownian bridge. In addition, the parametric bootstrap work for $\mathbb{D}_n$ since θ_n is regular.

Example 2 *If the maximum likelihood estimator θ_n of θ exists, then $\Theta_n = \mathcal{J}^{-1}\mathbb{W}_n + o_P(1)$, so $\eta_t = \mathcal{J}^{-1}\xi_t$ satisfies conditions (10) and (11).*

Remark 5 *As discussed in Bai (2003), in practice, a truncated conditional expectation is used, instead of the full conditional expectation, meaning that the information about $(Y_t)_{t\le 0}$ is replaced by setting $Y_t = 0$, for all $t \le 0$. In most dynamic models, that does not affect the methodology since for large t, the process usually forgets from where it starts (otherwise estimation would be impossible).*

The rest of the section is dedicated to particular examples.

4.2 Markovian models

Suppose that $(Y_t)_{t\ge 1}$ is a time series with values in $\mathbb{R}^d$. It is called p-Markov, $p \ge 1$, if the the process $\{Z_t = (Y_{t-p+1}, \ldots, Y_t)\}_{t\ge p}$ is Markov. In other words, if $t \ge p$, then the conditional law of Y_t given $\mathcal{F}_{t-1}$ depends only on Z_{t-1}. Suppose that under P_θ, that conditional law has a density $g_\theta(z,y)$ with respect to some reference measure λ. As a result, $f_{t,\theta}(y) = g_\theta(z_{t-1}, y)$, $t > p$. Assume also that under P_θ, the law of Z_p has density π_θ with respect to a reference measure λ_0, with π_θ bounded and continuous with respect to θ. Finally, suppose that the process is ergodic, with unique stationary measure ν. For simplicity set $g = g_{\theta_0}$ and let $X_t = (Z_{t-1}, Y_t) = (Y_{t-p}, \ldots, Y_t)$, $t > p$. Then

$$\mathbb{W}_n = n^{-1/2}\sum_{t=p+1}^n \frac{\dot g(X_t)}{g(X_t)} + o_P(1),$$

and Assumptions A3-A4 are met if A1-A2 holds for $f_{t,\theta} = g_\theta$.

Finally, let $\theta_n = T_n(Y_1, \ldots, Y_n)$ be an estimator of θ and introduce $\Theta_n = n^{1/2}(\theta_n - \theta_0)$.

4.2.1 Empirical process for goodness-of-fit

For simplicity set $Z_t = (Y_1, \ldots, Y_t)$ whenever $t \le p$, and $X_t = (Z_{t-1}, Y_t) = (Y_1, \ldots, Y_t)$ whenever $t \le p$. It follows that the Rosenblatt's transform $R_{t,\theta}$ is given by $R^{(1)}_{t,\theta}(y) = P\left(Y_{t1} \le y_1 | Z_{t-1}\right)$, and for $j = 2, \ldots, d$,

$$R^{(j)}_{t,\theta}(y) = P\left(Y_{tj} \le y_j | Z_{t-1}, Y_{tk} = y_k, k = 1, \ldots, j-1\right),$$

for all $t \ge 1$. When $t > p$, then $R_{t,\theta}(y) = G_\theta(Z_{t-1}, y)$, where $G_\theta(z, \cdot)$ is the Rosenblatt's transform associated with density $g_\theta(z, \cdot)$. As a result, condition (9) is satisfied with

$$\gamma(u) = \int \mathbf{1}\{G(z,y) \le u\}\, \dot{g}(z,y)\lambda(dy)\nu(dz) = E\{\mathbb{B}(u)\mathbb{W}\}, \qquad u \in [0,1]^d, \tag{12}$$

where $G = G_{\theta_0}$ and $g = g_{\theta_0}$.

Remark 6 *For simplicity, set $G = G_{\theta_0}$. $H_n(x) = G_{\theta_n}(x)$, $H(x) = R(x)$. It follows easily that if $\mathbb{H}_n = n^{1/2}(H_n - H)$, and $\Theta_n \rightsquigarrow \Theta$, then $\mathbb{H}_n \rightsquigarrow \mathbb{H}$, where $\mathbb{H}^{(j)}(x) = \Theta^\top \dot{G}^{(j)}(x)$, $x \in \mathbb{R}^{(p+1)d}$, $j \in \{1, \ldots, d\}$. Using the results in Ghoudi and Rémillard (2004), one can then conclude that $\mathbb{D}_n = n^{1/2}(D_n - C_\perp) \rightsquigarrow \mathbb{D}$, where*

$$\mathbb{D}(u) = \mathbb{B}(u) - \Theta^\top \left\{\sum_{j=1}^d \gamma_j(u)\right\}, \qquad u \in [0,1]^d,$$

with

$$\gamma_j(u) = E\left\{\dot{G}^{(j)}(X_{p+1})\mathbf{1}(V_{p+1} \le u) | V_{p+1,j} = u_j\right\}, \qquad j = 1, \ldots, d. \tag{13}$$

The next lemma, proven in Appendix B.1, shows that $\gamma = \sum_{j=1}^d \gamma_j$.

Lemma 1 *For all $u \in [0,1]^d$, $\gamma(u) = \mathrm{E}\{\mathbb{B}(u)\mathbb{W}\} = \sum_{j=1}^d \gamma_j(u)$, where γ and γ_j are respectively defined by (12) and (13).*

Finally, we state some conditions on the estimator θ_n so that it is regular and satisfies all conditions of Theorem 2.

Example 3 *Suppose that $\theta_n = T_n(Y_1, \ldots, Y_n)$ for $\theta \in \mathcal{O}$ satisfies*

$$\Theta_n = n^{-1/2} \sum_{t=p+1}^n J_\theta(X_t) + o_P(1) \tag{14}$$

where the score function $J_\theta : \mathbb{R}^d \to \mathbb{R}^s$ is square integrable with respect to $g_\theta(z,y)\lambda(dy)\nu(dz)$ and such that for all $\theta \in \mathcal{O}$, one has both

$$\int J_\theta(z,y) g_\theta(z,y)\lambda(dy) = 0 \quad \text{and} \quad \int J_\theta(z,y)\dot{g}_\theta^\top(z,y)\lambda(dy)\pi(dz) = I. \tag{15}$$

Then θ_n is regular and conditions (10)-(11) are met. Also, since X is ergodic, $\eta_t = J_\theta(X_t)$ satisfies the Lindeberg-Feller CLT for martingale differences (Appendix C).

As shown by Genest and Rémillard (2008), many well-known estimators are regular. With a few adaptations for the Markovian setting, they are also regular. In addition to the MLE, there are other well-known estimators satisfying (14)–(15).

Example 4 (Moments-based estimators) *Many moments estimators also satisfy (14)–(15). Assume that $\theta = \psi(\mu)$, where for all z, ν_θ a.s.,*

$$\int M(z,y) g_\theta(z,y)\lambda(dy) = \mu,$$

for some integrable function $M : \mathbb{R}^{(p+1)d} \to \mathbb{R}^d$ *that does not depend on* θ*. Suppose also that* ψ *is continuously differentiable and that the matrix* $\dot\psi$ *of derivatives is non-singular. Then* ψ^{-1} *exists and is continuously differentiable by the inverse function theorem. Furthermore, Slutsky's theorem implies that for all* $x \in \mathbb{R}^{(p+1)d}$*, condition* (14) *is satisfied with* $J_\theta(x) = \dot\psi\{\psi^{-1}(\theta)\}\{M(x) - \psi^{-1}(\theta)\}$*. Finally, condition* (15) *is also met since*

$$\int \{M(z,y) - \psi^{-1}(\theta)\} g_\theta(z,y)\lambda(dy) = 0$$

and

$$\begin{aligned} \mathrm{E}(\Theta\mathbb{W}^\top) &= \dot\psi\{\psi^{-1}(\theta_0)\} \int M(z,y)\dot g(z,y)\lambda(dy)\pi(dz) \\ &= \dot\psi\{\psi^{-1}(\theta_0)\} \left[\frac{\partial}{\partial\theta}\int\int M(z,y)g_\theta(z,y)\lambda(dx)\pi(dz)\right]_{\theta=\theta_0} \\ &= \dot\psi\{\psi^{-1}(\theta_0)\} \left[\frac{\partial}{\partial\theta}\psi^{-1}(\theta)\right]_{\theta=\theta_0} = I. \end{aligned}$$

4.3 Regime-switching Markovian models

Suppose that (τ_t) is a (non observable) Markov chain on $\{1,\ldots,m\}$ with transition matrix Q and (τ_t, Y_t) is a Markov process so that given $\tau_{t-1} = i$ and $Y_{t-1} = z$, (τ_t, Y_t) has density $Q_{ij} f_\theta(j,z,y)$ with respect to measure $\lambda \times \nu$, ν being the counting measure on $\{1,\ldots,m\}$. It means that for any bounded continuous function h on $\{1,\ldots,m\} \times \mathbb{R}^d$,

$$E\left\{h(\tau_t, Y_t)|\tau_{t-1} = i, Y_{t-1} = z\right\} = \sum_{j=1}^m Q_{ij}\int h(j,y) f_\theta(j,z,y)dy.$$

Even if (τ_t, Y_t) is a Markov process, the results of the previous section does not apply directly since the regime process τ is not observed, only Y being observed.

The next result shows that the parametric bootstrap works for regime-switching Markov models when parameters are estimated using the EM algorithm.

Proposition 1 *Suppose that the process* (τ_t, Y_t) *is ergodic with stationary measure* π *for the Markov chain* τ*. Under the smoothness conditions in Cappé et al. (2005), if* $(\mathbb{Q}_n, \theta_n)$ *are the estimated parameters of* (Q, θ) *using the EM algorithm, then these estimators are regular and parametric bootstrap works.*

The proof is given in Appendix B.3. Note that the conditions of Proposition 1 hold for the traditional HMM model with Gaussian densities. An implementation of the parametric bootstrap in that setting is illustrated in Rémillard et al. (2010). Note that because the process is ergodic, the values of τ_0 and Y_0 are not important.

Remark 7 *For the selection of the number* r *of regimes, it makes sense to choose the first* r_0 *for which the* P*-value of the test of goodness-of-fit is larger than* 5%*. That was proposed in Papageorgiou et al. (2008).*

4.4 Dynamic models with innovations

By a dynamic model with innovations, one means a model of the following form:

$$Y_t = \mu_t + \sigma_t\varepsilon_t,$$

where the ε_t are i.i.d. with mean 0, covariance matrix I and common distribution K_{θ_1}, with density g_{θ_1}, and with Rosenblatt's transform G_{θ_1}, $\mu_t \in \mathbb{R}^d$ and $\sigma_t \in \mathbb{R}^{d\times d}$ are $\mathcal{F}_{t-1}$ measurable and do not depend on parameter θ_1, only on parameter θ_2. Here, $\theta = (\theta_1, \theta_2)^\top$. In addition, σ_t is invertible.

In a model with innovations, it follows that $R_{t,\theta}(y) = G_{\theta_1}\left\{\sigma_t^{-1}(y-\mu_t)\right\}$, and $f_{t,\theta}(y) = g_{\theta_1}\left\{\sigma_t^{-1}(y-\mu_t)\right\}/|\sigma_t|$, where $|\sigma_t|$ is the determinant of σ_t. Note that $R_{t,\theta}(Y_t) = G_\theta(\varepsilon_t)$ and if $e_{n,t}$ is the residual estimating ϵ_t, depending on $\theta_{n,2}$, then $v_{n,t} = G_{\theta_{n,1}}(e_{n,t})$.

It follows that $f_t(y) = g\left\{\sigma_t^{-1}(y-\mu_t)\right\}/|\sigma_t|$, so

$$\partial_{\theta_1} \log f_{t,\theta}(y) = \frac{\dot{g}\left\{\sigma_t^{-1}(y-\mu_t)\right\}}{g\left\{\sigma_t^{-1}(y-\mu_t)\right\}}$$

and

$$\partial_{\theta_{2j}} \log f_{t,\theta}(y) = -\frac{g'\left\{\sigma_t^{-1}(y-\mu_t)\right\}}{g\left\{\sigma_t^{-1}(y-\mu_t)\right\}}\sigma_t^{-1}\left\{\partial_{\theta_{2j}}\sigma_t\sigma_t^{-1}(y-\mu_t) + \partial_{\theta_{2j}}\mu_t\right\} - \frac{\partial_{\theta_{2j}}|\sigma_t|}{|\sigma_t|},$$

$1 \le j \le s_2$, where g' is the (row) gradient vector of g with respect to x. As a result, $\mathbb{W}_n = (\mathbb{W}_{n,1}, \mathbb{W}_{n,2})^\top$, with

$$\mathbb{W}_{n,1} = n^{-1/2}\sum_{t=1}^n \frac{\dot{g}(\varepsilon_t)}{g(\varepsilon_t)} \tag{16}$$

and

$$\mathbb{W}_{n,2,j} = -n^{-1/2}\sum_{t=1}^n \left[\frac{g'(\varepsilon_t)}{g(\varepsilon_t)}\sigma_t^{-1}\left\{\partial_{\theta_{2j}}\sigma_t\varepsilon_t + \partial_{\theta_{2j}}\mu_t\right\} + \frac{\partial_{\theta_{2j}}|\sigma_t|}{|\sigma_t|}\right], \tag{17}$$

for all $1 \le j \le s_2$.

Remark 8 *Recall that* $\partial_v \log A_v = \mathrm{Trace}\left(A_v^{-1}\partial_v A_v\right)$. *As a result,* $\frac{\partial_{\theta_{2j}}|\sigma_t|}{|\sigma_t|} = \mathrm{Trace}\left(\sigma_t^{-1}\partial_{\theta_{2j}}\sigma_t\right)$, *for all* $1 \le j \le s_2$.

It is easy to check that if θ is estimated by the maximum likelihood method, then θ_n, if it exists, will be regular.

However, in applications, θ_2 is often estimated using the so-called quasi maximum likelihood method (QMLE), where the innovations are treated as Gaussian even if they are not. More precisely, $\theta_{n,2}$ is the value minimizing

$$\theta_{n,2} = \arg\min_{\theta_2}\left\{\sum_{t=1}^n (y_t-\mu_t)^\top h_t^{-1}(y_t-\mu_t) + \sum_{t=1}^n \log|h_t|\right\} = \arg\min_{\theta_2} L(\theta_2),$$

where $h_t = \sigma_t\sigma_t^\top$.

The following proposition in proven in the Appendix.

Proposition 2 *Under the conditions above, if* $\theta_{n,1}$ *is a regular estimator of* θ_1 *then* θ_n *is a regular estimator of* θ.

Note that unfortunately, the limiting process will depend on the unknown value θ_2, so one cannot just apply the parametric bootstrap to the innovations. One has to generate the process Y and estimate parameters $\theta = (\theta_1, \theta_2)$ for each replication.

Finally, in addition to Theorem 2, one can state a result that will be useful when dealing with tests of independence or goodness-of-fit for the innovations.

To do so, set $\mathbb{K}_n = (K_n - K)$, where

$$K_n(x) = \frac{1}{n}\sum_{t=1}^n \mathbf{1}(e_{n,t} \le x), \qquad x \in [-\infty,\infty]^d.$$

Further set

$$\alpha_n(x) = \frac{1}{n^{1/2}}\sum_{t=1}^n \left\{\mathbf{1}(\varepsilon_t \le x) - K(x)\right\}, \qquad x \in [-\infty,\infty]^d.$$

Theorem 3 *Suppose that Assumptions A1–A4 are met and that $\Theta_{n,2} = n^{-1/2}\sum_{t=1}^n \eta_t + o_P(1)$, where the η_t are square integrable martingale differences satisfying the conditions of the Lindeberg-Feller Theorem for martingale differences in Appendix C. If in addition,*

$$\frac{1}{n}\sum_{t=1}^n \mathbf{1}(\varepsilon_t \le x)\xi_{t,2} \xrightarrow{Pr} \gamma(x), \qquad x \in [-\infty,\infty]^d, \tag{18}$$

$$\frac{1}{n}\sum_{t=1}^n \eta_t \xi_{t,2}^\top \xrightarrow{Pr} \Gamma, \tag{19}$$

and

$$\frac{1}{n}\sum_{t=1}^n \mathbf{1}(\varepsilon_t \le x)\eta_t \xrightarrow{Pr} \psi(x), x \in [-\infty,\infty]^d, \tag{20}$$

where γ and ψ are continuous, then $(\mathbb{W}_{n,2}, \Theta_{n,2}, \alpha_n, \mathbb{K}_n) \rightsquigarrow (\mathbb{W}_2, \Theta_2, \alpha, \mathbb{K})$, with $\mathbb{K} = \alpha - \Theta_2^\top \gamma_2$, where $\gamma_2(x) = \mathrm{E}\{\alpha(x)\mathbb{W}_2\}$, $x \in [-\infty,\infty]^d$, and α is a K-Brownian bridge.

To see that one can recover Theorem 2 from Theorem 3, consider the case $d = 1$. In this case,

$$D_n(u) = n^{-1}\sum_{t=1}^n \mathbf{1}\{F_{\theta_n}(e_{n,t}) \le u\} = K_n \circ F_{\theta_n}^{-1}(u),$$

so $\mathbb{D}_n(u) = n^{1/2}\{D_n(u) - u\}$ can be written as

$$\begin{aligned}
\mathbb{D}_n(u) &= \mathbb{K}_n \circ F_{\theta_{n,1}}^{-1}(u) + n^{1/2}\{K \circ F_{\theta_{n,1}}^{-1}(u) - u\} \\
&\rightsquigarrow \mathbb{K} \circ F^{-1}(u) - \Theta_1^\top \dot{F} \circ F^{-1}(u) \\
&= \alpha \circ F^{-1}(u) - \Theta_2^\top \gamma_2 \circ F^{-1}(u) - \Theta_1^\top \dot{F} \circ F^{-1}(u) \\
&= \beta(u) - \Theta^\top \gamma(u),
\end{aligned}$$

where $\gamma(u) = \mathrm{E}\{\beta(u)\mathbb{W}\}$, $u \in [0,1]$, and $\beta = \alpha \circ F^{-1}$ is a standard Brownian bridge.

5 Goodness-of-fit tests for copula-based models

The second category of models one considers is the family of semiparametric models, where the null hypothesis is a parametric hypothesis about the serial and interdependence of the series. In most cases, these models are concerned with two-stage modeling: The first stage is the modeling of univariate series, while the second stage is the modeling of the dependence between the series. More precisely, it is often assumed that $Y_{t,j} = \mu_{t,j} + h_{t,j}^{1/2}\varepsilon_{t,j}$, where $\mu_t, h_t \in \mathcal{F}_{t-1}$ and $\varepsilon_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,d})^\top$ are i.i.d. with continuous marginal distributions $F_1, \ldots, F_d$ and copula C. These models appeared in van den Goorbergh et al. (2005), Patton (2006), Chen and Fan (2006a). Formal goodness-of-fit tests were treated in Rémillard (2010), so there is no need to pursue that subject here.

Another type of copula-based models for univariate time series were studied in Chen and Fan (2006b) and extended to the multivariate case in Rémillard et al. (2011). More precisely, assume that the time series Y is p-Markov and stationary. One is not interested in modeling the series Y but in modeling the series U, where

$$U_t = M(Y_t) = (M_1(Y_{t1}), \ldots, M_d(Y_{td}))^\top, \quad t \ge 1,$$

and where M_j is the (continuous) marginal distributions of Y_{tj}, $j = 1, \ldots, d$. As a result, each U_{tj} is uniformly distributed over $[0,1]$. For that reason, the p-Markov process U is said to be on natural scale. The copula C of interest in this case is defined as the joint distribution function of $V_{t-1} = (U_{t-p}, \ldots, U_{t-1})$ and U_t. For details on estimation and tests of goodness-of-fit, see, e.g., Rémillard et al. (2011). The rest of the section is devoted in proving the validity of the parametric bootstrap methodology proposed in Rémillard et al. (2011). Not all copulas can be used. In fact, by stationarity, one must have,

$$\mathcal{C}(u) = C(u,\mathbf{1},\ldots,\mathbf{1}) = C(\mathbf{1},u,\mathbf{1},\ldots,\mathbf{1}) = \cdots = C(\mathbf{1},\ldots,\mathbf{1},u), \quad \text{for all } u \in [0,1]^d.$$

Note that $\mathcal{C}$ is a stationary distribution for the p-Markov process U. Throughout the rest of the section, the null hypothesis takes the form

H_0: The distribution function C of (V_{t-1}, U_t), $t > 1$, belongs to the parametric family $\{C_\theta; \theta \in \mathcal{O}\}$.

From now on, assume that the density c_θ of C_θ is continuous and positive on $(0,1)^{(p+1)\times d}$. It then follows that the Markov chain V is irreducible and ergodic (Bradley, 2005)[Theorem 3.5]. As before, instead of working directly with the distribution functions, one will work with the Rosenblatt transforms. Therefore, the null hypothesis can be rewritten as

H_0: The Rosenblatt transform R_t of (V_{t-1}, U_t), belongs to the parametric family $\{R_{t,\theta}; \theta \in \mathcal{O}\}$.

Note that because V is a Markov process, it follows that for all $t > p$, $R_{t,\theta}(u) = G_\theta(v_{t-1}, u)$, for some parametric family G_θ defined on $(0,1)^{(p+1)\times d}$. For simplicity, set $G = G_{\theta_0}$, $\dot G = \dot G_{\theta_0}$, and $\ddot G = \ddot G_{\theta_0}$, where θ_0 is the true (unknown) parameter. For simplicity set $c = c_{\theta_0}$ and set $q = q_{\theta_0}$, where $q_\theta(v) = \int_{(0,1)^r} c_\theta(v,u)du$. Then

$$\mathbb{W}_n = n^{-1/2} \sum_{t=p+1}^n \left\{ \frac{\dot c(V_{t-1}, U_t)}{c(V_{t-1}, U_t)} - \frac{\dot q(V_{t-1})}{q(V_{t-1})} \right\} + o_P(1),$$

and Assumptions A3-A4 are met if A1-A2 holds for $f_{t,\theta}(v,u) = c_\theta(v,u)/q_\theta(v)$.

Before describing the test and the parametric bootstrap procedure, set $u_{n,t,j} = M_{n,j}(Y_{t,j}) = E_{n,j}(U_{t,j})$, where

$$E_{n,j}(u_j) = \frac{1}{n}\sum_{t=1}^n \mathbf{1}(U_{t,j} \le u_j) \tag{21}$$

is the j-th marginal of $E_n(u) = \frac{1}{n}\sum_{t=1}^n \mathbf{1}(U_t \le u)$, $u \in [0,1]^d$. Further set $\mathbb{E}_n = n^{1/2}(E_n - C)$, $F_n(u) = (E_{n1}(u_1), \ldots, E_{nd}(u_d))^\top$,

$$H_n(v) = \left(F_n^\top(v_{11},\ldots,v_{1d}),\ldots,F_n^\top(v_{p1},\ldots,v_{pd})\right)^\top,$$

$v = (v_{11},\ldots,v_{1d},\ldots,v_{p1},\ldots,v_{pd})^\top \in [0,1]^{p\times d}$, and $\mathbb{B}_n = n^{1/2}(B_n - C_\perp)$, where

$$B_n(u) = \frac{1}{n}\sum_{t=1}^n \mathbf{1}\left\{R_{t,\theta_0}(U_t) \le u\right\}, \qquad u \in [0,1]^d.$$

Define $e_{n,t} = R_{\theta_n}(u_{n,t})$, $t \in \{1,\ldots,n\}$. Then for $p < t \le n$, $e_{n,t} = G_{\theta_n}\{H_n(V_{t-1}), F_n(U_t)\}$. In the present context, recall that (U_t) is a stationary and ergodic Markov process so that $(V_{t-1}, U_t) \sim C$. Under the null hypothesis H_0, the empirical distribution function

$$D_n(u) = \frac{1}{n}\sum_{t=1}^n \mathbf{1}\left(e_{n,t} \le u\right), \qquad u \in [0,1]^d,$$

should be "close" to $C_\perp$, the d-dimensional independence copula. Based on the results in Genest et al. (2009), to test H_0, it was proposed in Rémillard et al. (2011) to use the Cramér-von Mises type statistic

$$\begin{aligned} S_n &= T(\mathbb{D}_n) = \int_{[0,1]^d} \mathbb{D}_n^2(u)du = n\int_{[0,1]^d} \left\{D_n(u) - C_\perp(u)\right\}^2 du \qquad (22)\\ &= \frac{n}{3^d} - \frac{1}{2^{d-1}}\sum_{t=1}^n\prod_{k=1}^d\left(1 - e_{n,tk}^2\right) + \frac{1}{n}\sum_{t=1}^n\sum_{j=1}^n\prod_{k=1}^d\left\{1 - \max\left(e_{n,tk}, e_{n,jk}\right)\right\},\end{aligned}$$

where $\mathbb{D}_n = n^{1/2}(D_n - C_\perp)$. Thus assume that $\theta_n = T_n(U_1, \ldots, U_n)$ and suppose that $S_n = \phi(\mathbb{D}_n)$ is a continuous functional of the empirical process $\mathbb{D}_n$. The parametric bootstrap procedure can be described as follows:

> For $k = 1, \ldots, N$, generate a stationary Markov process U^*, so that so that the joint law of $(U^*_{k,j}, \ldots, U^*_{k,j+p})$ is C_{θ_n}, $1 \le j \le n-p$ and estimate θ by $\theta^*_{k,n} = T_n\left(U^*_{k,1}, \ldots, U^*_{k,n}\right)$. Then define $u^*_{k,n,t} = F^*_{k,n}\left(U^*_{k,t}\right)$, and $e^*_{k,n,t} = R_{t,\theta^*_{n,k}}\left(u^*_{k,n,t}\right)$, $t \in \{1, \ldots, n\}$. Compute $S^*_{k,n} = \psi_n\left(\mathbb{D}^*_{k,n}\right)$ according to formula (22), where $D^*_{k,n}(u) = \frac{1}{n}\sum_{t=1}^n \mathbf{1}(e^*_{k,n,t} \le u)$.

Assuming that large values of S_n lead to the rejection of H_0, an approximate P-value for the test based on S_n is given by

$$\frac{1}{N}\sum_{k=1}^N \mathbf{1}(S^*_{k,n} > S_n).$$

5.1 Convergence of the empirical process $\mathbb{D}_n$

In addition to Assumptions A1–A4, assume that $\mathbb{E}_n \rightsquigarrow \mathbb{E}$, where $\mathbb{E}$ is a continuous centered Gaussian process. That condition yields the convergence of $\mathbb{D}_n$, as well as the convergence of pseudo-likelihood estimators of θ. See, e.g., Rémillard et al. (2011). In fact, a sufficient condition for the convergence of $\mathbb{E}_n$ is that the process U is α-mixing, with $\alpha(n) \le cn^{-a}$, for some $c > 0$ and $a > 1$. Most copula families satisfy this property (Rémillard et al., 2011). Under these conditions, it can also be shown that $\mathbb{C}_n = n^{1/2}(C_n - C) \rightsquigarrow \mathbb{C}$, where $\check{\mathbb{C}}_n = n^{1/2}(\check{C}_n - C) \rightsquigarrow \check{\mathbb{C}}$,

$$C_n(v,u) = n^{-1}\sum_{t=p+1}^n \mathbf{1}(H_n(V_{t-1}) \le v, F_n(U_t) \le u),$$

$$\check{C}_n(v,u) = n^{-1}\sum_{t=p+1}^n \mathbf{1}(V_{t-1} \le v, U_t \le u),$$

and $\mathbb{C}$ has representation $\mathbb{C}(v,u) = \check{\mathbb{C}}(v,u) - \mathbb{F}(u)^\top \nabla_u C(v,u) - \mathbb{H}(v)^\top \nabla_v C(v,u)$, $(v,u) \in [0,1]^{(p+1)d}$. Under the smoothness assumptions on G_θ, it then follows that for all $j = 1, \ldots, d$, $n^{1/2}\left[G^{(j)}_{\theta_n}\{H_n(v), F_n(u)\} - G^{(j)}(v,u)\right]$ converges weakly

$$\Theta^\top \dot{G}^{(j)}(v,u) + \sum_{k=1}^d \partial_{u_k} G^{(j)}(v,u)\mathbb{E}_k(u_k) + \sum_{l=1}^p \sum_{k=1}^d \partial_{v_{lk}} G^{(j)}(v,u)\mathbb{E}_k(v_{lk}).$$

Next, under Assumptions A1–A2, $\Delta M_t = \frac{\dot{c}(V_{t-1},U_t)}{c(V_{t-1},U_t)} - \frac{\dot{q}(V_{t-1})}{q(V_{t-1})}$, $t > p$, form a martingale difference sequence satisfying the conditions of the Lindeberg-Feller CLT for martingales (Appendix C). As a result, $\mathbb{W}_n = n^{-1/2}\sum_{t=p+1}^n \Delta M_t + o_P(1) \rightsquigarrow \mathbb{W} \sim N_p(0, \mathcal{J})$, where

$$\mathcal{J} = \int_{(0,1)^{(p+1)\times d}} \frac{\dot{c}_\theta(v,u)\dot{c}_\theta(v,u)^\top}{c_\theta(v,u)} dvdu - \int_{(0,1)^{p\times d}} \frac{\dot{q}_\theta(v)\dot{q}_\theta(v)^\top}{q_\theta(v)} dv,$$

if the chain is ergodic. As said before, the latter is true because $c_\theta(v,u) > 0$ for all $u, v \in (0,1)^{(p+1)\times d}$. By Lemma 1,

$$\sum_{j=1}^d \mathrm{E}\left\{\dot{G}^{(j)}(V,U)\mathbf{1}\{G(V,U) \le u\} | G^{(j)}(V,U) = u_j\right\} = \gamma(u) = \mathrm{E}\{\mathbb{B}(u)\mathbb{W}\}.$$

Next, for any $k \in \{1, \ldots, d\}$, and for any bounded continuous function h on $[0,1]$ so that $h(0) = h(1) = 0$, set

$$\begin{aligned}\mu_k(u,h) &= \sum_{j=1}^{d}\sum_{l=1}^{p} \mathrm{E}\left\{\partial_{v_{lk}} G^{(j)}(V,U) h(V_{lk}) \mathbf{1}\{G(V,U)\le u\} | G^{(j)}(V,U)=u_j\right\} \\ &\quad + \sum_{j=1}^{d} \mathrm{E}\left\{\partial_{u_k} G^{(j)}(V,U) h(U_k) \mathbf{1}\{G(V,U)\le u\} | G^{(j)}(V,U)=u_j\right\}.\end{aligned}$$

Therefore, using the results in Ghoudi and Rémillard (2004), one may conclude that $\mathbb{D}_n \rightsquigarrow \mathbb{D}$, where $\mathbb{D}(u) = \mathbb{B}(u) - \Theta^\top \gamma(u) - \sum_{k=1}^d \mu_k(u, \mathbb{E}_k)$.

5.2 Validity of the parametric bootstrap

It follows from Theorem 1 that if θ_n is regular for θ, then the parametric bootstrap work if and only if $\mathrm{E}\{\mathbb{D}(u)\mathbb{W}\} = 0$ for all $u \in [0,1]^d$. This is the content of the next result which is proven in Appendix B.2.

Lemma 2 $\mathrm{E}\{\mathbb{D}(u)\mathbb{W}\} = 0$ *for all* $u \in [0,1]^d$ *and* $\mathrm{E}\{\mathbb{C}(v,u)\mathbb{W}\} = \dot{C}(v,u)$ *for all* $(v,u) \in [0,1]^{(p+1)d}$.

We now give some examples of regular estimators.

Example 5 (Pseudo maximum likelihood) *An obvious extension of the pseudo maximum likelihood method (Genest et al., 1995) to the Markovian case consists is maximizing*

$$\sum_{t=p+1}^{n} \log\left\{\frac{c_\theta\left(\hat{V}_{t-1}, \hat{U}_t\right)}{q_\theta\left(\hat{V}_{t-1}\right)}\right\} \tag{23}$$

with respect to θ, *where* c_θ *is the density of* C_θ, *assumed to be non vanishing on* $(0,1)^{(p+1)\times d}$, *and* q_θ *is the density of* $C_\theta(v,\mathbf{1})$. *Note that* (23) *is the logarithm of the conditional density of* $U_{p+1},\ldots,U_n$, *given* $U_1,\ldots,U_p$. *Under assumptions A1–A4 in Rémillard et al. (2011), and if the sequence* U_t *is* α*-mixing, then the maximum likelihood estimator* $\tilde{\theta}_n$ *obtained by maximizing*

$$\sum_{t=p+1}^{n} \log\left\{\frac{c_\theta\left(V_{t-1}, U_t\right)}{q_\theta\left(V_{t-1}\right)}\right\} \tag{24}$$

with respect to θ *behaves nicely. In fact,* $n^{1/2}\left(\tilde{\theta}_n - \theta\right) = \mathcal{J}^{-1}\mathbb{W}_n + o_P(1) \rightsquigarrow \tilde{\Theta} = \mathcal{J}^{-1}\mathbb{W} \sim N_p\left(0, \mathcal{J}^{-1}\right)$. *It follows that* $n^{1/2}(\theta_n - \theta) \rightsquigarrow \tilde{\Theta} + \check{\Theta} \sim N_p(0, J)$, *for some covariance matric* J, *if* $L_\theta(v,u) = \frac{\dot{c}_\theta(v,u)}{c_\theta(v,u)} - \frac{\dot{q}_\theta(v)}{q_\theta(v)}$ *is continuously differentiable with respect to* (v,u) *and* $\check{\Theta} = \mathcal{J}^{-1}\int\left\{\nabla_u L_\theta^\top(v,u)\mathbb{F}(u) + \nabla_v L_\theta^\top(v,u)\mathbb{H}(v)\right\} dC_\theta(v,u)$. *It then follows that* θ_n *is a regular estimator of* θ, *since by the proof of Lemma 2,* $\mathrm{E}\{\mathbb{F}(u)\mathbb{W}\} = 0$ *and* $\mathrm{E}\{\mathbb{H}(v)\mathbb{W}\} = 0$, *for all* $(v,u) \in [0,1]^{(p+1)d}$.

Example 6 (Moments-based estimators) *Often, it can be shown that for some moments* $\mathcal{M}$, $\mathcal{M} = K(\theta)$, *with* K *invertible and differentiable, and* $n^{1/2}(\mathcal{M}_n - \mathcal{M}) = \mathcal{H}_\theta(\mathbb{C}_n) + o_P(1)$, *for some continuous linear function* $\mathcal{H}$ *with values in* $\mathbb{R}^s$. *It then follows that* $\theta_n = K^{-1}(\mathcal{M}_n)$ *is a regular estimator of* θ *if* $\mathcal{H}_\theta(\dot{C}_\theta) = K'(\theta)$. *In particular, Kendall's tau and Spearman's rho are moments that satisfy* $n^{1/2}(\mathcal{M}_n - \mathcal{M}) = \mathcal{H}_\theta(\mathbb{C}_n) + o_P(1)$.

6 An illustration

Consider a GARCH(1,1) model with GED innovations, i.e., $Y_t = \mu + \sigma_t\varepsilon_t$, where $\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, and ε_t has GED distribution with parameter $\nu > 0$, with density $h_\nu(x) = \frac{\nu}{2^{1+1/\nu} b_\nu \Gamma(1/\nu)} e^{-\frac{1}{2}\left(\frac{|x|}{b_\nu}\right)^\nu}$, and $b_\nu = 2^{-1/\nu}\sqrt{\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}}$. The innovation ε are independent and ε_t is independent of $\mathcal{F}_{t-1}$. Note that the case $\nu = 2$ corresponds to the Gaussian distribution. From the results of Section 4.4, using the MLE or QMLE estimates, the parametric bootstrap approach is valid.

Implementing Khmaladze's transform for Gaussian innovations is relatively easy, while implementing it for the GED distribution is very difficult. However, the parametric bootstrap approach is always easy to implement. Using the parametric bootstrap approach with the maximum likelihood estimator, we obtain the following parameters for the returns of Apple (appl) from January 14th 2009 to January 14thm 2011. $\hat{\mu} = 0.0028$, $\hat{\omega} = 7.12 \times 10^{-7}$, $\hat{\alpha} = 0.0817$, $\hat{\beta} = 0.8969$, and $\hat{\nu} = 1.3511$. The P-values corresponding to the Kolmogorov-Smirnov test statistic (K_n) and the Cramér-von Mises statistic (S_n) are respectively 14.4% and 28.4%, using $N = 1000$ bootstrap samples. Hence the null hypothesis of a GARCH(1,1) model with GED innovations is not rejected. For the Gaussian distribution however, corresponding to a GED with $\nu = 2$, the Kolmogorov-Smirnov and Cramér-von Mises statistics based on the Khmaladze's transform, and both our test statistics K_n and S_n yield P-values close to 0, rejecting the null hypothesis of a GARCH(1,1) model with Gaussian innovations.

A Proofs of the main results

A.1 Proof of Theorem 1

The proof is very similar to the proof of the analogous result in Genest and Rémillard (2008) obtained in the serial independent case; however it is included here for sake of completeness.

Suppose $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ is an independent copy of $(Y_1, \ldots, Y_n)$. Denote their joint law by P_n. Set $\tilde{\ell}_n = \ell_n(\tilde{Y}_1, \ldots, \tilde{Y}_n, \theta_n)$, and denote by P_n^* the joint law of $(Y_1, \ldots, Y_n, \tilde{Y}_1, \ldots, \tilde{Y}_n)$ under the change of measure defined by $\frac{dP_n^*}{dP_n} = \exp(\tilde{\ell}_n)$. It is easy to check that under P_n, and conditionally on $(Y_1, \ldots, Y_n)$, $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ has law P_{θ_n}, i.e., under P_n, $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ has the same law as the bootstrap sample $(Y_1^*, \ldots, Y_n^*)$. The rest of the proof is based on a very powerful result, called "Le Cam Third Lemma", that can be used to transfer any convergence result valid for statistics of $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ into a corresponding result for the bootstrap sample $(Y_1^*, \ldots, Y_n^*)$. However, it is much easier to work with $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$, since it is independent of $Y_1, \ldots, Y_n)$ and its law is the same. In particular, if S_n is any statistic of $(Y_1, \ldots, Y_n)$, i.e., $S_n = \psi_n(Y_1, \ldots, Y_n)$, let $\tilde{S}_n$ be its independent copy based on $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$, i.e., $\tilde{S}_n = \psi_n(\tilde{Y}_1, \ldots, \tilde{Y}_n)$, and let S_n^* be its bootstrapped version, i.e. $S_n^* = \psi_n(Y_1^*, \ldots, Y_n^*)$. In particular, define $(\tilde{\mathbb{W}}_n, \tilde{\theta}_n, \tilde{\mathbb{A}}_n)$, $(\mathbb{W}_n^*, , \theta_n^*, \mathbb{A}_n^*)$ accordingly. Further set $\tilde{\Theta}_n = n^{1/2}(\tilde{\theta}_n - \theta)$ and $\Theta_n^* = n^{1/2}(\theta_n^* - \theta)$.

By construction, under P_n, $(\tilde{\mathbb{W}}_n, \tilde{\Theta}_n, \tilde{\mathbb{A}}_n)$ is an independent copy of $(\mathbb{W}_n, \Theta_n, \mathbb{A}_n)$, so

$$(\mathbb{W}_n, \Theta_n, \mathbb{A}_n, \tilde{\mathbb{W}}_n, \tilde{\Theta}_n, \tilde{\mathbb{A}}_n) \rightsquigarrow (\mathbb{W}, \Theta, \mathbb{A}, \tilde{\mathbb{W}}, \tilde{\Theta}, \tilde{\mathbb{A}}),$$

where $(\tilde{\mathbb{W}}, \tilde{\Theta}, \tilde{\mathbb{A}})$ is an independent copy of $(\mathbb{W}, \Theta, \mathbb{A})$. Using the tightness of Θ_n and the joint convergence of $(\Theta_n, \tilde{\mathbb{W}}_n)$, it follows from (1) that

$$\tilde{\ell}_n = \ell_n(\tilde{Y}_1, \ldots, \tilde{Y}_n, \theta_n) = \Theta_n^\top \tilde{\mathbb{W}}_n - \frac{1}{2}\Theta_n^\top \mathcal{J} \Theta_n + o_P(1),$$

as $n \to \infty$. Consequently, setting $\zeta_n = \exp(\tilde{\ell}_n)$, one also gets

$$(\zeta_n, \mathbb{W}_n, \Theta_n, \mathbb{A}_n, \tilde{\mathbb{W}}_n, \tilde{\Theta}_n, \tilde{\mathbb{A}}_n) \rightsquigarrow (\zeta, \mathbb{W}, \Theta, \mathbb{A}, \tilde{\mathbb{W}}, \tilde{\Theta}, \tilde{\mathbb{A}}),$$

with $\zeta = \exp\left(\Theta^\top \tilde{\mathbb{W}} - \frac{1}{2}\Theta^\top \mathcal{J} \Theta\right)$. Note that $\zeta > 0$ and $\mathrm{E}(\zeta | \mathbb{W}, \Theta, \mathbb{A}) = 1$ since $\tilde{\mathbb{W}} \sim N(0, \mathcal{J})$ and is independent of $(\mathbb{W}, \Theta, \mathbb{A})$, so ζ defines a change of measure.

Invoking Le Cam's Third Lemma (van der Vaart and Wellner, 1996), one can now see that P_n^* is contiguous with respect to P_n, and one may conclude that

$$(\mathbb{W}_n, \Theta_n, \mathbb{A}_n, \mathbb{W}_n^*, \Theta_n^*, \mathbb{A}_n^*) \rightsquigarrow (\mathbb{W}, \Theta, \mathbb{A}, \Theta^*, \mathbb{A}^*),$$

where

$$\mathrm{E}\{L(\mathbb{W}, \Theta, \mathbb{A}, \Theta^*, \mathbb{A}^*)\} = \mathrm{E}\{\zeta L(\mathbb{W}, \Theta, \mathbb{A}, \tilde{\Theta}, \tilde{\mathbb{A}})\},$$

for any bounded continuous function L on $\mathbb{R}^s \times \mathbb{R}^s \times \mathcal{D}(T,\mathbb{R}^m) \times \mathbb{R}^s \times \mathcal{D}(T,\mathbb{R}^m)$. It remains to study the law of $(\mathbb{W},\Theta,\mathbb{A},\tilde\Theta,\tilde{\mathbb{A}})$ under the change of measure ζ. To do so, it is enough to study the law of any linear combination of

$$\mathbb{W},\Theta,\mathbb{A}(t_1),\ldots,\mathbb{A}(t_k),\tilde\Theta,\tilde{\mathbb{A}}(s_1),\ldots,\tilde{\mathbb{A}}(s_j).$$

Therefore, to complete the proof, it suffices to study the law of (ξ_1,ξ_2) under the change of measure, for any random variables ξ_1 and ξ_2, with ξ_2 independent of $(\mathbb{W},\Theta,\mathbb{A},\xi_1)$, $(\xi_1,\xi_2,\tilde{\mathbb{W}})$ centered Gaussian, with $\mathrm{E}(\xi_2^2)=\sigma^2$ and $\mathrm{E}\left(\xi_2\tilde{\mathbb{W}}\right)=\gamma$. To that end, let (ξ_1^*,ξ_2^*) denote the associated vector under the change of measure and note that for every $\lambda_1,\lambda_2\in\mathbb{R}$, one has

$$\begin{aligned}
\mathrm{E}\left\{\exp\left(i\lambda_1\xi_1^*+i\lambda_2\xi_2^*\right)\right\} &= \mathrm{E}\{\zeta\exp(i\lambda_1\xi_1+i\lambda_2\xi_2)\}\\
&= \mathrm{E}\{\exp(i\lambda_1\xi_1+i\lambda_2\xi_2+\Theta^\top\tilde{\mathbb{W}}-\Theta^\top\mathcal{J}\Theta/2)\}\\
&= \mathrm{E}\{\exp(i\lambda_1\xi-\lambda_2^2\sigma^2/2+i\lambda_2\gamma^\top\Theta)\}\\
&= \mathrm{E}[\exp\{i\lambda_1\xi_1+i\lambda_2(\xi_2+\gamma^\top\Theta)\}].
\end{aligned}$$

As a result, (ξ_1^*,ξ_2^*) has the same law as $(\xi_1,\xi_2+\gamma^\top\Theta)$. □

A.2 Proofs of Theorems 2–3

Let $\delta>0$ be given. First, note that because Θ_n is tight, one can find $M>0$ such that $P(\|\Theta\|>M)<\delta$. Also, one can find a finite number of vectors $a_1,\ldots,a_m$, m depending on M and δ, so that

$$\{\|\Theta_n\|\le M\}\subset\bigcup_{k=1}^m\{\|\theta_n-\theta_{n,k}\|<\delta/n^{1/2}\},$$

where $\theta_{n,k}=\theta+n^{-1/2}a_k$, $k=1,\ldots,m$. The trick now is to replace the random value Θ_n by the deterministic values $\theta_{n,k}$, using partitions of unity as in Ghoudi and Rémillard (2004). See also van der Vaart and Wellner (2007) who used the same trick.

Under $P=P_{\theta_0}$, $\mathbb{W}_n\rightsquigarrow\mathbb{W}\sim N(0,\mathcal{J})$. Since

$$\log\left(\left.\frac{dP_{\theta_{n,k}}}{dP}\right|_{\mathcal{F}_n}\right)\rightsquigarrow a_k^\top\mathbb{W}-\frac{1}{2}a_k^\top\mathcal{J}a_k,$$

then under $P_{\theta_{n,k}}$, $\mathbb{W}_n\rightsquigarrow\mathbb{W}^{(k)}=\mathbb{W}+\mathcal{J}a_k$, so under $P_{\theta_{n,k}}$,

$$\log\left(\left.\frac{dP}{dP_{\theta_{n,k}}}\right|_{\mathcal{F}_n}\right)\rightsquigarrow -a_k^\top\mathbb{W}^{(k)}+\frac{1}{2}a_k^\top\mathcal{J}a_k=-a_k^\top\mathbb{W}-\frac{1}{2}a_k^\top\mathcal{J}a_k.$$

Next, again invoking Le Cam's Third Lemma (van der Vaart and Wellner, 1996), under $P_{\theta_{n,k}}$, $\mathbb{D}_n^{(k)}\rightsquigarrow\mathbb{B}$, where $\mathbb{B}$ is a $C_\perp$-Brownian bridge, then under P, $\mathbb{D}_n\rightsquigarrow\mathbb{B}-a_k^\top\gamma$, where $\gamma(u)=\mathrm{E}\{\mathbb{B}(u)\mathbb{W}\}$. As a result, going back to Θ, under P, $\mathbb{D}_n\rightsquigarrow\mathbb{B}-\Theta^\top\gamma$, completing the proof of Theorem 2.

The proof of Theorem 3 goes along the same lines. Under $P_{\theta_{n,k}}$, $n^{1/2}(K_n-K_{\theta_{n,k}})\rightsquigarrow\mathbb{B}_K$, where $\mathbb{B}_K$ is a K-Brownian bridge and $\mathrm{E}\{B_K(x)\mathbb{W}\}=\gamma_K(x)$. Hence, by Le Cam's Third Lemma, one gets $\mathbb{K}_n\rightsquigarrow\mathbb{K}=\mathbb{B}_K-a_k^\top\gamma_K$. Going back to Θ, one may conclude that $\mathbb{K}_n\rightsquigarrow\mathbb{B}_K-\Theta^\top\gamma_K$, completing the proof of Theorem 3. □

B Auxiliary results

B.1 Proof of Lemma 1

It follows easily that

$$E\{\mathbb{B}(u)\mathbb{W}\} = E_\pi\left\{\mathbf{1}(U_t \le u)\frac{\dot g(X_t)}{g(X_t)}\right\} = \int \mathbf{1}\{G(z,y)\le u\}\dot g(z,y)\lambda(dy)\pi(dz),$$

for any $t > p$. Here E_π stands for the expectation under the stationary distribution π of Z_t.

Next, set $y^{(j)} = (y_1,\ldots,y_j)$, and define

$$g^{(j-1)}\left(z, y^{(j-1)}\right) = \int g^{(j)}(z, y^{(j)}, y_{j+1})\lambda_j(dy_j),$$

$1 \le j \le d$, with $g^{(d)}(z,y) = g(z,y)$. Note that for any $j = 1,\ldots,d$,

$$G^{(j)}(z,y) = \int_{-\infty}^{y_j} g^{(j)}\left(z, y^{(j-1)}, w\right)\lambda_j(dw)/g^{(j-1)}\left(z, y^{(j-1)}\right). \tag{25}$$

Next, using (25), one gets

$$\begin{aligned}
&\int \mathbf{1}\left\{G^{(d)}(z,y)\le u_d\right\} g_\theta^{(d)}\left(z, y^{(d-1)}, y_d\right)\lambda_d(dy_d)\\
&\quad = g_\theta^{(d-1)}\left(z, y^{(d-1)}\right) G_\theta^{(d)}\left[z, y^{(d-1)}, \left\{G^{(d)}\right\}^{-1}\left(z, y^{(d-1)}, u_d\right)\right],
\end{aligned}$$

so using Lebesgue's dominated convergence,

$$\begin{aligned}
&\int \mathbf{1}\left\{G^{(d)}(z,y)\le u_d\right\}\dot g\left(z, y^{(d-1)}, y_d\right)\lambda_d(dy_d)\\
&\quad = \partial_\theta\left[\int \mathbf{1}\left\{G^{(d)}(z,y)\le u_d\right\} g_\theta^{(d)}\left(z, y^{(d-1)}, y_d\right)\lambda_d(dy_d)\right]_{\theta=\theta_0}\\
&\quad = \partial_\theta\left[g_\theta^{(d-1)}\left(z, y^{(d-1)}\right) G_\theta^{(d)}\left\{z, y^{(d-1)}, \left\{G^{(d)}\right\}^{-1}\left(z, y^{(d-1)}, u_d\right)\right\}\right]_{\theta=\theta_0}\\
&\quad = \dot g^{(d-1)}\left(z, y^{(d-1)}\right) G^{(d)}\left[z, y^{(d-1)}, \left\{G^{(d)}\right\}^{-1}\left(z, y^{(d-1)}, u_d\right)\right]\\
&\qquad + g^{(d-1)}\left(z, y^{(d-1)}\right)\dot G^{(d)}\left[z, y^{(d-1)}, \left\{G^{(d)}\right\}^{-1}\left(z, y^{(d-1)}, u_d\right)\right]\\
&\quad = u_d \dot g^{(d-1)}\left(z, y^{(d-1)}\right)\\
&\qquad + g^{(d-1)}\left(z, y^{(d-1)}\right)\dot G^{(d)}\left[z, y^{(d-1)}, \left\{G^{(d)}\right\}^{-1}\left(z, y^{(d-1)}, u_d\right)\right].
\end{aligned}$$

As a result,

$$\begin{aligned}
E\{\mathbb{B}(u)\mathbb{W}\} &= \gamma_d(u)\\
&\quad + u_d \int\int \dot g^{(d-1)}\left((z, y^{(d-1)}\right)\prod_{k=1}^{d-1}\mathbf{1}\left\{G^{(k)}\left(z, y^{(k)}\right)\le u_k\right\}\\
&\qquad \times \lambda^{(d-1)}\left(dy^{(d-1)}\right)\pi(dz).
\end{aligned}$$

Using (25) and iterating the last procedure, one ends up with

$$
\begin{aligned}
E\{\mathbb{B}(u)\mathbb{W}\} &= \gamma_2(u)+\cdots+\gamma_d(u)\\
&\qquad + \left(\prod_{j=2}^d u_j\right)\int\int \dot g^{(1)}(z,y_1)\mathbf{1}\left\{G^{(1)}(z,y_1)\le u_1\right\}\lambda_1(dy_1)\pi(dz).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\int \dot g^{(1)}(z,y_1)\mathbf{1}\left\{G^{(1)}(z,y_1)\le u_1\right\}\lambda_1(dy_1)\\
&\quad = \partial_\theta\left[\int g_\theta^{(1)}(z,y_1)\mathbf{1}\left\{G^{(1)}(z,y_1)\le u_1\right\}\lambda_1(dy_1)\right]_{\theta=\theta_0}\\
&\quad = \partial_\theta\left[\int G_\theta^{(1)}\left\{z,\left\{G^{(1)}\right\}^{-1}(z,u_1)\right\}\right]_{\theta=\theta_0}\\
&\quad = \dot G^{(1)}\left\{z,\left\{G^{(1)}\right\}^{-1}(z,u_1)\right\}.
\end{aligned}
$$

As a result

$$
\gamma_1(u) = \left(\prod_{j=2}^d u_j\right)\int\int \dot g^{(1)}(z,y_1)\mathbf{1}\left\{G^{(1)}(z,y_1)\le u_1\right\}\lambda_1(dy_1)\pi(dz),
$$

so $E\{\mathbb{B}(u)\mathbb{W}\}=\gamma(u)$, completing the proof. □

B.2 Proof of Lemma 2

Since one already knows that $\mathrm{E}\{\mathbb{B}(u)\mathbb{W}\}=\gamma(u)$, $\mathrm{E}\{\mathbb{D}(u)\mathbb{W}\}=0$ is equivalent to $\sum_{k=1}^d \mathrm{E}\{\mu_k(u,\mathbb{E}_k)\mathbb{W}\}=0$, $u\in[0,1]^d$. From the very definition of μ_k, to prove the previous equality, it is sufficient to show that $\mathrm{E}\{\mathbb{E}_k(s)\mathbb{W}\}=0$, for any $s\in[0,1]$. That follows from the fact that $\mathrm{E}(\mathbb{C}\mathbb{W})=\dot C$, proven next, since

$$
\mathrm{E}\{\mathbb{E}_k(s)\mathbb{W}\} = \dot C(1,1,\ldots,1,s,1,\ldots,1) = \partial_\theta\left[s\right]_{\theta=\theta_0}=0.
$$

Since $\mathbb{C}(v,u)=\check{\mathbb{C}}(v,u)-\mathbb{F}(u)^\top\nabla_u C(v,u)-\mathbb{H}(v)^\top\nabla_v C(v,u)$, it follows from the proof of Lemma 2 that $\mathrm{E}\{\mathbb{C}(v,u)\mathbb{W}\}=\mathrm{E}\{\check{\mathbb{C}}(v,u)\mathbb{W}\}$, for all $(v,u)\in[0,1]^{(p+1)d}$.

Now

$$
\begin{aligned}
\mathrm{E}\{\check{\mathbb{C}}(v,u)\mathbb{W}\} &= \lim_{n\to\infty} n^{-1}\mathrm{E}\left[\sum_{i=p+1}^n\sum_{t=p+1}^n\left\{\frac{\dot c(V_{t-1},U_t)}{c(V_{t-1},U_t)}-\frac{\dot d(V_{t-1})}{d(V_{t-1})}\right\}\mathbf{1}(V_{t-1}\le v,U_t\le u)\right]\\
&= \lim_{n\to\infty} n^{-1}\sum_{t=p+1}^n \mathrm{E}\left[\left\{\frac{\dot c(V_{t-1},U_t)}{c(V_{t-1},U_t)}-\frac{\dot d(V_{t-1})}{d(V_{t-1})}\right\}\mathbf{1}(V_{t-1}\le v,U_t\le u)\right]\\
&\quad + \lim_{n\to\infty} n^{-1}\sum_{p+1\le t<i\le n} \mathrm{E}\left[\left\{\frac{\dot c(V_{t-1},U_t)}{c(V_{t-1},U_t)}-\frac{\dot d(V_{t-1})}{d(V_{t-1})}\right\}\mathbf{1}(V_{t-1}\le v,U_t\le u)\right].
\end{aligned}
$$

First, $\mathrm{E}\left\{\frac{\dot c(V_{t-1},U_t)}{c(V_{t-1},U_t)}\mathbf{1}(V_{t-1}\le v,U_t\le u)\right\}=\dot C(v,u)$, since

$$
\begin{aligned}
\int_{(0,1)^{(p+1)d}} \dot c(z,w)\mathbf{1}(z\le v,w\le u)dzdw &= \partial_\theta\left[\int_{(0,1)^{(p+1)d}} c(z,w)\mathbf{1}(z\le v,w\le u)dzdw\right]_{\theta=\theta_0}\\
&= \partial_\theta\left[C(v,u)\right]_{\theta=\theta_0}=\dot C(v,u).
\end{aligned}
$$

Next, since $\int_{(0,1)^d} c(u,v)du=q(v)$, one gets, whenever $p+1\le t<i$,

$$
\mathrm{E}\left\{\frac{\dot c(V_{t-1},U_t)}{c(V_{t-1},U_t)}\mathbf{1}(V_{i-1}\le v,U_i\le u)\right\} = \mathrm{E}\left\{\frac{\dot q(V_t)}{q(V_t)}\mathbf{1}(V_{i-1}\le v,U_i\le u)\right\},
$$

by using the Markov property and integrating with respect to U_{t-p}. As a result,

$$
\begin{aligned}
& \sum_{t=p+1}^{n} \mathrm{E}\left[\left\{\frac{\dot c(V_{t-1},U_t)}{c(V_{t-1},U_t)} - \frac{\dot q(V_{t-1})}{q(V_{t-1})}\right\}\mathbf{1}(V_{t-1}\le v, U_t \le u)\right] \\
& \quad + \sum_{p+1\le t<i\le n} \mathrm{E}\left[\left\{\frac{\dot c(V_{t-1},U_t)}{c(V_{t-1},U_t)} - \frac{\dot q(V_{t-1})}{q(V_{t-1})}\right\}\mathbf{1}(V_{i-1}\le v, U_i \le u)\right] \\
= & \ (n-p)\dot C(v,u) - \sum_{t=p+1}^{n} \mathrm{E}\left[\frac{\dot q(V_{t-1})}{q(V_{t-1})}\mathbf{1}(V_{t-1}\le v, U_t\le u)\right] \\
& \quad + \sum_{p+1\le t<i\le n} \mathrm{E}\left[\left\{\frac{\dot q(V_t)}{q(V_t)} - \frac{\dot q(V_{t-1})}{q(V_{t-1})}\right\}\mathbf{1}(V_{i-1}\le v, U_i\le u)\right] \\
= & \ (n-p)\dot C(v,u) - \sum_{i=p+1}^{n} \mathrm{E}\left\{\frac{\dot q(V_p)}{q(V_p)}\mathbf{1}(U_{ik}\le s)\right\}.
\end{aligned}
$$

Hence, since C_n converges in probability to C, it follows that

$$
\begin{aligned}
\mathrm{E}\{\mathbb{C}(v,u)\mathbb{W}\} &= \dot C(v,u) - \lim_{n\to\infty} n^{-1}\sum_{i=p+1}^{n} \mathrm{E}\left\{\frac{\dot q(V_p)}{q(V_p)}\mathbf{1}(V_{i-1}\le v, U_i\le u)\right\} \\
&= \dot C(v,u) - C(v,u)\mathrm{E}\left\{\frac{\dot q(V_p)}{q(V_p)}\right\} \\
&= \dot C(v,u) - C(v,u)\int_{(0,1)^{pd}} \dot q(v)dv = \dot C(v,u).
\end{aligned}
$$

Hence one may conclude that $\mathrm{E}\{\mathbb{C}(v,u)\mathbb{W}\} = \dot C(v,u)$ for all $(v,u)\in[0,1]^{(p+1)d}$. □

B.3 Proof of Proposition 1

Suppose $(\tau_t^*, Y_t^*)_{t=0}^n$ is an independent copy of the chain $(\tau_t, Y_t)_{t=0}^n$ with parameters (Q_0,θ_0). Denote by $\mathcal{F}_k^*$ the sigma-algebra generated by $(\tau_t^*, Y_t^*)_{t=0}^k$. Suppose that (Q_n,θ_n) are $n^{1/2}$-consistent estimates of (Q_0,θ_0) and define the law $P_{Q_0,\theta_0}\otimes P_{Q_n,\theta_n}$ on $\mathcal{F}_n\vee\mathcal{F}_n^*$ by

$$
\begin{aligned}
\ell_n &= \log\left(\frac{dP_{Q_0,\theta_0}\otimes P_{Q_n,\theta_n}}{dP_{Q_0,\theta_0}\otimes P_{Q_0,\theta_0}}\right)\Bigg|_{\mathcal{F}_n\vee\mathcal{F}_n^*} \\
&= \sum_{t=1}^{n}\log\left\{\frac{Q_n(\tau_{t-1}^*,\tau_t^*)}{Q_0(\tau_{t-1}^*,\tau_t^*)}\times\frac{f_{\theta_n}(\tau_t^*,Y_{t-1}^*,Y_t^*)}{f_{\theta_0}(\tau_t^*,Y_{t-1}^*,Y_t^*)}\right\} \\
&= \sum_{t=1}^{n}\log\left\{\frac{Q_n(\tau_{t-1}^*,\tau_t^*)}{Q_0(\tau_{t-1}^*,\tau_t^*)}\right\} + \sum_{t=1}^{n}\log\left\{\frac{f_{\theta_n}(\tau_t^*,Y_{t-1}^*,Y_t^*)}{f_{\theta_0}(\tau_t^*,Y_{t-1}^*,Y_t^*)}\right\}.
\end{aligned}
$$

Then under $P_{Q_0,\theta_0}\otimes P_{Q_n,\theta_n}$, given $\mathcal{F}_n$, the Markov chain (τ_t^*, Y_t^*) is determined by law P_{Q_n,θ_n}.

For simplicity, set $Q=Q_0$ and $f=f_{\theta_0}$. Further set $\mathcal{A}_m = \{(i,j); i,j\in\{1,\ldots,m\}, j\ne i\}$.

Then, setting $N_n^*(i,j)=\sum_{t=1}^n \mathbf{1}(\tau_{t-1}^*=i,\tau_t^*=j)$, and $N_n^*(i,\cdot)=\sum_{t=1}^n \mathbf{1}(\tau_{t-1}^*=i)$, one gets

$$
\begin{aligned}
\ell_{n,1} &= \sum_{t=1}^{n}\log\left\{\frac{Q_n(\tau_{t-1}^*,\tau_t^*)}{Q(\tau_{t-1}^*,\tau_t^*)}\right\} \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{t=1}^{n}\mathbf{1}(\tau_{t-1}^*=i,\tau_t^*=j)\log\left\{\frac{Q_n(i,j)}{Q(i,j)}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{m}\sum_{j=1}^{m} N_n^*(i,j)\log\left\{\frac{Q_n(i,j)}{Q(i,j)}\right\}\\
&= \frac{1}{n^{1/2}}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{N_n^*(i,j)}{Q(i,j)}\mathbb{Q}_n(i,j) - \frac{1}{2n}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{N_n^*(i,j)}{\{Q(i,j)\}^2}\{\mathbb{Q}_n(i,j)\}^2 + o_P(1)\\
&= \frac{1}{n^{1/2}}\sum_{(i,j)\in\mathcal{A}_m}\left\{\frac{N_n^*(i,j)}{Q(i,j)} - \frac{N_n^*(i,i)}{Q(i,i)}\right\}\mathbb{Q}_n(i,j) - \frac{1}{2n}\sum_{(i,j)\in\mathcal{A}_m}\frac{N_n^*(i,j)}{\{Q(i,j)\}^2}\{\mathbb{Q}_n(i,j)\}^2\\
&\quad -\frac{1}{2n}\sum_{i=1}^{m}\frac{N_n^*(i,i)}{\{Q(i,i)\}^2}\left\{-\sum_{j\neq i}\mathbb{Q}_n(i,j)\right\}^2 + o_P(1)\\
&= \mathbb{Q}_n^\top \mathbb{W}_{n,1}^* - \frac{1}{2}\mathbb{Q}_n^\top A_n \mathbb{Q}_n + o_P(1),
\end{aligned}
$$

where $\mathbb{Q}_n$ and $\mathbb{W}_{n,1}^*$ are respectively the $m(m-1)$-dimensional vectors with components $n^{1/2}\{Q_n(i,j) - Q(i,j)\}$, and $n^{-1/2}\left\{\frac{N_n^*(i,j)}{Q(i,j)} - \frac{N_n^*(i,i)}{Q(i,i)}\right\}$, $(i,j)\in\mathcal{A}_m$, Δ_n is the $m(m-1)\times m(m-1)$ diagonal matrix with element $\frac{N_n^*(i,j)}{n\{Q(i,j)\}^2}$ at $(i,j)\in\mathcal{A}_m$, and $A_n = \Delta_n + \frac{1}{n}\sum_{i=1}^m \frac{N_n^*(i,i)}{\{Q(i,i)\}^2}\mathbf{1}_i\mathbf{1}_i^\top$, where $\mathbf{1}_i$ is the $m(m-1)$-dimensional vector with zeros everywhere but at elements (i,j) where it is 1, whenever $j\neq i$. Note that $A_n \to A$ a.s., where $A = \Delta + \sum_{i=1}^m \frac{\pi_i}{Q(i,i)}\mathbf{1}_i\mathbf{1}_i^\top$, and Δ is the $m(m-1)\times m(m-1)$ diagonal matrix with element $\frac{\pi_i}{Q(i,j}$ at $(i,j)\in\mathcal{A}_m$.

As a result,

$$
\ell_{n,1} = \mathbb{Q}_n^\top \mathbb{W}_{n,1}^* - \frac{1}{2}\mathbb{Q}_n^\top A \mathbb{Q}_n + o_P(1). \tag{26}
$$

Further note that $\mathbb{W}_{n,1}^* \rightsquigarrow \mathbb{W}_1^*$, where $\mathbb{W}_1^*$ is centered Gaussian with covariance matrix A. To see that, set $\zeta_t = \frac{1}{Q(i,j)}\mathbf{1}(\tau_{t-1}^* = i, \tau_t^* = j) - \frac{1}{Q(i,i)}\mathbf{1}(\tau_{t-1}^* = i, \tau_t^* = i)$ and $\xi_t = \frac{1}{Q(\alpha,\beta)}\mathbf{1}(\tau_{t-1}^* = \alpha, \tau_t^* = \beta) - \frac{1}{Q(\alpha,\alpha)}\mathbf{1}(\tau_{t-1}^* = \alpha, \tau_t^* = \alpha)$. Then $E(\zeta_t|\mathcal{F}_{t-1}) = E(\xi_t|\mathcal{F}_{t-1}) = 0$, so $E\left(\zeta_t\xi_k\right) = 0$ if $t\neq k$ and

$$
E\left(\zeta_t\xi_t\right) = \frac{\pi_i}{Q(i,j)}I_{i\alpha}I_{j\beta} + \frac{\pi_i}{Q(i,i)}I_{i\alpha}.
$$

As a result, if λ is a $m(m-1)$-dimensional vector, then one has

$$
\begin{aligned}
\mathrm{Var}(\lambda^\top \mathbb{W}_1^*) &= \sum_{(i,j)\in\mathcal{A}_m}\sum_{(\alpha,\beta)\in\mathcal{A}_m}\lambda(i,j)\lambda(\alpha,\beta)\left\{\frac{\pi_i}{Q(i,j)}I_{i\alpha}I_{j\beta} + \frac{\pi_i}{Q(i,i)}I_{i\alpha}\right\}\\
&= \sum_{(i,j)\in\mathcal{A}_m}\frac{\pi_i\lambda^2(i,j)}{Q(i,j)} + \sum_{i=1}^m \frac{\pi_i}{Q(i,i)}\left\{\sum_{j\neq i}\lambda(i,j)\right\}^2\\
&= \lambda^\top A\lambda.
\end{aligned} \tag{27}
$$

Next, set

$$
\mathbb{W}_{n,2}^* = n^{-1/2}\sum_{t=1}^n \frac{\dot f(\tau_t^*, Y_{t-1}^*, Y_t^*)}{f(\tau_t^*, Y_{t-1}^*, Y_t^*)},
$$

and remark that $\frac{\dot f(\tau_t^*, Y_{t-1}^*, Y_t^*)}{f(\tau_t^*, Y_{t-1}^*, Y_t^*)}$ is a difference martingale since for any $i\in\{1,\ldots,m\}$ and any $z\in\mathbb{R}^d$,

$$
\int \dot f(i,z,y)dy = 0. \tag{28}
$$

Furthermore,

$$
\mathcal{J}_{f,n} = \sum_{t=1}^n\left\{\frac{\ddot f(\tau_t^*, Y_{t-1}^*, Y_t^*)}{f(\tau_t^*, Y_{t-1}^*, Y_t^*)} - \frac{\dot f(\tau_t^*, Y_{t-1}^*, Y_t^*)\dot f(\tau_t^*, Y_{t-1}^*, Y_t^*)^\top}{f^2(\tau_t^*, Y_{t-1}^*, Y_t^*)}\right\}
$$

converges in probability to $-cI_f$, where

$$\mathcal{J}_f = \sum_{i=1}^{m}\sum_{j-1}^{m} \pi_i Q(i,j) \int\int \frac{\dot f(j,z,y)\dot f(j,z,y)^\top}{f(j,z,y)} g(i,z)dzdy,$$

the sequence $(\tau_t^*, Y_{t-1}^*, Y_t^*)$ being ergodic.

By the central limit theorem for martingales, e.g., Durrett (1996), one may conclude that $\mathbb{W}_{n,2}^* \rightsquigarrow N(0,\mathcal{J}_f)$. One can also show that the covariance between $\mathbb{W}_{n,1}$ and $\mathbb{W}_{n,2}^*$ is zero, again because of (28). From the calculation in Section 4.2, it follows that

$$\begin{aligned}
\ell_{n,2} &= \sum_{t=1}^{n} \log\left\{ \frac{f_{\theta_n}(\tau_t^*, Y_{t-1}^*, Y_t^*)}{f(\tau_t^*, Y_{t-1}^*, Y_t^*)} \right\} \\
&= \Theta_n^\top \mathbb{W}_{n,2}^* - \frac{1}{2}\Theta_n^\top \mathcal{J}_{f,n}\Theta_n + o_P(1) \\
&= \Theta_n^\top \mathbb{W}_{n,2}^* - \frac{1}{2}\Theta_n^\top \mathcal{J}_{f}\Theta_n + o_P(1).
\end{aligned}$$

As a result, $\ell_n = \mathbb{Q}_n^\top \mathbb{W}_{n,1}^* - \frac{1}{2}\mathbb{Q}_n^\top A \mathbb{Q}_n + \Theta_n^\top \mathbb{W}_{n,2}^* - \frac{1}{2}\Theta_n^\top \mathcal{J}_f \Theta_n + o_P(1)$. Finally, by construction, $\mathbb{W}_{n,1}^*$ and $\mathbb{W}_{n,2}^*$ are independent of $(\mathbb{Q}_n, \Theta_n)$ and one may conclude that

$$(\mathbb{Q}_n, \Theta_n, \mathbb{W}_{n,1}^*, \mathbb{W}_{n,1}^*) \rightsquigarrow (\mathbb{Q}, \Theta, \mathbb{W}_1^*, \mathbb{W}_2^*),$$

where the latter is a centered Gaussian vector with $(\mathbb{W}_1^*, \mathbb{W}_2^*)$ independent of $(\mathbb{Q}, \Theta)$. Furthermore, since the covariance between $\mathbb{W}_{n,1}^*$ and $\mathbb{W}_{n,2}^*$ is zero, it follows that $\mathbb{W}_1^*$ and $\mathbb{W}_2^*$ are independent, with $\mathbb{W}_1^* \sim N(0,A)$ and $\mathbb{W}_2^* \sim N(0,\mathcal{J}_f)$. As a result,

$$(\mathbb{Q}_n, \Theta_n, \mathbb{W}_{n,1}^*, \mathbb{W}_{n,1}^*, \ell_n) \rightsquigarrow (\mathbb{Q}, \Theta, \mathbb{W}_1^*, \mathbb{W}_2^*, \ell),$$

with $\ell = \mathbb{Q}^\top \mathbb{W}_1^* - \frac{1}{2}\mathbb{Q}^\top A \mathbb{Q} + \Theta^\top \mathbb{W}_2^* - \frac{1}{2}\Theta^\top \mathcal{J}_f \Theta$. Furthermore, using (27) and independence of $\mathbb{W}_1^*$ and $\mathbb{W}_2^*$, one has $E\{\exp(\ell)|\mathbb{Q},\Theta\} = 1$, showing that $E\{\exp(\ell)\} = 1$.

It remains now to show that the MLE estimators (Q_n, θ_n) based of the EM algorithm, and depending only on $Y_1, \ldots, Y_n$ are regular. According to Cappé et al. (2005), one has

$$\begin{bmatrix} \mathbb{Q}_n \\ \Theta_n \end{bmatrix} = J^{-1} Z_n = J^{-1} \begin{bmatrix} Z_{n,1} \\ Z_{n,2} \end{bmatrix},$$

where $Z_{n,1}$ is a $m(m-1)$-dimensional vector with component $Z_{n,1}(i,j)$,$(i,j) \in \mathcal{A}_m$ given by

$$Z_{n,1}(i,j) = n^{-1/2} \sum_{t=1}^{n} \eta_t(i,j),$$

where the martingale differences η_t are defined by

$$\eta_t(i,j) = E\left[\left\{ \frac{N_t(i,j)}{Q(i,j)} - \frac{N_t(i,i)}{Q(i,i)} \right\} \middle| \mathcal{Y}_t \right] - E\left[\left\{ \frac{N_{t-1}(i,j)}{Q(i,j)} - \frac{N_{t-1}(i,i)}{Q(i,i)} \right\} \middle| \mathcal{Y}_{t-1} \right],$$

and $Z_{n,2} = n^{-1/2}\sum_{t=1}^n \Xi_t$, where the martingale differences Ξ_t are defined by

$$\Xi_t = E\left[\sum_{k=1}^{t} \frac{\dot f(\tau_k, Y_{k-1}, Y_k)}{f(\tau_k, Y_{k-1}, Y_k)} \middle| \mathcal{Y}_t \right] - E\left[\sum_{k=1}^{t-1} \frac{\dot f(\tau_k, Y_{k-1}, Y_k)}{f(\tau_k, Y_{k-1}, Y_k)} \middle| \mathcal{Y}_{t-1} \right],$$

where $\mathcal{Y}_t$ is the σ-algebra generated by $\tau_0, Y_0, \ldots, Y_t$. Moreover $Z_n \rightsquigarrow Z$, where Z is centered Gaussian with covariance matrix J. Therefore, to show that (Q_n, θ_n) are regular, it suffices to show that $E\left(\mathbb{W}Z^\top\right) = J$. First, note that $E\{N_n(i,j)|\mathcal{F}_t\} = N_t(i,j) + \sum_{l=1}^m \sum_{k=t+1}^n \mathbf{1}(\tau_t = l)\left(Q^{k-1-t}\right)_{li} Q_{ij}$. As a result, since $\mathcal{Y}_t \subset \mathcal{F}_t$, it follows that $E\{\mathbb{W}_{n,1}(i,j)|\mathcal{Y}_t\} - E\{\mathbb{W}_{n,1}(i,j)|\mathcal{Y}_{t-1}\} = n^{-1/2}\eta_t(i,j)$, $(i,j) \in \mathcal{A}_m$. Therefore, $Z_{n,1} = E(\mathbb{W}_{n,1}|\mathcal{Y}_n)$. Similarly, $E\{\mathbb{W}_{n,2}|\mathcal{Y}_t\} - E\{\mathbb{W}_{n,2}|\mathcal{Y}_{t-1}\} = n^{-1/2}\Xi_t$, so $Z_{n,2} = E(\mathbb{W}_{n,2}|\mathcal{Y}_n)$. Combining the two equalities, one obtains that $Z_n = E(\mathbb{W}_n|\mathcal{Y}_n)$. Hence, as $n \to \infty$, $E\left(\mathbb{W}_n Z_n^\top\right) = E\left(Z_n Z_n^\top\right) \to J$, completing the proof. □

B.4 Proof of Proposition 2

It follows that for $j,k \in \{1,\ldots,s_2\}$,

$$\begin{aligned}\partial_{\theta_{2,j}} L &= -2\sum_{t=1}^{n}(y_t-\mu_t)^\top h_t^{-1}\partial_{\theta_{2,j}}\mu_t - \sum_{t=1}^{n}(y_t-\mu_t)^\top h_t^{-1}\partial_{\theta_{2,j}}h_t h_t^{-1}(y_t-\mu_t)\\ &\quad + \sum_{t=1}^{n}\frac{\partial_{\theta_{2,j}}|h_t|}{|h_t|}\end{aligned}$$

and

$$\begin{aligned}\partial_{\theta_{2,j}}\partial_{\theta_{2,k}} L &= -2\sum_{t=1}^{n}(y_t-\mu_t)^\top h_t^{-1}\partial_{\theta_{2,j}}\partial_{\theta_{2,k}}\mu_t + 2\sum_{t=1}^{n}(y_t-\mu_t)^\top h_t^{-1}\partial_{\theta_{2,j}}h_t h_t^{-1}\partial_{\theta_{2,k}}\mu_t\\ &\quad +2\sum_{t=1}^{n}(y_t-\mu_t)^\top h_t^{-1}\partial_{\theta_{2,k}}h_t h_t^{-1}\partial_{\theta_{2,j}}\mu_t + 2\sum_{t=1}^{n}\partial_{\theta_{2,j}}\mu_t^\top h_t^{-1}\partial_{\theta_{2,k}}\mu_t\\ &\quad + \sum_{t=1}^{n}(y_t-\mu_t)^\top h_t^{-1}\left\{2\partial_{\theta_{2,j}}h_t h_t^{-1}\partial_{\theta_{2,k}}h_t - \partial_{\theta_{2,j}}\partial_{\theta_{2,k}}h_t\right\}h_t^{-1}(y_t-\mu_t)\\ &\quad + \sum_{t=1}^{n}\left[\frac{\partial_{\theta_{2,j}}\partial_{\theta_{2,k}}|h_t|}{|h_t|} - \frac{\partial_{\theta_{2,j}}|h_t|\partial_{\theta_{2,k}}|h_t|}{|h_t|^2}\right].\end{aligned}$$

Note that since $\mathrm{Trace}\left\{h_t^{-1}\partial_{\theta_{2,j}}\right\} = \partial_{\theta_{2,j}}\log|h_t|$, it follows that

$$\frac{\partial_{\theta_{2,j}}\partial_{\theta_{2,k}}|h_t|}{|h_t|} - \frac{\partial_{\theta_{2,j}}|h_t|\partial_{\theta_{2,k}}|h_t|}{|h_t|^2} = \mathrm{Trace}\left\{h_t^{-1}\partial_{\theta_{2,j}}\partial_{\theta_{2,k}}h_t - h_t^{-1}\partial_{\theta_{2,k}}h_t h_t^{-1}\partial_{\theta_{2,j}}h_t\right\}.$$

Set $\zeta_{t,j} = 2\varepsilon_t^\top\sigma_t^{-1}\partial_{\theta_{2,j}}\mu_t + \mathrm{Trace}\left\{\sigma_t^{-1}\partial_{\theta_{2,j}}h_t\left(\sigma_t^\top\right)^{-1}\left(\varepsilon_t\varepsilon_t^\top - I\right)\right\}$, and

$$\begin{aligned}V_{t,jk} &= 2\varepsilon_t^\top\sigma_t^{-1}\left\{\partial_{\theta_{2,j}}h_t h_t^{-1}\partial_{\theta_{2,k}}\mu_t + \partial_{\theta_{2,k}}h_t h_t^{-1}\partial_{\theta_{2,j}}\mu_t - \partial_{\theta_{2,j}}\partial_{\theta_{2,k}}\mu_t\right\}\\ &\quad +2\partial_{\theta_{2,j}}\mu_t h_t^{-1}\partial_{\theta_{2,k}}\mu_t + \mathrm{Trace}\left\{h_t^{-1}\partial_{\theta_{2,k}}h_t h_t^{-1}\partial_{\theta_{2,j}}h_t\right\}\\ &\quad +\mathrm{Trace}\left\{\sigma_t^{-1}\left(2\partial_{\theta_{2,j}}h_t h_t^{-1}\partial_{\theta_{2,k}}h_t - \partial_{\theta_{2,j}}\partial_{\theta_{2,k}}h_t\right)\left(\sigma_t^\top\right)^{-1}\left(\varepsilon_t\varepsilon_t^\top - I\right)\right\},\end{aligned}$$

and note that $\mathrm{E}(\zeta_t|\mathcal{F}_{t-1}) = 0$ and

$$\mathrm{E}(V_{t,jk}|\mathcal{F}_{t-1}) = 2\partial_{\theta_{2,j}}\mu_t h_t^{-1}\partial_{\theta_{2,k}}\mu_t + \mathrm{Trace}\left\{h_t^{-1}\partial_{\theta_{2,j}}h_t h_t^{-1}\partial_{\theta_{2,k}}h_t\right\},$$

$1 \le j,k \le s_2$. Assuming that $\frac{1}{n}\sum_{t=1}^n V_t \xrightarrow{Pr} V$ and $\frac{1}{n}\sum_{t=1}^n E(V_t|\mathcal{F}_{t-1}) \xrightarrow{Pr} V$, then $\Theta_{n,2} = n^{-1/2}\sum_{t=1}^n V^{-1}\zeta_t + o_P(1)$.

Finally, $\mathrm{E}\left(\zeta_t\xi_{t,1}|\mathcal{F}_{t-1}\right) = 0$ and

$$\begin{aligned}\mathrm{E}\left(\zeta_{t,j}\xi_{t,2,k}|\mathcal{F}_{t-1}\right) &= 2\partial_{\theta_{2,j}}\mu_t h_t^{-1}\partial_{\theta_{2,k}}\mu_t + 2\mathrm{Trace}\left(h_t^{-1}\partial_{\theta_{2,j}}h_t h_t^{-1}\partial_{\theta_{2,k}}\sigma_t\sigma_t^\top\right)\\ &= \mathrm{E}(V_{t,jk}|\mathcal{F}_{t-1}),\end{aligned}$$

since $h_t^{-1}\partial_{\theta_{2,j}}h_t h_t^{-1}$ is symmetric. As a result, if $(\Theta_{n,2},\mathbb{W}_n) \rightsquigarrow (\Theta_2,\mathbb{W})$, then $\mathrm{E}\left(\Theta_2\mathbb{W}^\top\right) = (0,I)^\top$. Hence, if $\theta_{n,1}$ is a regular estimator of θ_1, then θ_n is a regular estimator of θ. □

C Central limit theorem for dependent variables

The following theorem is proven in Durrett (1996).

Theorem 4 (Lindeberg-Feller CLT for Martingales) *Suppose that $E(X_{n,m}|\mathcal{F}_{n,m-1}) = 0$ and set $V_{n,k} = \sum_{m=1}^k E(X_{n,m}^2|\mathcal{F}_{n,m-1})$. Set $S_n(t) = \sum_{m=1}^{[nt]} X_{n,m}$, $t \in [0,1]$. If, for every $t \in [0,1]$, $V_{n,[nt]} \xrightarrow{Pr} t$ and $\sum_{m=1}^n \mathrm{E}\left\{X_{n,m}^2\mathbf{1}(|X_{n,m}| > \epsilon)|\mathcal{F}_{n,m-1}\right\} \xrightarrow{Pr} 0$, for every $\epsilon > 0$, then $S_n \rightsquigarrow \beta$, where β is a Brownian motion, so $S_n(1) \rightsquigarrow \beta(1) \sim N(0,1)$.*

References

Andrews, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica*, 65(5):1097–1128.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics*, 85(3):531–549.

Bai, J. and Chen, Z. (2008). Testing multivariate distributions in garch models. *Journal of Econometrics*, 143(1):19 – 36.

Beran, R. and Millar, P. W. (1987). Stochastic estimation and testing. *Ann. Statist.*, 15(3):1131–1154.

Beran, R. and Millar, P. W. (1989). A stochastic minimum distance test for multivariate parametric models. *Ann. Statist.*, 17:125–140.

Beran, R. J., Le Cam, L., and Millar, P. W. (1987). Convergence of stochastic empirical measures. *J. Multivariate Anal.*, 23(1):159–168.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144 (electronic). Update of, and a supplement to, the 1986 original.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York.

Chen, X. and Fan, Y. (2006a). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125–154.

Chen, X. and Fan, Y. (2006b). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130:307–335.

Corradi, V. and Swanson, N. R. (2006). Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics*, 133(2):779–806.

Delgado, M. A. and Stute, W. (2008). Distribution-free specification tests of conditional models. *Journal of Econometrics*, 143(1):37–55. Specification testing.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–883.

Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics. *J. Econometrics*, 133(2):443–477.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.*, 1(2):279–290.

Durrett, R. (1996). *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552.

Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. H. Poincaré Sect. B*, 44:1096–1127.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Omnibus goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom.*, 44:199–213.

Ghoudi, K. and Rémillard, B. (2004). Empirical processes based on pseudo-observations. II. The multivariate case. In *Asymptotic Methods in Stochastics*, volume 44 of *Fields Inst. Commun.*, pages 381–406. Amer. Math. Soc., Providence, RI.

Khmaladze, V. (1988). An innovation approach to goodness-of-fit tests in $\mathbf{R}^m$. *Ann. Statist.*, 16(4):1503–1516.

Khmaladze, V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.*, 21(2):798–829.

Kojadinovic, I. and Yan, J. (2010). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Stat. Comput.*

Li, B. (2009). Asymptotically distribution-free goodness-of-fit testing: A unifying view. *Econometric Reviews*, 28(6):632–657.

Li, F. and Tkacz, G. (2006). A consistent bootstrap test for conditional density functions with time-series data. *Journal of Econometrics*, 133(2):863–886.

Papageorgiou, N., Rémillard, B., and Hocquard, A. (2008). Replicating the properties of hedge fund returns. *Journal of Alternative Invesments*, 11:8–38.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Rémillard, B. (2010). Goodness-of-fit tests for copulas of multivariate time series. Technical report, SSRN Working Paper Series No. 1729982.

Rémillard, B. (2011). The multiplier central limit theorem for pseudo-observations and its applications. Technical Report G-2011-??, GERAD.

Rémillard, B., Hocquard, A., and Papageorgiou, N. A. (2010). Option Pricing and Dynamic Discrete Time Hedging for Regime-Switching Geometric Random Walks Models. Technical report, SSRN Working Paper Series No. 1591146.

Rémillard, B., Papageorgiou, N., and Soustra, F. (2011). Copula-based semiparametric models for multivariate time series. Technical report, SSRN Working Paper Series No. 1960564.

Rémillard, B. and Scaillet, O. (2009). Testing for equality between two copulas. *J. Multivariate Anal.*, 100:377–386.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23:470-472.

Scaillet, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. *Canad. J. Statist.*, 33(3):415–427.

Stute, W., Gonzáles Manteiga, W., and Presedo Quindimil, M. (1993). Bootstrap based goodness-of-fit tests. *Metrika*, 40:243–256.

van den Goorbergh, R. W. J., Genest, C., and Werker, B. J. M. (2005). Bivariate option pricing using dynamic copula models. *Insurance: Mathematics and Economics*, 37:101–114.

van der Vaart, A. and Wellner, J. (2007). Empirical processes indexed by estimated functions. In *Asymptotics: Particles, Processes and Inverse Problems*, volume 55 of *IMS Lectures Notes – Monograph Series*, pages 234–252. Institute of Mathematical Statistics.

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