Econometrics Journal (2018), volume **01**, pp. 1–13.
Article No. ectj??????

Supplement: Testing for constant correlation of filtered series under structural change

Matei Demetrescu† and Dominik Wied‡

†*Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Germany*
E-mail: mdeme@stat-econ.uni-kiel.de
‡*Institute for Econometrics and Statistics, University of Cologne, Germany*
E-mail: dwied@uni-koeln.de

Received:

A. PROOFS

Before providing the main proofs, we state and prove an auxiliary result.

Lemma A.1. *It holds under Assumptions 2.1 and 3.1 that*

$$\hat{\boldsymbol{\theta}}_i\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_i\left(\lambda_0\right) = o_p\left(n^{-1/2}\right), i = 1, 2,$$

as $n \to \infty$, provided that $\boldsymbol{\theta}_{1,0} \neq \boldsymbol{\theta}_{2,0}$.

Proof of Lemma A.1

Let us first discuss the behavior of

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_1\left(\lambda_0\right) &= \left(\sum_{j=1}^{\hat{\lambda}n} B'_{j,n} W_n \sum_{j=1}^{\hat{\lambda}n} B_{j,n}\right)^{-1} \sum_{j=1}^{\hat{\lambda}n} B'_{j,n} W_n \sum_{j=1}^{\hat{\lambda}n} A_{j,n} + R_{\hat{\lambda}n,n} \\ &\quad - \left(\sum_{j=1}^{\lambda_0 n} B'_{j,n} W_n \sum_{j=1}^{\lambda_0 n} B_{j,n}\right)^{-1} \sum_{j=1}^{\lambda_0 n} B'_{j,n} W_n \sum_{j=1}^{\lambda_0 n} A_{j,n} - R_{\lambda_0 n,n} \\ &= P_n^{-1}\left(\hat{\lambda}\right) Q_n\left(\hat{\lambda}\right) - P_n^{-1}\left(\lambda_0\right) Q_n\left(\lambda_0\right) + R_{\hat{\lambda}n,n} - R_{\lambda_0 n,n},\end{aligned}$$

where $P_n\left(\lambda\right) = \sum_{j=1}^{\lambda n} B'_{j,n} W_n \sum_{j=1}^{\lambda n} B_{j,n}$ and $Q_n\left(\lambda\right) = \sum_{j=1}^{\lambda n} B'_{j,n} W_n \sum_{j=1}^{\lambda n} A_{j,n}$, such that

$$P_n^{-1}\left(\lambda_0\right) = O_p\left(n^{-2}\right) \qquad \text{and} \qquad Q_n\left(\lambda_0\right) = O_p\left(n^{3/2}\right)$$

given the behavior of the individual components from Assumption 2.1 and 3.1. Since both λ_0 and $\hat{\lambda}$ (w.p. 1) are interior points of $[0,1]$, we also have from Assumption 3.1 that

$$\left|R_{\hat{\lambda}n,n} - R_{\lambda_0 n,n}\right| = o_p\left(n^{-1/2}\right)$$

for either $\hat{\lambda} \leq \lambda_0$ or $\hat{\lambda} > \lambda_0$. Furthermore,

$$\begin{aligned}&\left\|P_n^{-1}\left(\hat{\lambda}\right)Q_n\left(\hat{\lambda}\right) - P_n^{-1}\left(\lambda_0\right)Q_n\left(\lambda_0\right)\right\| \\ &\quad \leq \left\|P_n^{-1}\left(\hat{\lambda}\right) - P_n^{-1}\left(\lambda_0\right)\right\| \left\|Q_n\left(\hat{\lambda}\right)\right\| + \left\|P_n^{-1}\left(\lambda_0\right)\right\| \left\|Q_n\left(\hat{\lambda}\right) - Q_n\left(\lambda_0\right)\right\|.\end{aligned}$$

To assess $\left\|Q_n\left(\hat{\lambda}\right) - Q_n\left(\lambda_0\right)\right\|$, write

$$\left\|Q_n\left(\hat{\lambda}\right) - Q_n\left(\lambda_0\right)\right\| \leq \left\|\sum_{j=1}^{\hat{\lambda}n} B_{j,n}\right\| \|W_n\| \left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} A_{j,n}\right\| + \left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} B_{j,n}\right\| \|W_n\| \left\|\sum_{j=1}^{\lambda_0 n} A_{j,n}\right\|$$

where we make the convention that $\sum_{j=\hat{\lambda}n}^{\lambda_0 n} = -\sum_{j=\lambda_0 n}^{\hat{\lambda}n}$ if $\hat{\lambda} > \lambda$, such that

$$\left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} A_{j,n}\right\| \leq n\left|\lambda_0 - \hat{\lambda}\right| \sup_{1\leq j\leq n} \|A_{j,n}\| = O_p\left(n^{1/(2+\alpha)}\right).$$

(The uniform $L_{2+\alpha}$ boundedness of $A_{j,n}$ has been used to derive the magnitude of the maximum.) We the have analogously that

$$\left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} B_{j,n}\right\| = O_p\left(n^{1/(1+\alpha)}\right),$$

such that

$$\left\|\sum_{j=1}^{\hat{\lambda}n} B_{j,n}\right\| \leq \left\|\sum_{j=1}^{\lambda_0 n} B_{j,n}\right\| + \left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} B_{j,n}\right\| = O_p\left(n^{1/(1+\alpha)}\right) = O_p(n)$$

and, summing up, that

$$\left\|Q_n\left(\hat{\lambda}\right) - Q_n\left(\lambda_0\right)\right\| = O_p\left(\max\left\{n^{1+1/(2+\alpha)}, n^{1/2+1/(1+\alpha)}\right\}\right) = o_p\left(n^{3/2}\right).$$

Furthermore, this implies that

$$\left\|Q_n\left(\hat{\lambda}\right)\right\| \leq \left\|Q_n\left(\lambda_0\right)\right\| + \left\|Q_n\left(\hat{\lambda}\right) - Q_n\left(\lambda_0\right)\right\| = O_p\left(n^{3/2}\right).$$

Now, Lütkepohl (1996, Section 8.4.1, Eq. (11c)) implies that

$$\left\|n^2 P_n^{-1}\left(\hat{\lambda}\right) - n^2 P_n^{-1}\left(\lambda_0\right)\right\| \leq \left\|n^2 P_n^{-1}\left(\lambda_0\right)\right\| \frac{\left\|n^2 P_n^{-1}\left(\lambda_0\right)\right\| \left\|\frac{1}{n^2}P_n\left(\hat{\lambda}\right) - \frac{1}{n^2}P_n\left(\lambda_0\right)\right\|}{1 - \left\|n^2 P_n^{-1}\left(\lambda_0\right)\right\| \left\|\frac{1}{n^2}P_n\left(\hat{\lambda}\right) - \frac{1}{n^2}P_n\left(\lambda_0\right)\right\|}$$

if $\left\|n^2 P_n^{-1}\left(\lambda_0\right)\right\| \left\|\frac{1}{n^2}P_n\left(\hat{\lambda}\right) - \frac{1}{n^2}P_n\left(\lambda_0\right)\right\| < 1$ and $\left\|n^2 P_n\left(\lambda_0\right)\left(\frac{1}{n^2}P_n\left(\hat{\lambda}\right) - \frac{1}{n^2}P_n\left(\lambda_0\right)\right)\right\| < 1$, where

$$\left\|P_n\left(\hat{\lambda}\right) - P_n\left(\lambda_0\right)\right\| \leq 2\left\|\sum_{j=1}^{\hat{\lambda}n} B_{j,n}\right\| \|W_n\| \left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} B_{j,n}\right\| = O_p\left(n^{1+1/(1+\alpha)}\right) = o_p\left(n^2\right).$$

Consequently, $\left(\frac{1}{n^2}P_n\left(\hat{\lambda}\right)-\frac{1}{n^2}P_n\left(\lambda_0\right)\right)\xrightarrow{p}0$, so that the two conditions are fulfilled with probability approaching one and we have that

$$\left\|P_n^{-1}\left(\hat{\lambda}\right)-P_n^{-1}\left(\lambda_0\right)\right\|=o_p\left(n^{-2}\right).$$

Summing up,

$$\left\|P_n^{-1}\left(\hat{\lambda}\right)Q_n\left(\hat{\lambda}\right)-P_n^{-1}\left(\lambda_0\right)Q_n\left(\lambda_0\right)\right\|=o_p\left(n^{-1/2}\right)$$

and

$$\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)-\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)=o_p\left(n^{-1/2}\right).$$

The result for $\hat{\boldsymbol{\theta}}_2\left(\hat{\lambda}\right)-\hat{\boldsymbol{\theta}}_2\left(\lambda_0\right)$ is derived analogously and we omit the details.

Proof of Proposition 3.1

Use the mean value theorem to expand the vector function $\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\right)$ elementwise about $\boldsymbol{\theta}_0$ to obtain with $\boldsymbol{Z}_t^*=\boldsymbol{h}\left(\boldsymbol{X}_t,\ldots;\boldsymbol{\theta}^*\right)$

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}g_l\left(\hat{\boldsymbol{Z}}_t\right)=&\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}g_l\left(\boldsymbol{Z}_t\right)+\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\\&+\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}-\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\end{aligned}$$

where $\boldsymbol{\theta}^*$ is a convex combination of $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$. Since $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0=O_p\left(n^{-1/2}\right)$, $\boldsymbol{\theta}^*$ belongs to a $\sqrt{n}$-neighbourhood of $\boldsymbol{\theta}_0$ and thus to Φ_n; we pick $\boldsymbol{\theta}_t^*=\boldsymbol{\theta}^*$ $1\le t\le n$, and Assumption 2.1 ensures uniform negligibility of the third term on the r.h.s. for $l=1,\ldots,L$,

$$\begin{aligned}&\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}-\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\right\|\\&\qquad\le\left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\right\|\sup_{\boldsymbol{\theta}^*,t}\left\|\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}-\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|\\&\qquad\xrightarrow{p}0.\end{aligned}$$

The first result follows with Assumption 2.1 and the continuous mapping theorem (CMT).

Let us now examine the case of the recursive estimation scheme. Since $g_l\left(\tilde{\boldsymbol{Z}}_t\right)$ is a function of $\hat{\boldsymbol{\theta}}_t$, we have n convex combinations $\boldsymbol{\theta}_t^*$ $(t=1,\ldots,n)$ of $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_t$ in the mean-value expansion about $\boldsymbol{\theta}_0$, leading to

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}g_l\left(\tilde{\boldsymbol{Z}}_t\right)=&\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}g_l\left(\boldsymbol{Z}_t\right)+\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\left(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_0\right)\\&+\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^*}-\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\left(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_0\right).\end{aligned}$$

Since $\sup_{s\in[\epsilon,1]}\left\|\hat{\boldsymbol{\theta}}_{[sn]}-\boldsymbol{\theta}_0\right\|=O_p\left(n^{-1/2}\right)$ when $\boldsymbol{\Psi}$ is bounded in probability, the third term on the r.h.s. is immediately seen to vanish like before, such that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\boldsymbol{g}\left(\boldsymbol{Z}_t\right)+\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left.\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\left(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_0\right)+o_p\left(1\right)$$

where the o_p term is uniform on $[\epsilon,1]$, and the result is completed with Assumption 2.1 and the CMT.

Proof of Proposition 3.2

The desired asymptotic equivalence follows for the case of full-sample estimation from the condition

$$\sup_{s\in[0,1]}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\hat{\lambda}\right)\right)-\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\lambda_0\right)\right)\right)\right|=o_p\left(1\right).$$

Examining $\hat{\boldsymbol{Z}}_t\left(\hat{\lambda}\right)$, we have (writing explicitly only the dependence on $\boldsymbol{X}_t$ to simplify notation)

$$\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\hat{\lambda}\right)\right)=\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)\left(1-D_{t,\hat{\lambda}}\right)+\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_2\left(\hat{\lambda}\right)\right)\right)D_{t,\hat{\lambda}}$$

and analogously

$$\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\lambda_0\right)\right)=\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)\left(1-D_{t,\lambda_0}\right)+\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_2\left(\lambda_0\right)\right)\right)D_{t,\lambda_0}$$

such that

$$\begin{aligned}\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\hat{\lambda}\right)\right)-\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\lambda_0\right)\right)&=\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)\left(1-D_{t,\hat{\lambda}}\right)-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)\left(1-D_{t,\lambda_0}\right)\\&\quad+\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_2\left(\hat{\lambda}\right)\right)\right)D_{t,\hat{\lambda}}-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_2\left(\lambda_0\right)\right)\right)D_{t,\lambda_0}\\&=M_t+N_t.\end{aligned}$$

Then,

$$\begin{aligned}M_t&=\left(\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)\right)\left(1-D_{t,\hat{\lambda}}\right)+\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)\left(D_{t,\lambda_0}-D_{t,\hat{\lambda}}\right)\\&=M_{1t}+M_{2t}.\end{aligned}$$

Now, $D_{t,\hat{\lambda}}$ is either zero or unity, so we may focus on $\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)$ in discussing cumulated sums of M_{1t}, for which we resort to the mean value theorem elementwise and obtain like in the proof of Proposition 3.1 that, for each l, and $t\leq\lambda_0 n$,

$$\begin{aligned}g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)&=g_l\left(\boldsymbol{Z}_t\right)+\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\left(\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)-\boldsymbol{\theta}_{1,0}\right)\\&\quad+\left(\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^*}-\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\right)\left(\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)-\boldsymbol{\theta}_{1,0}\right)\end{aligned}$$

and

$$g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)=g_l(\boldsymbol{Z}_t)+\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\left(\hat{\boldsymbol{\theta}}_1(\lambda_0)-\boldsymbol{\theta}_{1,0}\right)$$
$$+\left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^{0*}}-\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\right)\left(\hat{\boldsymbol{\theta}}_1(\lambda_0)-\boldsymbol{\theta}_{1,0}\right)$$

for suitable $\boldsymbol{\theta}_t^*$ ($\boldsymbol{\theta}_t^{0*}$) between $\boldsymbol{\theta}_{1,0}$ and $\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)$ (between $\boldsymbol{\theta}_{1,0}$ and $\hat{\boldsymbol{\theta}}_1(\lambda_0)$), such that, for all $1\le t\le\lambda_0 n$,

$$\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)=\left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\left(\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)-\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)+o_p\left(\frac{1}{\sqrt{n}}\right)$$

where the $o_p\left(\frac{1}{\sqrt{n}}\right)$ term is uniform in t following Assumption 3.1. For $t>\lambda_0 n$, we expand $g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)$ and $g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)$ about the same $\boldsymbol{\theta}_{1,0}$, but note that $\boldsymbol{h}(\boldsymbol{X}_t,\boldsymbol{\theta}_{1,0})\neq\boldsymbol{Z}_t$ for t in the second regime. We obtain however similarly

$$\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)=\left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{X}_t,\boldsymbol{\theta}_{1,0})}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\left(\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)-\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)+o_p\left(\frac{1}{\sqrt{n}}\right)$$

thanks to Assumption 3.1. Using now Lemma A.1, we obtain immediately

$$\sup_{s\in[0,1]}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}M_{1t}\right|=o_p(1).$$

For M_{2t} we note that $\sum\left|D_{t,\hat{\lambda}}-D_{t,\lambda_0}\right|=O_p(1)$ since $\hat{\lambda}-\lambda_0=O_p\left(n^{-1}\right)$. Then, for each $t<\lambda_0 n$ and l, write again

$$g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)=g_l(\boldsymbol{Z}_t)+\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\left(\hat{\boldsymbol{\theta}}_1(\lambda_0)-\boldsymbol{\theta}_{1,0}\right)$$
$$+\left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}-\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\right)\left(\hat{\boldsymbol{\theta}}_1(\lambda_0)-\boldsymbol{\theta}_{1,0}\right)$$

where $\sup_{t=1,\ldots,n}|g_l(\boldsymbol{Z}_t)|=o_p(\sqrt{n})$ and $\sup_{t=1,\ldots,n}\left\|\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\right\|=o_p(n)$ thanks to Assumption 3.1, and the third summand on the r.h.s. can be dealt with using Assumption 3.1 such that

$$\sup_{t=1,\ldots,\lambda_0 n}\left\|\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)\right\|=o_p(\sqrt{n}).$$

For each $t\ge\lambda_0 n$ and l, we have like before

$$g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)=g_l(\boldsymbol{h}(\boldsymbol{X}_t,\boldsymbol{\theta}_{1,0}))+\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{X}_t,\boldsymbol{\theta}_{1,0})}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\left(\hat{\boldsymbol{\theta}}_1(\lambda_0)-\boldsymbol{\theta}_{1,0}\right)$$
$$+\left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{X}_t,\boldsymbol{\theta}^*)}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}-\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{X}_t,\boldsymbol{\theta}_{1,0})}\left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\right)\left(\hat{\boldsymbol{\theta}}_1(\lambda_0)-\boldsymbol{\theta}_{1,0}\right)$$

and Assumption 3.1 leads analogously to

$$\max_{\lambda_0 n\le t\le n}\left\|\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)\right\|=o_p(\sqrt{n})$$

such that, summing up,

$$\sup_{s\in[0,1]}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}M_{2t}\right| \leq \sup_{t=1,\dots,n}\left\|\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)\right\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|D_{t,\hat{\lambda}}-D_{t,\lambda_0}\right| = o_p(1).$$

The partial sums of N_t are evaluated in the same manner and the first result follows.

Analogously, one can show for the case of recursive estimation that, on $[\epsilon,\lambda_0]\cup[\lambda_0+\epsilon,1]$ for any $0<\epsilon<\min\{\lambda_0,1-\lambda_0\}$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_{t,\hat{\lambda}}\right)-\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right)\Rightarrow\Omega^{1/2}\boldsymbol{\Gamma}(s)+\left(\int_0^s\boldsymbol{\Theta}'_{\lambda_0}(r)\left(\Sigma_{\lambda_0}^{1/2}\right)'W'\Pi_{\lambda_0}(r)\left(\Pi'_{\lambda_0}(r)\,W\,\Pi_{\lambda_0}(r)\right)^{-1}\mathrm{d}\boldsymbol{\tau}'_{\lambda_0}(r)\right)'.$$

The result can be proved along the same lines as the result for full-sample estimation (but taking into account the fact that, at the beginning of the sample and after the break, the recursive estimator does not have proper asymptotics) and we omit the details.

B. FILTERING WITH BREAKS IN MARGINAL MEAN AND VARIANCE

Let us first consider testing the covariance of some bivariate $\boldsymbol{X}_t$ which has unknown mean but only the covariance (matrix) is subject to inference.

For the illustration, we take i.i.d. series $\boldsymbol{Z}_t$ in a location-scale model,

$$X_t=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}+\begin{pmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{pmatrix}\boldsymbol{Z}_t\quad\text{with}\quad\boldsymbol{Z}_t\sim\text{i.i.d.}\left(0,\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right).$$

Given that we work under i.i.d. sampling, the assumptions in Section 2 can easily be shown to hold, provided that enough moments of $\boldsymbol{Z}_t$ are finite and the parameter space is compact, so we do not spell out the details here to save space. Then,

$$g(\boldsymbol{z})=z_1z_2,\qquad\hat{\boldsymbol{Z}}_t=\boldsymbol{X}_t-\bar{\boldsymbol{X}}\qquad\text{and}\qquad\boldsymbol{h}(x)=\begin{pmatrix}x_1-\theta_1\\ x_2-\theta_2\end{pmatrix}$$

with $\hat{\theta}_1=\hat{\mu}_1$ and $\hat{\theta}_2=\hat{\mu}_2$. Hence

$$\frac{\partial g}{\partial z_1}=z_2\quad\frac{\partial g}{\partial z_2}=z_1,\qquad\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}=\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix},$$

$$\frac{1}{n}\sum_{t=1}^{[sn]}\left.\frac{\partial g}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial h}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}=\frac{1}{n}\sum_{t=1}^{[sn]}\left(-Z_{t2},-Z_{t1}\right)\Rightarrow\mathbf{0}.$$

Here the distribution is not asymptotically affected compared to the test based on $Z_{t,1}Z_{t,2}$.

Then again, if looking at the correlation ρ rather than the covariance of Z_{t1} and Z_{t2}, the residual effect is present. We have like before $g(\boldsymbol{z})=z_1z_2$, but, for $i=1,2$, we have that

$$\hat{Z}_{ti}=\frac{X_{ti}-\hat{\mu}_i}{\hat{\sigma}_i}$$

with $\hat{\mu}_i=\bar{X}_i$ and

$$\hat{\sigma}_i^2=\frac{1}{n}\sum_{t=1}^{n}\left(X_{ti}-\bar{X}_i\right)^2=\frac{1}{n}\sum_{t=1}^{n}\sigma_i^2\left(Z_{ti}-\bar{Z}_i\right)^2=\frac{1}{n}\sum_{t=1}^{n}\sigma_i^2Z_{ti}^2+O_p\left(n^{-1}\right),$$

such that, with $\theta_3=\sigma_1^2$ and $\theta_4=\sigma_2^2$, we write $\boldsymbol{h}(\boldsymbol{x})=\left(\frac{x_1-\theta_1}{\sqrt{\theta_3}}\;\frac{x_2-\theta_2}{\sqrt{\theta_4}}\right)'$. While $\frac{\partial g}{\partial\boldsymbol{z}}$ is the same as in the case of the covariance,

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -\frac{1}{\sigma_1} & 0 & -\frac{1}{2}\frac{x_1-\mu_1}{\sigma_1^3} & 0 \\ 0 & -\frac{1}{\sigma_2} & 0 & -\frac{1}{2}\frac{x_2-\mu_2}{\sigma_2^3} \end{pmatrix}$$

such that

$$\frac{1}{n}\sum_{t=1}^{[sn]} \left.\frac{\partial g}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n}\sum_{t=1}^{[sn]} (Z_{t2}, Z_{t1}) \begin{pmatrix} -\frac{1}{\sigma_{1,0}} & 0 & -\frac{1}{2}\frac{Z_{t1}-\mu_{1,0}}{\sigma_{1,0}^3} & 0 \\ 0 & -\frac{1}{\sigma_{2,0}} & 0 & -\frac{1}{2}\frac{Z_{t2}-\mu_{2,0}}{\sigma_{2,0}^3} \end{pmatrix}$$

$$\Rightarrow -\rho_0\, s \begin{pmatrix} 0 & 0 & \frac{1}{2\sigma_{1,0}^3} & \frac{1}{2\sigma_{2,0}^3} \end{pmatrix} \equiv \boldsymbol{\tau}(s)$$

and variance estimation matters whenever the correlation is nonzero, but demeaning does not. Kicking out the zero elements, $\boldsymbol{\tau}(s) = -\rho_0\, s \begin{pmatrix} \frac{1}{2\sigma_{1,0}^2} & \frac{1}{2\sigma_{2,0}^2} \end{pmatrix}$; the relevant Brownian motion is

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]} \begin{pmatrix} Z_{t1}Z_{t2} - \rho_0 \\ \sigma_{1,0}^2 Z_{t1}^2 - \sigma_{1,0}^2 \\ \sigma_{2,0}^2 Z_{t2}^2 - \sigma_{2,0}^2 \end{pmatrix} \Rightarrow \boldsymbol{\Psi}(s)$$

with quadratic covariation

$$[\boldsymbol{\Psi}](s) = s \begin{pmatrix} \mathrm{E}\left(Z_{t1}^2 Z_{t2}^2\right) - \rho_0^2 & \sigma_{1,0}^2\left(\mathrm{E}\left(Z_{t1}^3 Z_{t2}\right) - \rho_0\right) & \sigma_{2,0}^2\left(\mathrm{E}\left(Z_{t1} Z_{t2}^3\right) - \rho_0\right) \\ \sigma_{1,0}^2\left(\mathrm{E}\left(Z_{t1}^3 Z_{t2}\right) - \rho_0\right) & \sigma_{1,0}^4\left(\mu_{4,1,0} - 1\right) & \sigma_{1,0}^2\sigma_{2,0}^2\left(\mathrm{E}\left(Z_{t1}^2 Z_{t2}^2\right) - 1\right) \\ \sigma_{2,0}^2\left(\mathrm{E}\left(Z_{t1} Z_{t2}^3\right) - \rho_0\right) & \sigma_{1,0}^2\sigma_{2,0}^2\left(\mathrm{E}\left(Z_{t1}^2 Z_{t2}^2\right) - 1\right) & \sigma_{2,0}^4\left(\mu_{4,2,0} - 1\right) \end{pmatrix}.$$

If interested in tests on constant correlation, $\boldsymbol{\tau}$ is linear in s so the estimation effect cancels out.

This preliminary finding extends to the case of tests on the correlation if the breaks accounted for are only in the mean but not in the variance as follows. Let

$$\boldsymbol{X}_t = \boldsymbol{\mu}_{1,0}\left(1 - D_{t,\lambda_0}\right) + \boldsymbol{\mu}_{2,0} D_{t,\lambda_0} + \begin{pmatrix} \sigma_{1,0} & 0 \\ 0 & \sigma_{2,0} \end{pmatrix} \boldsymbol{Z}_t$$

with λ_0 known. We still have $g(\boldsymbol{z}) = z_1 z_2$, but

$$\hat{Z}_{ti} = \frac{X_{ti} - \hat{\mu}_{1,i}\left(1 - D_{t,\lambda}\right) - \hat{\mu}_{2,i} D_{t,\lambda}}{\hat{\sigma}_i} \tag{B.1}$$

such that, with $\theta_1 = \mu_1$, $\theta_2 = \mu_2$, $\theta_3 = \sigma_1^2$ and $\theta_4 = \sigma_2^2$, and defining for brevity $\bar{D}_{t,\lambda} = 1 - D_{t,\lambda}$, we obtain

$$\boldsymbol{h}_\lambda(\boldsymbol{x}) = \begin{pmatrix} \frac{x_1 - \theta_1 \bar{D}_{t,\lambda} - \theta_2 D_{t,\lambda}}{\sqrt{\theta_5}} \\ \frac{x_2 - \theta_3 \bar{D}_{t,\lambda} - \theta_4 D_{t,\lambda}}{\sqrt{\theta_6}} \end{pmatrix}.$$

While $\frac{\partial g}{\partial \boldsymbol{z}} = (z_2, z_1)$, we now have

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}} = -\begin{pmatrix} \frac{1}{\sigma_1}\bar{D}_{t,\lambda} & \frac{1}{\sigma_1}D_{t,\lambda} & 0 & 0 & \frac{1}{2}\frac{x_1-\mu_{1,1}\bar{D}_{t,\lambda}-\mu_{2,1}D_{t,\lambda}}{\sigma_1^3} & 0 \\ 0 & 0 & \frac{1}{\sigma_2}\bar{D}_{t,\lambda} & \frac{1}{\sigma_2}D_{t,\lambda} & 0 & \frac{1}{2}\frac{x_2-\mu_{2,1}-\mu_{2,2}D_{t,\lambda}}{\sigma_2^3} \end{pmatrix},$$

hence

$$\frac{1}{n}\sum_{t=1}^{[sn]} \left.\frac{\partial g}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\frac{1}{n}\sum_{t=1}^{[sn]} (Z_{t2}, Z_{t1}) \begin{pmatrix} \frac{1}{\sigma_{1,0}}\bar{D}_{t,\lambda_0} & \frac{1}{\sigma_{1,0}}D_{t,\lambda_0} & 0 & 0 & \frac{1}{2}\frac{Z_{t1}}{\sigma_{1,0}^2} & 0 \\ 0 & 0 & \frac{1}{\sigma_{2,0}}\bar{D}_{t,\lambda_0} & \frac{1}{\sigma_{2,0}}D_{t,\lambda_0} & 0 & \frac{1}{2}\frac{Z_{t2}}{\sigma_{2,0}^2} \end{pmatrix}$$

$$\Rightarrow -\rho_0\, s \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{2\sigma_{1,0}^2} & \frac{1}{2\sigma_{2,0}^2} \end{pmatrix} \equiv \boldsymbol{\tau}_{\lambda_0}(s)$$

and only the variance estimation has an effect on the limiting behavior of the partial sums, which would cancel out if testing the constancy of the correlation. The relevant Brownian motion is the same as for demeaning only, and breaks in the mean (accounted for) do not matter for testing the correlation either.[1]

Finally, if allowing for a break in the variance, say a model

$$\boldsymbol{X}_t = \begin{pmatrix} \sqrt{\sigma_{1,1}^2 \left(1 - D_{t,\lambda}\right) + \sigma_{1,2}^2 D_{t,\lambda}} & 0 \\ 0 & \sqrt{\sigma_{2,1}^2 \left(1 - D_{t,\lambda}\right) + \sigma_{2,2}^2 D_{t,\lambda}} \end{pmatrix} \boldsymbol{Z}_t$$

(for simplicity with known zero mean since demeaning does not have an asymptotic effect in this setup), we obtain

$$\hat{Z}_{ti} = \frac{X_{ti}}{\sqrt{\hat{\sigma}_{i,1}^2 \left(1 - D_{t,\lambda}\right) + \hat{\sigma}_{i,2}^2 D_{t,\lambda}}} \qquad \text{and} \quad \boldsymbol{h}\left(\boldsymbol{x}\right) = \begin{pmatrix} \frac{x_1}{\sqrt{\theta_1 \bar{D}_{t,\lambda} + \theta_2 D_{t,\lambda}}} \\ \frac{x_2}{\sqrt{\theta_3 \bar{D}_{t,\lambda} + \theta_4 D_{t,\lambda}}} \end{pmatrix}$$

and consequently

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}} = -\frac{1}{2} \begin{pmatrix} \frac{x_1 \bar{D}_{t,\lambda}}{\left(\sigma_{1,1}^2 \bar{D}_{t,\lambda} + \sigma_{1,2}^2 D_{t,\lambda}\right)^{3/2}} & \frac{x_1 D_{t,\lambda}}{\left(\sigma_{1,1}^2 \bar{D}_{t,\lambda} + \sigma_{1,2}^2 D_{t,\lambda}\right)^{3/2}} & 0 & 0 \\ 0 & 0 & \frac{x_2 \bar{D}_{t,\lambda}}{\left(\sigma_{2,1}^2 \bar{D}_{t,\lambda} + \sigma_{2,2}^2 D_{t,\lambda}\right)^{3/2}} & \frac{x_2 D_{t,\lambda}}{\left(\sigma_{2,1}^2 \bar{D}_{t,\lambda} + \sigma_{2,2}^2 D_{t,\lambda}\right)^{3/2}} \end{pmatrix}.$$

Then, we obtain for $\frac{1}{n} \sum_{t=1}^{[sn]} \frac{\partial g}{\partial \boldsymbol{z}}\Big|_{\boldsymbol{z}=\boldsymbol{Z}_t} \frac{\partial h}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ the expression

$$-\frac{1}{2n} \sum_{t=1}^{[sn]} \left(Z_{t2}, Z_{t1}\right) \begin{pmatrix} \frac{Z_{t1} \bar{D}_{t,\lambda_0}}{\sigma_{1,1,0}^2 \bar{D}_{t,\lambda_0} + \sigma_{1,2,0}^2 D_{t,\lambda_0}} & \frac{Z_{t1} D_{t,\lambda_0}}{\sigma_{1,1,0}^2 \bar{D}_{t,\lambda_0} + \sigma_{1,2,0}^2 D_{t,\lambda_0}} & 0 & 0 \\ 0 & 0 & \frac{Z_{t2} \bar{D}_{t,\lambda_0}}{\sigma_{2,1,0}^2 \bar{D}_{t,\lambda_0} + \sigma_{2,2,0}^2 D_{t,\lambda_0}} & \frac{Z_{t2} D_{t,\lambda_0}}{\sigma_{2,1,0}^2 \bar{D}_{t,\lambda_0} + \sigma_{2,2,0}^2 D_{t,\lambda_0}} \end{pmatrix}$$

$$\Rightarrow -\frac{\rho_0}{2} \left(\frac{\mathbb{I}(s<\lambda_0)}{\sigma_{1,1,0}^2} s + \frac{\mathbb{I}(s \geq \lambda_0)}{\sigma_{1,1,0}^2} \lambda_0 \quad \frac{\mathbb{I}(s \geq \lambda_0)}{\sigma_{1,2,0}^2} \left(s - \lambda_0\right) \quad \frac{\mathbb{I}(s<\lambda_0)}{\sigma_{2,1,0}^2} s + \frac{\mathbb{I}(s \geq \lambda_0)}{\sigma_{2,1,0}^2} \lambda_0 \quad \frac{\mathbb{I}(s \geq \lambda_0)}{\sigma_{2,2,0}^2} \left(s - \lambda_0\right) \right) \equiv \boldsymbol{\tau}_{\lambda_0}\left(s\right)$$

which is piecewise linear for $s \in [0, 1]$. Hence the effect of accounting for breaks in the variance is not negligible when concerned about the correlation, not even when testing the constancy, unless $\rho_0 = 0$. The corresponding process is also not a Brownian motion,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[sn]} \begin{pmatrix} Z_{t1} Z_{t2} - \rho_0 \\ \sigma_{1,1,0}^2 \left(Z_{t1}^2 - 1\right) \left(1 - D_{t,\lambda_0}\right) \\ \sigma_{1,2,0}^2 \left(Z_{t1}^2 - 1\right) D_{t,\lambda_0} \\ \sigma_{2,1,0}^2 \left(Z_{t2}^2 - 1\right) \left(1 - D_{t,\lambda_0}\right) \\ \sigma_{2,0,2}^2 \left(Z_{t2}^2 - 1\right) D_{t,\lambda_0} \end{pmatrix} \Rightarrow \boldsymbol{\Psi}_{\lambda_0}\left(s\right) \equiv \begin{pmatrix} \Omega^{1/2} \Gamma\left(s\right) \\ \Sigma^{1/2} \boldsymbol{\Theta}_{\lambda_0}\left(s\right) \end{pmatrix}.$$

Similar conclusions hold for the case of recursive parameter estimation. Summing up, breaks in the variances complicate the analysis of correlation tests, but breaks in the mean do not.

C. ADDITIONAL EXAMPLE: TESTING FOR NORMALITY

Let us consider testing hypotheses about the higher-order moments of a (univariate latent) i.i.d. series Z_t in a location-scale model,

$$X_t = \mu + \sigma Z_t \quad \text{with} \quad Z_t \sim \text{i.i.d.}\left(0, 1\right).$$

[1]A similar result can be shown for testing the constancy of variances, which is asymptotically not affected by changes in the mean, if residuals taking into account these changes are used. This justifies the procedure applied in Borowski et al. (2014).

Letting

$$\hat{Z}_t = \frac{X_t - \hat{\mu}}{\hat{\sigma}} \quad \text{with} \quad \hat{\sigma}^2 = \frac{1}{n}\sum (X_t - \hat{\mu})^2 \quad \text{and} \quad \hat{\mu} = \bar{X},$$

we may test hypotheses about the skewness μ_3 of Z_t (or equivalently the standardized skewness of X_t) building on the statistic

$$\mathcal{T} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\hat{Z}_t^3 - \mu_{3,0} \right).$$

The relevant quantities are

$$g(z) = z^3, \qquad \boldsymbol{\theta} = \left(\mu, \sigma^2\right)' \qquad \text{and} \qquad h(x) = \frac{x - \theta_1}{\sqrt{\theta_2}},$$

such that

$$\frac{\partial g}{\partial z} = 3z^2 \qquad \text{and} \qquad \frac{\partial h}{\partial \boldsymbol{\theta}} = \left(-\frac{1}{\sqrt{\theta_2}}, -\frac{1}{2}\frac{x - \theta_1}{\theta_2^{3/2}} \right),$$

leading to

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{[sn]} \left.\frac{\partial g}{\partial z}\right|_{z=Z_t} \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= \frac{1}{n}\sum_{t=1}^{[sn]} 3Z_t^2 \left(-\frac{1}{\sigma_0}, -\frac{1}{2}\frac{Z_t}{\sigma_0^3} \right) \\ &\Rightarrow -3s \left(\sigma_0, \frac{\mu_{3,0}}{2\sigma_0^3} \right) \equiv \boldsymbol{\tau}(s). \end{aligned}$$

Hence

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[sn]} \left(\hat{Z}_t^3 - \mu_{3,0} \right) \Rightarrow \Omega^{1/2}\Gamma(s) - 3s\left(\sigma_0, \frac{\mu_{3,0}}{2\sigma_0^3} \right) \Sigma^{1/2}\boldsymbol{\Theta}(1)$$

where

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[sn]} \begin{pmatrix} Z_t^3 - \mu_{3,0} \\ \sigma_0 Z_t \\ \sigma_0^2 Z_t^2 - \sigma_0^2 \end{pmatrix} \Rightarrow \boldsymbol{\Psi}(s) \equiv \begin{pmatrix} \Omega^{1/2}\Gamma(s) \\ \Sigma^{1/2}\boldsymbol{\Theta}(s) \end{pmatrix}$$

with $\boldsymbol{\Psi}$ a Brownian motion with quadratic covariation process

$$[\boldsymbol{\Psi}](s) = s \begin{pmatrix} \mu_{6,0} - \mu_{3,0}^2 & \sigma_0\mu_{4,0} & \sigma_0^2(\mu_{5,0} - \mu_{3,0}) \\ \sigma_0\mu_{4,0} & \sigma_0^2 & \sigma_0^3\mu_{3,0} \\ \sigma_0^2(\mu_{5,0} - \mu_{3,0}) & \sigma_0^3\mu_{3,0} & \sigma_0^4(\mu_{4,0} - 1) \end{pmatrix},$$

hence $\Omega = \mu_{6,0} - \mu_{3,0}^2$, $\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_0^3\mu_{3,0} \\ \sigma_0^3\mu_{3,0} & \sigma_0^4(\mu_{4,0} - 1) \end{pmatrix}$ and $\Lambda = \begin{pmatrix} \sigma\mu_4 \\ \sigma_0^2(\mu_{5,0} - \mu_{3,0}) \end{pmatrix}$. Also, $\Pi(s) = sI_2$ is this case, as we deal with estimators that are essentially sample averages. (This is the case for the following examples as well.)

We note that demeaning always has an effect on the partial sums, but whether estimating the variance has an effect or not depends explicitly on the true skewness $\mu_{3,0}$ of the considered DGP. If one is interested in testing the constancy of the skewness, both effects cancel out in the statistic according to Corollary 3.2.

Note also that Jarque and Bera (1980) claim that there is no effect when testing the null of normality in the Pearson family of distributions. Jarque and Bera (1980, p. 257) indicate $m_3^2/6m_2^3$ as unfeasible statistic, with $m_k = n^{-1}\sum_{t=1}^{n} Z_t^k$, and the analog $\hat{m}_3^2/6\hat{m}_2^3$, with $\hat{m}_k = n^{-1}\sum_{t=1}^{n} \hat{Z}_t^k$, as residual-based one. So, as it is known that the residual-based statistic works, their conclusion seems correct. However, since the 6th centered moment of the normal distribution is $15\sigma^6$, it is immediately seen that the statistic $m_3^2/6m_2^3$ is not χ_1^2 in the limit (and the correct unfeasible statistic would have been $m_3^2/15m_2^3$), so the residual effect is actually present, as discussed above.

Now, for testing the kurtosis of Z_t, h is the same but

$$g(z) = z^4 \qquad \text{and} \qquad \frac{\partial g}{\partial z} = 4z^3,$$

such that

$$\frac{1}{n}\sum_{t=1}^{[sn]} \left.\frac{\partial g}{\partial z}\right|_{z=Z_t} \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n}\sum_{t=1}^{[sn]} 4Z_t^3 \left(-\frac{1}{\sigma_0}, -\frac{1}{2}\frac{Z_t}{\sigma_0^2}\right)$$
$$\Rightarrow -4s\left(\frac{\mu_3}{\sigma_0}, \frac{\mu_4}{2\sigma_0^2}\right) \equiv \boldsymbol{\tau}(s).$$

The process $\boldsymbol{\Psi}(s)$ (in particular the component $\Gamma(s)$) is different,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]} \begin{pmatrix} Z_t^4 - \mu_{4,0} \\ \sigma_0 Z_t \\ \sigma_0^2 Z_t^2 - \sigma_0^2 \end{pmatrix} \Rightarrow \boldsymbol{\Psi}(s),$$

having a different quadratic covariation,

$$[\boldsymbol{\Psi}](s) = s\begin{pmatrix} \mu_{8,0} - \mu_{4,0}^2 & \sigma_0\mu_{5,0} & \sigma_0^2(\mu_{6,0} - \mu_{4,0}) \\ \sigma_0\mu_{5,0} & \sigma_0^2 & \sigma_0^3\mu_{3,0} \\ \sigma_0^2(\mu_{6,0} - \mu_{4,0}) & \sigma_0^3\mu_{3,0} & \sigma_0^4(\mu_{4,0} - 1) \end{pmatrix}.$$

Contrary to the case of the skewness, estimating the variance has an effect on the partial sums irrespective of the skewness, but the actual skewness $\mu_{3,0}$ controls now whether demeaning has an effect. Again, if interested in the constancy of the kurtosis, both effects cancel out and the asymptotics is not affected by the residual effect.

D. ASYMPTOTIC AND BOOTSTRAP CORRECTIONS

For the cases where there is a residual effect, corrections are required. We first discuss the more straightforward case of simple hypotheses, $\mathrm{E}(\boldsymbol{g}(\boldsymbol{Z}_t)) = \boldsymbol{\mu}^{(0)}$.

If basing the test on residuals with full-sample parameter estimation, we have under the null

$$\sqrt{n}\left(\bar{\boldsymbol{g}} - \boldsymbol{\mu}^{(0)}\right) \Rightarrow \Omega^{1/2}\boldsymbol{\Gamma}(1) + \boldsymbol{\tau}(1)\left(\Pi'(1)\,W\,\Pi(1)\right)^{-1}\Pi'(1)\,W\,\Sigma^{1/2}\boldsymbol{\Theta}(1)$$

which is actually multivariate normally distributed under Assumption 2.1, so making the distribution of this quadratic form pivotal is a matter of using the right covariance matrix estimator: $\hat{\Omega}$ is only correct when $\boldsymbol{\tau}$ is zero; see the corollaries above. Otherwise, one should have used

$$\left(I_L; W'\Pi(1)\left(\Pi'(1)\,W\,\Pi(1)\right)^{-1}\boldsymbol{\tau}'(1)\right)\hat{\Xi}\begin{pmatrix} I_L \\ \boldsymbol{\tau}(1)\left(\Pi'(1)\,W\,\Pi(1)\right)^{-1}\Pi'(1)\,W \end{pmatrix} \tag{D.2}$$

instead of $\hat{\Omega}$. Here, I_L denotes the $L \times L$ identity matrix. This situation is often encountered in the literature; see e.g. Bai and Ng (2005).

This correction is not available for recursive estimation of the parameters. The difference is that $\mathrm{Cov}(\tilde{\boldsymbol{\Gamma}}(1))$ depends on the entire path of $\boldsymbol{\tau}$ which makes a correct estimation of the required covariance matrix more demanding. In principle, one could simulate from the limiting distribution, given estimates for $\boldsymbol{\tau}$, Π and Ξ. While this is feasible, it would may be easier to bootstrap, as is not uncommon in the literature; see e.g. Zhou (2013) and Hansen (2000). This too is not without disadvantages; see the discussion on bootstrap implementations below.

Now, if $\boldsymbol{g}(\boldsymbol{Z}_t)$ is weakly stationary then $\Omega^{1/2}\boldsymbol{\Gamma}$ is a Brownian motion. Under time-varying 2nd moments of $\boldsymbol{g}(\boldsymbol{Z}_t)$, however, the process $\Omega^{1/2}\boldsymbol{\Gamma}$ would have nonlinear quadratic covariation. In this case $\boldsymbol{\Gamma}$ cannot be a vector of independent Wiener processes, and the test statistic is not asymptotically pivotal under the null. Provided that (consistent) estimates of $\boldsymbol{\tau}$, Π and Ξ, as well as of the nonlinear (co)variance profiles of the limiting process $\boldsymbol{\Psi}$, are available, one may simulate critical values from the limiting distribution. Again, it may be more convenient to resort to a suitable bootstrap. E.g. Zhou (2013) uses the block wild bootstrap.

Moving on to the case of moment constancy tests, it is worth asking the question whether $\hat{Q}_n$ or $\tilde{Q}_n$ could be corrected using the right covariance matrix estimator, like in the case of simple hypotheses. This is more difficult to achieve since the test statistic depends on the entire path of Ψ and not only on the properties of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Theta}$ at $s = 1$. For such a correction to work, one needs linear combinations of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Theta}$ to have the same properties as $\boldsymbol{\Gamma}$ only. This, as can be easily checked, is the case only when $\boldsymbol{\Gamma}$ and $\boldsymbol{\Theta}$ are Gaussian processes with covariance profile of the form $\eta(s)\Upsilon$ with $\eta(s)$ a suitable scalar function and Υ a constant positive definite matrix. Should the correction be applicable, this works immediately for $\hat{Q}_n$, but becomes decisively more complex for $\tilde{Q}_n$ where the integral of $\boldsymbol{\Theta}$ over $[0, s]$ is a Gaussian process, but no Brownian motion.

Finally, since analytical corrections may not be straightforward, and sometimes nonlinear quadratic covariations need to be accounted for, the bootstrap suggests itself to obtain critical values. Since the effect depends also on the properties of estimator $\hat{\boldsymbol{\theta}}$ (in particular on $A_{t,n}$ or $B_{t,n}$), on which it is difficult to get more precise without becoming too model-specific, a thorough analysis of bootstrap validity is out of the reach of this paper. Rather, we point out some pitfalls associated to standard (block) i.i.d. and wild bootstrap schemes.

Denote by $\boldsymbol{X}^*_{t,b}$ the bootstrapped sample (which may be obtained either by bootstrapping $\boldsymbol{X}_t$, or by bootstrapping $\hat{\boldsymbol{Z}}_t$ or $\tilde{\boldsymbol{Z}}_t$ and filtering through an estimated version of $\boldsymbol{f}$). For testing, we shall assume that the null is suitably imposed when bootstrapping.[2] Then, with "$\overset{p}{\Rightarrow}$" denoting weak convergence in probability[3] and E^* the bootstrap expectation, it must be ensured that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(\boldsymbol{g}\left(\hat{\boldsymbol{Z}}^*_{t,b}\right) - \mathrm{E}^*\left(\boldsymbol{g}\left(\boldsymbol{Z}^*_{t,b}\right)\right)\right) \overset{p}{\Rightarrow} \Omega^{1/2}\boldsymbol{\Gamma}(s) + \boldsymbol{\tau}(s)\left(\Pi'(1)\,W\,\Pi(1)\right)^{-1}\Pi'(1)\,W\,\Sigma^{1/2}\boldsymbol{\Theta}(1)$$

for the full sample estimation, and

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}^*_{t,b}\right) - \mathrm{E}^*\left(\boldsymbol{g}\left(\boldsymbol{Z}^*_{t,b}\right)\right)\right) \overset{p}{\Rightarrow} \Omega^{1/2}\boldsymbol{\Gamma}(s) + \left(\int_0^s \boldsymbol{\Theta}'(r)\left(\Sigma^{1/2}\right)' W'\Pi(r)\left(\Pi'(r)\,W\,\Pi(r)\right)^{-1}\mathrm{d}\boldsymbol{\tau}'(r)\right)'$$

for recursive estimation. I.e., the bootstrapped partial sums should converge to the same limit process as in Proposition 3.1, such that the residual effect is correctly replicated by the bootstrap.

This, however, is not guaranteed with any bootstrap scheme. Consider e.g. the well-understood case of the i.i.d. bootstrap performed on $\boldsymbol{X}_t$. Then, the bootstrap samples do not replicate serial correlation or nonstationarities of the DGP. One could of course use the block bootstrap to side-step the first issue, and resort to the residual i.i.d. bootstrap, if the source of the nonstationarity

[2] This may not be difficult if constancy is of interest, but one may have to go at some lengths to impose say zero skewness in the bootstrap population.

[3] A sequence of random functions $X^*_1, \ldots$ converges weak in probability conditionally on the original data $X_1, \ldots$ to some limit function X, if $E(f(X^*_n)|X_1, \ldots) \to_p E(f(X))$ for every bounded and continuous function f.

lies in the filter or in the structure of the estimator. If on the other hand the quantities $\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right)$ or $A_{t,n}$ are not stationary, but only piecewise locally stationary, one could use wild or block wild bootstraps as suggested by Hansen (2000) or Zhou (2013) in related contexts. A seminal reference for this bootstrap is Wu (1986). This too is not always going to lead to valid results. To see why, take $A_{t,n} = \boldsymbol{a}\left(\boldsymbol{X}_t\right)$. Then, wild bootstrapping $\boldsymbol{X}_t$ or $\hat{\boldsymbol{Z}}_t$ ($\tilde{\boldsymbol{Z}}_t$), even in block versions, does not produce the desired result in general: in an extreme case, $\boldsymbol{g}$ or $\boldsymbol{a}$ may e.g. be even functions, and using e.g. Rademacher random variables $R_{t,b}$ to generate bootstrap samples $\boldsymbol{X}^*_{t,b} = \boldsymbol{X}_t R_{t,b}$ would not give bootstrap sampling variability at all. But the issue is more subtle, because even if we don't use the Rademacher distribution, the covariance of $\boldsymbol{g}\left(\boldsymbol{X}^*_{t,b}\right)$ and $\boldsymbol{a}\left(\boldsymbol{X}^*_{t,b}\right)$ need not equal the covariance of $\boldsymbol{g}\left(\boldsymbol{X}_t\right)$ and $\boldsymbol{a}\left(\boldsymbol{X}_t\right)$.[4] (A related case of wild bootstrap failure is given in Brüggemann et al., 2016.) The solution here would be to block wild bootstrap $\boldsymbol{g}\left(\boldsymbol{Z}_t\right)$ and $A_{t,n}$ *jointly*, e.g. $\left(\boldsymbol{g}\left(\boldsymbol{X}_t\right), \boldsymbol{a}\left(\boldsymbol{X}_t\right)\right)^* = \left(\boldsymbol{g}\left(\boldsymbol{X}_t\right), \boldsymbol{a}\left(\boldsymbol{X}_t\right)\right) R_{t,b}$. The bottom line is that bootstrapping without understanding the asymptotics of the residual effect is likely to fail.

[4]Consider e.g. $g(u) = u$ and $a(u) = u^2$; then, unless $\mathrm{E}\left(R^3_{t,b}\right) = 1$, the wild bootstrap fails.

References

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*(1), 49–60.

Borowski, M., N. Rudak, B. Hussong, D. Wied, S. Kuhnt, and W. Tillmann (2014). On- and offline detection of structural breaks in thermal spraying processes. *Journal of Applied Statistics 41*(5), 1073–1090.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics 191*(1), 69–85.

Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics 97*(1), 93–115.

Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters 6*(3), 255–259.

Lütkepohl, H. (1996). *Handbook of matrices.* John Wiley and Sons.

Wu, C. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics 14*, 1261–1350.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*(502), 726–740.