Online Supplement to "TIME VARYING PARAMETER REGRESSIONS WITH STATIONARY PERSISTENT DATA"

Zhishui Hu
University of Science & Technology of China

Ioannis Kasparis
University of Cyprus

Qiying Wang
The University of Sydney

February 29, 2024

This supplement is organized as follows. Section A provides proofs of the results given in the main paper. In particular, Section A.1 introduces a preliminary lemma and the proofs of theorems in Sections 3-6 are given in A.2-A.5, respectively. In Section B, we provide some additional details to verify (3)-(4) and (6)-(10) in Section 2 of the main paper -see Lemmas 2 and 3 together with their proofs. A simulation study that investigates the finite sample properties of the estimators and related test statistics is given in Section C. Throughout the proofs, we assume that C is a positive constant that may take a different value in each appearance and define $K_{kn} = K\left[c_n\left(k/n - \tau\right)\right]$. Other notation is the same as in the main paper unless stated otherwise.

A Proofs of the main results

A.1 A preliminary Lemma

Let $\{v_k\}_{k\geq 1}$ be a $p \times p'$ matrix sequence of random variables and $K(x)$ be a Borel function on $\mathbb{R}$. Set

$$L_n(\tau) := \frac{c_n}{n}\sum_{k=1}^{n} v_k K\left[c_n(k/n - \tau)\right],$$

where $\{c_k\}_{k\geq 1}$ is a sequence of positive constants. The following lemma plays a key role in the proofs of the main results and provides an extension to Lemma 5.1 of Hu et al. (2021).

Lemma 1. *Suppose that*

(a) $\sup_{k\geq 1} E\|v_k\| < \infty$ *and there exist* $A_0 \in \mathbb{R}^{p\times p'}$ *and* $0 < m := m_n \to \infty$ *satisfying* $n/m \to \infty$ *so that* $\max_{m\leq j\leq n-m} E\left\|\frac{1}{m}\sum_{k=j+1}^{j+m} v_k - A_0\right\| = o(1)$;

(b) $K(x)$ *is locally Riemann integrable and eventually monotonic so that* $\int |K| < \infty$.

Then, for each $\tau \in (0,1)$, $c_n \to \infty$ and $c_n/n \to 0$, we have

$$\left\| L_n(\tau) - A_0 \int K \right\| = o_P(1). \tag{A.1}$$

Remark A.1. Strict stationarity and ergodicity for v_k are sufficient for the limit result of condition (a) above. Indeed, it follows from the stationarity requirement that $E\left\| \frac{1}{m}\sum_{k=j+1}^{j+m} v_k - A_0 \right\| = E\left\| \frac{1}{m}\sum_{k=1}^{m} v_k - A_0 \right\|$ for all j. This, together with the ergodicity and the finite moment condition, yields (with $A_0 = E(v_1)$)

$$\max_{m \le j \le n-m} E\left\| \frac{1}{m}\sum_{k=j+1}^{j+m} v_k - A_0 \right\| = E\left\| \frac{1}{m}\sum_{k=1}^{m} v_k - A_0 \right\| \to 0,$$

as $m \to \infty$, cf. Shiryaev (1996), Theorem 3, p. 413.

Remark A.2. It can be easily seen from the proof of Lemma 1 that (A.1) holds true for $\tau = 0$ and $\tau = 1$, if the one sided integral limits $\int_0^\infty K$ (when $\tau = 0$) and $\int_{-\infty}^0 K$ (when $\tau = 1$) are used in the place of $\int_{\mathbb{R}} K$. For instance, if $\tau = 0$ the counterpart of (A.5) is

$$\left| \frac{c_n}{n} \sum_{k=\delta_{1n}}^{\delta_{2n}} K\left[c_n(k/n - \tau)\right] - \int_0^\infty K \right| \to 0. \tag{A.2}$$

Remark A.3. Suppose that $\beta : [0,1] \to \mathbb{R}$ is Riemann integrable and v_t satisfies condition (a) in Lemma 1. Since, instead of (A.2),

$$\left| \frac{1}{n} \sum_{k=\delta_{1n}}^{\delta_{2n}} \beta(k/n) - \int_0^1 \beta \right| \to 0,$$

for any $\delta_{1n}/n \to 0$ and $\delta_{2n}/n \to 1$, a minor modification in the proof of Lemma 1 gives

$$\frac{1}{n}\sum_{k=1}^{n} v_k \beta(k/n) = A_0 \int_0^1 \beta(\tau) d\tau + o_P(1).$$

Proof of Lemma 1. We first assume that there exists an $A > 0$ such that $K(x) = 0$ if $|x| \ge A$ and $K(x)$ is Lipschitz continuous on $\mathbb{R}$. These restrictions on $K(x)$ will be removed later. Without loss of generality, suppose $A = 1$ and $K \ge 0$. Set $\delta_{1n} = [n(\tau - 1/c_n)] \vee 1$, $\delta_{2n} = [n(\tau + 1/c_n)] \vee 1$, and

$$L_n'(\tau) := \frac{c_n}{n} \sum_{k=\delta_{1n}}^{\delta_{2n}} v_k K\left[c_n(k/n - \tau)\right].$$

Since,

$$|c_n(k/n - \tau)| < 1 \quad \text{only if} \quad \delta_{1n} \le k \le \delta_{2n}, \tag{A.3}$$

we have $L_n = L'_n$. Result (A.1) will follow if we prove that for all $\tau \in (0,1)$,

$$E \left\| L'_n(\tau) - A_0 \int K \right\| \to 0. \tag{A.4}$$

Since, Euler summation yields

$$\left| \frac{c_n}{n} \sum_{k=\delta_{1n}}^{\delta_{2n}} K\left[c_n(k/n - \tau)\right] - \int K \right| \to 0 \tag{A.5}$$

for all $\tau \in (0,1)$, as $n \to \infty$, it suffices to show that $E\left\|R_n(\tau)\right\| \to 0$, where

$$R_n(\tau) = \frac{c_n}{n} \sum_{k=\delta_{1n}}^{\delta_{2n}} (v_k - A_0)\; K\left[c_n(k/n - \tau)\right].$$

Let $\gamma = \gamma_n$ be integers such that $\gamma \to \infty$ and $\gamma\, c_n/n \to 0$, $T_{1n} = [\delta_{1n}/\gamma]$ and $T_{2n} = [\delta_{2n}/\gamma]$. Noting (A.3), we may write

$$\begin{aligned}
\left\|R_n(\tau)\right\| &= \left\| \frac{c_n}{n} \sum_{k=\delta_{1n}}^{\delta_{2n}} (v_k - A_0)\, K\left[c_n(k/n - \tau)\right] \right\| \\
&= \left\| \frac{c_n}{n} \sum_{s=T_{1n}}^{T_{2n}} \sum_{k=s\gamma}^{(s+1)\gamma} (v_k - A_0) K\left[c_n(k/n - \tau)\right] \right\| \\
&\leq \frac{\gamma c_n}{n} \sum_{s=T_{1n}}^{T_{2n}} K\left[c_n(s\gamma/n - \tau)\right] \frac{1}{\gamma} \left\| \sum_{k=s\gamma}^{(s+1)\gamma} (v_k - A_0) \right\| \\
&\quad + \frac{c_n}{n} \sum_{s=T_{1n}}^{T_{2n}} \sum_{k=s\gamma}^{(s+1)\gamma} \left\|v_k - A_0\right\| \left|K\left[c_n(k/n - \tau)\right] - K\left[c_n(s\gamma/n - \tau)\right]\right| \\
&:= A_{1n}(\tau) + A_{2n}(\tau).
\end{aligned}$$

Recall that $\sup_{k\geq 1} E\left\|v_k\right\| < \infty$ by condition (b). In view of this it is readily seen from the Lipschitz condition on $K(x)$ that

$$E \sup_{\tau} A_{2n}(\tau) \leq C\, \frac{\gamma c_n}{n}\, \frac{c_n}{n} \sum_{k=\delta_{1n}}^{\delta_{2n}} E\left\|v_k - A_0\right\| \leq C \frac{\gamma c_n}{n} \to 0.$$

Similarly, using condition (b), we have

$$E A_{1n}(\tau) \leq \max_{\gamma \leq s \leq n-\gamma} E \left\| \frac{1}{\gamma} \sum_{k=s}^{s+\gamma} v_k - A_0 \right\| \sup_{\tau} A_{3n}(\tau) \to 0,$$

where

$$A_{3n}(\tau) = \frac{\gamma c_n}{n} \sum_{s=T_{1n}}^{T_{2n}} K\left[c_n(s\gamma/n - \tau)\right],$$

and we have used the fact that $\sup_{\tau\in(0,1)} \left|A_{3n}(\tau) - \int K\right| \to 0$. Combining all these facts, we complete the proof of $E\left\|R_n(\tau)\right\| \to 0$.

We next remove the restrictions on K and then conclude the proof of Lemma 1. Without loss of generality, we assume $K \geq 0$. Since $K \geq 0$ is eventually monotonic, for any $\epsilon > 0$, there exists a constant $A_{1\epsilon} > 0$ such that $K(x)$ is monotonic on $(-\infty, -A_{1\epsilon})$ and $(A_{1\epsilon}, \infty)$ and $\int_{|x|>A_{1\epsilon}} K(x)dx < \epsilon$. As a consequence, it follows from $\int K < \infty$ that, for any $\epsilon > 0$ and $A \geq A_{1\epsilon} + 1$, there is some $K_{\epsilon,A}(x)$ Lipschitz continuous on $\mathbb{R}$ such that

$$\int |K - K_{\epsilon,A}| \leq 2\epsilon, \tag{A.6}$$

and $K_{\epsilon,A}(x) = 0$, if $|x| \geq A$ (see e.g. Theorem 2.26 in Folland, 1999). It has been shown in the first part that, for any $\epsilon > 0$ and $A \geq A_{1\epsilon} + 1$,

$$\frac{c_n}{n} \sum_{k=1}^{n} v_k \, K_{\epsilon,A}\left[c_n(k/n - \tau)\right] = A_0 \int K_{\epsilon,A} + o_P(1).$$

To show (A.1), it suffices to show that, as $n \to \infty$ first and then $\epsilon \to 0$ (implying $A \to \infty$),

$$L_{n,\epsilon}(\tau) := \frac{c_n}{n} \sum_{k=1}^{n} v_k \widetilde{K}\left[c_n(k/n - \tau)\right] = o_P(1), \tag{A.7}$$

where $\widetilde{K}(x) = K(x) - K_{\epsilon,A}(x)$.

For any $\epsilon > 0$, let A be given as in (A.6). First note that, by the local Riemann integrability of $\tilde{K}(x)$, we have

$$\left|\frac{c_n}{n} \sum_{k=1}^{n} \widetilde{K}\left[c_n(k/n - \tau)\right] I(c_n|k/n - \tau| \leq A) - \int_{-A}^{A} \widetilde{K}(x)dx\right| \to 0,$$

for each τ when $n \to \infty$. Therefore, for n sufficiently large,

$$R_{1n} := \frac{c_n}{n} \sum_{k=1}^{n} \left|\widetilde{K}\left[c_n(k/n - \tau)\right] I(c_n|k/n - \tau| \leq A)\right| \leq \int |\widetilde{K}(x)|dx + \epsilon \leq 3\epsilon.$$

On the other hand, it follows from the monotonicity of $K(x)$ on $(-\infty, -A)$ and (A, ∞) that, whenever n is sufficiently large,

$$\begin{aligned} R_{2n} &:= \frac{c_n}{n} \sum_{k=1}^{n} \left|\widetilde{K}\left[c_n(k/n - \tau)\right]\right| I(c_n|k/n - \tau| > A) \\ &= \frac{c_n}{n} \sum_{k=1}^{n} K\left[c_n(k/n - \tau)\right] I(c_n|k/n - \tau| > A) \end{aligned}$$

$$\leq \int_{|x|>A-c_n/n} K(x)dx \leq \int_{|x|>A_{1\epsilon}} K(x)dx < \epsilon.$$

By using these facts, when n is sufficiently large, we have

$$\frac{c_n}{n}\sum_{k=1}^{n}\left|\widetilde{K}\left[c_n(k/n-\tau)\right]\right| \leq R_{1n}+R_{2n} \leq 4\epsilon.$$

In view of the above, as $n\to\infty$ first and then $\epsilon\to 0$,

$$E\left\|L_{n,\epsilon}\right\| \leq 4\epsilon \sup_{k\geq 1} E\left\|v_k\right\| \to 0,$$

as required. This completes the proof of Lemma 1. □

A.2 Proof of Theorem 1

We only consider M_n, i.e., (11) in the main paper, since the limit result for S_n follows easily from Lemma 1 with $v_k = \mathbf{x}_{k-1}\mathbf{x}_{k-1}'\sigma_k^m$.

Set $Q_{k,n} := \sqrt{\frac{c_n}{n}}\alpha'\mathbf{x}_{k-1}\sigma_k K\left[c_n(k/n-\tau)\right]$ where $\alpha\in\mathbb{R}^p$. Using Lemma 1 with $v_k = \left[\alpha' g(x_{k-1})\sigma_k\right]^2$, we have

$$\begin{aligned}\sum_{k=1}^{n} Q_{k,n}^2 &= \frac{c_n}{n}\sum_{k=1}^{n}\left[\alpha'\mathbf{x}_{k-1}\sigma_k\right]^2 K^2\left[c_n(k/n-\tau)\right] \\ &= E\left[\alpha'\mathbf{x}_0\sigma_1\right]^2\int K^2 + o_P(1), \qquad \text{(A.8)}\end{aligned}$$

where the second equation follows from Lemma 1 - $K(x)$ is replaced by $K^2(x)$ and A_0 is set $A_0 = E\left[\alpha'\mathbf{x}_0\sigma_1\right]^2$. In terms of (A.8), it follows from the classical martingale limit theorem (c.g., Hall and Heyde (1980), Theorem 3.2 or Wang (2014), Theorem 2.1) that, to prove (11) in the main paper, it suffices to show

$$\max_{1\leq k\leq n}|Q_{k,n}| = o_P(1). \qquad \text{(A.9)}$$

Note that for any $A>0$, we have the inequality

$$\begin{aligned}\max_{1\leq k\leq n}|Q_{k,n}| &\leq \max_{1\leq k\leq n}\left[\,|Q_{k,n}|\, I\left\{\|\mathbf{x}_{k-1}\sigma_k\|>A\right\}\right] + \max_{1\leq k\leq n}\left[\,|Q_{k,n}|\, I\left\{\|\mathbf{x}_{k-1}\sigma_k\|\leq A\right\}\right] \\ &\leq \left\{\sum_{k=1}^{n} Q_{k,n}^2 I\left\{\|\mathbf{x}_{k-1}\sigma_k\|>A\right\}\right\}^{1/2} + \left\{\sum_{k=1}^{n} Q_{k,n}^4 I\left\{\|\mathbf{x}_{k-1}\sigma_k\|\leq A\right\}\right\}^{1/4} \\ &=: II_{1n}(A)^{1/2} + II_{2n}(A)^{1/4}.\end{aligned}$$

Similar arguments used in (A.8) show that the first term

$$II_{1n}(A) \leq \|\alpha\|^2\frac{c_n}{n}\sum_{k=1}^{n}\|\mathbf{x}_{k-1}\sigma_k\|^2 I\left\{\|\mathbf{x}_{k-1}\sigma_k\|>A\right\} K^2\left[c_n(k/n-\tau)\right]$$

$$= \quad \|\alpha\|^2 E \left\|\mathbf{x}_0\sigma_1\right\|^2 I\left\{\|\mathbf{x}_0\sigma_1\| > A\right\} \int K^2 + o_P(1) = o_P(1),$$

where we take $n \to \infty$ first and then $A \to \infty$, and the second term

$$II_{2n}(A) \le \|\alpha\|^4 A^4 \left(\frac{c_n}{n}\right)^2 \sum_{k=1}^n K^4\big[c_n(k/n-\tau)\big] = o_P(1)$$

for each $A > 0$, as $n \to \infty$. Combining these facts together, we establish (A.9). The proof of Theorem 1 is now complete. □

A.3 Proofs of Theorems 2 and 3

We only prove Theorem 3. The proof of Theorem 2 is similar and therefore omitted. Recall $\widehat{\mathbf{x}}_k = (\mathbf{x}_k', \widetilde{\mathbf{x}}_k')'$, where $\widetilde{\mathbf{x}}_{k-1} = (k/n-\tau)\mathbf{x}_{k-1}$, and note that

$$\begin{bmatrix} \widetilde{\theta}(\tau) \\ \widetilde{\theta}^{(1)}(\tau) \end{bmatrix} = \left[\sum_{k=1}^n \widehat{\mathbf{x}}_{k-1}\widehat{\mathbf{x}}_{k-1}' K_{kn}\right]^{-1} \sum_{k=1}^n y_k \widehat{\mathbf{x}}_{k-1} K_{kn}.$$

We may write

$$D_n\left(\begin{bmatrix} \widetilde{\theta}(\tau) \\ \widetilde{\theta}^{(1)}(\tau) \end{bmatrix} - \begin{bmatrix} \theta(\tau) \\ \theta^{(1)}(\tau) \end{bmatrix}\right) = Q_n^{-1}\left(\mathcal{M}_n + R_n\right), \tag{A.10}$$

where $Q_n = D_n^{-1}\sum_{k=1}^n \widehat{\mathbf{x}}_{k-1}\widehat{\mathbf{x}}_{k-1}' K_{kn} D_n^{-1}$, $\mathcal{M}_n = D_n^{-1}\sum_{k=1}^n e_k \widehat{\mathbf{x}}_{k-1} K_{kn}$ and

$$R_n = D_n^{-1}\sum_{k=1}^n \begin{bmatrix} \mathbf{x}_{k-1} \\ \widetilde{\mathbf{x}}_{k-1} \end{bmatrix} K_{kn}\,\theta(k/n)'\mathbf{x}_{k-1} - Q_n D_n \begin{bmatrix} \theta(\tau) \\ \theta^{(1)}(\tau) \end{bmatrix}.$$

Let $K_j(x) = x^j K(x)$ and $K_{j,kn} = K_j\big[c_n(k/n-\tau)\big]$. As in the proof of Theorem 1, it follows from Lemma 1 that

$$Q_n = \begin{bmatrix} \frac{c_n}{n}\sum_{k=1}^n \mathbf{x}_{k-1}\mathbf{x}_{k-1}' K_{kn} & \frac{c_n}{n}\sum_{k=1}^n \mathbf{x}_{k-1}\mathbf{x}_{k-1}' K_{1,kn} \\ \frac{c_n}{n}\sum_{k=1}^n \mathbf{x}_{k-1}\mathbf{x}_{k-1}' K_{1,kn} & \frac{c_n}{n}\sum_{k=1}^n \mathbf{x}_{k-1}\mathbf{x}_{k-1}' K_{2,kn} \end{bmatrix} \to_P Q_2. \tag{A.11}$$

Similarly, the conditional variance matrix $[\mathcal{M}_n, \mathcal{M}_n]$ of the martingale $\mathcal{M}_n$ is

$$\begin{aligned} [\mathcal{M}_n, \mathcal{M}_n] &= D_n^{-1}\sum_{k=1}^n \sigma_k^2 \widehat{\mathbf{x}}_{k-1}\widehat{\mathbf{x}}_{k-1}' K_{kn}^2 D_n^{-1} \\ &= \begin{bmatrix} \frac{c_n}{n}\sum_{k=1}^n \sigma_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' K_{kn}^2 & \frac{c_n}{n}\sum_{k=1}^n \sigma_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' {}_{(1)}K_{kn}^2 \\ \frac{c_n}{n}\sum_{k=1}^n \sigma_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' {}_{(1)}K_{kn}^2 & \frac{c_n}{n}\sum_{k=1}^n \sigma_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' {}_{(2)}K_{kn}^2 \end{bmatrix} \to_P \mathbf{\Omega}_2, \end{aligned}$$

where ${}_{(\ell)}K^2(x) = x^\ell K^2(x)$ and ${}_{(\ell)}K_{kn}^2 = {}_{(\ell)}K^2\big[c_n(k/n-\tau)\big]$, indicating that $\mathcal{M}_n \to_d \mathbf{N}(\mathbf{0}, \mathbf{\Omega}_2)$ due to Theorem 1. Combining these facts and (A.10), Theorem 3 will follow if we prove

$$R_n = o_P(1). \tag{A.12}$$

In fact, by noting

$$\mathbf{x}_{k-1}'\theta(k/n) - \mathbf{x}_{k-1}'\theta(\tau) - \widetilde{\mathbf{x}}_{k-1}'\theta^{(1)}(\tau) = (1/2)\mathbf{x}_{k-1}'\theta^{(2)}(\overline{\tau})\,(k/n-\tau)^2\,,$$

where $\overline{\tau}$ is a mean value between k/n and τ (i.e., $0<\overline{\tau}\le 1$), it is readily seen that

$$\begin{aligned} R_n &= D_n^{-1}\sum_{k=1}^{n} K_{kn}\begin{bmatrix}\mathbf{x}_{k-1}\\ \widetilde{\mathbf{x}}_{k-1}\end{bmatrix}\left\{\mathbf{x}_{k-1}'\theta(k/n) - \begin{bmatrix}\mathbf{x}_{k-1}\\ \widetilde{\mathbf{x}}_{k-1}\end{bmatrix}'\begin{bmatrix}\theta(\tau)\\ \theta^{(1)}(\tau)\end{bmatrix}\right\} \\ &= D_n^{-1}\sum_{k=1}^{n} K_{kn}\begin{bmatrix}\mathbf{x}_{k-1}\\ \widetilde{\mathbf{x}}_{k-1}\end{bmatrix}\left\{\mathbf{x}_{k-1}'\theta(k/n) - \mathbf{x}_{k-1}'\theta(\tau) - \widetilde{\mathbf{x}}_{k-1}'\theta^{(1)}(\tau)\right\} \\ &= (1/2)D_n^{-1}\sum_{k=1}^{n}(k/n-\tau)^2 K_{kn}\begin{bmatrix}\mathbf{x}_{k-1}\\ \widetilde{\mathbf{x}}_{k-1}\end{bmatrix}\mathbf{x}_{k-1}'\theta^{(2)}(\overline{\tau}). \end{aligned}$$

Hence,

$$\begin{aligned} ||R_n|| &\le (1/2)D_n^{-1}\sum_{k=1}^{n}(k/n-\tau)^2 K_{kn}\left\|\begin{bmatrix}\mathbf{x}_{k-1}\\ \widetilde{\mathbf{x}}_{k-1}\end{bmatrix}\right\|\left|\mathbf{x}_{k-1}'\theta^{(2)}(\overline{\tau})\right| \\ &\le (1/2)\sum_{k=1}^{n}(k/n-\tau)^2 K_{kn}\left[\sqrt{\frac{c_n}{n}} + \sqrt{\frac{c_n^3}{n}}\,|k/n-\tau|\right]\|\mathbf{x}_{k-1}\|\left|\mathbf{x}_{k-1}'\theta^{(2)}(\overline{\tau})\right| \\ &\le C\left[c_n^{-2}\sqrt{\frac{c_n}{n}}\sum_{k=1}^{n}K_{2,kn}\|\mathbf{x}_{k-1}\|^2 + \sqrt{\frac{c_n^3}{n}}c_n^{-3}\sum_{k=1}^{n}|K_{3,kn}|\,\|\mathbf{x}_{k-1}\|^2\right] \\ &= C\sqrt{\frac{n}{c_n^5}}\frac{c_n}{n}\sum_{k=1}^{n}\left(K_{2,kn}+|K_{3,kn}|\right)\|\mathbf{x}_{k-1}\|^2 = o_P(1), \end{aligned}$$

where we have used the facts that $\theta^{(2)}(.)$ is uniformly bounded on $[0,1]$, $n/c_n^5\to 0$, and

$$\frac{c_n}{n}\sum_{k=1}^{n}\left(K_{2,kn}+|K_{3,kn}|\right)\|\mathbf{x}_{k-1}\|^2 \to_P \int\left(K_2+|K_3|\right)E||\mathbf{x}_0||^2,$$

the latter limit follows directly from Lemma 1. This proves (A.12) and also completes the proof of Theorem 3. □

A.4 Proofs of Theorems 4 and 5

We only prove Theorem 5. The proof of Theorem 4 is similar and therefore omitted.

By recalling (A.11) and using Theorem 3, it suffices to show that

$$D_n^{-1}\tilde{\mathbf{\Omega}}_n D_n^{-1} \;=\; \mathbf{\Omega}_2 + o_P(1). \tag{A.13}$$

Indeed, since $Q_n = D_n^{-1}\tilde{\mathcal{Q}}_n D_n^{-1}$, it follows from (A.11) and (A.13) that

$$A_n := \left(D_n^{-1}\tilde{\mathcal{Q}}_n D_n^{-1}\right)^{-1} D_n^{-1}\tilde{\mathbf{\Omega}}_n D_n^{-1}\left(D_n^{-1}\tilde{\mathcal{Q}}_n D_n^{-1}\right)^{-1} = Q_2^{-1}\mathbf{\Omega}_2 Q_2^{-1} + o_P(1).$$

As a consequence, for $i = 1, \cdots, p$ and $j = i + p$, we have

$$\frac{n}{c_n}\left[\tilde{Q}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{Q}_n^{-1}\right]_{ii} = (A_n)_{ii} \to_P \left[Q_2^{-1}\mathbf{\Omega}_2 Q_2^{-1}\right]_{ii}, \tag{A.14}$$

$$\frac{n}{c_n^3}\left[\tilde{Q}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{Q}_n^{-1}\right]_{jj} = (A_n)_{jj} \to_P \left[Q_2^{-1}\mathbf{\Omega}_2 Q_2^{-1}\right]_{jj}. \tag{A.15}$$

It follows from Theorem 3 and (A.14) that, under $H_0 : \theta_i(\tau) = \eta(\tau)$,

$$\tilde{t}_i(\tau) = \frac{\sqrt{\frac{n}{c_n}}\left(\tilde{\theta}_i(\tau) - \theta_i(\tau)\right)}{\sqrt{\frac{n}{c_n}\left[\tilde{Q}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{Q}_n^{-1}\right]_{ii}}} \to_d \mathbf{N}(0,1),$$

yielding (22) in the main paper. Similarly, it follows from Theorem 3 and (A.15) that, under $H_0 : \theta_i^{(1)}(\tau) = \eta(\tau)$

$$\tilde{t}_i^{(1)}(\tau) = \frac{\sqrt{\frac{n}{c_n^3}}\left(\tilde{\theta}_i^{(1)}(\tau) - \theta_i^{(1)}(\tau)\right)}{\sqrt{\frac{n}{c_n^3}\left[\tilde{Q}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{Q}_n^{-1}\right]_{jj}}} \to_d \mathbf{N}(0,1),$$

which gives (23) in the main paper, i.e. the second limit result of Theorem 5.

We next prove (A.13). It is readily seen that

$$D_n^{-1}\tilde{\mathbf{\Omega}}_n D_n^{-1} = \begin{bmatrix} \frac{c_n}{n}\sum_{k=1}^n \tilde{e}_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' \, {}_{(0)}K_{kn}^2 & \frac{c_n}{n}\sum_{k=1}^n \tilde{e}_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' \, {}_{(1)}K_{kn}^2 \\ \frac{c_n}{n}\sum_{k=1}^n \tilde{e}_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' \, {}_{(1)}K_{kn}^2 & \frac{c_n}{n}\sum_{k=1}^n \tilde{e}_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' \, {}_{(2)}K_{kn}^2 \end{bmatrix}, \tag{A.16}$$

where ${}_{(\ell)}K^2(x) = x^\ell K^2(x)$ and ${}_{(\ell)}K_{kn}^2 = {}_{(\ell)}K^2\left[c_n(k/n - \tau)\right], \ell = 0, 1, 2$ as defined in the proof of Theorem 3. Recalling $\tilde{e}_k = y_k - \tilde{\theta}(\tau)'\mathbf{x}_{k-1}$ and noting

$$|\left[\tilde{\theta}(\tau) - \theta(k/n)\right]'\mathbf{x}_{k-1}| \le C\left[|\tau - k/n| + o_P(1)\right] ||\mathbf{x}_{k-1}||,$$

due to Theorem 3 and the smoothing condition on $\theta(\tau)$, we have

$$\tilde{e}_k^2 = \left\{\sigma_k u_k - \left[\tilde{\theta}(\tau) - \theta(k/n)\right]'\mathbf{x}_{k-1}\right\}^2 = \sigma_k^2 u_k^2 + \Delta_{nk}, \tag{A.17}$$

where, uniformly in $k = 1, 2, ..., n$, and $0 \le \tau \le 1$

$$\begin{aligned} |\Delta_{nk}| &\le C|\sigma_k u_k|\left[|\tau - k/n| + o_P(1)\right] ||\mathbf{x}_{k-1}|| + C\left[|\tau - k/n| + o_P(1)\right]^2 ||\mathbf{x}_{k-1}||^2 \\ &\le C\left[|\tau - k/n| + o_P(1)\right]\sigma_k^2 u_k^2 + C\left[|\tau - k/n| + o_P(1)\right] ||\mathbf{x}_{k-1}||^2 \\ &:= \Delta_{1,nk}\, \sigma_k^2 u_k^2 + \Delta_{2,nk}. \end{aligned}$$

In view of this, it follows from (A.1) with $v_k = ||\mathbf{x}_{k-1}||^2$ (recalling $\int |x|^3 K^2 < \infty$) that, for $\ell = 0, 1, 2$,

$$\frac{c_n}{n}\sum_{k=1}^n |\Delta_{2,nk}| \, ||\mathbf{x}_{k-1}||^2 \, |{}_{(\ell)}K_{kn}^2|$$

$$\leq \frac{C\,c_n}{n}\sum_{k=1}^{n}||\mathbf{x}_{k-1}||^2\left[o_P(1)|_{(\ell)}K_{kn}^2|+c_n^{-1}|\,_{(\ell+1)}K_{kn}^2|\right]=o_P(1), \tag{A.18}$$

due to $c_n \to \infty$. Now (A.13) will follow if we prove, for $\ell \leq 2$ and any $\alpha \in \mathbb{R}^p$, that

$$\frac{c_n}{n}\sum_{k=1}^{n}\sigma_k^2u_k^2\left(\alpha'\mathbf{x}_{k-1}\right)^2\,_{(\ell)}K_{kn}^2=E\left[\sigma_1^2\left(\alpha'\mathbf{x}_0\right)^2\right]\int x^\ell K^2+o_P(1). \tag{A.19}$$

Indeed, by (A.19), we have

$$\begin{aligned}&\frac{c_n}{n}\sum_{k=1}^{n}\sigma_k^2u_k^2\,|\Delta_{1,nk}|\,||\mathbf{x}_{k-1}||^2|_{(\ell)}K_{kn}^2|\\ \leq\;& C\,\frac{c_n}{n}\sum_{k=1}^{n}\sigma_k^2u_k^2\,||\mathbf{x}_{k-1}||^2\left(o_P(1)|_{(\ell)}K_{kn}^2|+c_n^{-2}|_{(\ell+1)}K_{kn}^2|\right)=o_P(1),\end{aligned}$$

for $\ell = 0, 1, 2$. This, together with (A.18), yields that, for $\ell = 0, 1, 2$,

$$\begin{aligned}&\frac{c_n}{n}\sum_{k=1}^{n}|\Delta_{nk}|\,||\mathbf{x}_{k-1}||^2\,|_{(\ell)}K_{kn}^2|\\ \leq\;&\frac{c_n}{n}\sum_{k=1}^{n}\sigma_k^2u_k^2\,|\Delta_{1,nk}|\,||\mathbf{x}_{k-1}||^2\,|_{(\ell)}K_{kn}^2|+\frac{c_n}{n}\sum_{k=1}^{n}|\Delta_{2,nk}|\,||\mathbf{x}_{k-1}||^2|_{(\ell)}K_{kn}^2|\\ =\;&o_P(1).\end{aligned}$$

Now, by (A.17) and (A.19), we have

$$\begin{aligned}&\frac{c_n}{n}\sum_{k=1}^{n}\tilde{e}_k^2\mathbf{x}_{k-1}\mathbf{x}_{k-1}'\,_{(\ell)}K_{kn}^2\\ =\;&\frac{c_n}{n}\sum_{k=1}^{n}\sigma_k^2u_k^2\mathbf{x}_{k-1}\mathbf{x}_{k-1}'\,_{(\ell)}K_{kn}^2+\frac{c_n}{n}\sum_{k=1}^{n}\Delta_{nk}\,\mathbf{x}_{k-1}\mathbf{x}_{k-1}'\,_{(\ell)}K_{kn}^2\\ =\;&\mathbf{\Omega}\int x^\ell K^2+o_P(1),\end{aligned}$$

for $\ell = 0, 1, 2$. Taking this result into (A.16), we obtain (A.13).

We finally prove (A.19). Set $v_k = \sigma_k^2u_k^2[\alpha'\mathbf{x}_{k-1}]^2$, where $\alpha \in \mathbb{R}^p$ and recall that $E\left(u_k^2|\mathcal{F}_{k-1}\right) = 1$ and σ_k are $\mathcal{F}_{k-1}$ measurable. It is readily seen that $A_0 := E\left(\sigma_1^2[\alpha'\mathbf{x}_0]^2\right) = Ev_k$ for each $k \geq 1$. Using Lemma 1, it suffices showing that

$$\begin{aligned}\max_{m\leq j\leq n-m}E\left|\frac{1}{m}\sum_{k=j+1}^{j+m}v_k-A_0\right|\;\leq\;&\max_{m\leq j\leq n-m}E\left|\frac{1}{m}\sum_{k=j+1}^{j+m}\sigma_k^2[\alpha'\mathbf{x}_{k-1}]^2\left[u_k^2-E\left(u_k^2|\mathcal{F}_{k-1}\right)\right]\right|\\ &+\max_{m\leq j\leq n-m}E\left|\frac{1}{m}\sum_{k=j+1}^{j+m}\left\{\sigma_k^2[\alpha'\mathbf{x}_{k-1}]^2-E(\sigma_k^2[\alpha'\mathbf{x}_{k-1}]^2)\right\}\right|\to 0.\end{aligned}$$

The asymptotic negligibility of the second term on the r.h.s. above follows directly from Lemma 1 -recalling **A2**. To show the negligibility of the first term, first suppose that **A4** b(i) holds i.e. $\sup_k Eu_k^4 < \infty$. Set $\lambda_k = \sigma_k^2[\alpha'\mathbf{x}_{k-1}]^2$ and $U_k = u_k^2 - E\left(u_k^2|\mathcal{F}_{k-1}\right)$. Then for all $A > 0$ as $m \to \infty$ first and then as $A \to \infty$,

$$\begin{aligned} E\left|\frac{1}{m}\sum_{k=j+1}^{j+m}\lambda_k U_k\right| &\leq E\left|\frac{1}{m}\sum_{k=j+1}^{j+m}\lambda_k I\left\{\lambda_k \leq A\right\}U_k\right| + E\left|\frac{1}{m}\sum_{k=j+1}^{j+m}\lambda_k I\left\{\lambda_k > A\right\}U_k\right| \\ &\leq A\left\{\frac{1}{m^2}E\sum_{k=j+1}^{j+m}U_k^2\right\}^{1/2} + \frac{1}{m}E\sum_{k=j+1}^{j+m}\lambda_k I\left\{\lambda_k > A\right\}\left(E(u_k^2 \mid \mathcal{F}_{k-1}) + 1\right) \\ &\leq A\left\{\frac{1}{m}\sup_k Eu_k^4\right\}^{1/2} + 2E\lambda_1 I\left\{\lambda_1 > A\right\} \to 0. \end{aligned}$$

Next, suppose that **A4** b(ii) holds i.e. $Y_k = \sigma_k^2[\alpha'\mathbf{x}_{k-1}]^2\left[u_k^2 - E\left(u_k^2|\mathcal{F}_{k-1}\right)\right]$ is uniformly integrable. In this case we have

$$\max_{m\leq j\leq n-m} E\left|\frac{1}{m}\sum_{k=j+1}^{j+m}Y_k\right| \to 0 \text{ as } m \to \infty. \tag{A.20}$$

To see this note that $E\left(Y_k|\mathcal{F}_{k-1}\right) = 0$ and for all $A > 0$ set

$$X_k = Y_k 1\left\{|Y_k| < A\right\} - E\left(Y_k 1\left\{|Y_k| < A\right\}|\mathcal{F}_{k-1}\right)$$

and

$$Z_k = Y_k 1\left\{|Y_k| \geq A\right\} - E\left(Y_k 1\left\{|Y_k| \geq A\right\}|\mathcal{F}_{k-1}\right).$$

It can be easily checked that

$$Y_k = X_k + Z_k.$$

In view of this as $m \to \infty$ first and then as $A \to \infty$ we get

$$E\left(\frac{1}{m}\sum_{k=j+1}^{j+m}X_k\right)^2 = \frac{1}{m^2}\sum_{k=j+1}^{j+m}EX_k^2 \leq A^2/m \to 0,$$

and

$$E\left|\frac{1}{m}\sum_{k=j+1}^{j+m}Z_k\right| \leq \frac{2}{m}E\sum_{k=j+1}^{j+m}|Y_k|\,1\left\{|Y_k| \geq A\right\} \leq 2\max_{k\in\mathbb{N}} E\,|Y_k|\,1\left\{|Y_k| \geq A\right\} \to 0,$$

as required. The proof of Theorem 5 is now complete. □

A.5 Proofs of Theorems 6 and 7

We only prove Theorem 6. In relation to Theorem 6, the approach taken in the proof of Theorem 7 is similar to that of Theorem 5 and the details are omitted.

Note that the LLev-IV estimator is of the form

$$\hat{\theta}_{IV}(\tau) = \left[\sum_{k=1}^{n} \mathbf{z}_k \mathbf{x}_k' K_{kn}\right]^{-1} \sum_{k=1}^{n} y_k \mathbf{z}_k K_{kn}.$$

Set $Q_n = \frac{c_n}{n} \sum_{k=1}^{n} \mathbf{z}_k \mathbf{x}_k' K_{kn}$ and write

$$\hat{\theta}_{IV}(\tau) - \theta(\tau) = Q_n^{-1} \left(\mathcal{M}_n + R_n\right), \tag{A.21}$$

where $\mathcal{M}_n = \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} e_k \mathbf{z}_k K_{kn}$ and

$$R_n := \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \theta(k/n)' \mathbf{x}_k \mathbf{z}_k K_{kn} - \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \mathbf{z}_k \mathbf{x}_k' \theta(\tau) K_{kn}.$$

Using arguments similar to those used in the proof of Theorem 3 we get

$$\mathcal{M}_n \to_d \mathbf{N}\left(\mathbf{0}, \mathbf{\Omega}_3\right), \ \mathbf{\Omega}_3 = E\left(\sigma_1^2 \mathbf{z}_1 \mathbf{z}_1' \int K^2\right).$$

Further, it follows directly from Lemma 1 that

$$Q_n \to_P \int K \ E\mathbf{z}_1 \mathbf{x}_1' = Q_3,$$

and

$$\begin{aligned}
\|R_n\| &= \left\| \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \theta(k/n)' \mathbf{x}_k \mathbf{z}_k K_{kn} - \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \mathbf{z}_k \mathbf{x}_k' \theta(\tau) K_{kn} \right\| \\
&\leq \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \left| \left\{ \theta(k/n)' \mathbf{x}_k - \theta(\tau)' \mathbf{x}_k \right\} \right| \|\mathbf{z}_k\| K_{kn} \\
&\leq \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \left| \left\{ \theta(k/n)' \mathbf{x}_k - \theta(\tau)' \|\mathbf{x}_k\| \right\} \right| \|\mathbf{z}_k\| K_{kn} \\
&\leq C \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} |(k/n - \tau) \|\mathbf{x}_k\|| \|\mathbf{z}_k\| K_{kn} \\
&= C \sqrt{\frac{n}{c_n^{1+2\gamma}}} \frac{c_n}{n} \sum_{k=1}^{n} \|\mathbf{x}_k\| \|\mathbf{z}_k\| \left[c_n(k/n - \tau)\right]^{\gamma} K_{kn} = O_P\left(\sqrt{\frac{n}{c_n^{1+2\gamma}}}\right) = o_P(1),
\end{aligned}$$

where we have used condition (e) and the fact that $\frac{c_n}{n} \sum_{k=1}^{n} \|\mathbf{x}_k\| \|\mathbf{z}_k\| \left[c_n(k/n - \tau)\right]^{\gamma} K_{kn} \to_P \int x^{\gamma} K(x) dx E \|\mathbf{x}_1\| \|\mathbf{z}_1\|$ (cf. Lemma 1) and $n/c_n^{1+2\gamma} \to 0$. Taking these facts into (A.21), we establish (25) of the main paper. □

B Supporting Results for Section 2

In this section, we provide proofs for the results discussed in Section 2 of the main paper. We first assume model (2), i.e.,

$$y_k = \beta(k/n)x_{k-1} + u_k,$$

and **Assumption P** hold. The following result demonstrates the asymptotic power of OLS based t-tests for the predictability hypothesis, under neglected time variation in the slope parameter.

Lemma 2. *Under **Assumption P** equations (3) and (4) in main paper hold, i.e.,*

$$\tilde{\beta}_{OLS} \to_P \int_0^1 \beta(\tau)d\tau,$$

and

$$n^{-1/2}\tilde{t}_{OLS} = \frac{n^{-1/2}\tilde{\beta}_{OLS}}{\sqrt{\hat{\sigma}_u^2\left[\sum_{k=1}^n x_{k-1}^2\right]^{-1}}} \to_P \frac{\int_0^1 \beta(\tau)d\tau}{\sqrt{\sigma_*^2\left[Ex_1^2\right]^{-1}}},$$

where $\hat{\sigma}_u^2 = \frac{1}{n}\sum_{k=1}^n \left[y_k - \tilde{\beta}_{OLS}x_{k-1}\right]^2$ *and* σ_*^2 *is the pseudo-true value*

$$\sigma_*^2 = \left[\int_0^1 \beta^2(\tau)d\tau - \left(\int_0^1 \beta(\tau)d\tau\right)^2\right] Ex_1^2 + \sigma_u^2.$$

Proof. Note that

$$\tilde{\beta}_{OLS} = \frac{\sum_{k=1}^n y_k\, x_{k-1}}{\sum_{k=1}^n x_{k-1}^2} = \frac{\sum_{k=1}^n \beta(k/n)x_{k-1}^2}{\sum_{k=1}^n x_{k-1}^2} + \frac{\sum_{k=1}^n x_{k-1}u_k}{\sum_{k=1}^n x_{k-1}^2}.$$

It follows from **Remark A.3** that

$$\frac{1}{n}\sum_{k=1}^n \beta(k/n)x_{k-1}^2 \to_P \int_0^1 \beta(\tau)d\tau E\left(x_1^2\right). \tag{B.1}$$

Further, under the given conditions, it is readily seen that

$$E\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n x_{k-1}u_k\right)^2 = \sigma_u^2 E\left(x_1^2\right),$$

which in turn implies that

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n x_{k-1}u_k = O_P(1). \tag{B.2}$$

In view of (B.1) and (B.2), and the LLN for strictly stationary ergodic sequences (e.g. Shiryaev

(1996); Theorem 3, p. 413) we have

$$\begin{aligned}\tilde{\beta}_{OLS} &= \frac{\sum_{k=1}^{n}\beta(k/n)x_{k-1}^2}{\sum_{k=1}^{n}x_{k-1}^2}+\frac{\sum_{k=1}^{n}x_{k-1}u_k}{\sum_{k=1}^{n}x_{k-1}^2}\\ &= \frac{\sum_{k=1}^{n}\beta(k/n)x_{k-1}^2}{\sum_{k=1}^{n}x_{k-1}^2}+O_P(n^{-1/2})\to_P\int_0^1\beta(\tau)d\tau,\end{aligned}$$

as required for (3) in the main paper.

To prove (4) in the main paper, we first prove that $\hat{\sigma}_u^2 \to_P \sigma_*^2$. Set

$$\begin{aligned}\hat{\sigma}_u^2 &= T_{1n}+T_{2n}+T_{3n}\\ &=: \frac{1}{n}\sum_{k=1}^{n}\left\{\left[\beta(k/n)-\tilde{\beta}_{OLS}\right]x_{k-1}\right\}^2+\frac{2}{n}\sum_{k=1}^{n}\left\{\left[\beta(k/n)-\tilde{\beta}_{OLS}\right]x_{k-1}\right\}u_k+\frac{1}{n}\sum_{k=1}^{n}u_k^2.\end{aligned}$$

Notice that Riemann integrability of β on $[0,1]$ implies Riemann integrability of β^2 on the same set.[1] In view of this, a simple binomial expansion of the first term above together with **Remark A.3** yields

$$T_{1n}\to_P\left\{\int_0^1\beta(\tau)^2d\tau-\left[\int_0^1\beta(\tau)d\tau\right]^2\right\}E\left(x_1^2\right).$$

Further, noting that

$$E\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\beta(k/n)x_{k-1}u_k\right)^2=\sigma_u^2E\left(x_1^2\right)\frac{1}{n}\sum_{k=1}^{n}\beta(k/n)^2=\sigma_u^2E\left(x_1^2\right)\int_0^1\beta(\tau)^2d\tau+o(1)$$

together with (B.2) and the fact that $\tilde{\beta}_{OLS}=O_P(1)$ we get

$$T_{2n}=O_P(1).$$

Finally, a standard argument yields $T_{3n}\to_P\sigma_u^2$. In view of the above, $\hat{\sigma}_u^2\to_P\sigma_*^2$ and hence

$$n^{-1/2}\tilde{t}_{OLS}=\frac{\tilde{\beta}_{OLS}}{\sqrt{\hat{\sigma}_u^2\left[\frac{1}{n}\sum_{k=1}^{n}x_{k-1}^2\right]^{-1}}}\to_P\frac{\int_0^1\beta(\tau)d\tau}{\sqrt{\sigma_*^2\left[Ex_1^2\right]^{-1}}},$$

as required. □

We next consider model (5) in the main paper, i.e.,

$$y_k=\mu(k/n)+\beta x_{k-1}+u_k,$$

together with **Assumption S**. The following lemma demonstrates the size distortions associated with OLS t-tests under $H_0:\beta=\beta_0\in\mathbb{R}$ when time variation in the intercept is neglected and

[1] Riemann integrability of β implies that β is bounded. In view of this and Lebesgue's criterion for Riemann integrability (e.g. Apostol (1981; Thm 7.48)), it follows that β^2 is also Riemann integrable.

the predictor is a stationary long memory process.

Lemma 3. *Suppose that* ***Assumption S (a)-(d.i)*** *holds.*

Then equations (6)-(9) in the main paper hold. Furthermore, under $H_0 : \beta = \beta_0 \in \mathbb{R}$ *we have*

$$\frac{n^{1/2}}{\delta_n}\tilde{t}_{OLS} = \frac{\delta_n}{n^{1/2}}\frac{\tilde{\beta}_{OLS} - \beta_0}{\sqrt{\widetilde{\sigma}_u^2\left[\sum_{k=1}^n x_{k-1}^2\right]^{-1}}} \rightarrow_d \left[1, -\int_0^1 \mu(\tau)d\tau\right] \cdot \mathbf{N}\left(0, \frac{1}{\sigma_+^2 E\left(x_1^2\right)}\Psi\right), \tag{B.3}$$

where δ_n, Ψ *are given as in Section 2 of the main paper,* $\widetilde{\sigma}_u^2 = \frac{1}{n}\sum_{k=1}^n \left[y_k - \tilde{\mu}_{OLS} - \tilde{\beta}_{OLS}x_{k-1}\right]^2$ *and* σ_+^2 *is the pseudo-true value*

$$\sigma_+^2 = \int_0^1 \mu^2(\tau)d\tau - \left(\int_0^1 \mu(\tau)d\tau\right)^2 + \sigma_u^2.$$

As a consequence, result (10) in the main paper holds true as well.

Proof. We start with the verification of (6) and (7) that appear in the main paper. First note that

$$\begin{aligned}
&\begin{bmatrix} \tilde{\mu}_{OLS} \\ \tilde{\beta}_{OLS} \end{bmatrix} - \begin{bmatrix} n^{-1}\sum_{k=1}^n \mu(k/n) \\ \beta \end{bmatrix} \\
= &\left\{\sum_{k=1}^n \begin{bmatrix} 1 & x_{k-1} \\ x_{k-1} & x_{k-1}^2 \end{bmatrix}\right\}^{-1} \times \\
&\sum_{k=1}^n \begin{bmatrix} 1 \\ x_{k-1} \end{bmatrix} \left\{\mu(k/n) + \beta x_{k-1} + u_k - \begin{bmatrix} 1 & x_{k-1} \end{bmatrix} \begin{bmatrix} n^{-1}\sum_{j=1}^n \mu(j/n) \\ \beta \end{bmatrix}\right\} \\
= &\left\{\sum_{k=1}^n \begin{bmatrix} 1 & x_{k-1} \\ x_{k-1} & x_{k-1}^2 \end{bmatrix}\right\}^{-1} \cdot \sum_{k=1}^n \begin{bmatrix} 1 \\ x_{k-1} \end{bmatrix} \left\{\mu(k/n) + u_k - n^{-1}\sum_{j=1}^n \mu(j/n)\right\} \\
= &\left\{\sum_{k=1}^n \begin{bmatrix} 1 & x_{k-1} \\ x_{k-1} & x_{k-1}^2 \end{bmatrix}\right\}^{-1} \cdot \sum_{k=1}^n \begin{bmatrix} u_k \\ x_{k-1}\left\{u_k + \mu(k/n) - n^{-1}\sum_{j=1}^n \mu(j/n)\right\} \end{bmatrix}.
\end{aligned}$$

Set $\Delta_n := n^{-1}\sum_{k=1}^n \left[x_{k-1} - \left(n^{-1}\sum_{j=1}^n x_{j-1}\right)\right]^2$. It follows from Birkhoff's ergodic theorem (cf. Kallenberg (2002), Theorem 10.6) that $\Delta_n \rightarrow_P E(x_1^2)$ and

$$\begin{aligned}
\left\{\frac{1}{n}\sum_{k=1}^n \begin{bmatrix} 1 & x_{k-1} \\ x_{k-1} & x_{k-1}^2 \end{bmatrix}\right\}^{-1} &= \Delta_n^{-1} \begin{bmatrix} n^{-1}\sum_{k=1}^n x_{k-1}^2 & -n^{-1}\sum_{k=1}^n x_{k-1} \\ -n^{-1}\sum_{k=1}^n x_{k-1} & 1 \end{bmatrix} \\
&\rightarrow_P \begin{bmatrix} 1 & 0 \\ 0 & 1/E(x_1^2) \end{bmatrix}.
\end{aligned}$$

In view of these facts, standard arguments show that the intercept estimator $\tilde{\mu}_{OLS}$ satisfies that

$$\sqrt{n}\left(\tilde{\mu}_{OLS} - \int_0^1 \mu(\tau)d\tau\right) = [1 + o_P(1)]\, n^{-1/2} \sum_{k=1}^{n} u_k \to_d \mathbf{N}(0, \sigma_u^2),$$

where we have used the fact that the Euler sum $n^{-1}\sum_{k=1}^{n} \mu(k/n) - \int_0^1 \mu(\tau)d\tau = O(n^{-1})$ since $\mu(.)$ is a bounded variation function. This yields (6) in the main paper.

The verification of (7) in the main paper is similar. Indeed, it is readily seen that

$$\begin{aligned}
& \frac{n}{\delta_n}\left(\tilde{\beta}_{OLS} - \beta\right) \\
= \;& [1 + o_P(1)] \left[E(x_1^2)\right]^{-1} \Bigg\{ \delta_n^{-1} \sum_{k=1}^{n} x_{k-1} u_k + \delta_n^{-1} \sum_{k=1}^{n} \mu(k/n) x_{k-1} \\
& - \left(n^{-1} \sum_{j=1}^{n} \mu(j/n)\right) \left(\delta_n^{-1} \sum_{k=1}^{n} x_{k-1}\right) \Bigg\} \\
= \;& \left[E(x_1^2)\right]^{-1} \left\{ \delta_n^{-1} \sum_{k=1}^{n} \mu(k/n) x_{k-1} - \left(n^{-1} \sum_{j=1}^{n} \mu(j/n)\right) \left(\delta_n^{-1} \sum_{k=1}^{n} x_{k-1}\right) \right\} + o_P(1) \\
\to_d \;& \left(E x_1^2\right)^{-1} \left[1, \; -\int_0^1 \mu(\tau)d\tau\right] \cdot \mathbf{N}\left(\mathbf{0}, \Psi\right), \qquad \text{(B.4)}
\end{aligned}$$

where we have used the result:

$$\delta_n^{-1} \left[\sum_{k=1}^{n} \mu(k/n) x_{k-1}, \; \sum_{k=1}^{n} x_{k-1}\right] \to_d \mathbf{N}\left(\mathbf{0}, \Psi\right). \qquad \text{(B.5)}$$

We next show (B.5) under **Assumption S (d.i)** with the matrix Ψ defined in (8) of the main paper -a similar limit result holds under **Assumption S (d.ii)** but an explicit proof is omitted (see also footnote 7 of the main paper). We commence with the proof of (B.5). Recalling that $x_k = \sum_{j=0}^{\infty} \phi_j \xi_{k-j}$, with $\phi_j \sim c_0\, j^{-\nu}$, $\nu = 1 - d$ and $0 < d < 1/2$, we have

$$S_n' := \sum_{k=1}^{n} \mu\left(k/n\right) x_{k-1} = \sum_{k=-\infty}^{n-1} \sum_{s=k\vee 0}^{n-1} \mu\left((s+1)/n\right) \phi_{s-k} \xi_k,$$

and

$$S_n'' := \sum_{k=1}^{n} x_{k-1} = \sum_{k=-\infty}^{n-1} \sum_{s=k\vee 0}^{n-1} \phi_{s-k} \xi_k.$$

For any fixed $m \in \mathbb{N}$, define

$$S_{n,m}' := \sum_{k=-mn}^{n-1} \sum_{s=k\vee 0}^{n-1} \mu\left((s+1)/n\right) \phi_{s-k} \xi_k, \qquad S_{n,m}'' := \sum_{k=-mn}^{n-1} \sum_{s=k\vee 0}^{n-1} \phi_{s-k} \xi_k.$$

For λ_1, $\lambda_2 \in \mathbb{R}$, by applying Lindeberg-Feller CLT (e.g. Kallenberg (2002), Theorem 5.12)[2],

$$\lambda_1 \delta_n^{-1} S'_{n,m} + \lambda_2 \delta_n^{-1} S''_{n,m}$$

converges to a normal distribution that has asymptotic variance determined by the limit of

$$\begin{aligned}
&\sum_{k=-mn}^{n-1} E\left[\lambda_1 \delta_n^{-1} \sum_{s=k\vee 0}^{n-1} \mu\left((s+1)/n\right) \phi_{s-k}\xi_k + \lambda_2 \delta_n^{-1} \sum_{s=k\vee 0}^{n-1} \phi_{s-k}\xi_k\right]^2 \\
= \;& \sigma_\xi^2 \sum_{k=-mn}^{n-1} \left[\delta_n^{-1} \sum_{s=k\vee 0}^{n-1} (\lambda_1 \mu\left((s+1)/n\right) + \lambda_2)\phi_{s-k}\right]^2 \\
= \;& \frac{1}{nc(d)} \sum_{k=-mn}^{n-1} \left[\frac{1}{n} \sum_{s=k\vee 0}^{n-1} (\lambda_1 \mu\left((s+1)/n\right) + \lambda_2) \left(\frac{s-k}{n}\right)^{-\nu}\right]^2 + o(1) \\
= \;& \frac{1}{c(d)} \int_{-m}^{1} \left[\lambda_1 \int_{r\vee 0}^{1} \mu(s)(s-r)^{-\nu} ds + \lambda_2 \int_{r\vee 0}^{1} (s-r)^{-\nu} ds\right]^2 dr + o(1).
\end{aligned}$$

Noting that $\mu(.)$ is bounded on $[0,1]$ and applying Fatou's lemma yields that

$$\begin{aligned}
& E\left[\lambda_1 \delta_n^{-1}(S'_n - S'_{n,m}) + \lambda_2 \delta_n^{-1}(S''_n - S''_{n,m})\right]^2 \\
= \;& E\Big(\lim_{m'\to\infty} \left[\lambda_1 \delta_n^{-1}(S'_{n,m'} - S'_{n,m}) + \lambda_2 \delta_n^{-1}(S''_{n,m'} - S''_{n,m})\right]^2\Big) \\
\leq \;& \lim_{m'\to\infty} E\left[\lambda_1 \delta_n^{-1}(S'_{n,m'} - S'_{n,m}) + \lambda_2 \delta_n^{-1}(S''_{n,m'} - S''_{n,m})\right]^2 \\
= \;& \sigma_\xi^2 \sum_{k=-\infty}^{-mn-1} E\left[\delta_n^{-1} \sum_{s=0}^{n-1} (\lambda_1 \mu\left((s+1)/n\right) + \lambda_2)\phi_{s-k}\right]^2 \\
\leq \;& \frac{C(|\lambda_1| \max_{0\leq t\leq 1}|\mu(t)| + |\lambda_2|)^2}{n} \sum_{k=-\infty}^{-mn-1} \left[\frac{1}{n}\sum_{s=0}^{n-1}\left(\frac{s-k}{n}\right)^{-\nu}\right]^2 \\
\leq \;& C\int_{-\infty}^{-m} \left[\int_0^1 (s-r)^{-\nu} ds\right]^2 dr \to 0, \quad m\to\infty.
\end{aligned}$$

Combining the facts above, it follows from Theorem 4.28 of Kallenberg (2002) that

$$\lambda_1 \delta_n^{-1} S'_n + \lambda_2 \delta_n^{-1} S''_n \to_d N\left(0, [\lambda_1, \lambda_2]\,\Psi \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}\right)$$

[2]To see this, define

$$a_{n,k} = \delta_n^{-1} \sum_{s=k\vee 0}^{n-1} (\lambda_1 \mu\left((s+1)/n\right) + \lambda_2)\phi_{s-k}, \quad k = -mn, \cdots, n-1,$$

then $a_n := \max_{-mn\leq k\leq n-1} a_{n,k} \to 0$ and hence the Lindeberg condition holds: for any $\epsilon > 0$

$$\sum_{k=-mn}^{n-1} E(a_{nk}^2 \xi_k^2 I(a_{nk}|\xi_k| > \epsilon)) \leq E\Big(\xi_1^2 I\Big(|\xi_1| > \epsilon/a_n\Big)\Big) \sum_{k=-mn}^{n-1} a_{nk}^2 \to 0.$$

where the matrix Ψ is of the form as in (8) of the main paper under **Assumption S (d.i)**. Result (B.5) follows from the Cramér-Wold theorem.

We next verify (9) in the main paper. In fact, since $\tilde{u}_k$ are the OLS residuals, the required result follows from:

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n}\tilde{u}_k^2 &= \frac{1}{n}\sum_{k=1}^{n}\left[\mu(k/n)-\tilde{\mu}+\left(\beta-\tilde{\beta}_{OLS}\right)x_{k-1}+u_k\right]^2 \\
&= \frac{1}{n}\sum_{k=1}^{n}\left[\mu(k/n)-\tilde{\mu}+\left(\beta-\tilde{\beta}_{OLS}\right)x_{k-1}\right]^2+\frac{1}{n}\sum_{k=1}^{n}u_k^2 \\
&\quad +\frac{2}{n}\sum_{k=1}^{n}\left[\mu(k/n)-\tilde{\mu}+\left(\beta-\tilde{\beta}_{OLS}\right)x_{k-1}\right]u_k \\
&\to_P \int_0^1\mu(\tau)^2d\tau-\left(\int_0^1\mu(\tau)d\tau\right)^2+\sigma_u^2=:\sigma_+^2, \qquad \text{(B.6)}
\end{aligned}$$

where we have used the fact that

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n}\left[\mu(k/n)-\tilde{\mu}\right]^2 &= \frac{1}{n}\sum_{k=1}^{n}\mu(k/n)^2-\frac{2\tilde{\mu}}{n}\sum_{k=1}^{n}\mu(k/n)+\tilde{\mu}^2 \\
&\to_P \int_0^1\mu(\tau)^2d\tau-\left(\int_0^1\mu(\tau)d\tau\right)^2.
\end{aligned}$$

We finally prove (B.3) and then complete the proof of Lemma 3. This is simple since, by (B.4), (B.6) and the fact that $n^{-1}\sum_{k=1}^{n}x_k \to_P 0$, the OLS based t-statistic for the null hypothesis $H_0:\beta=\beta_0$, $\beta_0\in\mathbb{R}$ satisfies

$$\begin{aligned}
\frac{\sqrt{n}}{\delta_n}\tilde{t}_{OLS} &= \frac{\sqrt{n}}{\delta_n}\frac{\tilde{\beta}_{OLS}-\beta_0}{\sqrt{\left(\frac{1}{n}\sum_{k=1}^{n}\tilde{u}_k^2\right)\left[\sum_{k=1}^{n}x_k^2-n^{-1}\left(\sum_{k=1}^{n}x_k\right)^2\right]^{-1}}} \\
&= \frac{\sqrt{n}}{\delta_n}\frac{\sqrt{n}\left(\tilde{\beta}_{OLS}-\beta_0\right)}{\sqrt{\left(\frac{1}{n}\sum_{k=1}^{n}\tilde{u}_k^2\right)\left[n^{-1}\sum_{k=1}^{n}x_k^2-\left(n^{-1}\sum_{k=1}^{n}x_k\right)^2\right]^{-1}}} \\
&= \left[1+o_P(1)\right]\frac{\frac{n}{\delta_n}\left(\tilde{\beta}_{OLS}-\beta_0\right)}{\sqrt{\left(\frac{1}{n}\sum_{k=1}^{n}\tilde{u}_k^2\right)\left[n^{-1}\sum_{k=1}^{n}x_k^2\right]^{-1}}} \\
&\to_d \frac{1}{\sqrt{\sigma_+^2E\left(x_1^2\right)}}\left[1,-\int_0^1\mu(\tau)d\tau\right]\cdot\mathbf{N}\left(0,\frac{1}{\sigma_+^2E\left(x_1^2\right)}\Psi\right)
\end{aligned}$$

as required. □

C Simulation Study

We explore the finite sample properties of the proposed nonparametric estimators and related test statistics with the aid of a simulation study. In particular, we consider predictive TVP regressions of the form

$$y_k = \mu(k/n) + \beta(k/n)x_{k-1} + e_k, \tag{C.1}$$

and the following test hypotheses

$$H_0 : \beta(\tau) = 0 \text{ vs } H_1 : \beta(\tau) \neq 0,$$

and

$$H_0 : \partial\mu(\tau)/\partial\tau = 0 \text{ vs } H_1 : \partial\mu(\tau)/\partial\tau \neq 0,$$

with $\tau \in \mathcal{T} \subset (0,1)$. Note that the latter is a time invariance hypothesis about the intercept term. The theoretical results of Section 2 demonstrate that neglecting time variability in the intercept results in power loss and severe size distortions. These findings are corroborated by the simulations.

In all cases the significance level is set at 5% and the number of replication paths is 10,000. For the purposes of this experiment the following vector of innovations is generated

$$\begin{bmatrix} \xi_k \\ u_k \end{bmatrix} \sim i.d.\mathbf{N}\left(\mathbf{0}, \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}\right),$$

$\delta \in (-1,1)$. The predictor is a type II fractional process (e.g. Robinson and Hualde, 2003) of the form

$$(I - L)^d x_k = \xi_k 1\{k \geq 1\}. \tag{C.2}$$

The regression error is

$$e_k = \sigma_k u_k,$$

with either

$$\sigma_k^2 = 1,$$

or

$$\sigma_k^2 = 0.01 + 0.45\sigma_{k-1}^2 + 0.45e_{k-1}^2,\ \sigma_0^2 = 0.01, \tag{C.3}$$

which makes the regression error a strong GARCH(1,1).

We consider the following values for the memory parameter $d = \{0.25, 0.35, 0.45, 0.55\}$. The value $d = 0.55$ is slightly above the nonstationarity threshold ($d = 0.5$) that determines the maximal value of the memory parameter for which the limit distribution of the tests is $\mathbf{N}(0,1)$.[3] For nonstationary predictors, the nonparametric estimators under consideration do not possess mixed Gaussian limit distribution and therefore some size distortion is likely. It is reasonable to

[3] As mentioned before, some preliminary theoretical results suggest that the proposed methods are also valid for weakly nonstationary predictors i.e. long memory with $d = 0.5$ or mildly integrated processes.

expect size distortions become more severe for larger values of the memory and the endogeneity parameters. In certain data sets, some predictors (e.g. realised variance, inflation) appear to be long memory with memory parameter close to 0.5. We therefore consider the value $d = 0.55$ in order to assess the robustness of the proposed methods when predictors are close to the nonstationarity threshold.

The effects of bandwidth choice on the bias and the MSE of the LLev and LLin slope parameter estimators are illustrated in Figure 5 and Figure 6 respectively. The memory order has very little effect on bias, whilst bandwidth choice has a profound effect with substantial bias reduction when under-smoothing is employed. Further, the LLin estimator exhibits superior performance, relative to the LLev estimator for $c_n = n^{0.3}$. Higher memory order is associated with higher MSE, particularly for the LLev estimator. Under-smoothing results in substantial MSE gains for both estimators. Finally, LLin exhibits a better MSE performance relatively to that of the LLev estimator, particularly for smaller sample sizes.

Figure 5: Bias of TVP Slope Estimators (plotted against τ)

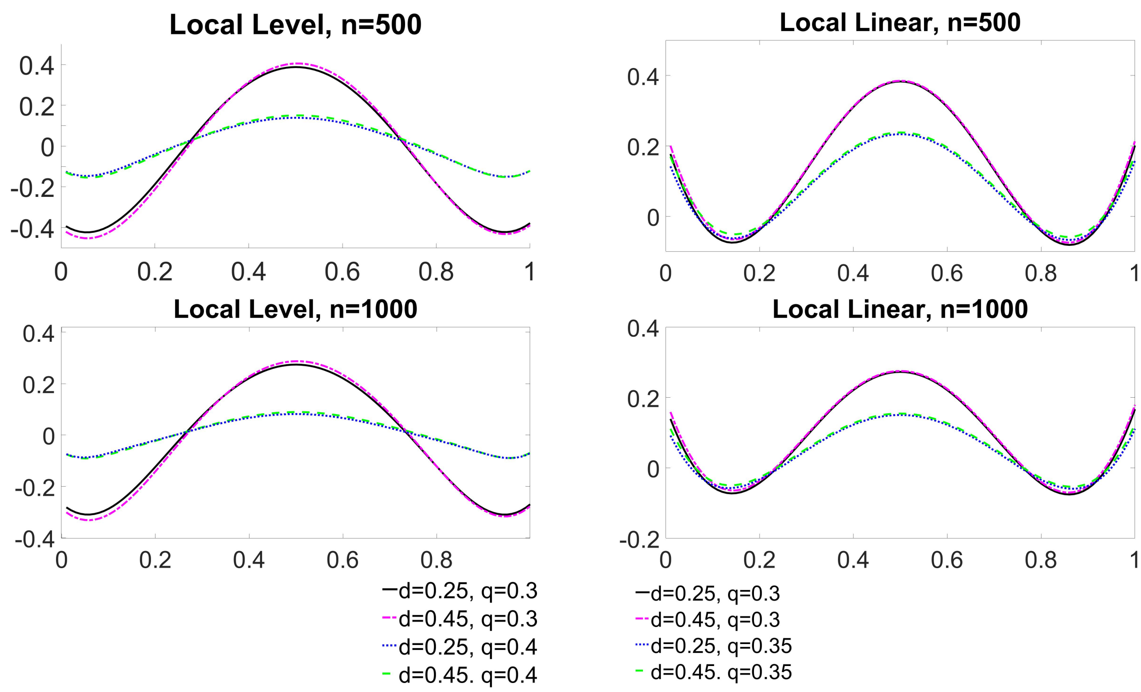
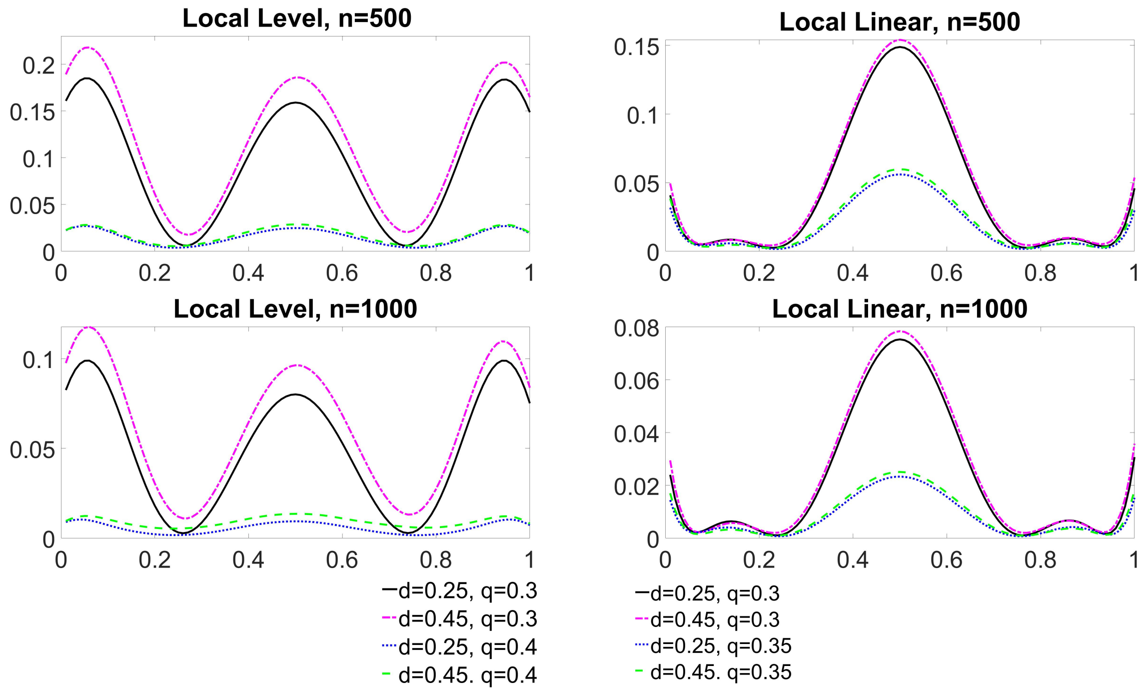


Figure 6: MSE of TVP Slope Estimators (plotted against τ)



Next, we report results for the finite sample performance of non parametric t-tests in the context of predictive regressions as per (28) in the main paper. As mentioned above, we consider two hypotheses. First, the no predictability hypothesis $H_0 : \beta(\tau) = 0,\ \tau \in (0,1)$ against $H_1 : \beta(\tau) \neq 0$. Under H_1 we choose $\beta(.)$ to be either a periodic function, capable of reproducing periodic episodic predictability events, or a smooth transition function that is more relevant when predictability is related to some regime switching event. For this kind of hypothesis we consider both LLev and LLin tests. Second, we test the time invariance hypothesis for the intercept $H_0 : \partial\mu(\tau)/\partial\tau = 0,\ \tau \in (0,1)$ against $H_1 : \partial\mu(\tau)/\partial\tau \neq 0$ using the LLin based test. We consider two possibilities for the exponent of the bandwidth parameter $c_n = n^q$. In particular,

$$q = \begin{cases} 0.3, 0.4, & \text{Local Level} \\ 0.3, 0.35, & \text{Local Linear} \end{cases} .$$

As mentioned before, larger values of for c_n (under-smoothing) provide better size control while smaller values (over-smoothing) result in better power. In preliminary simulations we have also considered additional possibilities for c_n (i.e. $q = \{0.1, 0.2\}$), however we only report results for bandwidth values that appear to yield superior size-power trade-off.

We next specify the intercept and slope parameter functions $\mu(\tau)$ and $\beta(\tau)$ utilised for the predictability hypothesis. Under both the null and the alternative hypothesis the intercept is given by

$$\mu(\tau) = 0.025 \cdot \sin(2\pi\tau).$$

On the other hand the slope parameter is

$$\beta(\tau)=\begin{cases} 0, \text{ under } H_0 \\ \left.\begin{array}{l} b\cdot\cos(2\pi\tau), \text{ or} \\ b\cdot\{1+\exp\left[-30\left(\tau-0.5\right)\right]\}^{-1} \end{array}\right\} \text{under } H_1 \end{cases},$$

with $b=\{0.033, 0.066, 0.099\}$. It should be emphasised that contrary to the fixed parameter case, the estimators under consideration are not numerically invariant to the value of the intercept when the latter is time varying. Therefore, the shape of the intercept function has an impact on the finite sample performance of the tests. Intercept functions that exhibit more abrupt variation are likely to result in more severe size distortions because of larger nonlinearity induced asymptotic bias (see Remark 2 in the main paper). On the other hand smaller variability in the intercept function is associated with smaller asymptotic bias (cf. condition (e) of Theorems 2 and 3 in the main paper. We therefore employ a time varying intercept in order to assess the performance of the proposed tests in situations when there is finite sample bias due to time variation in the intercept estimator. In particular, we choose a sinusoidal function that has period one over $(0,1)$ i.e. the domain of the TVPs. The maximal value of the intercept function in the simulation experiment, for the non predictability hypothesis, is relevant to the empirical application, where we consider TVP predictive regressions with the realised variance as a predictor. We find that the maximal estimates for the intercept are approximately 0.01, 0.02 and 0.05 for monthly, quarterly and annual data respectively. Therefore, 0.025 is a mid-range value. The choice for the slope parameter function is also relevant to our empirical application. In our empirical application, the maximal estimates for the slope parameter of realised variance are approximately, 1.25, 2 and 6 for monthly, quarterly and annual data respectively. Therefore, the particular choice for $\beta(\tau)$ (and b) is likely to give conservative asymptotic power results under the alternative hypothesis.

Figures 7-8 report the empirical size of the LLev and LLin based tests for the non predictability hypothesis for sample sizes $n=\{500, 1000\}$, and $d=\{0.35, 0.45, 0.55\}$. We consider two endogeneity scenarios. First, moderate endogeneity with $\delta=-0.55$ and then very strong endogeneity with $\delta=-0.95$. Size (vertical axis) is plotted against various values of $\tau\in(0,1)$ (horizontal axis). In general, higher values for the memory parameter and strong endogeneity lead to size distortions. It can be seen that size control is reasonably good even when $d=0.55$ (i.e. slightly above the stationarity boundary) with small oversizing when $\delta=-0.55$ and moderate oversizing when $\delta=-0.95$. Additional simulations, not reported here, show that when the intercept is fixed over time, size is slightly better than that in Figures 9 and 10. Moreover, for smaller values of d and $|\delta|$ preliminary simulations show that empirical size is closer to the nominal one. Finally, as mentioned above abrupt changes in the intercept parameter may cause size distortions. It seems however the tests perform reasonably well in this respect, in particular when under-smoothing is employed (see also Figure 16 and the discussion below).

Figure 7: Empirical Size of t-tests against τ: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 500$; $\delta = -0.55$; fractional regressor, GARCH(1,1) regression errors)

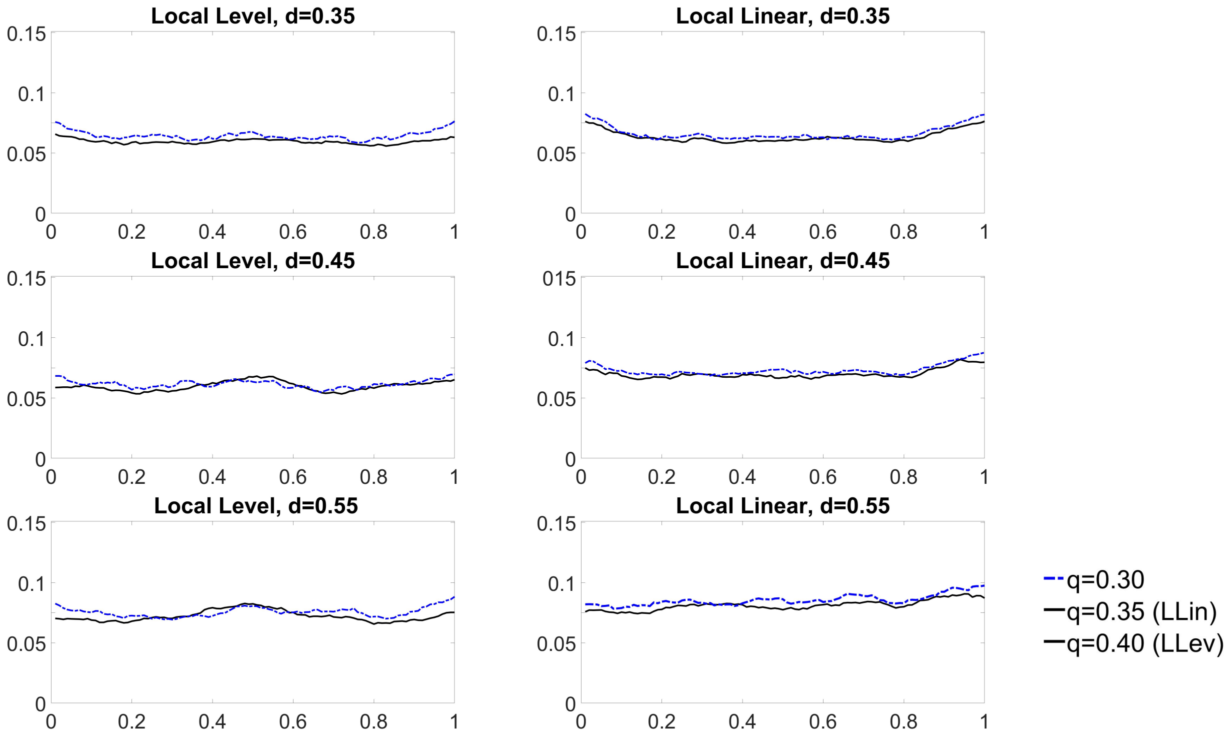


Figure 8: Empirical Size of t-tests against τ: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 1000$; $\delta = -0.55$; fractional regressor, GARCH(1,1) regression errors)

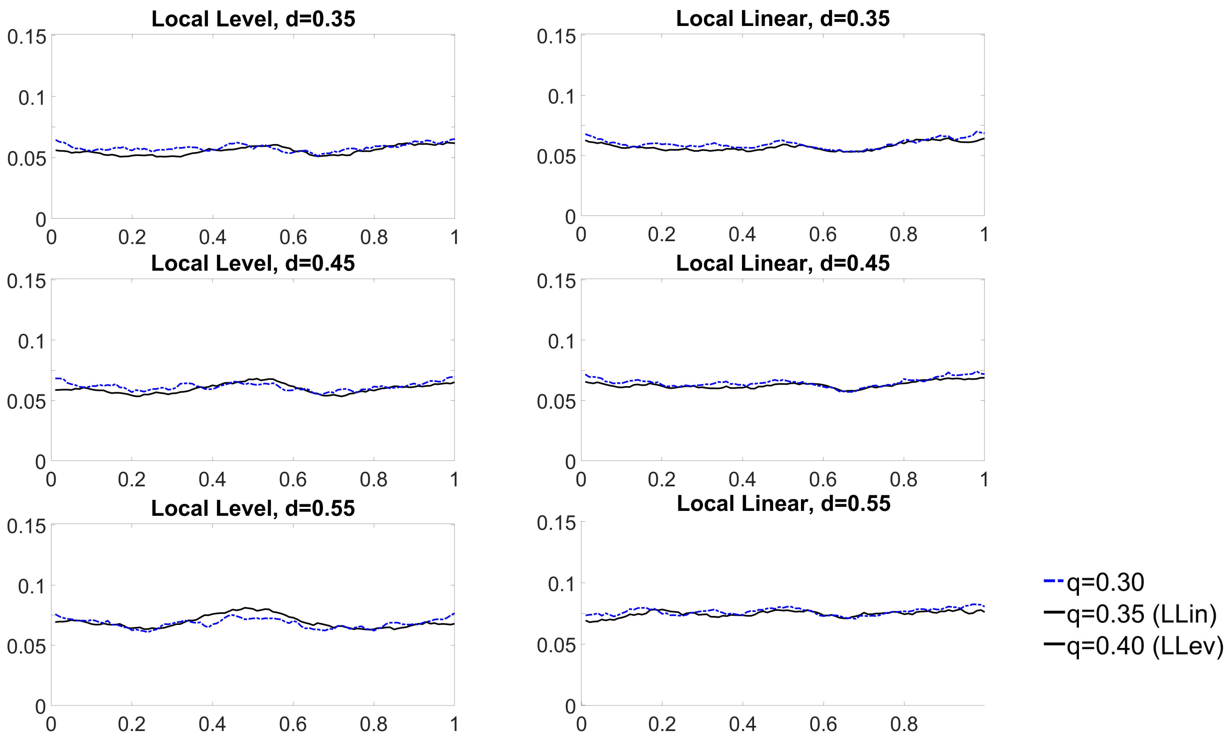


Figure 9: Empirical Size of t-tests against τ: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 500$; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

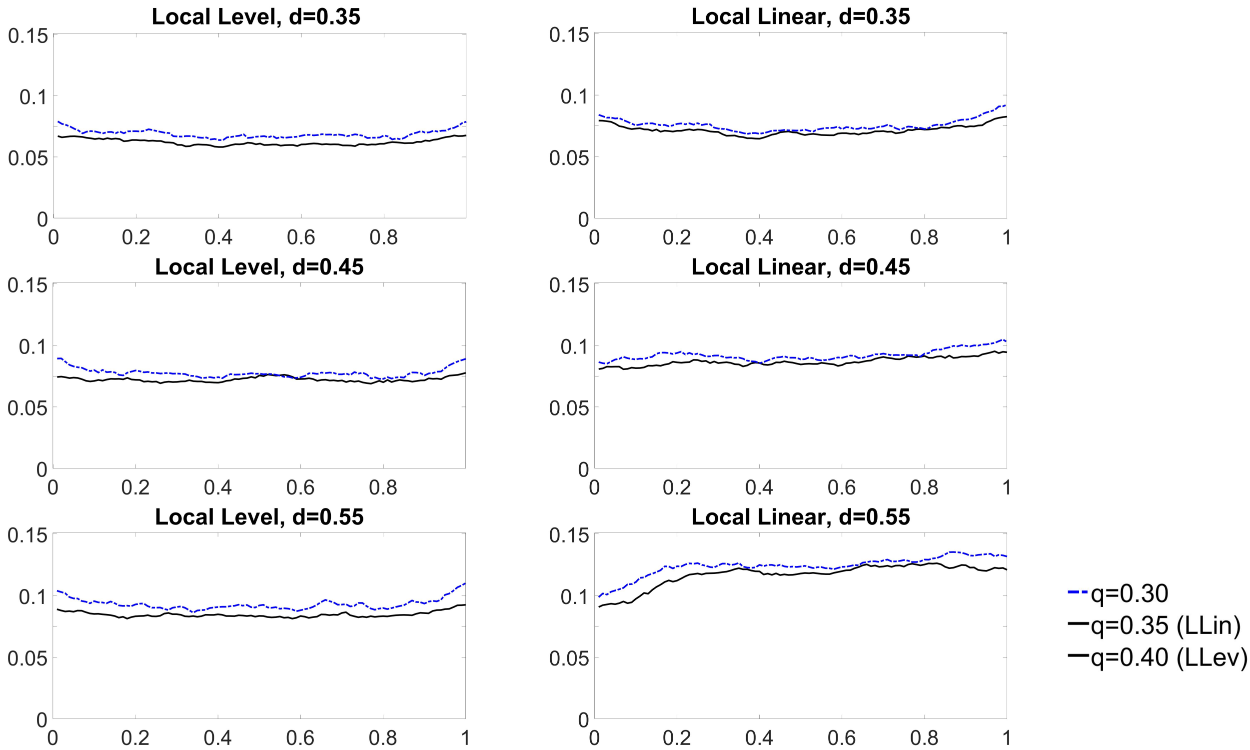
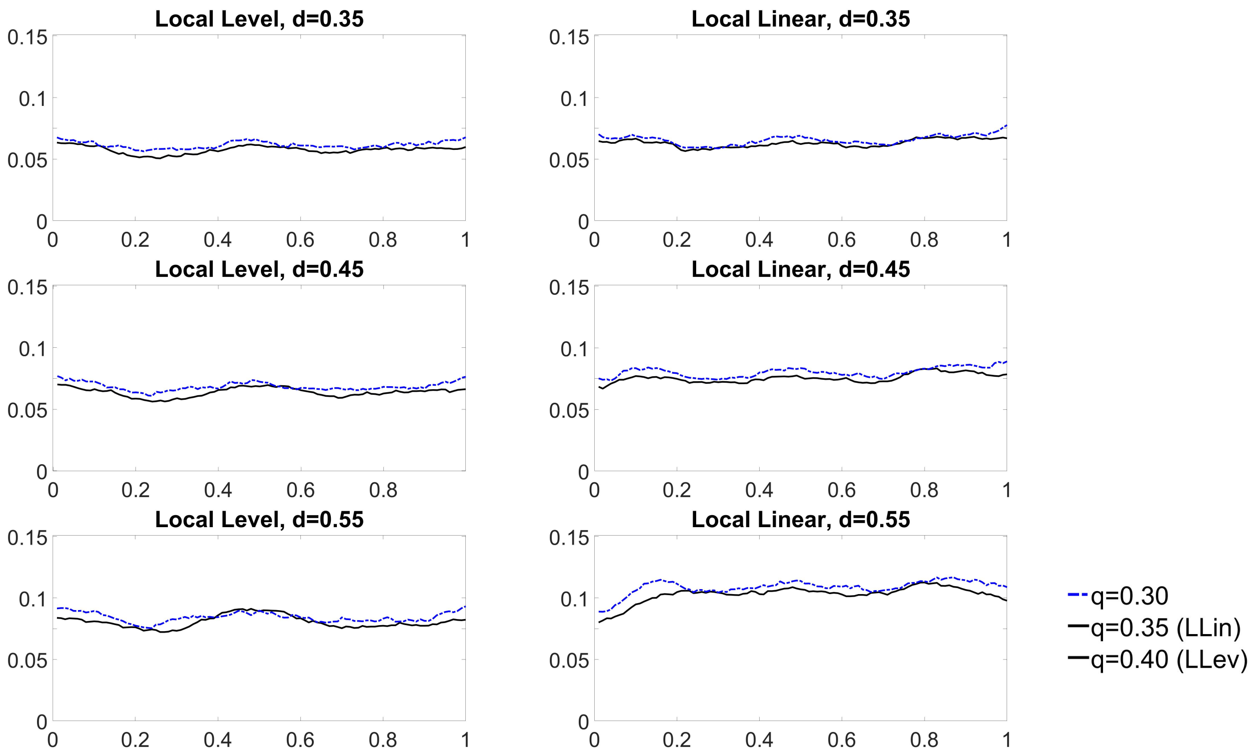


Figure 10: Empirical Size of t-tests: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 1000$; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)



The empirical power of both tests is reported in Figures 11 and 12, for $d = 0.45$. Under the alternative, for $\beta(\tau) = b \cdot \cos(2\pi\tau)$, power peaks at $\tau = 0, 0.5, 1$, approximately. These locations correspond to the extrema of the cosine slope parameter function. There are small differences between the LLev and LLin tests, and the two bandwidth choices. For $\beta(\tau) = b \cdot \{1 + \exp[-30(\tau - 0.5)]\}^{-1}$, it seems that the LLev performs better than the LLin test, particularly at boundary points. Note that the LLin test exhibits some power drop for τ close to one. In all cases power improves when sample increases, as expected.

Figure 11: Empirical Power of t-tests: $H_1 : \beta(\tau) = b \cdot \cos(2\pi\tau)$
(5% nominal size; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

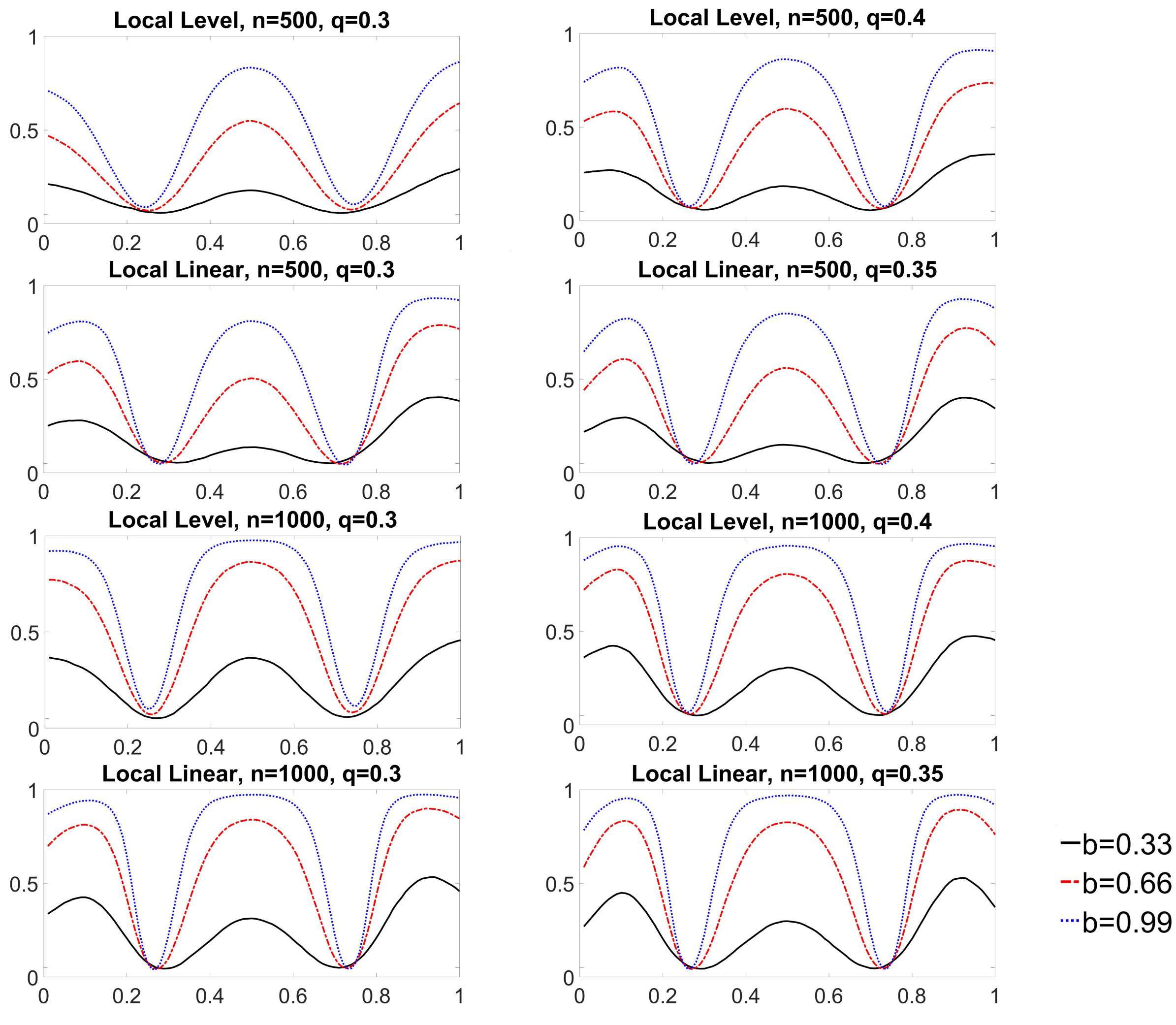
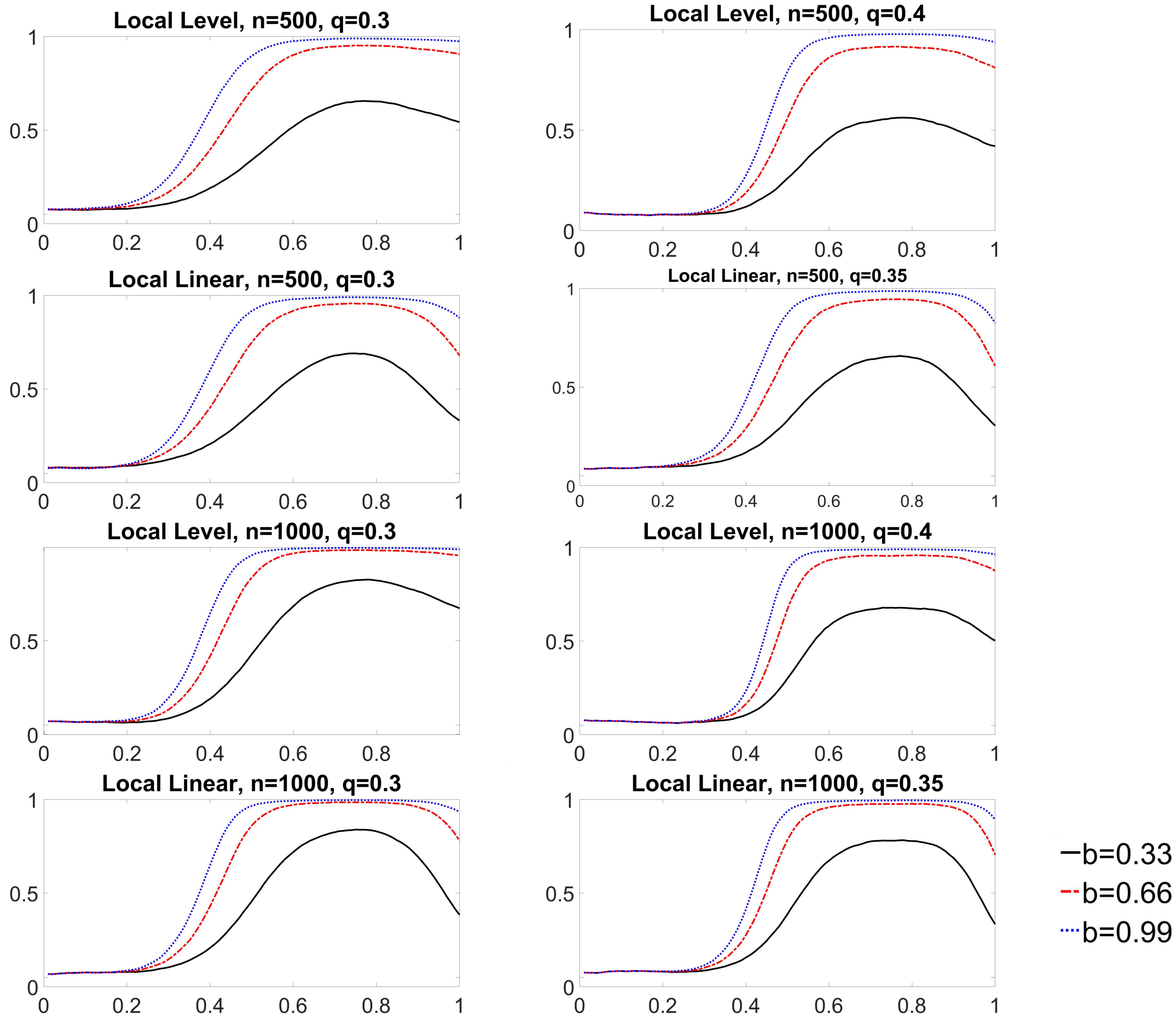


Figure 12: Empirical Power of t-tests: $H_1 : \beta(\tau) = b \cdot \{1 + \exp[-30(\tau - 0.5)]\}^{-1}$
(5% nominal size; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)



We next consider the finite sample performance of the LLin test for the hypotheses $H_0 : \partial\mu(\tau)/\partial\tau = 0$ i.e. the regression intercept is invariant with respect to time. The test statistic in this case relies on the estimator for the derivative of $\mu(\tau)$ which attains a slower convergence rate (i.e. $\sqrt{n/c_n^3}$) than that of the regression parameters $\mu(\tau)$ and $\beta(\tau)$. Therefore, it is reasonable to expect that the power of the time invariance test is inferior to that for the no predictability hypothesis considered earlier.

To assess the size of the test under the null hypothesis, we generate data from (28) in the main paper with $\mu(\tau) = 0.025$ and $\beta(\tau) = 0.66 \cdot \cos(2\pi\tau)$. Note that the slope parameter is chosen to be time varying. Time variation in the slope parameter induces nonlinearity asymptotic bias (see Remark 2 in the main paper) which is likely to result in some size distortions. Figure 13 reports the empirical size of the test for various values of the memory parameter and different sample sizes. As before, the exponent of the bandwidth term is $q = \{0.3, 0.35\}$. Size is in general close to the nominal one with somewhat more substantial over-sizing when the predictor is nonstationary. It is worth noting that some variation in empirical size with respect to time is evident that appears to resemble the time variation in the slope parameter. This is likely to be due to nonlinearity induced asymptotic bias in slope parameter estimates.

Figure 13: Empirical Size of t-tests: $H_0 : \partial\mu(\tau)/\partial\tau = 0$
(5% nominal size; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

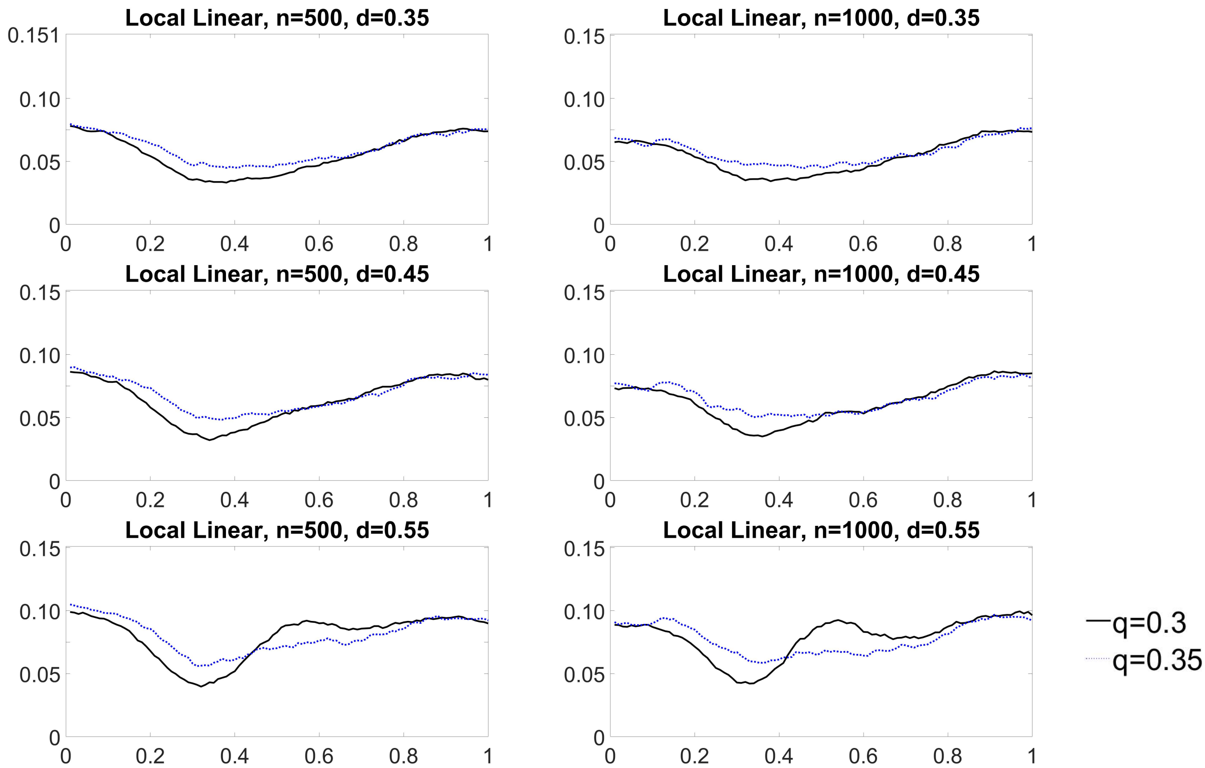
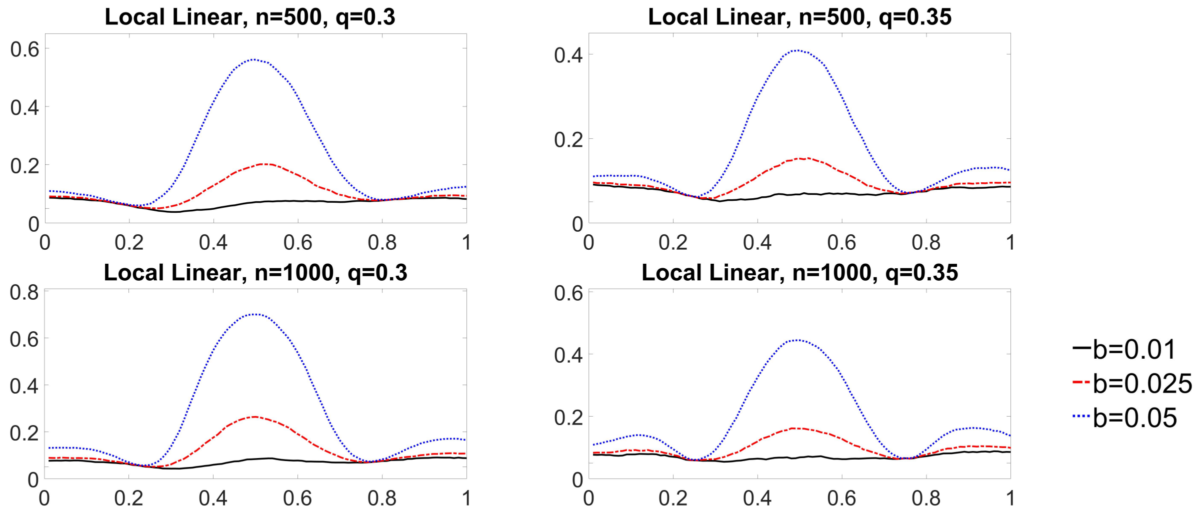


Figure 14 reports the rejection frequency of the latter test under the alternative hypothesis. In particular, the regression parameters are $\mu(\tau) = b \cdot \sin(2\pi\tau)$ with $b = \{0.01, 0.025, 0.05\}$, and $\beta(\tau) = 0.066 \cdot \cos(2\pi\tau)$. The memory of the predictor is $d = 0.45$ and as before we consider two sample sizes. The time invariance test is less powerful than the predictability test considered earlier. Notably, there is a substantial power drop at boundary points. Note that under $H_1 : \partial\mu(\tau)/\partial\tau = 2\pi b \cdot \cos(2\pi\tau)$. Therefore, the derivative function assumes its maximum values at $\tau = \{0, 0.5, 1\}$. At boundary points power is very poor. This is likely due to asymptotic bias in derivative estimation at boundary points (cf. Figure 1 in main paper). Hence, the test appears to be quite conservative in terms of power, when there is substantial variation in the parameter at boundary points. However, this test can be easily implemented in conjunction with the predictability test. Better performance could be possibly achieved with the utilisation of higher order kernels (e.g. local quadratic estimation) that may result in further bias reduction. Tests for time variation in the parameters of predictive regressions is an important topic on its own. We therefore leave further developments in this area for future work.

Figure 14: Empirical Power of t-tests: $H_1 : \partial\mu(\tau)/\partial\tau = 2\pi b \cdot \cos(2\pi\tau)$
(5% nominal size; $\delta = -0.95$; $d = 0.45$, GARCH(1,1) regression errors)



We conclude this section with some results for OLS based t-tests for the predictability hypothesis when time variability in regression parameters is neglected. We first consider the size of OLS based t-tests for the hypothesis $H_0 : \beta(\tau) = 0$, when $\mu(\tau) = 0.25 \cdot \sin(2\pi\tau)$ i.e. there is neglected variation in the regression intercept. We compare the size of the OLS based test with that based on the LLev estimator. It has been demonstrated in Section 2 that the conventional t-statistic is divergent in this case when the memory parameter is strictly greater than zero. Further, divergence rates are faster when memory is longer. These theoretical findings are confirmed by the empirical size reported in Figure 16. It is worth noting that the LLev exhibits some oversizing for $d = 0.45$ when over-smoothing is employed. Note that in this case the intercept parameter is more volatile than the one considered in Figures 7-8. It seems that long memory, in conjunction with high variation in the slope parameter, exacerbates finite sample bias. Nevertheless, when under-smoothing is employed (i.e. $q = 0.4$) empirical size is close the nominal one. Finally, Figure 15 reports asymptotic power when the slope parameter is either a sinusoid or a smooth transition function ($\mu(\tau)$ as above). Notice that in almost all cases the LLev test outperforms the OLS test by a substantial margin.

Figure 15: Empirical Power of OLS and LLev t-tests.
upper panel: $H_1 : \beta(\tau) = 0.2 \cdot cos(2\pi\tau)$
lower panel: $H_1 : \beta(\tau) = 0.15 \cdot \{1 + \exp[-30(\tau - 0.5)]\}^{-1}$
(5% nominal size; $\delta = -0.95$; $d = 0.35$, $i.d.N(0, 1)$ regression errors)

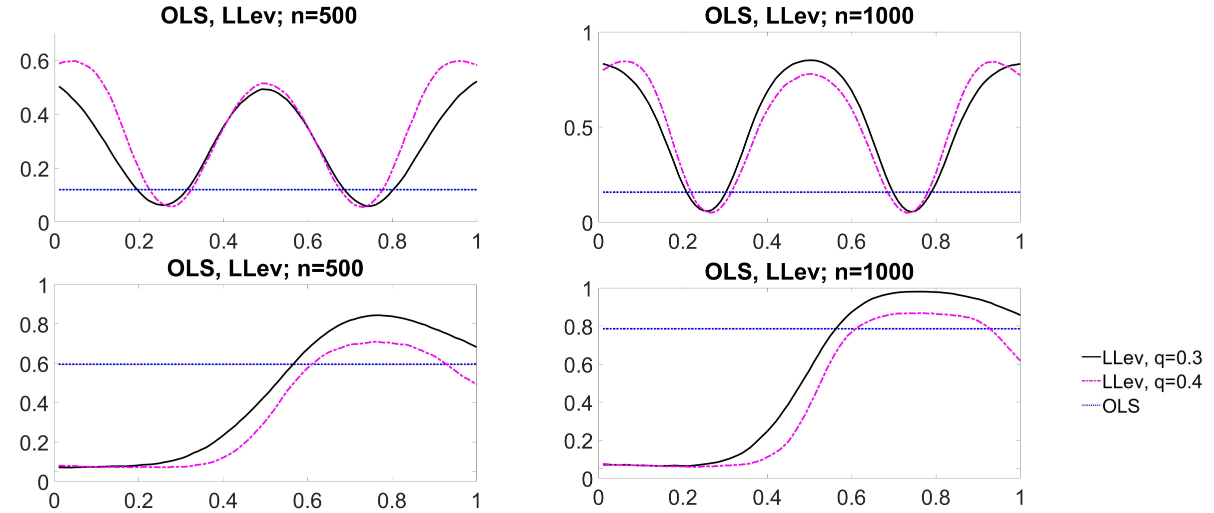
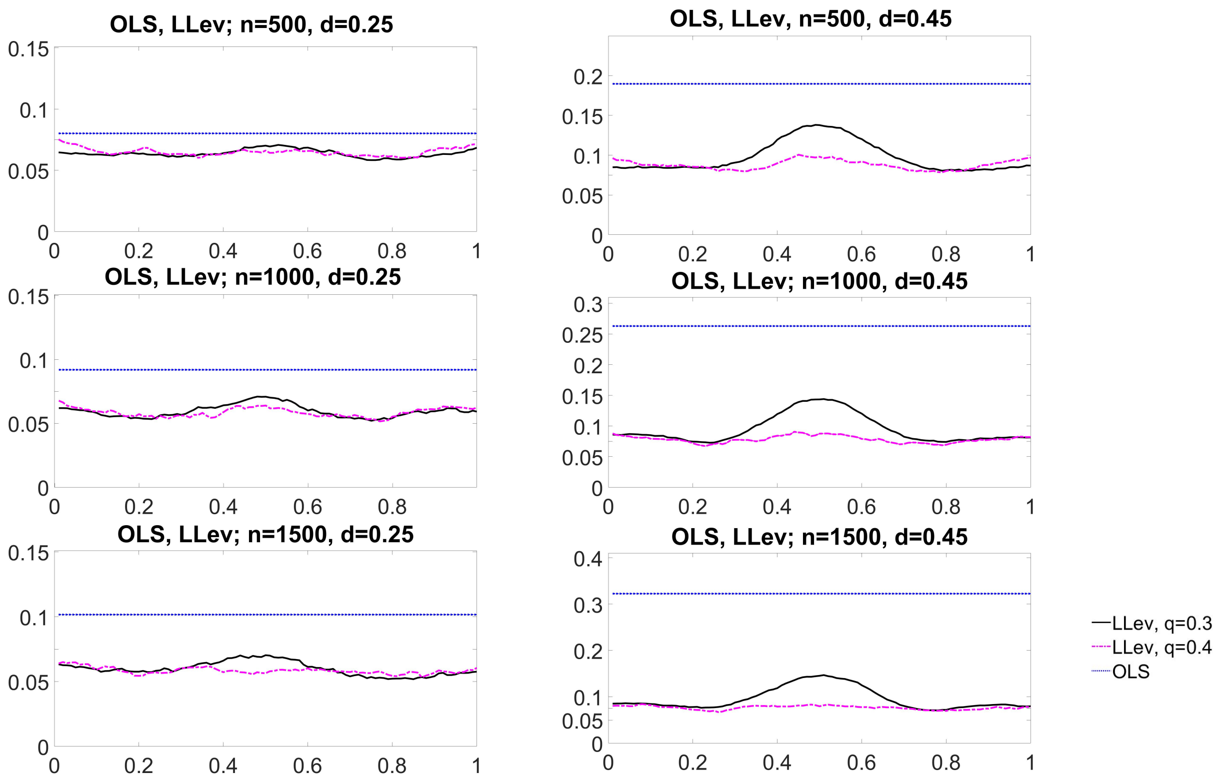


Figure 16: Empirical Size of OLS and LLev based t-tests: τ: $H_0 : \beta(\tau) = 0$
(5% nominal size; $\delta = -0.95$; $i.d.N(0, 1)$ regression errors)



References

[1] Apostol, T.M. (1981). Mathematical Analysis. Addison-Wesley.

[2] Folland, G.B. (1999). Real Analysis. John Wiley & Sons, New York.

[3] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

[4] Hu, Z., Phillips, P.C.B. and Wang, Q. (2021). Nonlinear cointegrating power function regression with endogeneity. *Econometric Theory*, **37**(6), 1173–1213.

[5] Kallenberg, O. (2002). *Foundations of Modern Probability.* Second Edition. Springer-Verlag, Berlin.

[6] Shiryaev, A.N. (1996). *Probability.* Second Edition. Springer-Verlag, New York.

[7] Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Theory*, **30**(3), 509–535.