TIME VARYING PARAMETER REGRESSIONS WITH STATIONARY PERSISTENT DATA *

Zhishui Hu
University of Science and Technology of China

Ioannis Kasparis
University of Cyprus

Qiying Wang
The University of Sydney

April 19, 2024

Abstract

We consider local level and local linear estimators for estimation and inference in TVP regressions with general stationary covariates. The latter estimator also yields estimates for parameter derivatives that are utilised for the development of time invariance tests for the regression coefficients. Our theoretical framework is general enough to allow for a wide range of stationary regressors, including stationary long memory. We demonstrate that neglecting time variation in the regression parameters has a range of adverse effects in inference, in particular when regressors exhibit long range dependence. For instance, parametric tests diverge under the null hypothesis when the memory order is strictly positive. The finite sample performance of the methods developed is investigated with the aid of a simulation experiment. The proposed methods are employed for exploring the predictability of SP500 returns by realised variance. We find evidence of time variability in the intercept as well as episodic predictability when realised variance is utilised as a predictor in TVP specifications.

JEL Classification: C13, C22.

Key words and phrases: time varying parameter regression; kernel regression; long memory; predictive regression, IV estimation; episodic predictability; persistent data.

---

*The authors thank the co-editor and two referees for their very helpful comments on the original version and revision of this paper. Hu acknowledges research support from NSFC (No. 11671373); Wang acknowledges research support from the Australian Research Council. Address correspondence to Ioannis Kasparis, University of Cyprus, Nicosia, Cyprus, email: kasparis@ucy.ac.cy.

1 Introduction

Structural change is the subject of a vast literature in statistics, econometrics, empirical economics, and finance. Early work in this area has mainly focused on abrupt changes that are typically modelled in terms of structural breaks in regression parameters. Nevertheless, *smooth* time varying parameter (TVP) models have gained a lot of attention recently -see for example Robinson (1989, 1991), Dahlhaus (2000) for earlier work in this area, and for more recent developments Kristensen (2012), Giraitis et al. (2013), Giraitis et al. (2014), Phillips et al. (2017), Dahlhaus et al. (2019), Petrova (2019), Giraitis et al. (2021), Demetrescu et al. (2022) among others. Most studies consider TVP models in the context of *locally nonstationary* time series autoregressions e.g. Giraitis et al. (2014), Dahlhaus et al. (2019), Petrova (2019). Although these processes are nonstationary due to TVPs, boundness restrictions on the autoregressive parameters ensure that they behave asymptotically as stationary sequences. TVPs can be estimated by kernel methods -see e.g. Giraitis et al. (2014), Dahlhaus et al. (2019) and estimators have Gaussian limit distributions. The latter implies that conventional inference applies (e.g. limit distributions of various test statistics are either $N(0,1)$ or χ^2). The recent work of Giraitis et al. (2021) considers IV estimation in structural non-autoregressive TVP regressions with nonstationary covariates that satisfy mixing conditions. Similar to locally nonstationary autoregressions, the methods proposed in the aforementioned study yield conventional inference due to weak dependence assumptions.

Despite these developments, TVP models with strongly dependent data have attracted less attention. In a recent work, Phillips et al. (2017) consider structural TVP regressions with I(1) processes. The aforementioned study establishes that kernel methods yield consistent estimation of regression parameters in the nonstationary case, but inference is non conventional. In the presence of unit roots, the limit distribution of kernel estimators is comparable to that of OLS for fixed parameter models with I(1) regressors i.e. limit distributions are determined by stochastic integrals. To overcome this problem, these authors consider FMLS-type (see e.g. Phillips (1995)) of kernel estimators that enjoy mixed Gaussian distributions and as a result yield conventional tests. In another recent work, Demetrescu et al. (2022) develop inferential methods for the predictability hypothesis in predictive regressions that allow for smooth time varying slope parameters under the alternative hypothesis (non predictability), and predictors that can be stationary or nearly integrated processes. The methods considered in the aforementioned work do not involve estimation of TVP parameters. Instead, these authors consider sup t-statistics based on parametric estimators. In particular, regression parameters are estimated by a combination of IVX instruments (see e.g. Magdalinos and Phillips, 2009; Kostakis et al. 2015; Yang et al. (2020)) and time trend variables (see Breitung and Demetrescu, 2015). Due to IVX instrumentation, limit distributions are nuisance parameter free, irrespective of the stationarity properties of the data. Therefore, despite the fact limit distributions are non conventional[1], simulation and bootstrap methods can be used for obtaining p-values.

In this work we consider deterministic TVP predictive and structural regressions with *general stationary* covariates. We derive the limit properties of *local level* (LLev hereafter) and *local linear* (LLin hereafter) nonparametric estimators, and related test statistics, under high level conditions that involve stationary, ergodicity and existence of moments, thereby avoiding mixing requirements that rule out stationary long memory (cf. Kolmogorov and Rozanov, 1960). In particular, the basic paradigm under consideration entails specifications of the form

$$y_k = \mu(k/n) + \sum_{j=1}^{p-1} \beta_j\,(k/n)\,x_{k-1,j} + \sigma_k u_k,\ \ k = 1, ..., n,$$

where the regressions error is a conditionally heteroscedastic martingale difference sequence (e.g. GARCH(p,q), ARCH(∞)) with respect to certain filtration, and $x_{k-1,j}$ strictly stationary and predetermined with respect to the regressions error. Our assumptions allow for a general class of stationary processes (see eq. (13) for more details), e.g.

$$x_{k,j} = f_j(w_{k,j}),\ \ w_{k,j} = \sum_{i=0}^{\infty} \phi_{i,j}\xi_{k-i,j} \text{ with } \xi_{k,j} \sim iid(0, \sigma^2_{\xi,j}) \text{ and } \sum_{i=0}^{\infty} \phi^2_{i,j} < \infty. \tag{1}$$

It can be readily seen that (1) allows for models nonlinear in variables i.e. regression functions of known form f_j when $w_{k,j}$ is observable (see e.g. Park and Phillips, 1999). Further, the square summability of the coefficients of the MA(∞) process encompasses stationary long memory.

Although our focus is on predictive regressions, the proposed methods can be easily extended to structural regressions, where the regressors and the regression error are contemporaneously generated, with the aid of conventional instruments (e.g. Giraitis et al. 2021).[2] Our framework provides a generalisation to Robinson (1989) who considers TVP regressions with stationary mixing covariates, and a partial generalisation to Kristensen (2012) and Giraitis et al. (2021) who consider nonstationary covariates that satisfy mixing assumptions. Further, the results are complementary to Phillips et al. (2017) who focus on a different part of the regression space i.e. I(1) models.

[1]Limit distributions are determined by sup functionals of Gaussian processes

[2]Limit theory for structural TVP regressions is provided in Section 5, Theorems 6 and 7 of this work.

We demonstrate (Section 2) that neglecting time variation in regression parameters has severe adverse effects on inference. It can be easily seen that neglecting time variation leads to inconsistent estimates. It is less obvious however that in the presence of TVPs, parametric test statistics are divergent under the null hypothesis, when there are covariates of long memory. Size distortions can be very severe with deteriorating test performance in larger sample sizes. For example, if there is time variation in the regression intercept and predictors are of long memory, parametric t-tests for the predictability hypothesis[3] diverge in probability, under the null, as the sample size tends to infinity.[4] Moreover, neglecting time variation in general undermines the power of tests. Therefore, it is important to account for time variation in parameters, particularly when regressors are persistent processes.

This work considers models with stationary covariates. Some preliminary results suggest that the proposed methods also provide valid inference when data are weakly nonstationary (e.g. fractional $d = 1/2$ or mildly integrated processes -see Phillips and Magdalinos (2007) and Duffy and Kasparis (2021). It should be emphasised however, that the methods under consideration do not yield pivotal tests under nonstationarity in general e.g. for covariates that are I(d), $d > 1/2$. As noted above, in this case limit distributions are determined by stochastic integrals as per Phillips et al. (2017), and different methods will be required for obtaining pivotal tests.

Existing methods in this area only consider LLev estimators. Another contribution of the current work is the study of LLin methods. LLin estimators are known to result in reduced bias and also provide derivative estimators for the TVPs. Moreover, these estimators can be easily used to construct nonparametric t-tests to test for time variation in the slope parameters. Nonparametric t-statistics based on LLev and LLin estimation are considered with an emphasis on the predictability hypothesis in the context of predictive regressions. Kristensen (2012) also provides a test for structural change for TVP models with smooth regression parameters. The test proposed in the aforementioned work is based on an F-statistic that compares the RSS of a fully non-parametric TVP fit to those of a partly nonparametric fit. We expect that the F-statistic attains faster divergence rates than the nonparametric derivative test proposed in this work.[5] However, the implementation of the proposed derivative test is very easy i.e. the test statistic is simply a studentised LLin estimator. Furthermore, the derivative test statistic can be calculated for each regression point yielding a series of rolling test statistics that can distinguish

[3]i.e. $H_0 : \beta_j = 0$ for some $j = 1, ..., p - 1$.

[4]We demonstrate this size distortion phenomenon under stationary long memory, however standard asymptotic arguments together with Hu, Kasparis and Wang (2021; Theorem 3), suggest that this is also true under nonstationary long memory and nearly integrated arrays.

[5]It is well known that nonparametric derivative estimators attain slower convergence rates. This in turn affects the asymptotic power rates for related tests.

between periods of parameter stability and parameter instability.

The proposed methods are relevant to TVP predictive regressions with stationary predictors. Predictive regressions is an important area of research in econometrics and empirical finance. Many studies in this area focus on predictive regressions with nonstationary persistent data -for recent developments in this area see for example Kostakis et al. (2015), Yang et al. (2020), Demetrescu et al. (2022) and the references therein. Nevertheless, there is evidence that certain predictors such as realised volatility and inflation are stationary long memory or very close to the stationarity boundary i.e. of memory parameter $d \approx 0.5$.[6] Amihud and Hurvich (2004), Christensen and Nielsen (2006), Ang and Bekaert (2007), Bollerslev et al. (2009), Chen and Deo (2009), Bollerslev et al. (2013), Bandi et al. (2019) among others, develop methods for predictive regressions with stationary predictors.

The remainder of this work is organised as follows. Section 2 provides some theoretical results for the performance of parametric inference in the presence of TVPs. Section 3 develops basic limit theory for kernel functionals of stationary processes. This limit theory is utilised in Sections 4 and 5 for the development of estimation and inferential procedures for predictive and structural regressions respectively. Finally, an empirical application to the predictability of stock returns is the subject of Section 6. Proofs of all the results in the paper and additional derivations for Section 2 are provided in the Online Supplement (Hu et al., 2024). An extensive simulation study appears in the Online Supplement as well.

Throughout this paper, we make use of the following notation. For two deterministic sequences a_n and b_n, $a_n \sim b_n$ denotes $\lim_{n\to\infty} a_n/b_n = 1$. $1\{A\}$ is the indicator function on set A. For a vector or a matrix A, A' denotes its transpose. The norm of a vector x is $\|x\| = (x'x)^{1/2}$. Further, for an $l \times m$-dimensional matrix $A = [a_{ij}]$, $\|A\| = \sum_{i=1}^{l}\sum_{j=1}^{m}|a_{ij}|$. By $[x]$ we denote the integer part of a positive number x. As usual, $\otimes$ denotes the Kronecker product and $\mathbf{N}(\mathbf{0}, \sum)$ the multivariate normal variate with mean $\mathbf{0}$ and covariance matrix $\sum$. For an integrable function K, $\int K$ stands for $\int_{\mathbb{R}} K(x)dx$, unless otherwise specified. Finally, $\text{diag}\{a_1, ..., a_p\}$ denotes a $p \times p$ diagonal matrix with elements $\{a_1, ..., a_p\}$ on the main diagonal.

2 Consequences of Neglecting Time Variation in Regression Parameters

In this section, we provide some theoretical results on the consequences of neglecting time variation in regression parameters of predictive models, using univariate specifications with a long

[6]See Section 6 for more details.

memory covariate. Our objective is to highlight the consequences of this type of misspecification on inference. For this reason our presentation is somewhat informal. Precise derivations are stated in Section B (e.g., Lemma 2 and Lemma 3) of the Online Supplement. Simulations that support these theoretical findings are reported in Section C of the Online Supplement.

Neglecting time variation in the parameters has consequences to both estimation and testing, even if time variation is present only in some nuisance regression parameter -i.e. not a focus of empirical interest such as the intercept or the slope coefficient of another covariate. In general neglecting time variation in the parameter of interest leads to inconsistent estimates, and undermines the power of tests. Surprisingly, neglecting time variation in a nuisance parameter may have even more severe consequences. It can be shown that the latter type of misspecification not only results in size distortions, but also renders t-statistics divergent under the null hypothesis when the predictor has memory parameter strictly greater than zero i.e. $d > 0$. We demonstrate the above for OLS based inference for regressions with stationary predictors, but we expect that similar phenomena also apply to other methods (e.g. IV), and in models with nonstationary long memory -see footnote 4 above.

Consequences on Power. We start with the consequences of neglecting time variation in the parameter of interest under the alternative hypothesis by considering the following simple linear regression

$$y_k = \beta(k/n)x_{k-1} + u_k, \tag{2}$$

where $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ is a conditionally homoscedastic martingale difference sequence. In particular, we will use the following set of technical conditions together with (2).

Assumption P (power):

(a) y_k *is generated by (2);*

(b) $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ *is a martingale sequence such that*

  (i) *for all* $k \geq 1$, $E\left(u_k^2 \mid \mathcal{F}_{k-1}\right) = \sigma_u^2 < \infty$ *a.s.*

  (ii) u_k^2 *is uniformly integrable.*

(c) $\beta : [0,1] \to \mathbb{R}$ *is Riemann integrable on* $[0,1]$*;*

(d) x_k *is* $\mathcal{F}_k$*-measurable, strictly stationary and ergodic with* $Ex_1^2 < \infty$*.*

It is shown in Lemma 2 of the Online Supplement (See Section B there) that, under **Assumption P**, the OLS estimator $\tilde{\beta}_{OLS}$ from regressing y_k on x_{k-1} satisfies

$$\tilde{\beta}_{OLS} = \frac{\sum_{k=1}^{n} y_k \, x_{k-1}}{\sum_{k=1}^{n} x_{k-1}^2} \rightarrow_P \int_0^1 \beta(\tau) d\tau. \tag{3}$$

The OLS estimator converges to the pseudo-true value $\int_0^1 \beta(\tau)d\tau$ which is a chronological average of the TVP. As a result, OLS based t-tests for the parameter significance hypothesis/ predictability hypothesis (i.e. $H_0 : \beta = 0$) are likely to have poor power in situations where predictability is episodic. Indeed, under the alternative hypothesis (predictability), for the OLS based t-statistic $\tilde{t}_{OLS}$, we have

$$n^{-1/2}\tilde{t}_{OLS} = \frac{\tilde{\beta}_{OLS}}{\sqrt{\hat{\sigma}_u^2 \left[n^{-1} \sum_{k=1}^{n} x_{k-1}^2\right]^{-1}}} \rightarrow_P \frac{\int_0^1 \beta(\tau) d\tau}{\sqrt{\sigma_*^2 \left[E x_1^2\right]^{-1}}}, \tag{4}$$

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 a pseudo-true value that is given by (see Lemma 2 of the Online Supplement)

$$\sigma_*^2 = \left[\int_0^1 \beta^2(\tau) d\tau - \left(\int_0^1 \beta(\tau) d\tau\right)^2\right] E x_1^2 + \sigma_u^2.$$

In the context of predictive regressions, the pseudo-true value $\int_0^1 \beta(\tau)d\tau$ will tend to be small as episodic predictability events are averaged out over time. Further, it is possible that positive predictability events (i.e. $\beta(.) > 0$) are cancelled out by negative ones (i.e. $\beta(\cdot) < 0$) -see Figure 15 given in Section C of the Online Supplement for simulation results that illustrate these adverse power effects.

Consequences on Size. We next illustrate the effects of neglecting time variation in the intercept under the null hypothesis, when the parameter of interest is the slope coefficient in the following model

$$y_k = \mu(k/n) + \beta x_{k-1} + u_k, \tag{5}$$

where u_k is given as in model (2) and x_k a linear process, possibly of long memory. The following assumption introduces a set of technical conditions that define (5) precisely.

Assumption S (size):

(a) *y_k is generated by (5);*

(b) *Condition (b) of **Assumption P** holds;*

(c) $\mu : [0,1] \to \mathbb{R}$ *is of bounded variation on* $[0,1]$*;*

(d) x_k *is an* $\mathcal{F}_k$*-measurable linear process of the form* $x_k = \sum_{i=0}^{\infty} \phi_i \xi_{k-i}$ *with* $\xi_k \sim iid(0, \sigma_\xi^2)$*, and either*

(i) $\phi_i \sim c_0 i^{d-1}$*, with* $0 < d < 1/2$*, or*

(ii) $0 < \sum_{i=0}^{\infty} |\phi_i| < \infty$.

Under **Assumption S (d.i)**, x_k is a type I fractional long memory process (**LM**), while under **Assumption S (d.ii)**, x_k is short memory (**SM**) -see, e.g. Phillips and Shimotsu (2004). We note that **Assumption S** entails a specific parametric model for the covariate (i.e. a linear process). The assumptions on x_k and the regression error will be relaxed in the subsequent sections.

Set

$$\delta_n^2 := \mathrm{Var}\Big(\sum_{k=1}^{n} x_k\Big) \sim \begin{cases} c_0^2 \, \sigma_\xi^2 \, c(d) \cdot n^{1+2d}, & \text{under } \mathbf{LM}, \\ \left(\sum_{i=0}^{\infty} \phi_i\right)^2 \sigma_\xi^2 n, & \text{under } \mathbf{SM} \end{cases}$$

(see e.g. Wang(2015), Example 2.12), where

$$c(d) = \frac{1}{d(1+2d)} \int_0^\infty (x(1+x))^{d-1} dx.$$

It is known that $\delta_n^{-1} \sum_{k=1}^{n} x_k \to_d \mathbf{N}(0,1)$ (see e.g. Ibragimov and Linnik (1971), Theorem 18.6.5 or Wang, 2015). Suppose that we are interested in testing $H_0 : \beta = \beta_0 \in \mathbb{R}$, using OLS based inference. It can be shown that, under **Assumption S (a)-(d.i)**, the OLS estimator

$$\sqrt{n}\left(\tilde{\mu}_{OLS} - \int_0^1 \mu(\tau) d\tau\right) \to_d \mathbf{N}(0, \sigma_u^2), \tag{6}$$

and

$$\frac{n}{\delta_n}\left(\tilde{\beta}_{OLS} - \beta\right) \to_d \left(E x_1^2\right)^{-1} \left[1, \; -\int_0^1 \mu(\tau) d\tau\right] \cdot \mathbf{N}\left(\mathbf{0}, \Psi\right), \tag{7}$$

where Ψ is the matrix

$$\Psi = \frac{1}{c(d)} \int_{-\infty}^{1} \begin{bmatrix} \left\{\int_{r\vee 0}^{1} \mu(s) (s-r)^{d-1} ds\right\}^2 & \int_{r\vee 0}^{1} \mu(s) (s-r)^{d-1} ds \cdot \int_{r\vee 0}^{1} (s-r)^{d-1} ds \\ \int_{r\vee 0}^{1} \mu(s) (s-r)^{d-1} ds \cdot \int_{r\vee 0}^{1} (s-r)^{d-1} ds & \left\{\int_{r\vee 0}^{1} (s-r)^{d-1} ds\right\}^2 \end{bmatrix} dr, \tag{8}$$

with $0 < d < 1/2$. See Lemma 3 in the Online Supplement for more details. An n/δ_n-convergence result also holds under **SM** with a different Ψ variance matrix[7]. It follows from the above that the OLS estimator for β is consistent but there is a reduction in the convergence rate when the predictor is a stationary fractional process with memory parameter strictly greater than zero. This reduction in the convergence rate does not affect asymptotic power rates[8] but nonetheless results in severe size distortions under the null hypothesis. To see this note first that the regression error variance estimator is

$$\frac{1}{n}\sum_{k=1}^{n}\tilde{u}_k^2 \to_P \sigma_u^2 + \int_0^1 \mu(\tau)^2 d\tau - \left(\int_0^1 \mu(\tau)d\tau\right)^2, \tag{9}$$

where $\tilde{u}_k$ denotes the OLS residuals. Combining (6)-(9) it follows that under the null hypothesis that for $d > 0$,

$$\left|\tilde{t}_{OLS}\right| \to_P \infty. \tag{10}$$

More details for the validity of (6)-(10) can be found in Lemma 3 of the Online Supplement. Figure 16 given in the Online Supplement provides simulation results that highlight these effects. The actual divergence rate of the t-statistic in (10), under the null hypothesis, is determined by the sequence

$$\delta_n/n^{1/2} \sim \begin{cases} C_1 n^d, & \text{under } \mathbf{LM}; \\ C_2, & \text{under } \mathbf{SM}, \end{cases}$$

where $0 < C_1, C_2 < \infty$. Clearly, when x_k is a short memory process, the test statistic is bounded under the null but fails to have a standard normal distribution[9]. Therefore, OLS based t-tests exhibit size distortions even in this case.

3 Asymptotics for Kernel Functionals

This section develops basic limit theory for functionals of stationary processes weighted by kernels of time trend variables. Let $\{\mathbf{x}_k\}_{k\geq 0}$ be an $\mathbb{R}^p$ time series process. Further, let K be an integrable function and $m \in \{0, 1, 2\}$, and set

$$S_n := \quad \frac{c_n}{n}\sum_{k=1}^{n}\mathbf{x}_{k-1}\mathbf{x}_{k-1}'\sigma_k^m K\left[c_n(k/n-\tau)\right],$$

[7] We do not provide any derivations for the **SM** case, but the proof of Lemma 3 in the Online Supplement can be easily extended to the case that x_k is iid (and therefore a short memory) random sequence.

[8] It follows from (6)-(9) that $\tilde{t}_{OLS}$, attains a $\sqrt{n}$-divergence rate under H_1.

[9] In the **SM** case the t-statistic is convergent with a normal limit distribution. Notice however that $\frac{1}{n}\sum_{k=1}^{n}\tilde{u}_k^2$ is inconsistent, therefore limit variance does not equal one.

$$M_n := \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \mathbf{x}_{k-1} K\left[c_n(k/n-\tau)\right] \sigma_k u_k,$$

where c_n is a diverging bandwidth sequence of positive constants, and u_k (scalar) together with an appropriate filtration $\{\mathcal{F}_k\}$ forms a martingale difference sequence so that $\mathbf{x}_k$ is $\mathcal{F}_k$-measurable and σ_k is (scalar) $\mathcal{F}_{k-1}$ -measurable. Note that we can alternatively formulate the kernel functionals in terms of vanishing bandwidth terms i.e. $c_n = h_n^{-1}$ with $h_n \to 0$. The asymptotics of $\{S_n, M_n\}$ are utilised in Sections 4 and 5 for the asymptotic analysis of the kernel estimators. The vector $\mathbf{x}_k$ and σ_k are set to be strictly stationary. In view of this, $\sigma_k u_k$ exhibits conditional heteroscedasticity and can be a GARCH(p,q) or an ARCH(∞) process. To facilitate our basic limit results, we make use of the following conditions.

A1 (innovations): $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ forms a martingale difference satisfying the following conditions:

(a) $\sup_{k\geq 1} E(u_k^2 I(|u_k| \geq M)|\mathcal{F}_{k-1}) = o_P(1)$, as $M \to \infty$;

(b) for all $k \geq 1$, $E(u_k^2|\mathcal{F}_{k-1}) = 1\ a.s.$

A2 (stationary process): $(\mathbf{x}_{k-1}, \sigma_k)_{k\geq 1}$ is a sequence of ergodic (strictly) stationary random vectors with $E\left\{\|\mathbf{x}_0\|^2 + \sigma_1^2\left(1 + \|\mathbf{x}_0\|^2\right)\right\} < \infty$ so that $\mathbf{x}_k$ is $\mathcal{F}_k$-measurable and σ_k is $\mathcal{F}_{k-1}$ -measurable, where $\mathcal{F}_k$ is defined as in **A1**.

A3 $K(x)$ is a positive, locally Riemann integrable and eventually monotonic function[10] with $0 < \int K < \infty$.

We remark that the innovation process $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ in **A1** is standard in the literature. Due to Assumption **A1**, M_n has a martingale structure (see e.g. Park and Phillips (1999, 2000, 2001), Wang (2014), Wang and Phillips (2009), Duffy and Kasparis (2021)). The uniform integrability condition (a) is weak in comparison with the high moments used in previous works. In **A1**(b), we impose $E(u_k^2|\mathcal{F}_{k-1}) = 1, a.s.$ for convenience of notation. In fact, if $\sigma_u^2 := E(u_k^2|\mathcal{F}_{k-1}) \neq 1$, it is routine to see that our results still hold when σ_k is replaced by $\sigma_k\, \sigma_u$. Ergodicity and strict stationarity (cf. Assumption **A2**), together with an existence of moments requirement postulate that the underlying processes (i.e. $\mathbf{x}_k$ and σ_k) satisfy a law of large numbers (see e.g. Shiryaev (1996), Theorem 3, p. 413). A wide range of time series models relevant to econometrics satisfy **A2**. For instance, under **A2** $\mathbf{x}_k$ can be a short/long memory linear process[11]. Furthermore,

[10]i.e., $K(x)$ is Riemann integrable in any finite interval and there exists an $A_1 > 0$ such that $K(x)$ is monotonic on $(-\infty, -A_1)$ and (A_1, ∞).

[11]e.g. $x_k = \sum_{i=0}^{\infty} \phi_i \xi_{k-i}$, $\xi_i \sim iid(0, \sigma_\xi^2)$, $\sum_{i=0}^{\infty} \phi_i^2 < \infty$.

under **A1** and **A2**, $\sigma_k u_k$ is allowed to be a strictly stationary GARCH or ARCH(∞) model (e.g. Francq and Zakonian, 2010; Section 2.2). More details about the time series models that can be accommodated by **A1**-**A3** are provided in the next Section. Finally, the monotonicity condition of **A3** is a technical requirement that is commonly satisfied in applications.

We now present the limit theory for the sample functionals S_n and M_n .

Theorem 1. *Suppose **A2** and **A3** hold. For each $\tau \in (0,1)$, we have*

$$S_n \to_P E[\sigma_1^m \mathbf{x}_0 \mathbf{x}_0'] \int K, \tag{11}$$

*as $c_n \to \infty$ and $c_n/n \to 0$. If in addition **A1** holds, then*

$$M_n \to_d \mathbf{N}\left(\mathbf{0},\ E\left[\sigma_1^2 \mathbf{x}_0 \mathbf{x}_0'\right] \int K^2\right). \tag{12}$$

Remark 1. (a) If $\{\mathbf{x}_k\}_{k\geq 0}$ is a scalar weakly nonstationary process (i.e., $I(1/2)$) and mildly integrated process, where FCLTs do not apply -see e.g. Phillips and Magdalinos (2007), Duffy and Kasparis (2021)- we conjecture that

$$\frac{c_n}{n}\sum_{k=1}^{n}\left(d_n^{-1}\mathbf{x}_{k-1}\right)^2 \sigma_k^2 K\left[c_n(k/n-\tau)\right] \to_d E(\sigma_1^2)\int_{\mathbb{R}}(x+X^-)^2 \varphi_{\sigma_+^2}(x)dx \int K,$$

where $d_n \to \infty$, $\varphi_{\sigma_+^2}(x)$ is the density of an $N\left(0,\sigma_+^2\right)$ variate ($\sigma_+^2 > 0$) and $X^- \sim N\left(0,\sigma_-^2\right)$ ($\sigma_-^2 \geq 0$). Analysis of this generalization is left for future work.

(b) The limit result of Theorem 1 also holds true for the boundary values $\tau = 0, 1$ with $\int K$ and $\int K^2$ replaced by one sided integrals e.g. (11) holds for $\tau = 0$ and $\tau = 1$ with $\int_0^\infty K(x)dx$ and $\int_{-\infty}^0 K(x)dx$ respectively. Following Phillips et al. (2017), we present results only for $\tau \in (0,1)$.

4 Estimation and Inference in TVP Models

In this section, we present estimation and inferential methods for models with TVP covariates by using the theoretical results of Section 3. We start our analysis with predictive (i.e. reduced form) models. Extensions to structural regressions are provided in the next Section. Consider the TVP predictive regression

$$y_k = \mu(k/n) + \sum_{j=1}^{p-1} \beta_j\left(k/n\right) x_{k-1,j} + e_k,\ e_k = \sigma_k u_k,\ k = 1, ..., n, \tag{13}$$

where $\mu, \beta_j : [0,1] \to \mathbb{R}$, and the predictor $x_{k,j}$ is a strictly stationary ergodic process. Define the vector

$$\mathbf{x}_k' = [1, x_{k,1}, ..., x_{k,p-1}].$$

In particular, we assume that u_k, $\mathbf{x}_k$ and σ_k satisfy assumptions **A1** and **A2**. It can be readily seen that under these assumptions, (13) is a reduced form regression where the predictors are pre-determined with respect to the regression errors. This assumption entails weak endogeneity. Generalizations to the case where covariates and e_k are contemporaneously determined (i.e. to strong endogeneity) will be provided in Section 5. Under the current assumptions the regressions error term e_k can be a stationary conditionally heteroscedatstic process e.g. ARCH(∞).

A similar specification has been considered by Robinson (1989) with covariates being stationary strong mixing. Further, Kristensen (2012) and Giraitis et al. (2021) assume nonstationary mixing covariates in the context of structural TVP models. Our moment requirements on the processes are quite weak. For deriving limit distribution theory for the LLev and LLin estimator we require existence of two moments for $\mathbf{x}_k$ and σ_k. This requirement will be strengthened to four moments for the obtaining limit theory for nonparametric t-tests and F-tests -see also Remark 7 below. In comparison Giraitis et al. (2021) assume more than eight moments for the model covariates. As remarked earlier, our assumptions are general enough to allow for nonlinear in variables regression models with regressors being for example strongly dependent or heavy tailed weakly dependent processes. For example, suppose that $x_{k,j} = f_j(w_{k,j})$, $j = 1, ..., p-1$, with

$$\mathbf{w}_k := (w_{k1}, ..., w_{k,p-1})' = \sum_{i=0}^{\infty} \Phi_i \xi_{k-i}, \ \Phi_i = diag\{\phi_{i,1}, ..., \phi_{i,p-1}\},$$

and $(\xi_{k,1}, ..., \xi_{k,p-1}, \sigma_{k+1}) := (\xi_k', \sigma_{k+1}) \sim iid$. Further, suppose that either **FR** or **HT** conditions below hold:

FR: $\xi_k \sim iid(\mathbf{0}, \Sigma)$ and $\sum_{i=0}^{\infty} \phi_{i,j}^2 < \infty$;

HT: (a) ξ_k follows a multivariate α-stable distribution with $\alpha \in (0,2)$, $\sum_{i=0}^{\infty} |\phi_{i,j}|^{\alpha'} < \infty$ for some $\alpha' \in (0, \alpha)$;

(b) $E|f_j(w_{1,j})|^q < \infty, \ q \geq 2$.

Note that under conditions **FR**, $w_{k,j}$ can be a fractional process of memory order $|d| < 1/2$. On the other hand, condition **HT**(a) entails that $w_{k,j}$ is a heavy tailed linear process that possess a finite α' moment.[12] In the latter case, $w_{k,j}$ may not have a finite mean or variance nevertheless

[12]Under **HT**, $\mathbf{w}_k$ exists *a.s.* and in L_p-sense with $p = \alpha'$, and is strictly stationary. To see this, first note

due to **HT**(b), the transformed process $f_j(w_{k,j})$ has a qth moment. In general, if f_j satisfies the reduced growth requirement $|f_j(x)| \leq C(1+|x|^{\eta})$, $C \in (0,\infty)$, $\eta \in (0,1)$, **HT** holds for all $\alpha \in (\eta q, 2)$. Further, $f_j(x) = \ln(x)_+$ satisfies the aforementioned requirement for all values of the tail parameter α in $(0,2)$.[13] Logarithmic transformations and reduced polynomial growth regression functions have been used in a number of studies in the predictability of stock returns -see e.g. Lewellen, 2004; Marmer, 2008; Bollerslev et al., 2013; Anderson and Varneskov, 2021. To some extent, our methods are complementary to those of Phillips et al. (2017) who focus on a different area of the regressor space. The regressor space under consideration is comparable to that of Christensen and Nielsen (2006), Bollerslev et al. (2013), Bandi et al. (2019) among others, who consider predictive regressions with stationary fractional predictors.

We first consider an LLev kernel regression estimator (cf. Li and Racine, 2006; Section 2.1) -see also Robinson (1989), Phillips et al. (2017), Giraitis et al. (2021) among others. Set

$$\theta(\tau)' := [\mu(\tau), \beta_1(\tau), ..., \beta_{p-1}(\tau)] .$$

We can write (13) as

$$y_k = \theta(k/n)'\mathbf{x}_{k-1} + e_k. \tag{14}$$

The LLev is obtained from minimisation of the following objective function:

$$\hat{\theta}(\tau) := \arg\min_{\mathbf{a}\in\mathbb{R}^p} \sum_{k=1}^{n} \left(y_k - \mathbf{a}'\mathbf{x}_{k-1}\right)^2 K\left[c_n\left(k/n-\tau\right)\right], \tag{15}$$

where c_n is a diverging bandwidth sequence of positive constants. Define

$$Q \ = \ E(\mathbf{x}_0\mathbf{x}_0') \text{ and } \mathbf{\Omega} \ = E\left(\sigma_1^2\mathbf{x}_0\mathbf{x}_0'\right).$$

that Samorodnitsky and Taqqu (1994, Theorem 2.3.1) implies that each component series $\xi_{k,j}$, $j = 1,...,p-1$, follows univariate alpha stable distribution with the same stability parameter and hence $E\left|\xi_{k,j}\right|^{\alpha'} < \infty$. Now we show that each component series in $\mathbf{w}_k$ exists in L_p-sense and $a.s.$ For the former it suffices showing that $\sum_{j=0}^{n} \phi_{i,j}\xi_{k-i,j}$ is a Cauchy sequence (due to the completeness of $L_p(\Omega, \mathcal{F}, P)$; e.g. Brockwell and Davis, 1991; p. 68-69). Suppressing the occurrence of the index j we have as $n, m \to \infty$

$$E\left|\sum_{i=m}^{n}\phi_i\xi_{k-i}\right|^{\alpha'} \leq_{(1)} 2\sum_{i=m}^{\infty}|\phi_i|^{\alpha'}\sup_k E\left|\xi_k\right|^{\alpha'} \to 0,$$

where $\leq_{(1)}$ follows from the Loève inequality for $\alpha' \in (0,1]$ and from the Bahr-Esseen inequality otherwise. Similarly, for $a.s.$ convergence note that

$$E\left|\sum_{i=0}^{\infty}\phi_i\xi_{k-i}\right|^{\alpha'} \leq \lim_{n\to\infty} E\sum_{i=0}^{n}|\phi_i|^{\alpha'}\left|\xi_{k-i}\right|^{\alpha'} < \infty.$$

Strict stationarity follows easily from the fact that ξ_k is i.i.d.

[13] $\ln(x)_+ := \max(\ln(x), 0)$.

The following theorem gives the limit distribution of the LLev estimator.

Theorem 2. *Suppose that:*

(*a*) $\{y_k\}_{k\in\mathbb{N}}$ *is generated by (13);*

(*b*) ***A1*** *and* ***A2*** *hold;*

(*c*) $\theta(.)$ *is Hölder continuous on* $[0,1]$ *of order* $\gamma \in (0,1]$*, i.e.*

$\|\theta(x)-\theta(y)\| \leq C\,\|x-y\|^{\gamma}$ *for* $x, y \in [0,1]$ *and some* $C \in (0,\infty)$*;*

(*d*) K *satisfies* ***A3*** *and* $\int x^{\gamma}K < \infty$*;*

(*e*) $c_n/n + n/c_n^{1+2\gamma} \to 0$*;*

(*f*) Q *is a positive definite matrix, i.e,* Q^{-1} *exists.*

Then, for each fixed $\tau \in (0,1)$*,*

$$\sqrt{\frac{n}{c_n}}\left(\hat{\theta}(\tau)-\theta(\tau)\right) \to_d \mathbf{N}\left(\mathbf{0}, Q_1^{-1}\mathbf{\Omega}_1 Q_1^{-1}\right), \tag{16}$$

where $Q_1 = Q\int K$ *and* $\mathbf{\Omega}_1 = \mathbf{\Omega}\int K^2$*.*

We next introduce an LLin estimator for $\theta(\tau)$. Write the vector of derivatives $\theta^{(1)}(\tau) := \partial\theta(\tau)/\partial\tau$. The LLin estimator is defined by

$$\begin{bmatrix} \widetilde{\theta}(\tau) \\ \widetilde{\theta}^{(1)}(\tau) \end{bmatrix} := \underset{(\mathbf{a}',\mathbf{b}')'\in\mathbb{R}^{2p}}{\arg\min} \sum_{k=1}^{n} \left(y_k - \mathbf{a}'\mathbf{x}_{k-1} - \mathbf{b}'\widetilde{\mathbf{x}}_{k-1}\right)^2 K\left[c_n\left(k/n-\tau\right)\right], \tag{17}$$

where $\widetilde{\mathbf{x}}_{k-1} = (k/n-\tau)\,\mathbf{x}_{k-1}$, and $\widetilde{\theta}(\tau)$ and $\widetilde{\theta}^{(1)}(\tau)$ are estimators for $\theta(\tau)$ and $\theta^{(1)}(\tau)$, respectively. LLin estimators in general exhibit reduced asymptotic bias, relative to LLev estimators (cf. Li and Racine, 2006; Section 2.5), and also provide estimates for the derivatives of the regression function when the latter exist. For the analysis of the LLin estimator we need to introduce some additional notation. Define

$$\mathcal{K}_1 = \begin{bmatrix} \int K & \int xK \\ \int xK & \int x^2K \end{bmatrix} \quad \text{and} \quad \mathcal{K}_2 = \begin{bmatrix} \int K^2 & \int xK^2 \\ \int xK^2 & \int x^2K^2 \end{bmatrix}.$$

The limit properties of the LLin estimator is given below.

Theorem 3. *Suppose that:*

(*a*) $\{y_k\}_{k\in\mathbb{N}}$ *is generated by (13);*

(*b*) ***A1*** *and* ***A2*** *hold;*

(*c*) $\theta(.)$ *has a uniformly bounded second derivative on* $[0,1]$*;*

(*d*) *in addition to* ***A3****,* K *satisfies* $\int x^3 K < \infty$*;*

(*e*) $c_n/n + n/c_n^5 \to 0$*;*

(*f*) Q^{-1} *and* $\mathcal{K}_1^{-1}$ *exist.*

Then, for each fixed $\tau \in (0,1)$*,*

$$D_n\left(\begin{bmatrix}\widetilde{\theta}(\tau)\\ \widetilde{\theta}^{(1)}(\tau)\end{bmatrix} - \begin{bmatrix}\theta(\tau)\\ \theta^{(1)}(\tau)\end{bmatrix}\right) \to_d \mathbf{N}\left(\mathbf{0}, Q_2^{-1}\mathbf{\Omega}_2 Q_2^{-1}\right), \tag{18}$$

where $D_n = diag\left\{\sqrt{\frac{n}{c_n}}, \sqrt{\frac{n}{c_n^3}}\right\} \otimes I_p$ *(*I_p *a p-dimensional identity matrix), and*

$$Q_2 = \mathcal{K}_1 \otimes Q \quad and \quad \mathbf{\Omega}_2 = \mathcal{K}_2 \otimes \mathbf{\Omega}.$$

Remark 2. The smoothness assumptions on the TVP $\theta(.)$ of Theorems 2 and 3 (i.e., (c) in both theorems) are standard. For the LLin estimator a more restrictive regularity condition is required so that it attains smaller 'asymptotic bias' in comparison with the LLev estimator. In general, kernel regression estimators entail nonlinearity induced asymptotic 'bias'. In the current framework, this type of bias is due to increments of the form

$$\sum_{k=1}^{n}\{\theta(\tau) - \theta(k/n)\}.$$

In particular, the l.h.s. of (16) entails an asymptotic bias term of order $O_P(\sqrt{n/c_n^{1+2\gamma}})$, whilst the corresponding bias term in (18) is $O_P(\sqrt{n/c_n^5})$, under the corresponding settings on the TVP $\theta(.)$. Note that both bias terms disappear in the given results due to the assumption (e) on c_n. An additional advantage of the LLin is that the derivatives of the TVPs can be estimated. These estimates can be readily utilised for testing hypotheses about parameter constancy with respect to time (see also Remark 9 below).

Remark 3. The condition in (e) of Theorem 2 is equivalent to $c_n/n \to 0$ and $c_n^{1+2\gamma}/n \to \infty$, where, as explained in Remark 2, the latter condition is used to remove the bias. There is a trade off between the selection of c_n and γ where the latter depends on (c) in that theorem. In practice, if γ is fixed, the optimal choice of c_n is that $c_n = n^{1/(1+2\gamma)}a_n$ where $a_n \to \infty$ so that (16) has a fast convergence rate. A similar explanation applies to (e) of Theorem 3.

Remark 4. The existence of Q^{-1} in (f) of Theorems 2 and 3 is natural when x_t is a stationary process. Phillips et al. (2017) show that, if x_k is an $I(1)$ process, the LLev estimator in Theorem 2 necessarily has a singular form with multiple rates of convergence arising from that singularity. This degeneracy is manifest for all fractional $d > 1/2$, as well as nearly integrated arrays. Indeed, under nonstationarity ($x_k \sim I(d)$, $d > 1/2$), it can be shown that the counterpart of the limit matrix Q_1 is of the form (see e.g. Hu, Kasparis and Wang, 2021; Theorem 3)

$$\begin{bmatrix} 1 & X_\tau' \\ X_\tau & X_\tau X_\tau' \end{bmatrix} \int K,$$

where X_τ is some limit Gaussian process (e.g. fractional Brownian Motion) at time τ. Notice that the matrix above is necessarily singular just as in Phillips et al. (2017).

Remark 5. Preliminary theoretical results suggest that Theorem 2 holds true when the predictor is a weakly non-stationary process (i.e. fractional $d = 1/2$ or MI). Suppose that $p = 2$. In particular, we conjecture that (16) holds with Q_1

$$Q_1 = \begin{bmatrix} 1 & X^- \\ X^- & \int_{\mathbb{R}} (x + X^-)^2 \varphi_{\sigma_+^2}(x)dx \end{bmatrix} \int K,$$

where $\varphi_{\sigma_+^2}(x)$ and X^- as in Remark 1. Note that Q_1 here is in general non singular. We therefore expect that the proposed methods are valid even if the data are weakly nonstationary.

Remark 6. It is worth noting that Theorems 2 and 3 demonstrate that the limit variances of the TVP estimators are independent of the regression point τ. This is in contrast to non-parametric density and regression estimators where limit variance does depend on location, and as a result there is a deterioration in estimation accuracy when functionals are estimated at regression points away from the origin. For TVP estimates however confidence intervals are not affected by the value of the chronological point τ even if the latter assumes boundary values. This theoretical result is also corroborated by our simulation study, that shows only minor oversizing close to boundary values in large sample sizes.

We next consider nonparametric t-tests, and F-tests based on the LLev and LLin estimators. Before presenting the test statistics under consideration, we introduce some notation. For a vector a let a_i, $i = 1, ..., p$, be its i^{th} element, and for a square matrix A, $[A]_{ii}$ denotes its i^{th} diagonal element. Further, $\theta_i(\tau)$ denotes i^{th} element of $\theta(\tau)$. For single restrictions we consider test hypotheses of the form

$$H_0 : \theta_i(\tau) = \eta(\tau), \tag{19}$$

and

$$H_0 : \theta_i^{(1)}(\tau) = \eta(\tau), \tag{20}$$

for some prespecified $\eta : (0, 1) \to \mathbb{R}$. In particular, (19) entails some hypothesis for $\mu(\tau)$ or $\beta_i(\tau)$, whilst (20) concerns the derivatives of the aforementioned parameters. The proposed tests utilise the estimators of (16) and (18). Set

$$\left\{\hat{\mathcal{Q}}_n, \hat{\mathbf{\Omega}}_n\right\} := \left\{\sum_{k=1}^{n} \mathbf{x}_{k-1}\mathbf{x}_{k-1}' K_{kn},\ \sum_{k=1}^{n} \hat{e}_k^2 \mathbf{x}_{k-1}\mathbf{x}_{k-1}' K_{kn}^2\right\}$$

with $\hat{e}_k := \hat{e}_k(\tau) := y_k - \hat{\theta}(\tau)'\mathbf{x}_{k-1}$, and recall that $K_{kn} = K\left[c_n\left(k/n - \tau\right)\right]$. Further, we let

$$\left\{\tilde{\mathcal{Q}}_n, \tilde{\mathbf{\Omega}}_n\right\} := \left\{\sum_{k=1}^{n} \widehat{\mathbf{x}}_{k-1}\widehat{\mathbf{x}}_{k-1}' K_{kn},\ \sum_{k=1}^{n} \tilde{e}_k^2\, \widehat{\mathbf{x}}_{k-1}\widehat{\mathbf{x}}_{k-1}' K_{kn}^2\right\},$$

where $\widehat{\mathbf{x}}_k = (\mathbf{x}_k', \widetilde{\mathbf{x}}_k')'$ and $\tilde{e}_k = y_k - \widetilde{\theta}(\tau)'\mathbf{x}_{k-1}$. The proposed test statistics are

$$\hat{t}_i(\tau) = \frac{\hat{\theta}_i(\tau) - \eta(\tau)}{\sqrt{\left[\hat{\mathcal{Q}}_n^{-1}\hat{\mathbf{\Omega}}_n\hat{\mathcal{Q}}_n^{-1}\right]_{ii}}}, \qquad \tilde{t}_i(\tau) = \frac{\tilde{\theta}_i(\tau) - \eta(\tau)}{\sqrt{\left[\tilde{\mathcal{Q}}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{\mathcal{Q}}_n^{-1}\right]_{ii}}}, \quad i = 1, ..., p,$$

for the null hypothesis (19), and

$$\tilde{t}_i^{(1)}(\tau) = \frac{\widetilde{\theta}_i^{(1)}(\tau) - \eta(\tau)}{\sqrt{\left[\tilde{\mathcal{Q}}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{\mathcal{Q}}_n^{-1}\right]_{jj}}}, \quad i = 1, ..., p,\ j = i + p,$$

for the null hypothesis (20).

Similarly, we can use nonparametric F-tests for testing multiple restrictions. In particular, in the context of LLev estimation we consider joint restrictions of the form

$$H_0 : R\theta(\tau) = \eta(\tau), \tag{21}$$

where R is a $q \times p$ $(q \leq p)$ matrix and $\eta(\tau)$ a predetermined function such that $\eta : (0, 1) \to \mathbb{R}^q$. Testing general restrictions based on the LLin estimator is technically challenging due to the fact that multiple convergence rates are attained in this case. Given than one of the advantages of LLin relative to the LLev is derivative estimation, we only consider restrictions on the parameter derivatives of the form

$$H_0 : R\theta^{(1)}(\tau) = \eta(\tau), \tag{22}$$

where R and $\eta(\tau)$ are as in (21) above. Consider the following nonparametric F-statistic

$$\hat{F}(\tau) = \left[R\hat{\theta}(\tau) - \eta(\tau)\right]' \left[R\hat{\mathcal{Q}}_n^{-1}\hat{\mathbf{\Omega}}_n\hat{\mathcal{Q}}_n^{-1}R'\right]^{-1} \left[R\hat{\theta}(\tau) - \eta(\tau)\right]$$

and

$$\tilde{F}(\tau) = \left[R\tilde{\theta}^{(1)}(\tau) - \eta(\tau)\right]' \left[[\mathbf{0}, R]\,\tilde{\mathcal{Q}}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{\mathcal{Q}}_n^{-1}\,[\mathbf{0}, R]'\right]^{-1} \left[R\tilde{\theta}^{(1)}(\tau) - \eta(\tau)\right],$$

with $\mathbf{0}$ being $q \times p$ matrix of zeros.

For the asymptotic analysis of the F-statistics we require an additional technical condition. In particular, to establish the consistency of $\hat{\mathbf{\Omega}}_n$ and $\tilde{\mathbf{\Omega}}_n$, Assumption **A4** below will be utilised.

A4 (a) $E\|\mathbf{x}_0\|^4 < \infty$;

(b) either (i) $\sup_{k\geq 1} Eu_k^4 < \infty$ or

(ii) $Y_k := \sigma_k^2[\alpha'\mathbf{x}_{k-1}]^2\left[u_k^2 - 1\right]$ is uniformly integrable for all $\alpha \in \mathbb{R}^p$.

Remark 7. (a) The additional moment condition **A4**(a) on the regressors is required for the estimation $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$, and can be relaxed to $E\|\mathbf{x}_1\|^2 < \infty$ when the regression error is homoscedastic. In the latter case we can employ the alternative variance estimator

$$\frac{1}{n}\sum_{k=1}^{n}\hat{e}_k^2\sum_{k=1}^{n}\mathbf{x}_{k-1}\mathbf{x}_{k-1}'$$

instead of $\hat{\mathbf{\Omega}}_n$ for the studentisation of LLev estimators.

(b) Condition **A4**(b) entails additional restrictions on the regression error u_k. In particular, u_k is required to have a fourth moment uniformly or Y_k being uniformly integrable. Note that a sufficient condition for the latter is $\sup_k E(|Y_k|^{1+\varepsilon}) < \infty$, for some $\varepsilon > 0$. Suppose for example that $\sup_k E(|u_k|^{2(1+\varepsilon)}|\mathcal{F}_{k-1}) \leq C < \infty$ *a.s.*, where C is a constant. Then $\sup_k E[\sigma_k^2\|\mathbf{x}_{k-1}\|^2]^{1+\varepsilon} < \infty$ is sufficient for uniform integrability.[14]

Theorem 4 next gives the limit properties of the LLev based tests.

[14]Note that there exists some constant $C_1 > 0$ such that

$$E(|Y_k|^{1+\varepsilon}) \leq C_1 E\left[[\sigma_k^2\left(\alpha'\mathbf{x}_{k-1}\right)^2]^{1+\varepsilon}[E(|u_k|^{2(1+\varepsilon)}|\mathcal{F}_{k-1}) + 1]\right].$$

Theorem 4. *Suppose that, in addition to the conditions of Theorem 2, $\int x^2K^2 < \infty$, and **A4** holds.*

(i) Under $H_0 : \theta_i(\tau) = \eta(\tau)$, for each fixed $\tau \in (0,1)$, we have

$$\hat{t}_i(\tau) \to_d \mathbf{N}(0,1).$$

(ii) If in addition $RQ^{-1}\mathbf{\Omega}Q^{-1}R'$ is of full rank, then under $H_0 : R\theta(\tau) = \eta(\tau)$, we have

$$\hat{F}(\tau) \to_d \chi^2_q.$$

The asymptotic distributions of LLin based nonparametric tests are shown in Theorem 5 below.

Theorem 5. *Suppose that, in addition to the conditions of Theorem 3, $\int x^4K^2 < \infty$, and **A4** holds.*

(i) Under $H_0 : \theta_i(\tau) = \eta(\tau)$, for each fixed $\tau \in (0,1)$, we have

$$\tilde{t}_i(\tau) \to_d \mathbf{N}(0,1), \tag{23}$$

and under $H_0 : \theta_i^{(1)}(\tau) = \eta(\tau)$,

$$\tilde{t}_i^{(1)}(\tau) \to_d \mathbf{N}(0,1). \tag{24}$$

(ii) If in addition $RQ^{-1}\mathbf{\Omega}Q^{-1}R'$ is of full rank, then under $H_0 : R\theta^{(1)}(\tau) = \eta(\tau)$ we have

$$\tilde{F}(\tau) \to_d \chi^2_q. \tag{25}$$

Remark 8. The asymptotic power rates of the test statistics under consideration are determined by the convergence rates of the LLev and LLin estimators involved. In particular, it can be easily checked that $\hat{t}_i(\tau), \tilde{t}_i(\tau)$ attain $\sqrt{n/c_n}$-divergence under H_1, whilst $\tilde{t}_i^{(1)}(\tau)$ a $\sqrt{n/c_n^3}$-divergence. Therefore, tests for the parameter derivatives are less powerful. These results are standard in the non-parametric literature (e.g. see Li and Racine, 2006).

Remark 9. Note that $\tilde{t}_i^{(1)}(\tau)$ and $\tilde{F}(\tau)$ can be used for testing parameter constancy with respect to time. In particular, the former statistic can be utilised for testing time variability in a single parameter (i.e. $H_0 : \theta_i^{(1)}(\tau) = 0$, for each $\tau \in (0,1)$) whilst the latter can be employed for testing time variability in multiple parameters (e.g. $\theta^{(1)}(\tau) = \mathbf{0}$, for each $\tau \in (0,1)$).

5 Extensions to Structural Regression

The basic limit theory provided in Section 3 is general enough for the asymptotic analysis of TVP estimators and related test statistics for structural regressions of the form

$$y_k = \theta(k/n)'\mathbf{x}_k + e_k, \tag{26}$$

where $\theta(.)$, $\mathbf{x}_k$ and e_k are as in (14). The actual difference between (14) and (26) is that in the former regressors are predetermined whereas in the latter they are contemporaneously generated with the regression error. We consider the limit properties of an IV-LLev type of estimator when a p-dimensional vector of instruments $\mathbf{z}_k$ is available. The underlying assumptions about the instruments are that they are uncorrelated with e_k, correlated with $\mathbf{x}_k$, and strictly stationary and ergodic. These are shown in detail below.

A5 $\mathbf{z}_k$ is a p-dimensional $\mathcal{F}_{k-1}$-measurable vector, where $\mathcal{F}_k$ is defined as in **A1.**

A6 $(\mathbf{z}_k, \mathbf{x}_k, \sigma_k)$ is (strictly) stationary, ergodic so that $E(\|\mathbf{x}_1\|\,\|\mathbf{z}_1\|) < \infty$,

$$E\left\{\|\mathbf{z}_1\|^2 + \sigma_1^2\left(1 + \|\mathbf{z}_1\|^2\right)\right\} < \infty$$

and $E\left(\mathbf{z}_1\mathbf{x}_1'\right)$ is of full rank.

The LLev-IV estimator under consideration is defined as

$$\hat{\theta}_{IV}(\tau) := \arg\min_{\mathbf{a}\in\mathbb{R}^p} G_n(\mathbf{a})'G_n(\mathbf{a}),\ G_n(\mathbf{a}) := \sum_{k=1}^{n}\left(y_k - \mathbf{a}'\mathbf{x}_k\right)\mathbf{z}_k K\left[c_n\left(k/n - \tau\right)\right].$$

Theorem 6. *Suppose that:*

(*a*) $\{y_k\}_{k\in\mathbb{N}}$ *is generated by (26);*

(*b*) ***A1-A2*** *and* ***A5-A6*** *hold;*

(*c*) $\theta(.)$ *is Hölder continuous on* $[0,1]$ *of order* $\gamma \in (0,1]$*;*

(*d*) K *satisfies* ***A3*** *and* $\int x^\gamma K < \infty$*;*

(*e*) $c_n/n + n/c_n^{1+2\gamma} \to 0$.

Then, for each fixed $\tau \in (0,1)$,

$$\sqrt{\frac{n}{c_n}}\left(\hat{\theta}_{IV}(\tau) - \theta(\tau)\right) \to_d \mathbf{N}\left(\mathbf{0}, Q_3^{-1}\mathbf{\Omega}_3 Q_3'^{-1}\right), \tag{27}$$

where $Q_3 = E(\mathbf{z}_1\mathbf{x}_1') \int K$ *and* $\mathbf{\Omega}_3 = E(\sigma_1^2\mathbf{z}_1\mathbf{z}_1') \int K^2$.

Remark 10. Giraitis et al. (2021) assume that the regressors are generated by a secondary regression model of the form

$$\mathbf{x}_k = \Phi_{kn}\mathbf{z}_k + \upsilon_k,$$

where $\mathbf{z}_k$ is a set of observable instruments, Φ_{kn} a TVP matrix, and υ_k an error term endogenous in the primary regression. Under our assumptions, an obvious set of instruments is regressors lagged by one period i.e. $\mathbf{z}_k = \mathbf{x}_{k-1}$. Note in this case $\mathbf{z}_k$ is predetermined with respect to the regression error e_k therefore satisfying the orthogonality requirement. Further, time dependence in $\mathbf{x}_k$ is a necessary condition for the relevance requirement i.e. invertibility of $E(\mathbf{z}_1\mathbf{x}_1')$.

Consider the following non-parametric test statistics based on $\hat{\theta}_{IV}(\tau)$

$$\hat{t}_{IV,i}(\tau) = \frac{\hat{\theta}_{IV,i}(\tau) - \eta(\tau)}{\sqrt{\left[\hat{\mathcal{Q}}_{zx,n}^{-1}\hat{\mathbf{\Omega}}_{zz,n}\hat{\mathcal{Q}}_{xz,n}^{-1}\right]_{ii}}}$$

and

$$\hat{F}_{IV}(\tau) = \left[R\hat{\theta}_{IV}(\tau) - \eta(\tau)\right]' \left[R\hat{\mathcal{Q}}_{zx,n}^{-1}\hat{\mathbf{\Omega}}_{zz,n}\hat{\mathcal{Q}}_{xz,n}^{-1}R'\right]^{-1} \left[R\hat{\theta}_{IV}(\tau) - \eta(\tau)\right],$$

where

$$\left\{\hat{\mathcal{Q}}_{zx,n}, \hat{\mathbf{\Omega}}_{zz,n}\right\} := \left\{\sum_{k=1}^{n} \mathbf{z}_k\mathbf{x}_k' K_{kn},\ \sum_{k=1}^{n} \hat{e}_k^2\mathbf{z}_k\mathbf{z}_k' K_{kn}^2\right\}, \text{ and } \{Q_{zx}, \mathbf{\Omega}_{zz}\} := \left\{E\mathbf{z}_1\mathbf{x}_1',\ E\mathbf{z}_1\mathbf{z}_1'\right\}.$$

Further, consider the following counterpart of **A4** for the structural case.

A4* (a) $E\|\mathbf{z}_1\|^4 < \infty$;

(b) either (i) $\sup_{k\geq 1} Eu_k^4 < \infty$ or

(ii) $Y_k := \sigma_k^2[\alpha'\mathbf{z}_k]^2\left[u_k^2 - 1\right]$ is uniformly integrable for all $\alpha \in \mathbb{R}^p$.

The limit distributions of these statistics are summarised in the following result.

Theorem 7. *Suppose that, in addition to the conditions of Theorem 6,* $\int x^2K^2 < \infty$*, and* ***A4**** *hold.*

(i) Under $H_0 : \theta_i(\tau) = \eta(\tau)$*, for each fixed* $\tau \in (0, 1)$*, we have*

$$\hat{t}_{IV,i}(\tau) \to_d \mathbf{N}(0, 1).$$

(ii) If in addition $RQ_{zx}^{-1}\mathbf{\Omega}_{zz}Q_{xz}^{-1}R'$ is of full rank, then under $H_0 : R\theta(\tau) = \eta(\tau)$, we have

$$\hat{F}_{IV}(\tau) \to_d \chi^2_q.$$

6 Application to the Predictability of Stock Returns

We utilise the proposed methods for an empirical investigation of the return-risk relationship. There is an extensive literature in this area -see e.g. Christensen and Nielsen (2006), Ang and Bekaert (2007), Bollerslev et al. (2009), Bollerslev et al. (2013), Bandi et al. (2019), Andersen and Varneskov (2021). Many research papers study predictive regressions of the form

$$r_k = \mu + \beta x_{k-1} + e_k, \tag{28}$$

where r_k are stock returns relating to some stock index, x_k some predictor capturing risk (e.g. realised variance, risk neutral return variation), and e_t a martingale difference regression error, and consider tests for the so called predictability hypothesis. i.e. $H_0 : \beta = 0$. In most datasets utilised in empirical work, there is strong evidence that various risk variables exhibit stationary long memory (see e.g. Bollerslev et al., 2013; p. 411). In a few cases e.g. Welch and Goyal (2008), Andersen and Varneskov (2021) memory estimates for volatility variables are slightly above the stationarity threshold (see e.g. Table 1). Further, inflation as another commonly used predictor appears to exhibit either stationary long memory or memory in the vicinity of $d = 0.5$ (see e.g. Hassler and Wolter, 1995; Baillie et al., 1996; Andersen et al., 2001; Hassler and Pohle, 2019). For example inflation memory estimates for the Welch-Goyal dataset range from $d = 0.19$ to $d = 0.48$ depending on the sampling frequency (see Table 1). The predictability hypothesis is typically tested with the aid of some t-stastistic based on parametric or semi-parametric estimators (e.g. Christensen and Nielsen, 2006) of the regression parameters that are largely assumed to be time invariant. In a recent paper Demetrescu et al. (2022) develop inferential methods for the predictability hypothesis that allow for a TVP slope coefficient, but not for a TVP intercept. The method can be applied to predictive regressions with either stationary or nonstationary predictors as long as IVX instruments are employed.

The proposed methods allow for testing the predictability hypothesis in situations where memory parameter is $|d| < 1/2$ and all regression parameters are TVP.[15] In particular, we

[15] As remarked before (e.g. Remark 5), we expect that Theorems 4 and 5 also hold for fractional $d = 1/2$ and MI processes.

Table 1: Memory Estimates (bandwidth $n^{0.65}$)

	Monthly		Quarterly		Annual	
	LW	ELW	LW	ELW	LW	ELW
Returns	0.07	0.069	0.14	0.15	-0.12	-0.08
SVAR	0.53	0.53	0.46	0.47	0.33	0.35
Inflation	0.46	0.47	0.48	0.48	0.19	0.22

consider the following specification

$$r_k = \mu(k/n) + \beta(k/n)x_{k-1} + e_k. \tag{29}$$

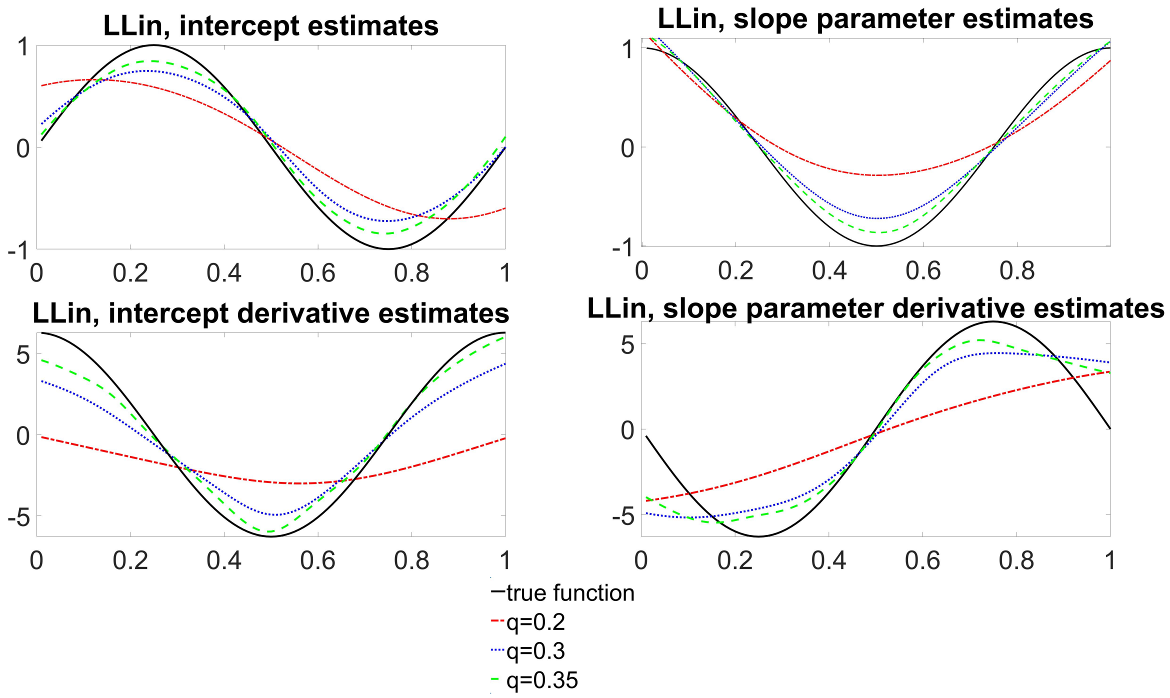
The tests under consideration can be either one-sided (t-tests) or two-sided. The implementation of the test procedures is straight forward. Statistics are based on studentised estimators and critical values are available from statistical tables. Further, the LLin estimator provides an easy method for testing for time invariance in the regression parameters e.g. the intercept.

The empirical study utilises the Welch-Goyal 2018 data set. The returns variable (r_k) is constructed by taking log differences of the SP500 index. The realised variance (SVAR) variable is the sum of squared daily returns on the SP500. We consider three different sampling frequencies of the aforementioned variables i.e. monthly, quarterly and annual. To get some idea about the persistent properties of the data we report memory estimates based on the local Whittle (LW; e.g. Robinson, 1995) and the exact local Whittle (ELW; cf. Shimotsu and Phillips, 2005) estimators. It can be seen from Table 1 that SP500 returns closely resemble an $I(0)$ process in all frequencies, while the predictive variable is persistent, exhibiting long memory. Note that for monthly data memory estimates are slightly above the nonstationarity threshold (i.e. $d = 1/2$), nevertheless the simulations show that for these values tests exhibit reasonably good size even in situations of very strong endogeneity.

As mentioned above, the independent variable has different memory characteristics than those of the predictor. This is a well known stylistic fact in the stock return predictability literature i.e. short-run returns exhibit very little persistence relative to various financial and macroeconomic predictors. This persistence mismatch casts some doubt on the plausibility of the commonly used linear specifications e.g. (28). Some authors suggest that nonlinearities in variables can provide a rebalancing mechanism. Certain nonlinear transformations (e.g. reduced growth, bounded) result in a reduction in the persistence of the underlying process. This approach has been discussed in Marmer (2008), Kasparis (2011), Kasparis et al. (2015), and Phillips

(2015) in the context of nonstationary predictive regressions. Other types of nonlinearities may facilitate a similar rebalancing effect. In particular, certain TVP specifications that induce episodic predictability events may result in a reduction in the persistence of the systematic part of the model i.e. $\beta(k/n)x_{k-1}$. For instance, a time varying slope parameter that is non zero for short time intervals is likely to result in a reduction in the signal of the process. Another approach to addressing misbalancing issues is to consider predictability over longer horizons. Long term returns are an accumulation of short-run terms and are therefore bound to exhibit higher persistence. In this work we focus on one period ahead returns leaving an investigation of long term predictability for future work.

Figure 1: LLin TVP estimates (averaged over replication paths)
(fractional regressor $d = 0.45$, $n = 1000$, GARCH(1,1) regression errors)



Bandwidth choice is very important for both estimation and testing. As mentioned before, in general there is a trade-off between size and power when it comes to bandwidth choice. Under-smoothing (i.e. larger values for $c_n = h_n^{-1}$) leads to smaller asymptotic bias and therefore to better size performance at the expense of slower divergent rates of test statistics under the alternative hypothesis. Nevertheless, the aforementioned trade-off is more subtle for non-parametric methods than for semi-parametric methods (e.g. IVX). In the current framework, there can be situations where under-smoothing may result in both better power and size. For instance, if the TVP varies wildly (e.g. when there are abrupt episodic predictability events) then over-smoothing may under estimate the variation in a TVP, and this may lead to power loss. This effect is illustrated in Figure 1 that shows LLin estimates of regression parameters and their derivatives for various bandwidth choices (i.e. $c_n = n^q, q = \{0.2, 0.3, 0.35\}$), and $\{\mu(\tau), \beta(\tau)\} = \{\sin(2\pi\tau), \cos(2\pi\tau)\}$. Note that this choice of TVPs entails periodic functions of period one over their domain (i.e. $(0, 1)$). It can be readily seen from Figure 1 that when over-smoothing is employed (e.g. $q = 0.2$) sudden changes in the TVPs are smoothed out i.e. the magnitude of slope parameters is understated. Additional simulations results that highlight the superior performance of tests when under-smoothing is employed appear in the Online Supplement (Hu et al., 2024).

We next provide estimates for $\mu(\tau)$ and $\beta(\tau)$, $\tau \in (0, 1)$ based on the LLev and the LLin estimators (see Figure 2). For the former we choose bandwidth $c_n = n^q$, with $q = 0.4$ and for the latter $q = 0.35$, which is slightly slower. These choices are close to the maximal under-smoothing allowed, given the theoretical constraints[16], and provide good performance both in terms of size and power according to the simulation study, as mentioned in the previous paragraph. It can be seen from Figure 2 that there is some time variation in both parameters for all sampling frequencies. First, the intercept parameter appears to be eventually increasing. Its maximal value for monthly returns is about 1% and for quarterly around 2%. The maximal value of the intercept for annual returns is higher, as expected due to compounding (5% for LLev and 15% for LLin). The slope parameter for monthly returns appears to be overall negative while for medium and long run returns (i.e. quarterly and annual) it oscillates around zero with some positive episodic events.

For a more rigorous investigation of episodic effects we utilise the LLev and LLin tests for the hypothesis $H_0 : \beta(\tau) = 0$. Rolling t-statistics are reported in Figure 3. We emphasise that these tests are pointwise for each $\tau \in (0, 1)$. For monthly returns, both test statistics indicate significant predictability for $\tau > 0.7$. In particular, for certain sub periods there is very strong evidence for negative predictability. In some cases the null hypothesis is rejected at 1% significance level even for two sided tests. For quarterly and annual returns there is some evidence of positive episodic predictability but is not as strong as those for monthly returns. In particular, for quarterly returns both tests reject the null at 5% significance (one sided tests) when τ is between 0.6 and 0.8 (approximately). For annual returns the null is rejected at 5% significance (one sided tests) only by the LLin test for $\tau < 0.15$ and $\tau > 0.9$. It should be noted however that the tests for quarterly and annual data are not that powerful due to sample size restrictions. Recall that the sample size for monthly data is $n = 1606$ while for quarterly and annual is $n = 535$ and $n = 133$ respectively.

[16] i.e. condition (e) in Theorem 2 and 3.

Figure 2: Local Level/Linear TVP Estimates

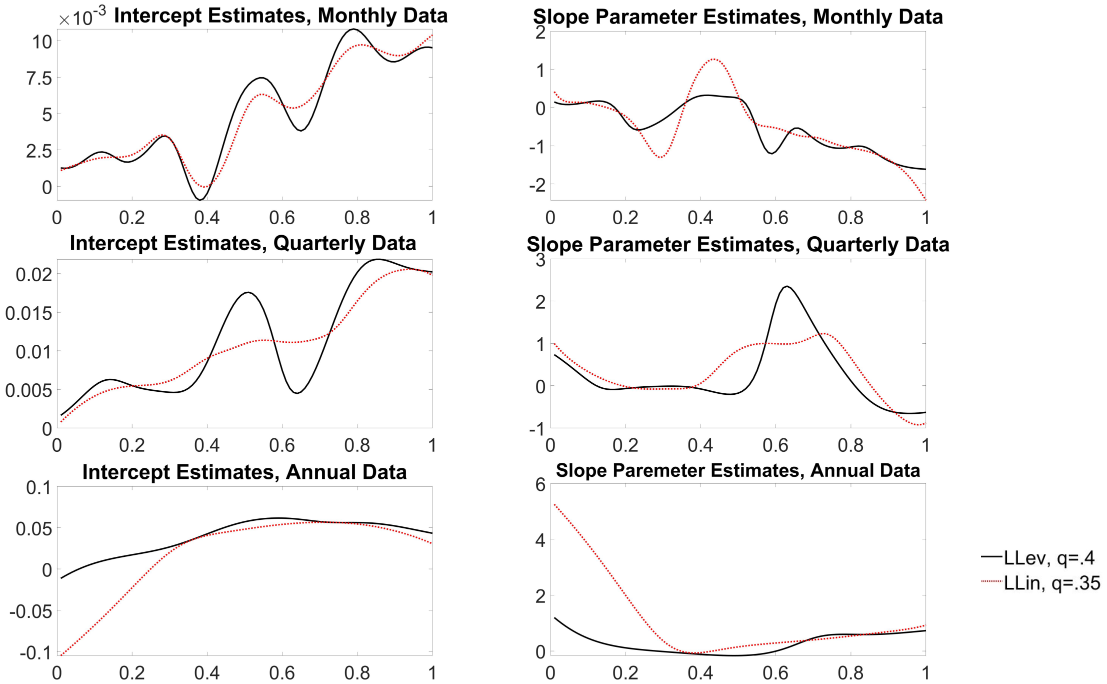


Figure 3: Local Level/Linear t-statistics for $H_0 : \beta(\tau) = 0$

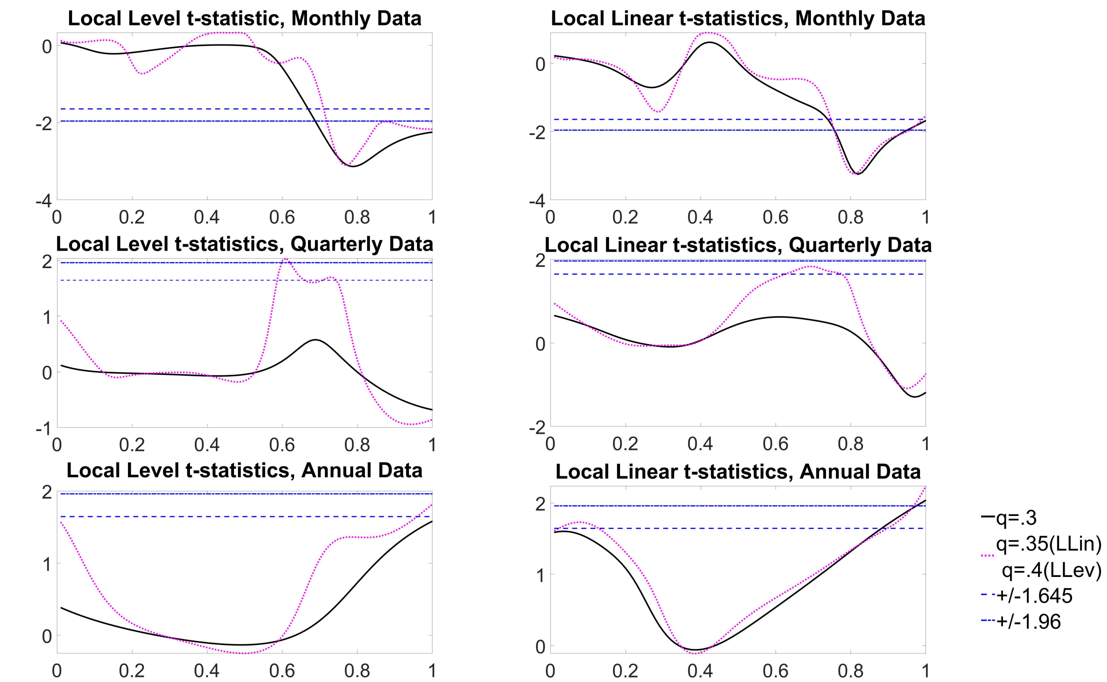
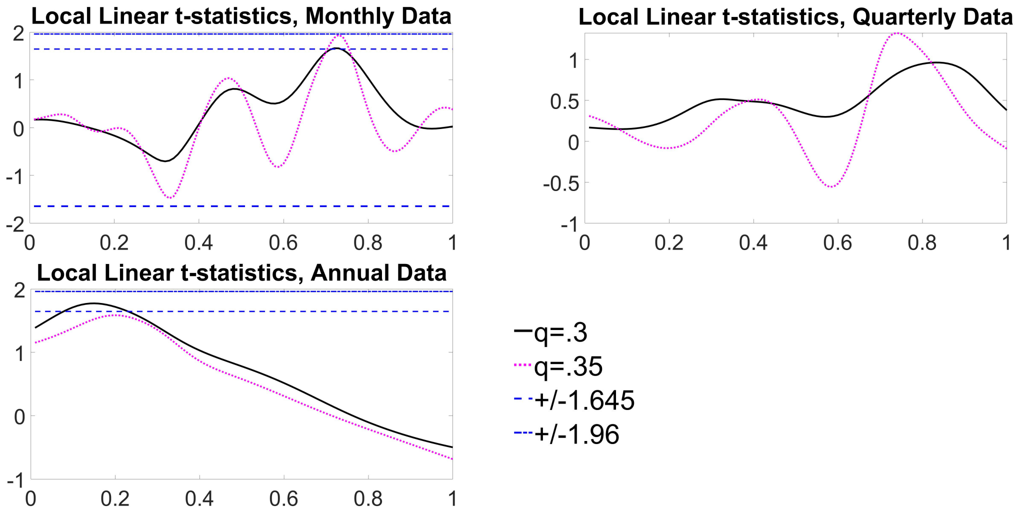


Figure 4: Local Linear t-statistics for $H_0 : \partial\mu(\tau)/\partial\tau = 0$



Finally, we utilise the LLin t-test for the hypothesis $H_0 : \partial\mu(\tau)/\partial\tau = 0$ i.e. no time variation in the intercept of (13). As explained in Section 2, neglecting time variation in the parameter of interest (i.e. β here) results in poor power. On the other hand, neglecting time variation in a "nuisance parameter" e.g. the intercept or the slope parameter of some other covariate is likely to result in inferior size control due to incorrect centering. Therefore, in practical work it is useful to know if there is time variation in the intercept. Figure 4 reports values for rolling t-statistics for the latter hypothesis. The tests show evidence for some episodic variation in the intercept for monthly and annual returns (significant at 5% level for one-sided tests). Our findings on the time variation of the regression intercept are likely to be conservative. As mentioned before, derivative estimators yield less powerful tests than those based on slope parameters due to slower convergence rates. In particular, the LLin derivative test attains $\sqrt{n/c_n^3}$-divergence while its slope parameter counterpart attains $\sqrt{n/c_n}$-divergence.

Supplementary Material

Hu, Z., Kasparis, I. and Wang, Q. (2024): Supplement to "Chronological trimming methods for nonlinear predictive regressions with stationary persistent data", Econometric Theory Supplementary Material. To view, please visit: [[doi will be inserted here by typesetter]]

References

[1] Amihud, Y. and Hurvich, C.M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, **39**(4), 813–841.

[2] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001) The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, **96**, 42–55.

[3] Andersen, T.G. and Varneskov, R.T. (2021). Consistent inference for predictive regressions in persistent economic systems. *Journal of Econometrics*, **224**(1), 215–244.

[4] Ang, A. and Bekaert, G. (2007). Stock return predictability: is it there? *Review of Financial Studies*, **20**(3), 651–707.

[5] Baillie, R. T., Chung, C.-F. and Tieslau, M. T. (1996) Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics*, **11**, 23–40.

[6] Bandi, F.M., Perron, B., Tamoni, A. and Tebaldi, C. (2019). The scale of predictability. *Journal of Econometrics*, **208**(1), 120–140.

[7] Bollerslev, T., Osterrieder, D., Sizova, N. and Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, **108**(2), 409–424.

[8] Bollerslev, T., Tauchen, G. and Zhou, H. (2009). Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies*, **22**(11), 4463–4492.

[9] Breitung, J. and Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics,* **187**(1), 358-375.

[10] Brockwell, P.J. and Davis, R. A. (1991) *Time Series: Theory and Methods*. Springer, New York.

[11] Chen, W.W. and Deo, R. (2009). Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood. *Econometric Theory*, **25**(5), 1143–1179.

[12] Christensen, B.J. and Nielsen M. Ø. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, **133**(1), 343–371.

[13] Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Annals of Statistics*, **28**(6), 1762–1794.

[14] Dahlhaus, R., Richter, S. and Wu, W.B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, **25**(2), 1013–1044.

[15] Demetrescu, M., Georgiev, I., Rodrigues, P. and Taylor, R. (2022). Testing for episodic predictability in stock returns. *Journal of Econometrics*, **227**(1), 85–113.

[16] Duffy J.A. and Kasparis, I. (2021). Estimation and inference in the presence of fractional $d = 1/2$ and weakly nonstationary processes. *Annals of Statistics*, **49**(2), 1195–1217.

[17] Francq, C. and Zakoian, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley & Sons.

[18] Giraitis, L., Kapetanios, G. and Marcellino, M. (2021). Time-varying instrumental variable estimation. *Journal of Econometrics*, **224**(2), 394–415.

[19] Giraitis L., Kapetanios G., Price, S. (2013). Adaptive forecasting in the presence of recent and ongoing structural change, *Journal of Econometrics*, **177**(2), 153-170.

[20] Giraitis, L., Kapetanios, G. and Yates, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, **179**(1), 46–65.

[21] Hassler, U. and Pohle, J.-M. (2019). Forecasting under long memory and nonstationarity. Working paper, Goethe University Frankfurt.

[22] Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: international evidence. *Journal of Business and Economic Statistics*, **13**(1), 37–45

[23] Hu, Z., Kasparis, I. and Wang, Q. (2021). Chronological trimming methods for nonlinear predictive regressions with persistent data. *Mimeo, University of Cyprus*

[24] Ibragimov, I.A. and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff Publishing, Groningen, The Netherlands.

[25] Kasparis, I. (2011). The Bierens test for certain nonstationary models. *Journal of Econometrics*, **158**(2), 221–230.

[26] Kasparis, I., Andreou, E. and Phillips, P.C.B. (2015). Nonparametric predictive regression. *Journal of Econometrics*, **185**(2), 468–494.

[27] Kolmogorov, A.N. and Rozanov, U.A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theory of Probability and Its Applications*, **5**(2), 204–208.

[28] Kostakis, A., Magdalinos, T. and Stamatogiannis, M.P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, **28**(5), 1506–1553.

[29] Kristensen, D. (2012). Non-parametric detection and estimation of structural change. *Econometrics Journal*, **15**(3) p. 420-461.

[30] Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, **74**(2), 209–235.

[31] Li, Q. and Racine, J.S. (2006). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

[32] Magdalinos, T. and Phillips, P.C.B. (2009). Econometric inference in the vicinity of unity. *Mimeo*, Singapore Management University.

[33] Marmer. V. (2008). Nonlinearity, nonstationarity and spurious forecasts. *Journal of Econometrics*, **142**(1), 1–27.

[34] Park, J.Y. and Phillips P.C.B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, **15**(3), 269–298.

[35] Park, J.Y. and Phillips P.C.B. (2000). Nonstationary binary choice, *Econometrica*, **68**(5), 1249-1280.

[36] Park, J.Y. and Phillips P.C.B. (2001). Nonlinear regressions with integrated time series, *Econometrica*, **69**(1), 117-161.

[37] Petrova, K. (2019). A quasi-Bayesian local likelihood approach to time varying parameter VAR models. *Journal of Econometrics*, **212**(1), 286–306.

[38] Phillips, P.C.B. (1995). Fully modified least squares and vector autoregression. *Econometrica*, **63**(5), 1023–78

[39] Phillips, P.C.B. (2015). Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, **13**(3), 521–555.

[40] Phillips, P.C.B., Li, D. and Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, **196**(1), 180–195.

[41] Phillips, P.C.B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, **136**(1), 115–130.

[42] Robinson, P.M. (1989). Nonparametric estimation of time varying parameters. In: Hackl, P. (Ed.), Economic Structural Change: Analysis and Forecasting. Springer Berlin, 253–264.

[43] Robinson, P.M. (1991). Time varying nonlinear regression. In: Hackl, P. (Ed.), Economic Structural Change: Analysis and Forecasting. Springer Berlin, 179–190.

[44] Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence. Annals of Statistics **23**(5), 1630–1661.

[45] Robinson, P.M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica*, **71**(6), 1727–1766.

[46] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes*. Chapman & Hall.

[47] Shimotsu, K. and Phillips, P.C.B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics*, **33**(4), 1890–1933.

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

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

[50] Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression*. World Scientific.

[51] Wang, Q. and Phillips, P.C.B. (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, **25**(3), 710–738.

[52] Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21**(4), 1455–1508.

[53] Yang, B., Long, W., Peng, L. and Cai, Z. (2020). Testing the Predictability of U.S. Housing Price Index Returns Based on an IVX-AR Model. *Journal of the American Statistical Association*, **115**(532), 1598–1619.

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) $\sup_{k\geq 1} E\left\|v_k\right\| < \infty$ *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

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

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

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

Similarly, using condition (b), we have

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

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

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

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

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

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,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\ell = 0, 1, 2$,

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

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

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

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

Indeed, by (A.19), we have

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

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

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

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

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

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

We finally prove (A.19). Set $v_k = \sigma_k^2 u_k^2 [\alpha' \mathbf{x}_{k-1}]^2$, 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 $A_0 := E\left(\sigma_1^2 [\alpha' \mathbf{x}_0]^2\right) = E v_k$ for each $k \geq 1$. Using Lemma 1, it suffices showing that

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

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

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

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

$$\begin{aligned}\tilde{\beta}_{OLS} &= \frac{\sum_{k=1}^{n}\beta(k/n)x_{k-1}^{2}}{\sum_{k=1}^{n}x_{k-1}^{2}}+\frac{\sum_{k=1}^{n}x_{k-1}u_{k}}{\sum_{k=1}^{n}x_{k-1}^{2}}\\ &= \frac{\sum_{k=1}^{n}\beta(k/n)x_{k-1}^{2}}{\sum_{k=1}^{n}x_{k-1}^{2}}+O_{P}(n^{-1/2})\rightarrow_{P}\int_{0}^{1}\beta(\tau)d\tau,\end{aligned}$$

as required for (3) in the main paper.

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

$$T_{1n}\rightarrow_{P}\left\{\int_{0}^{1}\beta(\tau)^{2}d\tau-\left[\int_{0}^{1}\beta(\tau)d\tau\right]^{2}\right\}E\left(x_{1}^{2}\right).$$

Further, noting that

$$E\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\beta(k/n)x_{k-1}u_{k}\right)^{2}=\sigma_{u}^{2}E\left(x_{1}^{2}\right)\frac{1}{n}\sum_{k=1}^{n}\beta(k/n)^{2}=\sigma_{u}^{2}E\left(x_{1}^{2}\right)\int_{0}^{1}\beta(\tau)^{2}d\tau+o(1)$$

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 $T_{3n} \rightarrow_P \sigma_u^2$. In view of the above, $\hat{\sigma}_u^2 \rightarrow_P \sigma_*^2$ and hence

$$n^{-1/2}\tilde{t}_{OLS}=\frac{\tilde{\beta}_{OLS}}{\sqrt{\hat{\sigma}_{u}^{2}\left[\frac{1}{n}\sum_{k=1}^{n}x_{k-1}^{2}\right]^{-1}}}\rightarrow_{P}\frac{\int_{0}^{1}\beta(\tau)d\tau}{\sqrt{\sigma_{*}^{2}\left[Ex_{1}^{2}\right]^{-1}}},$$

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*

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

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 $\Delta_n \to_P E(x_1^2)$ and

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

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

$$\begin{aligned}
&\frac{n}{\delta_n}\left(\tilde{\beta}_{OLS}-\beta\right)\\
=\;& [1+o_P(1)]\left[E(x_1^2)\right]^{-1}\left\{\delta_n^{-1}\sum_{k=1}^n x_{k-1}u_k+\delta_n^{-1}\sum_{k=1}^n \mu(k/n)x_{k-1}\right.\\
&\left.-\left(n^{-1}\sum_{j=1}^n \mu(j/n)\right)\left(\delta_n^{-1}\sum_{k=1}^n x_{k-1}\right)\right\}\\
=\;& \left[E(x_1^2)\right]^{-1}\left\{\delta_n^{-1}\sum_{k=1}^n \mu(k/n)x_{k-1}-\left(n^{-1}\sum_{j=1}^n \mu(j/n)\right)\left(\delta_n^{-1}\sum_{k=1}^n x_{k-1}\right)\right\}+o_P(1)\\
\to_d\;& \left(Ex_1^2\right)^{-1}\left[1,\;-\int_0^1 \mu(\tau)d\tau\right]\cdot \mathbf{N}\left(\mathbf{0},\Psi\right),
\end{aligned} \tag{B.4}$$

where we have used the result:

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

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

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

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

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

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

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

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

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

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)

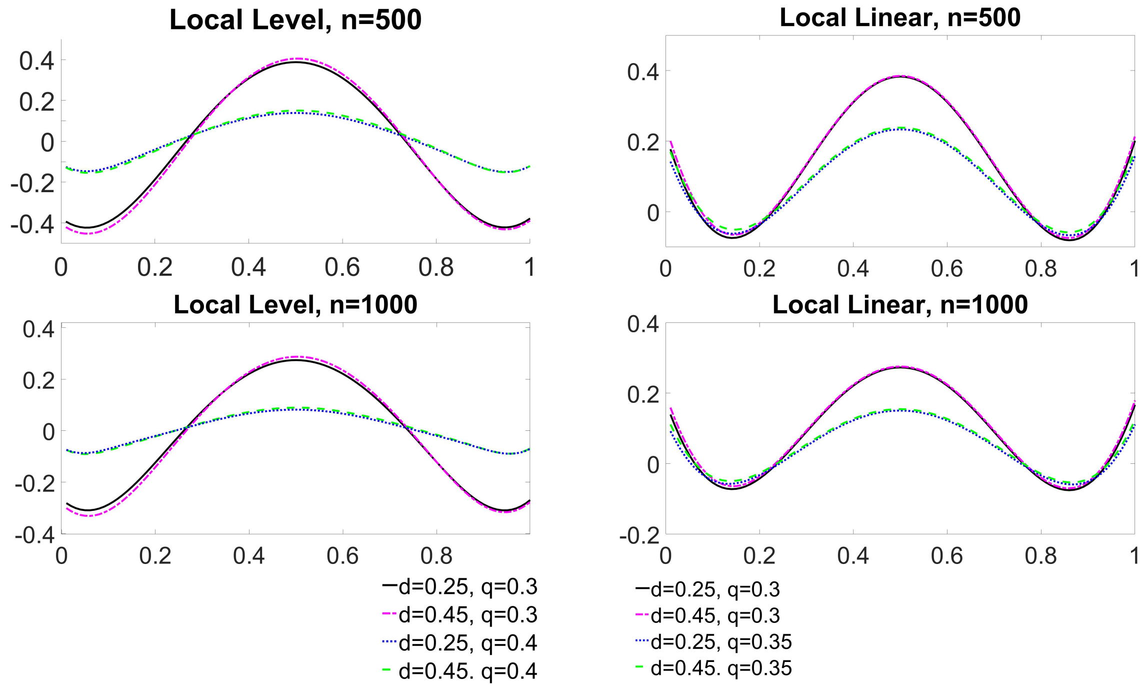
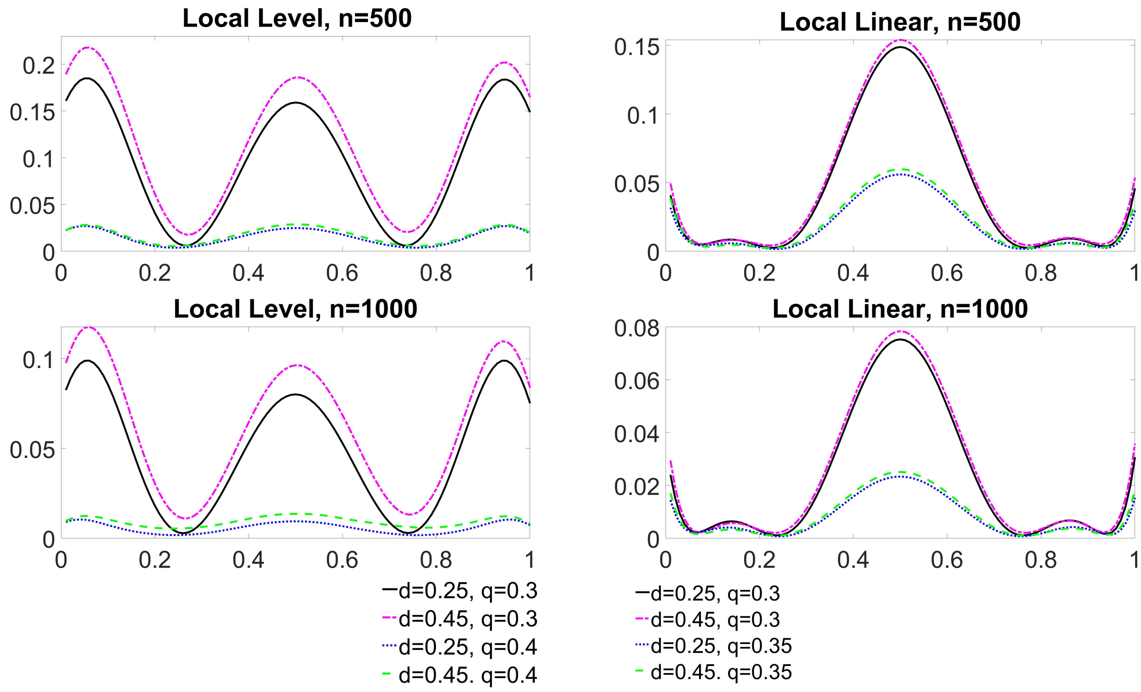


Figure 6: MSE of TVP Slope Estimators (plotted against τ)



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

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

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)

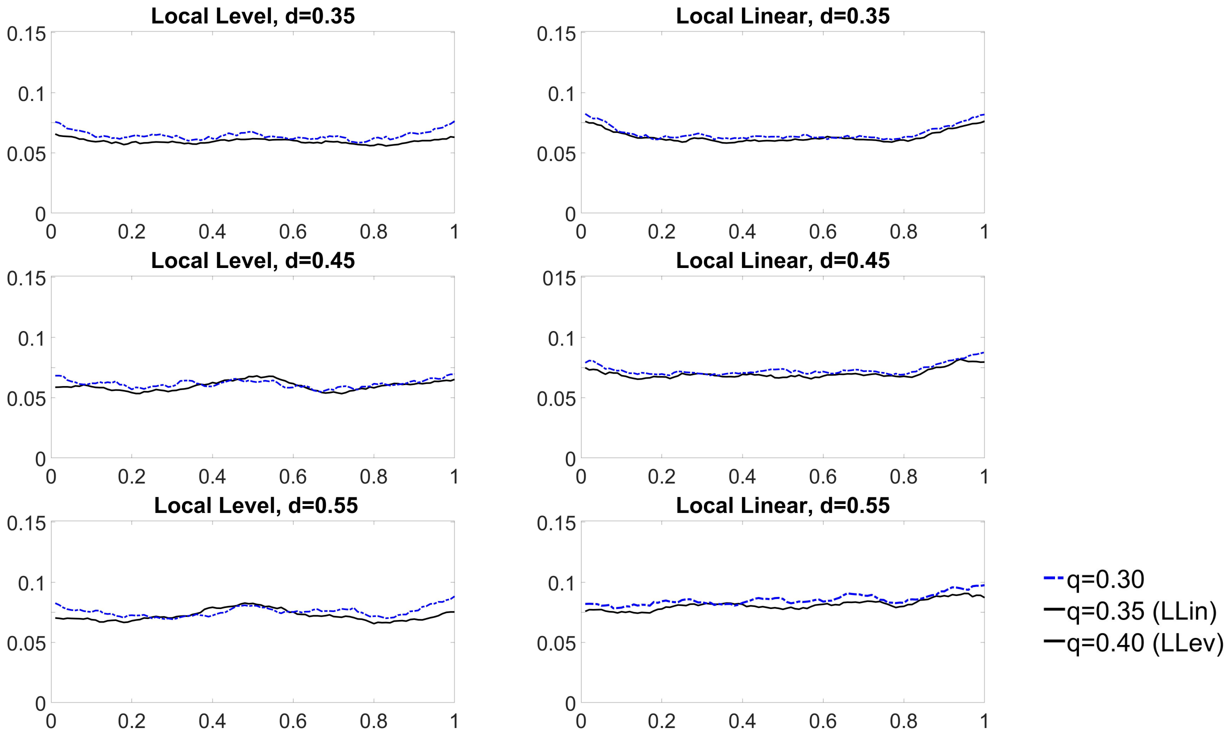


Figure 8: Empirical Size of t-tests against τ: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 1000$; $\delta = -0.55$; fractional regressor, GARCH(1,1) regression errors)

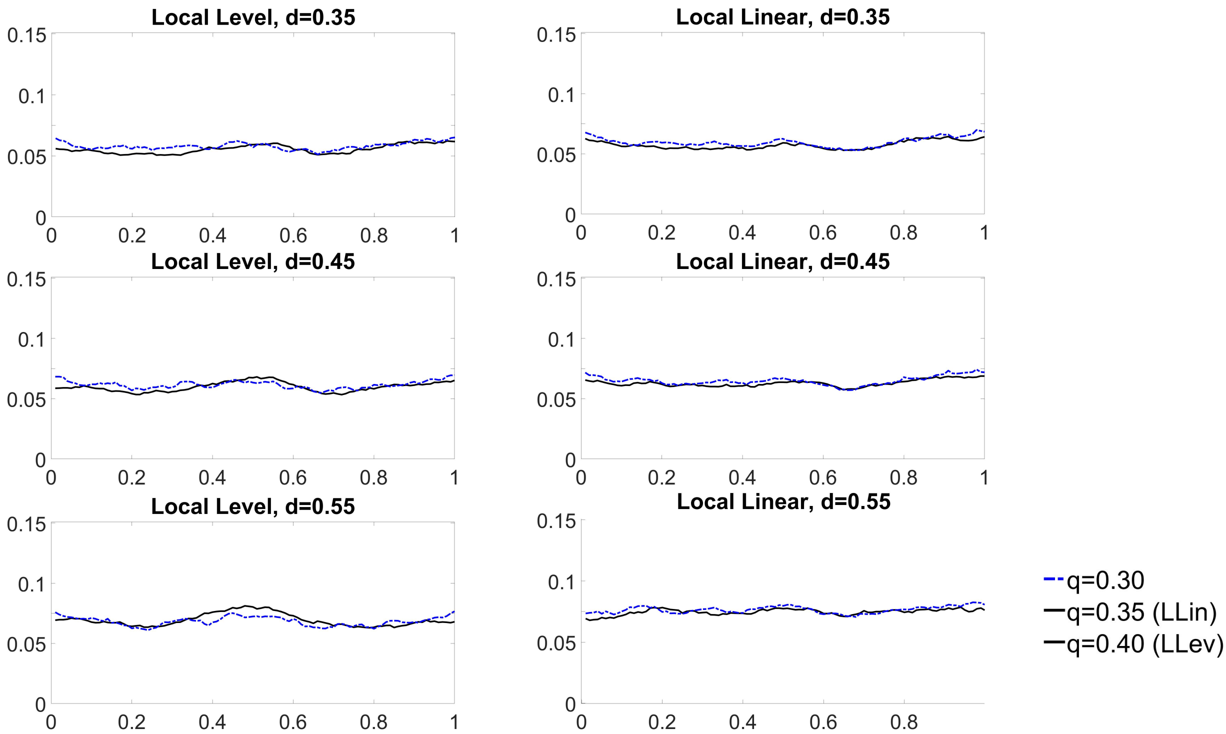


Figure 9: Empirical Size of t-tests against τ: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 500$; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

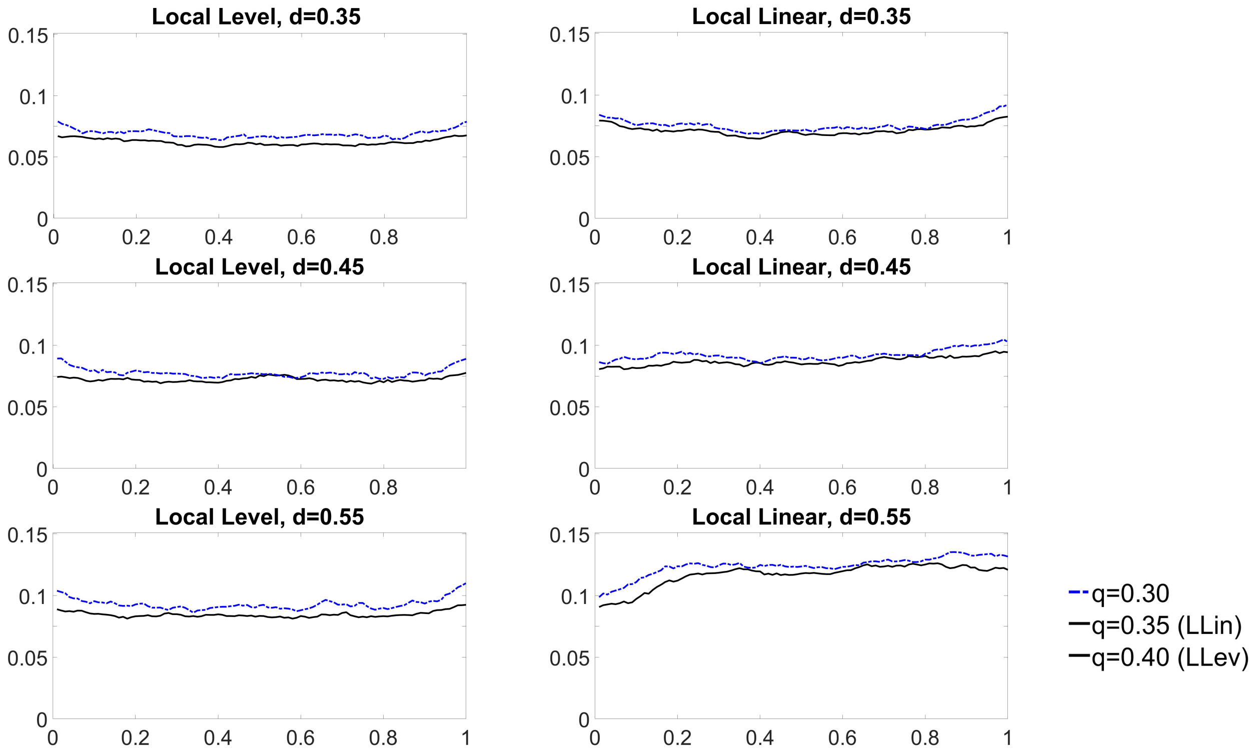
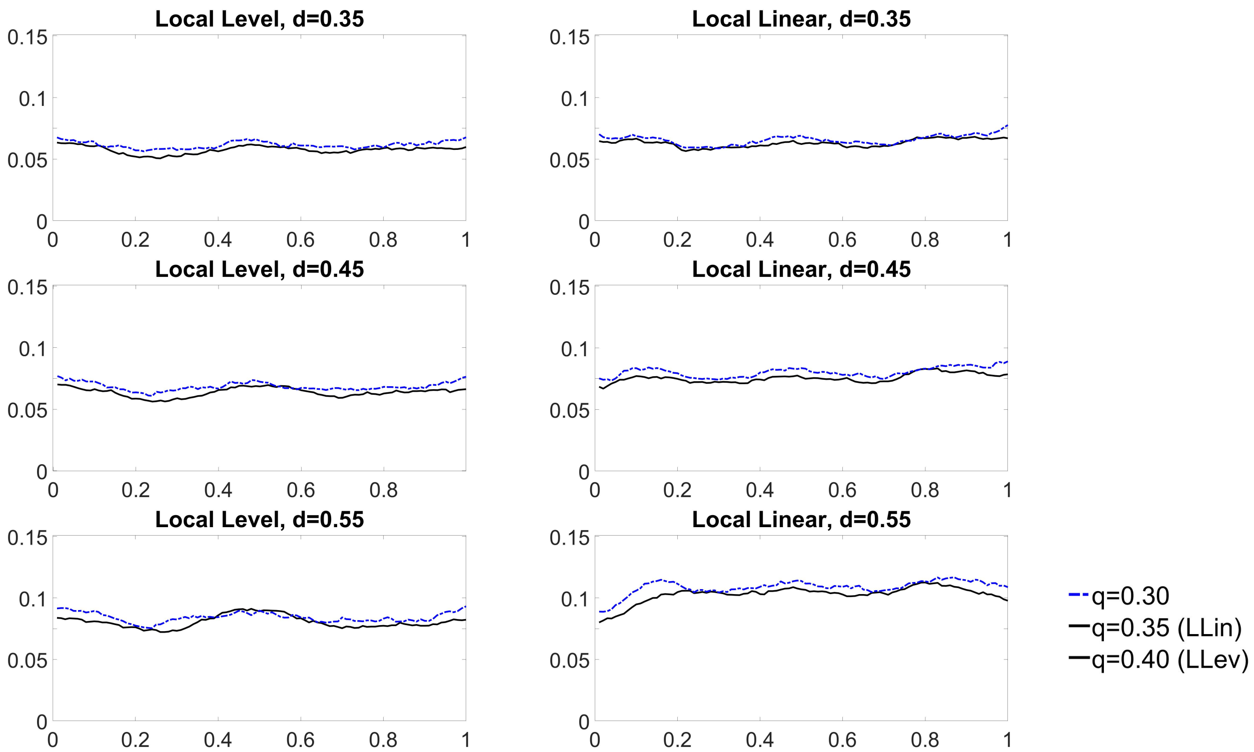


Figure 10: Empirical Size of t-tests: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 1000$; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)



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)

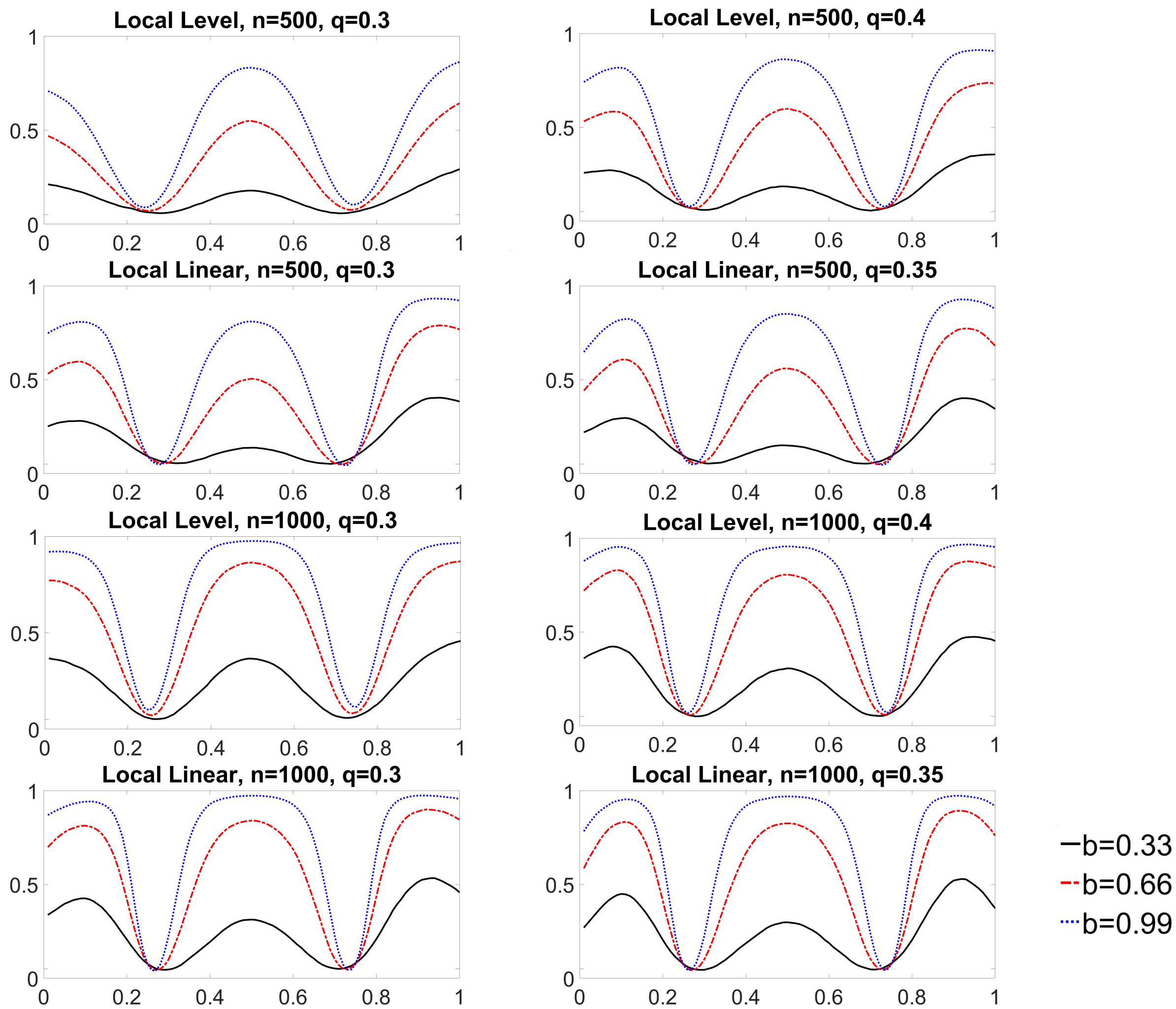
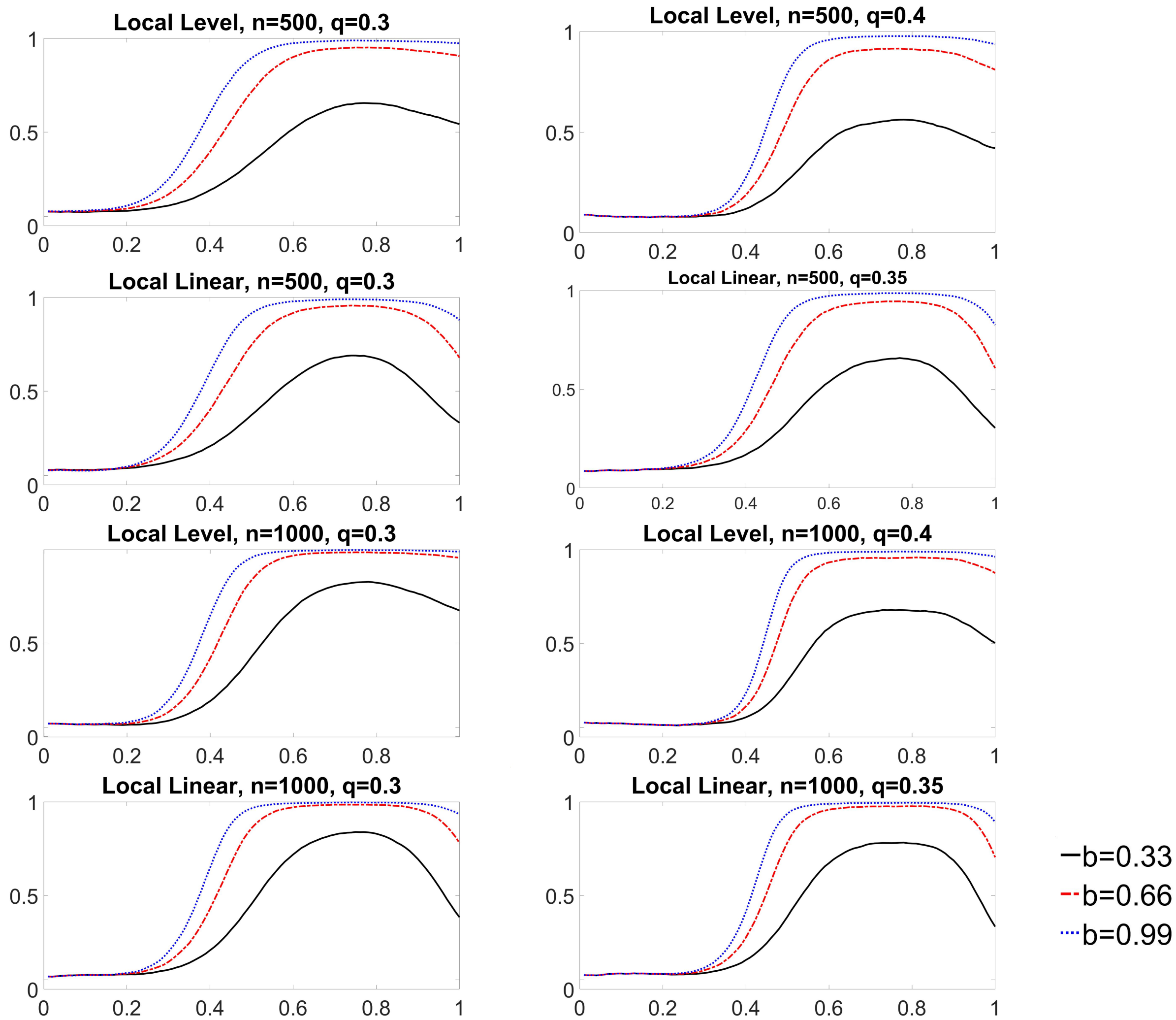


Figure 12: Empirical Power of t-tests: $H_1 : \beta(\tau) = b \cdot \{1 + \exp[-30(\tau - 0.5)]\}^{-1}$ (5% nominal size; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)



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)

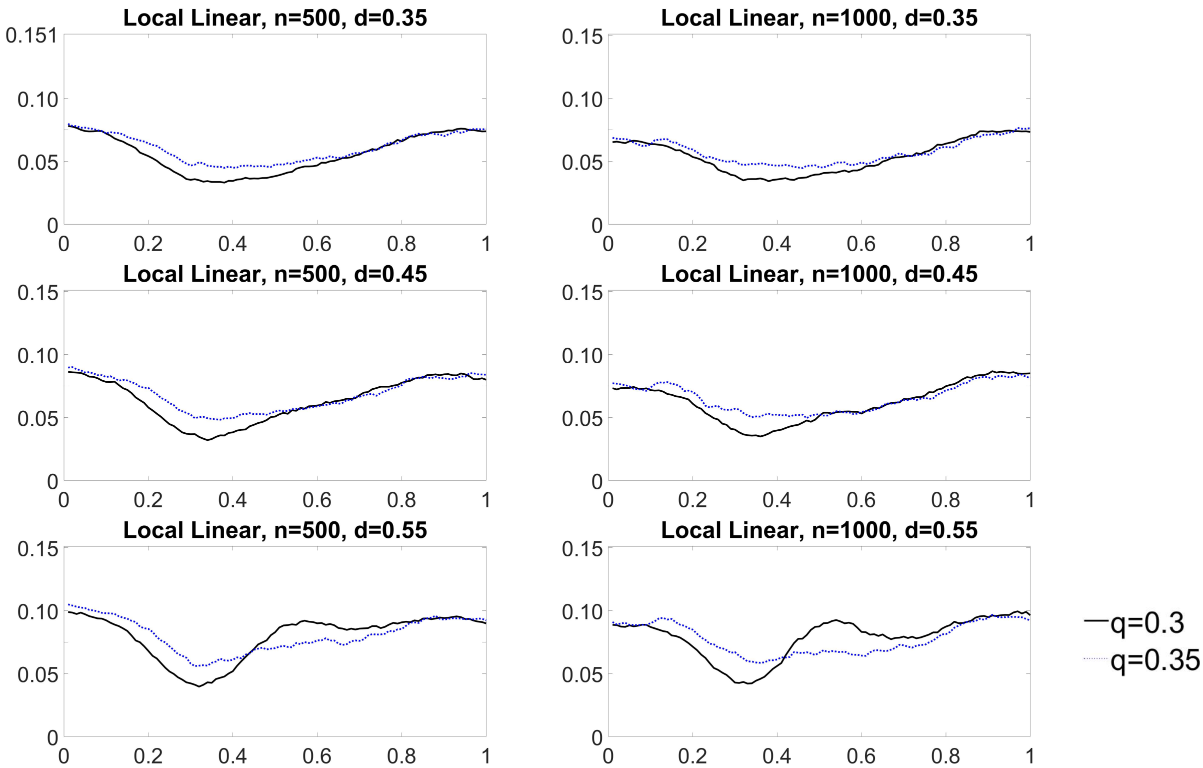
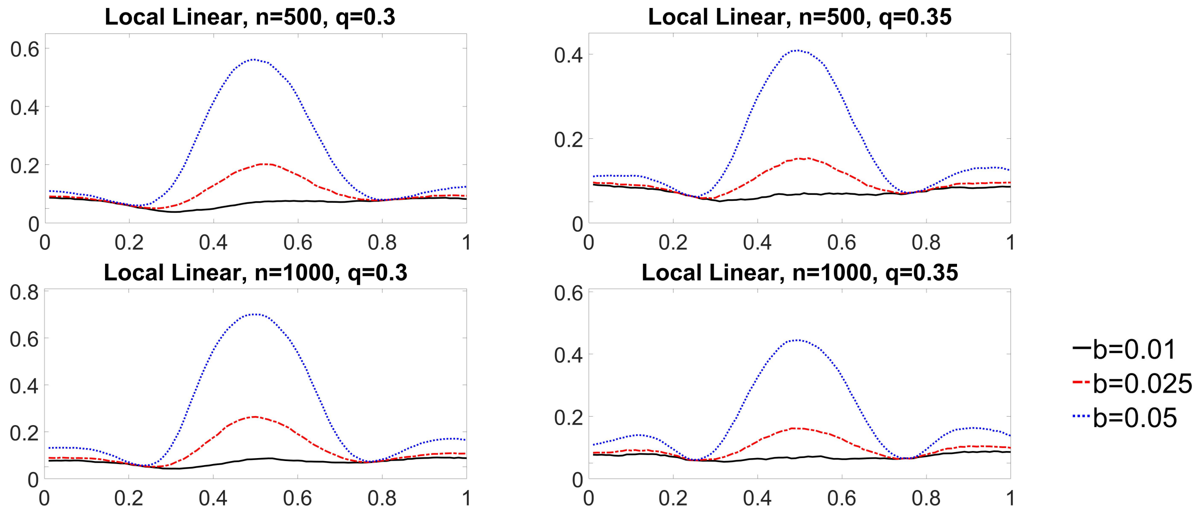


Figure 14 reports the rejection frequency of the latter test under the alternative hypothesis. In particular, the regression parameters are $\mu(\tau) = b \cdot \sin(2\pi\tau)$ with $b = \{0.01, 0.025, 0.05\}$, and $\beta(\tau) = 0.066 \cdot \cos(2\pi\tau)$. The memory of the predictor is $d = 0.45$ 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)

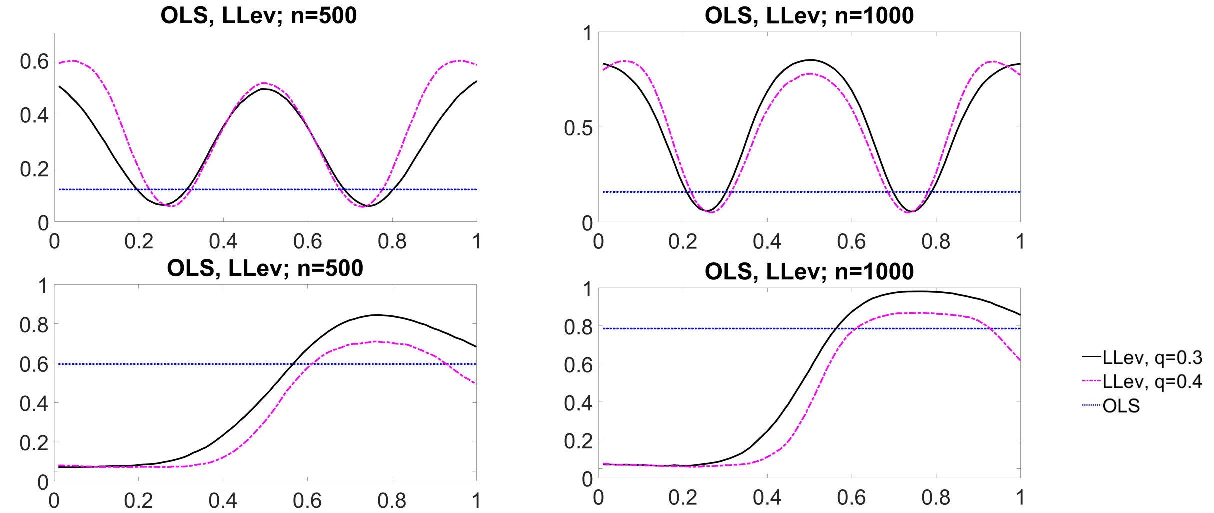
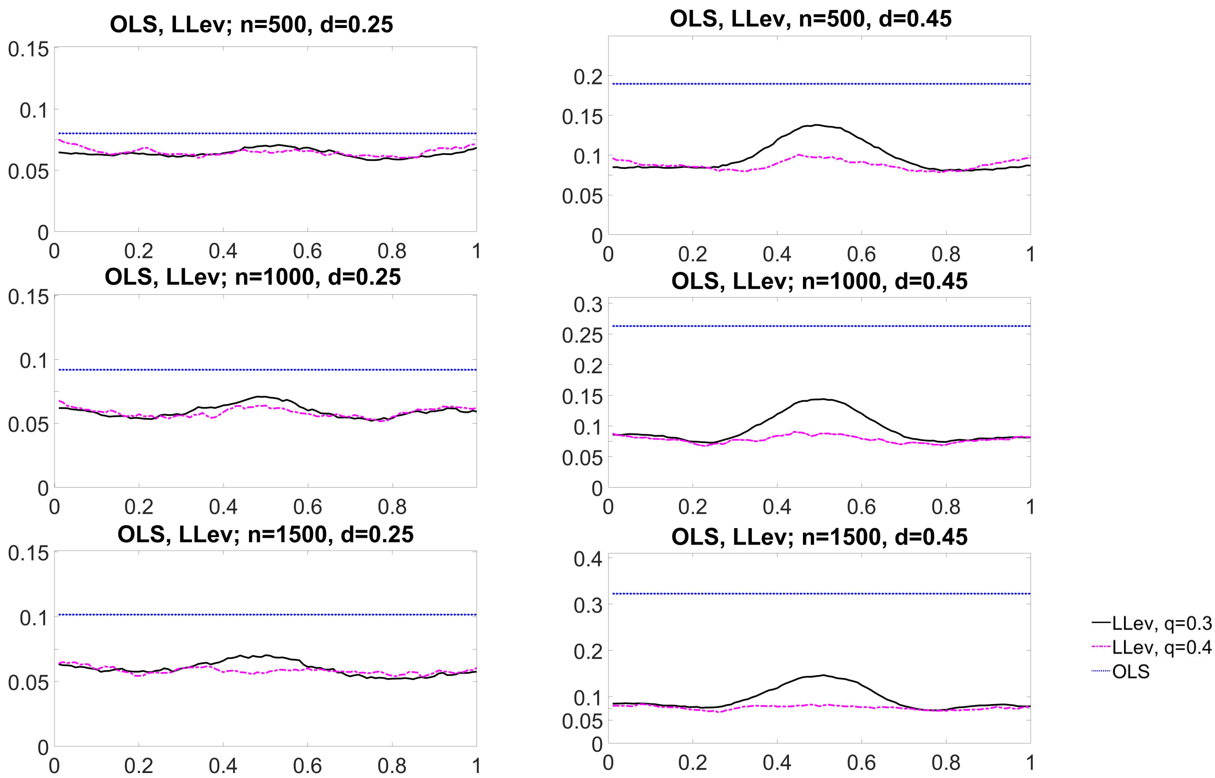


Figure 16: Empirical Size of OLS and LLev based t-tests: τ: $H_0 : \beta(\tau) = 0$
(5% nominal size; $\delta = -0.95$; $i.d.N(0, 1)$ regression errors)



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.