Econometric Theory, **0**, 2024, 1–27.
doi:10.1017/S0266466624000082

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

We consider local level and local linear estimators for estimation and inference in time-varying parameter (TVP) regressions with general stationary covariates. The latter estimator also yields estimates for parameter derivatives that are utilized 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 realized variance. We find evidence of time variability in the intercept as well as episodic predictability when realized variance is utilized as a predictor in TVP specifications.

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 modeled in terms of structural breaks in regression parameters. Nevertheless, *smooth* time-varying parameter (TVP) models have gained a lot of attention recently (see, e.g., Robinson, 1989, 1991; Dahlhaus, 2000 for earlier work in this area, and for more recent developments, Kristensen, 2012; Giraitis, Kapetanios, and Price, 2013; Giraitis, Kapetanios, and Yates, 2014; Phillips, Li, and Gao, 2017; Dahlhaus, Richter, and Wu, 2019;

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 the NSFC (Grant No. 11671373) and Wang acknowledges research support from the Australian Research Council. Address correspondence to Ioannis Kasparis, Department of Economics, University of Cyprus, Nicosia, Cyprus, email: kasparis@ucy.ac.cy.

Petrova, 2019; Giraitis, Kapetanios, and Marcellino, 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 nonconventional. 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, that is, 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 (nonpredictability) 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, Magdalinos, and Stamatogiannis, 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 nonconventional,[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

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

that rule out stationary long memory (cf. Kolmogorov and Rozanov, 1960). In particular, the basic paradigm under consideration entails specifications of the form

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

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), for example,

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

It can be readily seen that (1) allows for models nonlinear in variables, that is, 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 generalization to Robinson (1989) who considers TVP regressions with stationary mixing covariates and a partial generalization 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, that is, I(1) models.

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.

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

[3]That is, $H_0 : \beta_j = 0$ for some $j = 1,\ldots,p-1$.

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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 et al., 2017; Duffy and Kasparis, 2021). It should be emphasized, however, that the methods under consideration do not yield pivotal tests under nonstationarity in general, for example, 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 nonparametric 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, that is, the test statistic is simply a studentized LLin estimator. Furthermore, the derivative test statistic can be calculated for each regression point yielding a series of rolling test statistics that can distinguish 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, e.g., 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 realized volatility and inflation are stationary long memory or very close to the stationarity boundary, that is, of memory parameter $d \approx 0.5$.[6] Amihud and Hurvich (2004), Christensen and Nielsen (2006), Ang and Bekaert (2007), Bollerslev, Tauchen, and Zhou (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 organized as follows: Section 2 provides some theoretical results for the performance of parametric inference in the presence of

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

[6] See Section 6 for more details.

TVPs. Section 3 develops basic limit theory for kernel functionals of stationary processes. This limit theory is utilized 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 Supplementary Material (Hu et al., 2024). An extensive simulation study appears in the Supplementary Material 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 $\|x\| = \left(x'x\right)^{1/2}$. Further, for an $l \times m$-dimensional matrix $A = [a_{ij}]$, $\|A\| = \sum_{i=1}^{l}\sum_{j=1}^{m}\left|a_{ij}\right|$. 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, $\text{diag}\{a_1,\ldots,a_p\}$ denotes a $p \times p$ diagonal matrix with elements $\{a_1,\ldots,a_p\}$ 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 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., Lemmas 2 and 3) of the Supplementary Material. Simulations that support these theoretical findings are reported in Section C of the Supplementary Material.

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—that is, 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, that is, $d > 0$. 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).

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

2.1. 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 in the Supplementary Material (see Section B there) that, under *Assumption P*, the OLS estimator $\tilde{\beta}_{OLS}$ from regressing y_k on x_{k-1} satisfies

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

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

$$n^{-1/2}\tilde{t}_{OLS} = \frac{\tilde{\beta}_{OLS}}{\sqrt{\hat{\sigma}_u^2\left[n^{-1}\sum_{k=1}^n x_{k-1}^2\right]^{-1}}} \to_P \frac{\int_0^1 \beta(\tau)d\tau}{\sqrt{\sigma_*^2\left[Ex_1^2\right]^{-1}}}, \tag{4}$$

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 in the Supplementary Material)

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

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 canceled out by negative ones (i.e., $\beta(\cdot) < 0$) (see Figure 15 in Section C of the Supplementary Material for simulation results that illustrate these adverse power effects).

2.2. 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

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

(see, e.g., Wang, 2015, Exam. 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, Thm. 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

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

where Ψ is the matrix

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

with $0 < d < 1/2$ (see Lemma 3 in the Supplementary Material 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

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

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

$$\left|\tilde{t}_{OLS}\right| \to_P \infty. \quad \textbf{(10)}$$

More details for the validity of (6)–(10) can be found in Lemma 3 in the Supplementary Material. Figure 16 given in the Supplementary Material 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

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

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

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

[8] It follows from (6) to (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.

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

process. Further, let K be an integrable function and $m \in \{0,1,2\}$, and set

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

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

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, that is, $c_n = h_n^{-1}$ with $h_n \to 0$. The asymptotics of $\{S_n, M_n\}$ are utilized 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, Thm. 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, under *A1* and *A2*, $\sigma_k u_k$ is allowed to be a strictly stationary GARCH or ARCH(∞) model (e.g., Francq and Zakoian, 2010, Sect. 2.2). More details about the time series models

[10] That is, $K(x)$ 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] For example, $x_k = \sum_{i=0}^{\infty}\phi_i\xi_{k-i}, \xi_i \sim iid(0,\sigma_\xi^2), \sum_{i=0}^{\infty}\phi_i^2 < \infty$.

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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, for example, (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

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

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

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

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 heteroscedastic process, for example, 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). 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 $x_{k,j} = f_j(w_{k,j}), j = 1, \ldots, p-1$, with

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

and $\left(\xi_{k,1}, \ldots, \xi_{k,p-1}, \sigma_{k+1}\right) := \left(\xi_k', \sigma_{k+1}\right) \sim iid$. 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 $\alpha \in (0,2)$, $\sum_{i=0}^{\infty} \left|\phi_{i,j}\right|^{\alpha'} < \infty$ 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

[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 that Samorodnitsky and Taqqu (1994, Thm. 2.3.1) implies that each component series $\xi_{k,j}$, $j = 1, \ldots, p-1$, 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, pp. 68–69). Suppressing the occurrence of the index j, we have as $n, m \to \infty$

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

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

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

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

have a finite mean or variance nevertheless 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 $\left|f_j(x)\right| \leq C(1+|x|^{\eta})$, $C \in (0,\infty)$, $\eta \in (0,1)$, 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; Andersen 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, Sect. 2.1) (see also Robinson, 1989; Phillips et al., 2017; Giraitis et al., 2021 among others). Set

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

We can write (13) as

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

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

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

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

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

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* $\gamma \in (0,1]$, *that is,* $\|\theta(x)-\theta(y)\| \leq C\|x-y\|^{\gamma}$ *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, that is, Q^{-1} exists.*

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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

$$\sqrt{\frac{n}{c_n}}\left(\hat{\boldsymbol{\theta}}(\tau)-\theta(\tau)\right) \to_d \mathbf{N}\left(\mathbf{0}, Q_1^{-1}\Omega_1 Q_1^{-1}\right), \tag{16}$$

where $Q_1 = Q\int K$ *and* $\Omega_1 = \Omega\int K^2$.

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, Sect. 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

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

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* $\int x^3K < \infty$.
(e) $c_n/n + n/c_n^5 \to 0$*;*
(f) Q^{-1} *and* $\mathcal{K}_1^{-1}$ *exist.*

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

$$D_n\left(\begin{bmatrix} \widetilde{\theta}(\tau) \\ \widetilde{\theta}^{(1)}(\tau) \end{bmatrix} - \begin{bmatrix} \theta(\tau) \\ \theta^{(1)}(\tau) \end{bmatrix}\right) \to_d \mathbf{N}\left(\mathbf{0}, Q_2^{-1}\Omega_2 Q_2^{-1}\right), \tag{18}$$

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*

$$Q_2 = \mathcal{K}_1 \otimes Q \quad \textit{and} \quad \Omega_2 = \mathcal{K}_2 \otimes \Omega.$$

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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 $O_P(\sqrt{n/c_n^{1+2\gamma}})$, while 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 utilized for testing hypotheses about parameter constancy with respect to time (see also Remark 9).

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 et al., 2021, Thm. 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 nonstationary 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 nonsingular. We therefore expect that the proposed methods are valid even if the data are weakly nonstationary.

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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 nonparametric 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, $i = 1, \ldots, p$, be its ith element, and for a square matrix A, $[A]_{ii}$ denotes its ith diagonal element. Further, $\theta_i(\tau)$ denotes ith 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 $\beta_i(\tau)$, while (20) concerns the derivatives of the aforementioned parameters. The proposed tests utilize the estimators of (16) and (18). Set

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

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

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

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

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

for the null hypothesis (19), and

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

for the null hypothesis (20).

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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 $q \times p$ ($q \le p$) 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). Consider the following nonparametric F-statistic:

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

and

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

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 $\hat{\boldsymbol{\Omega}}_n$ and $\tilde{\boldsymbol{\Omega}}_n$, Assumption *A4* will be utilized.

A4 (a) $E\|\mathbf{x}_0\|^4 < \infty$;
(b) either (i) $\sup_{k\ge 1} Eu_k^4 < \infty$ 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 Ω_1 and Ω_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 $\hat{\boldsymbol{\Omega}}_n$ for the studentization 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 $\sup_k E(|u_k|^{2(1+\varepsilon)}|\mathcal{F}_{k-1}) \le C < \infty$ *a.s.*,

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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 gives the limit properties of the LLev-based tests.

THEOREM 4. *Suppose that, in addition to the conditions of Theorem 2, $\int x^2 K^2 < \infty$, 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, $RQ^{-1}\Omega Q^{-1}R'$ is of full rank, then under $H_0 : R\theta(\tau) = \eta(\tau)$, we have*

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

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

THEOREM 5. *Suppose that, in addition to the conditions of Theorem 3, $\int x^4 K^2 < \infty$, 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 $RQ^{-1}\Omega Q^{-1}R'$ is of full rank, then under $H_0 : R\theta^{(1)}(\tau) = \eta(\tau)$, we have*

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

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 nonparametric literature (see, e.g., 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 utilized for testing time variability in a single parameter (i.e., $H_0 : \theta_i^{(1)}(\tau) = 0$, for each

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

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

$\tau \in (0,1)$) while 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

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

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 $\tau \in (0,1)$,

$$\sqrt{\frac{n}{c_n}}\left(\hat{\boldsymbol{\theta}}_{IV}(\tau) - \theta(\tau)\right) \to_d \mathbf{N}\left(\mathbf{0}, Q_3^{-1}\Omega_3 Q_3'^{-1}\right), \tag{27}$$

where $Q_3 = E\left(\mathbf{z}_1\mathbf{x}_1'\right)\int K$ *and* $\Omega_3 = E\left(\sigma_1^2\mathbf{z}_1\mathbf{z}_1'\right)\int K^2$.

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

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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, that is, $\mathbf{z}_k = \mathbf{x}_{k-1}$. 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, that is, invertibility of $E\left(\mathbf{z}_1\mathbf{x}_1'\right)$.

Consider the following nonparametric test statistics based on $\hat{\boldsymbol{\theta}}_{IV}(\tau)$

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

and

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

where

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

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 summarized 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, $RQ_{zx}^{-1}\Omega_{zz}Q_{xz}^{-1}R'$ is of full rank, then under $H_0 : R\theta(\tau) = \eta(\tau)$, we have*

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

6. APPLICATION TO THE PREDICTABILITY OF STOCK RETURNS

We utilize 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

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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

where r_k are stock returns relating to some stock index, x_k some predictor capturing risk (e.g., realized variance, risk neutral return variation), and e_t a martingale difference regression error, and consider tests for the so-called predictability hypothesis. That is, $H_0 : \beta = 0$. In most datasets utilized 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 Wolters, 1995; Baillie, Chung, and Tieslau, 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-statistic 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 consider the following specification:

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

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 studentized 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, for example, the intercept.

[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.

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

The empirical study utilizes the Welch–Goyal 2018 dataset. The returns variable (r_k) is constructed by taking log differences of the SP500 index. The realized variance (SVAR) variable is the sum of squared daily returns on the SP500. We consider three different sampling frequencies of the aforementioned variables, that is, 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, that is, 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, for example, (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, Andreou, and Phillips (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, that is, $\beta(k/n)x_{k-1}$. For instance, a time-varying slope parameter that is nonzero 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.

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 nonparametric 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 underestimate the variation in a TVP, and this may lead to power loss. This effect is illustrated in Figure 1

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

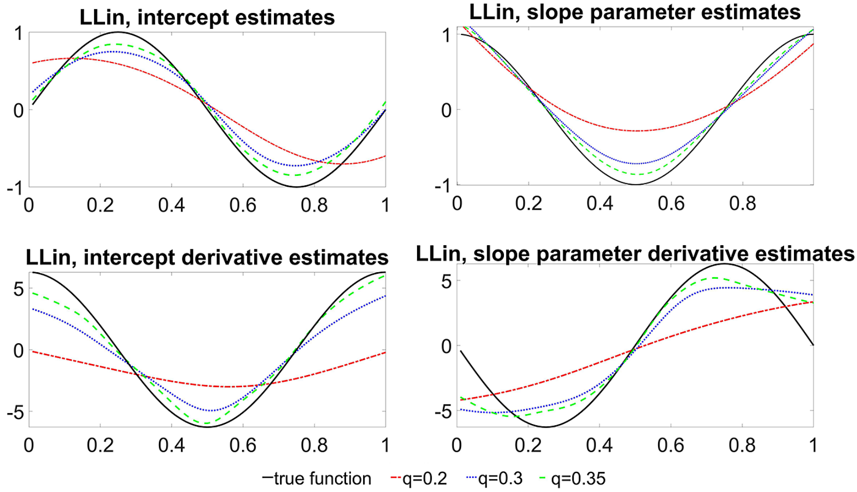

FIGURE 1. LLin TVP estimates (averaged over replication paths) (fractional regressor $d = 0.45$, $n = 1,000$, GARCH(1,1) regression errors).

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, that is, 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 Supplementary Material (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

[16]That is, condition (e) in Theorems 2 and 3.

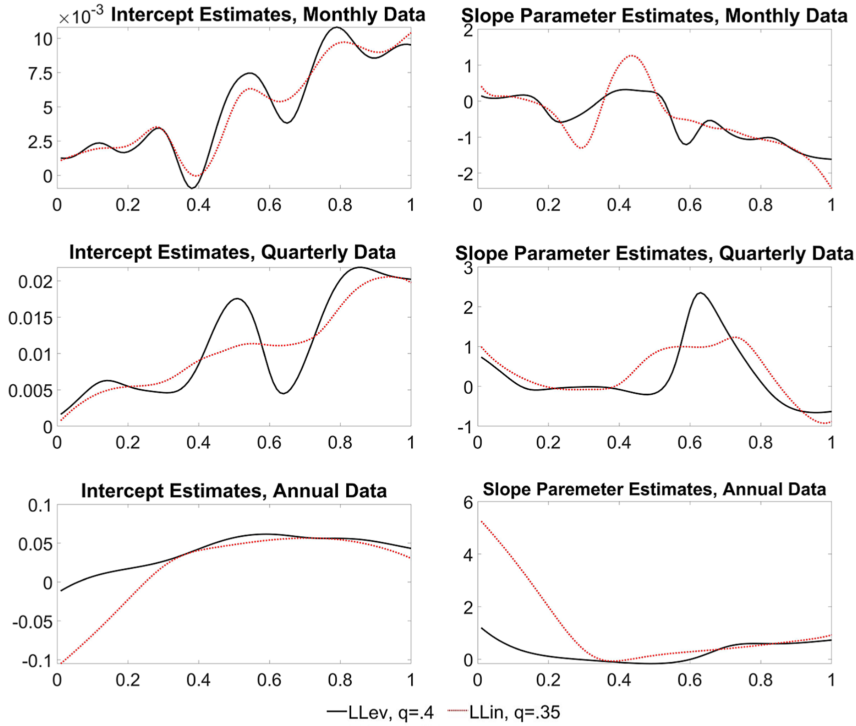

FIGURE 2. Local level/linear TVP estimates.

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 utilize the LLev and LLin tests for the hypothesis $H_0 : \beta(\tau) = 0$. Rolling t-statistics are reported in Figure 3. We emphasize 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 $n = 1,606$ while for quarterly and annual is $n = 535$ and $n = 133$, respectively.

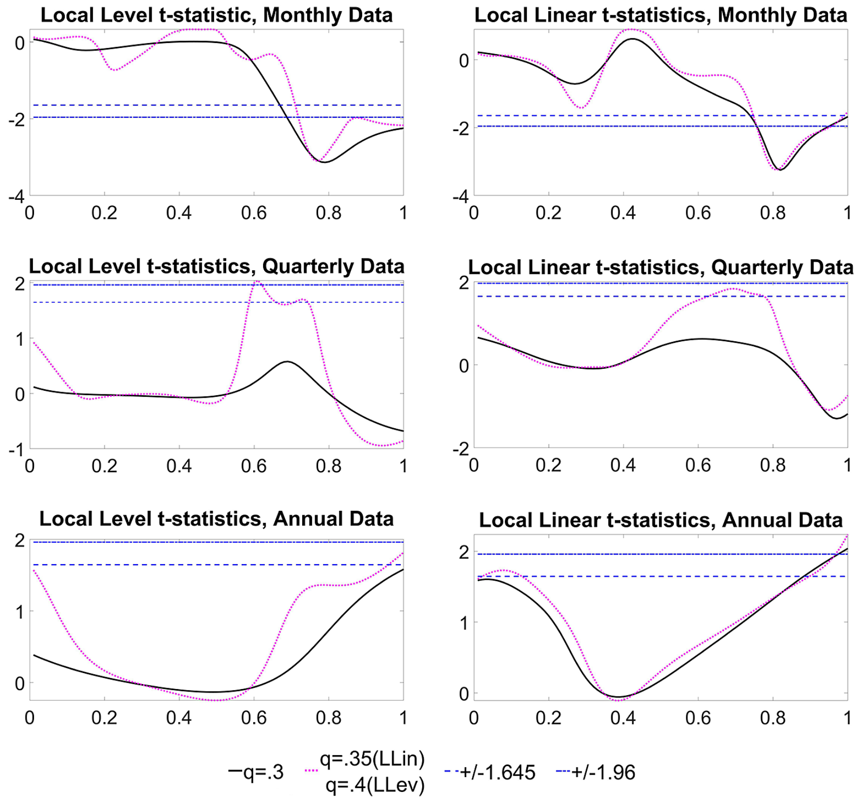

FIGURE 3. Local level/linear t-statistics for $H_0 : \beta(\tau) = 0$.

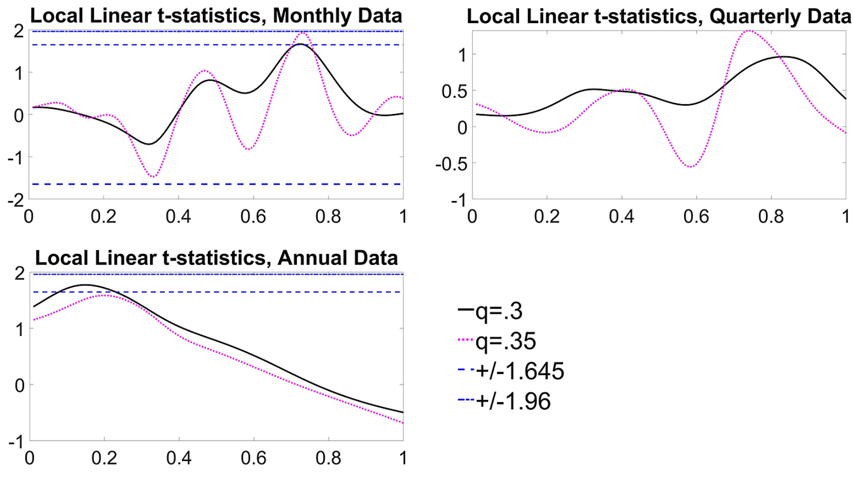

FIGURE 4. Local linear t-statistics for $H_0 : \partial\mu(\tau)/\partial\tau = 0$.

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

Finally, we utilize the LLin t-test for the hypothesis $H_0 : \partial\mu(\tau)/\partial\tau = 0$, that is, 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," for example, 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 "Time-Varying Parameter Regressions with Stationary Persistent Data," Econometric Theory Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466624000082.

REFERENCES

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

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

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

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3), 651–707.

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

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

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

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11), 4463–4492.

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

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods*. Springer.

Chen, W. W., & 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.

Christensen, B. J., & 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.

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

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

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

Duffy, J. A., & 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.

Francq, C., & Zakoian, J.-M. (2010). *GARCH models: Structure, statistical inference and financial applications*. John Wiley & Sons.

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

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.

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

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

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

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

Ibragimov, I. A., & Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing.

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

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

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

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

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

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

Li, Q., & Racine, J. S. (2006). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

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

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

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

Park, J. Y., & Phillips, P. C. B. (2000). Nonstationary binary choice, *Econometrica*, 68(5), 1249–1280.

Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series, *Econometrica*, 69(1), 117–161.

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

https://doi.org/10.1017/S0266466624000082 Published online by Cambridge University Press

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

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

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

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

Robinson, P. M. (1989). Nonparametric estimation of time varying parameters. In P. Hackl (Ed.), *Economic structural change: Analysis and forecasting* (pp. 253–264). Springer.

Robinson, P. M. (1991). Time varying nonlinear regression. In P. Hackl (Ed.), *Economic structural change: Analysis and forecasting* (pp. 179–190). Springer.

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

Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable non-Gaussian random processes*. Chapman & Hall.

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

Shiryaev, A. N. (1996). *Probability*. (2nd ed.) Springer.

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

Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression*. World Scientific.

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

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

Yang, B., Long, W., Peng, L., & 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.