CHRONOLOGICAL TRIMMING METHODS FOR NONLINEAR PREDICTIVE REGRESSIONS WITH PERSISTENT DATA

Zhishui Hu*, Ioannis Kasparis† and Qiying Wang‡

April 24, 2021

Abstract

We develop semi- and non-parametric methods for estimation and inference in nonlinear predictive regressions with persistent predictors. We first consider a semi-parametric method, that we term *Chronologically Trimmed LS (CTLS)*, for inference in regressions where the nature and extent of persistence in the data is uncertain. CTLS attains a sub-OLS convergence rate and has (mixed) Gaussian limit distribution in situations where the data may be weakly or strongly persistent. In particular, we allow for nonlinear predictive type of regressions where the regressor can be stationary short/long memory as well as nonstationary long memory process or a nearly integrated array. The resultant t-tests have conventional limit distributions (i.e. $\mathbf{N}(0,1)$) free of (near to unity and long memory) nuisance parameters. In the case where the regressor is a fractional process, no preliminary estimator for the memory parameter is required. Therefore, the practitioner can conduct inference while being agnostic about the exact dependence structure in the data. The CTLS estimator is obtained by applying certain *chronological trimming* to the OLS instrument via the utilisation of appropriate kernel functions of time trend variables. A specific version of CTLS also yields consistent non-parametric estimates of time varying parameters (TVPs). The resultant non-parametric estimator can be utilised for the development of non-parametric t-tests with conventional limit distributions when predictors are stationary. In particular, we allow for general stationary processes that can be of long memory, ARCH(∞) and in some cases heavy tailed. The finite sample performance of CTLS based t-tests is investigated with the aid of a simulation experiment.

*International Institute of Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China, email: huzs@ustc.edu.cn.

†University of Cyprus, Nicosia, Cyprus, email: kasparis@ucy.ac.cy.

‡School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia, email: qiying@maths.usyd.edu.au.

The proposed methods are employed for investigating the predictability of SP500 returns. We consider four widely used predictors (i.e. dividend yield, earning-to-price and book-to-market ratios, and realised variance) in the context of fixed parameter (FP) and TVP regressions. In the case of FP specifications, we find stronger predictability evidence when nonlinear regressions of reduced growth are employed with dividend yield and book-to-market ratio as predictors. Further, 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.

1 Introduction

Estimation and inference under temporal dependence is a challenging task. An enormous literature in time series econometrics and statistical time series is dedicated to this topic. Despite major advances in this area, relatively little progress has been made towards the development of a comprehensive framework for inference in general models that allow for flexible functional forms and regressors that may exhibit a wide range of persistence.

A major obstacle for a development of this kind has to do with the fact that parametric estimators, under nonstastionarity and mild endogeneity, exhibit drastically different limit distributions than those under stationarity. As a consequence, inferential procedures developed for stationary data are not applicable under nonstationarity and vice versa. A number of early studies in the area of nonstationary econometrics (e.g. Phillips and Hansen, 1990; Johansen, 1995; Phillips, 1995) develop inferential procedures suitable for nonstationary models with $I(1)$ covariates, however these methods not only there are not valid under stationarity, they are also non robust to local deviations from the unit root paradigm. In particular, when there are local or larger deviations from a unit root, nuisance parameters such as memory and near-to-unity feature in estimators' limit distributions, making inference challenging. Near-to-unity parameters are not estimable, rendering various statistical tests non pivotal. On the other hand memory parameters can be estimated in general, however more complicated procedures are required for valid inference.

Despite progress in recent years towards methodologies that partially robustify inference to the persistence properties of the data, a unifying framework for inference that allows for a wide range of persistence in the data and a wide range of model specifications remains elusive. Certain studies in this area develop procedures that provide robust inference in the presence of nearly integrated (NI) processes, in the context of reduced form type of regressions i.e. regressions where the covariate is predetermined with respect to the regression error. For example, Cavanaght, Elliot and Stock (1995), Campbell and Yogo (2006), Janson and Moreira (2006), Elliott, Müller and Watson (2015), develop procedures suitable for parametric models with a NI covariate. The aforementioned papers propose test statistics with limit distributions free of the nuisance near-to-unity parameter. This is achieved mainly[1] via conservative inferential methods e.g. Bonferroni methods or by considering test statistics averaged over a prespecified range for the nuisance parameter space -for a review see Mikusheva (2007) and Phillips (2014, 2015). Although these procedures provide valid inference under local deviations from a unit root, their emphasis is on nearly integrated (NI) models and may not be valid under large deviations from unity (see Phillips 2014). Further, their implementation is more involved than that of conventional tests based on studentised regression estimators (i.e. t-/F-tests). This is due to the fact that the related test statistics can be more complex, but more importantly because limit distributions are not conventional (e.g. $\mathbf{N}(0,1)$, χ^2). Therefore, critical values are not readily available from commonly used statistical tables. The implementation of these methods becomes even more difficult in situations where the dimensionality of the nuisance parameter space increases e.g. when the model involves multiple near unit roots and/or memory parameters (fractional data), tail parameters (heavy tailed data), time varying parameters (TVPs), different types of nonlinearity in the regression function etc.

Another related literature focuses on valid inference in fractionally cointegrated systems e.g. Robinson and Hualde (2003), Christensen and Nielsen (2006), Hualde and Robinson (2010), Andersen and Varneskov (2020). The specifications under consideration are in general structural (i.e. covariates may not be predetermined) and in some cases (e.g. Hualde and Robinson, 2010; Andersen and Varneskov, 2020) both stationary and nonstationary long memory is allowed. These methods are mainly semi-parametric with respect to the short memory components of the system, and may attain sub-OLS convergence rates due to bandwidth parameters. Regression estimators have mixed Gaussian or Gaussian limit distributions and therefore inference is conventional in this framework, nevertheless, preliminary memory estimators are required that makes implementation somewhat more involved. Further, although the specifications are quite general, nonlinearities and nearly integrated arrays are not ruled out. For instance, similarly to FMLS (e.g. Phillips, 1995), the spectral LS method of Robinson and Hualde (2003) relies on (fractionally) differencing the data. It is well known that this approach results in severe size distortions when there are near-unity-parameters.

Nonlinearities can complicate inference in the presence of persistent data further. Park and Phillips (1999, 2001), Chan and Wang (2015) and Hu, Phillips and Wang (2021) study nonstationary regressions with nonlinear regressions of known form, while Wang and Phillips (2009a,b) and Kasparis, Andreou and Phillips (2015), Lin, Tu and Yao (2020) consider fully non-parametric models with regression functions of unknown form. Nonlinearities in the regressions parameters are considered by Phillips, Li and Gao (2017) and Demetrescu, Georgiev, Rodrigues and Taylor (2020). In particular, the latter two studies consider infer-

[1] Janson and Moreira (2006) consider coditional inference rather than conservative tests.

ence in nonstationary regressions with time varying parameters (TVP). Although inference in kernel regressions is conventional (e.g. Wang and Phillips 2009a,b; Kasparis et al. 2015), for parametric models (e.g. Park and Phillips, 2001) and TVP models (Phillips et al., 2017) limit distributions involve nuisance parameters. For instance, Phillips et al. (2017) show that the limit distributions of non-paramertic estimators for TVP parameters in $I(1)$ regressions closely resemble that of OLS, and they propose a FMLS type of method (e.g. Phillips and Hansen, 1990). It is well known that although this approach yields mixed normality when data are exactly $I(1)$, tests exhibit severe size distortions under local deviations from unity. Implementation of inferential procedures in parametric nonlinear regressions can be complicated even if the functional form and the integration order of the data are known. Strong smoothness assumptions may be required for the validity of tests, and limit distributions may depend on the shape of the regression function (see for example Kasparis, 2008; Choi and Saikkonen, 2010). The relevance of nonlinear regression functions in predictive regressions has been emphasised in the recent work of Phillps (2015), among others, who points out that nonlinear regressions may alleviate misbalancing issues between persistent predictors and less persistent returns series. Further, economic theory models (e.g. Lettau and Ludvigson 2001; Menzly, Santos and Veronesi, 2004) suggest that the relationship between returns and predictors such as dividend yield involve time varying parameters. Neglecting nonlinearities in regression with persistent data may lead to substantial adverse effects in inference. For example, misspecifing functional form in regressions with nonstationary data typically leads to diverging or vanishing estimators (see e.g. Kasparis, 2011; Phillips, 2015). In this work we demonstrate that neglecting time variation in "nuisance" regression parameters (e.g. the regression intercept) leads to diverging t-statistics under the null hypothesis when regressors are long memory of order $0 < d < 1/2$.

The current paper develops estimation methods that yield conventional inference in predictive regressions that are nonlinear in variables, with nonlinearities of known form. In this respect we built on the work of Park and Phillps (1999, 2001). In particular, we consider two types of models: a) Models with fixed parameters that allow for a wide range of dependence in the data including stationary or nonstaionary long memory as well as fractionally nearly integrated arrays (e.g. Buchmann and Chan, 2007); b) models with TVPs and a general stationary covariate that can be a long memory linear process, a stationary $ARCH(\infty)$, and in some cases a heavy tailed linear process. Although these two types of specifications are substantially different with respect to the regression parameters, the proposed estimation procedures involve similar instrumentation methods. In fact, the proposed estimation method for TVP models is a special case of the instrumental variables method considered for FP models. For TVP models, the instrumentation methods entail the same type of kernel functionals utilised by Phillips et al. (2017) for nonparametric estimation of TVP functionals in the $I(1)$ case. To some extent, our results for TVP regressions are complementary to those of Phillips et al. (2017) who focus on a different area of the regressor space. Interestingly, although the instrumentation of Phillips et al. (2017) does not lead to conventional free of nuisance parameters inference for TVPs in the nonstationary case, certain generalisation of these instruments does lead to conventional inference for fixed parameter models. The particular instruments are of attenuated signal relative to OLS. This reduction in persistence leads to vanishing asymptotic endogeneity. As a result, estimators for fixed slope parameters have mixed Gaussian limit distributions and therefore conventional inference applies (see e.g. Phillips, 1991).

The proposed methods, that we term Chronologically Trimmed LS (CTLS hereafter), in the FP case share the same underlying principle with the IVX method of Magdalinos and Phillips (2009; MP hereafter). Both CTLS and IVX achieve mixed normal limit distributions via a reduction in the persistence of the instruments. Recent advancements in generalised martingale CLTs reveal that under certain conditions, a reduction in the persistence of instruments may result in vanishing asymptotic endogeneity in large samples.[2] This in turn induces mixed asymptotic normality. The key feature of IVX methods is the utilisation of instruments based on certain linear filtering of the regressors. The so called IVX instruments can be constructed to have arbitrarily weaker signal than that of the OLS instruments, and this is sufficient for martingale CLT to operate. As a result, contrary to the OLS estimator, in the presence of nonstationary data the IVX estimator has mixed Gaussian limit distribution and studentised IVX estimators either $\mathbf{N}(0,1)$ (t-tests) or χ^2 (F-tests) limit distributions. Therefore, conventional and nuisance parameter free inference is achieved for a wide range of persistence in the data at the expense of a slight reduction in the convergence rate. In particular, the IVX estimator attains a sub-OLS convergence rate.[3] MP consider multivariate regressions with mildly and nearly integrated data. The subsequent work of Kostakis, Magdalinos and Stamotogiannis (2015; KMS) extends MP to stationary short memory regressors, and also provides finite sample improvement methods relating to intercept demeaning. Demetrescu et al. (2020) develop predictability tests that allow for TVPs under the alternative hypothesis that utilise IVX and other IV methods. Magdalinos (2020) provides an explicit theory for the properties of IVX methods in predictive regressions with GARCH(p,q) regression errors.

Similarly to IVX, CTLS achieve a reduction in the instruments signal by certain linear filtering of the OLS instruments. In particular, this method entails OLS instruments weighted by integrable kernel functionals of time trend variables. It is well known that kernel methods are local (i.e. they extract information locally), and as a result they are typically associated with slower convergence rates. This is the case for instance in density estimation, kernel regression, long-run variance estimation. Similarly to IVX, CTLS in-

[2]Cf. Jeganathan (2008) and Wang (2014).

[3]By sub-OLS we mean the OLS rate less an arbitrary slow regularly varying rate.

struments can be chosen to yield sub-OLS convergence rates. Both methods can be seen as semi-parametric (for FP models) in terms of generality and convergence rates. In particular, for the implementation of inferential procedures based on CTLS no prior information about the near-to-unity or memory parameters, is required. On the other hand convergence rates are sub-parametric (e.g. slower than OLS) but faster than non-parametric regression (e.g. Nadaraya-Watson estimator; see e.g. Wang and Phillips 2009a,b).

To illustrate how CTLS instrumentation achieves a signal reduction consider the following FP predictive regression

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

where e_k together with some filtration $\mathcal{F}_k$ is a martingale difference error term, and x_k is $\mathcal{F}_k$-measurable. Further, for convenience suppose that $\mu = 0$. In this case CTLS instruments for the estimation of β are arrays Z_{kn} of the form

$$Z_{kn} = K\left[c_n\left(k/n - \tau\right)\right] f(x_{k-1}), \tag{2}$$

where $K > 0$ is an integrable kernel function, $0 < \tau < 1$ and c_n is a reciprocal bandwidth[4] term such that $c_n^{-1} + c_n n^{-1} \to 0$. It can be readily seen that Z_{kn} above utilises the OLS instrument $f(x_{k-1})$ weighted by certain kernel function. The weight function attenuates the signal of the OLS instrument. In particular, due to the integrability of the kernel function, less weight is given to the OLS instrument $f(x_{k-1})$ when the argument of $K(.)$ is away from zero. For example set $\tau = 1/2$ and $K(0) = 1$. In this case the kernel function fully extracts information from the OLS instrument for observations near the middle of the sample i.e. $Z_{kn} \approx f(x_{k-1})$, when $k \approx n/2$. Moreover, $Z_{kn} \approx 0$ when k is far from $n/2$. In other words certain *chronological trimming* applies around the "*chronological point* τ". By allowing the c_n sequence to diverge at an arbitrarily slow rate (i.e. $c_n \to \infty$), the resultant IV (CTLS) estimator attains an arbitrarily slower convergence rate relative to the OLS estimator. To see this note that for f linear and $x_k \sim I(1)$ the CTLS instrument is (using the convention $x_0 = 0$)

$$\sum_{k=1}^{n} \left|K\left[c_n\left(k/n - \tau\right)\right] x_{k-1}\right| = O_p(n^{3/2}c_n^{-1}), \tag{3}$$

while the OLS instrument

$$\sum_{k=1}^{n} |x_{k-1}| = O_p(n^{3/2}). \tag{4}$$

[4]Alternatively we can write

$$Z_{kn} = K\left\{\frac{k/n - \tau}{h_n}\right\} f(x_{k-1}), \text{ with } h_n = c_n^{-1}.$$

The reduction in persistence in (3) relative to (4) is due to the fact that for integrable K, $K(x) \to 0$ as $|x| \to \infty$. This reduction in the signal of the CTLS instruments, translates into vanishing endogeneity and therefore to mixed Gaussian limit distributions. Table 1 provides a comparison between the instrumentation of OLS and other methods that yield mixed Gaussian limit distributions via signal reduction.

As mentioned above, instruments similar to those shown in (2) have been also considered in the recent work of Phillips et al. (2017) who study estimation and inference in linear regressions with TVPs and $I(1)$ regressors. In the context of (1), TVPs can be formulated as $\beta(k/n)$, with $\beta : (0,1] \to \mathbb{R}$. Indeed, due to the presence of the kernel functionals in the instrument of (2), the functional form of $\beta(k/n)$ can be consistently estimated for each *chronological point* τ i.e. $\hat{\beta}(\tau) \to_P \beta(\tau)$, $\tau \in (0,1]$. Nevertheless, in models with multiple parameters and non stationary covariates, the utilisation of kernel functionals with a single *chronological point* τ (cf. (2)) yields estimators with a singular limit variance matrix[5]. Due to this degeneracy, estimators have limit distributions determined by stochastic integrals i.e. limit theory is similar to that of the OLS estimator. Therefore, limit distributions depend on nuisance parameters (e.g. near to unity), and statistical tests are not pivotal. Phillips et al. (2017) propose a FMLS type of modification to get usable statistical tests, which is known to lack robustness under local deviations from unity. Here we demonstrate that this degeneracy does not occur when regressors are stationary (e.g. $I(d)$, with $|d| < 1/2$). Therefore, CTLS based TVP estimators can be used for inference in the stationary case without complications. Nevertheless, CTLS estimators that utilise a single *chronological point* do not yield pivotal tests for $|d| > 1/2$ or in situations where data are NI arrays.

For consistent nonparametric estimation of TVPs, utilising a single *chronological point* τ is crucial and cannot be avoided. Nevertheless, for the estimation of FP models, multiple *chronological points* can be employed. In fact, in the context of FP models, we resolve the degeneracy in limit variance matrix mentioned above by considering a more general class of instruments that involve multiple chronological points (*cps* hereafter). In principle, it is possible to extract information around multiple *cps* of the form $0 < \tau_1 < ... < \tau_{l_n} < 1$, where l_n is either fixed or $l_n \to \infty$ such that $l_n = o(c_n)$. In this case the relevant instrument is

$$Z_{kn} = \sum_{j=1}^{l_n} K\left[c_n\left(k/n - \tau_j\right)\right] f(x_{k-1}). \tag{5}$$

An instrument that utilises multiple *cps* extracts information more evenly over the sample period, locally to each *cp*, in a piecewise fashion (see Figure 1). To see the effects of additional *cps* on the instruments' signal suppose that $x_k \sim I(1)$. In this case the order

[5]See Phillips et al. (2017) and Remark 12 below.

of magnitude of $\sum_{k=1}^{n}|Z_{kn}|$ in (5) for linear f is $O_p(c_n^{-1}l_n n)$.[6] If the number of *cps* is too large, i.e. $l_n = O(c_n)$, CTLS is asymptotically equivalent to OLS. However, under the requirement $c_n \to \infty$, $l_n = o(c_n)$, Z_{kn} exhibits a weaker signal than the OLS instrument $f(x_{k-1})$, and as a result asymptotic mixed Gaussianity applies. Figure 1 provides plots of the weights/trimming functionals $\sum_{j=1}^{l_n} K\left[c_n\left(k/n-\tau_j\right)\right]$ against k/n with $k = 1, ..., n$. Note that the weight function in OLS instrumentation in unity. From Figure 1 we can deduce the information utilised by various CTLS instruments relative to that by OLS. The total information extracted by CTLS instrumentation over the whole sample period corresponds to the dark area whilst the total information conveyed by OLS instrumentation is unity (i.e. the sum of the dark and grey area).

It should be emphasised that utilising multiple *cps*, in the estimation of FP models, is crucial for avoiding a singular limit variance matrix in situations where data are non stationary and multiple parameters need to be estimated. In particular, the number of *cps* needs to be at least the same as the number of the regression parameters. A non singular variance matrix in turn ensures the limit distribution is mixed Gaussian. In the current work we consider models with a single predictor and an intercept, therefore at least two *cps* will be required (i.e. $l_n \geq 2$). To ensure that the CTLS estimator has mixed Gaussian limit distribution, the number of *cps* must be sufficient to avoid a singularity in the limit variance matrix, but not too many in order to allow for a martingale CLT to operate i.e. $l_n = o(c_n)$. Overall, the following requirement is imposed on c_n, l_n

$$\left(1+l_n\right)c_n^{-1} + c_n n^{-1} \to 0, \tag{6}$$

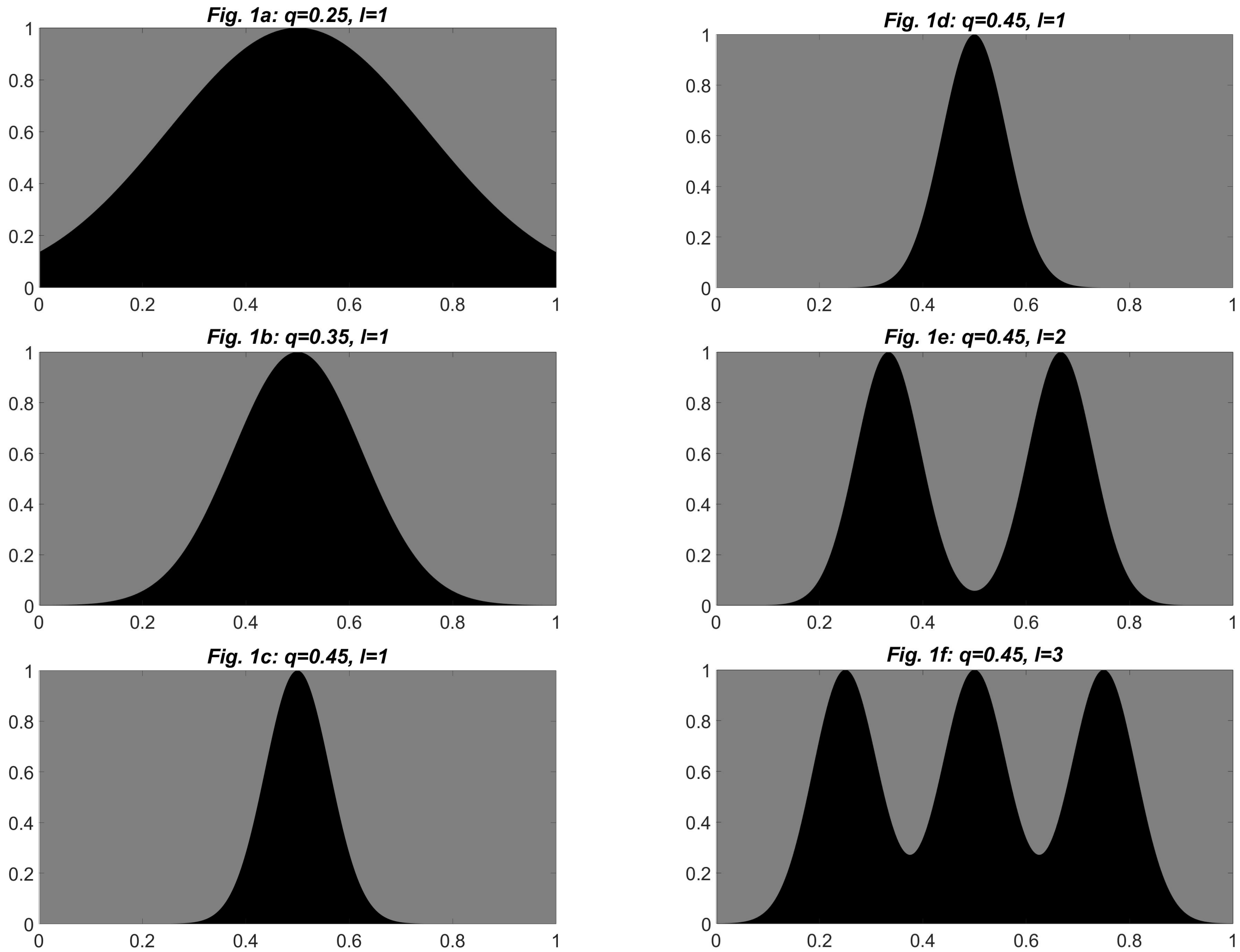
with the additional restriction $l_n \geq$ *# regression parameters*, for FP models. This requirement is similar to Assumption T of Andersen and Varneskov (2020) how also consider semi-parametric methods with dual bandwidth terms. It is readily obvious from (6) that the sequences c_n and l_n are independent of the regression covariates.

In general, there is trade-off between size and power when it comes to the choice of c_n and l_n. It follows from the above that smaller c_n (and larger l_n) leads to CTLS estimators that resemble the asymptotic behaviour of the OLS estimator which attains better convergence rates but has non conventional limit distribution. Better size control is achieved when c_n is large and l_n small at the expense of slower convergence rate and less powerful tests. This trade-off is illustrated diagrammatically in Figure 1. Note for smaller values of c_n, more information is extracted i.e. CTLS instruments exhibit stronger signal. More information

[6] In view of (3), it can be readily seen that if multiple *cps* are utilised,

$$\sum_{k=1}^{n} K\left[c_n\left(\frac{k}{n}-\tau_1\right)\right]x_k + ... + \sum_{k=1}^{n} K\left[c_n\left(\frac{k}{n}-\tau_{l_n}\right)\right]x_k = l_n \cdot O_P(n^{3/2}c_n^{-1}).$$

Figure 1: Signal of CTLS Instruments Vs OLS



The total signal of the CTLS instruments equals the dark area;
The total signal of the OLS instruments equals the dark + grey area;
$c_n = n^q$, $q = \{0.25, 0.35, 0.45\}$; $l = \{1, 2, 3\}$; $l =$ number of equispaced *cps*.

can be also extracted by utilising additional *cps*. In situations where the dark area is large CTLS estimators attain faster convergence rates at the expense of worse size control and vice versa.

It can be readily seen from Figure 1, that the same information (in terms of area) can be extracted by different combinations of c_n and l_n. Developing methods that yield optimal bandwidth selection for inferential purposes is very challenging from a technical point of view. For instance Sun, Phillips and Sainan (2008) develop optimal bandwidth choice methods for HAC studentisation of t-statistics, where the optimality criterion is based on type I and type II errors. Despite the fact that the aforementioned work considers only simple location problems, i.e. regressions that involve a single unknown parameter corresponding to an intercept, the technical exposition is substantial. Developing similar techniques for optimally choosing c_n and l_n is even more difficult given the complexity of our theoretical framework. Instead we recommend values for c_n and l_n based on simulations. Simulations results show that a better size-power trade-off is achieved when $l_n \to \infty$. Note that in general, a larger number of *cps* leads to a more even extraction of information over the sample period -see for instance Figure 1(d)-(f).

Table 1: Strength of various instruments and convergence rates for eq. (1) ($I(1)$ data; $\mu = 0$, known).

Estimation Method	Parameter of Interest	Instrument Z_{kn} ($x_k \sim I(1)$)	$\sum_{k=1}^{n} \lvert Z_{kn} \rvert$ Order of Magnitude	Mixed Normality	Estimator's Convergence Rate
OLS	β (f linear)	x_{k-1}	$O_p(n^{3/2})$	No	n (parametric)
IVX	β (f linear)	$\sum_{j=0}^{k-1} \left(1 + \frac{c_z}{n^b}\right)^j (x_{k-1-j} - x_{k-2-j})$, $c_z < 0$, $b \in (0,1)$	$O_p(n^{1+b/2})$	Yes	$n^{1/2+b/2}$ (semi-parametric)
Nadaraya-Watson	$\beta \cdot f(.)$	$K\left(\frac{x_{k-1}-x}{h_n}\right)$, K integrable, $h_n + h_n^{-1} n^{-1/2} \to 0$	$O_p(h_n n^{1/2})$	Yes	$h_n^{1/2} n^{1/4}$ (non-parametric)
CTLS (single trimming point)	β (f linear)	$K\left(\frac{k/n-\tau}{h_n}\right) x_{k-1}$, K integrable, $h_n + h_n^{-1} n^{-1} \to 0$, $\tau \in (0,1)$	$O_p(h_n n^{3/2})$	Yes	$h_n^{1/2} n$ (semi-parametric)
CTLS (mutliple trimming points)	β (f linear)	$\sum_{j=1}^{l_n} K\left(\frac{k/n-\tau_j}{h_n}\right) x_{k-1}$, K integrable, $(1 + l_n) h_n + h_n^{-1} n^{-1} \to 0$, $\{\tau_j\} \in (0,1)$	$O_p(l_n h_n n^{3/2})$	Yes	$(l_n h_n)^{1/2} n$ (semi-parametric)

$h_n = c_n^{-1}$

In Section 3 we explore the properties of the CTLS estimator and related t-tests in nonlinear regressions like the one given in eq. (1) with e_t being a martingale difference error possibly conditionally heteroscedastic e.g. $ARCH(\infty)$, and (μ, β) either FP or TVPs of the form $(\mu, \beta) : (0,1]^2 \to \mathbb{R}^2$. In particular, for fixed slope parameters, we consider CTLS based inference with multiple *cps* ($l_n \geq 2$) and the predictive variable can either be a stationary or nonstationary fractional processes (e.g. $I(d)$, $-1/2 < d < 3/2$) or a NI fractional array. For TVP models we consider CTLS with a single *cp* and stationary predictors. In all cases the CTLS estimator attains a sub-OLS convergence rate and has a (mixed) Gaussian limit distribution. To summarise the limit properties of CTLS suppose for the sake of simplicity that f is linear and $\mu = 0$ (known). In this case the CTLS estimator is of the form $\hat{\beta} = \sum_{k=1}^{n} Z_{kn} y_k / \sum_{k=1}^{n} Z_{kn} x_{k-1}$ for both FP and TVP models. For the former first suppose that x_k is a nonstationary process such that for $t \in [0,1]$ and some $d_n \to \infty$, $d_n^{-1} x_{\lfloor nt \rfloor} \Rightarrow X_t$ in $D[0,1]$ where X_t is a continuous process (i.e. an FCLT holds). For example, X_t can be a fractional BM or a fractional Ornstein-Uhlenbeck process, depending on some memory and/or some near-to-unity nuisance parameter that are unknown. Then for c_n and l_n as in (6), the CTLS estimator for β satisfies

$$d_n \sqrt{\frac{n l_n}{c_n}} \left(\hat{\beta} - \beta \right) \to_d \mathbf{MN} \left(0, \frac{E\left(e_1^2\right) \int_{\mathbb{R}} K^2(x) dx}{\left(\int_{\mathbb{R}} K(x) dx \right)^2 \frac{1}{l_n} \sum_{j=1}^{l_n} X_{\tau_j}^2} \right),$$

with $\frac{1}{l_n} \sum_{j=1}^{l_n} X_{\tau_j}^2 \equiv \int_0^1 X_t^2 dt$, if l_n is not fixed and diverging. Because $c_n \to \infty$, $c_n = o(n)$ and $l_n = o(c_n)$, the convergence rate of the CTLS is slower than that of the OLS estimator ($d_n \sqrt{n}$). Nuisance parameters (e.g. the near-to-unity, fractional parameters) affect the limit distribution only via the mixing variate $\left[\frac{1}{l_n} \sum_{j=1}^{l_n} X_{\tau_j}^2 \right]^{-1}$ and as a consequence the studentised CTLS estimator has a $\mathbf{N}(0,1)$ limit distribution. Note that the limit process X_t is allowed to be a fractional Ornstein–Uhlenbeck process that depends on dual nuisance parameters i.e.

$$X_t = \int_0^t e^{c(t-r)} dW_d(r),$$

with $c \in \mathbb{R}$ being a near to unity parameter, and $W_d(r)$, $d > 1/2$, a fractional Brownian motion (see e.g. Bunchmann and Chan, 2007). For stationary x_k (e.g. $I(d)$, $|d| < 1/2$) and FP models we have

$$\sqrt{\frac{n l_n}{c_n}} \left(\hat{\beta} - \beta \right) \to_d \mathbf{N} \left(0, \frac{E\left(x_1^2 e_2^2\right) \int_{\mathbb{R}} K^2(x) dx}{\left(\int_{\mathbb{R}} K(x) dx \right)^2 \left[E\left(x_1^2\right) \right]^2} \right),$$

It should be emphasised that the restrictions on sequences c_n and l_n of eq. (6) are independent of x_k and therefore practitioners can implement the method while being agnostic about the persistence level in the data.

For TVP models we consider estimators and related nonparametric t-statistics for TVPs and their derivatives. Derivative estimators are useful for testing time invariance hypotheses of the form $H_0 : \partial\mu(\tau)/\partial\tau = 0$ and $H_0 : \partial\beta(\tau)/\partial\tau = 0$. To illustrate the limit theory for TVP models, suppose that (1) holds with $\mu = 0$ (known) and slope parameter $\beta(k/n)$. The for a stationary regressor the TVP estimator (single cp τ) is

$$\sqrt{\frac{n}{c_n}}\left(\hat{\beta}(\tau) - \beta(\tau)\right) \to_d \mathbf{N}\left(0, \frac{E\left(x_1^2 e_2^2\right)\int_{\mathbb{R}} K^2(x)dx}{\left(\int_{\mathbb{R}} K(x)dx\right)^2 \left[E\left(x_1^2\right)\right]^2}\right),$$

for each $\tau \in (0, 1]$. Further, the TVP estimator for the derivative of the slope parameter satisfies a limit result of the form

$$\sqrt{\frac{n}{c_n^3}}\left(\frac{\partial\hat{\beta}(\tau)}{\partial\tau} - \frac{\partial\beta(\tau)}{\partial\tau}\right) \to_d \mathbf{N}\left(0, \vartheta\right),$$

for some $\vartheta > 0$ that can be consistently estimated. Both estimators can be utilised for construction of non parametric tests. The derivative estimator attains a slower convergence rate[7], and therefore yields less powerful tests, nevertheless, the implementation of time invariance tests is very easy and can be done in conjunction with tests for the predictability hypothesis.

Before proceeding to the next section, we provide some further discussion about the CTLS, for FP models, relative to IVX. First, as explained earlier, both estimators enjoy (mixed) Gaussian distribution by utilising instruments of reduced signal, and in this sense they belong to the same class of estimators. The treatment of the regression intercept is however different and this has some practical consequences in the implementation of tests. IVX is utilising conventional demeaning for the intercept but requires additional studentisation based on long run variance estimators (see Kostakis et al., 2015 for more details). One the other hand CTLS requires CTLS-type of instrumentation for the intercept as well. MP show that IVX can accommodate NI, MI covariates while the most recent work of Kostakis et al. (2015) extends the method to stationary short memory processes. Further, some preliminary theoretical results suggest[8] that IVX, probably after some minor modification, is also valid for fractional processes. For FP models we only consider univariate regressions. From this point of view our framework is less general than Kostakis et al. (2015). Nevertheless, our theoretical framework readily allows for fractional predictors and nonlinear regression functions. The current results could be generalised to multivariate regressions but this would require more complicated limit theory. We leave an extension to this direction for future work.[9] In terms of implementation, both methods appear to be of similar complexity.

[7]This result is standard in for non-parametric derivative estimators see e.g. Li and Racine (2006).
[8]See Theorem 3.2 in Duffy and Kasparis (2018).
[9]An extention of TVP methods to multi-covariate regressions is provided in the Appendix.

Both procedures employ studentised estimators. CTLS involves an additional bandwidth type of parameter (l_n) that determines the number *cps*, however it does not require long-run variance estimators that also involve bandwidth terms.

The remaining of this work is organised as follows. Section 2 provides basic limit theory for chronologically trimmed functionals of stationary and nonstationary processes. This limit theory is utilised in Section 3 for the development of estimation and inferential procedures for predictive regressions with persistent predictors. In particular, Section 3.1 considers nonlinear predictive regressions with fixed parameters and general predictors that can be either stationary or nonstationary, whilst Section 3.2 focuses on nonlinear TVP predictive regressions with stationary predictors. Section 3.3 provides some discussion about the consequences of ignoring time variation in regression parameters. A simulation study for the methods of Section 3.1 and 3.2. is reported in Section 4. Finally, empirical applications to the predictability of stock returns is the subject of Section 5. All proofs are provided in the Appendix (Sections 6-8).

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. We may write the integral $\int_{\mathbb{R}} f(x)dx$ as $\int f$. $\Rightarrow$ denotes weak convergence in the space $D[0,1]$. For a vector x, $\|x\|$ is its inner product norm and x' its transpose. By $[x]$ we denote the integer part of a positive number x. Finally, $\text{diag}\{a_1, ..., a_p\}$ denotes a $p \times p$ diagonal matrix with elements $\{a_1, ..., a_p\}$ on the main diagonal, $\to_d$ denotes the convergence in distribution and $Y := \mathbf{MN}(\mathbf{0}, \Sigma)$ denotes a Gaussian variate (mixing normal) with characteristic function $\psi(t) = Ee^{it'Y} = Ee^{-t'\Sigma t/2}$.

2 Asymptotics for Chronologically Trimmed Sample Functionals

This section develops basic limit theory for chronologically trimmed (CT hereafter) sample functionals of stationary and nonstationary processes. Let $\{x_k\}_{1\le k\le n}$ be a scalar time series process and $\{X_{nk}\}_{1\le k\le n, n\ge 1}$ be some scalar random array. Further, let K be an integrable kernel function and $g(.) = [g_1(.), ..., g_p(.)]'$, where, for each $i = 1, ..., p$, g_i is a measurable function. For $l \in \mathbb{N}$, $0 < \tau_1 < ... < \tau_l < 1$ and $m = 0, 1$ or 2, set

$$\begin{aligned} S_{1n,l}^{(m)} &= \frac{c_n}{n}\sum_{k=1}^{n} g(x_{k-1})\sigma_k^m \left\{\frac{1}{l}\sum_{j=1}^{l} K\left[c_n(k/n - \tau_j)\right]\right\}, \\ M_{1n,l} &= \sqrt{\frac{c_n}{n}}\sum_{k=1}^{n} g(x_{k-1})\left\{\frac{1}{\sqrt{l}}\sum_{j=1}^{l} K\left[c_n(k/n - \tau_j)\right]\right\}\sigma_k u_k, \end{aligned}$$

$$S_{2n,l}^{(m)} = \frac{c_n}{n}\sum_{k=1}^{n} g(X_{n,k-1})\sigma_k^m \left\{\frac{1}{l}\sum_{j=1}^{l} K\left[c_n(k/n-\tau_j)\right]\right\},$$
$$M_{2n,l} = \sqrt{\frac{c_n}{n}}\sum_{k=1}^{n} g(X_{n,k-1})\left\{\frac{1}{\sqrt{l}}\sum_{j=1}^{l} K\left[c_n(k/n-\tau_j)\right]\right\}\sigma_k u_k,$$

where c_n is a sequence of positive constants, l either fixed or $l\to\infty$ as $n\to\infty$, and u_k together with an appropriate filtration $\{\mathcal{F}_k\}$ forms a martingale difference sequence so that X_{nk}, x_k are $\mathcal{F}_k$-measurable and σ_k is $\mathcal{F}_{k-1}$-measurable.

The asymptotics of $\left\{S_{jn,l}^{(m)}, M_{jn,l}\right\}_{j=1}^{2}$ are utilised in Section 3 for the asymptotic analysis of the CTLS estimators. In particular, limit theory for the functionals $\left\{S_{1n,l}^{(m)}, M_{1n,l}\right\}$ is relevant for stationary regressors whilst $\left\{S_{2n,l}^{(m)}, M_{2n,l}\right\}$ for nonstationary. Indeed, it is assumed that x_k is a stationary random sequence, but X_{nk} satisfies some functional central limit theorems (FCLT). σ_k is set to be strictly stationary so that the asymptotics of $\left\{S_{jn,l}^{(m)}, M_{jn,l}\right\}_{j=1}^{2}$ are applicable when the regression errors exhibit conditional heteroscedasticity (e.g. GARCH, ARCH(∞) etc). It should be mentioned that the term $S_{2n,l}^{(0)}$ resembles certain functionals considered by Phillips, Li and Gao (2017) who study the estimation of cointegrated models with smooth time varying parameters (TVP). The aforementioned work utilises statistics of the form

$$\frac{c_n}{n}\sum_{k=1}^{n} X_{nk}^2 K\left[c_n(k/n-\tau)\right],\ 0<\tau<1,$$

where X_{nk} is an $I(1)$ process normalised by $\sqrt{n}$. Under our assumptions, X_{nk} can be an appropriately normalised $I(d)$, $d>1/2$, process or a NI array (possibly driven by fractional errors). Therefore, the limit results provided in this section are also relevant to the estimation of TVP models for the case where the covariate is either a stationary process or a general nonstationary process satisfying some FCLT as set in Assumption **A3** below.

To facilitate basic limit results, we make use of the following conditions.

A1 (innovations): $\{\eta_k, \mathcal{F}_k\}_{k\geq 1}$, where $\eta_k' = (\xi_k, u_k)$ and $\mathcal{F}_k = \sigma(u_k, u_{k-1}, ..., u_1; \xi_j, j\leq k)$, forms a 2-dimensional 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) $\sup_{k\geq 1} E(\xi_k^2 I(|\xi_k|\geq M)|\mathcal{F}_{k-1}) = o_P(1)$, as $M\to\infty$;

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

A2 (stationary process): x_k is a functional of $\xi_k, \xi_{k-1}, ...$ and σ_k is adapted to $\mathcal{F}_{k-1}$, where $\mathcal{F}_k$ is defined as in **A1**, so that $g(x_{k-1})\sigma_k^m$ ($m=0,1$ or 2, respectively) is an ergodic (strictly) stationary random sequence with $E\big\{\|g(x_1)\| + \sigma_2^2\big[1+\|g(x_1)\|^2\big]\big\}<\infty$.

A3 (nonstationary process and invariance principle):

(a) $X_{nk} = d_n^{-1} x_k$, where $0 < d_n^2 = var(x_n) \to \infty$ and x_k is a functional of $\xi_k, \xi_{k-1}, ...$ (depending on n is allowed) so that, on $D_{\mathbb{R}^3}[0,1]$,

$$\left[\frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \xi_k, \; \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \xi_{-k}, \; X_{n,[nt]} \right] \Rightarrow [B_{1t}, B_{2t}, X_t] , \tag{7}$$

where B_{1t} and B_{2t} are two independent Gaussian process with mean zero and stationary independent increments, and X_t is a continuous process that depends only on functionals of $\{B_{1t}\}_{0 \le t \le 1}$ and $\{B_{2t}\}_{0 \le t \le 1}$;

(b) σ_k is adapted to $\mathcal{F}_{k-1}$ and a sequence of ergodic (strictly) stationary variables satisfying $E\sigma_1^4 < \infty$, where $\mathcal{F}_k$ is defined as in **A1**.

A4 (kernel function and restrictions on τ_j, l_n and c_n):

(a) $K(x)$ is a positive real function having a compact support with $0 < \int K < \infty$;

(b) $0 < c_n \to \infty$ and $c_n/n \to 0$;

(c) $\tau_j = j/(l_n + 1)$ where $j = 1, ..., l_n$ with $c_n^{-1} l_n + l_n^{-1} \to 0$.

We remark that the innovation process $\{\eta_k, \mathcal{F}_k\}_{k \ge 1}$ used in **A1** is standard in literature so that both $M_{1n,l}$ and $M_{2n,l}$ have a martingale structure. The uniform integrability conditions (a) and (b) are weak in comparison with the high moments used in previous works. See, for instance, Wang (2014) and Wang and Phillips (2009a, b). In **A1**(c), we impose $E(u_k^2|\mathcal{F}_{k-1}) = 1$ for the convenience of notation. In fact, if $\sigma_{1k}^2 := E(u_k^2|\mathcal{F}_{k-1}) \neq 1$, it is routine to see that our results still hold when σ_k is replaced by $\sigma_k \sigma_{1k}$. Examples of processes that satisfy **A2** include short and long memory (fractional) processes. [10] Indeed, when x_k and σ_k are (strictly) stationary relying on $\xi_k, \xi_{k-1}, ...,$ we also have $g(x_{k-1})\sigma_k^m$ is an ergodic (strictly) stationary random sequence. Typical examples on nonstationary processes satisfying **A3**(a) have the form:

$$x_k = \rho_n x_{k-1} + w_k,$$

where $\rho_n = 1 + \kappa/n$ with $\kappa \in \mathbb{R}$ and w_k being a stationary linear process, possibly of long memory, with innovations ξ_k). Under some additional regularity conditions, (7) holds with X_t being a possibly fractional Ornstein-Uhlenbeck process e.g. Buchmann and Chan (2007), Wang and Phillips (2009a, b) and Wang (2015). We assume strictly stationary for σ_k of **A3**(b) to avoid the complicity in notation. It is readily seen that **A3**(b) allows for strictly stationary GARCH, ARCH(∞) models (e.g. Francq and Zakonian, 2010; Section 2.2). The

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

compact support requirement of **A4**, for the kernel function $K(x)$, can be relaxed under some additional conditions on l_n as follows:

A4* (kernel function and restrictions on τ_j, l_n and c_n):

(a) $K(x)$ is a bounded positive and eventually monotonic real function[11] with $0 < \int K < \infty$;

(b) $0 < c_n \to \infty$ and $c_n/n \to 0$;

(c) $\tau_j = j/(l_n+1)$ where $j = 1, ..., l_n$ with $c_n^{-1} l_n \log n + l_n^{-1} \to 0$.

We next present the limit theory for CT sample functionals. Limit results for stationary and nonstationary functionals are given by Theorem 1 and Theorem 2 respectively.

Theorem 1. *Suppose **A2** and **A4** or **A4*** hold. Then, as $n \to \infty$, we have*

$$S_{1n,l_n}^{(m)} = E[\sigma_2^m g(x_1)] \int K + o_P(1). \tag{8}$$

*If in addition **A1** holds, then, as $n \to \infty$,*

$$M_{1n,l_n} \to_d \mathbf{N}\left(\mathbf{0},\ E\left[\sigma_2^2 g(x_1) g(x_1)'\right] \int K^2\right). \tag{9}$$

Theorem 2. *Suppose that **A3** and **A4** or **A4*** hold and $g(.)$ is continuous. Then, as $n \to \infty$, we have*

$$S_{2n,l_n}^{(m)} = E\sigma_1^m \int_0^1 g(X_{n,\lfloor nt \rfloor}) dt \int K + o_P(1) \to_d E\sigma_1^m \int_0^1 g(X_t) dt \int K. \tag{10}$$

*If in addition **A1**, jointly with (10), we have*

$$M_{2n,l_n} \to_d \mathbf{MN}\left(\mathbf{0},\ E\sigma_1^2 \int_0^1 g(X_t) g(X_t)' dt \int K^2\right). \tag{11}$$

Remark 1. Theorem 1 can be trivially extended to the case where x_k is a p-dimensional process and $g : \mathbb{R}^p \to \mathbb{R}^p$, . We omit the details since this only involves some routine calculations. Theorem 2 could be also generalised to multivariate nonstationary processes at the expense of somewhat more involved exposition. We leave the latter extension for future work.

Remark 2. If we are only interested in limit results for the functionals $S_{1n,l_n}^{(m)}$ and $S_{2n,l_n}^{(m)}$, conditions **A2** and **A3** can be reduced. For instance, the result (10) still holds if only (7) is replaced by $X_{n,[nt]} \Rightarrow X_t$ on $D_{\mathbb{R}}[0,1]$. A unified general result is presented in Lemma 1

[11] i.e., there exists an $A_1 > 0$ such that $K(x)$ is monotonic on $(-\infty, -A_1)$ and (A_1, ∞).

of Section 4. Furthermore, if x_k is a weakly nonstationary process (i.e., $I(1/2)$ and mildly integrated processes, where FCLTs do not apply) as considered in Phillips and Magdalinos (2007) and Duffy and Kasparis (2021), some preliminary calculations suggest (see also Theorem 3.2 in Duffy and Kasparis, 2018) that

$$\frac{c_n}{n}\sum_{k=1}^{n} g\big(d_n^{-1}x_k\big)\Big\{\frac{1}{l_n}\sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right]\Big\} \to_d \int_{\mathbb{R}} g(x+X^-)\varphi_{\sigma_+^2}(x)dx\int K,$$

where $\varphi_{\sigma_+^2}(x)$ is the density of a $\mathbf{N}\left(0,\sigma_+^2\right)$ variate ($\sigma_+^2>0$) and $X^-\sim\mathbf{N}\left(0,\sigma_-^2\right)$ ($\sigma_-^2\geq 0$). Discussions toward this kind of generalization are left for future work.

Remark 3. The continuity requirement on $g(x)$ in Theorem 2 is not essential for (10) and (11). These results can be extended to the case where g is locally Lebesgue integrable, if we impose more smoothness conditions on X_{nk}. This kind of generalisation involves more complicated derivations and will not be pursued here in order to keep the paper under reasonable length.

Remark 4. Following the proof of Theorem 1, it is easy to see that results (8) and (9) still hold if **A4** (c) or **A4*** (c) is replaced by $\tau_j=j/(l+1)$ where $j=1,...,l$, i.e., if $l_n\equiv l$ is fixed. As for (10) and (11), if **A4** (c) or **A4*** (c) is replaced by $\tau_j=j/(l+1)$ where $j=1,...,l$, we have

$$\left[S_{2n,l}^{(m)},\ M_{2n,l}\right]\ \to_d\ \left[\frac{E\sigma_1^m}{l}\sum_{j=1}^{l} g(X_{\tau_j})\int K,\ \ \mathbf{MN}\Big(\mathbf{0},\ \frac{E\sigma_1^2}{l}\sum_{j=1}^{l} g(X_{\tau_j})g(X_{\tau_j})'\int K^2\Big)\right].$$

Theorem 2 provides a limit theory for rescaled functionals of nonstationary processes (i.e. $X_{nk}=d_n^{-1}x_k$ as given in **A3**). For the purposes of regression analysis, limit theory for non rescaled processes (i.e., X_{nk} is replaced by x_k) is more relevant. Following Park and Phillips (1999, 2001), we assume that the function $g(.)=[g_1(.),...,g_p(.)]'$ is asymptotically homogeneous, i.e. for large λ

$$g_i(\lambda x)\approx\pi_i(\lambda)H_i(x),\ i=1,...,p$$

where π_i (positive real valued function) is the "asymptotic order" of g_i and H_i is the "asymptotic homogeneous function" of g_i that is assumed continuous. Several specifications of interest satisfy this conditions e.g. polynomial functions, logarithmic, indicator functions and distribution type of functions e.g. see Park and Phillips (2001) for more details. Set $\pi(.):=\text{diag}\{\pi_1(.),...,\pi_p(.)\}$ and $H(.)=[H_1(.),...,H_p(.)]'$. The following result is the counterpart of Theorem 2 for additive transformations of non rescaled sequences.

Theorem 3. *Suppose that:*

(*a*) ***A1, A3** and **A4** or **A4*** hold;*

(b) for each $i = 1, .., p$, there exist a continuous function H_i on $\mathbb{R}$ and a real function $\pi_i : (0, \infty) \to (0, \infty)$ so that

$$g_i(\lambda x) = \pi_i(\lambda) H_i(x) + R_i(\lambda, x),$$

where $|R_i(\lambda, x)| \le a_i(\lambda)(1 + |x|^\delta)$ for some $\delta > 0$ and $a_i(\lambda)/\pi_i(\lambda) \to 0$, as $\lambda \to \infty$.

Then, as $n \to \infty$, we have

$$\begin{aligned}
&\sum_{k=1}^{n} \pi\,(d_n)^{-1}\, g(x_{k-1}) \Big\{ \sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right] \Big\} \left[\frac{c_n}{nl_n}\, \sigma_k^m\,,\ \sqrt{\frac{c_n}{nl_n}}\, \sigma_k\, u_k \right] \\
= \;\; & \sum_{k=1}^{n} H(X_{n,k-1}) \Big\{ \sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right] \Big\} \left[\frac{c_n}{nl_n}\, \sigma_k^m\,,\ \sqrt{\frac{c_n}{nl_n}}\, \sigma_k\, u_k \right] + o_P(1) \qquad (12) \\
\to_d \;\; & \Big[E\sigma_1^m \int_0^1 H(X_t)dt \int K,\ \mathbf{MN}\Big(\mathbf{0},\ E\sigma_1^2 \int_0^1 H(X_t)H(X_t)'dt \int K^2\Big)\Big]. \qquad (13)
\end{aligned}$$

Remark 5. As in Remark 4, if **A4** (c) or **A4*** (c) is replaced by $\tau_j = j/(l+1)$ where $j = 1, ..., l$, we have

$$\begin{aligned}
&\sum_{k=1}^{n} \pi\,(d_n)^{-1}\, g(x_{k-1})\, \sigma_k \Big\{ \sum_{j=1}^{l} K\left[c_n(k/n - \tau_j)\right] \Big\} \left[\frac{c_n}{nl}\, \sigma_k^m,\ \sqrt{\frac{c_n}{nl}}\, \sigma_k\, u_k \right] \\
\to_d \;\; & \left[\frac{E\sigma_1^m}{l} \sum_{j=1}^{l} H(X_{\tau_j}) \int K,\ \ \mathbf{MN}\Big(\mathbf{0},\ \frac{E\sigma_1^2}{l} \sum_{j=1}^{l} H(X_{\tau_j}) H(X_{\tau_j})' \int K^2\Big) \right].
\end{aligned}$$

Furthermore, if K^* is a real function satisfying **A4** (a) or **A4*** (a), similar arguments as in the proof of Theorems 2 and 3 show that, under the conditions of Theorem 3 with $g(.) = [g_1(.), g_2(.)]'$,

$$\Big(\int_0^1 H_1(X_{n,[nt]})dt,\ U_{1n},\ U_{2n} \Big) \ \to_d \ \Big(\int_0^1 H_1(X_t)dt,\ \sqrt{E\sigma_1^2}\, \mathbf{MN}\,(\mathbf{0},\ V) \Big), \qquad (14)$$

where

$$\begin{aligned}
U_{1n} &= \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \pi_2\,(d_n)^{-1}\, g_2(x_{k-1})\, \sigma_k \left\{ \frac{1}{\sqrt{l_n}} \sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right] \right\} u_k, \\
U_{2n} &= \sqrt{\frac{c_n}{n}} \sum_{k=1}^{n} \sigma_k \left\{ \frac{1}{\sqrt{l_n}} \sum_{j=1}^{l_n} K^*\left[c_n(k/n - \tau_j)\right] \right\} u_k, \\
V &= \begin{bmatrix} \int_0^1 H_2^2\,(X_t)\,dt \int K^2 & \int_0^1 H_2\,(X_t)\,dt\, \int KK^* \\ \int_0^1 H_2\,(X_t)\,dt\, \int KK^* & \int (K^*)^2 \end{bmatrix}.
\end{aligned}$$

The limit result of (14), together with Theorems 1-3, will be utilised in Section 3 next.

3 CTLS Estimation and Inference in Predictive Regressions with Persistent Data

This section utilises the limit theory presented in Section 2 for the asymptotic analysis of the CTLS estimator and a related t-statistics in predictive regressions. We first develop methods for a general CTLS estimator that is utilising multiple *cps* that is relevant for inference in FP models with a general covariate (i.e. stationary or nonstationary). Subsequently, we show that certain versions of this estimator that employ a single *cp* can be used for non-parametric estimation and inference of TVP parameters. In particular, CTLS with a single *cp* (CTLS_1, hereafter) can provide a consistent estimation in predictive regression with a general covariate, and conventional inference when the predictor is restricted to be stationary process. While the later restriction on the regressor space is substantial, the regressor space is general enough to accommodate a wide range of specifications that are relevant for applied work. First, our framework allows for strictly stationary long memory processes and transformations of such processes. Linear and nonlinear predictive regressions with stationary long memory predictors have been considered by Christensen and Nielsen (2006, 2007) and more recently by Bollerslev et al. (2013) among others. Further, we can allow for transformations of heavy tailed predictors i.e. stationary covariates that may not possess second or first moments. For a review of the relevance of heavy tailed processes in finance and economics see for example Ibragimov, Ibragimov and Walden (2015). Moreover it should be mentioned that some preliminary theoretical results show that the proposed inferential methods for TVP models are also valid for weakly nonstationary processes i.e. MI and fractional $d = 1/2$ predictors (see e.g. Phillips and Magalinos (2007), MP and Duffy and Kasparis (2021). Finally, we conclude this section by providing some theoretical considerations that are relevant to applied work with regard to the use of FP versus TVP specifications.

3.1 FP models with a general covariate

Consider the following FP nonlinear model

$$y_k = \mu + \beta f(x_{k-1}) + e_k \text{ with } e_k = \sigma_k u_k,\ k = 1, ..., n, \tag{15}$$

where f is a known regression function (μ, β) unknown parameters and the covariate x_k can be a nonstationary or a stationary stationary process amenable to the limit theory of Theorem 1 or Theorem 2 respectively. The process u_k together with some filtration $\mathcal{F}_k$ forms a martingale difference sequence such that $\mathbb{E}\left(u_k^2 \mid \mathcal{F}_k\right) = 1$ *a.s.*, and x_k is $\mathcal{F}_k$-measurable. Finally, σ_k is a volatility process allowing for stationary GARCH effects (cf. Assumptions **A1-A3**). σ_k is assumed to be $\mathcal{F}_{k-1}$-measurable (i.e. predetermined w.r.t. to $\mathcal{F}_k$). The exact properties of the these processes will be specified in detail later. Similar nonlinear models with a predetermined covariate have been considered for example by Park and Phillips (1999, 2001) and Chan and Wang (2015), in a parametric set up, and by Wang and Phillips (2009a,b, 2011, 2012) in a nonparametric set-up.[12] For recent related results, we refer to Wang (2021) and Hu et al. (2021) and references therein.

This section considers CTLS estimation of (15) with multiple *cps*. Let K be a kernel function satisfying **A4**(a) or **A4***(a), and $\tau_j = j/(l_n+1), j = 1, ..., l_n$, c_n and l_n be deterministic sequences satisfying **A4**(b,c) or **A4***(b,c). The number l_n of *cps* is allowed to be a fixed w.r.t. to n or $l_n \to \infty$. As explained in Section 1, a minimum of two *cps* will be required to ensure that the CTLS estimator of μ, β has a full rank limit covariance matrix. Set

$$K_{kn} := \sum_{j=1}^{l_n} K\left[c_n\left(k/n - \tau_j\right)\right]. \tag{16}$$

Our aim is to estimate the unknown parameter β in (15) by using the following instrument for $f(x_{k-1})$

$$Z_{kn} := f_k K_{kn} := f(x_{k-1}) K_{kn}.$$

A chronological trimming instrumentation is also utilised for demeaning $\{y_k\}$, i.e. taking into account the unknown intercept μ. Let K^*_{kn}, $k = 1, ..., n$ be additive functionals of certain integrable kernel function defined by

$$K^*_{kn} := \sum_{j=1}^{l_n} K^*\left[c_n\left(k/n - \tau_j\right)\right], \tag{17}$$

where $K^*(x)$ is a kernel function that is specified in later, whilst $\tau_j = j/(l_n+1), j = 1, 2, ..., l_n$, are given as above.[13] For any sequence $\{a_k\}_{k=1}^n$, let

$$\overline{a} := \frac{\sum_{k=1}^n a_k K^*_{kn}}{\sum_{k=1}^n K^*_{kn}} \qquad \text{and} \qquad \overline{a}_k := a_k - \overline{a}. \tag{18}$$

[12] Here we consider nonlinear models in x_k only. Our results can be generalised to models that are both nonlinear in x_k and the parameters along the lines of Chan and Wang (2015) for instance.

[13] It is also possible choosing $K = K^*$. For purposes of generality and presentation we will assume a distinct kernel functional for intercept estimation. For the latter issue, a distinct kernel functional formulation provides better illustration of the consequences of intercept instrumentation to the limit theory.

Define the CTLS estimator[14] for β as

$$\hat{\beta} := \frac{\sum_{k=1}^{n} Z_{kn}\overline{y}_k}{\sum_{k=1}^{n} Z_{kn}\overline{f}_k}. \tag{19}$$

It should be mentioned that, when model (15) involves nonstationary components, the employment of a chronologically trimmed sample mean ($\overline{y}_k$) is crucial for obtaining mixed Gaussian limit theory. To see the reason, we rewrite $\hat{\beta}$ as

$$\hat{\beta} = \beta + \frac{1}{\sum_{k=1}^{n} Z_{kn}\overline{f}_k}\left\{\sum_{k=1}^{n} f_k K_{kn}\sigma_k u_k - \frac{\left(\sum_{k=1}^{n} f_k K_{kn}\right)\sum_{k=1}^{n} K_{kn}^{*}\sigma_k u_k}{\sum_{k=1}^{n} K_{kn}^{*}}\right\}.$$

The asymptotic behaviour of $\hat{\beta}$ is clearly determined by two martingale terms in the representation above, i.e. $\sum_{k=1}^{n} f_k K_{kn}\sigma_k u_k$ and $\sum_{k=1}^{n} K_{kn}^{*}\sigma_k u_k$. When x_k is nonstationary (satisfying **A3**, say), and K^* satisfies **A4**(a) or **A4***(a), the two martingale terms converge jointly to a bivariate mixed Gaussian limit (e.g. see the proof of Theorem 5), which in turn ensures mixed Gaussian limit theory for the CTLS estimator $\hat{\beta}$. However, if instead standard demeaning is employed (i.e. $K^* = 1$), mixed Gaussian limit theory for $\hat{\beta}$ is not applicable since it **is not true** that

$$\left[\sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^{n}\pi(d_n)^{-1} f_k K_{kn}\sigma_k u_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sigma_k u_k\right] \to_d \mathbf{MN}\left(\mathbf{0}, V\right),$$

for some random matrix V (i.e. a joint mixed Gaussian limit does not exist), despite the fact that each of the components on the l.h.s. above converges weakly to some (mixed) Gaussian limit.

We next state the asymptotic properties of the CTLS estimator $\hat{\beta}$. Theorem 4 considers a stationary regressor while Theorem 5 provides limit theory for the nonstationary case.

Theorem 4. *Suppose that:*

(*a*) ***A1**, **A2** with $g = f$ and **A4** or **A4*** hold;*

(*b*) *K^* satisfies **A4**(a) or **A4***(a).*

Then, as $n \to \infty$, we have

$$\sqrt{\frac{nl_n}{c_n}}\left(\hat{\beta} - \beta\right) \to_d \mathbf{N}\left(0, C_1^{-2} A_1 \mathbf{V}_1 A_1'\right),$$

[14]An equivalent formulation of the CTLS estimator is

$$\begin{bmatrix} \widehat{\mu} \\ \widehat{\beta} \end{bmatrix} = \left[\sum_{k=1}^{n}\begin{bmatrix} 1 \\ f_k \end{bmatrix}\begin{bmatrix} K_{kn}^{*} \\ f_k K_{kn} \end{bmatrix}'\right]^{-1}\cdot\sum_{k=1}^{n}\begin{bmatrix} K_{kn}^{*} \\ f_k K_{kn} \end{bmatrix} y_k$$

where $C_1 := \left\{Ef^2(x_1) - [Ef(x_1)]^2\right\} \int K$, $A_1 := \left(1, \ -Ef(x_1)\int K / \int K^*\right)$ *and*

$$\mathbf{V}_1 := E\left[\begin{array}{cc} \sigma_2^2 f(x_1)^2 \int K^2 & \sigma_2^2 f(x_1)\int KK^* \\ \sigma_2^2 f(x_1)\int KK^* & \sigma_2^2 \int (K^*)^2 \end{array}\right].$$

Remark 6. The limit result of Theorem 4 allows for conditional heteroscedasticity (e.g. GARCH) in the regression error. Under conditional homoscedasticity (i.e. $\sigma_1^2 \equiv \sigma^2$ for some non random $\sigma^2 \in (0,\infty)$), the matrix $\mathbf{V}_1$ simplifies to

$$\mathbf{V}_1 = \sigma^2 E\left[\begin{array}{cc} f(x_1)^2 \int K^2 & f(x_1)\int KK^* \\ f(x_1)\int KK^* & \int (K^*)^2 \end{array}\right].$$

With a stationary regressor in model (15), Theorem 4 holds under nearly minimal assumption on the regression function f (c.g. **A2** with $g = f$). It is well known however that for nonstationary regressions (e.g. Park and Phillips, 1999, 2001), different regression functions lead to different convergence rates and limit theory. Following Park and Phillips (2001), for nonstationary regressions we focus on locally integrable regression functions (that are not integrable) and exhibit asymptotic homogeneity. This family of functions is the most relevant to empirical work allowing for polynomial, logarithmic, threshold and smooth transition (distribution type) transformations.

Theorem 5. *Suppose that*

(*a*) ***A1**, **A3** and **A4** or **A4*** hold with* $l_n \to \infty$*;*

(*b*) *there exist a continuous function* H *on* $\mathbb{R}$ *satisfying* $|H(x)| \leq C(1+|x|^\alpha)$ *for some* $\alpha > 0$ *and a real function* $\pi : (0,\infty) \to (0,\infty)$ *so that*

$$f(\lambda x) = \pi(\lambda)H(x) + R(\lambda, x),$$

where $|R(\lambda,x)| \leq a(\lambda)(1+|x|^\delta)$ *for some* $\delta > 0$ *and* $a(\lambda)/\pi(\lambda) \to 0$*, as* $\lambda \to \infty$*;*

(*c*) K^* *satisfies **A4**(a) or **A4***(a).*

Then, as $n \to \infty$*,*

$$\sqrt{\frac{nl_n}{c_n}}\pi(d_n)\left(\hat{\beta} - \beta\right) \to_d \sqrt{E\sigma_1^2}\,\mathbf{MN}\left(0, C_2^{-2}\,A_2\mathbf{V}_2A_2'\right),$$

where

$$C_2 = \left\{\int_0^1 H^2(X_t)dt - \left[\int_0^1 H(X_t)dt\right]^2\right\}\int K, \quad A_2 = \left[1, \ -\int_0^1 H(X_t)dt\int K/\int K^*\right],$$

$$\mathbf{V}_2 = \begin{bmatrix} \int_0^1 H^2(X_t)dt \int K^2 & \int_0^1 H(X_t)dt \int KK^* \\ \int_0^1 H(X_t)dt \int KK^* & \int (K^*)^2 \end{bmatrix}.$$

Remark 7. Standard arguments (e.g. proof of Theorem 3.2 of Wang, 2021) yield that the OLS estimator $\widetilde{\beta}_{OLS}$ is $\sqrt{n}(\widetilde{\beta}_{OLS} - \beta) = O_P(1)$ and $\sqrt{n}\pi(d_n)(\widetilde{\beta}_{OLS} - \beta) = O_P(1)$, under the conditions of Theorems 4 and 5, respectively [15]. Therefore, convergence rate of CTLS estimators for both stationary and nonstationary regressors is slower by a $(l_n/c_n)^{1/2}$ rate.

Remark 8. Theorem 5 holds under the assumption $l_n \to \infty$. When $l := l_n \geq 2$ is fixed, we have the same result with the limit terms $\int_0^1 H(X_t)dt$, $\int_0^1 H^2(X_t)dt$ replaced by $\frac{1}{l}\sum_{j=1}^{l} H\left(X_{\tau_j}\right)$ and $\frac{1}{l}\sum_{j=1}^{l} H\left(X_{\tau_j}\right)^2$ respectively.

Remark 9. It can be readily seen from Theorem 5 that under nonstationarity, conditional heteroscedasticity in the regression error does not affect the limit variance of the CTLS estimator in a material way. In particular, the volatility term $E\sigma_1^2$ is scaled out. As a result conventional estimators for the limit variance can be employed for the construction of t-statistics (see also Remark 11 below). This result is comparable with the recent findings of Magdalinos (2020) how demonstrates that conditional heteroscedasticity has a material effect in the limit distribution of the IVX estimator only under stationarity.

Next, we consider the following t-statistics for the hypothesis $H_0 : \beta = \beta_0$ (for some $\beta_0 \in \mathbb{R}$)

$$\hat{T} := \mathcal{C}_n \frac{\hat{\beta} - \beta_0}{\sqrt{\mathcal{A}_n \mathcal{V}_n \mathcal{A}_n'}}, \tag{20}$$

where

$$\mathcal{A}_n := \left[1, \quad -\frac{\sum_{k=1}^n f_k K_{kn}}{\sum_{k=1}^n K_{kn}^*}\right], \quad \mathcal{C}_n := \sum_{k=1}^n Z_{kn}\overline{f}_k,$$

$$\mathcal{V}_n := \begin{bmatrix} \sum_{k=1}^n \check{e}_k^2 K_{kn}^2 f_k^2 & \sum_{k=1}^n \check{e}_k^2 K_{kn}^* K_{kn} f_k \\ \sum_{k=1}^n \check{e}_k^2 K_{kn}^* K_{kn} f_k & \sum_{k=1}^n \check{e}_k^2 \left(K_{kn}^*\right)^2 \end{bmatrix},$$

with $\check{e}_k = y_k - \widetilde{\mu}_{OLS} - \widetilde{\beta}_{OLS} f_k$, and $(\widetilde{\mu}_{OLS}, \widetilde{\beta}_{OLS})$ the OLS estimator of (μ, β). The limit properties of $\hat{T}$ under the null hypothesis are demonstrated by following theorem.

Theorem 6. *Suppose that $H_0 : \beta = \beta_0$ is true, either the conditions of Theorem 4 or Theorem 5 hold, and* $\sup_{k\geq 1} Eu_k^4 < \infty$. *Then*

$$\hat{T} \to_d \mathbf{N}(0,1).$$

[15] Asymptotic distribution can be explicitly obtained by using some additional notation (c.g. Theorem 3.2 of Wang, 2021). Since it is beyond the scope of this paper, we omit the details.

Remark 10. The limit distribution of the test statistic under the null hypothesis is standard normal for both stationary and nonstationary regressors. Under the alternative hypothesis, the divergence rate of both t-statistics is determined by the convergence rate of the CTLS estimator. In particular, for stationary x_k it can be easily seen that $\hat{T} = O_P(\sqrt{nl_n/c_n})$. On the other hand in the nonstationary case we have $\hat{T} = O_P(\pi(d_n)\sqrt{nl_n/c_n})$, where $d_n = \sqrt{n}$ for x_k NI and $d_n = n^d$, x_k for $I(d)$, $1/2 < d < 3/2$. Therefore, faster divergence rate is attained for more persistence processes. This fact is also corroborated by our simulation results (see Figure 3). In the nonstationary case, asymptotic power is affected by the asymptotic order (i.e. growth rate) of f. Note that for logarithmic, or lower order polynomial (e.g. $f(x) = |x|^p$, $p < 1$) regression functions, slower power rates are attained relative to linear and higher order polynomial transformations.

Remark 11. Note that the normalising matrix $\mathcal{V}_n$ allows for conditional heterscedasticity. If however $\sigma_k^2 = \sigma^2$ for all k then the following estimator can be used instead.

$$\check{\mathcal{V}}_n := \check{\sigma}^2 \begin{bmatrix} \sum_{k=1}^n K_{kn}^2 f_k^2 & \sum_{k=1}^n K_{kn}^* K_{kn} f_k \\ \sum_{k=1}^n K_{kn}^* K_{kn} f_k & \sum_{k=1}^n (K_{kn}^*)^2 \end{bmatrix}, \ \check{\sigma}^2 := n^{-1} \sum_{k=1}^n \check{e}_k^2. \tag{21}$$

In view of Remark 9, $\check{\mathcal{V}}_n$ provides also a consistent estimator under nonstationarity even if the regression errors are conditionally heteroscedastic.

3.2 TVP models with a stationary covariate

We next focus on CTLS estimation of TVP predictive regressions of the form

$$y_k = \mu(k/n) + \beta(k/n) \cdot f(x_{k-1}) + e_k, \ k = 1, ..., n, \tag{22}$$

where $\mu, \beta : (0, 1] \to \mathbb{R}$, the predictor x_k is a strictly stationary process that may exhibit long memory or could be heavy tailed. The error term e_k is a martingale difference term as in (15). Stochastic and deterministic TVP models have gained a lot of attention recently in both econometrics and statistical time series due to their ability to accommodate structural change e.g. see Dahlhaus (2000), Giraitis, Kapetanios and Yates (2014, 2018), Phillips, Li and Gao (2017), Dahlhaus, Richter and Wu (2019), Demetrescu et al. (2020), among others. A number of papers in this area consider (possibly vector) autoregressive type TVP models with the autoregressive parameters' modulus bounded below unity (e.g. Giraitis et al., 2014; Dahlhaus et al. 2019). These models behave like stable autoregressive processes, when it comes to estimation, and therefore conventional inference applies.

The theoretical framework of Phillips et al. (2017) is more closely related to the specification of (22). The aforementioned work considers estimation and inference in non autoregressive TVP models with deterministic parameters and multiple $I(1)$ covariates. The estimation methods of Phillips et al. (2017) are also related to the current work. In particular, these authors propose certain CTLS type of estimation that utilises a single *cp* (CTLS_1). As explained in the introduction, in multi-parameter regressions, CTLS_1 results in a singular limit covariance matrix and this renders inference non conventional with limit distributions comparable to those OLS in regressions with $I(1)$ covariates. To get pivotal statistical tests, Phillips et al. (2017) propose CTLS_1 regression of a "*fully modified*" version of the dependent variable y_k (cf. Phillips and Hansen 1990). A similar specification to that of (22) has been also considered by Demetrescu et al. (2020) who develop predictability tests for regressions with a deterministic and possibly time varying slope parameter under the alternative hypothesis (predictability). These predictability tests are based on sup-functionals of studentised IV estimators that utilise a combination of IVX and other time trend instruments.

We consider CTLS_1 estimation and related t-tests in the context of (22). An advantage of the CTLS approach in this framework is that it can be used for both estimation and testing. In particular, this method provides direct non-parametric estimators of TVP functionals. In addition, studentised estimators can form the basis of non-parametric t-tests. Implementation of these tests is very simple since test statistics are not complex, limit distributions are free of nuisance parameters, and critical values readily available from statistical tables. In order to keep technical complexity simple, in this Section we consider models with a single regressor. Nevertheless the results can be easily extended to multivariate models (see Remark 1). A generalisation to the multivariate case is provided in the Appendix 8.

The limit properties of TVP estimators and related test statistics are based on the asymptotic theory developed in Section 2 for CT functionals of strictly stationary processes. CTLS_1 methods can provide consistent estimation of TVPs even for nonstationary covariates however, as mentioned earlier, we need to restrict the regressors to be stationary processes in order to ensure that the test statistics have conventional distributions free of nuisance parameters. The regressor space under consideration is general enough to accommodate several data generating processes relevant in empirical marcoeconomics and finance such as long memory linear processes, and GARCH, ARCH(∞). Further, due the the fact that we are considering models that are nonlinear in variables, in some cases it is possible to allow for heavy tailed linear processes.[16] For instance, we can allow stationary linear processes of the form

$$x_k = \sum_{i=0}^{\infty} \phi_i \xi_{k-i}, \tag{23}$$

with either

[16] In situatations where the predictor x_k is heavy tailed, a nonlinear regression function of reduced growth may ensure that $f(x_k)$ has sufficient moments for the validity of Theorem .

FR: $\xi_i \sim iid(0, \sigma_\xi^2), \sum_{i=0}^{\infty} \phi_i^2 < \infty$; or

HT: (a) $\xi_i \sim iid$ in the domain of attraction of an α-stable law with $\alpha \in (0, 2]$, $\sum_{i=0}^{\infty} |\phi_i|^{\min\{\alpha', 1\}} < \infty$, $\alpha' \in (0, \alpha)$; and (b) $Ef(x_1)^4 < \infty$.

Condition **FR**, x_k allows for stationary fractional process of memory parameter $|d| < 1/2$. Condition **HT(a)** ensures that x_k is a well defined possibly heavy tailed process (e.g. Astrauskas, 1983) that possesses a finite α' moment. In the latter case, x_k may not have a finite mean or variance. Condition **HT(b)** is a technical requirement that is utilised for obtaining the limit distribution of CTLS_1 estimators. In practice, $f(x_1)^4$ may have a finite moment even if the predictor is heavy tailed. For example, if f satisfies the reduced growth requirement $|f(x)| \leq C(1 + |x|^p)$, $C \in (0, \infty)$, $p \in (0, 1/4)$, **HT** holds for all $\alpha \in (4p, 2]$. Further, $f(x) = \ln(x)_+$ satisfies the aforementioned requirement for all values of the tail parameter α in $(0, 2]$.[17] Logarithmic transformations and reduced polynomial growth regression functions have been used a number of studies in the predictability of stock returns (see Section 5 for more details). To some extent, our methods on TVP models are complementary to Phillips et al. (2017) who focus on different area of the regressor space. In fact the regressor space under consideration is comparable to that of Christensen and Nielsen (2006, 2007), Bandi and Perron (2008), Bollerslev et al. (2013), Bandi et al. (2018) among others, who consider models with stationary fractional predictors.

The CTLS_1 estimator can be formulated as a *local-level* kernel regression estimator (cf. Li and Racine, 2006; Section 2.1). In particular, for $K = K^*$ estimators for the TVPs of (22) can be obtained from the minimisation of the objective function

$$\begin{bmatrix} \hat{\mu}(\tau) \\ \hat{\beta}(\tau) \end{bmatrix} =: \hat{\theta}(\tau) := \arg\min_{\mathbf{a} \in \mathbb{R}^2} \sum_{k=1}^{n} (y_k - \mathbf{a}'\mathbf{f}_k)^2 K \left[c_n (k/n - \tau)\right], \tag{24}$$

where $\tau \in (0, 1]$ and $\mathbf{f}_k' := [1, f(x_{k-1})]$. This estimator is closely related to that considered by Phillips et al. (2017).[18] Set $\theta(\tau)' := [\mu(\tau), \beta(\tau)]$ and define the vector of derivatives $\theta^{(1)}(\tau) := \partial\theta(\tau)/\partial\tau$. We also consider the following TVP estimator:

$$\begin{bmatrix} \widetilde{\theta}(\tau) \\ \widetilde{\theta}^{(1)}(\tau) \end{bmatrix} := \arg\min_{(\mathbf{a}', \mathbf{b}')' \in \mathbb{R}^4} \sum_{k=1}^{n} \left(y_k - \mathbf{a}'\mathbf{f}_k - \mathbf{b}'\mathbf{f}_{1k}\right)^2 K \left[c_n (k/n - \tau)\right], \tag{25}$$

where $\mathbf{f}_{1k} = (k/n - \tau)\,\mathbf{f}_k$, and $\widetilde{\theta}(\tau)$ and $\widetilde{\theta}^{(1)}(\tau)$ are CTLS_1 type estimators for $\theta(\tau)$ and $\theta^{(1)}(\tau)$, respectively. The latter is a *local-linear* estimator (cf. Li and Racine, 2006; Section

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

[18] Phillips et al. (2017) consider a local level estimator for mulivariate TVP (linear in variable) models, and a fully modified version of the local level estimator that involves of modified version of the dependent variable along the lines of Phillips (1995).

2.5) that exhibits reduced asymptotic bias, relative to the local-level of (24). Write

$$Q = \begin{bmatrix} 1 & Ef(x_1) \\ Ef(x_1) & Ef(x_1)^2 \end{bmatrix} \quad \text{and} \quad \mathbf{\Omega} = \begin{bmatrix} E\sigma_2^2 & E\{\sigma_2^2 f(x_1)\} \\ E\{\sigma_2^2 f(x_1)\} & E\{\sigma_2^2 f(x_1)^2\} \end{bmatrix}.$$

The following theorem demonstrates the limit distribution of the local-level CTLS_1 estimator (LLev, hereafter).

Theorem 7. *Suppose that:*

(*a*) $\{y_k\}_{k\in\mathbb{N}}$ *is generated by (22);*

(*b*) ***A1*** *holds and, in addition to* ***A2*** *with* $g = f$, $P[f(x_1) \neq Ef(x_1)] \neq 0$;

(*c*) K *satisfies* ***A4****(a) or* ***A4******(a);*

(*d*) $\theta(.)$ *is Holder continuous on* $(0, 1]$ *of order* $\gamma \in [0, 1]$;[19]

(*e*) $c_n/n + n/c_n^{1+2\gamma} \to 0$, *where* γ *is defined as in (d).*

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

$$\sqrt{\frac{n}{c_n}} \left(\hat{\theta}(\tau) - \theta(\tau)\right) \to_d \mathbf{N}\left(\mathbf{0}, Q_1^{-1}\mathbf{\Omega}_1 Q_1^{-1}\right), \tag{26}$$

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

Remark 12. Note that $P[f(x_1) \neq Ef(x_1)] \neq 0$ implies that $[Ef(x_1)]^2 < Ef^2(x_1)$, i.e. $[Ef(x_1)]^2 \neq Ef^2(x_1)$. The latter condition insures that Q (Q_1) is of full rank under the assumption that x_k is a stationary process. Phillips et al. (2017) show that, if x_k is an $I(1)$ process, a CTLS_1 (local level) estimator has necessarily a singular covariance matrix. 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 follows directly from Lemma 2 (see also Remark 4) that the counterpart of the limit matrix Q is of the form

$$\begin{bmatrix} 1 & f(X_\tau) \\ f(X_\tau) & f(X_\tau)^2 \end{bmatrix} \int K \text{ with } X_t \in D[0, 1],$$

which is necessarily singular.

[19] i.e. $\|\theta(x) - \theta(y)\| \leq C \|x - y\|^\gamma$, for $x, y \in (0, 1]^2$ and $C \in (0, \infty)$. Note for constant $\theta(.)$, $\gamma = 0$.

Remark 13. Preliminary theoretical results suggest that CTLS_1 method is also valid in situations where the the predictor is weakly non-stationary process (i.e. fractional $d = 1/2$ or MI). In this case (26) holds with Q_1

$$Q_1 = \left[\begin{array}{cc} 1 & \int_{\mathbb{R}} f(x+X^-)\varphi_{\sigma_+^2}(x)dx \\ \int_{\mathbb{R}} f(x+X^-)\varphi_{\sigma_+^2}(x)dx & \int_{\mathbb{R}} f(x+X^-)^2\varphi_{\sigma_+^2}(x)dx \end{array} \right] \int_{\mathbb{R}} K(x)dx,$$

where $\varphi_{\sigma_+^2}(x)$ and X^- as in Remark 2. Note that Q_1 here is in general non singular.

The limit properties of the local-linear estimator (LLin, hereafter) are given by the following result.

Theorem 8. *Suppose that:*

(*a*) $\{y_k\}_{k\in\mathbb{N}}$ *is generated by (22)*

(*b*) ***A1*** *holds and, in addition to* ***A2*** *with* $g = f$, $P\left[f(x_1) \neq Ef(x_1)\right] \neq 0$;

(*c*) *in addition to that* K *satisfies* ***A4****(a) or* ***A4******(a),* $\int x^2K^2 < \infty$;

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

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

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

$$D_n \left(\left[\begin{array}{c} \widetilde{\theta}(\tau) \\ \widetilde{\theta}^{(1)}(\tau) \end{array} \right] - \left[\begin{array}{c} \theta(\tau) \\ \theta^{(1)}(\tau) \end{array} \right] \right) \to_d \mathbf{N}\left(\mathbf{0}, Q_2^{-1}\mathbf{\Omega}_2 Q_2^{-1}\right), \tag{27}$$

where $D_n = diag\left\{\sqrt{\frac{n}{c_n}}, \sqrt{\frac{n}{c_n}}, \sqrt{\frac{n}{c_n^3}}, \sqrt{\frac{n}{c_n^3}}\right\}$,

$$Q_2 = \left[\begin{array}{cc} Q\int K & Q\int xK \\ Q\int xK & Q\int x^2K \end{array} \right] \quad \textit{and} \quad \mathbf{\Omega}_2 = \left[\begin{array}{cc} \mathbf{\Omega}\int K^2 & \mathbf{\Omega}\int xK^2 \\ \mathbf{\Omega}\int xK^2 & \mathbf{\Omega}\int x^2K^2 \end{array} \right].$$

Remark 14. The smoothness assumptions on the TVP $\theta(.)$ of Theorems 7 and 8 are standard. Note that for the LLin estimator more restrictive assumptions are required. Nevertheless, LLin attains smaller 'asymptotic bias', relative to the LLev estimator. In general, kernel regression type of 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 (26) entails an asymptotic bias term of order $O_P(\sqrt{n/c_n^3})$, whilst the corresponding bias term in (27) is $O_P(\sqrt{n/c_n^5})$. Note that when the parameters are fixed, with respect to time, the bias terms equal zero for all n. In can be readily seen that for both estimators there is a trade-off between convergence rate and bias. An additional advantage of the LLin is that entails estimates for the derivatives of the TVPs. These estimates can be readily utilised for testing hypotheses about parameter constancy with respect to time (see also Remark 19 below).

Remark 15. As noted in Remark 12, Q (Q_1) is of full rank. This, together with $\left(\int xK\right)^2 \neq \int K \int x^2 K$ due to $0 < \int K < \infty$, implies that Q_2 is also of full rank, indicating the limitation in (27) is well defined. If K is symmetric, i.e., $\int xK = 0$ and $\int xK^2 = 0$, we further have that the limit distributions of $\widetilde{\theta}(\tau)$ and $\widetilde{\theta}^{(1)}(\tau)$ are independent.

Remark 16. As mentioned above the limit results of Theorems 7 and 8 can be readily generalised to additively separable multi-covariate models with stationary regressors of the form

$$y_k = \mu(k/n) + \sum_{j=1}^{p-1} \beta_j\,(k/n) \cdot f_j(x_{k-1,j}) + e_k,\;\; k = 1, ..., n,$$

with $p \geq 2$, $x_{k,j} \sim I(d_j)$, $|d_j| < 1/2$. More details on this generalisation are provided in the Appendix (Section 8).

Remark 17. It can be readily seen from Theorems 7 and 8 that the limit variance of the TVP estimators is 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 t-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 be its i^{th} element, and for a square matrix A, $[A]_{ii}$ denotes its i^{th} diagonal element. The test hypothesis under consideration is of the form

$$H_0 : \theta_i(\tau) = \eta(\tau), \tag{28}$$

and

$$H_0 : \theta_i^{(1)}(\tau) = \eta(\tau), \tag{29}$$

for $i = \{1, 2\}$, some prespecified $\eta : (0, 1] \to \mathbb{R}$ and $\tau \in (0, 1]$. In particular, (28) entails a hypothesis for $\mu(\tau)$ and $\beta(\tau)$, whilst (29) concerns the derivatives of the aforementioned parameters. The proposed tests utilise the estimators of (26) and (27). Set

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

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

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

where $\widetilde{\mathbf{f}}_k = \left(\mathbf{f}_k', \mathbf{f}_{1k}'\right)'$ with $\mathbf{f}_{1k} = (k/n - \tau)\,\mathbf{f}_k$ and $\tilde{e}_k := \tilde{e}_k(\tau) := y_k - \tilde{\theta}(\tau)'\mathbf{f}_k$. The proposed test statistics can be constructed as

$$\hat{t}_i(\tau) = \frac{\hat{\theta}_i(\tau) - \eta(\tau)}{\sqrt{\left[\hat{\mathcal{Q}}_n^{-1} \hat{\mathbf{\Omega}}_n \hat{\mathcal{Q}}_n^{-1}\right]_{ii}}}, \quad \tilde{t}_i(\tau) = \frac{\tilde{\theta}_i(\tau) - \eta(\tau)}{\sqrt{\left[\tilde{\mathcal{Q}}_n^{-1} \tilde{\mathbf{\Omega}}_n \tilde{\mathcal{Q}}_n^{-1}\right]_{ii}}}, \quad i = 1, 2,$$

for the null hypothesis (28), and

$$\tilde{t}_i^{(1)}(\tau) = \frac{\tilde{\theta}_i^{(1)}(\tau) - \eta(\tau)}{\sqrt{\left[\tilde{\mathcal{Q}}_n^{-1} \tilde{\mathbf{\Omega}}_n \tilde{\mathcal{Q}}_n^{-1}\right]_{jj}}}, \quad i = 1, 2,\ j = i + 2,$$

for the null hypothesis (29). The following theorems establish the limit properties of the these test statistics.

Theorem 9. *Suppose that, in addition to the conditions of Theorem 7, $\int x^2 K^2 < \infty$, $\sup_{k \geq 1} E u_1^4 < \infty$ and $E f^4(x_1) + E\left\{\sigma_2^2\left[1 + f^2(x_1)\right]\right\} < \infty$. Under $H_0 : \theta_i(\tau) = \eta(\tau)$, we have*

$$\hat{t}_i(\tau) \to_d \mathbf{N}(0, 1). \tag{30}$$

Theorem 10. *Suppose that, in addition to the conditions of Theorem 8, $\int x^4 K^2 < \infty$, $\sup_{k \geq 1} E u_1^4 < \infty$ and $E f^4(x_1) + E\left\{\sigma_2^2\left[1 + f^2(x_1)\right]\right\} < \infty$. Under $H_0 : \theta_i(\tau) = \eta(\tau)$, we have*

$$\tilde{t}_i(\tau) \to_d \mathbf{N}(0, 1), \tag{31}$$

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

$$\tilde{t}_i^{(1)}(\tau) \to_d \mathbf{N}(0, 1). \tag{32}$$

Remark 18. The alternative hypothesis of all tests can be either two-sided or one-sided. The asymptotic power rates of the test statistics under consideration are determined by the convergence rates of the CTLS_1 estimators involved. In particular, it can be easily checked that $\hat{t}_i(\tau), \tilde{t}_i(\tau) = O_P\left(\sqrt{n/c_n}\right)$ under H_1, whilst $\tilde{t}_i^{(1)}(\tau) = O_P\left(\sqrt{n/c_n^3}\right)$. 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 19. Note that $\tilde{t}_i^{(1)}(\tau)$ can be used for testing parameter constancy with respect to time i.e. $H_0 : \theta_i^{(1)}(\tau) = 0$, for each $\tau \in (0, 1]$.

3.3 Neglecting time variation in regression parameters: some theoretical considerations

We conclude this section with some brief discussion about the consequences of neglecting time variation in regression parameters of predictive models. The issues pointed out here, are useful for the interpretation of the empirical results of Section 5. For instance, in our empirical application we find that in FP models realised variance does not provide significant predictability, while in TVP models it does. A plausible explantation for these conflicting results is provided by certain theoretical facts discussed next. Sketch proofs for the subsequent theoretical results are provided in the Appendix (Section 9).

Although substantial progress has been made recently in the development statistical methods for TVP models, time variation in parameters has attracted little attention in the returns predictability literature. They key advantage of TVP models is that they can accommodate structural change. It is reasonable to expect that structural change may occur due external shocks or during different phases of the business cycle. In practice, structural change is more likely to occur in situations where the data sets' span is very long, as it is in the case of Welch and Goyal (2008) who consider predictability of stock returns using data from 1926 to 2005. In this work we use an updated data set due to this authors from 1926 to 2018.

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. in some regression parameter that is not the focus practitioner's analysis e.g. the regression intercept or the slope parameter of some other covariate-. In general neglecting time variation in the parameter of interest leads to inconsistent estimates, and undermines the power of predictability 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 test statistics divergent under the null hypothesis when the predictor has memory parameter strictly greater than zero. We demonstrate the above for OLS based inference for regressions with stationary predictors, but we expect that similar phenomena also apply to other FP methods e.g. CTLS (multiple *cps*), IVX, conservative predictability tests, when covariates are either stationary or nonstationary.

We first demonstrate the consequences of neglecting time variation in the parameter of interest by considering the following simple linear regression

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

where, x_k is stationary long memory satisfying (23) and condition **FR**. Then under certain regularity conditions, it follows easily from Lemma 1, and some additional arguments that the OLS estimator from regressing y_k on x_{k-1} is

$$\sqrt{n}\left(\tilde{\beta}_{OLS} - \int_0^1 \beta(\tau)d\tau\right) \to_d \left(Ex_1^2\right)^{-1} \cdot \mathbf{N}\left(0, E\left(e_2^2 x_1^2\right) + \int_0^1 \beta(\tau)^2 d\tau \cdot Var(x_1^2)\right).$$

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-test are likely to have poor power in situations where predictability is episodic. To see this note that under the alternative hypothesis (predictability)

$$\tilde{t}_{OLS} = \sqrt{n}\int_0^1 \beta(\tau)d\tau \cdot O_P(1)$$

The value of 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. $\beta(.) < 0$).

Next, we illustrate the effects of neglecting time variation in the intercept when the parameter of interest is the slope coefficient in the following model

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

Suppose that we are interested testing $H_0 : \beta = 0$, using OLS based inference. For convenience suppose that the coefficients in (23) for long memory x_k (i.e. $0 < d < 1/2$) are $\phi_i \sim cons. \cdot i^{d-1}$ (see e.g. Johansen and Nielsen, 2012; p. 673). Further, without loss of generality suppose that innovations with negative index in (23) are zero.[20] In this case, it is well known $\delta_n := [Var(\sum_{k=1}^n x_k)]^{1/2} \sim cons. \cdot n^{1/2+d}$, with $0 \le d < 1/2$, and $\delta_n^{-1}\sum_{k=1}^n x_k \to_d \mathbf{N}(0,1)$ (for the latter see e.g. Ibragimov and Linnik, 1971; Thm 18.6.5 or Peligrad and Utev, 1997). Then it can be shown that the OLS estimator

$$\sqrt{n}\left(\tilde{\mu}_{OLS} - \int_0^1 \mu(\tau)d\tau\right) \to_d \mathbf{N}(0, E\left(e_1^2\right)) \tag{33}$$

[20] In this case x_k is a type II long memory process see e.g. Phillips and Shimotsu, 2004.

and

$$\frac{n}{\delta_n}\left(\tilde{\beta}_{OLS}-\beta\right)\rightarrow_d \left(Ex_1^2\right)^{-1}\left[1,-\int_0^1 \mu(\tau)d\tau\right]\cdot \mathbf{N}\left(\mathbf{0},E\left(\xi_1^2\right)\Psi\right), \tag{34}$$

with

$$\Psi=\int_0^1 \begin{bmatrix} \left\{\int_r^1 \mu\left(1-s\right)\left(s-r\right)^{d-1}ds\right\}^2 & 2\int_r^1 \mu\left(1-s\right)\left(s-r\right)^{d-1}ds\cdot\int_r^1\left(s-r\right)^{d-1}ds \\ 2\int_r^1 \mu\left(1-s\right)\left(s-r\right)^{d-1}ds\cdot\int_r^1\left(s-r\right)^{d-1}ds & \left\{\int_r^1 \left(s-r\right)^{d-1}ds\right\}^2 \end{bmatrix} dr. \tag{35}$$

for $0<d<1/2$. Therefore, OLS estimator for β is consistent however there is a reduction in the converge rate when the predictor is (stationary) fractional with memory parameter strictly greater than zero. This reduction in the convergence rate does not effect asymptotic power[21], nevertheless it results in severe size distortions under the null hypothesis. To see this note first that the regression error variance estimator is

$$\frac{1}{n}\sum_{k=1}^{n}\tilde{e}_k^2\rightarrow_P Ee_1^2+\int_0^1\mu(\tau)^2d\tau-\left(\int_0^1\mu(\tau)d\tau\right)^2. \tag{36}$$

Combining (33)-(36) it follows that under the null hypothesis,

$$\left|\tilde{t}_{OLS}\right|\rightarrow_P\infty. \tag{37}$$

In fact the divergence rate of the t-statistic is $\delta_n/n^{1/2}=n^d$, $0\leq d<1/2$. Clearly, when x_k is a short memory process the test statistic is bounded under the null, nevertheless it does not have a standard normal distribution and therefore OLS based t-tests exhibit size distortions even in this case.

4 Simulations

We next explore the finite sample properties of CTLS inferential methods with the aid of a simulation study. First, we consider the no predictability hypothesis

$$H_0:\beta=0 \text{ vs } H_1:\beta\neq 0$$

for FP regressions of the form[22]

$$y_k=\beta x_{k-1}+e_k. \tag{38}$$

[21] It follows from (33)-(36) that $\tilde{t}=O_P(\sqrt{n})$, under H_1.

[22] Without loss of generality we set $\mu=0$. For FP models, estimators are numerically invariant to the value of the intercept.

Further, in the context of TVP models of the form

$$y_k = \mu(k/n) + \beta(k/n)x_{k-1} + e_k, \tag{39}$$

we consider 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. As discussed in Section 3.3 neglecting time variability in the intercept could result in severe size distortions.

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 either a NI array of the form

$$x_k = \left(1 + \frac{c}{n}\right) x_{k-1} + \xi_k, \tag{40}$$

with $c \leq 0$ and $x_0 = 0$ or a type II fractional process (e.g. see Robinson and Hualde, 2003) of the form

$$(I - L)^d x_k = \xi_k 1\{k \geq 1\}. \tag{41}$$

The regression error is

$$e_k = \sigma_k u_k,$$

with either

$$\sigma_k^2 = 1,$$

or

$$\sigma_k^2 = 0.01 + 0.45\sigma_{k-1}^2 + 0.45e_{k-1}^2, \; \sigma_0^2 = 0.01, \tag{42}$$

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

Simulations for FP models. We first consider the finite sample performance of CTLS based t-tests of Section 3.1 for models with fixed regression parameters. In particular, we report empirical size and power results for $\hat{T}$ given in (20), and $\check{T}$ which is a CTLS t-stastistic that is utilising the variance estimator of (21). Note that the former provides valid inference in the presence of GARCH regression errors while the second is in general relevant when the errors are (conditionally) homoscedastic. CTLS methods involve various tuning parameters such as bandwidths and kernel functions that do affect finite sample performance. As explained in the Section 1 and 2, there is in general a trade-off between size and power when it comes to the choice of c_n and l_n, with better size control achieved in general for larger values for c_n and smaller for l_n. We have conducted extensive preliminary simulations involving various choices of tuning parameters. We only report results for the set-up that attains the best size-power trade-off, according to the preliminary simulations.[23] Let $\varphi_\varsigma(x)$ be the density of $\mathbf{N}(0,\varsigma)$ variate. For FP models we consider the following kernel functions and bandwidth terms

- $K_{kn} = \sum_{j=1}^{l_n} K\left[c_n\left(k/n-\tau_j\right)\right]$, $K_{kn}^* = \sum_{j=1}^{l_n} K^*\left[c_n\left(k/n-\tau_j\right)\right]$, $K(x) = \varphi_{0.1}(x)^{1/2}$, $K(x)^* = \varphi_1(x)^{1/2}$;
- $c_n = n^{0.95}$, $l_n = c_n^{0.7}$,

where $\{\tau_j\}_{j=1}^{l_n}$ are equispaced points on $(0,1)$.

Table 2 reports the empirical size of CTLS tests for NI predictors and $\sigma_k^2 = 1$. For comparison with also consider an IVX test, that incorporates the finite sample correction of Kostakis et al. (2015), and an OLS based t-test. The IVX method appear to have superior finite sample performance relative to other methods for NI -see Kostakis et al., (2015) and Kasparis et al. (2015). Contrary to CTLS and IVX estimators, OLS does not have a mixed normal distribution under nonstationarity, nevertheless OLS based methods -or similar procedures appropriate only for stationary models (e.g. Gaussian MLE), are routinely used in empirical work (see e.g. Stambaugh, 1999; Amihud and Hurvich, 2004; Bandi and Perron 2008; Chen and Deo (2009), Bandi et al., 2018). Further, OLS provides a natural benchmark for assessing the benefits in empirical size when mixed normality is induced. It can be seen from Table 2 that in general, both CTLS test statistics result in good size control with empirical size close to nominal in most cases. CTLS is somewhat oversized relative to IVX when the near to unity parameter is $c = 0$, and endogeneity is strong. Size however improves as sample size increases. We further investigate the empirical size of $\hat{T}$ for NI predictors and GARCH regression errors in Table 3. Simulations show that the empirical size of $\hat{T}$ in this case is comparable to that reported in Table 2 under conditional homoscedasticity. Simulation results not reported here suggest that $\check{T}$ and IVX exhibit some moderate oversizing under GARCH regression errors, and large deviations from unity. Finally, we consider (see Table 4) the size performance of $\hat{T}$ for a wide range of fractional predictors and conditionally homoscedastic regression errors. Again size control is in general

[23]In a previous version of this paper (arXiv.org> arXiv:2006.12595) we also consider a more complicated data driven method for choosing the tuning parameters that appears to attain superior finite sample performance for a wide range of configurations. However, the particular approach requires further theoretical investigation that we leave for future work.

good with some oversizing evident when the memory parameter is above unity. In all cases the size of CTLS tests can be improved by choosing smaller l_n or larger c_n at the expense of power performance. Additional simulation results, not reported here, show that the size performance of $\hat{T}$ in regressions with fractional predictors and GARCH regression errors is very similar to those shown in Table 4.

Finally, we explore the empirical power of CTLS tests. We focus on the $\hat{T}$ statistic that is robust to conditional heteroscedasticity in the regression error. We first consider power performance for when the predictor is NI. Figure 2 reports rejection probabilities for $\hat{T}$ and Kostakis et al. (2015) IVX t-statistic against various values for the slope parameter for under strong endogeneity (i.e. $\delta = -0.95$), $c = 0, -10, -50$ and two different sample sizes. Regression errors are conditionally homoscedastic. All tests are more powerful when the persistence in stronger and sample size larger, as expected, with IVX attaining a better performance. Figure 3 next, reports rejection probabilities for $\hat{T}$ for the fractional case. Again our limit theory is corroborated since superior performance is attained in situations where persistence is stronger and sample sizes larger.

Overall CTLS tests appear to have reasonably good sample size performance. IVX tests appear to be more powerful, nevertheless the CTLS procedures under consideration are readily available for fractional predictors, nonlinear regressions and in situations where there is conditional heteroscedasticity in the regression error. Some preliminary simulations show that IVX also has good performance in the fractional case. This can be partly explained by Theorem 3.2 of Duffy and Kasparis (2018) that yields basic limit theory for functionals of MI processes driven my long memory innovations. A formal investigation of IVX methods in presence of fractional processes is under development by the authors of the aforementioned work.

Table 2: Empirical Size of FP CTLS Tests (nominal size 5%; NI regressor, cond. homoscedastic regression errors)

δ		-0.95				-0.5				0				0.5				0.95			
	n	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS
$c=0$	250	0.084	0.089	0.059	0.278	0.059	0.062	0.056	0.117	0.051	0.055	0.050	0.053	0.061	0.063	0.056	0.113	0.087	0.091	0.061	0.295
	500	0.077	0.080	0.062	0.287	0.059	0.061	0.054	0.114	0.054	0.054	0.054	0.054	0.060	0.061	0.058	0.116	0.080	0.084	0.055	0.279
	750	0.076	0.078	0.058	0.272	0.059	0.058	0.052	0.109	0.052	0.052	0.050	0.051	0.059	0.061	0.055	0.111	0.080	0.081	0.057	0.277
	1000	0.070	0.069	0.053	0.278	0.054	0.056	0.051	0.111	0.049	0.049	0.051	0.053	0.059	0.060	0.050	0.108	0.075	0.077	0.053	0.277
δ		-0.95				-0.5				0				0.5				0.95			
	n	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS
$c=-5$	250	0.061	0.068	0.062	0.116	0.051	0.054	0.056	0.072	0.050	0.052	0.050	0.051	0.057	0.063	0.059	0.074	0.068	0.074	0.066	0.123
	500	0.060	0.064	0.063	0.117	0.051	0.053	0.059	0.073	0.051	0.053	0.052	0.054	0.056	0.057	0.057	0.071	0.062	0.065	0.058	0.116
	750	0.063	0.066	0.060	0.116	0.058	0.060	0.059	0.070	0.056	0.058	0.056	0.053	0.059	0.059	0.058	0.073	0.065	0.067	0.062	0.119
	1000	0.058	0.058	0.060	0.116	0.049	0.051	0.054	0.066	0.047	0.048	0.050	0.051	0.050	0.052	0.052	0.066	0.059	0.061	0.058	0.115
δ		-0.95				-0.5				0				0.5				0.95			
	n	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS
$c=-10$	250	0.058	0.064	0.062	0.086	0.051	0.056	0.055	0.063	0.049	0.056	0.051	0.052	0.056	0.062	0.057	0.063	0.063	0.069	0.065	0.090
	500	0.058	0.060	0.063	0.088	0.051	0.052	0.058	0.065	0.047	0.049	0.052	0.052	0.050	0.054	0.055	0.060	0.056	0.059	0.057	0.085
	750	0.058	0.059	0.060	0.087	0.058	0.059	0.056	0.064	0.055	0.056	0.056	0.053	0.056	0.058	0.055	0.062	0.058	0.063	0.062	0.088
	1000	0.053	0.056	0.058	0.084	0.049	0.051	0.053	0.059	0.046	0.048	0.050	0.051	0.049	0.047	0.051	0.058	0.054	0.056	0.058	0.088
δ		-0.95				-0.5				0				0.5				0.95			
	n	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS
$c=-20$	250	0.056	0.062	0.060	0.069	0.052	0.059	0.051	0.057	0.051	0.056	0.050	0.050	0.055	0.061	0.055	0.058	0.061	0.066	0.060	0.071
	500	0.054	0.055	0.060	0.072	0.050	0.051	0.054	0.058	0.048	0.050	0.051	0.052	0.049	0.051	0.055	0.058	0.053	0.059	0.056	0.067
	750	0.053	0.053	0.059	0.071	0.056	0.059	0.060	0.060	0.052	0.057	0.056	0.053	0.056	0.057	0.055	0.058	0.057	0.059	0.062	0.074
	1000	0.052	0.055	0.057	0.071	0.047	0.049	0.050	0.056	0.048	0.049	0.048	0.049	0.048	0.050	0.049	0.053	0.052	0.054	0.055	0.070
δ		-0.95				-0.5				0				0.5				0.95			
	n	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS	$\check{T}$	$\hat{T}$	IVX	OLS
$c=-50$	250	0.053	0.059	0.054	0.058	0.052	0.058	0.050	0.051	0.049	0.058	0.049	0.049	0.052	0.060	0.050	0.053	0.055	0.062	0.055	0.058
	500	0.052	0.055	0.054	0.059	0.052	0.053	0.051	0.053	0.048	0.051	0.047	0.048	0.050	0.053	0.050	0.050	0.053	0.056	0.055	0.059
	750	0.051	0.053	0.059	0.064	0.053	0.055	0.055	0.055	0.053	0.056	0.053	0.052	0.057	0.056	0.056	0.058	0.057	0.059	0.059	0.063
	1000	0.054	0.056	0.055	0.061	0.051	0.054	0.053	0.053	0.050	0.051	0.050	0.050	0.050	0.052	0.049	0.050	0.051	0.053	0.053	0.058

$\check{T}$: CTLS test statistic for conditionally homoscedastic errors; $\hat{T}$: CTLS test statistic for conditionally heteroscedastic errors

Table 3: Empirical Size of FP-CTLS tests: $\hat{T}$
(nominal size 5%; NI regressor, GARCH(1,1) regression errors)

	$c = 0$					$c = -5$				
δ	-0.95	-0.5	0	0.5	0.95	-0.95	-0.5	0	0.5	0.95
n=250	0.083	0.060	0.052	0.062	0.087	0.062	0.054	0.050	0.060	0.069
500	0.077	0.057	0.050	0.061	0.077	0.060	0.051	0.051	0.055	0.059
750	0.070	0.059	0.052	0.058	0.070	0.060	0.056	0.058	0.057	0.061
1000	0.065	0.054	0.048	0.056	0.069	0.054	0.051	0.045	0.049	0.055
	$c = -10$					$c = -20$				
δ	-0.95	-0.50	0.00	0.50	0.95	-0.95	-0.50	0.00	0.50	0.95
n=250	0.057	0.053	0.055	0.058	0.062	0.056	0.054	0.056	0.056	0.058
500	0.055	0.049	0.048	0.050	0.055	0.053	0.049	0.047	0.049	0.053
750	0.055	0.055	0.055	0.057	0.055	0.050	0.051	0.053	0.056	0.055
1000	0.052	0.049	0.048	0.046	0.053	0.052	0.048	0.045	0.047	0.053

Table 4: Empirical Size of FP-CTLS tests: $\hat{T}$ (nominal size 5%; fractional regressor, cond. homoscedastic regression errors)

		$d = 0.25$			$d = 0.5$			$d = 0.75$			$d = 0.8$		
	δ	-0.95	-0.5	0	-0.95	-0.5	0	-0.95	-0.5	0	-0.95	-0.5	0
CTLS	n=250	0.053	0.052	0.053	0.056	0.051	0.049	0.067	0.053	0.049	0.071	0.056	0.049
	500	0.052	0.052	0.050	0.055	0.050	0.047	0.064	0.053	0.049	0.069	0.054	0.052
	750	0.051	0.055	0.056	0.056	0.055	0.055	0.066	0.057	0.055	0.067	0.057	0.056
	1000	0.051	0.052	0.049	0.051	0.049	0.049	0.059	0.052	0.050	0.061	0.052	0.050
OLS	n=250	0.050	0.052	0.052	0.074	0.059	0.053	0.158	0.085	0.051	0.184	0.093	0.052
	500	0.052	0.050	0.048	0.072	0.055	0.051	0.161	0.085	0.054	0.184	0.091	0.055
	750	0.052	0.051	0.052	0.068	0.058	0.053	0.155	0.081	0.053	0.178	0.086	0.051
	1000	0.050	0.048	0.049	0.067	0.053	0.047	0.155	0.077	0.049	0.183	0.086	0.049
		$d = 0.9$			$d = 1$			$d = 1.1$			$d = 1.2$		
	δ	-0.95	-0.5	0	-0.95	-0.5	0	-0.95	-0.5	0	-0.95	-0.5	0
CTLS	n=250	0.077	0.057	0.051	0.084	0.059	0.051	0.089	0.060	0.052	0.089	0.063	0.053
	500	0.073	0.058	0.052	0.077	0.059	0.054	0.081	0.063	0.054	0.082	0.061	0.051
	750	0.073	0.058	0.054	0.076	0.059	0.052	0.078	0.060	0.051	0.079	0.061	0.051
	1000	0.066	0.054	0.050	0.070	0.054	0.049	0.072	0.055	0.050	0.072	0.054	0.051
OLS	n=250	0.235	0.107	0.053	0.278	0.117	0.053	0.308	0.121	0.052	0.325	0.126	0.052
	500	0.242	0.102	0.054	0.287	0.114	0.054	0.319	0.120	0.055	0.337	0.123	0.056
	750	0.230	0.098	0.052	0.272	0.109	0.051	0.301	0.117	0.051	0.322	0.119	0.053
	1000	0.229	0.102	0.053	0.278	0.111	0.053	0.310	0.118	0.054	0.327	0.120	0.055

Figure 2: Empirical Power of CTLS-FP tests: $\hat{T}$ (5% nominal size; $\delta = -0.95$; NI regressor, cond. homoscedastic regression errors)

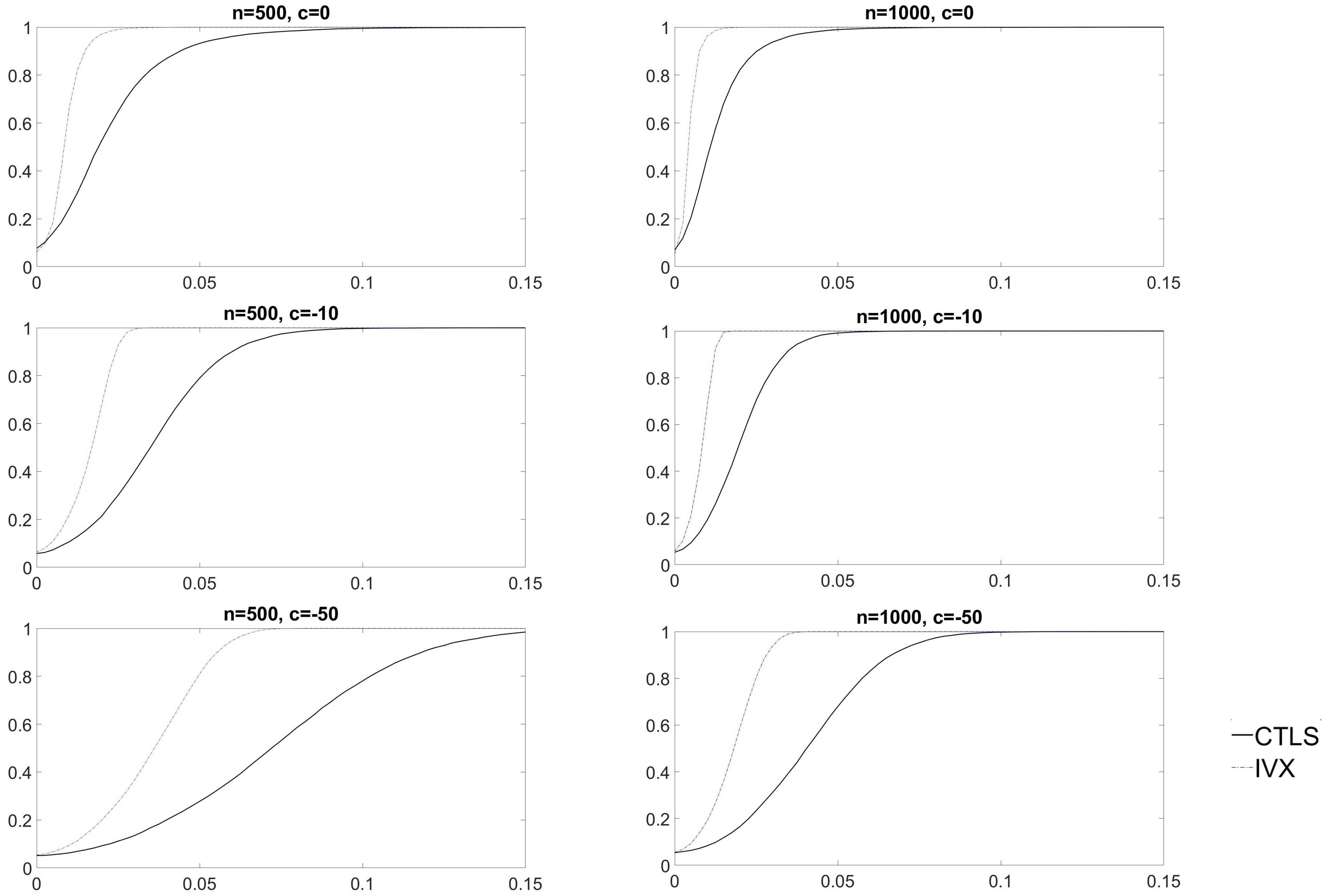
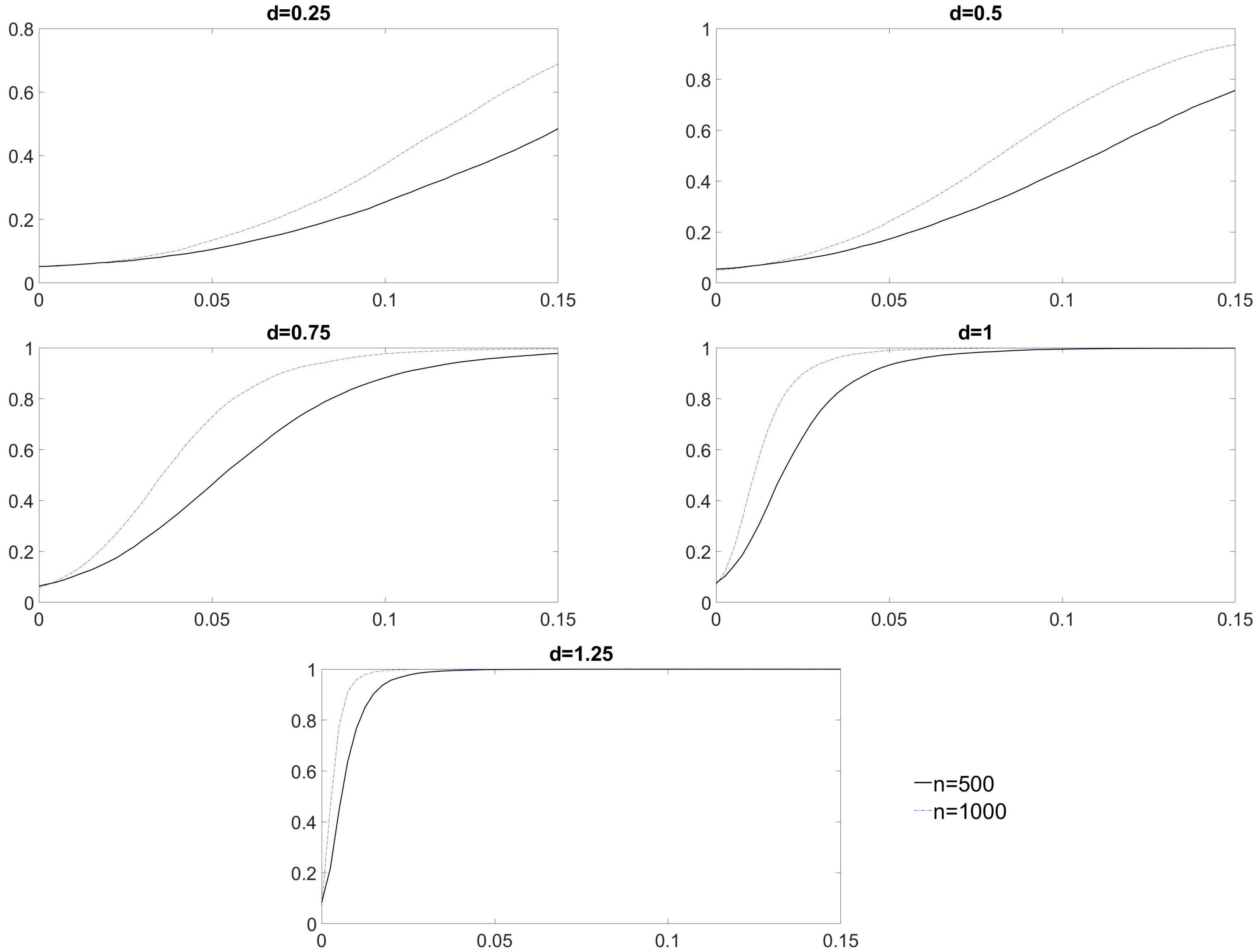


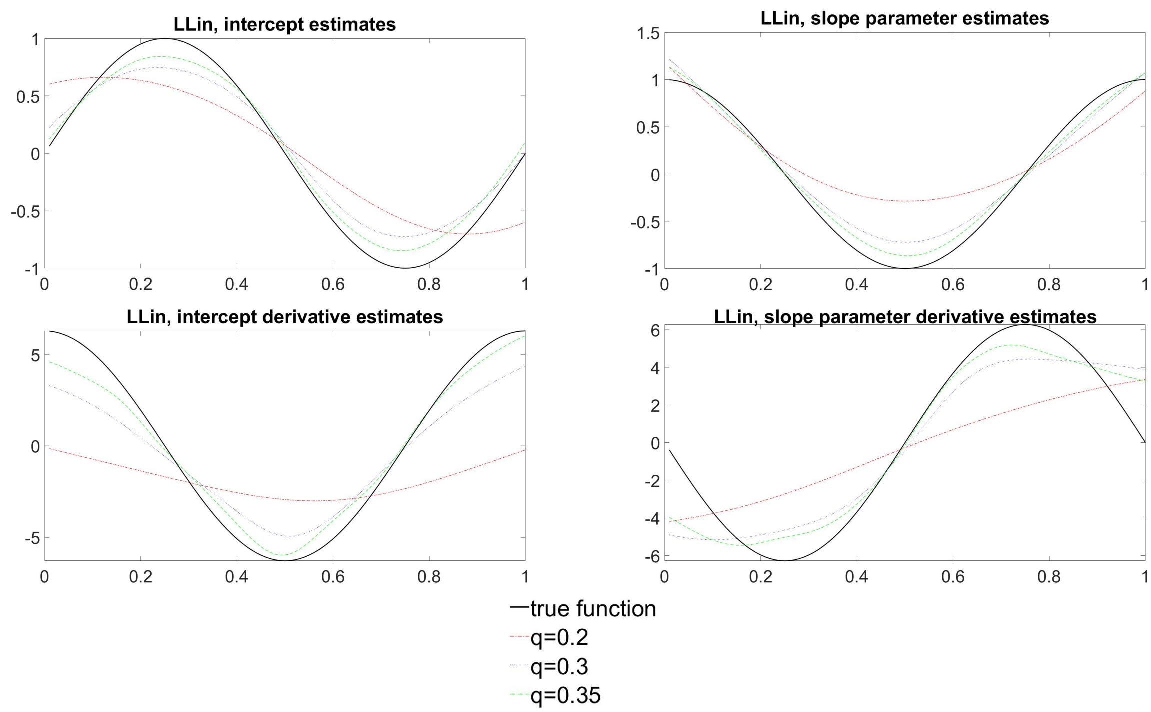
Figure 3: Empirical Power of CTLS-FP tests: $\hat{T}$
(5% nominal size; $\delta = -0.95$; fractional regressor, cond. homoscedastic regression errors)



Simulations for TVP models. We next report results for the finite sample performance of CTLS_1 based tests in the context of predictive regressions as per (39). 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 stationary fractional predictors of memory parameter $d = 0.35$ and $d = 0.45$. We also consider the case $d = 0.55$ which is slightly above the nonstationarity threshold ($d = 0.5$) that determines the minimal value of the memory parameter for which the limit distribution of the tests is $\mathbf{N}(0,1)$.[24] For nonstationary predictors, the CTLS_1 estimators under consideration do not possess mixed Gaussian limit distribution and therefore some size distortion in likely. It is reasonable to expect size distortions become more severe for larger values of the memory parameter. 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.

Figure 4: LLin TVP estimates
($\delta = -0.95$; fractional regressor $d = 0.45$, $n = 1000$, GARCH(1,1) regression errors)



[24] 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.

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. Nevertheless, the aforementioned trade-off is more subtle for non-parametric methods (e.g. CTLS for TVP models) than for semi-parametric methods (e.g. IVX, CTLS for FP models, etc). Other things being equal, for CTLS methods larger values of c_n (under-smoothing) result in better size control, while smaller values of values of c_n (over-smoothing) result in better power, because TVP estimators attain faster convergence rates in the latter case. There are however situations where under-smoothing may result in both better size and power. 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 4 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 4 that when over-smoothing is employed (e.g. $q = 0.2$) sudden changes in the TVPs are smoothed out. Another finding from Figure 4, that is worth noting, is that the derivative estimators appear to exhibit non trivial asymptotic bias at boundary points i.e. for $\tau \approx 0, 1$. This appears to impact inference related to TVP derivatives, and for this reason this issue will be revisited later.

For the finite sample evaluation the tests we consider the following two possibilities for the bandwidth parameter $c_n = n^q$

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

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

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

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 14). On the other hand smaller variability in the intercept function is associated with smaller asymptotic bias (cf. condition (e) of Theorems 7 and 8). 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. 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 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 5 and 6 report the empirical size of LLev and LLin based tests for the non predictability hypothesis for sample sizes $n = 500$ and $n = 1000$. We only consider $\delta = -0.95$ i.e. strong endogeneity. Size (vertical axis) is plotted against various values of $\tau \in (0, 1)$ (horizontal axis). It can be seen that size control is reasonably good with small oversizing when $d = 0.45$ and moderate oversizing when $d = 0.55$. Additional simulations, not reported here, show that when the intercept is fixed over time, size is slightly better than that in Figures 5 and 6. Moreover, for smaller values of d and $|\delta|$ preliminary simulations show that empirical size is closer to the nominal one.

The empirical power of both tests is reported in Figures 7 and 8, 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, in particular 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.

Finally, we 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 for 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 (39) 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 14) which is likely to result in some size distortions. Figure 9 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.

We conclude with the empirical power of the test. Figure 10 reports the rejection frequency of the test for the case where 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 appears to be less powerful than the predictability test considered earlier. Notably, there is 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 $\tau = \{0, 0.5, 1\}$. Nevertheless at the boundary points of its domain power is very poor. This likely due asymptotic bias in derivative estimation at boundary points (cf. Figure 4). Hence, the test appears to be quite conservative in terms of power, when there is substantial variation in the parameter at boundary points, nevertheless it can be easily implemented in conjunction with the predictability test. Possibly, better performance could be 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 5: Empirical Size of CTLS-TVP tests against τ: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 500$; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

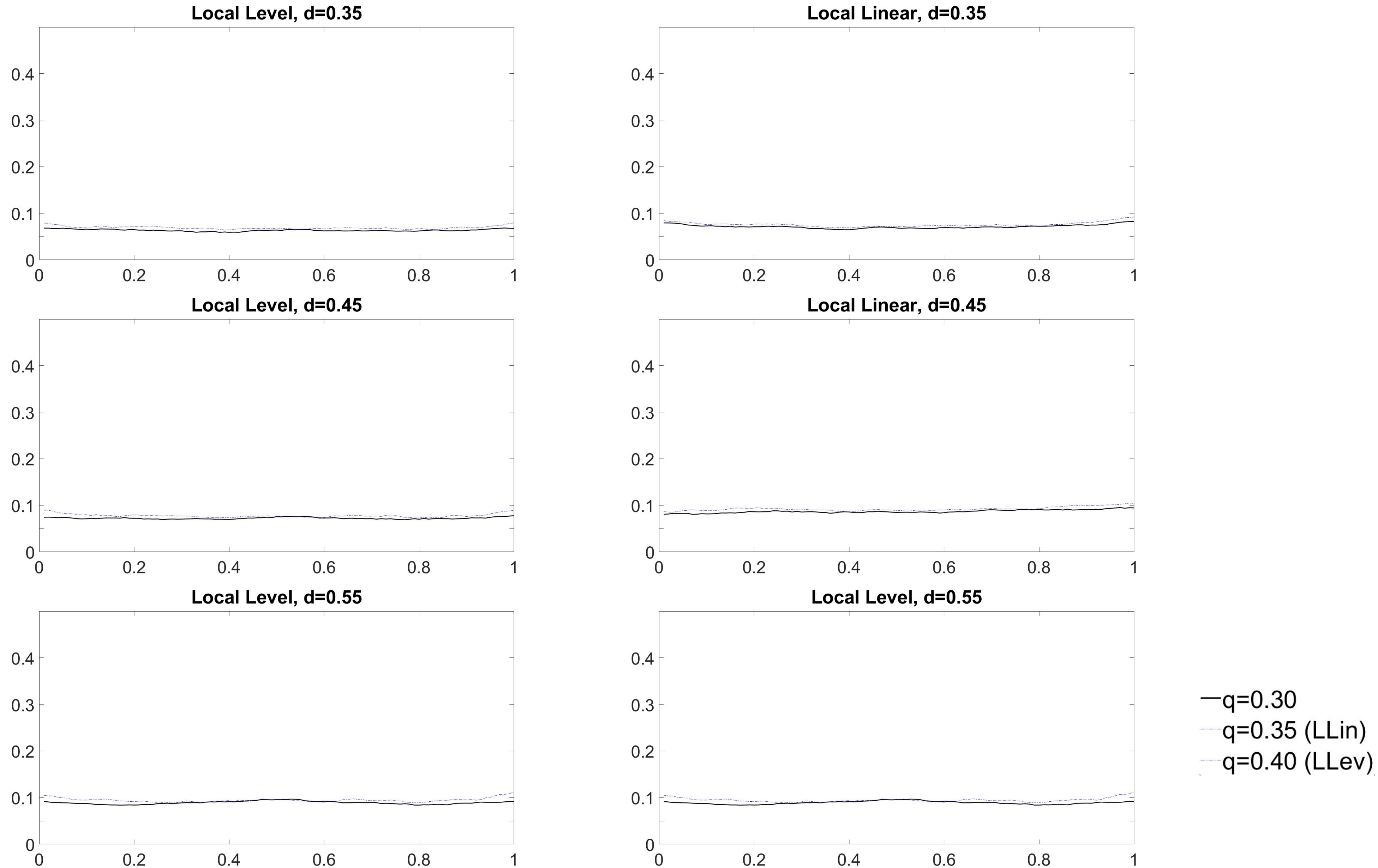


Figure 6: Empirical Size of CTLS-TVP tests: $H_0 : \beta(\tau) = 0$
(5% nominal size; $n = 1000$; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

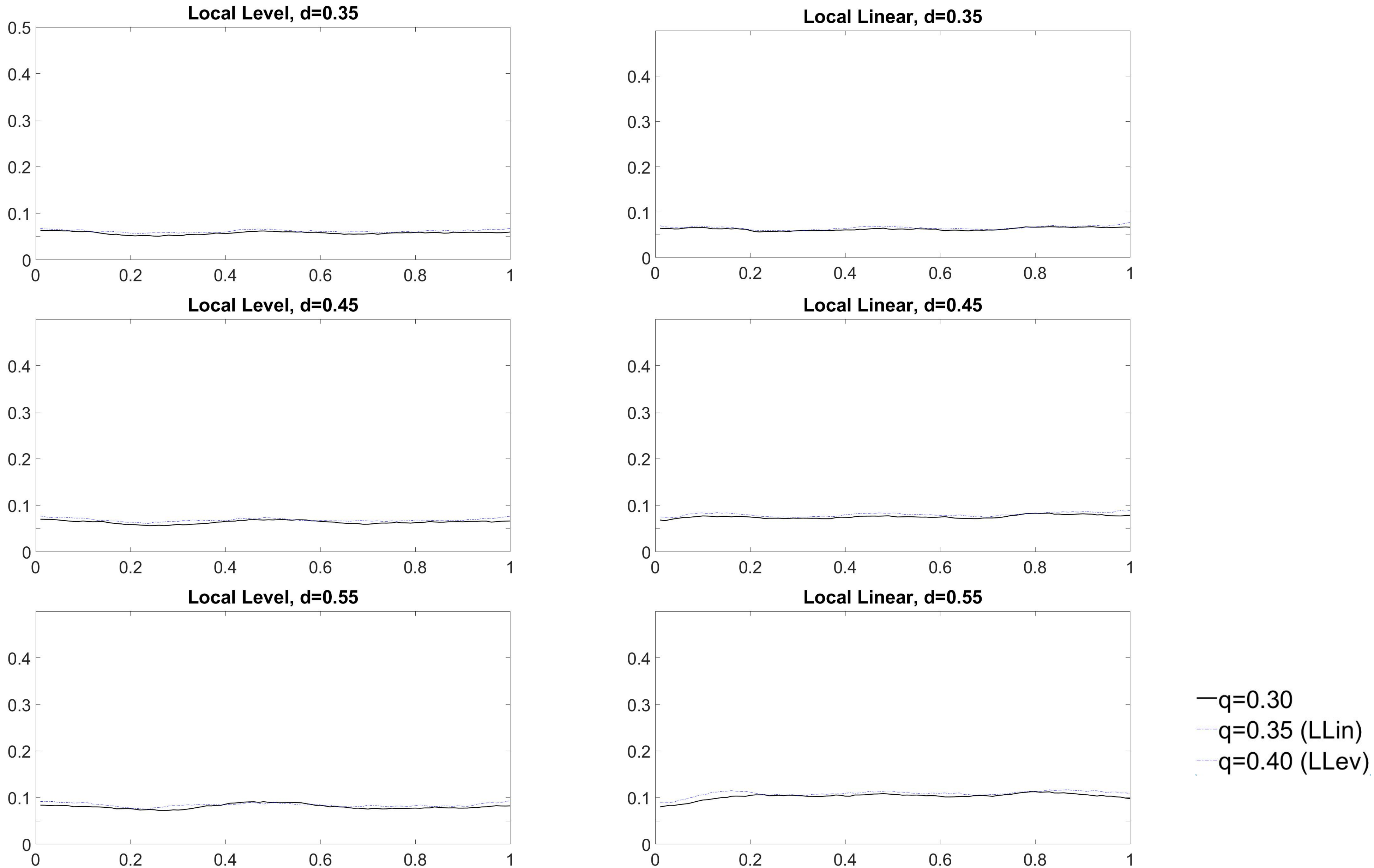


Figure 7: Empirical Power of CTLS-TVP tests: $H_1 : \beta(\tau) = b \cdot \cos(2\pi\tau)$
(5% nominal size; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

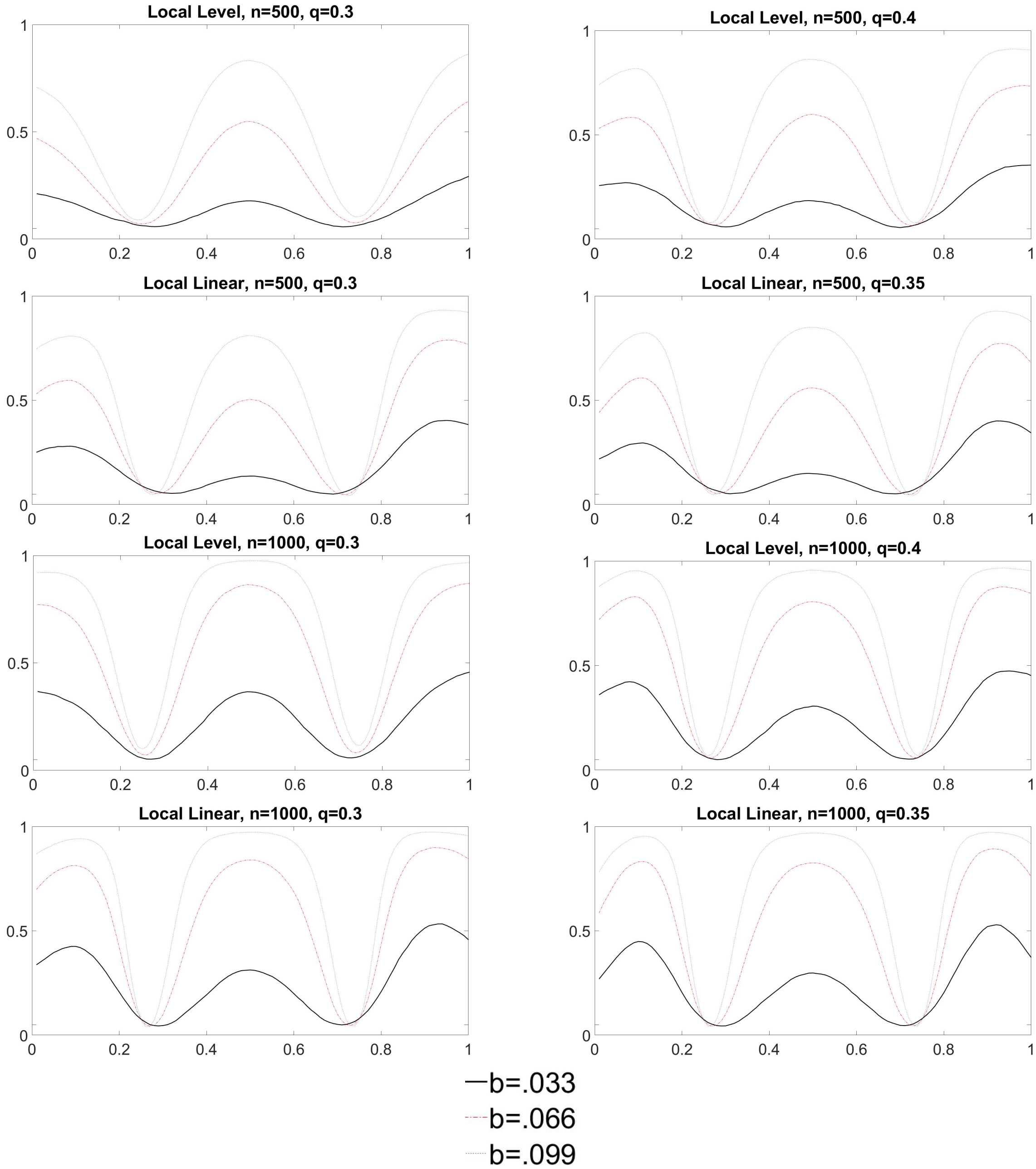


Figure 8: Empirical Power of CTLS-TVP tests: $H_1 : \beta(\tau) = b \cdot \{1 + \exp[-30(\tau - 0.5)]\}^{-1}$ (5% nominal size; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

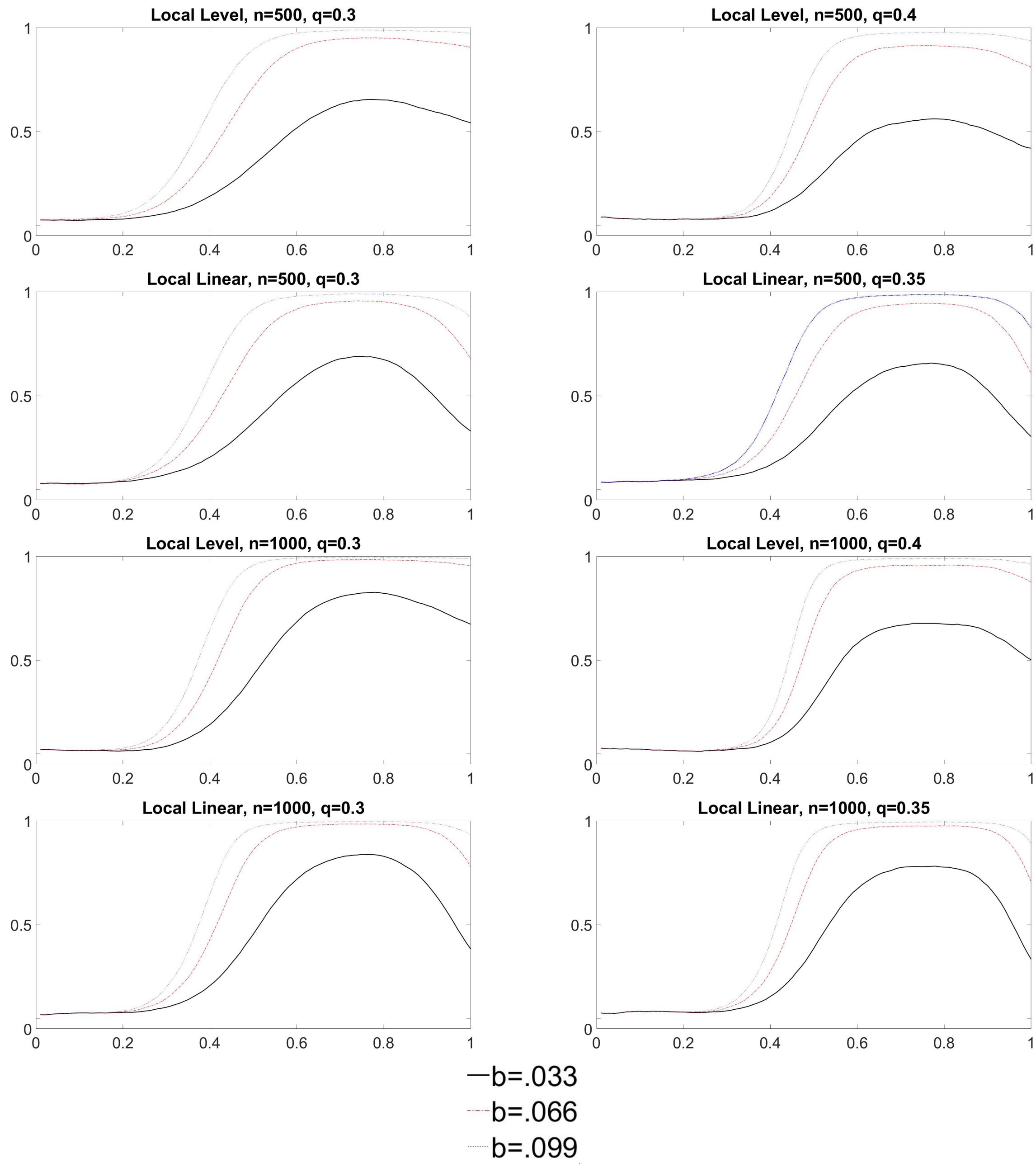


Figure 9: Empirical Size of CTLS-TVP tests: $H_0 : \partial\mu(\tau)/\partial\tau = 0$
(5% nominal size; $\delta = -0.95$; fractional regressor, GARCH(1,1) regression errors)

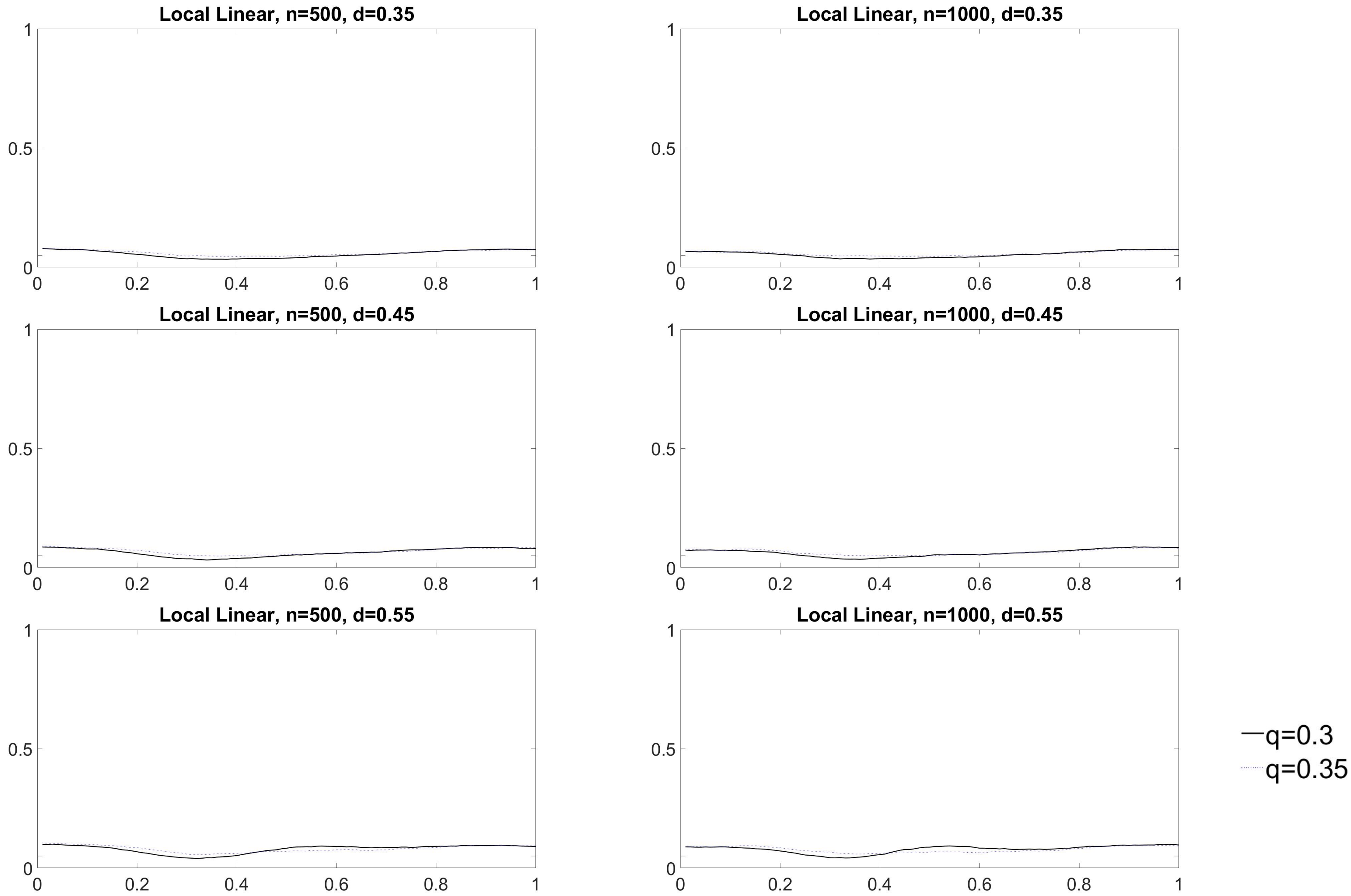
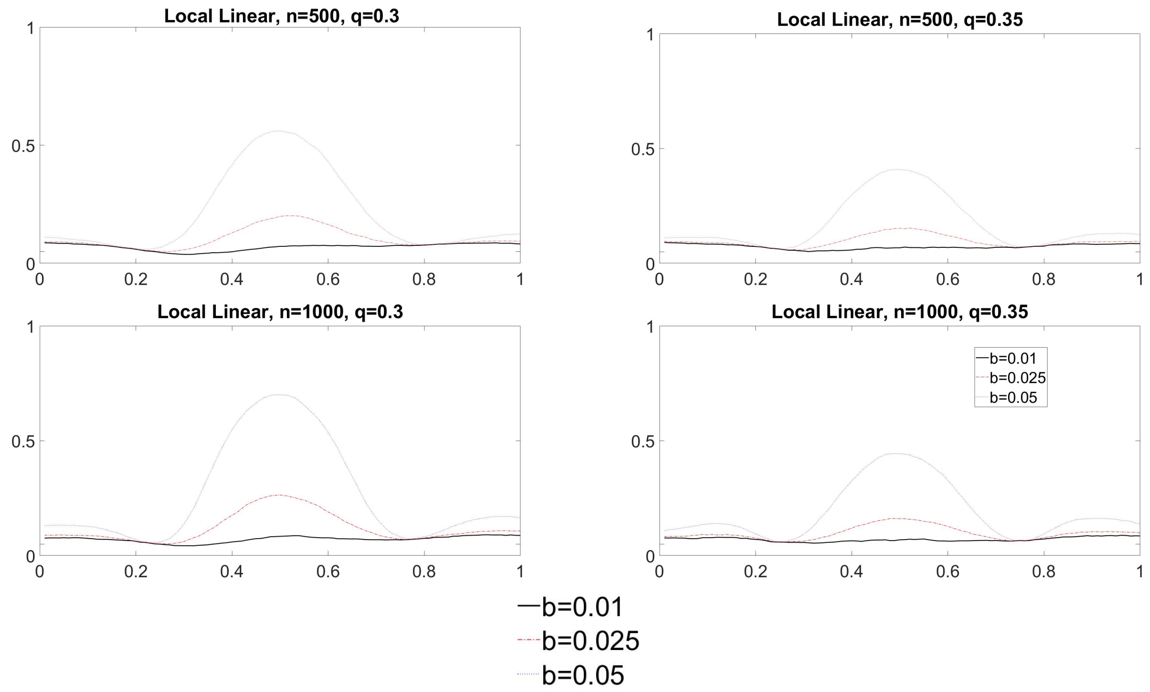


Figure 10: Empirical Power of CTLS-TVP tests: $H_1 : \partial\mu(\tau)/\partial\tau = 2\pi b \cdot \cos(2\pi\tau)$ (5% nominal size; $\delta = -0.95$; $d = 0.45$, GARCH(1,1) regression errors)



5 Application to the predictability of stock returns

A vast literature in empirical finance and econometrics is devoted to the investigation of the hypothesis that stock returns can be predicted with publicly available information. There are two main approaches in this area. First, certain studies (e.g. Welch and Goyal, 2008; Bollerslev et al., 2013) investigate the *in sample* or *out of sample* predictive ability of predictive regressions with the aid of some forecast adequacy test (e.g. McCraken, 2007) or some goodness of fit statistic (e.g. R^2). Typically predictive regressions take the form

$$r_k = \mu + \beta x_{k-1} + e_k, \tag{43}$$

where r_k are stock returns relating to some stock index, x_k some predictive variable and e_t a martingale difference regression error. Another approach is to test the predictability hypothesis $H_0 : \beta = 0$ using appropriate (in sample) inferential procedures (e.g. Valkanov, 2003; Lewellen, 2004; Campbell and Yogo, 2006; Hjalmarsson (2011), Kostakis et al., 2015). Usually some financial variable (e.g. dividend yield, earnings to price ratio, book to market ratio, realised variance) or some macroeconomic variable (e.g. inflation) is considered as a possible predictor for future returns. Many studies in this area investigate the predictability hypothesis under the assumption that predictor is a stationary AR(1) processes driven by i.i.d. innovations, and employ techniques that are in general valid only under stationarity.[25] Nevertheless, there is strong evidence that in certain data sets various financial and macroeconomic variables are persistent i.e. consistent with stationary long memory processes (see e.g. Bollerslev et al., 2013) or nonstationary long memory processes (e.g. see Kostakis et al., 2015; Table 4). Christensen and Nielsen (2007) (see also Chistensen and Nielsen, 2006), Bollerslev et al. (2013), Bandi, Perron, Tamoni and Tebaldi (2019) develop methods that allow for stationary long memory predictors. Campbell and Yogo (2006), Hjalmarsson (2011) consider *conservative* testing procedures that allow for a NI predictor. The latter two papers utilise Bonferroni bounds with confidence intervals based on the inversion of unit root tests (see also Cavanagh et al., 1995). Phillips (2014) shows that testing procedures based on the inversion of unit roots provide good robustification for local deviations from unity, but do not perform that well under larger deviations (e.g. Mildly Integrated or stationary data). The most recent work of Kostakis et al. (2015) investigates the predictability hypothesis utilising the IVX method of Magdalinos and Phillips (2009). Magdalinos and Phillips (2009) provide conventional inference in regressions with NI or MI covariates. Kostakis et al. (2015) demonstrate that the IVX method is also valid under larger deviations from unity i.e. when the data are generated short memory linear processes. Further, they provide a finite sample correction for IVX based test statistics that relates to intercept demeaning. The IVX method has been also utilised by a number of other studies in the context of predictive regressions. Gonzalo and Pitarakis (2012) investigate regime specific predictability in the context of threshold regressions while Demetrescu et al. (2020) examine episodic predictability in TVP predictive regressions. Both of the aforementioned papers develop predictability tests that utilise IVX instrumentation. A comprehensive review of the econometric methodology utilised in this area can be found in Phillips (2015).

In this section apply the methods of Section 3.1 and 3.2 to investigate the return predictability hypothesis. In particular, we consider the specification of (15) with a general predictor and (22) with a stationary predictor (e.g. fractional $|d| < 1/2$). Both models allow for nonlinear regressions functions with the latter providing a more flexible functional form due to TVPs. Therefore, the former specification is more general with respect to regression space while the latter is more general in terms of functional form.

Functional form is an aspect of modelling that can potentially address *misbalancing* in predictive regressions for returns. A well known stylistic fact about short term returns is that they exhibit very weak persistence with paths closely resembling those of martingale differences (i.e. $I(d)$, with $d = 0$). On the other hand most commonly used predictors are very persistent exhibiting either stationary or non stationary long memory. Misbalancing is an important issue that has received relative little attention in the predictability literature.

[25] For instance Stambaugh (1999), Amihud and Hurvich (2004), Chen and Deo (2009), assume that the predictor is a stationary AR(1) process driven by i.i.d. or Gaussian i.i.d. errors.

As emphasised by Phillips (2015) -see also Kasparis (2011)-, misbalancing may result to asymptotically vanishing estimators. For instance if $r_k \sim I(d)$ with $d < 1/2$ (stationary long memory) and $x_k \sim I(d)$ with $d > 1/2$, then then OLS estimator for β in (43) is $\tilde{\beta} \to_P 0$.

A departure from the usual linear in levels specification can potentially address misbalancing issues. For instance Christensen and Nielsen (2007) and Bollerslev et al. (2013) consider predictive models for returns where the systematic part of the model is of the form $\mu + g(x_{k-1})$ with g being the fractional difference operator $(I - L)^d$ and d the memory parameter of volatility (x_k). Note that in this case $g(x_{k-1})$ is $I(0)$. A similar but more general approach is considered by Andersen and Varneskov (2020) who assume predictive relationships between the short memory components for fractional series of the form

$$(1-L)^{d_y} y_t = \mu + \mathcal{B}'(1-L)^{d_x} \mathbf{x}_{t-1} + \eta_t,$$

where $y_t \sim I(d_y)$ and $\mathbf{x}_t \sim I(d_x)$ are scalar and vector processes respectively, possibly nonstationary. The regression parameters are estimated by a narrow band type of method, that also trims frequencies around zero in finite samples. The resultant estimator attains semi-pametric rates. In particular, it is sub-$\sqrt{n}$ consistent when the regression error is η_t is $I(0)$ but can be faster when η_t is of negative memory. Similarly to other spectral LS methods -e.g. Robinson and Hualde and Christensen and Nielsen (2006)- this approach requires plug-in estimates for the memory parameters. Marmer (2007), Kasparis (2010), Kasparis, Andreou and Phillips (2015) and Phillips (2015) suggest that nonlinear regression functions can potentially addresses misbalancing issues. It is well known (e.g. Park and Phillips, 1999; 2001) that nonlinear transformations can significantly attenuate the signal of persistent processes. In fact, a transformed nonstationary process may exhibit a weaker signal than that of a stationary one. For example, for some measurable function g and x_k stationary we have $\sum_{k=1}^{n} g(x_k) = O_P(n)$, in general. On the other hand for $x_k \sim I(1)$ the following orders apply (e.g. see Park and Phillips, 2001; Berenguer-Rico and Gonzalo, 2014)

$$\sum_{k=1}^{n} g(x_k) = \begin{cases} O_P(n^{1+p}), \text{ for } g \text{ polynomial of order } p > -1 \\ O_P(n\ln(n)), \text{ for } g \text{ logarithmic} \\ O_P(n), \text{ for } g \text{ bounded} \\ O_P(\sqrt{n}), \text{ for } g \text{ integrable} \end{cases}$$

It can be readily seen from the orders shown above that certain nonlinear transformation of $I(1)$ processes, may exhibit a very weak signal that can be equal (bounded functions) or smaller (integrable and reciprocal functions) than that of a stationary process. These orders can be smaller when $x_t \sim I(d)$ with $1/2 < d < 1$. Figure 11 provides a graphical illustration of the effects of certain nonlinear transformations on the paths of an $I(1)$, process relatively to those of stationary GARCH(1,1) (i.e. a martingale difference process). It can be seen that the aggregated paths of a GARCH with those of a transformed nonstationary processes, may resemble those of martingale difference processes.

TVP and threshold regression models (Gonzalo and Pitarakis, 2012; Demetrescu et al. 2020) entail an alternative form of nonlinearity that can also address misbalancing. In particular, these type of models allow for *episodic* predictability events when some time or state variable is in some regime. For instance, Gonzalo and Pitarakis (2012) find evidence that stock returns can be predicted by financial ratios when there "*bad news*" i.e. inflation exceeds certain level. Transformations relating to threshold models and TVP specifications can also attenuate the signal of a persistent predictor, and produce sample paths similar to those shown in Figure 11.

In this Section we will investigate predictability of stock returns using the Welch and Goyal, 2018 data set. The returns variable (r_k) is constructed by taking log differences of the SP500 index. Further, we consider 4 alternative predictors: Dividend Yield (DY), Earrings-to-Price ration (EP), Book-to-Market (BM), and a realised variance (SVAR) variable (the sum of squared daily returns on the SP500). Moreover, 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 5 that SP500 returns closely resemble an $I(0)$ process in all frequencies, while the predictive variables are persistent, exhibiting either stationary long memory (SVAR in lower frequencies), or nonstationary long memory - particularly DY, EP and BM.

We investigate return predictability by utilising the inferential techniques of Section 3.1 and 3.2 for FP and TVP models respectively. The CTLS techniques for FP models allow for a very general regressor space, capable of handling all predictors under consideration. On the other hand CTLS techniques for TVP models allow only for stationary or close to stationary long memory. Therefore for will only consider the SVAR predictor for TVP models that appears to have memory characteristic closer to the permissible regression space. 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.

FP regressions. First, we consider FP regressions with DY, EP, BM and SVAR as a possible predictor, and five alternative regression functions (cf. eq. (15)):

$$f(x) = \left\{x, \ln(x), x^{0.25}, x^{0.5}, x^{0.75}\right\}.$$

In particular, we consider the linear specification that is widely used in practice and four

Figure 11: Paths of GARCH(1,1) Vs Transformations of $I(1)$

P1: $0.5x_k$; P2: $0.5\,|x_k|^{0.25}$; P3: $0.5\ln(|x_k|)$; $x_k - x_{k-1} \sim i.d.\mathbf{N}(0,1)$; GARCH(1,1) with param. $0.01, 0.45, 0.45$.

additional regression functions that exhibit reduced growth rates relative the linear one. Nonlinear transformations of various predictors such as logarithmic and square roots have been employed by a number of studies (see e.g. Lewellen, 2004; Bollerslev et al., 2013; Anderson and Varneskov, 2020). Specifications of reduced growth could alleviate issues of misbalancing. Misbalancing is more likely to be committed in situations where the predictor exhibits extreme persistence as in the case of DY that appears to be far more persistent than a unit root process, particularly in higher sampling frequencies.[26]

Table 6 reports the values of $\hat{T}$ for the hypothesis $H_0 : \beta = 0$. The alternative hypothesis can be either one-sided or two-sided. We utilise the CTLS test statistic of (20) that is

[26]The EWL estimates suggest that DY has memory parameters significantly above unity for all sampling frequencies. Further, EP annual observations appear to be more persistent than a unit root. Note that the LW estimates are slightly below unity, nevertheless it is well known that the LW estimator converges to unity for $d > 1$ i.e. it is inconsistent (cf. Phillips and Shimotsu, 2004).

Table 5: 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
DY	0.96	1.76	0.91	1.55	0.83	1.31
EP	0.92	1.22	0.79	0.85	0.72	0.81
BM	0.99	1.02	0.74	0.77	0.63	0.65
SVAR	0.53	0.53	0.46	0.47	0.33	0.35

capable of accommodating GARCH effects in the regression error. It can be seen that for monthly data, there is some evidence of predictability for EP (10% level for one-sided test) and BM (5% level for one sided) test. Interestingly, for the BM predictor $\hat{T}$ attains its maximal value for the reduced growth rate regression function $x^{0.25}$. The CTLS procedure suggests no predictability for quarterly data. Finally, for annual observations there is limited predictability evidence for EP (10% level for one-sided) and some robust findings for DY (5% level for one sided test). For the latter predictor, the maximal value of $\hat{T}$ is attained under the logarithmic specification. Overall, the strongest predictability evidence (significant at 5% level - one sided tests) are associated with the regression functions $f(x) = x^{0.25}, x^{0.5}$ for the BM variable and $f(x) = \ln(x)$ for DY.

Table 6: Values of $\hat{T}$ for the hypothesis $H_0 : \beta = 0$

$f(x)$	x	$\ln(x)$	$x^{0.25}$	$x^{0.5}$	$x^{0.75}$
Monthly					
DY ($n = 1775$)	1.135	0.639	0.85	1.08	1.16
EP ($n = 1775$)	1.637^{C}	1.024	1.099	1.464^{C}	1.615
BM ($n = 1173$)	1.482^{C}	-0.887	$1.843^{\mathrm{c,B}}$	$1.705^{\mathrm{c,B}}$	1.581^{C}
SVAR ($n = 1606$)	-0.47	N/A	0.037	-0.157	-0.338
Quarterly					
DY ($n = 591$)	-0.859	-0.613	-1.127	-1.006	-0.914
EP ($n = 591$)	-0.077	-0.279	-0.864	-0.555	-0.306
BM ($n = 391$)	0.535	0.829	0.044	0.228	0.393
SVAR ($n = 535$)	0.198	-0.322	0.189	0.154	0.164
Annual					
DY ($n = 147$)	1.58^{C}	$1.885^{\mathrm{c,B}}$	0.613	1.384^{C}	1.618^{C}
EP ($n = 147$)	1.177	1.333^{C}	0.462	1.094	1.266
BM ($n = 97$)	-0.307	1.036	-0.625	-0.517	-0.41
SVAR ($n = 133$)	-0.426	-0.153	-0.068	-0.22	-0.351

a, b, c: signficant at 1%, 5% and 10% level respectively for two-sided test
A, B, C: signficant at 1%, 5% and 10% level respectively for one-sided test
N/A: statistic cannot be computed due to near singularity of some matrix

TVP regressions. We next consider TVP models with SVAR as a predictor as per (22). The memory estimates for the SVAR variable are between 0.35 and 0.53. Recall that our theoretical results demonstrate that CTLS_1 inferential methods for TVP models are valid for predictors that exhibit stationary long memory. As remarked before (e.g. Remark 13), we expect that Theorems 9 and 10 are also hold for fractional $d = 1/2$ and MI processes. Further, the simulation experiment suggests the that proposed tests perform reasonably well for memory parameters slightly above $d = 1/2$.

We first provide estimates for $\mu(\tau)$ and $\beta(\tau)$, $\tau \in (0, 1]$ based on the LLev and the LLin CTLS_1 estimators (see Figure 12). 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 the maximal over-smoothing allowed, given the theoretical constrains[27], and provide good performance both in terms of size and power according to the simulation study. It can be seen from Figure 12 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, and equal 5% approximately. The value of the slope parameter for monthly returns appears to be overall negative while for medium and long run returns (i.e. quarterly and annual) the slope

[27] i.e. condition (e) in Theorem 7 and 8.

parameter varies 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 shown in Figure 13. 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 monthly 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.

It is worth comparing the findings for the SVAR predictor in the context of FP models to those for TVP models. All the FP-CTLS tests retain the null hypothesis for non predictability when SVAR is utitilised as a predictor. In fact, the values of $\left|\hat{T}\right|$ are very small in this case. The maximal value of $\left|\hat{T}\right|$ corresponds to a p-value equal 0.66, which is substantial. On the other hand all the CTLS_1 tests for TVP models suggest that there is evidence for predictability with respect to SVAR. This is discrepancy between FP and TVP models is expected in situations where time variation in regression parameters is neglected. As mentioned before (i.e. Section 3.3), parametric (e.g. OLS) and semi-parametric estimators (e.g. CTLS, IVX), under certain regularity conditions, converge to pseudo-true values of the form $\int_0^1 \beta(\tau)d\tau$ when there is time variation in the slope parameters. This integral functional is a chronological average of the TVP parameters. As a consequence, predictability episodes tend to be averaged out when FP empirical specifications are utilised. As a consequence, inferential procedures based on FP estimators exhibit poor power performance in situations of neglected time variations in the parameters.

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 (22). As explained in Section 3.3, 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 14 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). In practice, it could be the case that time variation in the intercept is more substantial than what these test suggest. First, note that inference based on derivative estimators yields less powerful tests because derivative estimators attain slower convergence rates than those attained by ordinary estimators. In particular, the divergence rate of the derivative based LLin test statistic is $O_p\left(\sqrt{n/c_n^3}\right)$ while its non derivative counterpart attains divergence rate of order $O_p\left(\sqrt{n/c_n}\right)$. Therefore, our findings on the time variation of the regression intercept are likely to be conservative.

Figure 12: Local Level/Linear TVP Estimates

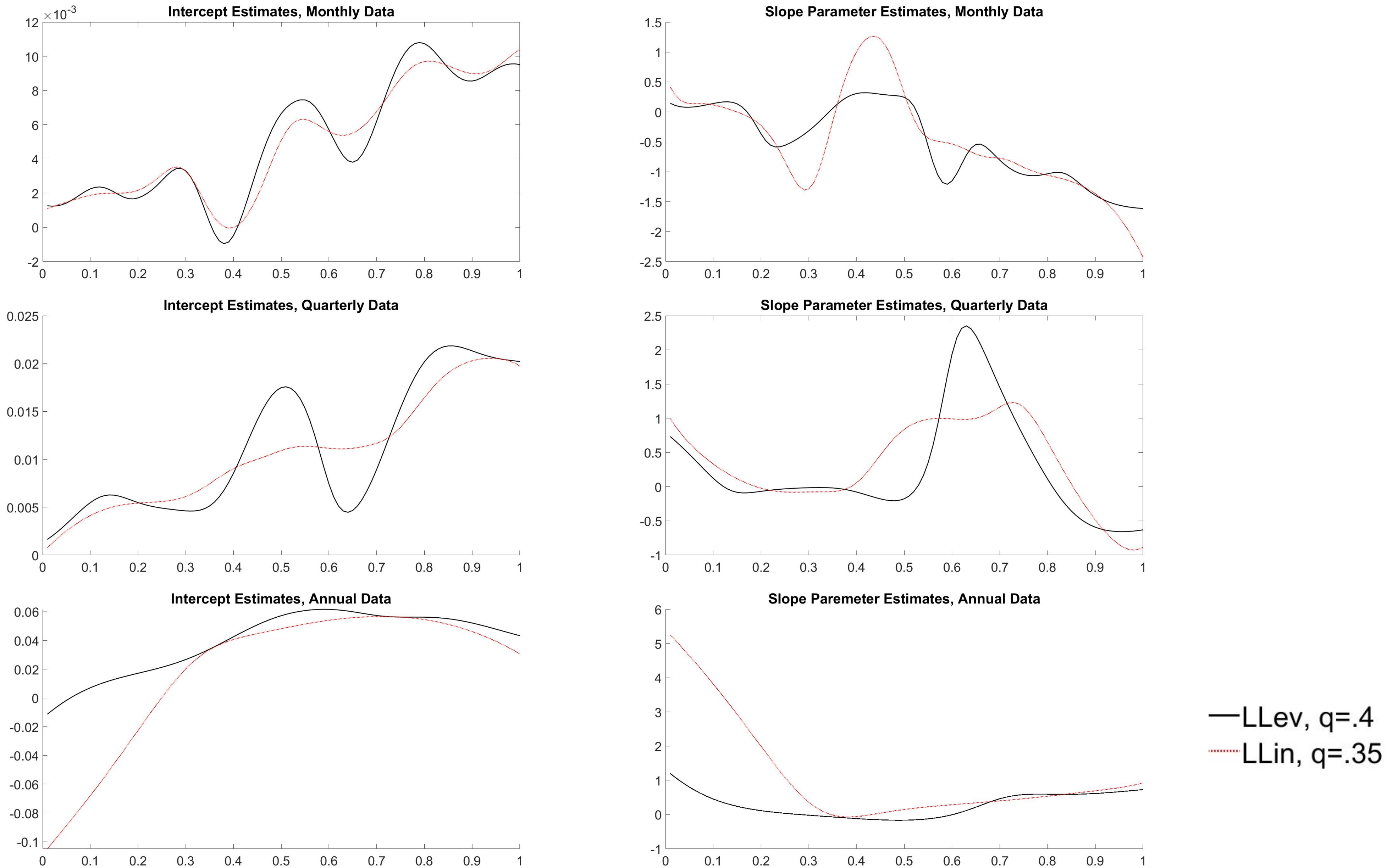


Figure 13: Local Level/Linear t-statistics for $H_0 : \beta(\tau) = 0$

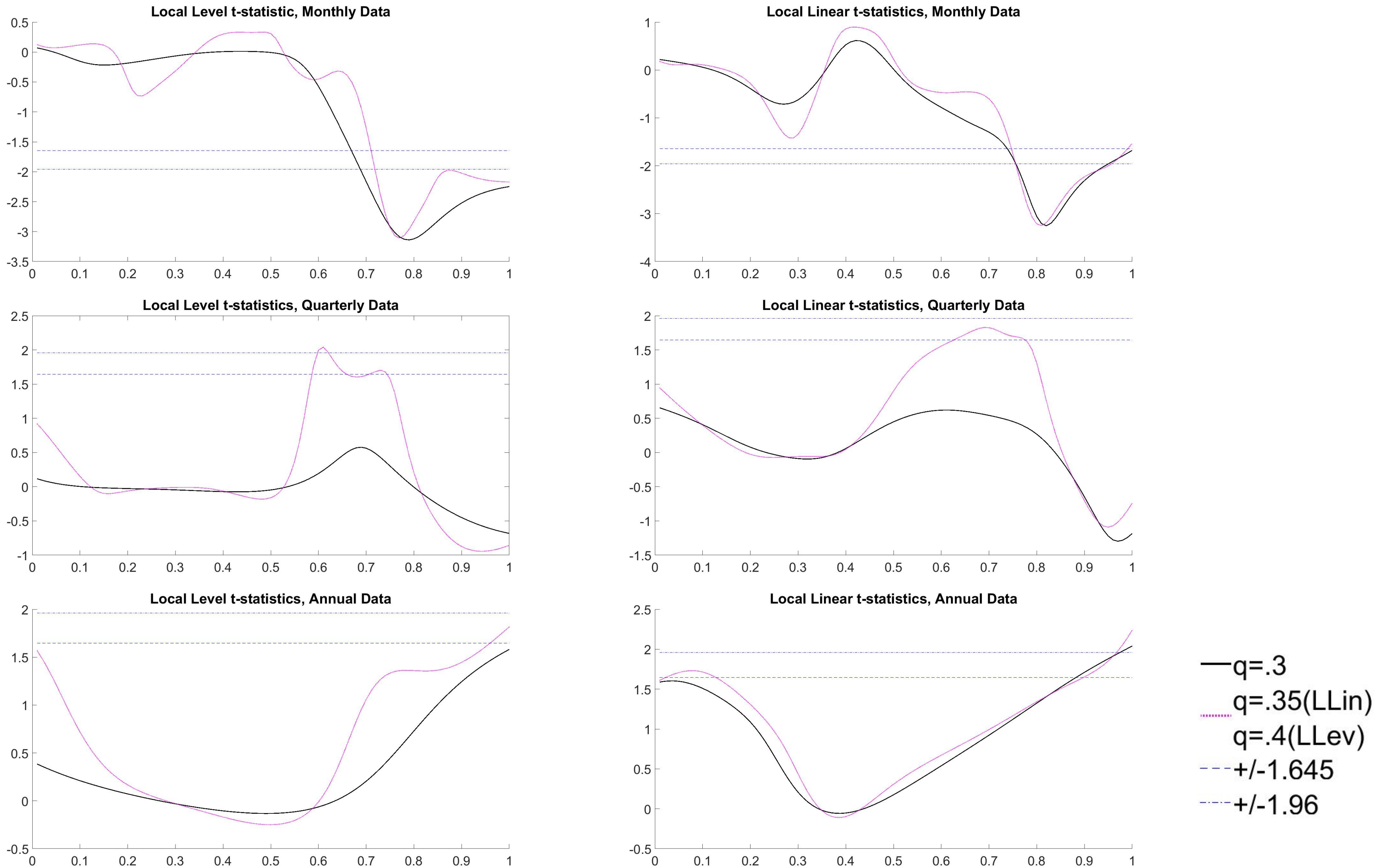
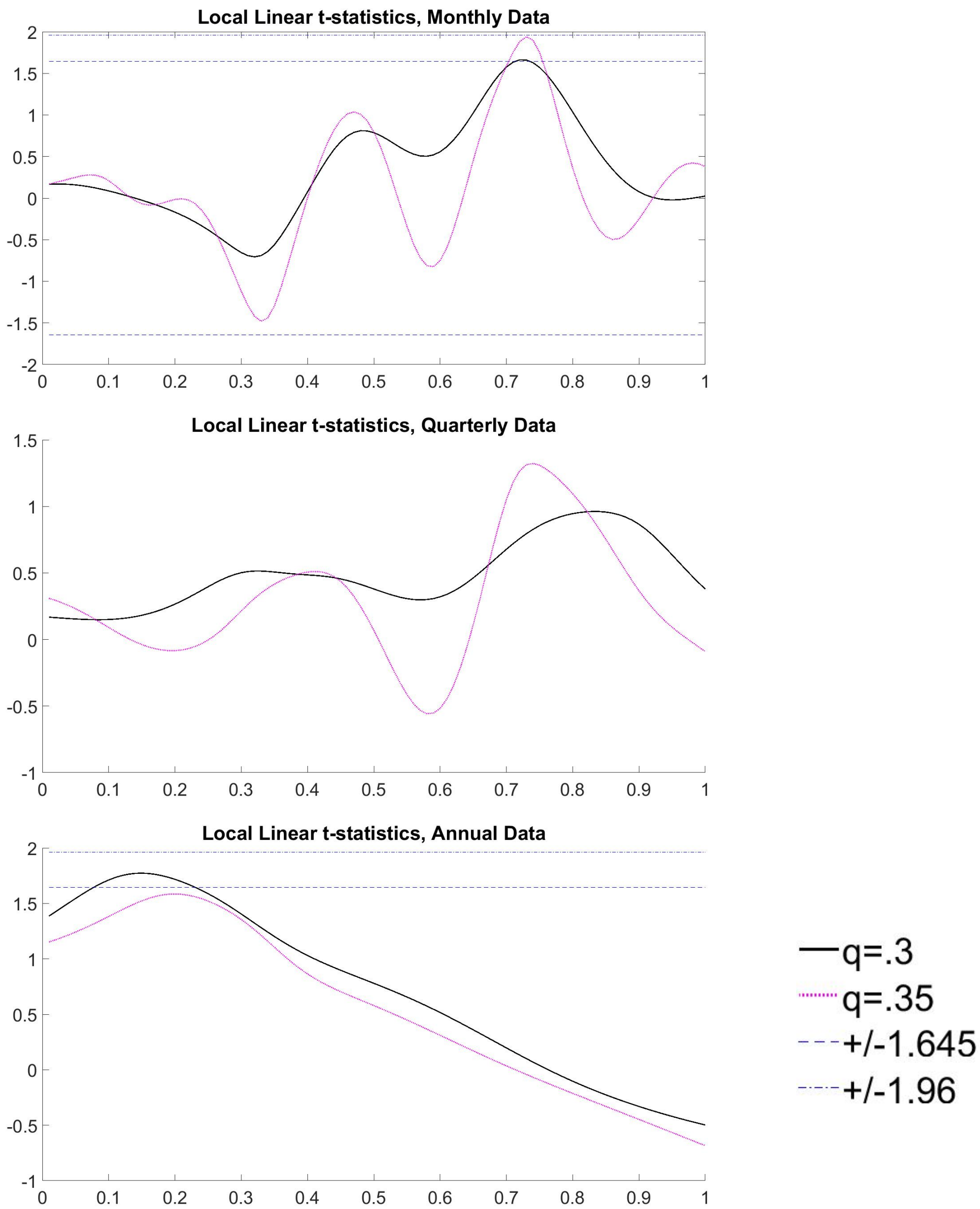


Figure 14: Local Linear t-statistics for $H_0 : \partial\mu(\tau)/\partial\tau = 0$



APPENDIX

Throughout the remaining paper, we assume that $C, C_0, C_1, C_2, ...$ are positive constants that may take a different value in each appearance and let $K_{kn} := \sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right]$ as in (16).

6 Proofs for Section 2

6.1 Preliminaries

We start with two preliminary lemmas, which provide significant extension to Lemma 4.1 of Hu, Phillips and Wang (2021) and include (8) and (10) as a corollary. The proofs of these two lemmas will be given in Sections 6.6 and 6.7, respectively.

Let $\{X_{n,k}\}_{k\geq 1, n\geq 1}$, where $X_{n,k}$, be a vector random array. When there is no confusion, we also use the notation $X_{nk} = X_{n,k}$. Let $\{v_k\}_{k\geq 1}$ be a sequence of random variables, and $G(q))$ and $K(x)$ be Borel functions on $\mathbb{R}$. For $0 < \tau_1 < \tau_2 < ... < \tau_l < 1$, set

$$S_{n,l} = \frac{c_n}{n}\sum_{k=1}^{n} G(X_{nk})v_k \frac{1}{l}\sum_{j=1}^{l} K\big[c_n(k/n - \tau_j)\big],$$

where $\{c_n\}_{n\geq 1}$ is a sequence of positive constants. Our first result investigates the asymptotics of $S_{n,l}$.

Lemma 1. *Suppose that*

(a) there is a continuous limiting process X_t such that $X_{n,[nt]} \Rightarrow X_t$ on $D_{\mathbb{R}}[0,1]$;

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

(c) $G(q)$ is continuous;

(d) $K(x)$ has a compact support or $K(x)$ is eventually monotonic so that $\int |K| < \infty$.

Then, for any fixed $l \geq 1$, $c_n \to \infty$ and $c_n/n \to 0$, we have

$$\begin{aligned} S_{n,l} &= \frac{1}{l}\sum_{j=1}^{l} G(X_{n,[n\tau_j]})\, A_0 \int K + o_P(1) \\ &\to_d \frac{1}{l}\sum_{j=1}^{l} G(X_{\tau_j})\, A_0 \int K. \end{aligned} \tag{44}$$

If in addition $\tau_j = j/(l_n+1), j = 1, 2, ..., l_n$, where $l_n^{-1} + l_n/c_n \to 0$, then

$$S_{n,l_n} = \int_0^1 G(X_{n,[nt]})dt\, A_0 \int K + o_P(1) \to_d \int_0^1 G(X_t)dt\, A_0 \int K. \tag{45}$$

Remark 20. Weak convergence in (a) and continuity of $G(q)$ are essentially necessary for this kind of result. The result can be extended to the case that $G(q)$ is locally Lebesgue integrable if we impose additional smoothness conditions on X_{nk}, but it involves more complicated calculations. We do not pursue the extension to keep this paper under reasonable length. It is worth to mention that no relationship is imposed between v_k and X_{nk} and condition (b) is satisfied with $A_0 = Ev_1$ whenever v_t is ergodic (strictly) stationary satisfying $E|v_1| < \infty$ and $\frac{1}{n}\sum_{k=1}^n v_k \to_{L_1} Ev_1$. This fact will be used in subsequent sections without further explanation.

If we are only interested in the boundedness of $S_{n,l}$, condition (b) can be reduced as seen in the following result.

Lemma 2. *Suppose that conditions (a), (c) and (d) of Lemma 1 hold and $\{v_k\}_{k\geq 1}$ is an arbitrary random sequence satisfying $\sup_{k\geq 1} E|v_k| < \infty$. Then, for any $l \geq 1$ (allowing for $l = l_n \to \infty$), $c_n \to \infty$ and $c_n/n \to 0$, we have*

$$\frac{c_n}{n}\sum_{k=1}^n |G(X_{nk})|\,|v_k|\,\frac{1}{l}\sum_{j=1}^l K\big[c_n(k/n-\tau_j)\big] = O_P(1). \tag{46}$$

If in addition $\tau_j = j/(l_n+1), j = 1, 2, ..., l_n$, where $l_n \log l_n/c_n + l_n^{-1} \to 0$, then

$$\frac{c_n}{n}\sum_{k=1}^n |G(X_{nk})|\,|v_k|\,\frac{1}{l_n}\sum_{1\leq i<j\leq l_n} K\big[c_n(k/n-\tau_i)\big]K\big[c_n(k/n-\tau_j)\big] = o_P(1), \tag{47}$$

$$\frac{c_n}{n}\sum_{k=1}^n |G(X_{nk})|\,|v_k|\,\Big(\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n} K\big[c_n(k/n-\tau_j)\big]\Big)^2 = O_P(1), \tag{48}$$

$$\Big(\frac{c_n}{n}\Big)^2\sum_{k=1}^n |G(X_{nk})|\,|v_k|\,\Big(\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n} K\big[c_n(k/n-\tau_j)\big]\Big)^4 = o_P(1). \tag{49}$$

Remark 21. The results in Lemmas 1 and 2 still hold in case that $K(x)$ is replaced by $x^jK^l(x)$ for any $j \geq 0$ and any $l \geq 1$ under additional condition $\int |x^jK^l| < \infty$. This claim is obvious from the proof of lemmas with minor modifications and will be used in the proofs of main results without further explanation. Furthermore, by letting $K^*(x)$ be an another positive function satisfying the same condition as that of $K(x)$, the same argument as in the proof of (47) yields

$$\frac{c_n}{n}\sum_{k=1}^n |G(X_{nk})|\,|v_k|\,\frac{1}{l_n}\sum_{1\leq i<j\leq l_n} K\big[c_n(k/n-\tau_i)\big]K^*\big[c_n(k/n-\tau_j)\big] = o_P(1). \tag{50}$$

This, together with Lemma 1, implies that

$$\begin{aligned}&\frac{c_n}{n}\sum_{k=1}^n G(X_{nk})\,v_k\,\frac{1}{l_n}\sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right]\sum_{j=1}^{l_n} K^*\left[c_n(k/n-\tau_j)\right]\\ &= \int_0^1 G(X_{n,[nt]})dt\,A_0\int KK^* + o_P(1) \to_d \int_0^1 G(X_t)dt\,A_0\int KK^*.\end{aligned} \tag{51}$$

Equation (51) shows the effect of employing "*double trimming*" i.e. sample functionals that involve two kernel functions, which will be used in the proofs of Theorems 4 and 5, and (14).

6.2 Proof of Theorem 1

We only consider M_{1n,l_n}, i.e., (9), since the limit result for $S_{1n,l_n}^{(m)}$ given in (8) follows easily from Lemma 1 with $G(x) \equiv 1$ and $v_k = \alpha' g(x_{k-1})\sigma_k^m$ for any $\alpha \in \mathbb{R}^p$.

Set $Q_{k,n} := \sqrt{\frac{c_n}{n}}\alpha' g(x_{k-1})\sigma_k \frac{1}{\sqrt{l_n}} \sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right]$ where $\alpha \in \mathbb{R}^p$. Using (47) in Lemma 2 with $G(x) \equiv 1$ and $v_k = \left[\alpha' g(x_{k-1})\sigma_k\right]^2$, we have

$$\begin{aligned}
\sum_{k=1}^{n} Q_{k,n}^2 &= \frac{c_n}{n}\sum_{k=1}^{n}\left[\alpha' g(x_{k-1})\sigma_k\right]^2 \frac{1}{l_n}\sum_{j=1}^{l_n} K^2\left[c_n(k/n-\tau_j)\right] + o_P(1) \\
&= E\left[\alpha' g(x_1)\sigma_2\right]^2 \int K^2 + o_P(1) \qquad (52)
\end{aligned}$$

where the second equation follows from Lemma 1 [$K(x)$ is replaced by $K^2(x)$] with additional $A_0 = E\left[\alpha' g(x_{k-1})\sigma_k\right]^2$. In terms of (52), 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 (9), it suffices to show

$$\max_{1\le k\le n} |Q_{k,n}| = o_P(1). \qquad (53)$$

Note that for any $A > 0$,

$$\begin{aligned}
\max_{1\le k\le n} |Q_{k,n}| &\le \left\{\sum_{k=1}^{n} Q_{k,n}^2 I\left\{\|g(x_{k-1})\sigma_k\| > A\right\}\right\}^{1/2} + \left\{\sum_{k=1}^{n} Q_{k,n}^4 I\left\{\|g(x_{k-1})\sigma_k\| \le A\right\}\right\}^{1/4} \\
&=: II_{1n}(A)^{1/2} + II_{2n}(A)^{1/4}.
\end{aligned}$$

Similar arguments used in (52) show that the first term

$$\begin{aligned}
II_{1n}(A) &\le \|\alpha\|^2 \frac{c_n}{n}\sum_{k=1}^{n}\|g(x_{k-1})\sigma_k\|^2 I\left\{\|g(x_{k-1})\sigma_k\| > A\right\}\frac{1}{l_n}\sum_{j=1}^{l_n} K^2\left[c_n(k/n-\tau_j)\right] + o_P(1) \\
&= \|\alpha\|^2 E\,\|g(x_1)\sigma_2\|^2 I\left\{\|g(x_1)\sigma_2\| > A\right\}\int K^2 + o_P(1) = o_P(1),
\end{aligned}$$

where we take $n \to \infty$ first and then $A \to \infty$. On the other hand, by using (49) in Lemma 2 with $G(x) \equiv 1$ and $v_k = 1$, the second term

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

for each $A \geq 1$, as $n \to \infty$. Combining these facts together, we establish (53). The proof of Theorem 1 is now complete. $\square$

6.3 Proof of Theorem 2

As in the proof of Theorem 1, we only consider M_{2n,l_n}, i.e., (11), since the result for $S_{2n,l_n}^{(m)}$ given in (10) follows from Lemma 1 with $v_k \equiv \sigma_k^m$, $m = 0, 1$ or 2, respectively.

Set $Q_{k,n} := \sqrt{\frac{c_n}{n}}\alpha' g(X_{n,k-1})\sigma_k \frac{1}{\sqrt{l_n}} \sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right]$ where $\alpha \in \mathbb{R}^p$. Noting that $\int_0^1 g(X_{n,[nt]})dt$ is a continuous functional of $X_{n,[nt]}$, the limit result of (11), jointly with (10), will follow if we prove that, for any $\alpha \in \mathbb{R}^p$.

$$\Big\{X_{n,[nt]},\ \sum_{k=1}^{n} Q_{k,n} u_k\Big\} \Rightarrow \Big\{X_t,\ \mathbf{MN}\Big(0, E\sigma_1^2 \int_0^1 \left[\ \alpha' g(X_t)\right]^2 dt \int K^2\Big)\Big\} \tag{54}$$

on $D_{\mathbb{R}^2}[0, 1]$. First note that, by using (47) with $v_k \equiv \sigma_k^2$ and $G(.) = \alpha' g(.)$ first and then (45),

$$\begin{aligned}\sum_{k=1}^{n} Q_{k,n}^2 &= \frac{c_n}{n}\sum_{k=1}^{n}\left[\alpha' g(X_{n,k-1})\right]^2 \sigma_k^2 \frac{1}{l_n}\sum_{j=1}^{l_n} K^2\left[c_n(k/n-\tau_j)\right] + o_P(1) \\ &= E\sigma_1^2 \int_0^1 \left[\alpha' g(X_{n,[nt]})\right]^2 dt \int K^2 + o_P(1). \end{aligned}\tag{55}$$

It follows from **A3**(a) and the continuous mapping theorem that

$$\begin{aligned}&\Big\{\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\xi_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\xi_{-k},\ X_{n,[nt]},\ \sum_{k=1}^{n} Q_{k,n}^2\Big\} \\ &\qquad \Rightarrow \Big\{B_{1t}, B_{2t}, X_t, \sigma_1^2\int_0^1 \left[\alpha' g(X_t)\right]^2 dt \int K^2\Big\},\end{aligned}$$

on $D_{R^4}[0, 1]$. Recall that **A1** and $Q_{k,n}$ is a functional of $\xi_k, \xi_{k-1}, \ldots$. By using Theorem 2.1 of Wang (2014) or Theorem 3.14 of Wang (2015), the limit result of (54) will follow, if we prove

$$\max_{1\leq k\leq n} |Q_{k,n}| = o_P(1), \tag{56}$$

and

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} |Q_{k,n}| = o_P(1). \tag{57}$$

In fact, by recalling the fact that $||g||^4$ is still continuous, it follows from (49) with

$|v_k| = \sigma_k^4$ in Lemma 2 that

$$\begin{aligned}
\Big[\max_{1\le k\le n}|Q_{k,n}|\Big]^4 &\le \sum_{k=1}^n Q_{k,n}^4 \\
\le \quad \|\alpha\|^4 \Big(\frac{c_n}{nl_n}\Big)^2 & \sum_{k=1}^n \|g(X_{n,k-1})\|^4\, \sigma_k^4 \Big(\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right]\Big)^4 = o_P(1),
\end{aligned}$$

yielding (56). Similarly, by recalling $l_n/c_n \to 0$ and using (46) in Lemma 2, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{k=1}^n |Q_{k,n}| &\le \|\alpha\| \frac{1}{\sqrt{n}}\sqrt{\frac{c_n}{nl_n}} \sum_{k=1}^n \|g(X_{n,k-1})\|\,|\sigma_k|\, \frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right] \\
&= \|\alpha\| \sqrt{\frac{l_n}{c_n}}\frac{c_n}{n}\sum_{k=1}^n \||\sigma_k| g(X_{n,k-1})\|\, \frac{1}{l_n}\sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right] \\
&= O_P\Big(\sqrt{\frac{l_n}{c_n}}\Big) = o_P\left(1\right),
\end{aligned}$$

which shows (57). The proof of Theorem 2 is complete. □

6.4 Proof of Theorem 3

To show Theorem 3, we only prove (12) since (13) is a direct consequence of (12) and Theorem 2.

Recall $K_{kn} = \sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right]$. Notice that, by the condition (b), we may write

$$\begin{aligned}
&\frac{c_n}{n}\sum_{k=1}^n \pi\left(d_n\right)^{-1} g(x_{k-1})\, \sigma_k^m \Big\{\frac{1}{l_n}\sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right]\Big\} \\
&\qquad = \frac{c_n}{nl_n}\sum_{k=1}^n H(X_{n,k-1})\sigma_k^m\, K_{kn} + \Delta_{1n}, \\
&\sqrt{\frac{c_n}{n}}\sum_{k=1}^n \pi\left(d_n\right)^{-1} g(x_{k-1})\, \sigma_k \Big\{\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right]\Big\}\, u_k\Big] \\
&\qquad = \sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^n H(X_{n,k-1})\sigma_k\, K_{kn} u_k + \Delta_{2n},
\end{aligned}$$

where $R(\lambda,x) = [R_1(\lambda,x),...,R_p(\lambda,x)]'$ and

$$\begin{aligned}
\Delta_{1n} &= \frac{c_n}{nl_n}\sum_{k=1}^n \pi\left(d_n\right)^{-1} R(d_n, X_{n,k-1})\, \sigma_k\, K_{kn}, \\
\Delta_{2n} &= \sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^n \pi\left(d_n\right)^{-1} R(d_n, X_{n,k-1})\, \sigma_k\, K_{kn}\, u_k.
\end{aligned}$$

Result (12) follows from Theorem 2 with $g(x) = H(x)$ if we prove

$$|\alpha'\Delta_{in}| \;=\; o_P(1), \quad i = 1, 2, \tag{58}$$

for any $\alpha = (\alpha_1, ..., \alpha_p)' \in \mathbb{R}^p$.

We only prove (58) with $i = 2$ since the proof of $|\alpha'\Delta_{1n}| = o_P(1)$ is similar except simpler. Set, for $A > 0$,

$$\widetilde{R}_{n,l_n}(A) = \sqrt{\frac{c_n}{nl_n}} \sum_{k=1}^{n} \alpha'\pi\,(d_n)^{-1}\, R(d_n, X_{n,k-1}) I\,\{|X_{n,k-1}| \le A\}\, \sigma_k\, K_{kn} u_k.$$

Note that as $n \to \infty$ first and then $A \to \infty$

$$P\left(\alpha'\Delta_{2n} \ne \widetilde{R}_{n,l_n}(A)\right) \le P\left(\max_{1\le k\le n} |X_{n,k-1}| \ge A\right) \to 0. \tag{59}$$

For any $\epsilon > 0$ and $A > 0$, we have

$$P\left(|\alpha'\Delta_{2n}| \ge \epsilon\right) \le P\left(\alpha'\Delta_n \ne \widetilde{R}_{n,l_n}(A)\right) + \epsilon^{-2} E\left[\widetilde{R}_{n,l_n}(A)\right]^2.$$

Now $|\alpha'\Delta_{2n}| = o_P(1)$ follows from (59) and the fact that as $n \to \infty$ for any $A > 0$

$$\begin{aligned} E\left[\widetilde{R}_{n,l_n}(A)\right]^2 &\le \frac{c_n}{nl_n} C \sum_{k=1}^{n} E\left|\alpha'\pi\,(d_n)^{-1}\, R(d_n, X_{n,k-1})\right|^2 I\,\{|X_{n,k-1}| \le A\}\, \sigma_k^2\, K_{kn}^2 \\ &\le \frac{c_n}{n} C \|\alpha\|^2 \left(1 + A^{\delta}\right)^2 \epsilon_n^2 \frac{1}{l_n} \sum_{k=1}^{n} \sigma_k^2\, K_{kn}^2 \to 0, \end{aligned}$$

where $\epsilon_n = \max_{1\le i\le p} |[\pi_i(d_n)]^{-1} a_i(d_n)| \to 0$ and we have used (48) of Lemma 2 (with $G(x) \equiv 1$ and $v_k \equiv \sigma_k^2$). The proof of Theorem 3 is now complete. $\square$

6.5 Proof of (14)

Proof of (14) is essentially the same as that of (12). We only provide a outline. For any $\alpha, \beta \in \mathbb{R}$, let

$$\widetilde{Q}_{k,n} = \sqrt{\frac{c_n}{nl_n}} \left(\alpha H_2(X_{n,k-1}) K_{kn} + \beta\, K_{kn}^*\right) \sigma_k\,,$$

where $K_{kn}^* := \sum_{j=1}^{l_n} K^*\left[c_n\,(k/n - \tau_j)\right]$. As in the proof of (12), we have

$$\alpha\, U_{1n} + \beta\, U_{2n} = \sum_{k=1}^{n} \widetilde{Q}_{k,n}\, u_k + o_P(1).$$

Note that, by using (51) and Lemmas 1 and 2,

$$\frac{1}{E\sigma_1^2}\sum_{k=1}^{n}\widetilde{Q}_{k,n}^2 = \alpha^2\int_0^1 H_2^2(X_{n,[nt]})dt\int K^2+2\alpha\beta\int_0^1 H_2(X_{n,[nt]})dt\int KK^* +\beta^2\int (K^*)^2+o_P(1),$$

as in the proof of (55). It follows from **A3**(a) and the continuous mapping theorem that, for any α and $\beta\in\mathbb{R}$,

$$\Big\{\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\xi_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\xi_{-k},\ X_{n,[nt]},\ \sum_{k=1}^{n}\widetilde{Q}_{k,n}^2\Big\} \Rightarrow \big\{B_{1t},\ B_{2t},\ X_t,\ E\sigma_1^2\,[\alpha,\beta]\,V\,[\alpha,\beta]'\big\},$$

on $D_{\mathbb{R}^4}[0,1]$. Similarly, we may prove that (56) and (57) hold with Q_{kn} being replaced by $\widetilde{Q}_{k,n}$. As a consequence, (14) follows from Wang (2014) as in the proof of Theorem 2. □

6.6 Proof of Lemma 1

We only prove (45), as the proof of (44) is similar except more simpler. We start with the proof of (45) by assuming that there exists an $A>0$ such that $K(x)=0$ if $|x|\geq A$ and $K(x)$ is Lipschitz continuous on $\mathbb{R}$. This restriction will be removed later.

Without loss of generality, suppose $A=1$. Set $\delta_{1n,j}=[n(\tau_j-1/c_n)]\vee 1$, $\delta_{2n,j}=[n(\tau_j+1/c_n)]\vee 1$ and $\delta_{n,j}=[n\tau_j]$. Recall $\tau_j=j/(l_n+1)$. Since

$$|c_n(k/n-\tau_j)|<1 \quad\text{only if}\quad \delta_{1n,j}\leq k\leq \delta_{2n,j},\quad j=1,...,l_n, \tag{60}$$

by letting $R_{1n,j}=\frac{c_n}{n}\sum_{k=\delta_{1n,j}}^{\delta_{2n,j}} v_k\ K\big[c_n(k/n-\tau_j)\big]$ and

$$R_{2n,j} = \frac{c_n}{n}\sum_{k=\delta_{1n,j}}^{\delta_{2n,j}}\big[G\big(X_{nk}\big)-G\big(X_{n,\delta_{n,j}}\big)\big]\ v_k\ K\big[c_n(k/n-\tau_j)\big],$$

we have

$$\begin{aligned} S_{n,l_n} &= \frac{1}{l_n}\sum_{j=1}^{l_n}\frac{c_n}{n}\sum_{k=1}^{n}G(X_{nk})\,v_k\,K\big[c_n(k/n-\tau_j)\big] \\ &= \frac{1}{l_n}\sum_{j=1}^{l_n}G\big(X_{n,\delta_{n,j}}\big)\frac{c_n}{n}\sum_{k=\delta_{1n,j}}^{\delta_{2n,j}} v_k\ K\big[c_n(k/n-\tau_j)\big] \end{aligned}$$

$$
\begin{aligned}
& +\frac{1}{l_n}\sum_{j=1}^{l_n}\frac{c_n}{n}\sum_{k=\delta_{1n,j}}^{\delta_{2n,j}}\big[G\big(X_{nk}\big)-G\big(X_{n,\delta_{n,j}}\big)\big]\ v_k\ K\big[c_n(k/n-\tau_j)\big] \\
= & \ \frac{1}{l_n}\sum_{j=1}^{l_n}G\big(X_{n,\delta_{n,j}}\big)\,R_{1n,j}+\frac{1}{l_n}\sum_{j=1}^{l_n}R_{2n,j} \\
= & \ \frac{1}{l_n}\sum_{j=1}^{l_n}G\big(X_{n,\delta_{n,j}}\big)\,A_0\int K\ +\frac{1}{l_n}\sum_{j=1}^{l_n}G\big(X_{n,\delta_{n,j}}\big)\,\big[R_{1n,j}-A_0\int K\big]+\frac{1}{l_n}\sum_{j=1}^{l_n}R_{2n,j} \\
:= & \ \frac{1}{l_n}\sum_{j=1}^{l_n}G\big(X_{n,\delta_{n,j}}\big)\,A_0\int K\ +R_{1n}+R_{2n}.
\end{aligned}
$$

Since $\frac{1}{l_n}\sum_{j=1}^{l_n}G\big(X_{n,\delta_{n,j}}\big)=\int_0^1 G(X_{n,[nt]})dt+o_P(1)\to_d\int_0^1 G(X_t)dt$, it suffices to show that

$$R_{jn}\ =\ o_P(1),\quad j=1,\ 2. \tag{61}$$

To prove (61), we start with some preliminaries. Recalling $X_{n,[nt]}\Rightarrow X_t$ on $D_{\mathbb{R}}[0,1]$ and the limit process $X(t)$ is path continuous, we have $X_{n,[nt]}\Rightarrow X_t$ on $D_{\mathbb{R}}[0,1]$ in the sense of uniform topology. See, for instance, Section 18 of Billingsley (1968). This fact implies that

$$\limsup_{N\to\infty}\limsup_{n\to\infty}P\Big(\max_{1\le k\le n}|X_{nk}|\ge N\Big)=0, \tag{62}$$

and by the tightness of $\{X_{n,[nt]}\}_{0\le t\le 1}$, for any $\varepsilon>0$ and $\delta>0$, there is some $\tilde\delta=\tilde\delta(\varepsilon,\delta)>0$ such that

$$P(\sup_{|s-t|\le\tilde\delta}|X_{n,[nt]}-X_{n,[ns]}|\ge\delta)\le\varepsilon \tag{63}$$

holds for all sufficiently large n. In terms of (63), for any $\delta>0$, we have

$$\lim_{n\to\infty}P(\max_{1\le j\le l_n}\max_{\delta_{1n,j}\le l\le k\le\delta_{2n,j}}|X_{nk}-X_{nl}|\ge\delta)=0. \tag{64}$$

We are now ready to prove (61), starting with $j=1$.

For any $N>0$, we let $G_N(x)=G(x)\xi_N(x)$ with

$$\xi_N(x)=\begin{cases}1, & |x|\le N,\\ 2-|x|/N, & N<|x|<2N,\\ 0, & |x|\ge 2N,\end{cases}$$

and

$$\widetilde{R}_{1n}\ =\ \frac{1}{l_n}\sum_{j=1}^{l_n}G_N\big(X_{n,\delta_{n,j}}\big)\,\big[R_{1n,j}-A_0\int K\big].$$

Note that, as $n \to \infty$ first and then $N \to \infty$,

$$P(R_{1n} \neq \widetilde{R}_{1n}) \leq P\Big(\max_{1\leq k\leq n} |X_{nk}| \geq N\Big) \to 0, \tag{65}$$

and

$$|\widetilde{R}_{1n}| \;\; \leq \;\; \frac{C_N}{l_n} \sum_{j=1}^{l_n} \big| R_{1n,j} - A_0 \int K \big|, \tag{66}$$

where $C_N := \sup_x |G_N(x)| < \infty$ is a constant depending only on N, due to the continuity of $G(x)$. Result (61) with $j = 1$ will follow if we prove

$$\max_{1\leq j\leq l_n} E\big|R_{1n,j} - A_0 \int K\big| \;\; \to \;\; 0, \tag{67}$$

as $n \to \infty$. Indeed, by virtue of (66) and (67), we have $E|\widetilde{R}_{1n}| \to 0$ and then $\widetilde{R}_{1n} = o_P(1)$ for each $N \geq 1$. This, together with (65), yields $R_{1n} = o_P(1)$.

Since, as $n \to \infty$,

$$\max_{1\leq j\leq l_n} \Big| \frac{c_n}{n} \sum_{k=\delta_{1n,j}}^{\delta_{2n,j}} K\big[c_n(k/n - \tau_j)\big] - \int K \Big| \to 0, \tag{68}$$

to prove (67), it suffices to show that $\max_{1\leq j\leq l_n} E|A_n(\tau_j)| \to 0$, where

$$A_n(\tau_j) \;\; = \;\; \frac{c_n}{n} \sum_{k=\delta_{1n,j}}^{\delta_{2n,j}} (v_k - A_0)\; K\big[c_n(k/n - \tau_j)\big].$$

Let $\gamma = \gamma_n$ be integers such that $\gamma \to \infty$ and $\gamma\, c_n/n \to 0$, $T_{1n,j} = [\delta_{1n,j}/\gamma]$ and $T_{2n,j} = [\delta_{2n,j}/\gamma]$. Noting (60), we may write

$$\begin{aligned}
A_n(\tau_j) \;\; &= \;\; \frac{c_n}{n} \sum_{k=\delta_{1n,j}}^{\delta_{2n,j}} (v_k - A_0)\, K\big[c_n(k/n - \tau_j)\big] \\
&= \;\; \frac{c_n}{n} \sum_{s=T_{1n,j}}^{T_{2n,j}} \sum_{k=s\gamma}^{(s+1)\gamma} (v_k - A_0)\, K\big[c_n(k/n - \tau_j)\big] \\
&\leq \;\; \frac{\gamma c_n}{n} \sum_{s=T_{1n,j}}^{T_{2n,j}} K\big[c_n(s\gamma/n - \tau_j)\big] \frac{1}{\gamma} \Big| \sum_{k=s\gamma}^{(s+1)\gamma} (v_k - A_0) \Big| \\
&\quad + \frac{c_n}{n} \sum_{s=T_{1n,j}}^{T_{2n,j}} \sum_{k=s\gamma}^{(s+1)\gamma} |v_k - A_0| \Big| K\big[c_n(k/n - \tau_j)\big] - K\big[c_n(s\gamma/n - \tau_j)\big] \Big| \\
:= \;\; & A_{1n}(\tau_j) + A_{2n}(\tau_j).
\end{aligned}$$

Recall $\sup_{k\geq 1} E|v_k| < \infty$ by condition (b), it is readily from the Lipschitz condition on $K(x)$ that

$$EA_{2n}(\tau_j) \leq C\,\frac{\gamma c_n}{n}\,\frac{c_n}{n}\sum_{k=\delta_{1n,j}}^{\delta_{2n,j}} E|v_k - A_0| \leq C\,\frac{\gamma c_n}{n} \to 0,$$

uniformly in $1 \leq j \leq l_n$. Similarly, by using condition (b), we have

$$\max_{1\leq j\leq l_n} EA_{1n}(\tau_j) \leq \max_{\gamma\leq s\leq n-\gamma} E\,\Big|\frac{1}{\gamma}\sum_{k=s}^{s+\gamma} v_k - A_0\Big| \max_{1\leq j\leq l_n} A_{4n}(\tau_j) \to 0,$$

where

$$A_{4n}(\tau_j) = \frac{\gamma c_n}{n}\sum_{s=T_{1n,j}}^{T_{2n,j}} K\big[c_n(s\gamma/n - \tau_j)\big],$$

and we have used the fact that $\max_{1\leq j\leq l_n}\Big|A_{4n}(\tau_j) - \int K\Big| \to 0$. Combining all these facts, we prove (67), and complete the proof of $R_{1n} = o_P(1)$.

We next show $R_{2n} = o_P(1)$. Let $\widetilde{R}_{2n} = \frac{1}{l_n}\sum_{j=1}^{l_n}\widetilde{R}_{2n,j}$, where

$$\widetilde{R}_{2n,j} = \frac{c_n}{n}\sum_{k=\delta_{1n,j}}^{\delta_{2n,j}}\big[G_N\big(X_{nk}\big) - G_N\big(X_{n,\delta_{n,j}}\big)\big]\; v_k\; K\big[c_n(k/n - \tau_j)\big].$$

In terms of (62), we have

$$P(R_{2n} \neq \widetilde{R}_{2n}) \leq P\Big(\max_{1\leq k\leq n}|X_{nk}| \geq N\Big) \to 0,$$

as $n \to \infty$ first and then $N \to \infty$. Result $R_{2n} = o_P(1)$ will follow if we prove $\widetilde{R}_{2n} = o_P(1)$, for each fixed $N \geq 1$.

Recall that $G_N(x)$ is continuous with compact support. For any $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ so that $|G_N(x) - G_N(y)| \leq \epsilon$ whenever $|x - y| \leq \delta_\epsilon$. Write

$$\Omega_{\delta_\epsilon} \;=\; \{\omega : \max_{1\leq j\leq l_n}\max_{\delta_{1n,j}\leq l\leq k\leq\delta_{2n,j}} |X_{nk} - X_{nl}| \leq \delta_\epsilon\}.$$

By virtue of the facts above and (68), it is readily seen that

$$\begin{aligned}
&\max_{1\leq j\leq l_n} E\big[|\widetilde{R}_{2n,j}| I(\Omega_{\delta_\epsilon})\big] \\
\leq\;& E\Big\{\max_{1\leq j\leq l_n}\max_{\delta_{1n,j}\leq l\leq k\leq\delta_{2n,j}} |G_N(X_{nk}) - G_N(X_{nl})|\,\frac{c_n}{n}\sum_{k=\delta_{1n,j}+1}^{\delta_{2n,j}} |v_k|\,K\big[c_n(k/n - \tau_j)\big]\Big\} \\
\leq\;& \epsilon\,\sup_{k\geq 1} E|v_k|\,\frac{c_n}{n}\sum_{k=\delta_{1n,j}+1}^{\delta_{2n,j}} K\big[c_n(k/n - \tau_j)\big] \leq C_N\epsilon,
\end{aligned}$$

where C_N is a constant depending only on N. Now, for any $\eta_1 > 0$ and $\eta_2 > 0$, let $\epsilon = \eta_1\eta_2$ and n_0 be large enough so that, for all $n \geq n_0$ [recall (64)],

$$P\big(\max_{1\leq j\leq l_n}\max_{\delta_{1n,j}\leq l\leq k\leq\delta_{2n,j}}|X_{nk}-X_{nl}|\geq\delta_\epsilon\big)\leq\eta_2.$$

It is readily seen that, for all $n \geq n_0$,

$$P(|\widetilde{R}_{2n}|\geq\eta_1) \;\leq\; P\big(\bar{\Omega}_{\delta_\epsilon}\big)+\eta_1^{-1}\frac{1}{l_n}\sum_{j=1}^{l_n}E\big[|\widetilde{R}_{2n,j}|I(\Omega_{\delta_\epsilon})\big] \;\leq\; C_N\,\eta_2$$

where $\bar{\Omega}_{\delta_\epsilon}$ denotes the complementary set of Ω_{δ_ϵ} and C_N is a constant depending only on N. This yields $\widetilde{R}_{2n} = o_P(1)$, for each fixed $N \geq 1$, and completes the proof of $R_{2n} = o_P(1)$.

We finally remove the restriction on K and then conclude the proof of Lemma 1.

If K has a compact support, there exists $A_1 > 0$ such that $K(x) = 0$ holds for all $|x| \geq A_1$. If K is eventually monotonic (without loss of generality, we assume $K \geq 0$), for any $\epsilon > 0$, we can also choose 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 $A \geq \max\{A_1, A_{1\epsilon}\} + 1$, there exists an $K_{\epsilon,A}(x)$ such that

$$\int|K-K_{\epsilon,A}|\leq 2\epsilon, \tag{69}$$

where $K_{\epsilon,A}(x) = 0$ if $|x| \geq A$ and $K_{\epsilon,A}(x)$ is Lipschitz continuous on $\mathbb{R}$. It has been shown in the first part that, for any $\epsilon > 0$ and $A \geq \max\{A_1, A_{1\epsilon}\} + 1$,

$$\begin{aligned}&\frac{1}{l_n}\sum_{j=1}^{l_n}\frac{c_n}{n}\sum_{k=1}^{n}G(X_{nk})\,v_k\,K_{\epsilon,A}\big[c_n(k/n-\tau_j)\big]\\ =\;&\int_0^1 G(X_{n,[nt]})dtA_0\int K_{\epsilon,A}+o_P(1)\to_d\int_0^1 G(X_t)dtA_0\int K_{\epsilon,A}.\end{aligned}$$

To show (45), it suffices to show that, as $n \to \infty$ first and then $\epsilon \to 0$ (implying $A \to \infty$),

$$S_{n,\epsilon} \;:=\; \frac{1}{l_n}\sum_{j=1}^{l_n}\frac{c_n}{n}\sum_{k=1}^{n}G(X_{nk})\,v_k\,\widetilde{K}\big[c_n(k/n-\tau_j)\big]=o_P(1), \tag{70}$$

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

The proof of (70) is similar to that of (61). For any $\epsilon > 0$, take A be given as in (69). First note that, as in (68),

$$\sup_{1\leq j\leq l_n}\Big|\frac{c_n}{n}\sum_{k=1}^{n}\big|\widetilde{K}\big[c_n(k/n-\tau_j)\big]\big|I(c_n|k/n-\tau_j|\leq A)-\int_{-A}^{A}|\widetilde{K}(x)|dx\Big|\to 0,$$

when $n \to \infty$, i.e., whenever n is sufficiently large,

$$A_{1j} := \frac{c_n}{n} \sum_{k=1}^{n} \big|\widetilde{K}\big[c_n(k/n - \tau_j)\big]\big| I(c_n|k/n - \tau_j| \le A) \le \int |\widetilde{K}(x)| dx + \epsilon \le 3\epsilon,$$

uniformly for $1 \le j \le l_n$. On the other hand, it follows from the monotonicity of $K(x)$ on $(-\infty, -A)$ and (A, ∞) that, whenever n is sufficiently large,

$$\begin{aligned} A_{2j} &:= \frac{c_n}{n} \sum_{k=1}^{n} \big|\widetilde{K}\big[c_n(k/n - \tau_j)\big]\big| I(c_n|k/n - \tau_j| > A) \\ &= \frac{c_n}{n} \sum_{k=1}^{n} K\big[c_n(k/n - \tau_j)\big] I(c_n|k/n - \tau_j| > A) \\ &\le \int_{|x| > A - c_n/n} K(x) dx \le \int_{|x| > \max\{A_1, A_{1\epsilon}\}} K(x) dx < \epsilon, \end{aligned}$$

uniformly for $1 \le j \le l_n$. By using these facts, when n is sufficiently large, we have

$$\frac{1}{l_n} \sum_{j=1}^{l_n} \frac{c_n}{n} \sum_{k=1}^{n} \big|\widetilde{K}\big[c_n(k/n - \tau_j)\big]\big| \le \frac{1}{l_n} \sum_{j=1}^{l_n} \big(A_{1j} + A_{2j}\big) \le 4\epsilon.$$

Now, for any $\delta > 0$, by letting

$$S_{n,\epsilon,N} = \frac{1}{l_n} \sum_{j=1}^{l_n} \frac{c_n}{n} \sum_{k=1}^{n} G_N(X_{nk})\, v_k\, \widetilde{K}\big[c_n(k/n - \tau_j)\big]$$

and noting the uniformed boundedness of $G_N(x)$, we have

$$\begin{aligned} P\big(|S_{n,\epsilon}| \ge \delta\big) &\le P\big(S_{n,\epsilon} \ne S_{n,\epsilon,N}\big) + P\big(|S_{n,\epsilon,N}| \ge \delta\big) \\ &\le P\big(\max_{1 \le k \le n} |X_{nk}| \ge N\big) + \delta^{-1} E|S_{n,\epsilon,N}| \\ &\le P\big(\max_{1 \le k \le n} |X_{nk}| \ge N\big) + \delta^{-1} C_N \sup_k E|v_k| \frac{1}{l_n} \sum_{j=1}^{l_n} \frac{c_n}{n} \sum_{k=1}^{n} \big|\widetilde{K}\big[c_n(k/n - \tau_j)\big]\big| \\ &\le P\big(\max_{1 \le k \le n} |X_{nk}| \ge N\big) + C_{1N}\, \epsilon \delta^{-1} \to 0 \end{aligned}$$

as $n \to \infty$ first, $N \to \infty$ second and then $\epsilon \to 0$. This proves (70) and hence completes the proof of Lemma 1. □

6.7 Proof of Lemma 2

We first prove (47) and, without loss of generality, assume $K \geq 0$. Using similar arguments as in the proof of (61) or (70), it suffices to show that, as $n \to \infty$,

$$I_n := \frac{c_n}{n}\sum_{k=1}^{n}\frac{1}{l_n}\sum_{1\leq i<j\leq l_n} K\big[c_n(k/n-\tau_i)\big]K\big[c_n(k/n-\tau_j)\big] \to 0.$$

Take $\eta_{n,i,j} = \frac{1}{2}n(\tau_i+\tau_j)$. Note that $c_n(k/n-\tau_i) \geq c_n(j-i)/(2(l_n+1))$ if $k \geq \eta_{n,i,j}$ and $|c_n(k/n-\tau_j)| \geq c_n(j-i)/(2(l_n+1))$ if $k \leq \eta_{n,i,j}$. It follows from $K(x) \leq 1/|x|$ as x is sufficiently large [28] that

$$\begin{aligned}
I_n &= \frac{1}{l_n}\sum_{1\leq i<j\leq l_n}\frac{c_n}{n}\sum_{k=1}^{n} K\big[c_n(k/n-\tau_i)\big]K\big[c_n(k/n-\tau_j)\big] \\
&\leq \frac{C}{l_n}\sum_{1\leq i<j\leq l_n}\frac{l_n+1}{c_n(j-i)}\,\frac{c_n}{n}\sum_{k=1}^{n}\Big(K\big[c_n(k/n-\tau_i)\big]+K\big[c_n(k/n-\tau_j)\big]\Big) \\
&\leq \frac{C}{c_n}\sum_{1\leq i<j\leq l_n}\frac{1}{j-i} \leq C\, l_n \log l_n/c_n \to 0,
\end{aligned}$$

as required.

The proof of (46) is similar to that of (47) and hence the details are omitted. Result (48) follows easily from (46) and (47). As for (49), it follows from the similar arguments as in the proof of (61) and the fact: as $n \to \infty$,

$$\begin{aligned}
&\Big(\frac{c_n}{n}\Big)^2\sum_{k=1}^{n}\Big(\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n}K\big[c_n(k/n-\tau_j)\big]\Big)^4 \\
\leq\ & 2\Big(\frac{c_n}{n}\Big)^2\sum_{k=1}^{n}\Big(\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n}K^2\big[c_n(k/n-\tau_j)\big]\Big)^2 \\
&+8\Big(\frac{c_n}{n}\Big)^2\sum_{k=1}^{n}\Big(\frac{1}{l_n}\sum_{1\leq i<j\leq l_n}K\big[c_n(k/n-\tau_i)\big]K\big[c_n(k/n-\tau_j)\big]\Big)^2 \\
\leq\ & 2C^2\Big(\frac{c_n}{n}\Big)^2\sum_{k=1}^{n}\Big(\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n}K\big[c_n(k/n-\tau_j)\big]\Big)^2 + 8I_n^2 \to 0,
\end{aligned}$$

due to (47) and (48). $\Box$

---

[28] Since $\int K < \infty$ and $K \geq 0$ is eventually monotonic, we have that K is decreasing on (A_1, ∞) for some $A_1 > 0$, and

$$xK(x)/2 \leq \int_{x/2}^{x} K(t)dt \to 0, \quad x \to +\infty,$$

Similarly $\lim_{x\to-\infty} xK(x) = 0$. Hence $K(x) \leq 1/|x|$ as x is sufficiently large.

7 Proofs for Section 3

We next provide proofs for the results of Section 3.

7.1 Proofs of Theorems 4 and 5

We only prove Theorem 5 since the proof of Theorem 4 is similar except simpler. Let

$$\begin{aligned}
A_{1n} &= \frac{c_n}{nl_n}\sum_{k=1}^n \pi(d_n)^{-2} f^2(x_{k-1})\, K_{kn}, \quad A_{2n} = \frac{c_n}{nl_n}\sum_{k=1}^n \pi(d_n)^{-1} f(x_{k-1})\, K_{kn}, \\
A_{3n} &= \frac{c_n}{nl_n}\sum_{k=1}^n \pi(d_n)^{-1} f(x_{k-1})\, K^*_{kn}, \\
B_{1n} &= \sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^n \pi(d_n)^{-1} f(x_{k-1})\, K_{kn}\, \sigma_k\, u_k, \quad B_{2n} = \sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^n K^*_{kn}\, \sigma_k\, u_k.
\end{aligned}$$

Recall (18), $K_{kn} = \sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right]$, $K^*_{kn} = \sum_{j=1}^{l_n} K^*\left[c_n(k/n-\tau_j)\right]$ and $Z_{kn} = f(x_{k-1})K_{kn}$. We have $\frac{c_n}{nl_n}\sum_{k=1}^n K^*_{kn} = \int K^* + o(1)$ (c.g, Lemma 1) and it follows from (12) of Theorem 3 [29] and (10) of Theorem 2 that

$$\begin{aligned}
&\frac{c_n}{nl_n}\frac{1}{\pi^2(d_n)}\sum_{k=1}^n Z_{kn}\overline{f}_k \\
&= \frac{c_n}{nl_n}\sum_{k=1}^n \pi(d_n)^{-1} f(x_{k-1})\, K_{kn}\left[\pi(d_n)^{-1} f(x_{k-1}) - \frac{\sum_{k=1}^n \pi(d_n)^{-1} f(x_{k-1}) K^*_{kn}}{\sum_{k=1}^n K^*_{kn}}\right] \\
&= A_{1n} - A_{2n}A_{3n}\Big/\int K^* + o_P(1) \\
&= \widehat{C}_n \int K + o_P(1), \qquad (71)
\end{aligned}$$

where $\widehat{C}_n = \int_0^1 H^2(X_{n,[nt]})dt - \left[\int_0^1 H(X_{n,[nt]})dt\right]^2$. Similarly, by using Theorems 2 and 3, we have

$$\sqrt{\frac{c_n}{nl_n}}\frac{1}{\pi(d_n)}\sum_{k=1}^n Z_{kn}\overline{e_k}$$

[29] We mention that $f^m(x), m = 2, 3, ...$ still are asymptotic homogeneous, i.e., $f^m(x)$ for each $m = 2, 3, ...$ still satisfies the condition (b) of Theorem 5. Indeed, by the definition of $f(x)$, we have

$$f^2(\lambda x) = \pi^2(\lambda)\, H^2(x) + R_1(\lambda, x),$$

where

$$|R_1(\lambda, x)| \leq 2\pi(\lambda)\,|H(x)|\,|R(\lambda, x)| + R^2(\lambda, x) \leq a_1(\lambda)(1 + |x|^{\alpha+2\lambda})$$

with $\alpha_1(\lambda) = 2\pi(\lambda)a(\lambda) + a^2(\lambda)$ satisfying $a_1(\lambda)/\pi^2(\lambda) \to 0$ as $\lambda \to \infty$. The proof for $m \geq 3$ is similar, we omit the details.

$$
\begin{aligned}
&= \sqrt{\frac{c_n}{nl_n}} \left\{ \sum_{k=1}^n \pi(d_n)^{-1} f(x_{k-1}) K_{kn}\, \sigma_k\, u_k - \frac{\left[\sum_{k=1}^n \pi(d_n)^{-1} f(x_{k-1}) K_{kn}\right] \left[\sum_{k=1}^n K_{kn}^*\, \sigma_k\, u_k\right]}{\sum_{k=1}^n K_{kn}^*} \right\} \\
&= B_{1n} - A_{2n}\, B_{2n} \Big/ \int K^* + o_P(1) \\
&= \widehat{A}_n\, B_n + o_P(1),
\end{aligned} \tag{72}
$$

where $\widehat{A}_n = \left[1, \ -\frac{\int_0^1 H\left(X_{n,[nt]}\right)dt \int K}{\int K^*}\right]$ and $B_n = \left(B_{1n}, B_{2n}\right)'$. Since both $\widehat{C}_n$ and $\widehat{A}_n$ are continuous functionals of $X_{n,[nt]}$, a simple application of (14) yields that

$$
\begin{aligned}
\sqrt{\frac{nl_n}{c_n}} \pi(d_n) \left(\hat{\beta} - \beta\right) &= \frac{\sqrt{\frac{c_n}{nl_n}} \frac{1}{\pi(d_n)} \sum_{k=1}^n Z_{kn} \overline{e_k}}{\frac{c_n}{nl_n} \frac{1}{\pi(d_n)^2} \sum_{k=1}^n Z_{kn} \overline{f}_k} \\
&= \left(\widehat{C}_n \int K\right)^{-1} \widehat{A}_n\, B_n + o_P(1) \\
&\to_d \sqrt{E\sigma_1^2}\, \mathbf{MN}\left(\mathbf{0}, \ C_2^{-2}\, A_2 V_2 A_2'\right),
\end{aligned} \tag{73}
$$

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

7.2 Proof of Theorem 6

We only prove Theorem 6 under the conditions of Theorem 5. The other is similar. Define

$$
V_n = \begin{bmatrix} \int_0^1 H^2\left(X_{n,[nt]}\right) dt \int K^2 & \int_0^1 H\left(X_{n,[nt]}\right) dt \int KK^* \\ \int_0^1 H\left(X_{n,[nt]}\right) dt \int KK^* & \int (K^*)^2 \end{bmatrix} =: \begin{bmatrix} V_{n,11} & V_{n,12} \\ V_{n,21} & V_{n,22} \end{bmatrix}.
$$

Letting $D_n = \text{diag}\left\{\pi(d_n)\sqrt{\frac{nl_n}{c_n}}, \ \sqrt{\frac{nl_n}{c_n}}\right\}$, we first claim that

$$
\begin{aligned}
D_n^{-1} \mathcal{V}_n D_n^{-1} &= \begin{bmatrix} \frac{c_n}{nl_n} \pi(d_n)^{-2} \sum_{k=1}^n \check{e}_k^2 K_{kn}^2 f_k^2 & \frac{c_n}{nl_n} \pi(d_n)^{-1} \sum_{k=1}^n \check{e}_k^2 K_{kn}^* K_{kn} f_k \\ \frac{c_n}{nl_n} \pi(d_n)^{-1} \sum_{k=1}^n \check{e}_k^2 K_{kn}^* K_{kn} f_k & \frac{c_n}{nl_n} \sum_{k=1}^n \check{e}_k^2 \left(K_{kn}^*\right)^2 \end{bmatrix} \\
&:= \begin{bmatrix} \mathcal{V}_{n,11} & \mathcal{V}_{n,12} \\ \mathcal{V}_{n,21} & \mathcal{V}_{n,22} \end{bmatrix} = E\sigma_1^2\, V_n + o_P(1).
\end{aligned} \tag{74}
$$

To prove (74), it suffices to show that, for $i, j = 1$ and 2,

$$
\mathcal{V}_{n,ij} = E\sigma_1^2\, V_{n,ij} + o_P(1). \tag{75}
$$

We only prove (75) for $i = j = 1$. Others are similar and hence the details are omitted. Recall that $\check{e}_k = y_k - \widetilde{\theta}'\mathbf{f}_k = e_k + (\theta - \widetilde{\theta})'\mathbf{f}_k$, where $\theta = (\mu, \beta)'$ and $\mathbf{f}_k' = [1, f(x_{k-1})]$. Since, by Remark 7, $||(\theta - \hat{\theta})'\mathbf{f}_k|| \le |\tilde{u} - u| + |\tilde{\beta} - \beta|\, |f_k| = o_P(1)\left[1 + \pi(d_n)^{-1} f_k\right]$, it follows that, uniformly for $k=1,2,...,n$,

$$\breve{e}_k^2 = \sigma_k^2 u_k^2 + o_P(1)\big[1+\sigma_k^2 u_k^2 + \pi(d_n)^{-2} f_k^2\big].$$

As a consequence, we have

$$\begin{aligned}
\mathcal{V}_{n,11} &= \frac{nl_n}{c_n}\,\pi(d_n)^{-2}\sum_{k=1}^n \breve{e}_k^2 K_{kn}^2 f_k^2 \\
&= \big[1+o_P(1)\big]\,\frac{nl_n}{c_n}\,\pi(d_n)^{-2}\sum_{k=1}^n \sigma_k^2 u_k^2 K_{kn}^2 f_k^2 \\
&\qquad + o_P(1)\,\frac{nl_n}{c_n}\,\pi(d_n)^{-2}\sum_{k=1}^n \big[1+\pi(d_n)^{-2}f_k^2\big]\,K_{kn}^2 f_k^2 \\
&=: \big[1+o_P(1)\big]\,R_{1n} + o_P(1)\,R_{2n}. \qquad (76)
\end{aligned}$$

Let $v_k=\sigma_k^2 u_k^2$. By recalling that $E\big(u_k^2|\mathcal{F}_{k-1}\big)=1$ and σ_k are $\mathcal{F}_{k-1}$ measurable, it is readily seen from **A3**(b) and $\sup_{k\ge 1} Eu_k^4<\infty$ that $A_0:=E\sigma_1^2 = Ev_k$ for each $k\ge 1$ and

$$\begin{aligned}
\max_{m\le j\le n-m} E\,\Big|\frac{1}{m}\sum_{k=j+1}^{j+m} v_k - A_0\Big| &\le \max_{m\le j\le n-m} E\,\Big|\frac{1}{m}\sum_{k=j+1}^{j+m} \sigma_k^2\big(u_k^2 - E\big(u_k^2|\mathcal{F}_{k-1}\big)\Big| \\
&\qquad + \max_{m\le j\le n-m} E\,\Big|\frac{1}{m}\sum_{k=j+1}^{j+m}(\sigma_k^2 - E\sigma_k^2)\Big| \to 0,
\end{aligned}$$

for any $0<m=m_n\to\infty$ satisfying $n/m\to\infty$. Due to this fact, it follows from Lemmas 1 and 2 that (taking the same argument as in the proof of (55))

$$R_{1n} = E\sigma_1^2\int_0^1 H^2(X_{n,[nt]})dt\int K^2 + o_P(1).$$

Similarly, $R_{2n}=\int_0^1\big[1+H^2(X_{n,[nt]})\big]H^2(X_{n,[nt]})dt\int K^2 + o_P(1)$. Taking these estimates into (76), we have

$$\mathcal{V}_{n,11} = E\sigma_1^2\int_0^1 H^2(X_{n,[nt]})dt\int K^2 + o_P(1) = E\sigma_1^2\,V_{n,11} + o_P(1),$$

i.e., (75) for $i=j=1$ is proved.

We now turn back to the proof of Theorem 6, making use of the same notation in the proof of Theorem 5. First note that, by using similar arguments as above,

$$\sqrt{\frac{c_n}{nl_n}}\pi(d_n)^{-1}\,\mathcal{A}_n D_n = \left[1,\ -\frac{\frac{c_n}{nl_n}\sum_{k=1}^n \pi(d_n)^{-1} f_k K_{kn}}{\frac{c_n}{nl_n}\sum_{k=1}^n K_{kn}^*}\right]$$

$$= \left[1, \ -\frac{\int_0^1 H\left(X_{n,[nt]}\right) dt \int K}{\int K^*}\right] + o_P(1) \ =: \ \widehat{A}_n + o_P(1),$$

This, together with (74), yields that

$$\begin{aligned}
\frac{c_n}{nl_n}\pi(d_n)^{-2}\mathcal{A}_n\mathcal{V}_n\mathcal{A}_n' &= \sqrt{\frac{c_n}{nl_n}}\pi(d_n)^{-1}\,\mathcal{A}_n D_n\left(D_n^{-1}\mathcal{V}_n D_n^{-1}\right)\sqrt{\frac{c_n}{nl_n}}\pi(d_n)^{-1}\,D_n\mathcal{A}_n' \\
&= E\sigma_1^2\,\widehat{A}_n\,V_n\,\widehat{A}_n' + o_P(1). \qquad (77)
\end{aligned}$$

Now, by using the same arguments as in the proofs of (72) and (73), it follows from (77) that

$$\hat{T} = \frac{\sqrt{\frac{c_n}{nl_n}}\pi(d_n)^{-1}\sum_{k=1}^n Z_{kn}\overline{e_k}}{\sqrt{\frac{c_n}{nl_n}\pi(d_n)^{-2}\mathcal{A}_n\mathcal{V}_n\mathcal{A}_n'}} = \left(E\sigma_1^2\,\widehat{A}_n\,V_n\,\widehat{A}_n'\right)^{-1/2}\widehat{A}_n\,B_n + o_P(1) \to_d N(0,1),$$

as required. □

7.3 Proofs of Theorems 7 and 8

We only prove Theorem 8. The proof of Theorem 7 is similar and therefore omitted.

Recall $\widetilde{\mathbf{f}}_k = \left(\mathbf{f}_k', \mathbf{f}_{1k}'\right)'$, where $\mathbf{f}_{1k} = (k/n-\tau)\,\mathbf{f}_k$, and note that $\begin{bmatrix}\widetilde{\theta}(\tau)\\ \widetilde{\theta}^{(1)}(\tau)\end{bmatrix} = \frac{\sum_{k=1}^n y_k\widetilde{\mathbf{f}}_k K_{kn}}{\sum_{k=1}^n \widetilde{\mathbf{f}}_k\widetilde{\mathbf{f}}_k{}' K_{kn}}$. We may write

$$D_n\left(\begin{bmatrix}\widetilde{\theta}(\tau)\\ \widetilde{\theta}^{(1)}(\tau)\end{bmatrix} - \begin{bmatrix}\theta(\tau)\\ \theta^{(1)}(\tau)\end{bmatrix}\right) = Q_n^{-1}\left(\mathcal{M}_n + R_n\right), \qquad (78)$$

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

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

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

$$Q_n = \begin{bmatrix}\frac{c_n}{n}\sum_{k=1}^n \mathbf{f}_k\mathbf{f}_k' K_{kn} & \frac{c_n}{n}\sum_{k=1}^n \mathbf{f}_k\mathbf{f}_k' K_{1,kn}\\ \frac{c_n}{n}\sum_{k=1}^n \mathbf{f}_k\mathbf{f}_k' K_{1,kn} & \frac{c_n}{n}\sum_{k=1}^n \mathbf{f}_k\mathbf{f}_k' K_{2,kn}\end{bmatrix} \to_P Q_2. \qquad (79)$$

Similarly, the conditional matrix $[\mathcal{M}_n, \mathcal{M}_n]$ of the martingale $\mathcal{M}_n$ has the property:

$$[\mathcal{M}_n, \mathcal{M}_n] = D_n^{-1}\sum_{k=1}^n \sigma_k^2\widetilde{\mathbf{f}}_k\widetilde{\mathbf{f}}_k{}' K_{kn}^2\,D_n^{-1}$$

$$= \left[\begin{array}{cc} \frac{c_n}{n}\sum_{k=1}^n \sigma_k^2 \mathbf{f}_k\mathbf{f}_k' K_{kn}^2 & \frac{c_n}{n}\sum_{k=1}^n \sigma_k^2 \mathbf{f}_k\mathbf{f}_k'{}_{(1)}K_{kn}^2 \\ \frac{c_n}{n}\sum_{k=1}^n \sigma_k^2 \mathbf{f}_k\mathbf{f}_k'{}_{(1)}K_{kn}^2 & \frac{c_n}{n}\sum_{k=1}^n \sigma_k^2 \mathbf{f}_k\mathbf{f}_k'{}_{(2)}K_{kn}^2 \end{array} \right] \to_P \mathbf{\Omega}_2,$$

where ${}_{(\ell)}K^2 = x^\ell K^2(x)$ 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 the (9) in Theorem 1 . Combining these facts and (78), Theorem 8 will follow if we prove

$$R_n = o_P(1). \tag{80}$$

In fact, by noting

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

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

$$\begin{aligned} R_n &= D_n^{-1}\sum_{k=1}^n K_{kn}\left[\begin{array}{c}\mathbf{f}_k\\ \mathbf{f}_{1k}\end{array}\right]\left\{\mathbf{f}_k'\theta(k/n) - \left[\begin{array}{c}\mathbf{f}_k\\ \mathbf{f}_{1k}\end{array}\right]'\left[\begin{array}{c}\theta(\tau)\\ \theta^{(1)}(\tau)\end{array}\right]\right\} \\ &= D_n^{-1}\sum_{k=1}^n K_{kn}\left[\begin{array}{c}\mathbf{f}_k\\ \mathbf{f}_{1k}\end{array}\right]\left\{\mathbf{f}_k'\theta(k/n) - \mathbf{f}_k'\theta(\tau) - \mathbf{f}_{1k}'\theta^{(1)}(\tau)\right\} \\ &= (1/2)D_n^{-1}\sum_{k=1}^n (k/n-\tau)^2 K_{kn}\left[\begin{array}{c}\mathbf{f}_k\\ \mathbf{f}_{1k}\end{array}\right]\mathbf{f}_k'\theta^{(2)}(\overline{\tau}), \end{aligned}$$

indicating that

$$||R_n|| \leq C\sqrt{\frac{n}{c_n^5}}\,\frac{c_n}{n}\sum_{k=1}^n ||\mathbf{f}_k\,\mathbf{f}_k'||K_{2,kn} = O_P\big(\sqrt{n/c_n^5}\big) = o_P(1)$$

due to $n/c_n^5 \to 0$, where we have used (79) and the fact that $\theta^{(2)}(.)$ is uniformly bounded on $(0,1]$. This proves (80) and also completes the proof of Theorem 8. □

7.4 Proofs of Theorems 9 and 10

We only prove Theorem 10, The proof of Theorem 9 is similar and therefore omitted.

By recalling (79) and using Theorem 8, it suffices to show that

$$D_n^{-1}\tilde{\mathbf{\Omega}}_n D_n^{-1} \quad = \quad \mathbf{\Omega}_2 + o_P(1). \tag{81}$$

Indeed, since $Q_n = D_n^{-1}\tilde{\mathcal{Q}}_n D_n^{-1}$, it follows from (79) and (81) that

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

As a consequence, for $i=1,2$ and $j=i+2$, we have

$$\frac{n}{c_n}\left[\tilde{\mathcal{Q}}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{\mathcal{Q}}_n^{-1}\right]_{ii} = (A_n)_{ii} \to_p \left[Q_2^{-1}\mathbf{\Omega}_2 Q_2^{-1}\right]_{ii}, \tag{82}$$

$$\frac{n}{c_n^3}\left[\tilde{\mathcal{Q}}_n^{-1}\tilde{\mathbf{\Omega}}_n\tilde{\mathcal{Q}}_n^{-1}\right]_{jj} = (A_n)_{jj} \to_p \left[Q_2^{-1}\mathbf{\Omega}_2 Q_2^{-1}\right]_{jj}. \tag{83}$$

It follows from (27) and (82) 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 (31). Similarly, it follows from (27) and (83) 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 (32) i.e. the second limit result of Theorem 10.

We next prove (81). It is readily seen that

$$D_n^{-1}\tilde{\mathbf{\Omega}}_n D_n^{-1} = \begin{bmatrix} \frac{c_n}{n}\sum_{k=1}^n \tilde{e}_k^2\mathbf{f}_k\mathbf{f}_k'\ {}_{(0)}K_{kn}^2 & \frac{c_n}{n}\sum_{k=1}^n \tilde{e}_k^2\mathbf{f}_k\mathbf{f}_k'\ {}_{(1)}K_{kn}^2 \\ \frac{c_n}{n}\sum_{k=1}^n \tilde{e}_k^2\mathbf{f}_k\mathbf{f}_k'\ {}_{(1)}K_{kn}^2 & \frac{c_n}{n}\sum_{k=1}^n \tilde{e}_k^2\mathbf{f}_k\mathbf{f}_k'\ {}_{(2)}K_{kn}^2 \end{bmatrix}, \tag{84}$$

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 8. Recalling $\tilde{e}_k = y_k - \tilde{\theta}(\tau)'\mathbf{f}_k$ and noting

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

due to Theorem 8 and the smoothing condition on $\theta(\tau)$, we have

$$\tilde{e}_k^2 = \big\{\sigma_k u_k - \big[\tilde{\theta}(\tau) - \theta(k/n)\big]'\mathbf{f}_k\big\}^2 = \sigma_k^2 u_k^2 + \Delta_{nk}, \tag{85}$$

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

$$\begin{aligned} |\Delta_{nk}| &\le 2|\sigma_k u_k|\big[|\tau - k/n| + o_P(1)\big]\, ||\mathbf{f}_k|| + \big[|\tau - k/n| + o_P(1)\big]^2\, ||\mathbf{f}_k||^2 \\ &\le \big[|\tau - k/n| + o_P(1)\big]\sigma_k^2 u_k^2 + 3\big[|\tau - k/n| + o_P(1)\big]\, ||\mathbf{f}_k||^2 \\ &:= \Delta_{1,nk}\,\sigma_k^2 u_k^2 + \Delta_{2,nk}. \end{aligned}$$

It follows from (46) with $G(.) = 1$, $v_k = ||\mathbf{f}_k||^4$ and Remark 21 (recalling $\int |x|^3 K^2 < \infty$ and

$E||\mathbf{f}_k||^4 < \infty$)that, for $\ell = 0, 1, 2,$

$$\begin{aligned} & \frac{c_n}{n} \sum_{k=1}^{n} |\Delta_{2,nk}| \, ||\mathbf{f}_k \mathbf{f}_k'|| \ {}_{(\ell)}K_{kn}^2 \\ \leq \ & \frac{C\, c_n}{n} \sum_{k=1}^{n} ||\mathbf{f}_k||^4 \, || \big[o_P(1) \ {}_{(\ell)}K_{kn}^2 + c_n^{-1} \ {}_{(\ell+1)}K_{kn}^2 \big] = o_P(1), \end{aligned} \tag{86}$$

due to $c_n \to \infty$. Now (81) will follow if we prove, for $\ell \leq 3$ and any $\alpha \in \mathbb{R}^2$,

$$\frac{c_n}{n} \sum_{k=1}^{n} \sigma_k^2 u_k^2 \left(\alpha' \mathbf{f}_k\right)^2 \ {}_{(\ell)}K_{kn}^2 = E\big[\sigma_1^2 \, (\alpha' \mathbf{f}_1)^2\big] \int x^\ell K^2 + o_P(1). \tag{87}$$

Indeed, by (87), we have

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

for $\ell = 0, 1, 2$. This, together with (86), yields that, for $\ell = 0, 1, 2,$

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

Now, by (85) and (87), we have

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

for $\ell = 0, 1, 2$. Taking this result into (84), we obtain (81).

We finally prove (87). In fact, by letting $v_k = \sigma_k^2 u_k^2 [\alpha' \mathbf{f}_k]^2$, where $\alpha \in \mathbb{R}^2$ and recalling that $E\big(u_k^2 | \mathcal{F}_{k-1}\big) = 1$ and σ_k are $\mathcal{F}_{k-1}$ measurable, it is readily seen that $A_0 :=$

$E\big(\sigma_1^2[\alpha'\mathbf{f}_1]^2\big) = Ev_k$ for each $k \geq 1$ and (recalling $\sup_{k\geq 1} Eu_k^4 < \infty$ and **A2** with $g = f$)

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

for any $0 < m = m_n \to \infty$ satisfying $n/m \to \infty$. Due to this fact, result (87) follows from Lemmas 1 with $G(.) \equiv 1$. The proof of Theorem 10 is now complete. $\square$

8 Additional workings for Section 3.2

We next provide some additional discussion of how the methods of Section 3.2 for TVP models generalise to multi-covariate regressions of the form

$$y_k = \mu(k/n) + \sum_{j=1}^{p-1} \beta_j\,(k/n)\cdot f_j(x_{k-1,j}) + e_k,\;\; k = 1, ..., n,$$

$p \geq 2$. Set $\theta(\tau)' := [\mu(\tau), \beta_1(\tau), ..., \beta_{p-1}(\tau)]$. From technical point of view this kind of generalisation is straightforward and can be established along the lines of existing proofs and the fact that Theorem 1 applies to multivariate stationary processes (see else Remark 1). In the multivariate case, the LLev estimator $\theta(\tau)$ is defined as

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

where $\tau \in (0, 1]$ and $\mathbf{f}_k' := [1, f_1(x_{k-1,1}), ..., f_{p-1}(x_{k-1,p-1})]$. Further, the LLin estimator is given by

$$\left[\begin{array}{c}\widetilde{\theta}(\tau)\\ \widetilde{\theta}^{(1)}(\tau)\end{array}\right] := \arg\min_{(\mathbf{a}\times\mathbf{b})\in\mathbb{R}^p\times\mathbb{R}^p} \sum_{k=1}^{n} \big(y_k - \mathbf{a}'\mathbf{f}_k - \mathbf{b}'\mathbf{f}_{1k}\big)^2 K\left[c_n\left(k/n - \tau\right)\right],$$

where as before $\theta^{(1)}(\tau) := \partial\theta(\tau)/\partial\tau$ and $\mathbf{f}_{1k} := (k/n - \tau)\,\mathbf{f}_k$. Set $\mathbf{F}_k := [f_1(x_{k,1}), ..., f_{p-1}(x_{k,p-1})]'$ and redefine Q and $\mathbf{\Omega}$ as

$$Q = \left[\begin{array}{cc} 1 & E\mathbf{F}_1' \\ E\mathbf{F}_1 & E\mathbf{F}_1\mathbf{F}_1'\end{array}\right] \quad \text{and} \quad \mathbf{\Omega} = \left[\begin{array}{cc} E\sigma_2^2 & E\big\{\sigma_2^2\mathbf{F}_1'\big\} \\ E\big\{\sigma_2^2\mathbf{F}_1\big\} & E\big\{\sigma_2^2\mathbf{F}_1\mathbf{F}_1'\big\}\end{array}\right].$$

Then we have the following generalisation of Theorem 7

$$\sqrt{\frac{n}{c_n}}\left(\hat{\theta}(\tau)-\theta(\tau)\right)\to_d \mathbf{N}\left(\mathbf{0},Q_1^{-1}\mathbf{\Omega}_1Q_1^{-1}\right),$$

with $Q_1=Q\int K$ and $\mathbf{\Omega}_1=\mathbf{\Omega}\int K^2$. Moreover let $\otimes$ be the Kronecker product and redefine $D_n=diag\left\{\sqrt{\frac{n}{c_n}},\sqrt{\frac{n}{c_n^3}}\right\}\otimes I_p$ where I_p is a p-dimensional identity matrix. Then Theorem 8 generalises to

$$D_n\left(\begin{bmatrix}\widetilde{\theta}(\tau)\\ \widetilde{\theta}^{(1)}(\tau)\end{bmatrix}-\begin{bmatrix}\theta(\tau)\\ \theta^{(1)}(\tau)\end{bmatrix}\right)\to_d \mathbf{N}\left(\mathbf{0},Q_2^{-1}\mathbf{\Omega}_2Q_2^{-1}\right),$$

with,

$$Q_2=\begin{bmatrix}Q\int K & Q\int xK\\ Q\int xK & Q\int x^2K\end{bmatrix}\quad\text{and}\quad\mathbf{\Omega}_2=\begin{bmatrix}\mathbf{\Omega}\int K^2 & \mathbf{\Omega}\int xK^2\\ \mathbf{\Omega}\int xK^2 & \mathbf{\Omega}\int x^2K^2\end{bmatrix}.$$

9 Additional workings for Section 3.3

Consider

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

Set $\Delta_n := n^{-1}\sum_{k=1}^n \left[x_{k-1} - \left(n^{-1}\sum_{j=1}^n x_{j-1}\right)\right]^2$, and note that by Theorem 1, $\Delta_n \to_P Ex_{k-1}^2$, and

$$\left\{\frac{1}{n}\sum_{k=1}^n \begin{bmatrix} 1 & x_{k-1} \\ x_{k-1} & x_{k-1}^2 \end{bmatrix}\right\}^{-1} = \Delta_n^{-1}\begin{bmatrix} n^{-1}\sum_{k=1}^n x_{k-1}^2 & -n^{-1}\sum_{k=1}^n x_{k-1} \\ -n^{-1}\sum_{k=1}^n x_{k-1} & 1 \end{bmatrix}$$

$$\to_P \begin{bmatrix} 1 & 0 \\ 0 & 1/Ex_1^2 \end{bmatrix}.$$

In view of the above, standard arguments show that the intercept estimator is

$$\sqrt{n}\left(\tilde{\mu}_{OLS} - \int_0^1 \mu(\tau)d\tau\right) = [1 + o_P(1)]\, n^{-1/2}\sum_{k=1}^n e_k \to_d \mathbf{N}(0, Ee_1^2),$$

where we have used the fact that the Euler sum $n^{-1}\sum_{k=1}^n \mu(k/n) - \int_0^1 \mu(\tau)d\tau = O(n^{-1})$. Next, we consider the OLS estimator for the case $0 < d < 1/2$. The derivations for the short memory case are similar and will be omitted. In view of the above

$$\begin{aligned}\frac{n}{\delta_n}\left(\tilde{\beta} - \beta\right) &= [1 + o_P(1)]\left[Ex_{k-1}^2\right]^{-1}\left\{\delta_n^{-1}\sum_{k=1}^n x_{k-1}e_k + \delta_n^{-1}\sum_{k=1}^n x_{k-1}\mu(k/n)\right. \\ &\quad \left. - \left(n^{-1}\sum_{j=1}^n \mu(j/n)\right)\left(\delta_n^{-1}\sum_{k=1}^n x_{k-1}\right)\right\}\end{aligned}$$

$$= \left[Ex_{k-1}^2\right]^{-1}\left\{\delta_n^{-1}\sum_{k=1}^n x_{k-1}\mu(k/n) - \left(n^{-1}\sum_{j=1}^n \mu(j/n)\right)\left(\delta_n^{-1}\sum_{k=1}^n x_{k-1}\right)\right\} + O_P\left(\frac{\sqrt{n}}{\delta_n}\right)$$

$$\to_d \left(Ex_1^2\right)^{-1}\left[1, -\int_0^1 \mu(\tau)d\tau\right] \cdot \mathbf{N}\left(\mathbf{0}, E\left(\xi_1^2\right)\Psi\right),$$

with Ψ being a variance matrix such that

$$\delta_n^{-1}\left[\sum_{k=1}^n x_{k-1}\mu(k/n), \sum_{k=1}^n x_{k-1}\right] \to_d \mathbf{N}\left(\mathbf{0}, E\left(\xi_1^2\right)\Psi\right). \tag{88}$$

The latter limit distribution result follows from standard argument e.g. Lindeberg-Feller CLT. We provide some key derivations that yield the form of Ψ. Without loss of generality set $E\left(\xi_1^2\right) = 1$. Recall that under our assumptions $x_k = \sum_{j=0}^{k-1}\phi_k\xi_{k-j}$, with $\phi_j \sim cons.j^{-\nu}$, $\nu = 1 - d$ and $0 < d < 1/2$. Set

$$S_n' = \sum_{k=1}^n \mu\left(k/n\right)x_{k-1} = \sum_{k=1}^n\sum_{s=k}^n \mu\left(1 - s/n\right)\phi_{s-k}\xi_k$$

and

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

For λ_1, $\lambda_2 \in \mathbb{R}$, we shall show that the martingale array

$$\left[\lambda_1 \delta_n^{-1} S_n' + \lambda_2 \delta_n^{-1} S_n''\right]$$

converges to a normal that has asymptotic variance determined by the limit of (e.g. Hall and Heyde, 1980; Corollary 3.1)

$$\sum_{k=1}^{n} E_{\mathcal{F}_{k-1}} \left[\lambda_1 \delta_n^{-1} \sum_{s=k}^{n} \mu\left(1 - j/n\right) \phi_{j-k}\xi_k + \lambda_2 \delta_n^{-1} \sum_{s=k}^{n} \phi_{j-k}\xi_k\right]^2$$

$$= \sum_{k=1}^{n} \left[\lambda_1 \delta_n^{-1} \sum_{s=k}^{n} \mu\left(1 - j/n\right) \phi_{j-k} + \lambda_2 \delta_n^{-1} \sum_{s=k}^{n} \phi_{j-k}\right]^2$$

$$= \frac{1}{n}\sum_{k=1}^{n} \left[\frac{\lambda_1}{n} \sum_{j=k}^{n} \mu\left(1 - j/n\right) \left(\frac{j-k}{n}\right)^{-\nu} + \frac{\lambda_2}{n} \sum_{j=k}^{n} \left(\frac{j-k}{n}\right)^{-\nu}\right]^2 + o(1)$$

$$= \int_0^1 \left[\lambda_1 \int_r^1 \mu\left(1-s\right)\left(s-r\right)^{-\nu} ds + \lambda_2 \int_r^1 \left(s-r\right)^{-\nu} ds +\right]^2 dr + o(1).$$

Indeed in view of the above together with a Lindeberg condition we get (88).

Next we consider the limit behaviour of the estimator for Ee_1^2. Let $\tilde{e}_k$ be the OLS residuals. Then

$$\frac{1}{n}\sum_{k=1}^{n} \tilde{e}_k^2 = \frac{1}{n}\sum_{k=1}^{n} \left[\mu(k/n) - \tilde{\mu} + \left(\beta - \tilde{\beta}\right) x_{k-1} + e_k\right]^2$$

$$= \frac{1}{n}\sum_{k=1}^{n} \left[\mu(k/n) - \tilde{\mu} + \left(\beta - \tilde{\beta}\right) x_{k-1}\right]^2 + \frac{1}{n}\sum_{k=1}^{n} e_k^2$$

$$+ \frac{2}{n}\sum_{k=1}^{n} \left[\mu(k/n) - \tilde{\mu} + \left(\beta - \tilde{\beta}\right) x_{k-1}\right] e_k$$

$$\to_P \int_0^1 \mu(\tau)^2 d\tau - \left(\int_0^1 \mu(\tau) d\tau\right)^2 + Ee_1^2,$$

where we have used the fact that

$$\frac{1}{n}\sum_{k=1}^{n} \left[\mu(k/n) - \tilde{\mu}\right]^2 = \frac{1}{n}\sum_{k=1}^{n} \mu(k/n)^2 - \frac{2\tilde{\mu}}{n}\sum_{k=1}^{n} \mu(k/n) + \tilde{\mu}^2$$

$$\to_P \int_0^1 \mu(\tau)^2 d\tau - \left(\int_0^1 \mu(\tau) d\tau\right)^2.$$

Finally, consider the OLS based statistic for the null hypothesis $H_0 : \beta = \beta_0$, $\beta_0 \in \mathbb{R}$. Using (34) and (36) we get

$$\begin{aligned}
\left|\tilde{t}_{OLS}\right| &= \left| \frac{\tilde{\beta} - \beta}{\sqrt{\left(\frac{1}{n}\sum_{k=1}^n \tilde{e}_k^2\right)\left[\sum_{k=1}^n x_k^2 - n^{-1}\left(\sum_{k=1}^n x_k\right)^2\right]^{-1}}} \right| + o_P(1) \\
&= \left| \frac{\sqrt{n}\left(\tilde{\beta} - \beta\right)}{\sqrt{\left(\frac{1}{n}\sum_{k=1}^n \tilde{e}_k^2\right)\left[n^{-1}\sum_{k=1}^n x_k^2 - \left(n^{-1}\sum_{k=1}^n x_k\right)^2\right]^{-1}}} \right| \\
&= [1 + o_P(1)] \frac{\delta_n}{\sqrt{n}} \left| \frac{\frac{n}{\delta_n}\left(\tilde{\beta} - \beta\right)}{\sqrt{\left(\frac{1}{n}\sum_{k=1}^n \tilde{e}_k^2\right)\left[n^{-1}\sum_{k=1}^n x_k^2\right]^{-1}}} \right| \to_P \infty.
\end{aligned}$$

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. and Varneskov, R.T. (2020). Consistent inference for predictive regressions in persistent economic systems. *Journal of Econometrics*, in press.

[3] Astrauskas, A. (1983). Limit theorems for sums of linearly generated random variables. *Lithuanian Mathematical Journal*, **23**, 127–134.

[4] Bandi, F.M, Perron, B, Tamoni, A. and Tebaldi, C. (2018). The scale of predictability. *Journal of Econometrics*, **208**(1), 120–140.

[5] Berenguer-Rico, V. and Gonzalo, J. (2014). Summability of stochastic processes - A generalization of integration for non-linear processes. *Journal of Econometrics*, **178**(2), 331–341.

[6] Billingsley P. (1968). *Convergence of probability measures.* Wiley, New York.

[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**, 409–424.

[8] Buchmann, B. and Chan, N.H. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *Annals of Statistics*, **35**(5), 2001–2017.

[9] Campbell, J.Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics,* **81**, 27–60.

[10] Cavanaght, C.L., Elliot, G. and Stock, J. (1995). Inference in models with nearly integrated regressors. *Econometric Theory*, **11**, 1131–1147.

[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] Chan, N. and Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics*, **185**, 182–195.

[13] Choi, I. and Saikkonen, P. (2010). Tests for nonlinear cointegration. *Econometric Theory*, **26**(2), 682–709.

[14] 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**, 343–371.

[15] Christensen, B.J. and Nielsen M.Ø. (2007). The effect of long memory in volatility on stock market fluctuations. *Review of Economics and Statistics*, **89**, 684–700.

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

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

[18] Demetrescu, M., Georgiev, I., Rodrigues, P. and Taylor, R. (2020). Testing for episodic predictability in stock returns. *Journal of Econometrics*, in press.

[19] Duffy J.A. and Kasparis, I. (2018). Regressions with Fractional $d = 1/2$ and Weakly Nonstationary Processes. arXiv:1812.07944v1

[20] 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

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

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

[23] Giraitis, L., Kapetanios, G., Yates, T. (2018). Inference on multivariate heteroscedastic stochastic time varying coefficient models. *Journal of Time Series Analysis*, **39**, 129–149.

[24] Gonzalo, J. and Pitarakis, J-Y. (2012). Regime-Specific Predictability in Predictive Regressions. *Journal of Business and Economic Statistics*, **30**(2), 229–241.

[25] Elliott, G., Müller, U.K. and Watson, M.W. (2015) Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, **83**, 771–811.

[26] Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application.* Academic Press, New York.

[27] Hjalmarsson, E. (2011). New methods for inference in long-horizon regressions. *Journal of Financial and Quantitative Analysis*, **46**, 815-839

[28] Hu, Z., Phillips, P.C.B. and Wang, Q. (2021). Nonlinear cointegrating power regression with endogeneity. *Econometric Theory*, https://doi.org/10.1017/S0266466620000560

[29] Hualde, J. and Robinson, P.M. (2010). Semiparametric inference in multivariate fractionally cointegrated systems. *Journal of Econometrics*, **157**(2), 492-511.

[30] Janson, M. and Moreira, J.M. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, **74**(3), 681–714.

[31] Jeganathan, P. (2008). Limit theorems for functionals of sums that converge to fractional Brownian and stable motions. Cowles Foundation Discussion Paper 1649.

[32] Johansen, S. (1995). *Likelihood-based inference in cointegrated vector auto-regressive models.* Oxford University Press, New York.

[33] Johansen, S. and Nielsen, M.Ø. (2012). A necessary moment condition for the fractional functional central limit theorem. *Econometric Theory*, **28**(3), 671–679.

[34] Ibragimov, I.A. and Linnik, Y.V. (1971) *Independent and stationary sequences of random variables.* Wolters-Noordhoff Publishing, Groningen, The Netherlands.

[35] Ibragimov, M., Ibragimov, R. and Walden, J. (2015). *Heavy-Tailed Distributions and Robustness in Economics and Finance.* Lecture Notes in Statistics Vol. 214, Springer.

[36] Kasparis, I. (2010). The Bierens test for certain nonstationary models. *Journal of Econometrics*, **158**, 221–230.

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

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

[39] Lettau, M. and Ludvigson, S. (2001). Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying. *Journal of Political Economy*, **109**(6), 1238–1287.

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

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

[42] Lin, Y., Tu, Y. and Yao, Q. (2020). Estimation for double-nonlinear cointegration. *Journal of Econometrics*, **216**(1), 175–191.

[43] Magdalinos, T. (2020). Least squares and IVX limit theory in systems of predictive regressions with GARCH innovations. *Econometric Theory*, forthcoming.

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

[45] Marmer. V. (2007). Nonlinearity, nonstationarity and spurious forecasts. *Journal of Econometrics*, 142. 1–27.

[46] Menly, L., Santos, T. and Veronesi, P. (2004). Understanding Predictability. *Journal of Political Economy*, **112**(1), 1–47 .

[47] Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, **75**(5), 1411–1452.

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

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

[50] Peligrad, M. and Utev, S. (1997). Central limit theorem for linear processes. *Annals of Probability*, **25**(1), 443–456.

[51] Phillips, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica*, **59**(2), 283–306.

[52] Phillips, P.C.B. (1995). Fully modified least squares and vector autoregression. *Econometrica*, **63**(5) 1023-1078.

[53] Phillips, P.C.B. (2014). On confidence intervals for autoregressieve roots and predictive regressions. *Econometrica*, **82**(3), 1177-1195.

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

[55] Phillips, P.C.B. and Hansen, B. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, **57**(1), 99-125

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

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

[58] Phillips, P.C.B. and Shimotsu, K. (2004) Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics*, **32**(2), 656–692.

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

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

[61] Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics* **54**, 783–820.

[62] Sun, Y., Phillips, P.C.B. and Sainan, J. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, **76**(1), 175–194.

[63] Valkanov, R. (2003). Long-horizon regressions: Theoretical results and applications. *Journal of Financial Economics* **68**, 201–232.

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

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

[66] Wang, Q. (2021). Least squares estimation for nonlinear regression models with heteroscedasticity. *Econometric Theory*, https://doi.org/10.1017/S0266466620000493.

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

[68] Wang, Q. and Phillips P.C.B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, **77**(6), 1901-1948.

[69] Wang, Q. and Phillips, P.C.B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory*, **27**(2), 235-259.

[70] Wang, Q. and Phillips, P.C.B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, **40**(2), 727-758.

[71] Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21**, 1455–508.