# CHRONOLOGICALLY TRIMMED LS FOR NONLINEAR PREDICTIVE REGRESSIONS WITH PERSISTENCE OF UNKNOWN FORM *

Zhishui Hu
*University of Science and Technology of China*

Ioannis Kasparis
*University of Cyprus*

Qiying Wang
*The University of Sydney*

May 9, 2024

**Abstract**

Relatively recent work by Jeganathan (2008) and Wang (2014) on generalized martingale central limit theorems *(MCLTs)* implicitly introduces a new class of instrument arrays that yield (mixed) Gaussian limit theory irrespective of the persistence level in the data. Motivated by these developments, we propose a new semi-parametric method for estimation and inference in nonlinear predictive regressions with persistent predictors. The proposed method that we term *Chronologically Trimmed Least Squares (CTLS)* is comparable to the IVX method of Magdalinos and Phillips (2009) and yields conventional inference in regressions where the nature and extent of persistence in the data is uncertain. In terms of model generality, our contribution to the existing literature is twofold. First, our covariate model space allows for both nearly integrated (NI) and fractional processes (stationary and nonstationary) as a special case, whilst the vast majority of papers in this area only consider NI arrays. Second, we allow for nonlinear regression functions. The CTLS estimator is obtained by applying certain *chronological trimming* to the OLS instruments using appropriate kernel functions of time trend variables. In particular, the instruments under consideration are a generalized (averaged) version of those widely used for time-varying parameter (TVP) models. For the purposes of our analysis, we develop a novel asymptotic theory for sample averages of various processes weighted by such kernel functionals which is of independent interest and highly relevant to the TVP literature. Leveraging our nonlinear framework, we also provide an investigation on the effects of misbalancing on the predictability hypothesis. A new methodology is proposed to mitigate misbalancing effects. These methods are used for exploring the predictability of SP500 returns.

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

# 1 Introduction

Estimation and inference under temporal dependence are challenging tasks. 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 toward 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.

The 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 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 make 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. In general, two main approaches have been proposed to address this issue in the predictive regression literature. The first approach relies on so-called conservative methods; see, e.g. Mikusheva (2007), Phillips (2014), for a review. An alternative approach that has gained a lot of attention lately is the semi-parametric IVX method first proposed by Magdalinos and Phillips (2009, MP hereafter). This is an IV method that yields conventional inference for a general class of regressors via signal reduction in the instruments. Further, it should be mentioned that there is related literature for fractional cointegration systems (e.g. Hualde and Robinson, 2010) that has received little attention in the predictability area. More recently, Jin and Wang (2024) investigated self-weighted estimation for nonlinear cointegrating regression. A more detailed review of existing methods for robust inference under temporal dependence can be found in Section S1 of the Online Supplement.

In this article, we consider a reduced signal IV method comparable to IVX. We develop estimation methods that yield conventional inference in predictive regressions that are nonlinear in variables, with nonlinearities of known form -see, e.g. Park and Phillips (1999, 2001). In particular, we consider linear in parameter models that allow for a wide range of dependence in the data including stationary or nonstationary long memory as well as NI fractional arrays (e.g., Buchmann and Chan, 2007). The proposed methods, which we term Chronologically Trimmed LS (CTLS), share the same underlying principle as the IVX method.

The CTLS method is partly encouraged by recent developments in generalized MCLTs (e.g., Jeganathan, 2008, Wang, 2014, 2015). These generalized MCLTs entail asymptotic orthogonality conditions between triangular arrays and martingale difference terms that correspond to instruments and regression errors, respectively. The validity of the aforementioned orthogonality conditions requires instruments of weaker signal than that of the OLS instrumentation. These high level requirements effectively define a class of instrument arrays that, similarly to IVX, yield conventional limit theory (normal or mixed normal) irrespective of the persistence level in the data. To illustrate how CTLS instrumentation achieves signal reduction, consider the simple predictive regression

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

where $e_k$ is a regression error term. In this case, the proposed CTLS instruments for the estimate of $\beta$ are a generalized version of

$$Z_{kn}(\tau) = K\left[c_n\left(k/n - \tau\right)\right] f(x_{k-1}),\ 0 < \tau < 1,\ c_n^{-1} + c_n n^{-1} \to 0 \tag{2}$$

with $K > 0$ being an integrable kernel function, and $c_n$ is a (reciprocal) bandwidth term. Notice that for $f(x) = x$, the term in eq. (2), is exactly the instrument utilized for the estimation of TVPs in a number of studies[1], e.g. Robinson (1991), Giraitis et al. (2014), Phillips et al. (2017), Giraitis et al. (2021), Hu et al. (2024). The term $Z_{kn}(\tau)$, is a reduced signal version of the OLS instrument $f(x_{k-1})$. To see this, without loss of generality, assume that $\lim_{x\to\infty} K(|x|) = 0$. It then follows from (2) that for $k/n \neq \tau$, $K\left[c_n\left(k/n - \tau\right)\right] \to 0$ as $n \to \infty$. The particular instrument extracts information locally around the "*chronological point* $\tau$", whilst trimming applies for values of $k/n$ away from $\tau$. 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 that of the OLS estimator.

Despite the fact that $Z_{kn}(\tau)$ is a reduced signal instrument, its utilization does not yield conventional inference under nonstationarity when there is more than one regression parameter, e.g., eq. (1). In fact, Phillips et al. (2017) demonstrate that when covariates are $I(1)$ and the dimensionality of parameter space is strictly greater than unity, limit distributions associated with $Z_{kn}(\tau)$ instrumentation are comparable to those of OLS i.e. they are non-conventional determined by stochastic integrals. This phenomenon occurs because the *"Hessian"* is singular with multiple convergence rates arising from this singularity. We refer to Phillips et al. (2017) for more details. Alternatively, the CTLS estimator proposed here involves averaging over a multitude of chronological points. In particular, CTLS instruments are of the form

$$Z_{kn} = \sum_{j=1}^{l} Z_{kn}(\tau_j),\ 0 < \tau_1 < ... < \tau_l < 1,$$

with $Z_{kn}(.)$ defined in (2) and $l$ is either fixed or $l = l_n \to \infty$, as $n \to \infty$. We demonstrate that under nonstationarity (e.g., $x_t$ is an NI fractional process), $Z_{kn}$ yields conventional limit theory as long as the number of chronological points (i.e., $l$) is no smaller than the dimensionality of the parameter space and $l_n$ does not diverge too fast. The former restriction rules out a singularity in the Hessian, while the second restriction ensures that non-trivial trimming (i.e. sufficient signal reduction) applies. The signal reduction achieved by CTLS results in vanishing endogeneity, between $Z_{kn}$ and $e_t$, and therefore to mixed Gaussian limit distributions.

The contribution of this paper can be summarized as follows.

**(i)** First, we provide a general theoretical contribution. Recent developments in general-

[1]The usual assumption in the TVP literature is that regression coeffients are of the form $\beta(k/n)$ with $\beta$ being a function $\beta : [0, 1] \to \mathbb{R}$.

ized MCLTs reveal that there is a whole class of reduced signal instruments that yield conventional inference similar to IVX. Therefore, more research is required to identify other potential methods and evaluate their relative merits.

**(ii)** Second, the current work allows for a substantially more general regressor space than that considered by most research articles in this area. Existing studies focus primarily only on NI predictors. For example, the ARFIMA class has received very little attention in the predictive regression literature, if any, despite the fact that it can accommodate precesses exhibiting far more general levels of persistence than those of the usual NI processes. Note that if $\{x_t\}_{t=1}^n$ is an NI array, we have the following *single* order of magnitude $\sum_{t=1}^n x_t = O_p(n^{3/2})$. On the other hand, if $x_t \sim I(d)$ with memory parameter $d > -1/2$ we get $\sum_{t=1}^n x_t = O_p\left(n^{1/2+d}\right)$, which gives a continuum of asymptotic rates.[2] In the current work, we allow for wide range of stationary and nonstationary covariates that encompass both NI and fractional processes as special cases, e.g. the NI fractional (NIF, hereafter) nonstationary array[3] (see also Buchmann and Chan, 2007; Kasparis et al., 2015)

$$x_t = (1 + c/n)\, x_{t-1} + v_t,\;\; v_t \sim I(\delta),\; \delta > -1/2. \tag{3}$$

**(iii)** Further, this article allows for nonlinear regression models (e.g. Park and Phillips, 1999, 2001), generalizing the linear predictive regressions commonly used in practice. Nonlinear specifications can potentially address misbalancing issues that are typically encountered in the applied predictive literature, i.e. situations where the predictor is more persistent than the dependent variable -see, e.g. Kasparis, Andreou and Phillips (2015), Phillips (2015), and the discussion below. A new methodology, which entails rolling test statistics of sequences of flexible functional forms, is proposed for mitigating misbalancing effects related to the predictability hypothesis.

**(iv)** Finally, this work provides a novel limit theory of independent interest. For example, we generalize asymptotics for kernel functionals related to the literature on TVP models in various directions. Among other things, we consider kernel functionals that entail substantially more general nonstationary covariates than those in the existing studies, e.g., Phillips et al. (2017). Furthermore, we consider kernel sample functionals that involve stationary and nonstationary processes, possibly of long memory, at the same time.

Finally, we provide some comparison between CTLS and IVX. MP show that IVX can accommodate NI and mildly integrated (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, see Theorem 3.2 in Duffy and Kasparis (2018), that IVX, probably after some minor modification, is also valid for fractional processes. Our theoretical framework does not allow for MI processes but readily allows for fractional predictors and nonlinear regression functions. The current results could be generalized to MI

[2]Notice that an $I(d)$ process of memory $d > -1/2$, can be either nonstationary or stationary, possibly of negative memory.

[3]The memory of $x_t$ under this specification is $d = 1 + \delta$. Note that for $\delta = 0$, $x_t$ is the usual NI process considered in the predictability literature. On the other hand, for $c = 0$, we get a fractional process with memory parameter $d > 1/2$.

covariates, but generalization would require a drastically different asymptotic machinery, e.g., a generalization of the asymptotic theory provided by Duffy and Kasparis (2021). We leave an extension towards this direction for future work. In terms of implementation, both methods are of comparable complexity, i.e. they both rely on studentized IV estimators. The simulation study provided here indicates that, for the usual linear specifications with NI covariates, IVX-based inference has better finite sample performance relative to CTLS. Nevertheless, as mentioned above, CTLS is readily available to nonlinear models and in situations where covariates are nonstationary fractional arrays.

The remainder of this work is organized as follows. Section 2 introduces the model and the main assumptions. Section 3 presents the CTLS instruments and related estimators. Section 4 considers CTLS-based predictive tests in a univariate setting. CTLS inference in a multicovariate setting is provided in Section S2 of the Online Supplement. Section 5 provides a theoretical investigation into the consequences of misbalancing due to functional form misspecification on CTLS predictability tests, and proposes a new methodology for mitigating adverse effects due to this phenomenon. Section 6 is a simulation study. An empirical application to the predictability of stock returns is the subject of Section 7. Our basic limit theory is presented in the Appendix, i.e., Sections 8 and 9. All proofs are provided in the Online Supplement.

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 $\lim_{n\to\infty} a_n/b_n = 1$. $I(A)$ is the indicator function on set $A$. We may write the integral $\int_{\mathbb{R}} f(x)dx$ ($\int_A f(x)dx$) as $\int f$ ($\int_A f$). $\Rightarrow$ denotes weak convergence in the space $D[0,1]$. For a vector $x$, $\|x\|$ is its inner product norm and $x'$ its transpose. For a matrix $A$, $\|A\|$ denotes its matrix norm. For a matrix $A$, $[A]_{ij}$ denotes its $(i,j)$ element. By $[x]$ we denote the integer part of a positive number $x$. Definitional (distributional) equality is denoted as $:=$ ($=_d$). Finally, $\operatorname{diag}\{a_1, ..., a_p\}$ denotes a (block) diagonal matrix with (blocks) elements $\{a_1, ..., a_p\}$ on the main diagonal, $\to_d$ denotes convergence in distribution, and $Y := \mathbf{MN}(\mathbf{0}, \Sigma)$ denotes a mixed Gaussian variate with characteristic function $\psi(t) = Ee^{it'Y} = Ee^{-t'\Sigma t/2}$.

# 2 Model and Assumptions

We consider the nonlinear in variables regression model

$$y_k = \mu + \boldsymbol{\beta}'\mathbf{f}(\mathbf{x}_{k-1}) + e_k, \quad e_k = \sigma_k u_k, \quad k = 1, ..., n, \tag{4}$$

where $\mu$ is an intercept, $\boldsymbol{\beta} = [\beta_1, ..., \beta_p]'$ a vector of slope parameters, $\mathbf{x}_k = [x_{k,1}, ..., x_{k,p}]'$ a vector of covariates, and $\mathbf{f} : \mathbb{R}^p \to \mathbb{R}^p$ known regression functions with $\mathbf{f}(\mathbf{x_k}) = [f_1(x_{k,1}), ..., f_p(x_{k,p})]'$. Without loss of generality, it is convenient for presentation purposes to use the following initialization for the covariate vector $\mathbf{x}_0 = \mathbf{0}$. The process $u_k$ together with some filtration $\mathcal{F}_k$ forms a martingale difference sequence such that $\mathbb{E}(u_k^2 \mid \mathcal{F}_k) = 1$ *a.s.*, and $\mathbf{x}_k$ is $\mathcal{F}_k$-measurable. Finally, $\sigma_k$ is a volatility process allowing for a variety of conditionally heteroscedastic regression errors ($e_k$), e.g. strictly stationary ARCH($\infty$) or GARCH. The exact properties of these processes will be specified in Assumptions **A1-A3** below. Similar nonlinear models with a predetermined covariate have been considered for example by Park and Phillips (1999, 2001), Chang et al. (2001), and Chan and Wang (2015), in a parametric setup, and by Wang and Phillips (2009a,b, 2011, 2012), Kasparis et al. (2015) in a nonparametric setup. For recent related results, we refer to Wang (2021), Hu et al. (2021b), Duffy and Kasparis (2021) and the references therein. It should be mentioned that the model of eq. (4) can be rewritten in the following more compact form:

$$y_k = \boldsymbol{\theta}'\mathbf{F}(\mathbf{x}_{k-1}) + e_k, \tag{5}$$

where $\boldsymbol{\theta} := [\mu, \boldsymbol{\beta}']'$, and $\mathbf{F}(\mathbf{x}_{k-1}) := [1, \mathbf{f}(\mathbf{x}_{k-1})']'$. To keep our technical exposition and notation relatively simple, we assume that either all covariates $\mathbf{x}_k$ are stationary or they are all nonstationary.[4] Furthermore, following Park and Phillips (1999, 2001), we assume that under nonstationarity the regression functions $\{f_i\}_{i=1}^{p}$, are *asymptotically homogeneous functions.* This class of functions is specified in detail by the following definition.

**Definition AHF (asymptotically homogeneous function).** *A measurable function $g : \mathbb{R} \to \mathbb{R}$ is an asymptotically homogeneous function (AHF, hereafter), if there is some continuous function $H_g : \mathbb{R} \to \mathbb{R}$ and $\pi_g : (0, \infty) \to (0, \infty)$ such that for all $\lambda > 0$,*

$$g(\lambda x) = \pi_g(\lambda)H_g(x) + R_g(\lambda; x),$$

*where $|R_g(\lambda; x)| \leq a_g(\lambda)P_g(x)$, with $P_g(x) = 1+|x|^{\delta_g}$ for some $\delta_g > 0$, and $a_g(\lambda)/\pi_g(\lambda) \to 0$, as $\lambda \to \infty$. $H_g$ and $\pi_g$ are the limit homogeneous function and asymptotic order of $g$ respectively.*

*Remark* 1. An **AHF** $g(x)$ postulates that for large $\lambda$, $g(\lambda x) \approx \pi_g(\lambda)H_g(x)$. Many popular functions satisfy this property e.g. polynomial functions, logarithmic, step functions, distribution type of functions - see, e.g. Park and Phillips (1999, 2001) for further discussion. We impose continuity on $H_g$ in order to simplify our derivations. The results presented in this work can be extended to locally integrable functions (e.g. functions with integrable poles) at the expense of more complicated exposition - see, e.g. Christopeit (2009), Wang and Phillips (2009a,b), Duffy and Kasparis (2021) and the references therein.

The following assumptions specify in detail the covariates and the regression error in (4).

**A1** (innovations): $\eta_k \in \mathbb{R}^{p+1}$ is a random sequence of the form $\eta_k = [\xi_k', u_k]'$, $\xi_k \in \mathbb{R}^p$, and $\mathcal{F}_k = \sigma(u_k, u_{k-1}, ..., u_1; \xi_j, j \leq k)$ a sequence of sigma-fields. $\{\eta_k, \mathcal{F}_k\}_{k\geq 1}$ is a $(p+1)$-dimensional martingale difference sequence 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): $\mathbf{x}_k = [x_{k,1}, ..., x_{k,p}]'$ is a functional of $\{\xi_k, \xi_{k-1}, ...\}$, and $\sigma_k$ is adapted to $\mathcal{F}_{k-1}$, where $\mathcal{F}_k$ is defined in **A1** so that $\mathbf{F}(\mathbf{x}_{k-1})$ and $\mathbf{F}(\mathbf{x}_{k-1})\sigma_k$ are strictly stationary ergodic sequences with $\mathbf{F}$ measurable and $E\left[\|\mathbf{F}(\mathbf{x}_1)\|^2 + \sigma_2^2\left(1 + \|\mathbf{F}(\mathbf{x}_1)\|^2\right)\right] < \infty$.

**A3** (nonstationary process & invariance principle):

[4] Our basic asymptotic theory in Appendix can readily accommodate models where some regressors are stationary and some are nonstationary. However, to keep our presentation simple, we do not explicitly allow for both regimes at the same time in a regression setup. A generalization towards this direction would require substantially more complicated notation; see, e.g. Chang et al. (2001).

(a) $X_{n,k} := D_n^{-1}\mathbf{x}_k$, with $D_n = diag\{d_{1n}, ..., d_{pn}\}$, $0 < d_{in}^2 = \text{Var}(x_{n,i}) \to \infty$, for each $i = 1, ..., p$, and $\mathbf{x}_k$ is a functional of $\{\xi_k, \xi_{k-1}, ...\}$, possibly depending on $n$, such that on $D_{\mathbb{R}^{3p}}[0,1]$,

$$\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\xi_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\xi_{-k},\ X_{n,[nt]}\right] \Rightarrow [B_{1,t}, B_{2,t}, \mathcal{X}_t], \tag{6}$$

with $B_{1,t}$, $B_{2,t}$ two independent vector Brownian motions, and $\mathcal{X}_t$ is a continuous vector process that depends only on functionals of $\{B_{1,t}\}_{0\le t\le 1}$ and $\{B_{2,t}\}_{0\le t\le 1}$;

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

The martingale assumption for the innovation process $\{\eta_k, \mathcal{F}_k\}_{k\ge 1}$ under **A1** is standard in the literature. The uniform integrability conditions (a) and (b) are weak in comparison to higher moment assumptions used in previous studies -see, e.g. Wang (2014) and Wang and Phillips (2009a,b). In **A1**(c), we impose $E(u_k^2|\mathcal{F}_{k-1}) = 1$ for convenience of notation. In fact, if $\sigma_u^2 := E(u_k^2|\mathcal{F}_{k-1}) \neq 1$, it is routine to see that our results still hold when $\sigma_k$ is replaced by $\sigma_k\,\sigma_u$. Examples of processes that satisfy **A2** include short and long memory linear processes, e.g., $x_k = \sum_{i=0}^{\infty}\phi_i\xi_{k-i}$, $\xi_i \sim iid(0, \sigma_\xi)$, $\sum_{i=0}^{\infty}\phi_i^2 < \infty$. For the purposes of the subsequent analysis, it is worth noting that when $[\mathbf{x}_k, \sigma_k]$ are (strictly) stationary relying on $\xi_k, \xi_{k-1}, ...$, we also have that $\mathbf{f}(\mathbf{x}_{k-1})\sigma_k$ is an ergodic strictly stationary sequence. The NIF array of (3) is an example of a nonstationary process that satisfies **A3**(a). For (scalar) $x_k$ as in (3), under certain additional technical conditions, we have the following weak limit to a *fractional Ornstein-Uhlenbeck* process (see, e.g., Bunchmann and Chan, 2007; Kasparis et al., 2015)

$$X_{n,[nt]} \Rightarrow \int_0^t e^{c(t-r)}dW_d(r),$$

where $W_d(r)$, $d > 1/2$ is a fractional Brownian motion. Note that the limit process depends on dual nuisance parameters i.e. the near-to-unity parameter $c \in \mathbb{R}$ and the memory parameter $d > 1/2$. Finally, we note that the strict stationarity requirement for $\sigma_k$ of **A3**(b) is general enough to allow a for strictly stationary GARCH, ARCH($\infty$) regression error $e_k$ (e.g., Francq and Zakonian, 2010; Section 2.2).

# 3 CTLS Instruments and Estimation

This section presents the CTLS estimator in the context of the regression model (4), together with its asymptotic properties. Let $K$ be some integrable kernel function and $c_n$ a (reciprocal) bandwidth term. For an integer valued sequence $l_n$ and $0 < \tau_1 < ... < \tau_{l_n} < 1$, define the vector of instruments

$$\mathbf{Z}_{kn} := \sum_{j=1}^{l_n} Z_{kn}(\tau_j),\ Z_{kn}(\tau_j) := K\left[c_n\left(k/n - \tau_j\right)\right]\mathbf{f}(\mathbf{x}_{k-1}).$$

Given that the main focus of our analysis is inference for the slope parameters, it is convenient to consider an estimator for $\boldsymbol{\beta}$ only. For this reason, certain intercept demeaning for

$y_k$ is used below. For any $\{a_k\}_{k=1}^n$ ($a_k$ can be a vector), define

$$\overline{a} := \frac{\sum_{k=1}^n a_k K_{kn}}{\sum_{k=1}^n K_{kn}} \quad \text{and} \quad \overline{a}_k := a_k - \overline{a}, \tag{7}$$

where $K_{kn} := \sum_{j=1}^{l_n} K[c_n(k/n - \tau_j)]$. The proposed CTLS estimator for $\boldsymbol{\beta}$ in the model of (4) is

$$\hat{\boldsymbol{\beta}} = \left[\sum_{k=1}^n \mathbf{Z}_{kn}\overline{\mathbf{f}}(\mathbf{x}_{k-1})'\right]^{-1} \sum_{k=1}^n \mathbf{Z}_{kn}\overline{y}_k. \tag{8}$$

We note that, when $l_n = 1$, the CTLS estimator reduces to the usual local level kernel estimator that has been widely used in the TVP literature; for example, see Phillips et al. (2017), Hu et al. (2024) and the references therein. Although it may not be directly obvious, the CTLS estimator given in (8) also entails instrumentation for the intercept. In fact, $K_{kn}$ is implicitly used as an instrument for the intercept. This instrumentation is manifest in the demeaning of $y_k$, i.e.

$$\bar{y}_k = y_k - \frac{\sum_{k=1}^n y_k K_{kn}}{\sum_{k=1}^n K_{kn}}.$$

Notice that if we set $K_{kn} = 1$, the term above yields the OLS demeaning (instrumentation) as a special case. It should be mentioned that instrumentation for the intercept is necessary for CTLS to achieve mixed normality in the nonstationary case. In practice, it is possible to choose different kernel functions and bandwidth terms for each element of $\mathbf{f}(\mathbf{x}_{k-1})$. In order to keep the technical exposition simple, in this section, we only consider the case where the kernel function $K$ and $c_n$ is the same for all elements of $\mathbf{f}(\mathbf{x}_{k-1})$. We refer to Hu et al. (2021a) for more technical details about the properties of CTLS when different kernel functions are employed.

The following two assumptions specify in detail the properties of the kernel function $K$ and the sequences $c_n$, $l_n$, and $\{\tau_j\}_{j=1}^{l_n}$.

**A4** (kernel function and restrictions on $\tau_j, l_n$ 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 $c_n^{-1} l_n + l_n^{-1} \to 0$.

The 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 $\tau_j, l_n$ and $c_n$):

(a) $K(x)$ is a bounded positive and eventually monotonic real function[5] with $0 < \int K < \infty$;

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

(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 $c_n^{-1} l_n \log l_n + l_n^{-1} \to 0$.

Assumptions **A4** and **A4**$^*$ postulate that the bandwidth $c_n$ is divergent and dominated by $n$, and that $l_n$ is divergent and dominated by $c_n$. It should be emphasized that these restrictions are not model specific, that is, they are independent of the properties of $\mathbf{x}_k$. It is explained in more detail below that in practice we can also assume that $l_n$ is fixed, but with the additional restriction $l \geq p$, under nonstationarity.

Our first result provides the limit distribution of the CTLS estimator when **A2** is satisfied, i.e., for stationary $\mathbf{x}_k$. The basic limit theory for stationary and nonstationary functionals is provided by Theorem 5 and Theorem 6, respectively, in the Appendix.

**Theorem 1.** *Suppose that:*

***(a)*** *$y_k$ is generated by (4);*

***(b)*** ***A1**, **A2** and either **A4** or **A4**$^*$ hold;*

***(c)*** *$\Phi_1 := E\left\{[\mathbf{f}(\mathbf{x}_1) - E\mathbf{f}(\mathbf{x}_1)][\mathbf{f}(\mathbf{x}_1) - E\mathbf{f}(\mathbf{x}_1)]'\right\}$ is non-singular.*

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

$$\sqrt{\frac{nl_n}{c_n}}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \to_d \mathbf{N}\left(\mathbf{0},\ \frac{\int K^2}{\left(\int K\right)^2} \cdot \Phi_1^{-1}\Phi_0\Phi_1^{-1}\right),$$

*with $\Phi_0 := E\left\{\sigma_2^2 [\mathbf{f}(\mathbf{x}_1) - E\mathbf{f}(\mathbf{x}_1)][\mathbf{f}(\mathbf{x}_1) - E\mathbf{f}(\mathbf{x}_1)]'\right\}$.*

We note that the multiplicative structure of the limit variance matrix in Theorem 1 is due to conditional heteroscedasticity assumption for the regression error $e_k$. Under conditional homoscedasticity, e.g. $\sigma_k = \sigma$ for all $k$, we have $\Phi_0 = \sigma^2\Phi_1$ which gives the limit variance matrix $\left[\int K^2/\left(\int K\right)^2\right] \cdot \sigma^2\Phi_1^{-1}$.

Theorem 2 below gives the limit distribution of the CTLS estimator for nonstationary covariates, i.e., $\mathbf{x}_k$ satisfies **A3** instead of **A2**.

**Theorem 2.** *Suppose that:*

***(a)*** *$y_k$ is generated by (4).*

***(b)*** ***A1**, **A3** and either **A4** or **A4**$^*$ hold.*

***(c)*** *For each $i = 1, ..., p$, $f_i$ is an AHF with limit homogeneous function $H_{f_i}$ and asymptotic order $\pi_{f_i}$. Furthermore, $\mathcal{D}_n := diag\left\{\pi_{f_1}(d_{1n}), ..., \pi_{f_p}(d_{pn})\right\}$.*

***(d)*** *$\Phi_2 := \int_0^1 \tilde{H}_{\mathbf{f}}(\mathcal{X}_t)\tilde{H}_{\mathbf{f}}(\mathcal{X}_t)'dt$ is non-singular a.s., where $H_{\mathbf{f}}(\mathcal{X}_t)' := \left[H_{f_1}(\mathcal{X}_{t,1}), ..., H_{f_p}(\mathcal{X}_{t,p})\right]$, with $\mathcal{X}_t = [\mathcal{X}_{t,1}, ..., \mathcal{X}_{t,p}]'$ and $\tilde{H}_{\mathbf{f}}(\mathcal{X}_t) = H_{\mathbf{f}}(\mathcal{X}_t) - \int_0^1 H_{\mathbf{f}}(\mathcal{X}_s)ds$.*

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

$$\sqrt{\frac{nl_n}{c_n}}\mathcal{D}_n\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \to_d \mathbf{MN}\left(\mathbf{0},\ \frac{\int K^2}{\left(\int K\right)^2} \cdot E(\sigma_1^2)\Phi_2^{-1}\right).$$

The limit distribution in Theorem 2 is mixed normal and, therefore, conventional inference applies. Furthermore, unlike the stationary case, the volatility term $E\left(\sigma_1^2\right)$ separates from the variance matrix, ensuring easy implementation in applications.

*Remark* 2. Both under stationarity and nonstationarity, the CTLS estimator attains the OLS convergence rate ($\sqrt{n}$ and $\sqrt{n}\mathcal{D}_n$ respectively) times $\sqrt{l_n/c_n} \to 0$. The term $\sqrt{l_n/c_n}$ can be chosen to vanish at an arbitrarily slow rate e.g. slowly varying. In the nonstationary case, the convergence rate depends on the asymptotic order of the regression functions, i.e. $\{\pi_{f_i}\}_{i=1}^p$, as well as the memory characteristics of the covariates, i.e. $\{d_{in}\}_{i=1}^p$ - see also Park and Phillips (1999, 2001) and Chang et al. (2001) for comparable results. It should be mentioned that the requirement $c_n n^{-1} \to 0$ ensures that there is sufficient signal in the instruments for consistent estimation, while the restrictions ${c_n}^{-1}, l_n c_n^{-1} \to 0$ ensure that there is sufficient signal reduction in the instruments to facilitate mixed normality under nonstationarity. If ${c_n}^{-1} \to \infty$, it can be shown that CTLS and OLS are in general asymptotically equivalent[6]. In the stationary case, a fixed value for $c_n$ also yields an CTLS estimator that is asymptotically equivalent to OLS - see e.g. Hu et al. (2024, Remark A1).

*Remark* 3. If a single chronological point, say $\tau$, is utilized (i.e. $l_n = 1$), then

$$\Phi_2 = H_{\mathbf{f}}(\mathcal{X}_\tau)H_{\mathbf{f}}(\mathcal{X}_\tau)',$$

which in necessarily singular. Phillips et al. (2017) demonstrate that in this case the limit distributions of kernel regression estimators for TVPs are determined by stochastic integrals, and therefore limit theory is not conventional. In fact, the limit distribution in this case is comparable to that of OLS. Phillips et al. (2017) propose a kernel variant of FMLS to obtain pivotal tests. However, it is known that this method is not robust to local deviations from unity.

*Remark* 4. The limit result of Theorem 2 applies with a non degenerate variance matrix for a fixed number of chronological point $\{\tau_j\}_{j=1}^l$ provided $l \geq p$ i.e. the number of regression parameters does not exceed $l$. In this case, the variance matrix is of the form

$$\Phi_2 = \frac{1}{l}\sum_{j=1}^{l} H_{\mathbf{f}}(\mathcal{X}_{\tau_j})H_{\mathbf{f}}(\mathcal{X}_{\tau_j})'.$$

It can be shown that if there are *a.s.* $p$-many linearly independent vectors $H_{\mathbf{f}}(\mathcal{X}_{\tau_j})$, then the term above is non-singular *a.s.*

*Remark* 5. Kostakis et al. (2015) show that the IVX estimator attains a $\sqrt{n}$-convergent rate for stationary short memory covariates. However, CTLS attains a sub-OLS convergence rate in this case. This reduction in the convergence rate has an impact on the asymptotic power of CTLS based tests. A solution to this problem is to fine-tune the bandwidth $c_n$ using some memory estimator for the covariates. For example, consider the alternative bandwidth

$$\hat{c}_n = \begin{cases} c_n, \text{ if } \hat{d} \geq 0.5 \\ 1, \text{ if } \hat{d} < 0.5 \end{cases}, \tag{9}$$

[6]i.e. under stationarity $\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = \sqrt{n}\left(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}\right) + o_p(1)$, and under nonstationarity $\sqrt{n}\mathcal{D}_n\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = \sqrt{n}\mathcal{D}_n\left(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}\right) + o_P(1)$.

where $\hat{d}$ is some memory estimator for a covariate, e.g., local Whittle. It follows from Remark 2 above that if $\hat{c}_n = 1$, $\hat{\beta}$ is asymptotically equivalent to OLS. In fact, if $\hat{c}$ is employed, CTLS trimming is "switched off" when the process is in the stationary region. As a consequence, there is an improvement in the convergence rate of $\hat{\beta}$. Some preliminary results show that this bandwidth selection method delivers valid inference even if the covariate is a NIF array i.e. equation (3).[7] Simulations provided in Section 6 show that the bandwidth of (9) yields substantial power improvements when the covariate is in the stationary region. In particular, for values of the memory parameter $d \leq 0.4$, the CTLS predictive tests are indistinguishable from the OLS tests. A similar bandwidth selection method has been considered by Kasparis et al. (2015) for improving the performance of kernel based inference in the presence of fractional covariates. Furthermore, this method is comparable to the "sliding" statistics considered by Kostakis et al. (2015), Elliot et al. (2015) and Magdalinos and Petrova (2022). Kostakis et al. (2015) provide a finite sample improvement for the IVX method by considering a sliding studentization for the IVX t-statistic. This method employs plug-in estimators for long-run variance matrices that yield a different studentization in situations of weak and strong endogeneity. Elliot et al. (2015) consider an inferential method for predictive regressions with an NI predictor. The method requires the near-to-unit parameter to be on a bounded prespecified interval, therefore ruling out stationary processes. A preliminary autoregressive estimator is utilized to switch to a conventional test statistic when the predictor is in the stationary region. Magdalinos and Petrova (2022) propose a similar switching mechanism to switch test statistics between stationary, NI, and explosive regions in predictive regressions.

# 4 CTLS Inference

Next, we consider statistical inference for regression parameters based on CTLS estimation. It follows from the results in Section 3 that the CTLS estimator delivers conventional inference (e.g. $\mathbf{N}(0,1)$ and $\chi^2$). In order to keep our exposition relatively simple, in the remainder of the main paper, we restrict our analysis to t-tests for single restriction in the context of univariate regressions. Inference in the context of multicovariate models with possibly multiple restrictions is provided in Section S2 of the Online Supplement.

Setting the dimension of the slope coefficient vector $p = 1$, and suppressing the second index on the r.h.s., we can rewrite (4) as

$$y_k = \mu + \beta f(x_{k-1}) + e_k, \;\; k = 1, ..., n, \tag{10}$$

where $x_k$ and $e_k$ are defined as in **A1-A3**. In this case, the slope parameter estimator of $\beta$ in (8) reduces to

$$\hat{\beta} \;\; = \;\; \frac{\sum_{k=1}^{n} Z_{kn} \overline{y}_k}{\sum_{k=1}^{n} Z_{kn} \overline{f}_k}, \tag{11}$$

[7] The NIF array of (3) has the same memory characteristics as those of an ARFIMA with $d > 1/2$, in the sense that both processes are $O_p(n^{d-1/2})$, $d = 1 + \delta$ ($\delta$ is defined in (3)). Preliminary results show that the local Whittle and log periodogram estimators both converge to $d$ (for $d \leq 1$), even if there is some near-to-unity parameter, therefore correctly capturing the order of magnitude of the process. We expect that the exact local Whittle (Shimotsu and Phillips, 2005) is also valid, even if $d > 1$.

where $Z_{kn} = K_{kn}f(x_{k-1}), K_{kn} = \sum_{j=1}^{l_n} K\left[c_n\left(k/n - \tau_j\right)\right], \overline{y}_k = y_k - \frac{\sum_{k=1}^n y_k K_{kn}}{\sum_{k=1}^n K_{kn}}$ and

$$\overline{f}_k = f_k - \frac{\sum_{k=1}^n f_k K_{kn}}{\sum_{k=1}^n K_{kn}}, \quad \text{with } f_k = f(x_{k-1}).$$

To test the hypothesis $H_0 : \beta = \beta_0$ (for some $\beta_0 \in \mathbb{R}$), we consider the CTLS based $t$-statistic defined by

$$\hat{T} = \mathcal{H}_n \frac{\hat{\beta} - \beta_0}{\sqrt{\mathcal{A}_n \hat{\mathcal{V}}_n \mathcal{A}_n'}}, \tag{12}$$

where

$$\mathcal{A}_n := \left[-\bar{f},\ 1\right], \quad \mathcal{H}_n := \sum_{k=1}^n Z_{kn}\overline{f}_k, \quad \hat{\mathcal{V}}_n := \sum_{k=1}^n K_{kn}^2 \check{e}_k^2 \begin{bmatrix} 1 & f_k \\ f_k & f_k^2 \end{bmatrix},$$

and $\check{e}_k$ are the OLS residuals $\check{e}_k = y_k - \hat{\mu}_{LS} - \hat{\beta}_{LS} f_k$. Note that for the estimation of the regression error variance, we use residuals based on the OLS estimator, which is more efficient than the CTLS estimator (see also Kostakis et al. (2015) for a similar approach). The normalizing matrix $\hat{\mathcal{V}}_n$ allows for conditional heteroscedasticity. However, if $\sigma_k^2 = \sigma^2$ for all $k$, then the following estimator can be used instead.

$$\check{\mathcal{V}}_n := \check{\sigma}^2 \sum_{k=1}^n K_{kn}^2 \begin{bmatrix} 1 & f_k \\ f_k & f_k^2 \end{bmatrix}, \ \check{\sigma}^2 := n^{-1} \sum_{k=1}^n \check{e}_k^2. \tag{13}$$

Under nonstationarity, conditional heteroscedasticity in the regression error does not affect the limit variance of the CTLS estimator in a material way as seen in Theorem 2. In particular, the volatility term $E\sigma_1^2$ is scaled out. As a result, conventional estimators for the limit variance (i.e. $\check{\mathcal{V}}$) can be employed for the construction of t-statistics (see also Remark 8 below). This fact is comparable with the recent findings of Magdalinos (2022) who demonstrates that conditional heteroscedasticity has a material effect in the limit distribution of the IVX estimator only under stationarity.

The subsequent result gives the limit distribution of $\hat{T}$ under the null hypothesis when the regressor is either stationary satisfying **A2** or nonstationary satisfying **A3**.

**Theorem 3.** *Suppose that the conditions of Theorem 1 or Theorem 2 hold with $p = 1$. In addition, $\sup_{k\geq 1} Eu_k^4 < \infty$. Then, under $H_0 : \beta = \beta_0$, we have*

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

*Remark* 6. 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 the t-statistic is determined by the convergence rate of the CTLS estimator. In particular, for stationary $x_k$ straightforward arguments show that $\hat{T} = O_P(\sqrt{nl_n/c_n})$. On the other hand in the nonstationary case we have $\hat{T} = O_P(\pi_f(d_n)\sqrt{nl_n/c_n})$, where $d_n$ is the normalizing sequence of **A3(a)** for the case $p = 1$, e.g. $d_n = \sqrt{n}$ for $x_k$ NI and $d_n = n^{d-1/2}$, $x_k$ for $I(d)$, $1/2 < d < 3/2$. Therefore, a faster divergence rate is attained for more persistent processes. This fact is also corroborated by our simulation results (see Figure S1 in the Online Supplement). 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 -see also Park and Phillips (1999, 2001) for comparable results.

*Remark* 7. It is well known that the choice of bandwidth in semi-parametric (and nonparametric) methods involves a trade-off between size and power. This is true for CTLS as well. In general, choices of $l_n$, $c_n$ that yield better asymptotic power are associated with inferior size and vice versa. An optimal choice of these bandwidth terms would require second order limit theory comparable to that of Sun et al. (2008), which is very challenging from a technical point of view under our theoretical framework. In the current work, we recommend values for the bandwidth terms that appear to give a good size/power trade-off based on a simulation study; see Section 6 below.

*Remark* 8. The fourth moment requirement for $u_k$ in Theorem 3 can be dispensed with when the regression errors are conditionally homoscedastic, i.e. in situations where $\check{\mathcal{V}}_n$, of (13), can be utilized in the place of $\hat{\mathcal{V}}_n$.

# 5 Consequences and Mitigation of Misbalancing in Predictability Tests

Regression *misbalancing* (cf. Phillips, 2015) is an important issue that has received relatively little attention in the predictability literature. In this section, we provide some theory about the consequences of misbalancing to the predictability hypothesis. These results are supported by the simulation study in the next section. A well known stylistic fact in the empirical studies of stock return predictability is that stock returns and various predictors exhibit vastly different memory characteristics. Stock returns are typically consistent with $I(0)$ processes, whilst various financial and macroeconomic predictors (e.g. dividend yield, inflation) appear to be long memory and in most cases nonstationary (see, e.g. Kostakis et al. 2015). Phillips (2015) referred to this memory mismatch as *misbalancing*. Misbalancing makes linear specifications that dominate the empirical literature less plausible. We show that for many specifications of interest, misbalancing due to incorrect functional form still leads to consistent tests as long as the true regression function is diverging. However, there is a reduction in the asymptotic power rate. Overall, the power loss is more substantial if the model is badly misspecified, the persistence of the predictor is more intense, and the noise (regression error) less substantial. In addition, if the true regression function is of vanishing order, statistical tests can be bounded under the alternative hypothesis.

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 $\mu + f(x_{k-1})$ with $f$ being the fractional difference operator $(I - L)^d$ and $d$ the memory parameter of some volatility predictor, $x_k$ say. Note that in this case $f(x_{k-1})$ is $I(0)$ -see also Andersen and Varneskov (2021). Similarly to spectral LS methods, e.g., Robinson and Hualde (2003) and Christensen and Nielsen (2006), this approach requires plug-in estimates for the memory parameters.

Marmer (2008), Kasparis (2010), Kasparis et al. (2015) and Phillips (2015) suggest that nonlinear regression functions can potentially address 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 a stationary one. For example, for some measurable function $f$ and $x_k$ stationary we have $\sum_{k=1}^{n} |f(x_k)| = O_P(n)$, in general. On the other hand for $x_k \sim I(1)$ the following orders apply (e.g. see Park and Phillips, 2001)

$$\sum_{k=1}^{n} |f(x_k)| = \begin{cases} O_P(n^{1+p/2}), & f \text{ polynomial of order } p > -1 \\ O_P(n \ln(n)), & f \text{ logarithmic} \\ O_P(n), & f \text{ bounded} \\ O_P(\sqrt{n}), & f \text{ integrable} \end{cases}$$

It is easy to see from the orders shown above that certain nonlinear transformations of $I(1)$ processes exhibit very weak signals that can be equal (bounded functions) or smaller (integrable and reciprocal functions) than those of a stationary process. Figure 1 provides a graphical illustration of the effects of certain nonlinear transformations on the paths of an $I(1)$, process relative 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 process may resemble those of martingale difference processes.

As emphasized by Phillips (2015), -see also Kasparis (2011)- misbalancing may result to asymptotically vanishing estimators.[8] Nevertheless, it is not directly obvious how this affects the predictability tests. Interestingly, for many specifications of interest (e.g. AHF of diverging homogeneous order), t-statistics diverge under the alternative hypothesis even if misbalancing has been committed, albeit at slower rates. This is due to the fact that the test statistic studentization partially offsets misbalancing effects. However, if the fitted model is badly misspecified, the t-tests can be bounded under the alternative hypothesis.

Next, we consider the effects of misbalancing in CTLS t-tests due to an incorrect functional form. Some comparable analysis is provided by Kasparis et al. (2015) who investigate the asymptotic power rates of FMLS t-tests and the Janson and Moirera (2006) test under functional form misspecification. In this work, we provide some deeper analysis of the effects of misbalancing due to functional form misspecification with a focus on CTLS fits. In particular, we assume that the true model is given by (10) with $f$ AHF of increasing asymptotic order $\pi_f(\lambda)$ (i.e. $\pi_f(\lambda)^{-1} \to 0$) and the fitted model is

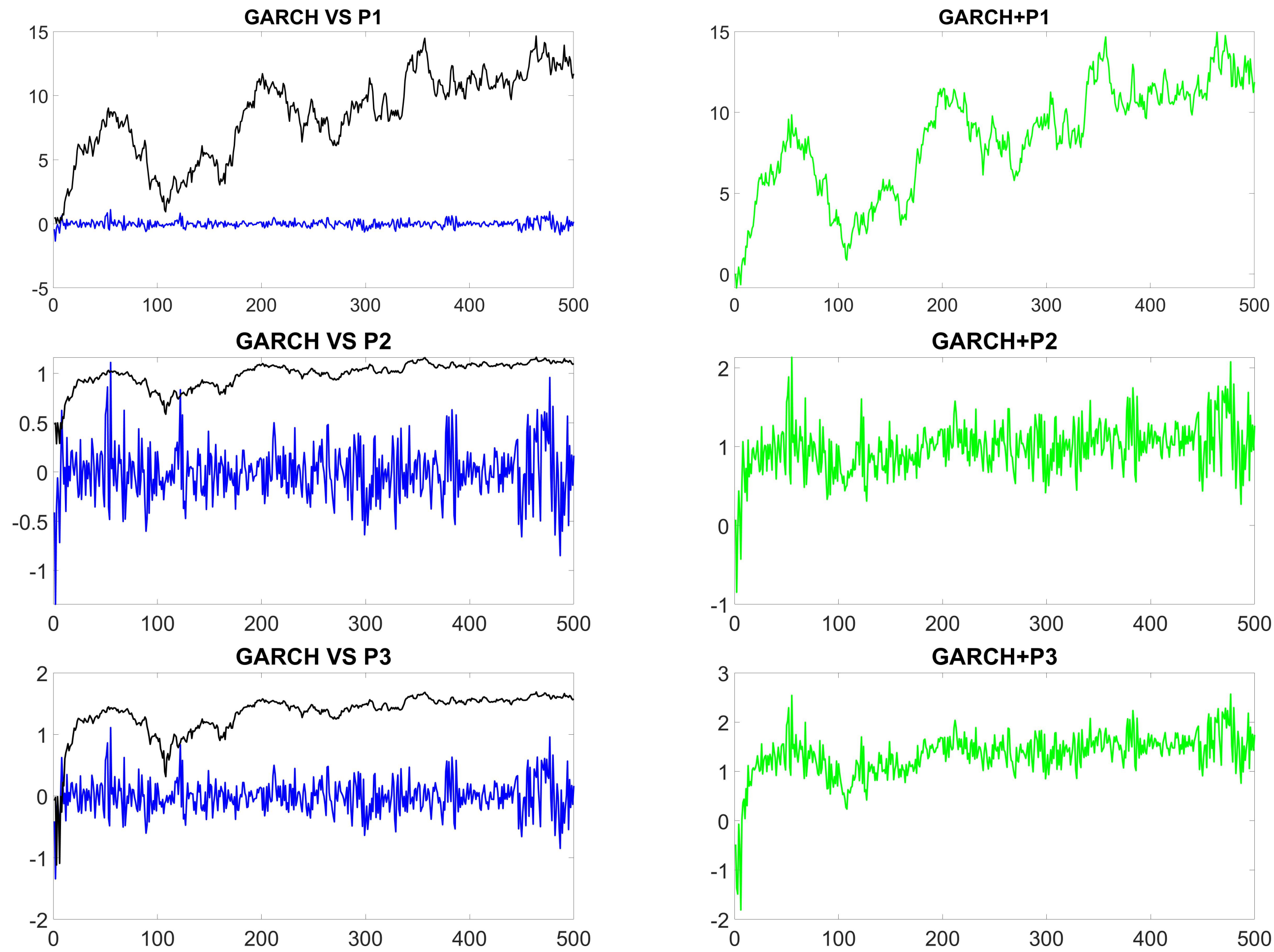
$$y_k = \hat{\mu} + \hat{\beta} f_M(x_{k-1}) + \hat{e}_k, \tag{14}$$

with $\hat{\mu}$, $\hat{\beta}$ being CTLS estimates, and $f_M$ AHF of asymptotic order $\pi_{f_M}(\lambda)$ such that $\pi_f(\lambda)$ $\pi_{f_M}(\lambda)^{-1} \to 0$ i.e. a) the fitted model is misbalanced because the fitted regression function dominates the true one; b) $f_M$ is of diverging asymptotic order. To keep the paper within manageable length, we do not provide theoretical results for the cases where $f$ is integrable or AHF with $\pi_f(\lambda) \to 0$. However, this type of misspecification is considered in the simulation study of the following section.

Before presenting the next theoretical result, we introduce some notation. Let $\hat{T}$ and $\hat{T}_M$ be the t-statistics of (12) based on the true and misbalanced models, respectively, for the null hypothesis $H_0 : \beta = 0$. Further, recall that for any locally integrable function $q$ we

[8] Consider for instance linear regression of $y_k$ on $x_k$ with $y_k \sim I(d_y)$, $d_y < 1/2$ , and $x_k \sim I(d_x)$ with $d_x > 1/2$.

Figure 1: 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. $0.01, 0.45, 0.45$.

define

$$\tilde{q}(\mathcal{X}_t) := q(\mathcal{X}_t) - \int_0^1 q(\mathcal{X}_s)ds.$$

Furthermore, by $d_n$ we denote the normalizing sequence of Assumption **A3(a)** for the case $p = 1$. The following result gives the divergence rate of $\hat{T}_M$ as well as the relative orders of $\hat{T}$ and $\hat{T}_M$.

**Theorem 4.** *Suppose that*

*(a)* $y_k$ *is generated by (4) with* $p = 1$, *and* $\sigma_k^2 = \sigma^2 \in (0, \infty)$ *for all* $k$;

*(b) the conditions of Theorem 2 hold;*

*(c)* $f_M$ *is AHF with limit homogeneous function* $H_{f_M}$ *and asymptotic order* $\pi_{f_M}$;

*(d) the matrix* $\int_0^1 \begin{bmatrix} 1 & H_{f_M}(\mathcal{X}_t) \\ H_{f_M}(\mathcal{X}_t) & H^2_{f_M}(\mathcal{X}_t) \end{bmatrix} dt$ *is nonsingular a.s.;*

*(e)* $\pi_f(\lambda)^{-1} + \pi_f(\lambda)\pi_{f_M}(\lambda)^{-1} \to 0.$

*Under* $H_1 : \beta \neq 0$*, we have*

$$\sqrt{\frac{c_n}{nl_n}}\hat{T}_M \to_d \frac{\beta \int K \cdot \int_0^1 \tilde{H}_{f_M}(\mathcal{X}_t)H_f(\mathcal{X}_t)dt}{\sqrt{\sigma_*^2 \int K^2 \int_0^1 \tilde{H}_{f_M}(\mathcal{X}_t)H_{f_M}(\mathcal{X}_t)dt}} \tag{15}$$

*with* $\sigma_*^2 := \int_0^1 \left[H(\mathcal{X}_t) - \mu_* - \beta_* H_{f_M}(\mathcal{X}_t)\right]^2 dt$ *and* $\mu_*$*,* $\beta_*$ *the pseudo-true value limits*[9] *of the OLS estimator. Furthermore, under* $H_1 : \beta \neq 0$

$$\pi_f(d_n)^{-1}\frac{\hat{T}}{\hat{T}_M} \to_d \frac{\sqrt{\sigma_*^2 \int_0^1 \tilde{H}_{f_M}(\mathcal{X}_t)H_{f_M}(\mathcal{X}_t)dt \int_0^1 \tilde{H}_f(\mathcal{X}_t)H_f(\mathcal{X}_t)dt}}{\sqrt{\sigma^2}\int_0^1 \tilde{H}_{f_M}(\mathcal{X}_t)H_f(\mathcal{X}_t)dt}. \tag{16}$$

It can be easily seen from Theorem 4 that if the true regression function is AHF of lower asymptotic order than the fitted one, there is a reduction in the asymptotic power rate by a factor of $\pi_f(d_n)$. Notice that $\pi_f(.)$ is the asymptotic order of *true regression function.* The ratio $\left|\hat{T}/\hat{T}_{f_M}\right|$ captures the relative asymptotic power of the test statistics based on balanced and misbalanced specifications. At first glance, the second part of Theorem 4 suggests that misbalancing has more adverse power effects when $\pi_f(.)$ is of higher order (i.e. closer to the fitted model), which is counter intuitive. A closer reading of Theorem 4 shows that $\pi_f(.)$ is not the only term that affects the relative asymptotic power. It follows from the second part of the theorem above that we have the following approximate behavior.

$$\begin{aligned} \left|\hat{T}/\hat{T}_M\right| &\approx \frac{\sqrt{\int_0^1 \tilde{H}_{f_M}(\mathcal{X}_t)H_{f_M}(\mathcal{X}_t)dt \int_0^1 \tilde{H}_f(\mathcal{X}_t)H_f(\mathcal{X}_t)dt}}{\left|\int_0^1 \tilde{H}_{f_M}(\mathcal{X}_t)H_f(\mathcal{X}_t)dt\right|} \cdot \sqrt{\frac{\pi_f^2(d_n)\sigma_*^2 + \sigma^2}{\sigma^2}} \\ &=: R_1 \cdot R_{2n}\left(\sigma^2, \sigma_*^2\right). \end{aligned} \tag{17}$$

The relative performance of the two statistics is therefore determined by

i) $\pi_f(.)$: the asymptotic order of the true regression function;

ii) $d_n$: the persistence of the predictor;

iii) $R_1$: the asymptotic distance between the true and fitted regression functions;[10]

iv) $\sigma_*^2$: the asymptotic $L^2$-distance between the true and fitted regression models;

v) $\sigma^2$: noise.[11]

[9] A full characterisation of $\mu_*$, $\beta_*$ is provided in the Online Supplement.

[10] Note that by the Cauchy-Schwarz inequality $R_1 \geq 1$. If $f \approx f_M$, then $R_1 \approx 1$. On the other hand, if there is a large discrepancy between $f$ and $f_M$, $R_1$ is significantly greater than unity.

[11] Notice that $R_{2n}\left(\sigma^2, \sigma_*^2\right)$ is decreasing in $\sigma^2$. Therefore, other things being equal, misbalancing has less severe power effects when noise is substantial.

Misbalancing has more severe effects when the terms $\pi_f(d_n)$, $R_1$, $\sigma_*^2$ assume larger values. Notice, however, that the latter two terms tend to be larger when there is substantial discrepancy between $f$ and $f_M$, while the former term is more substantial when the opposite is true. Simulation results provided in the next section show that in finite samples misbalancing has more severe effects when there is a larger discrepancy between $f$ and $f_M$ indicating that $R_1$ are $\sigma_*^2$ more important than $\pi_f(d_n)$.

**Mitigating Misbalancing Effects**. The discussion above shows that even if misbalancing is committed, predictability tests may have nontrivial asymptotic power. In many cases t-statistics are divergent under $H_1$, albeit at slower rates, even if the fitted model is misbalanced due to incorrect functional form.[12] In view of this, we consider a strategy that potentially mitigates misbalancing issues, when it comes to the predictability hypothesis. More specifically, we consider rolling t-statistics incorporating regression fits that span various rates. For example, consider the flexible functional forms given by the following set of regression functions

$$f(x,\theta) = \frac{x}{1+|x|^{\theta}},\ \theta \in \Theta, \tag{18}$$

with $\Theta$ being a discrete subset of positive real numbers. For each $\theta \in \Theta$, consider the CTLS t-statistic $\hat{T}(\theta)$ based on the empirical model

$$y_k = \hat{\mu} + \hat{\beta} f(x_{k-1},\theta) + \hat{e}_k \tag{19}$$

for the hypothesis $H_0 : \beta = 0$ -cf. eq. (12). For example, for $\Theta \subset [0,3]$ we get a wide range of growth rates for $f$ including linear ($\theta = 0$) bounded ($\theta = 1$), asymptotically reciprocal ($\theta > 1$), and integrable ($\theta > 2$). Under the null hypothesis of no predictability we have $\hat{T}(\theta) \to_d \mathbf{N}(0,1)$ for each $\theta \in \Theta$ by virtue of Theorem 3. Under $H_1$ the test statistics tend to be divergent for a wide range of values of $\theta$ (even if the fitted model is misspecified in terms of functional form) with maximal asymptotic power attained when $f(x,\theta)$ is closer to the true regression function.

An alternative approach for addressing misbalancing issues is testing based on kernel regression estimators; see, e.g. Kasparis et al. (2015). Nonparametric methods provide more generality at the expense of slower convergence rates. Further, an alternative parametric approach could be obtained by considering test functionals of the form $\max_{\theta\in\Theta}\hat{T}(\theta)^2$, $\sum_{\theta\in\Theta}\hat{T}(\theta)^2$ (see, e.g. Hansen, 1996). Developing limit theory for sup and integral functionals under our assumptions is challenging from a technical point of view. We leave exploration in this direction for future work. In Section 7 below, we employ rolling test statistics based on the flexible functional form of (18) to test the predictability hypothesis for the SP500 returns.

[12]Recall that if the fitted model is linear and the true regression function AHF of increasing order, t-statistics have asymptotic power one. However, the tests may not have power if the true regression function is AHF of decreasing order or integrable -see Kasparis et al. (2015) and the simulation study in this work.

## 6 Simulations

This section explores the finite sample properties of CTLS inferential methods with the aid of a simulation study. We consider the no predictability hypothesis

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

The DGP is of the form

$$y_k = \beta x_{k-1} + e_k. \tag{20}$$

Without loss of generality we set $\mu = 0$. Note that estimators are numerically invariant to the value of the intercept. 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

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

$\varrho \in (-1, 1)$. The predictor is either an NI array of the form

$$x_k = \left(1 + \frac{c}{n}\right) x_{k-1} + \xi_k, \tag{21}$$

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

$$(I - L)^d x_k = \xi_k I \{k \geq 1\}. \tag{22}$$

The regression error is

$$e_k = \sigma_k u_k,$$

with either

$$\sigma_k^2 = 1,$$

or

$$\sigma_k^2 = 0.01 + 0.45\sigma_{k-1}^2 + 0.45e_{k-1}^2, \ \sigma_0^2 = 0.01, \tag{23}$$

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

We report empirical size and power results for $\hat{T}$ given in (12), and $\check{T}$ which is a CTLS t-stastistic that is utilizing the variance estimator of (13). 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 affect finite sample performance. As explained in Section 3 and 4, there is 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 that involved 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. Let $\varphi_\varsigma(x)$ be the density of an $\mathbf{N}(0,\varsigma)$ variate. We consider the following bandwidth terms:

- $c_n = n^{0.95}$, $l_n = c_n^{0.7}$, with $\{\tau_j\}_{j=1}^{l_n}$ being equispaced points on $(0,1)$.

Further, we use the kernel function $K(x) = \varphi_{0.1}(x)^{1/2}$ for CTLS instrumentation for the predictor, and $K(x)^* = \varphi_1(x)^{1/2}$ for intercept instrumentation[13]. Notice that the latter kernel has a larger variance and, as a result, it entails more limited trimming compared to $\varphi_{0.1}(x)^{1/2}$. Preliminary simulations show that using different kernel functions provides a slightly better size-power trade-off. A possible explanation for this is that $x_k$ is the main source of endogeneity, and consequently more substantial signal reduction is required for the estimation/instrumentation of the slope parameter than for the intercept.

## 6.1 Finite Sample Performance of CTLS Tests

Table 1 reports the empirical size of CTLS tests for NI predictors and $\sigma_k^2 = 1$. For comparison we also consider an IVX test, which incorporates the finite sample improvement of Kostakis et al. (2015), and an OLS based t-test. Contrary to the CTLS and IVX estimators, the OLS estimator 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), Ang and Bekaert (2007), Bandi and Perron (2008), Chen and Deo (2009), Bandi et al. (2019). In addition, OLS provides a natural benchmark for assessing the benefits in empirical size when mixed normality is induced via attenuated signal instrumentation.

It can be seen from Table 1 that for NI predictors, both CTLS test statistics result in good size control in general 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 the endogeneity is strong. However, the size improves as the sample size increases. Additional simulations for the size performance of CTLS under GARCH regression errors and fractional predictors are provided in the Online Supplement (Table S1 and Table S2 respectively). These results indicate that the empirical size of $\hat{T}$ under GARCH errors is comparable to that reported in Table 1 under conditional homoscedasticity.[14] Further, CTLS tests exhibit good size control under fractional predictors, even in situations of strong endogeneity and memory parameters in excess of unity.

Next, 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 when the predictor is NI. Figure 2 reports the rejection probabilities for $\hat{T}$ and the IVX t-statistic of Kostakis et al. (2015) against various values for the slope parameter ($\beta$) under strong endogeneity (i.e. $\varrho = -0.95$), $c = 0, -10, -50$ and two different sample sizes. Regression errors are conditionally homoscedastic. All tests are more powerful when persistence is stronger and the sample size larger, as expected, with IVX achieving better

[13]As mentioned in Section 3, different kernel functions can be employed for each model parameter. The utilisation of $K^*$ implies that the intercept demeaning of the dependent variable is of the form $\sum_{k=1}^n y_k K_{kn}^* / \sum_{k=1}^n K_{kn}^*$. For more technical details, see Hu et al. (2021a).

[14]Simulation results not reported here suggest that $\check{T}$ and IVX exhibit some moderate oversizing under GARCH regression errors, and large deviations from unity.

performance. Additional simulation results in the Online Supplement (see Figure S1) highlight the power of $\hat{T}$ for the fractional case. Again, our limit theory is corroborated since superior performance is attained in situations where persistence is stronger and sample sizes are larger.

As mentioned in Section 3, the CTLS estimator attains a slower convergence rate than that of IVX under stationarity. A data-driven bandwidth, such as that of (9), can potentially alleviate this problem. Recall that the particular bandwidth scheme uses a preliminary memory estimator that detects if the predictor is in the stationary region. We consider the finite sample power performance of the CTLS statistic with $c_n$ replaced by the data driven bandwidth $\hat{c}_n$ of (9). The exact local Whittle (ELW) estimator of Shimotsu and Phillips (2005) is utilized for memory estimation. $\hat{T}_{SB}$ is the CTLS t-statistic that employs the stochastic bandwidth term $\hat{c}_n$ (i.e. $\hat{T}_{SB}$ is the same as $\hat{T}$, but with $c_n$ replaced by $\hat{c}_n$). To evaluate the performance of $\hat{T}_{SB}$, we consider fractional predictors with $d = 0.7, 0.6, 0.5, 0.4$. Figure 3 provides plots of $\hat{T}_{SB}$, $\hat{T}$ and the OLS based t-statistic for the predictability hypothesis against various values of $\beta$. It can be seen that for values of the memory parameter $d \geq 0.6$, $\hat{T}_{SB}$ and $\hat{T}$ have almost identical performance, while for smaller values of $d$, $\hat{T}_{SB}$ exhibits superior power. In particular, for $d \leq 0.4$, $\hat{T}_{SB}$ is practically indistinguishable from the OLS t-statistic. Some preliminary theoretical results show that semi-parametric memory estimators (e.g. local Whittle, exact local Whittle, log-periodogram) can also distinguish a mildly integrated process from an NI process, if the bandwidth of the memory estimator is chosen appropriately. Therefore, it seems possible to devise a bandwidth selection method comparable to that of (9) that can also provide asymptotic power improvements when the predictor is a MI process. We leave further exploration in this area for future work.

In general, CTLS tests appear to have reasonably good size performance. IVX tests appear to be more powerful; however, 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 by long memory innovations. The authors of the aforementioned work are working on a formal investigation of IVX methods in the presence of fractional processes.

Table 1: Empirical Size of CTLS Tests (nominal size 5%; NI regressor, cond. homoscedastic regression errors)

| $\varrho$ | | -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 |
| $\varrho$ | | -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 |
| $\varrho$ | | -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 |
| $\varrho$ | | -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 |
| $\varrho$ | | -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

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

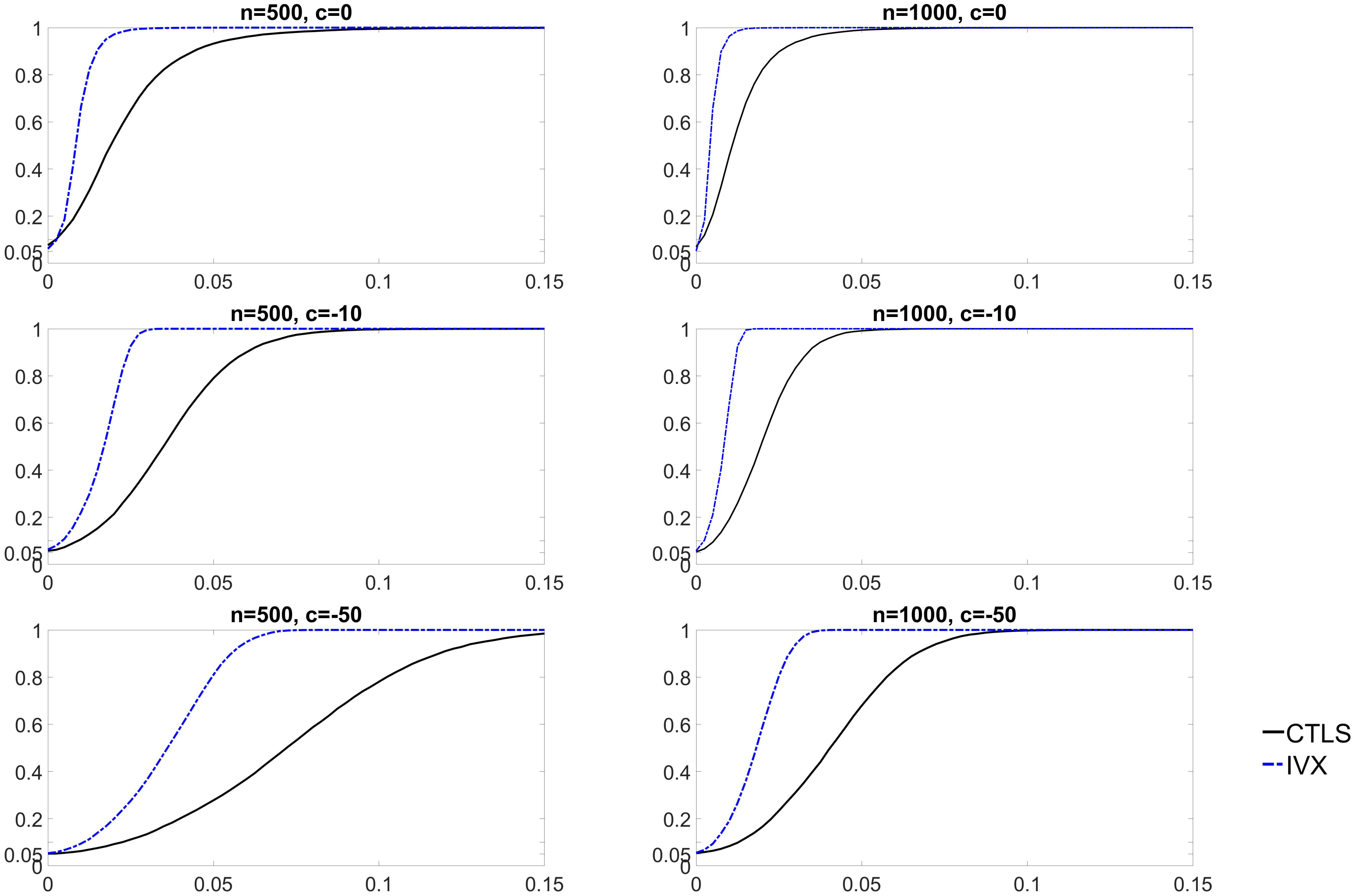
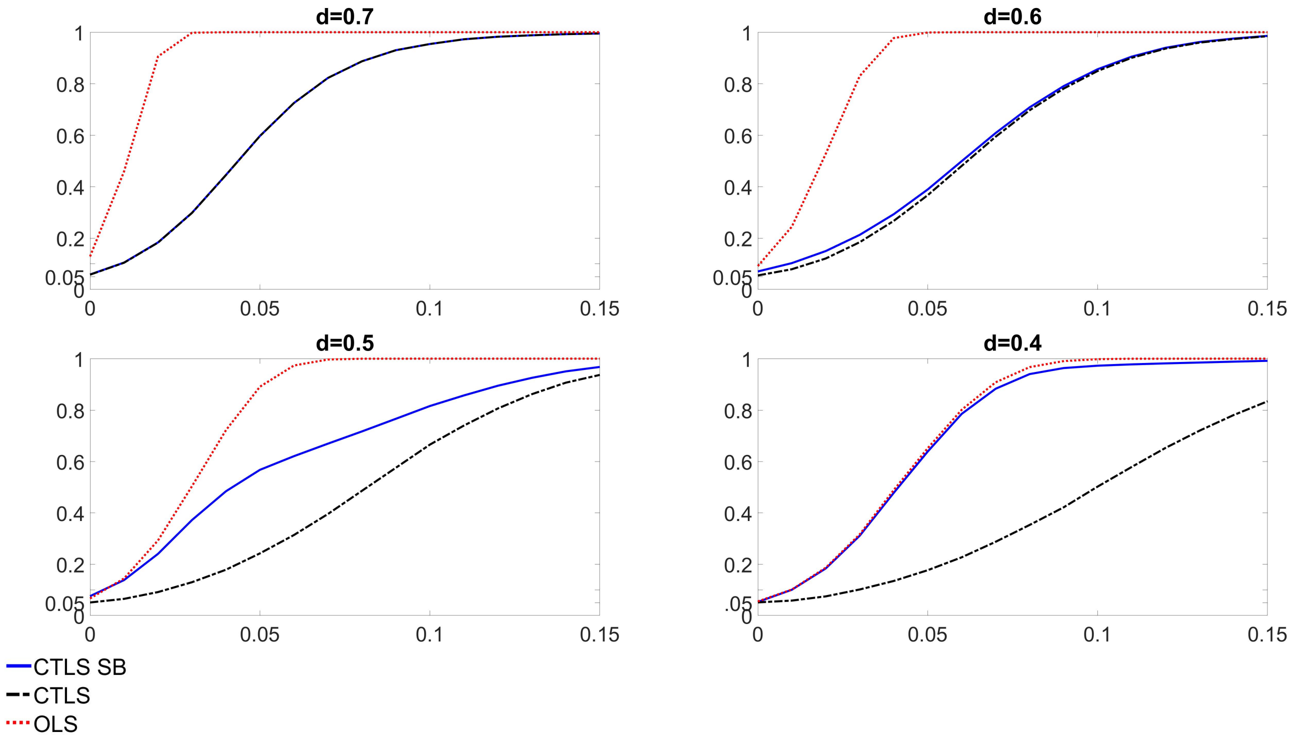


Figure 3: Empirical Power of CTLS tests plotted against $\beta$: $\hat{T}_{SB}$, $\hat{T}$, OLS (5% nominal size; $\varrho = -0.95$; fractional regressor, cond. homoscedastic regression errors)




## 6.2 Effects of Misbalancing on the Power of CTLS Tests

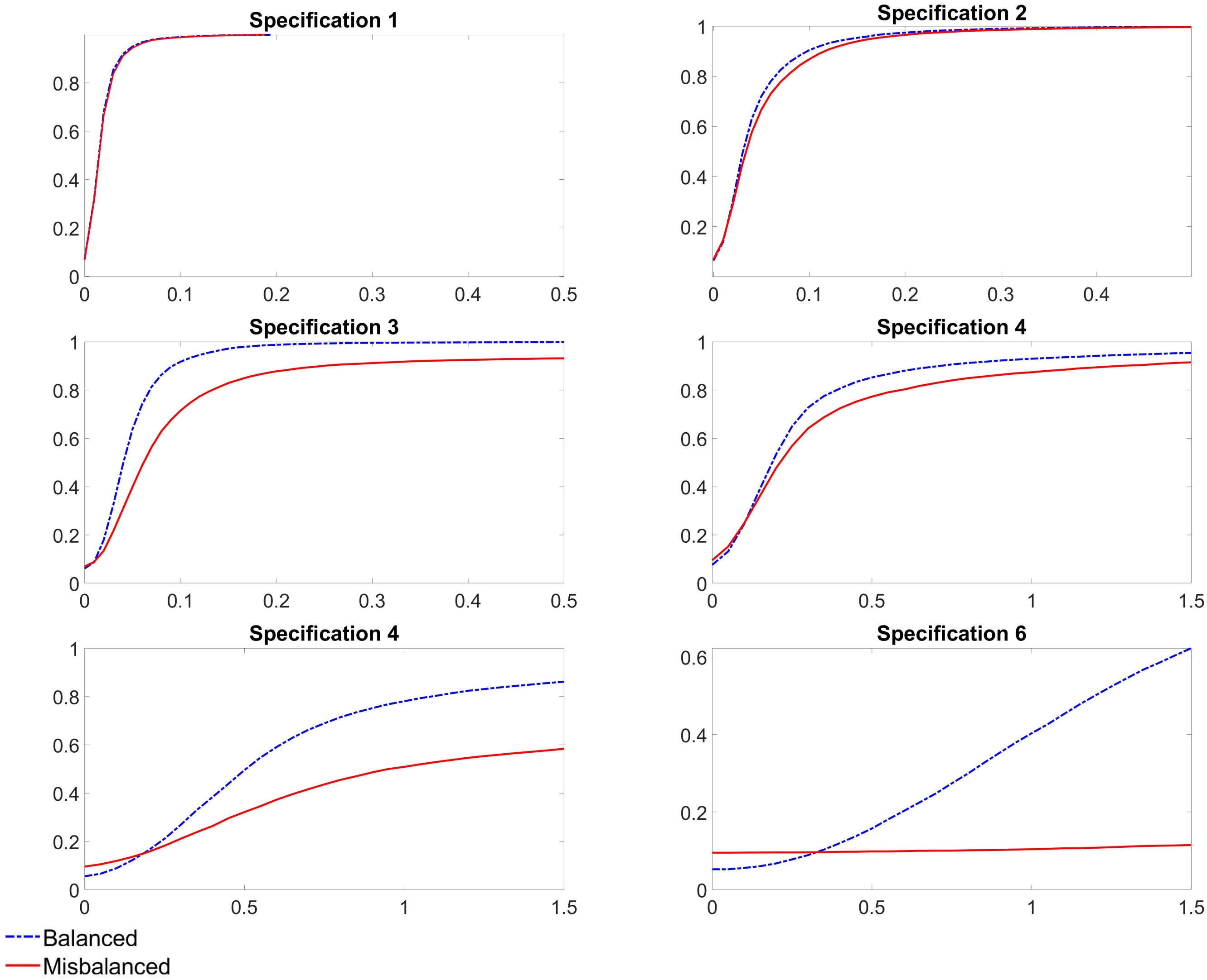
We next provide finite sample power results for the case where misbalancing has been committed due to functional form misspecification. We assume that the true regression model is given by (10) with $\sigma_k = 1$, and we consider 6 alternative DGPs for the true regression function:

| | |
|---|---|
| Specification 1: $f(x) = \text{sign}(x)\,|x|^{0.5}$ | Specification 4: $f(x) = \frac{x}{1+|x|}$ |
| Specification 2: $f(x) = \text{sign}(x)\,|x|^{0.25}$ | Specification 5: $f(x) = \frac{x}{1+|x|^{1.5}}$ |
| Specification 3: $f(x) = \ln|x|$ | Specification 6: $f(x) = \frac{x}{1+|x|^{3}}$ |

We report asymptotic power results for CTLS $\hat{T}$-statistics that are either based on a balanced fit (i.e. the regression function is correctly specified) and on a linear fit (i.e. the regression function is misspecified). The values of the $\hat{T}$-statistics are plotted against various values of the slope parameter $\beta$ in Figure 4. Notice that in all cases, the fitted regression function is of higher order of magnitude than the true one. For the computation of the $\hat{T}$-statistics we choose the following configuration: $K(x) = K(x)^* = \varphi_{0.1}(x)^{1/2}$, $c_n = n^{0.95}$, $l_n = c_n^{0.7}$. $\varrho = -0.95$. Additionally, the sample size is $n = 1000$.

It can be seen in Figure 4 that there is no substantial power drop when the true regression model is a polynomial of increasing order. Nevertheless, there is a substantial power reduction when the true regression function is logarithmic, bounded, or of decreasing order (i.e., vanishing at infinity). The poorest performance is evident in the case where $f(x)$ is integrable (Specification 6). For the integrable DGP, the test statistic appears to be bounded

Figure 4: Empirical Power: Balanced VS Misbalanced Models; $\hat{T}$ Plotted against $\beta$. (5% nominal size; unit root regressor; $\varrho = -0.95$; $i.d.\mathbf{N}(0,1)$ regression errors)




in probability if misbalancing is committed. This finding is consistent with some theoretical results provided by Kasparis et al. (2015).

# 7 Application to the predictability of stock returns

A substantial literature in empirical finance and econometrics investigates whether 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* predictability 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 of

$$r_k = \mu + \beta x_{k-1} + e_k, \tag{24}$$

where $r_k$ are stock returns relating to some stock index, $x_k$ some predictive variable and $e_k$ 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; Kostakis et al., 2015). Usually some financial variable (e.g., dividend yield, earnings-to-price ratio, book-to-market ratio, realized 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 the predictor is a stationary AR(1) process driven by i.i.d. innovations, and employ techniques that are generally valid only under stationarity. 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. Furthermore, Ang and Bekaert (2007) assume that the dividend yield is covariance stationary. However, 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 (see, e.g., Kostakis et al., 2015; Table 4). Christensen and Nielsen (2007) (see also Chistensen and Nielsen, 2006), Bollerslev et al. (2013), Bandi et al. (2019) develop methods that allow for stationary long memory predictors.

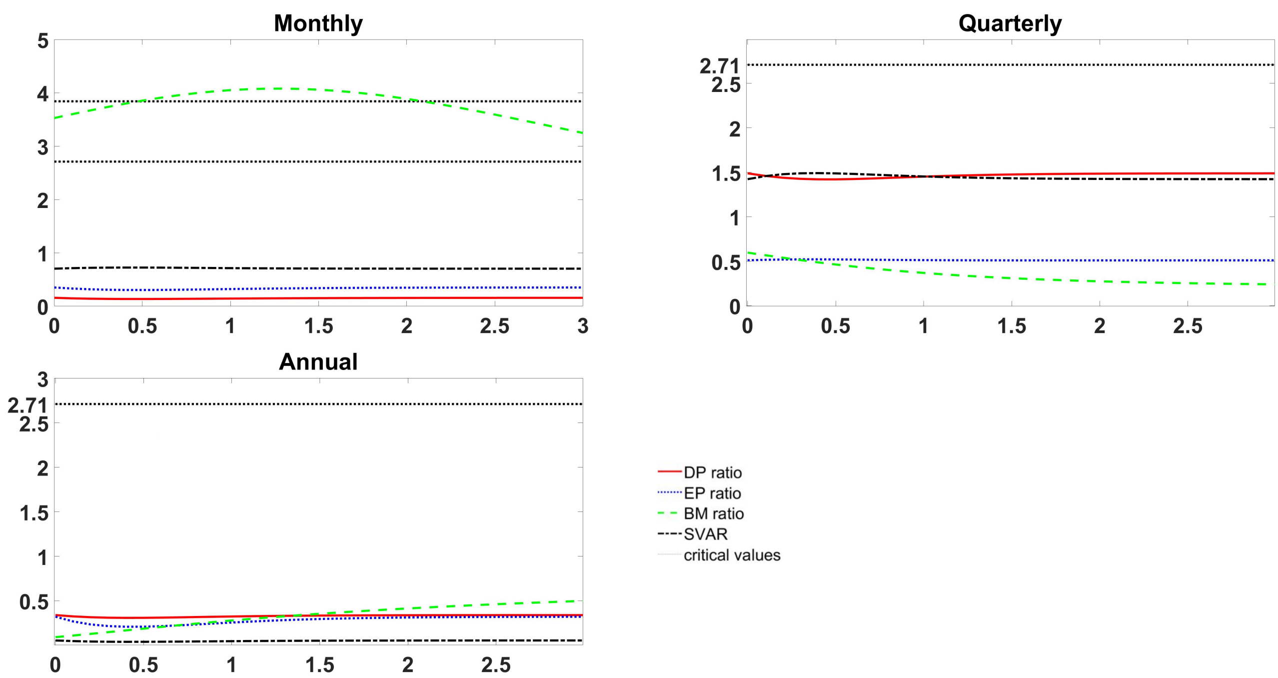
Several other articles focus on predictive regressions with nonstationary predictors. For example, Campbell and Yogo (2006) consider *conservative* testing procedures that allow for an NI predictor. The latter paper utilizes Bonferroni bounds with confidence intervals based on the inversion of unit root tests (see also Cavanagh et al., 1995). Phillips (2014) shows that testing procedures based on this approach provide good robustification for local deviations from unity but may 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 utilizing the IVX method of MP. MP 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. Several other studies have also used the IVX method in the context of predictive regressions. Gonzalo and Pitarakis (2012) investigate regime specific predictability in the context of threshold regressions while Demetrescu et al. (2022) examine episodic predictability in TVP predictive regressions. Both of the aforementioned papers develop predictability tests that utilize IVX instrumentation. A comprehensive review of the econometric methodology used in this area can be found in Phillips (2015).

Next, we investigate the predictability of stock returns using the Welch and Goyal data set, 2018. The returns variable ($r_k$) is the log differenced index. We consider four alternative predictors: Dividend-to-Price ratio (DP), Earnings-to-Price ratio (EP), Book-to-Market (BM), and a realized 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 of the memory properties of the data, we report memory estimates based on the local Whittle (LW) and the exact local Whittle (ELW; cf. Shimotsu and Phillips, 2005) estimators. It can be seen from Table 2 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.

Table 2: Memory Estimates (bandwidth $n^{0.65}$)

| | Monthly | | Quarterly | | Annual | |
|---|---|---|---|---|---|---|
| | LW | ELW | LW | ELW | LW | ELW |
| Returns | 0.07 | 0.07 | 0.14 | 0.15 | -0.12 | -0.08 |
| DP | 0.90 | 0.89 | 0.62 | 0.62 | 0.59 | 0.64 |
| EP | 0.99 | 1.00 | 0.68 | 0.68 | 0.44 | 0.47 |
| BM | 0.99 | 1.03 | 0.74 | 0.77 | 0.63 | 0.65 |
| SVAR | 0.53 | 0.53 | 0.46 | 0.47 | 0.33 | 0.35 |

Figure 5: Rolling CTLS F-statistics, $\hat{T}(\theta)^2$, plotted against $\theta \in [0, 3]$




$K(x) = \varphi_{0.1}(x)^{1/2}$, $K(x)^* = \varphi_1(x)^{1/2}$; $c_n = n^{0.95}$, $l_n = c_n^{0.7}$

We utilize the inferential techniques of Section 3. More specifically, we compute rolling CTLS t-statistics as per (19), employing the set of nonlinear regression functions $f(x,\theta) = \frac{x}{1+|x|^{\theta}}$, $\theta \in [0,3]$. We only report results based on $\hat{T}$. Tests based on $\hat{T}_{SB}$ lead to the same conclusions. Figure 5 reports the values of the CTLS F-statistics $\hat{T}(\theta)^2$ against $\theta$ for three sample frequencies (monthly, quarterly, annual). We find evidence of predictability only when monthly BM is used as a predictor. The null hypothesis is rejected at 10% significance level for all choices of $\theta$, and at 5% level for $0.5 \leq \theta \leq 2.1$ approximately. Interestingly, power is maximal for $\theta \approx 1.3$, which indicates a nonlinear (reciprocal) relationship between returns and BM of the form $f(x) \approx 1/x^{0.3}$, for large $x$. Our findings on the BM predictor are comparable with those of Kostakis et al. (2015), who also find predictability evidence for monthly data.

# APPENDIX

## 8 Asymptotics for Chronologically Trimmed Sample Functionals

In this Section of the Appendix, we develop basic limit theory for chronologically trimmed sample functionals of stationary and nonstationary processes that are relevant for regression analysis. This asymptotic theory is of independent interest and provides a generalization of existing limit results, e.g. Phillips et al. (2017) and Hu et al. (2024) who consider kernel methods for time varying parameter models. The notation in this section is the same as that used in Sections 2 and 3. In particular, $\{\mathbf{x}_k\}_{1\leq k\leq n}$ is a $p$-dimensional time series process and $\{X_{n,k}\}_{1\leq k\leq n, n\geq 1}$ is a $p$-dimensional random array that satisfies **A2** and **A3**, respectively, in Section 2. We assume that $K$ is an integrable kernel function, and for each $\mathbf{x} \in \mathbb{R}^p, \mathbf{f}(\mathbf{x}) = [f_1(x_1), ..., f_p(x_p)]'$, $\mathbf{q}(\mathbf{x}) = [q_1(x_1), ..., q_p(x_p)]'$ where $f_i$ and $q_i$ $(i = 1, ..., p)$ are measurable functions. Furthermore, let $\mathbf{F}(\mathbf{x}) = [1, \mathbf{f}(\mathbf{x})']'$ and $\mathbf{Q}(\mathbf{x}) = [1, \mathbf{q}(\mathbf{x})']'$, and set, for $l \in \mathbb{N}$, $\;0 < \tau_1 < ... < \tau_l < 1$ and $m = \{0, 2\}$,

$$
\begin{aligned}
S_{1n,l}^{(m)} &:= \frac{c_n}{n}\sum_{k=1}^{n}\mathbf{F}(\mathbf{x}_{k-1})\mathbf{F}(\mathbf{x}_{k-1})'\sigma_k^m\left\{\frac{1}{l}\sum_{j=1}^{l}K\left[c_n(k/n-\tau_j)\right]\right\}, \\
M_{1n,l} &:= \sqrt{\frac{c_n}{n}}\sum_{k=1}^{n}\mathbf{F}(\mathbf{x}_{k-1})\left\{\frac{1}{\sqrt{l}}\sum_{j=1}^{l}K\left[c_n(k/n-\tau_j)\right]\right\}\sigma_k u_k, \\
S_{2n,l}^{(m)} &:= \frac{c_n}{n}\sum_{k=1}^{n}\mathbf{F}(X_{n,k-1})\mathbf{F}(X_{n,k-1})'\sigma_k^m\left\{\frac{1}{l}\sum_{j=1}^{l}K\left[c_n(k/n-\tau_j)\right]\right\}, \\
S_{3n,l}^{(m)} &:= \frac{c_n}{n}\sum_{k=1}^{n}\mathbf{Q}(X_{n,k-1})\sigma_k^m\left\{\frac{1}{l}\sum_{j=1}^{l}K\left[c_n(k/n-\tau_j)\right]\right\}, \\
M_{2n,l} &:= \sqrt{\frac{c_n}{n}}\sum_{k=1}^{n}\mathbf{F}(X_{n,k-1})\left\{\frac{1}{\sqrt{l}}\sum_{j=1}^{l}K\left[c_n(k/n-\tau_j)\right]\right\}\sigma_k u_k,
\end{aligned}
$$

where $c_n$ is a sequence of positive constants, $l$ is 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 $X_{n,k}$, $\mathbf{x}_k$ are $\mathcal{F}_k$-measurable and $\sigma_k$ is $\mathcal{F}_{k-1}$-measurable (cf. Assumptions **A1** – **A3**).

The asymptotics of the functionals given above are utilized in Sections 3-5 for the analysis of the CTLS estimators and various test statistics. In particular, the limit theory for $\left\{S_{1n,l}^{(m)}, M_{1n,l}\right\}$ is relevant for stationary regressions, whilst asymptotics for $\left\{S_{2n,l}^{(m)}, S_{3n,l}^{(m)}, M_{2n,l}\right\}$ are relevant for nonstationary regressions. It is worth mentioning that the terms $S_{1n,1}^{(0)}$ and $S_{2n,1}^{(0)}$ resemble certain functionals considered by Hu et al. (2024) and Phillips et al. (2017) respectively. The latter paper studies cointegrated TVP models, whilst the former considers TVP models with stationary covariates. More specifically, Phillips et al. (2017) utilize statistics of the form

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

where $X_{n,k}$ is an $I(1)$ vector process normalized by $\sqrt{n}$. Under our assumptions, $X_{n,k}$ can be an appropriately normalized $I(d)$, $d > 1/2$, process or a NI array possibly driven by fractional errors, see e.g. (3). Therefore, the limit theory of this section is also relevant for the estimation of TVP models with stationary and nonstationary covariates.

We have the following Theorem 5 and Theorem 6, providing the limit results for stationary and nonstationary functionals, respectively.

**Theorem 5.** *Suppose **A2** and **A4** or **A4**$^*$ hold. Then, as $n \to \infty$,*

$$S_{1n,l_n}^{(m)} = E[\sigma_2^m \mathbf{F}(\mathbf{x}_1)\mathbf{F}(\mathbf{x}_1)'] \int K + o_P(1). \tag{25}$$

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

$$M_{1n,l_n} \to_d \mathbf{N}\left(\mathbf{0},\ E\left[\sigma_2^2 \mathbf{F}(\mathbf{x}_1)\mathbf{F}(\mathbf{x}_1)'\right]\int K^2\right). \tag{26}$$

**Theorem 6.** *Suppose that **A3** and **A4** or **A4**$^*$ hold, and $\mathbf{f}(.)$, $\mathbf{q}(.)$ are continuous. Then, as $n \to \infty$,*

$$\begin{aligned}\left[S_{2n,l_n}^{(m)}, S_{3n,l_n}^{(m)}\right] &= \left[\int_0^1 \mathbf{F}(X_{n,[nt]})\mathbf{F}(X_{n,[nt]})'dt,\ \int_0^1 \mathbf{Q}(X_{n,[nt]})dt\right] E(\sigma_1^m)\int K + o_P(1)\\ &\to_d \left[\int_0^1 \mathbf{F}(\mathcal{X}_t)\mathbf{F}(\mathcal{X}_t)'dt,\ \int_0^1 \mathbf{Q}(\mathcal{X}_t)dt\right] E(\sigma_1^m)\int K. \end{aligned} \tag{27}$$

*If in addition **A1** holds, we have jointly with (27)*

$$M_{2n,l_n} \to_d \mathbf{MN}\left(\mathbf{0},\ E\left(\sigma_1^2\right)\int_0^1 \mathbf{F}(\mathcal{X}_t)\mathbf{F}(\mathcal{X}_t)'dt \int K^2\right). \tag{28}$$

*Remark* 9. If we are only interested in the limit behaviour of $S_{1n,l_n}^{(m)}$ and $S_{2n,l_n}^{(m)}$, conditions **A2** and **A3** can be relaxed. For instance, the result of (27) still holds, if (6) is replaced by $X_{n,[nt]} \Rightarrow X_t$ on $D_{\mathbb{R}^p}[0,1]$.

*Remark* 10. The continuity requirement for $\mathbf{f}$ in Theorem 6 is not essential for (27) and (28). These results can be extended to the case where $\mathbf{f}$ is locally Lebesgue integrable if we impose more smoothness conditions on $X_{n,k}$. See Remark 1 for further discussion.

*Remark* 11. It is easy to see from the proof of Theorem 5 that (25) and (26) still hold, if **A4** (c) or **A4**$^*$ (c) is replaced by $\tau_j = j/(l+1)$, where $j = 1, ..., l$ (i.e., $l_n \equiv l$ is fixed). As for (27) and (28), if **A4** (c) or **A4**$^*$ (c) is replaced by $\tau_j = j/(l+1)$ $(j = 1, ..., l)$, we have

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

*Remark* 12. Theorem 5 provides a generalization of the limit theory of Hu et al. (2024) who consider kernel functionals with a single *chronological point*. Furthermore, Theorem 6 provides a generalization of a comparable result as in Phillips et al. (2017) who consider kernel functionals of time trend weighted by $I(1)$ processes. Theorem 6 is more general than the existing limit results in several ways:

**(i)** We consider more general nonstationary processes. In fact, our assumptions on $X_{n,k}$ are minimal and are not model specific. The main requirement is that $X_{n,k}$ converges in the $D_{\mathbb{R}^p}[0,1]$ space.

**(ii)** We allow for nonlinear transformations of $X_{n,k}$.

**(iii)** The $S_{2n,l}^{(m)}$, $M_{2n,l}$ functionals entail both stationary and nonstationary processes. This is relevant for example for the asymptotic analysis of TVP models that have both stationary and nonstationary covariates. Further, the assumptions for the stationary process $\sigma_k^m$ are quite general -e.g. $\sigma_k^m$ is allowed to be a strictly stationary long memory process.

**(iv)** The kernel functionals allow for multiple *chronological points.*

Theorem 6 provides a limit theory for rescaled functionals of nonstationary processes (i.e., $X_{n,k} = D_n^{-1}\mathbf{x}_k$ as in **A3**). For the purposes of regression analysis, limit theory for non-rescaled processes (i.e., $X_{nk}$ is replaced by $\mathbf{x}_k$) is more relevant. Following Park and Phillips (1999, 2001), we next assume that $\mathbf{f}(.)$ and $\mathbf{q}(.)$ are vectors of *AHFs*. Recall that $K_{kn} := \sum_{j=1}^{l_n} K\left[c_n\left(k/n - \tau_j\right)\right]$. The following result is the counterpart of Theorem 6 for transformations of non-rescaled sequences.

**Theorem 7.** *Suppose that:*

(*a*) ***A1, A3** and **A4** or **A4**$^*$ hold.*

(*b*) *For each $i = 1, ..., p$, $f_i$ is an AHF with limit homogeneous function $H_{f_i}$ and asymptotic order $\pi_{f_i}$. Further, for $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{H}_F(\mathbf{x}) := \left[1, H_{f_1}(x_1), ..., H_{f_p}(x_p)\right]'$, and $\mathcal{L}_n :=$ diag $\left\{1, \pi_{f_1}(d_{1n}), ..., \pi_{f_p}(d_{pn})\right\}$.*

(*c*) *For each $i = 1, ..., p$, $q_i$ is an AHF with limit homogeneous function $H_{q_i}$ and asymptotic order $\pi_{q_i}$. Further, for $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{H}_Q(\mathbf{x}) := \left[1, H_{q_1}(x_1), ..., H_{q_p}(x_p)\right]'$, and $\mathcal{C}_n :=$ diag $\left\{1, \pi_{q_1}(d_{1n}), ..., \pi_{q_p}(d_{pn})\right\}$.*

*Then, as $n \to \infty$, the following weak limits hold jointly*

$$\begin{aligned}
\frac{c_n}{nl_n}\sum_{k=1}^{n}\mathcal{C}_n^{-1}\mathbf{Q}(\mathbf{x}_{k-1})\sigma_k^m K_{kn} &= \frac{c_n}{nl_n}\sum_{k=1}^{n} H_{\mathbf{Q}}(X_{n,k})\sigma_k^m K_{kn} + o_P(1) \\
&\to_d E\left(\sigma_k^m\right)\int_0^1 H_{\mathbf{Q}}(\mathcal{X}_t)dt\int K, \qquad (29)\\
\frac{c_n}{nl_n}\sum_{k=1}^{n}\mathcal{L}_n^{-1}\mathbf{F}(\mathbf{x}_{k-1})\mathbf{F}(\mathbf{x}_{k-1})'\mathcal{L}_n^{-1}\sigma_k^m K_{kn} &= \frac{c_n}{nl_n}\sum_{k=1}^{n} H_{\mathbf{F}}(X_{n,k})H_{\mathbf{F}}(X_{n,k})'\sigma_k^m K_{kn} + o_P(1) \\
&\to_d E\left(\sigma_k^m\right)\int_0^1 H_{\mathbf{F}}(\mathcal{X}_t)H_{\mathbf{F}}(\mathcal{X}_t)'dt\int K, \qquad (30)\\
\sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^{n}\mathcal{L}_n^{-1}\mathbf{F}(\mathbf{x}_{k-1})\sigma_k u_k K_{kn} &= \sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^{n} H_{\mathbf{F}}(X_{n,k})\sigma_k u_k K_{kn} + o_P(1) \\
&\to_d \mathbf{MN}\left[\mathbf{0},\ E\left(\sigma_k^2\right)\int_0^1 H_{\mathbf{F}}(\mathcal{X}_t)H_{\mathbf{F}}(\mathcal{X}_t)'dt\int K^2\right]. \qquad (31)
\end{aligned}$$

The proof of Theorem 7 follows easily from Theorem 6 and the AHF assumption. Proofs are provided in the Online Supplement.

# 9 Asymptotics for a General Class of Sample Functionals

This section provides a significant extension to Lemma 5.1 of Hu et al. (2021b), establishing asymptotics for a general class of sample functionals. This extension includes Theorems 5 and 6 as corollaries. We start with some notation:

- $\{X_{n,k}\}_{k\geq 1, n\geq 1}$ is a random $q$-dimensional vector array;
- $\{v_k\}_{k\geq 1}$ is a $q\times q$ random matrix sequence;
- $G:\mathbb{R}^q \to \mathbb{R}^{q\times q}$ is a continuous matrix function, i.e., for every $r, s = 1, 2, \cdots, q$, the function $[G(\mathbf{x})]_{rs}$ is continuous on $\mathbb{R}^q$;
- $K:\mathbb{R}\to\mathbb{R}$ is a Borel function; and
- for $0 < \tau_1 < ... < \tau_l < 1$, set

  $$\mathcal{S}_{n,l} = \frac{c_n}{n}\sum_{k=1}^{n} G(X_{n,k})v_k\frac{1}{l}\sum_{j=1}^{l}K\left[c_n(k/n-\tau_j)\right],$$

  where $\{c_n\}_{n\geq 1}$ is a sequence of positive constants.

For the asymptotic behavior of $\mathcal{S}_{n,l}$, we have the following result.

**Lemma 1.** *Suppose that*

*(a) there is a continuous process $\mathcal{X}_t$ such that $X_{n,[nt]} \Rightarrow \mathcal{X}_t$ on $D_{\mathbb{R}^q}[0,1]$;*

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

*(c) $K(x)$ 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*

$$\begin{aligned}\mathcal{S}_{n,l} &= \frac{1}{l}\sum_{j=1}^{l} G(X_{n,[n\tau_j]})\, A_0 \int K + o_P(1) \\ &\to_d \frac{1}{l}\sum_{j=1}^{l} G(\mathcal{X}_{\tau_j})\, A_0 \int K. \end{aligned} \tag{32}$$

*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*

$$\mathcal{S}_{n,l_n} = \int_0^1 G(X_{n,[nt]})dt\, A_0 \int K + o_P(1) \to_d \int_0^1 G(\mathcal{X}_t)dt\, A_0 \int K. \tag{33}$$

*Remark* 13. In view of the weak convergence requirement in (a), the continuity of $G$ is sufficient for this result, but not necessary. The result can be extended to the case where $G$ is locally Lebesgue integrable if we impose additional smoothness conditions on $X_{n,k}$. However, this generalization would involve more complicated calculations. We will not pursue this extension in this paper. It is worth mentioning that no relationship is imposed between $v_k$ and $X_{n,k}$. Moreover, condition (b) is satisfied with $A_0 = Ev_1$ whenever $v_k$ is (strictly) stationary and ergodic 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 the proofs of the main results.

If we are only concerned with the boundedness of $\mathcal{S}_{n,l}$, weaker conditions are required compared to those used in the previous lemma.

**Lemma 2.** *Suppose that conditions (a) and (c) 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*

$$\frac{c_n}{n}\sum_{k=1}^{n} \|G(X_{n,k})\|\,\|v_k\|\,\frac{1}{l}\sum_{j=1}^{l} K\big[c_n(k/n - \tau_j)\big] = O_P(1). \tag{34}$$

*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*

$$\frac{c_n}{n}\sum_{k=1}^{n} \|G(X_{n,k})\|\,\|v_k\|\,\frac{1}{l_n}\sum_{1\leq i<j\leq l_n} K\big[c_n(k/n - \tau_i)\big]K\big[c_n(k/n - \tau_j)\big] = o_P(1), \tag{35}$$

$$\frac{c_n}{n}\sum_{k=1}^{n} \|G(X_{n,k})\|\,\|v_k\|\,\Big(\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n} K\big[c_n(k/n - \tau_j)\big]\Big)^2 = O_P(1), \tag{36}$$

$$\big(\frac{c_n}{n}\big)^2\sum_{k=1}^{n} \|G(X_{n,k})\|\,\|v_k\|\,\Big(\frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n} K\big[c_n(k/n - \tau_j)\big]\Big)^4 = o_P(1). \tag{37}$$

*Remark* 14. The results of Lemmas 1 and 2 hold true, if $K(x)$ is replaced by $x^a K^b(x)$ for any $a \geq 0$ and any $b \geq 1$ under the additional condition $\int |x^a K^b| < \infty$. This claim is obvious from the proofs of the aforementioned lemmas and will be used in the proofs of the main results. Let $K^*$ be another kernel function, possibly different from $K$, that satisfies the same conditions as $K$ does. Then the same argument used for the proof of (35) yields

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

This, together with Lemma 1, implies that

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

# 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. (2021). Consistent inference for predictive regressions in persistent economic systems. *Journal of Econometrics*, **224**(1), 215–244.

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

[4] Bandi, F.M. and Perron, B. (2008) Long-run risk-return trade-offs. *Journal of Econometrics*, **143**, 349–374.

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

[6] 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.

[7] 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.

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

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

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

[11] Chang, Y., Park, J.Y. and Phillips P.C.B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal*, **4**, 1–36.

[12] 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.

[13] 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.

[14] 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.

[15] Christopeit, N. (2009). Weak convergence of nonlinear transformations of integrated processes: The multivariate case. *Econometric Theory*, **25**, 1180–1207.

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

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

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

[19] 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.

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

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

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

[23] Gonzalo, J. and Pitarakis, J-Y. (2012). Regime-specific predictability in predictive regressions. *Journal of Business and Economic Statistics*, **30**(2), 229–241.

[24] Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, **64**, 413–430.

[25] Hu, Z., Kasparis, I. and Wang, Q. (2021a). Chronologically trimmed LS for nonlinear predictive regressions with persistence of unknown form. *mimeo.*

[26] Hu, Z., Kasparis, I. and Wang, Q. (2024). Time varying parameter models with stationary persistent data. *Econometric Theory*, forthcoming.

[27] Hu, Z., Phillips, P.C.B. and Wang, Q. (2021b). Nonlinear cointegrating power regression with endogeneity. *Econometric Theory*, **37**(1), 138–168.

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

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

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

[31] Jin, C and Wang, Q. (2024). Weighted nonlinear regression with nonstationary time series. *Statistica Sinica*, **34**, forthcoming. doi:https://doi.org/10.5705/ss.202021.0426

[32] Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models*. Oxford University Press, New York.

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

[34] Kasparis, I. (2011). Functional form misspecification in regressions with a unit root. *Econometric Theory*, **27**, 285–311.

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

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

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

[38] Magdalinos, T. (2022). Least squares and IVX limit theory in systems of predictive regressions with GARCH innovations. *Econometric Theory*, **38**(5), 875–912.

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

[40] Magdalinos, T. and Petrova, K. (2022). Uniform and distribution-free inference with general autoregressive processes, Working Paper 1344, Barcelona School of Economics.

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

[42] McCracken, M.W. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, **140**(2), 719–752.

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

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

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

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

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

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

[49] 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

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

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

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

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

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

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

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

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

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

[59] Wang, Q. (2021). Least squares estimation for nonlinear regression models with heteroscedasticity. *Econometric Theory*, **37**(6), 1173–1213.

[60] 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.

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

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

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

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

# Online Supplement to "CHRONOLOGICALLY TRIMMED LS FOR NONLINEAR PREDICTIVE REGRESSIONS WITH PERSISTENCE OF UNKNOWN FORM"

Zhishui Hu
*University of Science & Technology of China*

Ioannis Kasparis
*University of Cyprus*

Qiying Wang
*The University of Sydney*

May 8, 2024

This supplement is organized as follows. Section S1 provides a more detailed review of existing robust inferential methods under temporal dependence. In Section S2, we investigate CTLS inference in multi-covariate models, providing an extension of Theorem 3 given in Section 4 of the main paper. Section S3 provides the proofs of Lemmas 1 and 2 of the main paper. The proofs of the theorems in the main paper are given in Sections S4 to S10. Since Theorems 5 and 6 provide the basic tools for other proofs, we arrange the proofs of Theorem 5 and 6 in Sections S4 and S5, respectively. The proof of Theorem S1 of this supplement is given in Section S11. Section S12 gives some additional simulation results. Throughout the Supplement, unless mentioned explicitly, we use the same notation and equations as those given in the main paper.

## S1 A Detailed Review of Existing Inferential Methods

The first approach put forward for addressing the dichotomy in inference, between stationary and nonstationary regimes, relies on so-called conservative methods. In particular, a number of studies develop procedures that yield robust inference in the presence of NI processes, in the context of reduced form regressions where the covariate is predetermined with respect to the regression error. For example, Cavanagh et al. (1995), Campbell and Yogo (2006), Janson and Moreira (2006), Elliott et al. (2015) study parametric models with a NI covariate. The aforementioned papers propose test statistics with limit distributions free of nuisance near-to-unity parameters. 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

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

procedures provide valid inference under local deviations from a unit root, their emphasis is on 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 studentized 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. $\mathbf{N}(0,1)$, $\chi^2$). 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, tail parameters (heavy tailed data), TVPs, different types of nonlinearities in the regression function, etc.

The possibility of fractional predictors has received little attention in the literature on robust predictive regressions, despite substantial related work on fractional cointegrated systems, e.g., Robinson and Hualde (2003), Christensen and Nielsen (2006), Hualde and Robinson (2010), and Andersen and Varneskov (2021). The specifications considered in the aforementioned studies are in general structural (i.e. covariates may not be predetermined) and in some cases (e.g., Hualde and Robinson, 2010; Andersen and Varneskov, 2021) both stationary and nonstationary long memory are allowed. These methods are mainly semi-parametric (spectral OLS) with respect to the short memory components of the system and may attain sub-OLS[2] convergence rates due to bandwidth parameters. Regression estimators have mixed Gaussian or Gaussian limit distributions, and therefore inference is conventional in this framework. However, preliminary memory estimators are required, which makes implementation somewhat more involved. Further, although these models are quite general, nonlinearities and nearly integrated arrays are 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 may result in severe size distortions when there are near-to-unity parameters.

The relatively recent work of Magdalinos and Phillips (2009, MP hereafter) provides an alternative approach to inference in regressions with possible nonstationary covariates. This study proposes instrumentation based on certain linear filtering of the regressors. The resultant IVX instruments exhibit weaker signals than those of the covariates and, as a result, induce mixed normal limit distributions in situations where independent variables are unit roots or NI arrays. IVX instrumentation yields asymptotically vanishing endogeneity, and this is sufficient for a martingale CLT to operate. Hence, contrary to the OLS estimator, in the presence of nonstationary data the IVX estimator has mixed Gaussian limit distribution and studentized IVX estimators have $\mathbf{N}(0,1)$ (t-tests) or $\chi^2$ (F-tests) limit distributions. 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. 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

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

relating to intercept demeaning. For further work on the IVX method, see, e.g. Yang et al. (2020), Demetrescu et al. (2022), Magdalinos (2022), Magdalinos and Petrova (2022) and the references therein.

## S2 CTLS Inference in Multi-Covariate Models

In this section we extend CTLS based inference to multi-covariate models as per (4) in the main paper. Recall that the specification of the aforementioned equation is as follows.

$$y_k = \mu + \boldsymbol{\beta}'\mathbf{f}(\mathbf{x}_{k-1}) + e_k,$$

where the covariate and the slope parameter $\boldsymbol{\beta}$ are $p$-dimensional. As in (5) of the main paper, we may rewrite the model above as

$$y_k = \boldsymbol{\theta}'\mathbf{F}(\mathbf{x}_{k-1}) + e_k, \tag{S2.1}$$

where $\mathbf{F}(\mathbf{x}_{k-1}) = [1, \mathbf{f}(\mathbf{x}_{k-1})']'$ and $\boldsymbol{\theta} = [\mu, \boldsymbol{\beta}']'$. Define $\mathbf{f}_{k-1} = \mathbf{f}(\mathbf{x}_{k-1}), \mathbf{F}_{k-1} = \mathbf{F}(\mathbf{x}_{k-1})$ and

$$\left\{\mathcal{H}_n,\ \hat{\mathcal{V}}_n,\ \mathcal{A}_n\right\} := \left\{\sum_{k=1}^{n} \mathbf{f}_{k-1}\overline{\mathbf{f}}'_{k-1}K_{kn},\ \sum_{k=1}^{n} \check{e}_k^2\mathbf{F}_{k-1}\mathbf{F}'_{k-1}K_{kn}^2,\ \left[-\overline{\mathbf{f}},\ I_p\right]\right\}, \tag{S2.2}$$

where $\check{e}_k := y_k - \hat{\boldsymbol{\theta}}'_{LS}\mathbf{F}_{k-1}$ are the OLS residuals, $I_p$ is the $p$-dimensional identity matrix, $\overline{\mathbf{f}}_{k-1} = \mathbf{f}_{k-1} - \overline{\mathbf{f}}$, and $\overline{\mathbf{f}} = \sum_{k=2}^{n} \mathbf{f}_{k-1}K_{kn} / \sum_{k=2}^{n} K_{kn}$. For the single restriction hypothesis

$$H_0 : \beta_i = \eta \in \mathbb{R}, \quad i = 1, ..., p, \tag{S2.3}$$

we have the following general formulation of the CTLS t-statistic

$$\hat{T}_i = \frac{\hat{\beta}_i - \eta}{\sqrt{\left[\mathcal{H}_n^{-1}\mathcal{A}_n\hat{\mathcal{V}}_n\mathcal{A}'_n\mathcal{H}_n^{-1}\right]_{ii}}},$$

where recall $[.]_{ii}$ stands for the $i^{th}$ diagonal element of some matrix. We also consider multiple restrictions of the form

$$H_0 : R\boldsymbol{\beta} = \boldsymbol{\eta}, \tag{S2.4}$$

where $R$ is a $q \times p$ ($q \leq p$) matrix and $\boldsymbol{\eta}$ a predetermined $q$-dimensional vector. For the latter type of restrictions we consider the CTLS F-statistic

$$\hat{F} = \left[R\hat{\boldsymbol{\beta}} - \boldsymbol{\eta}\right]' \left[\mathcal{H}_n^{-1}\mathcal{A}_n\hat{\mathcal{V}}_n\mathcal{A}'_n\mathcal{H}_n^{-1}\right]^{-1} \left[R\hat{\boldsymbol{\beta}} - \boldsymbol{\eta}\right].$$

In the presence of nonstationary regressors, the CTLS estimators attains multiple convergence rates due to a variation in the degree of persistence, between various covariates, and nonlinearities arising from the regression model (see Theorem 2 in the main paper). This phenomenon requires matrix normalization for various components in $\hat{F}$. Matrix normalization creates technical difficulties due to non commutability of matrix products -for a discussion see Magdalinos and Phillips (2018). To avoid these technical difficulties, under nonstationarity we assume $q = p$ and $R = I_p$. This is general enough to allow for the joint predictability restrictions $\boldsymbol{\beta} = \mathbf{0}$.

We now state the main results in this multi-covariate Models under the null hypothesis when the regressor is either stationary or nonstationary as discussed in Theorem 3 of the main paper. Its proof is given in Section S11.

**Theorem S1.** *Suppose that, in addition to the conditions of Theorem 1 or Theorem 2 in the main paper,* $\sup_{k\geq 1} Eu_k^4 < \infty$. *Under* $H_0 : \beta_i = \eta$, *we have*

$$\hat{T}_i \to_d \mathbf{N}(0,1). \tag{S2.5}$$

*Furthermore, under* $H_0 : R\boldsymbol{\beta} = \boldsymbol{\eta}$

$$\hat{F} \to_d \chi_q^2. \tag{S2.6}$$

As explained in Remark 8 of the main paper, the requirement $\sup_{k\geq 1} Eu_k^4 < \infty$ can be dispensed with when the regression errors are conditionally homoscedastic, i.e. $\sigma_k^2 = \sigma^2$ for all $k$.

# S3 Proofs of Lemmas 1 and 2

## S3.1 Proof of Lemma 1

We only prove (32). The proof of (33) is similar and relatively simple. We shall first assume 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 that $A = 1$. 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 that $\tau_j = j/(l_n + 1)$. Since

$$|c_n(k/n - \tau_j)| < 1 \quad \text{only if} \quad \delta_{1n,j} \leq k \leq \delta_{2n,j}, \quad j = 1, ..., l_n, \tag{S3.1}$$

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

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

we have

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

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

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

$$
R_{jn}=o_P(1),\quad j=1,\ 2. \tag{S3.2}
$$

To prove (S3.2), we start with some preliminaries. Recalling $X_{n,[nt]}\Rightarrow \mathcal{X}_t$ on $D_{\mathbb{R}^q}[0,1]$ and that the limit process $\mathcal{X}_t$ is path continuous, we have $X_{n,[nt]}\Rightarrow \mathcal{X}_t$ on $D_{\mathbb{R}^q}[0,1]$ in the sense of uniform topology. See, for instance, Section 18 of Billingsley (1968). This fact implies that

$$
\limsup_{N\to\infty}\limsup_{n\to\infty} P\left(\max_{1\le k\le n}\|X_{n,k}\|\ge N\right)=0, \tag{S3.3}
$$

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

$$
P\left(\sup_{|s-t|\le\tilde\delta}\big\|X_{n,[nt]}-X_{n,[ns]}\big\|\ge\delta\right)\le\varepsilon \tag{S3.4}
$$

holds for all sufficiently large $n$. In view of (S3.4), for any $\delta>0$, we have

$$
\lim_{n\to\infty} P\left(\max_{1\le j\le l_n}\ \max_{\delta_{1n,j}\le l\le k\le\delta_{2n,j}}\|X_{n,k}-X_{n,l}\|\ge\delta\right)=0. \tag{S3.5}
$$

We are now ready to prove (S3.2), starting with $j=1$.

For any $N>0$, and any real $\mathbf{x}\in\mathbb{R}^q$ define $G_N(\mathbf{x})=\xi_N(\mathbf{x})G(\mathbf{x})$ with

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

Set

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

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

$$P(R_{1n} \neq \widetilde{R}_{1n}) \leq P\left(\max_{1\leq k\leq n} \|X_{n,k}\| \geq N\right) \to 0. \tag{S3.6}$$

Moreover,

$$\left\|\widetilde{R}_{1n}\right\| \leq \frac{C_N}{l_n} \sum_{j=1}^{l_n} \left\|R_{1n,j} - A_0 \int K\right\|, \tag{S3.7}$$

where $C_N := \sup_{\mathbf{x}\in\mathbb{R}^q} \|G_N(\mathbf{x})\| < \infty$, due to fact that $G_N$ is continuous with compact support. The result (S3.2) with $j = 1$ will follow if we prove

$$\max_{1\leq j\leq l_n} E\left\|R_{1n,j} - A_0 \int K\right\| \to 0, \tag{S3.8}$$

as $n \to \infty$. Indeed, by virtue of (S3.7) and (S3.8), we have $E\left\|\widetilde{R}_{1n}\right\| \to 0$ for each $N \geq 1$. This, together with (S3.6), yields $R_{1n} = o_P(1)$.

Since, as $n \to \infty$,

$$\max_{1\leq j\leq l_n} \left|\frac{c_n}{n} \sum_{k=\delta_{1n,j}}^{\delta_{2n,j}} K\left[c_n(k/n - \tau_j)\right] - \int K\right| \to 0, \tag{S3.9}$$

to prove (S3.8), it suffices to show that $\max_{1\leq j\leq l_n} E\|A_n(\tau_j)\| \to 0$, where

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

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 (S3.1), we may write

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

$$+\frac{c_n}{n}\sum_{s=T_{1n,j}}^{T_{2n,j}}\sum_{k=s\gamma}^{(s+1)\gamma}\|v_k-A_0\|\,|K\left[c_n(k/n-\tau_j)\right]-K\left[c_n(s\gamma/n-\tau_j)\right]|+o_P(1)$$

$$:= \quad A_{1n}(\tau_j)+A_{2n}(\tau_j)+o_P(1).$$

Recall that $\sup_{k\geq 1}E\|v_k\|<\infty$ by condition (b). Therefore, it follows from the Lipschitz continuity of $K(x)$ 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, in view of condition (b) and the fact that $\max_{1\leq j\leq l_n}\Big|A_{3n}(\tau_j)-\int K\Big|\to 0$ we have

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

where

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

In view of the results above, (S3.5) holds true, and this completes the proof of $R_{1n}=o_P(1)$.

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

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

In view of (S3.6), we have

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

as $n\to\infty$ first and then $N\to\infty$. For the asymptotic negligibility of $R_{2n}$ it suffices to show that $\widetilde{R}_{2n}=o_P(1)$, for each fixed $N\geq 1$.

By definition, $G_N(\mathbf{x})$ is continuous with compact support. Hence, for any $\epsilon>0$ there exists a $\delta_\epsilon>0$ such that $\|G_N(\mathbf{x})-G_N(\mathbf{y})\|\leq\epsilon$ whenever $\|\mathbf{x}-\mathbf{y}\|\leq\delta_\epsilon$. Write

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

By virtue of the facts above and (S3.9), it is readily seen that

$$\max_{1\leq j\leq l_n}E\left[\left\|\widetilde{R}_{2n,j}\right\|I(\Omega_{\delta_\epsilon})\right]$$

$$\leq \ E\Big\{\max_{1\leq j\leq l_n}\max_{\delta_{1n,j}\leq l\leq k\leq\delta_{2n,j}}\|G_N(X_{n,k})-G_N(X_{n,l})\|\ \frac{c_n}{n}\sum_{k=\delta_{1n,j}+1}^{\delta_{2n,j}}\|v_k\|\,|K\big[c_n(k/n-\tau_j)\big]\Big\}$$

$$\leq \ \epsilon\sup_{k\geq 1}E\,\|v_k\|\ \frac{c_n}{n}\sum_{k=\delta_{1n,j}+1}^{\delta_{2n,j}}K\big[c_n(k/n-\tau_j)\big]\leq C_N\epsilon,$$

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 (S3.5)],

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

Hence, for all $n\geq n_0$,

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

where $\bar{\Omega}_{\delta_\epsilon}$ denotes the complementary set of $\Omega_{\delta_\epsilon}$ 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 $R_{2n}=o_P(1)$.

Finally, we 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 a $K_{\epsilon,A}(x)$ such that

$$\int|K-K_{\epsilon,A}|\leq 2\epsilon, \tag{S3.10}$$

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 $A\geq\max\{A_1,A_{1\epsilon}\}+1$,

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

To show (32) it suffices proving that as $n\to\infty$ first and then $\epsilon\to 0$ (implying $A\to\infty$),

$$S_{n,\epsilon}:=\frac{1}{l_n}\sum_{j=1}^{l_n}\frac{c_n}{n}\sum_{k=1}^{n}G(X_{n,k})\,v_k\,\widetilde{K}\big[c_n(k/n-\tau_j)\big]=o_P(1), \tag{S3.11}$$

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

The proof of (S3.11) is similar to that of (S3.2). For any $\epsilon>0$, set $A$ as in (S3.10). First, note that as in (S3.9),

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

when $n\to\infty$. Hence, for $n$ 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 in $1\le j\le l_n$. On the other hand, it follows from the monotonicity of $K(x)$ on $(-\infty,-A)$ and $(A,\infty)$ 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 in $1\le j\le l_n$. Using these facts, when $n$ is sufficiently large, we have

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

Now, for any $\delta>0$, let

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

Using the fact that $G_N(\mathbf{x})$ is uniformly bounded, we have

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

as $n\to\infty$ first, $N\to\infty$ second and then $\epsilon\to 0$. This proves (S3.11) and hence completes the proof of Lemma 1. □

## S3.2 Proof of Lemma 2

We first prove (35) and, without loss of generality, assume $K \geq 0$. Using similar arguments as in the proof of (S3.2) or (S3.11), it suffices to show that, as $n \to \infty$,

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

Set $\eta_{n,i,j} := \frac{1}{2}n(\tau_i+\tau_j)$. 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 $k \leq \eta_{n,i,j}$. In view of the fact that $K(x) \leq 1/|x|$ for $x$ sufficiently large[3], we have

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

as required.

The proof of (34) is similar to that of (35) and, hence, the details are omitted. The result of (36) follows easily from (34) and (35). Finally, (37) follows from similar arguments as those used in the proof of (S3.2) and the fact that as $n \to \infty$,

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

due to (35) and (36). □

[3]Since $\int K < \infty$ and $K \geq 0$ is eventually monotonic, we have that $K$ is decreasing on $(A_1, \infty)$ 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, $K(x) \leq 1/|x|$ when $x$ is sufficiently large.

## S4 Proofs of Theorem 5

We only consider $M_{1n,l_n}$, i.e., (26), since the limit result for $S^{(m)}_{1n,l_n}$ given in (25) follows easily from Lemma 1 with $G(x) \equiv I_{p+1}$ and $v_k = \mathbf{F}(\mathbf{x}_{k-1})\mathbf{F}(\mathbf{x}_{k-1})'\sigma_k^m$.

Set $Q_{k,n} := \sqrt{\frac{c_n}{n}}\alpha'\mathbf{F}(\mathbf{x}_{k-1})\sigma_k \frac{1}{\sqrt{l_n}} \sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right]$, where $\alpha \in \mathbb{R}^{p+1}$. Using (35) in Lemma 2 with $G(x) \equiv 1$ and $v_k \equiv [\alpha'\mathbf{F}(\mathbf{x}_{k-1})\sigma_k]^2$, we have

$$\begin{aligned}\sum_{k=1}^{n} Q_{k,n}^2 &= \frac{c_n}{n}\sum_{k=1}^{n}\left[\alpha'\mathbf{F}(\mathbf{x}_{k-1})\sigma_k\right]^2 \frac{1}{l_n}\sum_{j=1}^{l_n} K^2\left[c_n(k/n-\tau_j)\right] + o_P(1) \\ &= E\left[\alpha'\mathbf{F}(\mathbf{x}_1)\sigma_2\right]^2 \int K^2 + o_P(1), \qquad \text{(S4.1)}\end{aligned}$$

where the second equation follows from Lemma 1, with $K(x)$ replaced by $K^2(x)$, and $A_0 = E\left[\alpha'\mathbf{F}(\mathbf{x}_1)\sigma_2\right]^2$. In view of (S4.1), it follows from the classical martingale limit theorem (e.g., Hall and Heyde (1980), Theorem 3.2 or Wang (2014), Theorem 2.1) that to prove (26), it suffices to show that

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

Note that for any $A > 0$,

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

Similar arguments used in (S4.1) show that the first term above is

$$\begin{aligned}II_{1n}(A) &\le \|\alpha\|^2\frac{c_n}{n}\sum_{k=1}^{n}\|\mathbf{F}(\mathbf{x}_{k-1})\sigma_k\|^2 I\left\{\|\mathbf{F}(\mathbf{x}_{k-1})\sigma_k\| > A\right\}\frac{1}{l_n}\sum_{j=1}^{l_n}K^2\left[c_n(k/n-\tau_j)\right] + o_P(1) \\ &= \|\alpha\|^2 E\|\mathbf{F}(\mathbf{x}_1)\sigma_2\|^2 I\left\{\|\mathbf{F}(\mathbf{x}_1)\sigma_2\| > A\right\}\int K^2 + o_P(1) = o_P(1),\end{aligned}$$

where we take $n \to \infty$ first and then $A \to \infty$. On the other hand, using (37) in Lemma 2 with $G(x) \equiv 1$ and $v_k \equiv 1$, the second term

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

for each $A \ge 1$, as $n \to \infty$. Combining these facts, we establish (26). The proof of Theorem 5 is now complete. □

## S5 Proof of Theorem 6

We only consider $M_{2n,l_n}$, i.e., (28). The result for $\left[S_{2n,l_n}^{(m)}, S_{3n,l_n}^{(m)}\right]$ given in (27) follows directly from Lemma 1 with $G(X_{n,k}) \equiv \mathbf{F}(X_{n,k})\mathbf{F}(X_{n,k})'$ or $G(X_{n,k}) \equiv \mathbf{Q}(X_{n,k})$, and $v_k \equiv \sigma_k^m$, $m = 0$ or 2.

Set $Q_{k,n} := \sqrt{\frac{c_n}{n}}\alpha'\mathbf{F}(X_{n,k-1})\sigma_k \frac{1}{\sqrt{l_n}}\sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right]$, where $\alpha \in \mathbb{R}^{p+1}$. Noting that $\int_0^1 \mathbf{F}(X_{n,[nt]})dt$ is a continuous functional of $X_{n,[nt]}$, the limit result of (28), jointly with (27), will follow if we prove that, for any $\alpha \in \mathbb{R}^{p+1}$.

$$\left[X_{n,[nt]},\ \sum_{k=1}^{n} Q_{k,n}u_k\right] \Rightarrow \left[\mathcal{X}_t,\ \mathbf{MN}\left(0, E(\sigma_1^2)\int_0^1 \left[\alpha'\mathbf{F}(\mathcal{X}_t)\right]^2 dt \int K^2\right)\right] \tag{S5.1}$$

on $D_{\mathbb{R}^p\times\mathbb{R}}[0,1]$. First, note that by using (35) with $v_k \equiv \sigma_k^2$ and $G(.) \equiv \alpha'\mathbf{F}(.)$ first, and then (32),

$$\begin{aligned}\sum_{k=1}^{n} Q_{k,n}^2 &= \frac{c_n}{n}\sum_{k=1}^{n}\left[\alpha'\mathbf{F}(X_{n,k-1})\right]^2\sigma_k^2\frac{1}{l_n}\sum_{j=1}^{l_n}K^2\left[c_n(k/n-\tau_j)\right] + o_P(1)\\ &= E\left(\sigma_1^2\right)\int_0^1\left[\alpha'\mathbf{F}(X_{n,[nt]})\right]^2 dt\int K^2 + o_P(1).\end{aligned} \tag{S5.2}$$

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

$$\begin{aligned}&\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor nt\rfloor}\xi_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor nt\rfloor}\xi_{-k},\ X_{n,[nt]},\ \sum_{k=1}^{n}Q_{k,n}^2\right]\\ &\qquad\Rightarrow\left[B_{1t}, B_{2t}, \mathcal{X}_t, E\left(\sigma_1^2\right)\int_0^1\left[\alpha'\mathbf{F}(\mathcal{X}_t)\right]^2 dt\int K^2\right],\end{aligned}$$

on $D_{\mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}}[0,1]$. Recall **A1** and that $Q_{k,n}$ is a functional of $\xi_k, \xi_{k-1}, ....$ By Theorem 2.1 of Wang (2014) or Theorem 3.14 of Wang (2015), the limit result of (S5.1) will follow if we prove

$$\max_{1\le k\le n}|Q_{k,n}| = o_P(1), \tag{S5.3}$$

and

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}|Q_{k,n}| = o_P(1), \tag{S5.4}$$

which is what we set out to do next. In view of the continuity of $||\mathbf{F}(.)||^4$, it follows from (37) with $v_k \equiv \sigma_k^4$ that

$$\left[\max_{1\le k\le n}|Q_{k,n}|\right]^4 \le \sum_{k=1}^{n} Q_{k,n}^4$$

$$\le \|\alpha\|^4 \left(\frac{c_n}{n}\right)^2 \sum_{k=1}^{n} \|\mathbf{F}(X_{n,k-1})\|^4 \sigma_k^4 \left(\frac{1}{\sqrt{l_n}} \sum_{j=1}^{l_n} K\left[c_n(k/n-\tau_j)\right]\right)^4 = o_P(1),$$

yielding (S5.3). Similarly, using the fact that $l_n/c_n \to 0$ and (34) in Lemma 2, we have

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

which shows (S5.4). The proof of Theorem 6 is complete. □

## S6 Proof of Theorem 7

We start with the following lemma.

**Lemma S1.** *Suppose that:*

(*a*) *for each* $i = 1, ..., q$, $g_i : \mathbb{R} \to \mathbb{R}$ *is an AHF function with limit homogeneous function* $H_{g_i}$ *and asymptotic order* $\pi_{g_i}$;

(*b*) $\mathbf{x}_k = [x_{k,1}, ..., x_{k,q}]'$, $\mathcal{X}_t$ *is a* $\mathbb{R}^q$*-valued continuous process, and in* $D_{\mathbb{R}^q}[0,1]$ *there are deterministic sequences* $d_{in} \to \infty$, $i = 1, ..., q$ *such that* $X_{n,[nt]} \Rightarrow \mathcal{X}_t$, *where*

$$X_{n,k} = diag\{d_{1n}, ..., d_{qn}\}^{-1} \mathbf{x}_k;$$

(*c*) *either* $e_k^2 = \sigma_k^m$, *where* $m = 0, 2$, *and Assumptions* **A3**(*b*) *and* **A4**$^*$ *hold, or*

$e_k = \sigma_k u_k$ *with* $\sup_{k \ge 1} E u_k^4 < \infty$, *and Assumptions* **A1**(*c*), **A3**(*b*) *and* **A4**$^*$ *hold;*

(*d*) $h : \mathbb{R}^q \to \mathbb{R}$ *is a continuous function, and there exist* $c_0 > 0, \alpha > 0$ *and* $\nu \ge 0$ *so that for every* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^q$,

$$|h(\mathbf{x}+\mathbf{y}) - h(\mathbf{x})| \le c_0 \|\mathbf{y}\|^\alpha (1 + \|\mathbf{x}\| + \|\mathbf{y}\|)^\nu. \tag{S6.1}$$

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

$$\frac{c_n}{n l_n} \sum_{k=1}^{n} e_k^2 K_{kn} h(\widetilde{\mathbf{x}}_{n,k-1}) = \frac{c_n}{n l_n} \sum_{k=1}^{n} e_k^2 K_{kn} h(\widetilde{\mathbf{y}}_{n,k-1}) + o_P(1),$$

*where*

$$\widetilde{\mathbf{x}}_{nk} := \left[\frac{g_1(x_{k,1})}{\pi_{g_1}(d_{1n})}, \cdots, \frac{g_q(x_{k,q})}{\pi_{g_q}(d_{qn})}\right] \quad \textit{and} \quad \widetilde{\mathbf{y}}_{nk} := \left[H_{g_1}\left(\frac{x_{k,1}}{d_{1n}}\right), \cdots, H_{g_q}\left(\frac{x_{k,q}}{d_{qn}}\right)\right].$$

*Proof.* We only prove Lemma S1 with $e_k = \sigma_k u_k$. The proof for $e_k^2 = \sigma_k^m$ is similar but simpler. Let $\tilde{\mathbf{z}}_{nk} = \tilde{\mathbf{x}}_{nk} - \tilde{\mathbf{y}}_{nk}$. It follows from the definition AHF that

$$\|\tilde{\mathbf{z}}_{nk}\| \le a_n \sum_{i=1}^{q} (1 + |x_{k,i}/d_{in}|^{\delta_{g_i}}) \le 2a_n q(1 + \|X_{n,k}\|)^{\delta},$$

where $a_n = \max_{1\le j\le q} \frac{a_{g_j}(d_{jn})}{\pi_{g_j}(d_{jn})} \to 0$ and $\delta = \max_{1\le j\le q} g_j$. Observe that there is a continuous function $h_0 : \mathbb{R}^q \to \mathbb{R}^q$ such that $\tilde{\mathbf{y}}_{nk} = h_0(X_{n,k})$. Therefore, by the condition (d), we have

$$\frac{c_n}{nl_n} \sum_{k=1}^{n} e_k^2 K_{kn} |h(\tilde{\mathbf{x}}_{n,k-1}) - h(\tilde{\mathbf{y}}_{n,k-1})| = o_P(1) \cdot \frac{c_n}{nl_n} \sum_{k=1}^{n} e_k^2 K_{kn} h_1(X_{n,k-1}), \tag{S6.2}$$

where

$$h_1(\mathbf{x}) = (1 + \|\mathbf{x}\|)^{\alpha\delta} [(1 + \|\mathbf{x}\|)^{\delta} + \|h_0(\mathbf{x})\|]^{\nu}$$

is continuous. Recall that $E\left(u_k^2|\mathcal{F}_{k-1}\right) = 1$, and $\sigma_k$ is $\mathcal{F}_{k-1}$ measurable and stationary. It is readily seen from **A3**(b) and $\sup_{k\ge1} Eu_k^4 < \infty$ that

$$\begin{aligned}
\max_{m\le j\le n-m} E\left|\frac{1}{m}\sum_{k=j+1}^{j+m} e_k^2 - E\left(\sigma_1^2\right)\right| &= E\left|\frac{1}{m}\sum_{k=1}^{m} e_k^2 - E\left(\sigma_1^2\right)\right| \\
&\le E\left|\frac{1}{m}\sum_{k=1}^{m} \sigma_k^2 \left[u_k^2 - E\left(u_k^2|\mathcal{F}_{k-1}\right)\right]\right| + E\left|\frac{1}{m}\sum_{k=1}^{m} \left[\sigma_k^2 - E\left(\sigma_k^2\right)\right]\right| \to 0,
\end{aligned}$$

for any $0 < m = m_n \to \infty$. Hence, condition (b) of Lemma 1 is satisfied with $v_k \equiv e_k^2$ and $A_0 \equiv E\left(\sigma_1^2\right)$. The desired result follows from (S6.2) and (32) in Lemma 1. □

We now turn to the proof of Theorem 7. It suffices to prove the $o_P(1)$ approximations in (30)-(31). The weak convergence results in the aforementioned equations are a direct consequence of Theorem 6. Further, the proof of (29) is identical to that for (30).

The proof of the $o_P(1)$ approximation in (30) is simple. Indeed, by recalling that $K_{kn} = \sum_{j=1}^{l_n} K\left[c_n(k/n - \tau_j)\right]$ it follows from the condition (b) in Theorem 7 that

$$\begin{aligned}
&\frac{c_n}{nl_n} \sum_{k=1}^{n} \mathcal{L}_n^{-1}\mathbf{F}(\mathbf{x}_{k-1})\mathbf{F}(\mathbf{x}_{k-1})'\mathcal{L}_n^{-1}\, \sigma_k^m K_{kn} \\
&\qquad =: \frac{c_n}{nl_n} \sum_{k=1}^{n} H_{\mathbf{F}}(X_{n,k-1}) H_{\mathbf{F}}(X_{n,k-1})' \sigma_k^m K_{kn} + \Delta_{1n},
\end{aligned}$$

where $\Delta_{1n}$ is a $(p+1)\times(p+1)$ matrix that is determined by the definition AHF. It follows from

Lemma S1 with $e_k^2 \equiv \sigma_k^m$ that $\Delta_{1n} = o_P(1)$. Therefore, (30) follows directly from Theorem 6.

We next prove (31). We write

$$\sqrt{\frac{c_n}{nl_n}} \sum_{k=1}^{n} \mathcal{L}_n^{-1} \mathbf{F}(\mathbf{x}_{k-1}) \sigma_k u_k \; K_{kn} = \sqrt{\frac{c_n}{nl_n}} \sum_{k=1}^{n} H_{\mathbf{F}}(X_{n,k-1}) \sigma_k \, K_{kn} u_k + \Delta_{2n},$$

where $\Delta_{2n}$ is a $(p+1)$-dimensional vector. In particular, the first element of $\Delta_{2n}$ is zero, and the $j+1$ element

$$[\Delta_{2n}]_j = \sqrt{\frac{c_n}{nl_n}} \pi_{f_j}^{-1}(d_{jn}) \sum_{k=1}^{n} R_{f_j}(d_{jn}; x_{k-1,j}) \sigma_k \, K_{kn} \, u_k, \; , j = 1, ..., p,$$

with $R_{f_j}$ given in the definition AHF. For $A > 0$, set

$$[\Delta_{2n}]_j (A) = \sqrt{\frac{c_n}{nl_n}} \pi_{f_j}(d_{jn})^{-1} \sum_{k=1}^{n} R_{f_j}(d_{jn}; x_{k-1,j}) I \left\{ |x_{k-1,j}/d_{jn}| \le A \right\} \sigma_k \, K_{kn} \, u_k.$$

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

$$P\Big( [\Delta_{2n}]_j \neq [\Delta_{2n}]_j (A) \Big) \le P\left( \max_{1 \le k \le n} |x_{k-1,j}/d_{jn}| > A \right) \to 0. \tag{S6.3}$$

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

$$P\left( \left| [\Delta_{2n}]_j \right| \ge \epsilon \right) \le P\Big( [\Delta_{2n}]_j \neq [\Delta_{2n}]_j (A) \Big) + \epsilon^{-2} E\Big[ [\Delta_{2n}]_j (A) \Big]^2. \tag{S6.4}$$

Furthermore, for any $A > 0$, as $n \to \infty$ we have

$$\begin{aligned} E\Big[ [\Delta_{2n}]_j (A) \Big]^2 &= \frac{c_n}{nl_n} \pi_{f_j}^{-2}(d_{jn}) \sum_{k=1}^{n} E\Big( R_{f_j}(d_{jn}; x_{k-1,j})^2 I \left\{ |x_{k-1,j}/d_{jn}| \le A \right\} \sigma_k^2 \Big) K_{kn}^2 \\ &\le \left[ \frac{a_{f_j}(d_{jn})}{\pi_{f_j}(d_{jn})} \right]^2 P_{f_i}^2(A) \frac{c_n}{nl_n} \sum_{k=1}^{n} E(\sigma_k^2) \, K_{kn}^2 \to 0, \end{aligned} \tag{S6.5}$$

where $P_{f_i}$ is given in the definition AHF, and we have used (36) of Lemma 2 with $G(x) \equiv 1$ and $v_k \equiv E(\sigma_k^2)$. In view of (S6.3)-(S6.5), as $n \to \infty$

$$P\left( \left| [\Delta_{2n}]_j \right| \ge \epsilon \right) \to 0,$$

for all $j = 1, ..., p+1$. The proof of Theorem 7 is now complete. □

## S7 Proof of Theorem 1

The CTLS estimator for $\boldsymbol{\beta}$ is

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left[\sum_{k=1}^{n} \mathbf{Z}_{kn}\overline{\mathbf{f}}_{k-1}'\right]^{-1} \sum_{k=1}^{n} \mathbf{Z}_{kn}\bar{y}_k \\
&= \left[\sum_{k=1}^{n} \mathbf{Z}_{kn}\overline{\mathbf{f}}_{k-1}'\right]^{-1} \sum_{k=1}^{n} \mathbf{Z}_{kn}\left[\overline{\mathbf{f}}_{k-1}'\boldsymbol{\beta} + e_k - \frac{\sum_{s=1}^{n} e_s K_{sn}}{\sum_{s=1}^{n} K_{sn}}\right] \\
&= \boldsymbol{\beta} + \left[\sum_{k=1}^{n} \mathbf{Z}_{kn}\overline{\mathbf{f}}_{k-1}'\right]^{-1} \cdot \sum_{k=1}^{n} \mathbf{Z}_{kn}\left[e_k - \frac{\sum_{s=1}^{n} e_s K_{sn}}{\sum_{s=1}^{n} K_{sn}}\right],
\end{aligned}$$

which gives

$$\sqrt{\frac{nl_n}{c_n}}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \left[\frac{c_n}{nl_n}\sum_{k=1}^{n} \mathbf{Z}_{kn}\overline{\mathbf{f}}_{k-1}'\right]^{-1} \cdot \sqrt{\frac{c_n}{nl_n}} \sum_{k=1}^{n} \mathbf{Z}_{kn}\left[e_k - \frac{\sum_{s=1}^{n} e_s K_{sn}}{\sum_{s=1}^{n} K_{sn}}\right]. \quad \text{(S7.1)}$$

It follows from Theorem 5 that

$$\begin{aligned}
\frac{c_n}{nl_n}\sum_{k=1}^{n} \mathbf{Z}_{kn}\overline{\mathbf{f}}_{k-1}' &= \frac{c_n}{nl_n}\sum_{k=1}^{n} \mathbf{Z}_{kn}\mathbf{f}_{k-1}' - \frac{c_n}{nl_n}\sum_{k=1}^{n} \mathbf{Z}_{kn}\frac{\frac{c_n}{nl_n}\sum_{s=1}^{n} \mathbf{f}_{s-1}' K_{sn}}{\frac{c_n}{nl_n}\sum_{s=1}^{n} K_{sn}} \\
&= \frac{c_n}{nl_n}\sum_{k=1}^{n} \mathbf{f}_{k-1}\mathbf{f}_{k-1}' K_{kn} - \frac{c_n}{nl_n}\sum_{k=1}^{n} \mathbf{f}_{k-1} K_{kn}\frac{\frac{c_n}{nl_n}\sum_{s=1}^{n} \mathbf{f}_{s-1}' K_{sn}}{\frac{c_n}{nl_n}\sum_{s=1}^{n} K_{kn}} \\
&\to_P \int K \cdot \left[E\left\{\mathbf{f}(\mathbf{x}_1)\mathbf{f}(\mathbf{x}_1)'\right\} - E\left\{\mathbf{f}(\mathbf{x}_1)\right\} E\left\{\mathbf{f}(\mathbf{x}_1)'\right\}\right] \\
&= \int K \cdot \Phi_1. \qquad \text{(S7.2)}
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\sqrt{\frac{c_n}{nl_n}} \sum_{k=1}^{n} \mathbf{Z}_{kn}\left[e_k - \frac{\sum_{s=1}^{n} e_s K_{sn}}{\sum_{s=1}^{n} K_{sn}}\right] \\
&= \left[-\frac{1}{\frac{c_n}{nl_n}\sum_{k=1}^{n} K_{kn}}\frac{c_n}{nl_n}\sum_{k=1}^{n} \mathbf{Z}_{kn},\ I_p\right] \cdot \sqrt{\frac{c_n}{nl_n}} \sum_{k=1}^{n} K_{kn}\begin{bmatrix} 1 \\ \mathbf{f}_{k-1}\end{bmatrix} e_k \\
&=: \mathbf{A}_n \cdot \sqrt{\frac{c_n}{nl_n}} \sum_{k=1}^{n} K_{kn}\mathbf{F}_{k-1}e_k.
\end{aligned}$$

Again by Theorem 5 we get

$$\mathbf{A}_n \to_P \mathbf{A} := \left[-E\mathbf{f}(\mathbf{x}_1),\ I_p\right],$$

and

$$\sqrt{\frac{c_n}{nl_n}}\sum_{k=2}^{n} K_{kn}\mathbf{F}_{k-1}e_k \to_d \mathbf{N}\left(\mathbf{0}, \int K^2 \cdot E\left(\sigma_2^2\mathbf{F}(\mathbf{x}_1)\mathbf{F}(\mathbf{x}_1)'\right)\right).$$

In view of these facts, we have

$$\begin{aligned}\sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^{n}\mathbf{Z}_{kn}\left[e_k - \frac{\sum_{s=1}^n e_s K_{sn}}{\sum_{s=1}^n K_{sn}}\right] &= \mathbf{A}_n \cdot \sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^{n} K_{kn}\mathbf{F}_{k-1}e_k \\ &\to_d \mathbf{N}\left(0, \int K^2 \cdot \mathbf{A}E(\sigma_2^2\mathbf{F}(\mathbf{x}_1)\mathbf{F}(\mathbf{x}_1)')\mathbf{A}'\right) =_d \mathbf{N}\left(0, \int K^2 \cdot \Phi_0\right), \qquad \text{(S7.3)}\end{aligned}$$

where we have used the fact that $\mathbf{A}E(\sigma_2^2\mathbf{F}(\mathbf{x}_1)\mathbf{F}(\mathbf{x}_1)')\mathbf{A}' = \Phi_0$. The desired result follows by combining (S7.1), (S7.2) and (S7.3). □

## S8 Proof of Theorem 2

Similarly to (S7.1), we may write

$$\sqrt{\frac{nl_n}{c_n}}\mathcal{D}_n\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \left[\frac{c_n}{nl_n}\mathcal{D}_n^{-1}\sum_{k=1}^{n}\mathbf{Z}_{kn}\bar{\mathbf{f}}_{k-1}'\mathcal{D}_n^{-1}\right]^{-1}\mathcal{D}_n^{-1}\sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^{n}\mathbf{Z}_{kn}\left[e_k - \frac{\sum_{s=1}^n e_s K_{sn}}{\sum_{s=1}^n K_{sn}}\right]. \qquad \text{(S8.1)}$$

It follows from Theorem 7 that

$$\begin{aligned}&\frac{c_n}{nl_n}\mathcal{D}_n^{-1}\sum_{k=1}^{n}\mathbf{Z}_{kn}\bar{\mathbf{f}}_{k-1}'\mathcal{D}_n^{-1} \\ &= \frac{c_n}{nl_n}\mathcal{D}_n^{-1}\sum_{k=1}^{n}\mathbf{Z}_{kn}\mathbf{f}_{k-1}'\mathcal{D}_n^{-1} - \frac{c_n}{nl_n}\mathcal{D}_n^{-1}\sum_{k=1}^{n}\mathbf{Z}_{kn}\frac{\frac{c_n}{nl_n}\sum_{s=1}^n \mathbf{f}_{s-1}'K_{sn}}{\frac{c_n}{nl_n}\sum_{s=1}^n K_{sn}}\mathcal{D}_n^{-1} \\ &= \frac{c_n}{nl_n}\mathcal{D}_n^{-1}\sum_{k=1}^{n}\mathbf{f}_{k-1}\mathbf{f}_{k-1}'K_{kn}\mathcal{D}_n^{-1} - \frac{c_n}{nl_n}\mathcal{D}_n^{-1}\sum_{k=1}^{n}\mathbf{f}_{k-1}K_{kn}\frac{\frac{c_n}{nl_n}\sum_{s=1}^n \mathbf{f}_{s-1}'K_{sn}}{\frac{c_n}{nl_n}\sum_{s=1}^n K_{kn}}\mathcal{D}_n^{-1} \\ &\to_d \int K\left[\int_0^1 H_{\mathbf{f}}(\mathcal{X}_t)H_{\mathbf{f}}(\mathcal{X}_t)'dt - \int_0^1 H_{\mathbf{f}}(\mathcal{X}_t)dt\int_0^1 H_{\mathbf{f}}(\mathcal{X}_t)'dt\right] \\ &= \int K \cdot \int_0^1 \tilde{H}_{\mathbf{f}}(\mathcal{X}_t)\tilde{H}_{\mathbf{f}}(\mathcal{X}_t)'dt = \int K \cdot \Phi_2. \qquad \text{(S8.2)}\end{aligned}$$

Further, by letting $\mathcal{L}_n = diag\{1, \mathcal{D}_n\}$, we have

$$\begin{aligned}&\mathcal{D}_n^{-1}\sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^{n}\mathbf{Z}_{kn}\left[e_k - \frac{\sum_{s=1}^n e_s K_{sn}}{\sum_{s=1}^n K_{sn}}\right] \\ &= \left[-\frac{1}{\frac{c_n}{nl_n}\sum_{k=1}^n K_{kn}}\mathcal{D}_n^{-1}\frac{c_n}{nl_n}\sum_{k=1}^{n}\mathbf{Z}_{kn},\ I_p\right]\sqrt{\frac{c_n}{nl_n}}\mathcal{L}_n^{-1}\sum_{k=1}^{n}K_{kn}\begin{bmatrix}1\\ \mathbf{f}_{k-1}\end{bmatrix}e_k\end{aligned}$$

$$=: \mathbf{B}_n\sqrt{\frac{c_n}{nl_n}}\mathcal{L}_n^{-1}\sum_{k=1}^{n}K_{kn}\mathbf{F}_{k-1}e_k.$$

Again by Theorem 7 we get jointly with (S8.2)

$$\mathbf{B}_n \to_d \mathbf{B} := \left[-\int_0^1 H_{\mathbf{f}}(\mathcal{X}_t)dt,\ I_p\right], \tag{S8.3}$$

and

$$\sqrt{\frac{c_n}{nl_n}}\mathcal{L}_n^{-1}\sum_{k=1}^{n}K_{kn}\mathbf{F}_{k-1}e_k \to_d \mathbf{MN}\left(\mathbf{0}, \int K^2\cdot E(\sigma_1^2)\int_0^1\left[\begin{array}{cc}1 & H_{\mathbf{f}}(\mathcal{X}_t)'\\ H_{\mathbf{f}}(\mathcal{X}_t) & H_{\mathbf{f}}(\mathcal{X}_t)H_{\mathbf{f}}(\mathcal{X}_t)'\end{array}\right]dt\right).$$

Therefore,

$$\begin{aligned}\mathcal{D}_n^{-1}\sqrt{\frac{c_n}{nl_n}}\sum_{k=1}^{n}\mathbf{Z}_{kn}\left[e_k-\frac{\sum_{s=1}^{n}e_sK_{sn}}{\sum_{s=1}^{n}K_{sn}}\right] &= \mathbf{B}_n\sqrt{\frac{c_n}{nl_n}}\mathcal{L}_n^{-1}\sum_{k=1}^{n}K_{kn}\mathbf{F}_{k-1}e_k\\ &\to_d \mathbf{MN}\left(\mathbf{0}, \int K^2\cdot E(\sigma_1^2)\mathbf{B}\int_0^1\left[\begin{array}{cc}1 & H_{\mathbf{f}}(\mathcal{X}_t)'\\ H_{\mathbf{f}}(\mathcal{X}_t) & H_{\mathbf{f}}(\mathcal{X}_t)H_{\mathbf{f}}(\mathcal{X}_t)'\end{array}\right]dt\mathbf{B}'\right)\\ &=_d \mathbf{MN}\left(\mathbf{0}, \int K^2\cdot E(\sigma_1^2)\Phi_2\right).\end{aligned} \tag{S8.4}$$

In view of (S8.1), (S8.2) and (S8.4) the CTLS estimator for $\boldsymbol{\beta}$

$$\sqrt{\frac{nl_n}{c_n}}\mathcal{D}_n\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \to_d \mathbf{MN}\left(\mathbf{0}, \frac{\int K^2}{\left(\int K\right)^2}\cdot E(\sigma_1^2)\Phi_2^{-1}\right),$$

as required. □

## S9 Proof of Theorem 3

See Theorem S1 and hence the details are omitted. □

## S10 Proof of Theorem 4 and additional explanation for (15)

We start with some preliminary results. We first derive the pseudo-true limits of the OLS and CTLS estimators under misbalancing (MB) assuming that the conditions of Theorem 4 hold. Recall that $d_n$ denotes the normalizing sequence of Assumption A3(a) for the case $p=1$. Set $f_k := f(x_{k-1})$, $f_{M,k} := f_M(x_{k-1})$, and $Q_n := diag\{\sqrt{n}, \sqrt{n}\pi_{f_M}(d_n)\}$. Furthermore, for a sequence $a_k$ let $\tilde{a}_k := a_k - n^{-1}\sum_{j=1}^{n}a_j$. Similarly, for a function $A(t)$, $\tilde{A}(t) := A(t) - \int_0^1 A(s)ds$.

**OLS under MB**. Let $\left[\tilde{\mu}, \tilde{\beta}\right]$ be the OLS estimator for the parameters of (10) when the fitted model is given by (14). Suppose that the conditions of Theorem 4 hold. Then by Assumption **A3**(a), there is a sequence $d_n$ such that $d_n^{-1}x_{[nt]} \Rightarrow \mathcal{X}_t$ in $D[0,1]$. In view of this and the continuity of the limit homogeneous functions of $f$ and $f_M$, standard arguments (e.g. Park and Phillips, 2001; Theorem 5.2) give

$$
\begin{aligned}
&\frac{1}{\pi_f(d_n)\sqrt{n}} Q_n \begin{bmatrix} \tilde{\mu} \\ \tilde{\beta} \end{bmatrix} \\
&= \left[ Q_n^{-1} \sum_k \begin{bmatrix} 1 \\ f_{M,k} \end{bmatrix} \begin{bmatrix} 1 \\ f_{M,k} \end{bmatrix}' Q_n^{-1} \right]^{-1} \frac{1}{\pi_f(d_n)\sqrt{n}} Q_n^{-1} \sum_k \begin{bmatrix} 1 \\ f_{M,k} \end{bmatrix} \beta f_k + o_P(1) \\
&\to_d \left\{ \int_0^1 \begin{bmatrix} 1 & H_{f_M}(\mathcal{X}_t) \\ H_{f_M}(\mathcal{X}_t) & H_{f_M}^2(\mathcal{X}_t) \end{bmatrix} dt \right\}^{-1} \beta \int_0^1 \begin{bmatrix} H_f(\mathcal{X}_t) \\ H_{f_M}(\mathcal{X}_t) H_f(\mathcal{X}_t) \end{bmatrix} dt \\
&= \left[ \int_0^1 \tilde{H}_{f_M} H_{f_M} \right]^{-1} \begin{bmatrix} \int_0^1 H_f \int_0^1 H_{f_M}^2 - \int_0^1 H_{f_M} \int_0^1 H_{f_M} H_f \\ \int_0^1 \tilde{H}_{f_M} H_f \end{bmatrix} \beta =: \begin{bmatrix} \mu_* \\ \beta_* \end{bmatrix}.
\end{aligned}
$$

In fact, the following joint weak limit holds

$$
\frac{1}{\pi_f(d_n)} \tilde{\mu} \to_d \mu_* \quad \text{and} \quad \frac{\pi_{f_M}(d_n)}{\pi_f(d_n)} \tilde{\beta} \to_d \beta_*.
$$

**CTLS under MB**. Similarly, consider the CTLS estimator (in this case, the CTLS instruments are: $Z_{kn} = f_{M,k} K_{kn}$). Using similar arguments as those used above together with Theorem 7 we get

$$
\begin{aligned}
\frac{\pi_{f_M}(d_n)}{\pi_f(d_n)} \hat{\beta} &= \frac{\frac{1}{\pi_{f_M}(d_n)\pi_f(d_n)}}{\frac{1}{\pi_{f_M}^2(d_n)}} \hat{\beta} = \beta \frac{\frac{c_n}{n l_n \pi_{f_M}(d_n)\pi_f(d_n)} \sum_k Z_{kn} \bar{f}_k}{\frac{c_n}{n l_n \pi_{f_M}^2(d_n)} \sum_k Z_{kn} \bar{f}_{M,k}} + o_P(1) \\
&\to_d \beta \frac{\int_0^1 \tilde{H}_f(\mathcal{X}_t) H_{f_M}(\mathcal{X}_t) dt}{\int_0^1 \tilde{H}_{f_M}(\mathcal{X}_t) H_{f_M}(\mathcal{X}_t) dt} = \beta_*.
\end{aligned}
$$

**OLS estimator for $\sigma^2$ under MB**. Next, we consider the variance estimator for $\sigma^2$ based on the OLS residuals $\check{e}_k$. Write

$$
\begin{aligned}
\pi_f^{-2}(d_n)\check{\sigma}^2 &= \pi_f^{-2}(d_n) n^{-1} \sum_k \check{e}_k^2 = \pi_f^{-2}(d_n) n^{-1} \sum_k \left( f_k - \tilde{\mu} - \tilde{\beta} f_{M,k} \right)^2 + o_P(1) \\
&= n^{-1} \sum_k \left( \pi_f^{-1}(d_n) f_k - \pi_f^{-1}(d_n)\tilde{\mu} - \frac{\pi_{f_M}(d_n)}{\pi_f(d_n)} \tilde{\beta} \pi_{f_M}^{-1}(d_n) f_{M,k} \right)^2 + o_P(1) \\
&\to_d \int_0^1 [H_f - \mu_* - \beta_* H_{f_M}]^2 =: \sigma_*^2. \qquad \text{(S10.1)}
\end{aligned}
$$

Hence, we roughly have the following approximation

$$
\check{\sigma}^2 \approx \pi_f^2(d_n)\sigma_*^2 + \sigma^2. \qquad \text{(S10.2)}
$$

**CTLS t-statistics.** Next, we introduce some additional notation. For any asymptotic homogeneous function $g:\mathbb{R}\to\mathbb{R}$ (and some kernel function $K$) set

$$\mathbf{A}_g := \left[-\int_0^1 H_g(\mathcal{X}_t)dt, 1\right], \ V_g := \left[\begin{array}{cc} 1 & \int_0^1 H_g(\mathcal{X}_t)dt \\ \int_0^1 H_g(\mathcal{X}_t)dt & \int_0^1 H_g^2(\mathcal{X}_t)dt \end{array}\right] \int K^2.$$

Note that for some AHF function $f$, the definitions above and straightforward calculations yield

$$\mathbf{A}_f V_f \mathbf{A}_f' = \int K^2 \int_0^1 \tilde{H}_f H_f. \tag{S10.3}$$

Furthermore, for any function $g:\mathbb{R}\to\mathbb{R}$ set $g_k = g(x_{k-1})$ and define

$$\mathcal{A}_{g,n} := [-\bar{g}, 1]\,, \ \mathcal{V}_{g,n} := \sum_{k=1}^{n} K_{kn}^2 \left[\begin{array}{cc} 1 & g_k \\ g_k & g_k^2 \end{array}\right], \ \check{\sigma}^2 = n^{-1}\sum_{k=1}^{n} \check{e}_k^2,$$

where $\bar{g} = \sum_{k=1}^n K_{kn} g(x_{k-1}) / \sum_{k=1}^n K_{kn}$ (cf. (7)), and $\check{e}_k$ are the OLS residuals based on the fitted regression function $g(x_{k-1})$. In the following, we assume $g = f$ (correct functional form) and $g = f_M$ (misbalanced model).

Without loss of generality, we shall consider t-statistics that utilize the studentization of (13). Hence, under correct functional form the CTLS t-statistic is

$$\hat{T} = \sum_{k=1}^{n} K_{kn} \bar{f}_k \frac{\hat{\beta} - \beta_0}{\sqrt{\check{\sigma}^2 \mathcal{A}_{f,n} \mathcal{V}_{f,n} \mathcal{A}_{f,n}'}},$$

where $\hat{\beta}$ is the CTLS estimator of $\beta$ in (10). Under misbalancing, the CTLS t-statistic is of the form

$$\hat{T}_M = \sum_{k=1}^{n} K_{kn} \bar{f}_{M,k} \frac{\hat{\beta} - \beta_0}{\sqrt{\check{\sigma}^2 \mathcal{A}_{f_M,n} \mathcal{V}_{f_M,n} \mathcal{A}_{f_M,n}'}},$$

with $\hat{\beta}$ being the CTLS estimator based on the fitted model (14).

We now turn to the proof of Theorem 4. We start with (15). Using Theorem 7 the CTLS t-statistic under misbalancing is

$$\begin{aligned}
\sqrt{\frac{c_n}{n l_n}} \hat{T}_M &:= \sqrt{\frac{c_n}{n l_n}} \sum_{k=1}^{n} K_{kn} \bar{f}_{M,k} \frac{\hat{\beta}}{\sqrt{\check{\sigma}^2 \mathcal{A}_{f_M,n} \mathcal{V}_{f_M,n} \mathcal{A}_{f_M,n}'}} \\
&= \sqrt{\frac{c_n}{n l_n}} \left[ \frac{\beta \sum_k Z_{kn} \bar{f}_k}{\sqrt{\check{\sigma}^2 \mathcal{A}_{f_M,n} \mathcal{V}_{f_M,n} \mathcal{A}_{f_M,n}'}} + \frac{\sum_k Z_{kn} \bar{u}_k}{\sqrt{\check{\sigma}^2 \mathcal{A}_{f_M,n} \mathcal{V}_{f_M,n} \mathcal{A}_{f_M,n}'}} \right] \\
&= \frac{\frac{c_n}{n l_n \pi_{f_M}(d_n) \pi_f(d_n)} \beta \sum_k Z_{kn} \bar{f}_k}{\sqrt{\pi_f^{-2}\ \check{\sigma}^2 \frac{c_n}{n l_n \pi_{f_M}^2(d_n)} \mathcal{A}_{f_M,n} \mathcal{V}_{f_M,n} \mathcal{A}_{f_M,n}'}}
\end{aligned}$$

$$+\sqrt{\frac{c_n}{nl_n}}\frac{1}{\pi_f(d_n)}\frac{\sqrt{\frac{c_n}{nl_n\pi^2_{f_M}(d_n)}}\sum_k Z_{kn}\bar{u}_k}{\sqrt{\pi_f^{-2}\breve{\sigma}^2\frac{c_n}{nl_n\pi^2_{f_M}(d_n)}\mathcal{A}_{f_M,n}\mathcal{V}_{f_M,n}\mathcal{A}'_{f_M,n}}}$$

$$\to_d \quad \frac{\beta\int K\cdot\int_0^1\tilde{H}_f H_{f_M}}{\sqrt{\sigma_*^2\mathbf{A}_{f_M}V_{f_M}\mathbf{A}'_{f_M}}}+O_P\left(\sqrt{\frac{c_n}{nl_n}}\frac{1}{\pi_f(d_n)}\right), \tag{S10.4}$$

where $\sigma_*^2$ is defined in (S10.1). (15) follows directly from (S10.4).

Next, we show (16). Recall that the divergence rate under the correct specification is $\pi_f(d_n)\sqrt{\frac{nl_n}{c_n}}$. In fact, under correct FF and under $H_1$ we have

$$\begin{aligned}
\sqrt{\frac{c_n}{\pi_f^2(d_n)nl_n}}\hat{T} \quad &:= \quad \sqrt{\frac{c_n}{\pi_f^2(d_n)nl_n}}\sum_{k=1}^n K_{kn}\bar{f}_k\frac{\hat{\beta}}{\sqrt{\breve{\sigma}^2\mathcal{A}_{f,n}\mathcal{V}_{f,n}\mathcal{A}'_{f,n}}}\\
&= \quad \sqrt{\frac{c_n}{\pi_f^2(d_n)nl_n}}\sum_{k=1}^n K_{kn}\bar{f}_k\frac{\hat{\beta}-\beta}{\sqrt{\breve{\sigma}^2\mathcal{A}_{f,n}\mathcal{V}_{f,n}\mathcal{A}'_{f,n}}}\\
&\qquad +\frac{c_n}{nl_n\pi_f^2(d_n)}\sum_k Z_{kn}\bar{f}_k\frac{\beta}{\sqrt{\breve{\sigma}^2\frac{c_n}{nl_n\pi_f^2}\mathcal{A}_{f,n}\mathcal{V}_{f,n}\mathcal{A}'_{f,n}}}\\
&\to_d \quad O_P\left(\sqrt{\frac{c_n}{\pi_f^2(d_n)nl_n}}\right)+\frac{\beta\int K\cdot\int_0^1\tilde{H}_f(\mathcal{X}_t)H_f(\mathcal{X}_t)dt}{\sqrt{\sigma^2\mathbf{A}_fV_f\mathbf{A}'_f}}.
\end{aligned}\tag{S10.5}$$

Result (16) follows directly from (S10.3), (S10.4) and (S10.5). This completes the proof of Theorem 4. □

**We finally consider supporting arguments for (17)**. First, note that (S10.2) together with (S10.4) postulate the following approximate behavior.

$$\hat{T}_M\approx\pi_f(d_n)\sqrt{\frac{nl_n}{c_n}}\frac{\beta\int K\cdot\int_0^1\tilde{H}_fH_{f_M}}{\sqrt{(\pi_f^2(d_n)\sigma_*^2+\sigma^2)\mathbf{A}_{f_M}V_{f_M}\mathbf{A}'_{f_M}}}. \tag{S10.6}$$

Furthermore, by (S10.5) we have.

$$\hat{T}\approx\pi_f(d_n)\sqrt{\frac{nl_n}{c_n}}\frac{\beta\int K\cdot\int_0^1\tilde{H}_fH_f}{\sqrt{\sigma^2\mathbf{A}_fV_f\mathbf{A}'_f}}. \tag{S10.7}$$

Combining (S10.5) and (S10.6), the ratio of the two test statistics is

$$\hat{T}/\hat{T}_M \quad\approx\quad \frac{\sqrt{\left(\pi_f^2(d_n)\sigma_*^2+\sigma^2\right)\mathbf{A}_{f_M}V_{f_M}\mathbf{A}'_{f_M}}\int_0^1\tilde{H}_fH_f}{\sqrt{\sigma^2\mathbf{A}_fV_f\mathbf{A}'_f}\int_0^1\tilde{H}_{f_M}H_f}$$

$$= \sqrt{\frac{\pi_f^2(d_n)\sigma_*^2 + \sigma^2}{\sigma^2}} \frac{\sqrt{\int_0^1 \tilde{H}_{f_M} H_{f_M} \int_0^1 \tilde{H}_f H_f}}{\sqrt{\int_0^1 \tilde{H}_f H_f \int_0^1 \tilde{H}_{f_M} H_{f_M}}}$$

$$= \sqrt{\frac{\pi_f^2(d_n)\sigma_*^2 + \sigma^2}{\sigma^2}} \frac{\sqrt{\int_0^1 \tilde{H}_{f_M} H_{f_M} \int_0^1 \tilde{H}_f H_f}}{\int_0^1 \tilde{H}_{f_M} H_f} \geq \sqrt{\frac{\pi_f^2(d_n)\sigma_*^2 + \sigma^2}{\sigma^2}},$$

where the lower bound above is due to the Cauchy-Schwarz inequality.

## S11 Proof of Theorem S1

We only prove (S2.6), as the proof of (S2.5) is similar.

We first assume that the conditions of Theorem 2 hold, together with $\sup_{k\geq 1} Eu_k^4 < \infty$. Define $\mathcal{L}_n := diag\left\{1, \pi_{f_1}(d_{1n}), ..., \pi_{f_p}(d_{pn})\right\}$ and $\mathcal{V}_n = \sum_{k=1}^n e_k^2 \mathbf{F}_{k-1}\mathbf{F}_{k-1}' K_{kn}^2$. Since the the OLS residuals in model (S2.1) satisfy

$$\check{e}_k^2 = \left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{LS}\right)' \mathbf{F}_{k-1}\right]^2 + 2\left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{LS}\right)' \mathbf{F}_{k-1}\right] e_k + e_k^2,$$

it follows that

$$\hat{\mathcal{V}}_n = \sum_{k=1}^n \check{e}_k^2 \mathbf{F}_{k-1}\mathbf{F}_{k-1}' K_{kn}^2 = \mathcal{V}_n + R_{1n} + 2R_{2n} \tag{S11.1}$$

where

$$R_{1n} = \sum_{k=1}^n \left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{LS}\right)' \mathbf{F}_{k-1}\right]^2 \mathbf{F}_{k-1}\mathbf{F}_{k-1}' K_{kn}^2,$$

$$R_{2n} = \sum_{k=1}^n e_k \left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{LS}\right)' \mathbf{F}_{k-1}\right] \mathbf{F}_{k-1}\mathbf{F}_{k-1}' K_{kn}^2.$$

We next show that, for $j = 1$ and 2,

$$\left\| \frac{c_n}{nl_n} \mathcal{L}_n^{-1} R_{jn} \mathcal{L}_n^{-1} \right\| = o_P(1). \tag{S11.2}$$

In fact, given that the covariates satisfy an FCLT and the regression function is continuous, standard arguments (see, e.g., Park and Phillips, 2001; Chang, Park and Phillips, 2001) give

$$\sqrt{n}\mathcal{L}_n\left(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}\right) = O_P(1). \tag{S11.3}$$

This result implies that

$$\left\| \frac{c_n}{nl_n} \mathcal{L}_n^{-1} R_{1n} \mathcal{L}_n^{-1} \right\| = \left\| \frac{c_n}{nl_n} \sum_{k=1}^n \left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{LS}\right)' \mathcal{L}_n \mathcal{L}_n^{-1} \mathbf{F}_{k-1}\right]^2 \mathcal{L}_n^{-1} \mathbf{F}_{k-1}\mathbf{F}_{k-1}' \mathcal{L}_n^{-1} K_{kn}^2 \right\|$$

$$
\begin{aligned}
&\le n^{-1}\left\|\sqrt{n}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{LS}\right)'\mathcal{L}_n\right\|^2\cdot\frac{c_n}{nl_n}\sum_{k=1}^{n}\left\|\mathcal{L}_n^{-1}\mathbf{F}_{k-1}\right\|^4K_{kn}^2\\
&= O_P\left(n^{-1}\right)\cdot\frac{c_n}{nl_n}\sum_{k=1}^{n}\left\|\mathcal{L}_n^{-1}\mathbf{F}_{k-1}\right\|^4K_{kn}^2=O_P(n^{-1}),
\end{aligned}
$$

where we have used Lemma S1 with $h(x)=||x||^4$ and the similar argument as in the proof of Theorem 6 (c.g. the proof of (S5.2)), yielding

$$
\frac{c_n}{nl_n}\sum_{k=1}^{n}\left\|\mathcal{L}_n^{-1}\mathbf{F}_{k-1}\right\|^4K_{kn}^2\to_d\int K^2\cdot\int_0^1\left\|H_{\mathbf{F}}\left(\mathcal{X}_t\right)\right\|^4dt.
$$

Hence (S11.2) is true with $j=1$. Similarly, using (S11.3) again, we have

$$
\begin{aligned}
\left\|\frac{c_n}{nl_n}\mathcal{L}_n^{-1}R_{2n}\mathcal{L}_n^{-1}\right\| &= \frac{2c_n}{nl_n}\left\|\sum_{k=1}^{n}e_k\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{LS}\right)'\mathbf{F}_{k-1}\right]\mathcal{L}_n^{-1}\mathbf{F}_{k-1}\mathbf{F}_{k-1}'\mathcal{L}_n^{-1}K_{kn}^2\right\|\\
&\le 2n^{-1/2}\left\|\sqrt{n}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{LS}\right)'\mathcal{L}_n\right\|\cdot\frac{c_n}{nl_n}\sum_{k=1}^{n}|e_k|\left\|\mathcal{L}_n^{-1}\mathbf{F}_{k-1}\right\|^3K_{kn}^2\\
&\le O_P\left(n^{-1/2}\right)\cdot\frac{c_n}{nl_n}\sum_{k=1}^{n}\left(e_k^2+1\right)\left\|\mathcal{L}_n^{-1}\mathbf{F}_{k-1}\right\|^3K_{kn}^2\\
&= O_P\left(n^{-1/2}\right)\cdot O_P\left(1\right)=o_P(1),
\end{aligned}
$$

i.e., (S11.2) is also true with $j=2$.

In terms of (S11.1) and (S11.2), we claim that, as $n\to\infty$,

$$
\frac{c_n}{nl_n}\mathcal{L}_n^{-1}\hat{\mathcal{V}}_n\mathcal{L}_n^{-1} = \frac{c_n}{nl_n}\mathcal{L}_n^{-1}\mathcal{V}_n\mathcal{L}_n^{-1}+o_P(1). \tag{S11.4}
$$

Now (S2.6) under the conditions of Theorem 2 is a direct consequence of Theorem 2 and (S11.4). To see this, set $\mathcal{D}_n^*:=\sqrt{\frac{nl_n}{c_n}}\mathcal{D}_n$, $\mathcal{L}_n^*:=\sqrt{\frac{nl_n}{c_n}}\mathcal{L}_n$, with $\mathcal{D}_n$ defined in Theorem 2 and recall $\mathcal{L}_n=diag\{1,\mathcal{D}_n\}$. Under the null hypothesis,

$$
\begin{aligned}
\hat{F} &= \left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\left[\mathcal{H}_n^{-1}\mathcal{A}_n\hat{\mathcal{V}}_n\mathcal{A}_n'\mathcal{H}_n^{-1}\right]^{-1}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\\
&= \left[\mathcal{D}_n^*\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right]'\mathcal{M}_n^{-1}\left[\mathcal{D}_n^*\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right]
\end{aligned}
$$

with

$$
\mathcal{M}_n:=\left(\mathcal{D}_n^{*-1}\mathcal{H}_n\mathcal{D}_n^{*-1}\right)^{-1}\cdot\left(\mathcal{D}_n^{*-1}\mathcal{A}_n\mathcal{L}_n^*\right)\cdot\mathcal{L}_n^{*-1}\hat{\mathcal{V}}_n\mathcal{L}_n^{*-1}\cdot\mathcal{L}_n^*\mathcal{A}_n'\mathcal{D}_n^{*-1}\cdot\left(\mathcal{D}_n^{*-1}\mathcal{H}_n\mathcal{D}_n^{*-1}\right)^{-1}.
$$

Next, note that by (S11.4) and Theorem 2

$$
\mathcal{L}_n^{*-1}\hat{\mathcal{V}}_n\mathcal{L}_n^{*-1}=\mathcal{L}_n^{*-1}\mathcal{V}_n\mathcal{L}_n^{*-1}+o_P(1)\to_d E(\sigma_1^2)\int K^2\cdot\int_0^1H_{\mathbf{F}}\left(X_t\right)H_{\mathbf{F}}\left(X_t\right)'dt. \tag{S11.5}
$$

Further, by (S8.2)

$$\mathcal{D}_n^{*-1}\mathcal{H}_n\mathcal{D}_n^{*-1} \to_d \int K \cdot \Phi_2, \tag{S11.6}$$

with $\Phi_2$ defined in Theorem 2. Moreover, using (S8.3)

$$\mathcal{D}_n^{*-1}\mathcal{A}_n\mathcal{L}_n^* = \mathbf{B}_n \to_d \mathbf{B}, \tag{S11.7}$$

with $\mathbf{B}$ defined in (S8.3). Combining (S11.5)-(S11.7),

$$\mathcal{M}_n \to_d E(\sigma_1^2)\frac{\int K^2}{\left(\int K\right)^2}\Phi_2^{-1}\mathbf{B}\int_0^1 H_{\mathbf{F}}(X_t)\, H_{\mathbf{F}}(X_t)'\, dt\, \mathbf{B}'\Phi_2^{-1} = E(\sigma_1^2)\frac{\int K^2}{\left(\int K\right)^2}\Phi_2^{-1}, \tag{S11.8}$$

where we have used the fact that

$$\Phi_2 = \mathbf{B}\int_0^1 H_{\mathbf{F}}(\mathcal{X}_t)\, H_{\mathbf{F}}(\mathcal{X}_t)'\, dt\, \mathbf{B}',$$

see e.g. (S8.4). In view of (S11.8) and Theorem 2,

$$\mathcal{M}_n^{-1/2}\mathcal{D}_n^*\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \to_d \mathbf{N}\left(\mathbf{0}, I_q\right),$$

and the result follows.

We next prove (S2.6) under the conditions of Theorem 1, together with $\sup_{k\geq 1} Eu_k^4 < \infty$. By using Theorem 1 and similar arguments as in the first part, it suffices to show that

$$\frac{c_n}{nl_n}\hat{\mathcal{V}}_n \quad \to_P \quad \int K^2 \cdot E\left[\sigma_2^2\mathbf{F}(\mathbf{x}_1)\,\mathbf{F}(\mathbf{x}_1)'\right]. \tag{S11.9}$$

Note that (S11.1) still holds. (S11.9) will follow if we prove

$$\left\|\frac{c_n}{nl_n}R_{jn}\right\| \quad = \quad o_P(1), \quad j = 1, 2, \tag{S11.10}$$

and

$$\frac{c_n}{nl_n}\mathcal{V}_n \to_P \int K^2 \cdot E\left[\sigma_2^2\mathbf{F}(\mathbf{x}_1)\mathbf{F}(\mathbf{x}_1)'\right]. \tag{S11.11}$$

The proof of (S11.10) is simple. In fact, by recalling **A2** (i.e., $\mathbf{x}_k$ and $\mathbf{F}_k = \mathbf{F}(\mathbf{x}_{k-1})$ are stationary with $E\left\|\mathbf{F}_1\right\|^2 < \infty$), the OLS estimator in this case satisfies

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta}\right) = O_P(1) \tag{S11.12}$$

and $\max_{1\le k\le n}\left\|n^{-1/2}\mathbf{F}_{k-1}\right\| = o_P(1)$. It follows from these facts that

$$\begin{aligned}
\left\|\frac{c_n}{nl_n}R_{1n}\right\| &= \left\|\frac{c_n}{nl_n}\sum_{k=1}^{n}\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{LS}\right)'\mathbf{F}_{k-1}\right]^2\mathbf{F}_{k-1}\mathbf{F}_{k-1}'K_{kn}^2\right\| \\
&\le n^{-1}\left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{LS}-\boldsymbol{\theta}\right)\right\|^2\frac{c_n}{nl_n}\sum_{k=1}^{n}\left\|\mathbf{F}_{k-1}\right\|^4K_{kn}^2 \\
&\le \left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{LS}-\boldsymbol{\theta}\right)\right\|^2\left[\max_{1\le k\le n}\left\|n^{-1/2}\mathbf{F}_{k-1}\right\|\right]^2\frac{c_n}{nl_n}\sum_{k=1}^{n}\left\|\mathbf{F}_{k-1}\right\|^2K_{kn}^2 \\
&= o_P(1),
\end{aligned}$$

where, in the last equality, we have used the result:

$$\frac{c_n}{nl_n}\sum_{k=1}^{n}\left\|\mathbf{F}_{k-1}\right\|^2K_{kn}^2\to_P\int K^2\cdot E\left\|\mathbf{F}_1\right\|^2$$

as explained in Remark 14 with $G(x)\equiv 1$. Similarly, we have

$$\begin{aligned}
\frac{1}{2}\left\|\frac{c_n}{nl_n}R_{2n}\right\| &= \frac{c_n}{nl_n}\left\|\sum_{k=1}^{n}e_k\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{LS}\right)'\mathbf{F}_{k-1}\right]\mathbf{F}_{k-1}\mathbf{F}_{k-1}'K_{kn}^2\right\| \\
&\le \left[\left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{LS}-\boldsymbol{\theta}\right)\right\|\max_{1\le k\le n}\left\|n^{-1/2}\mathbf{F}_{k-1}\right\|\right]\frac{c_n}{nl_n}\sum_{k=1}^{n}\left|e_k\right|\left\|\mathbf{F}_{k-1}\right\|^2K_{kn}^2 \\
&\le \left[\left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{LS}-\boldsymbol{\theta}\right)\right\|\max_{1\le k\le n}\left\|n^{-1/2}\mathbf{F}_{k-1}\right\|\right]\frac{c_n}{nl_n}\sum_{k=1}^{n}\left(1+e_k^2\right)\left\|\mathbf{F}_{k-1}\right\|^2K_{kn}^2 \\
&= o_P(1),
\end{aligned}$$

where, in the last equality, we have used the fact that

$$\frac{c_n}{nl_n}\sum_{k=1}^{n}\left(1+e_k^2\right)\left\|\mathbf{F}_{k-1}\right\|^2K_{kn}^2\to_P\int K^2\cdot E\left[\left(1+\sigma_2^2\right)\left\|\mathbf{F}_1\right\|^2\right].$$

Next, we prove (S11.11). Let $[A]_{rs}$ denote the $(r,s)$ element of a matrix $A$. In view of Lemmas 1 and 2 with $G(x)=1$ (cf. Remark 14), it suffices to show that, for any $m:=m_n\to\infty$, $n/m_n\to\infty$,

$$\Delta_{rs,n} := \max_{m\le j\le n-m}\frac{1}{m}E\left|\sum_{k=j+1}^{j+m}\left[e_k^2\mathbf{F}_{k-1}\mathbf{F}_{k-1}'\right]_{rs}-E\left[\sigma_k^2\mathbf{F}_{k-1}\mathbf{F}_{k-1}'\right]_{rs}\right|\to 0. \quad \text{(S11.13)}$$

Note that $\Delta_{rs,n}\le R_{3n}+R_{4n}$, where

$$R_{3n} = \max_{m\le j\le n-m}\frac{1}{m}E\left|\sum_{k=j+1}^{j+m}\left[\sigma_k^2\mathbf{F}_{k-1}\mathbf{F}_{k-1}'\right]_{rs}\left[u_k^2-E\left(u_k^2|\mathcal{F}_{k-1}\right)\right]\right|$$

$$R_{4n} = \max_{m \le j \le n-m} E \left| \frac{1}{m} \sum_{k=j+1}^{j+m} \left\{ \left[ \sigma_k^2 \mathbf{F}_{k-1} \mathbf{F}'_{k-1} \right]_{rs} - E \left[ \sigma_k^2 \mathbf{F}_{k-1} \mathbf{F}'_{k-1} \right]_{rs} \right\} \right|.$$

Since $\sigma_k^2 \mathbf{F}_{k-1} \mathbf{F}'_{k-1}$ is strict stationarity and ergodic, it is readily seen that $R_{4n} \to 0$.

To consider $R_{1n}$, set $\lambda_k := \left[ \sigma_k^2 \mathbf{F}_{k-1} \mathbf{F}'_{k-1} \right]_{rs}$ and $U_k := \left[ u_k^2 - E\left( u_k^2 | \mathcal{F}_{k-1} \right) \right]$. Then for all $A > 0$ as $m \to \infty$ first and then $A \to \infty$,

$$\begin{aligned} R_{3n} &\le \max_{m \le j \le n-m} \frac{1}{m} E \left| \sum_{k=j+1}^{j+m} \lambda_k I \left\{ |\lambda_k| \le A \right\} U_k \right| + \max_{m \le j \le n-m} \frac{1}{m} E \left| \sum_{k=j+1}^{j+m} \lambda_k I \left\{ |\lambda_k| > A \right\} U_k \right| \\ &\le A \max_{m \le j \le n-m} \left\{ \frac{1}{m^2} E \sum_{k=j+1}^{j+m} U_k^2 \right\}^{1/2} \\ &\quad + \max_{m \le j \le n-m} \frac{1}{m} E \sum_{k=j+1}^{j+m} |\lambda_k| I \left\{ |\lambda_k| > A \right\} \left[ E\left( u_k^2 | \mathcal{F}_{k-1} \right) + u_k^2 \right] \\ &\le A \left\{ \frac{1}{m} \sup_k E u_k^4 \right\}^{1/2} + 2E|\lambda_1| I \left\{ |\lambda_1| > A \right\} \to 0. \end{aligned}$$

Combining all these estimates, we establish (S11.13). This completes the proof. □

## S12 Additional Simulation Results

Table S1: Empirical Size of CTLS tests: $\hat{T}$
(nominal size 5%; NI regressor, GARCH(1,1) regression errors)

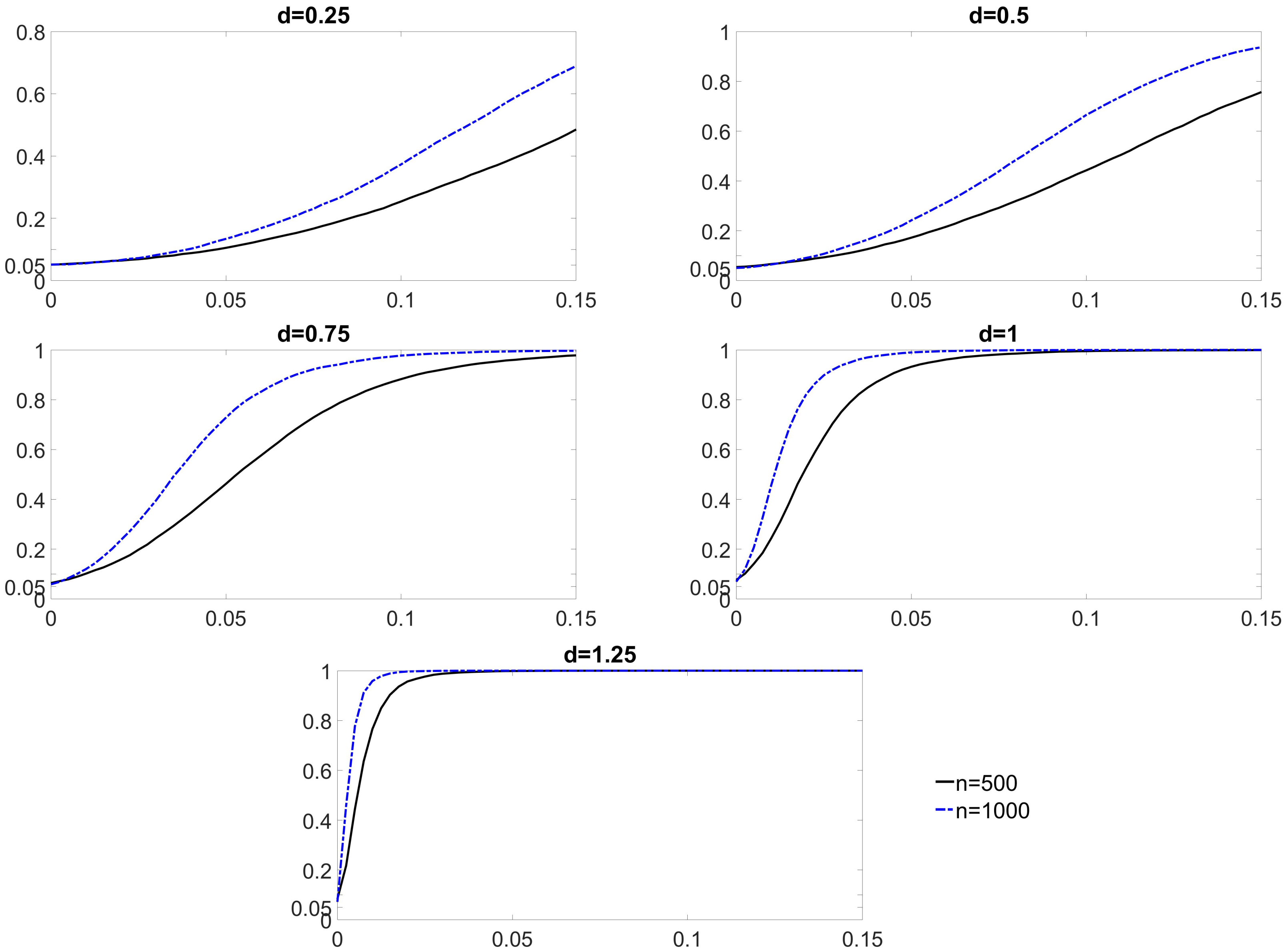
| | $c=0$ | | | | | $c=-5$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\varrho$ | -0.95 | -0.5 | 0 | 0.5 | 0.95 | -0.95 | -0.5 | 0 | 0.5 | 0.95 |
| $n$=250 | 0.083 | 0.060 | 0.052 | 0.062 | 0.087 | 0.062 | 0.054 | 0.050 | 0.060 | 0.069 |
| 500 | 0.077 | 0.057 | 0.050 | 0.061 | 0.077 | 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 |
| | $c=-10$ | | | | | $c=-20$ | | | | |
| $\varrho$ | -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 S2: Empirical Size of CTLS tests: $\hat{T}$ (nominal size 5%; fractional regressor, cond. homoscedastic regression errors)

| | | $d = 0.25$ | | | $d = 0.5$ | | | $d = 0.75$ | | | $d = 0.8$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\varrho$ | -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 |

| | | $d = 0.9$ | | | $d = 1$ | | | $d = 1.1$ | | | $d = 1.2$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\varrho$ | -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 S1: Empirical Power of CTLS tests: $\hat{T}$ Plotted against $\beta$.
(5% nominal size; $\varrho = -0.95 = -0.95$; fractional regressor, cond. homoscedastic regression errors)



## References

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

[2] Billingsley P. (1968). *Convergence of Probability Measures.* Wiley, New York.

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

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

[5] Chang, Y., Park, J.Y. and Phillips P.C.B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal*, **4**, 1–36.

[6] 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.

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

[8] 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.

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

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

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

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

[13] Magdalinos, T. (2022). Least squares and IVX limit theory in systems of predictive regressions with GARCH innovations. *Econometric Theory*, **38**(5), 875–912.

[14] Magdalinos, T. and Petrova, K. (2022). Uniform and distribution-free inference with general autoregressive processes, Working Paper 1344, Barcelona School of Economics.

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

[16] Magdalinos, T. and Phillips, P.C.B. (2018). Wald testing with matrix normalization. Working Paper, University of Southampton.

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

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

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

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

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

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

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

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

[25] Yang, B., Long W., Peng, L. and Cai, Z. (2020). Testing the predictability of U.S. housing price index returns based on an IVX-AR model. *Journal of the American Statistical Association*, **115**(532), 1–34.