Essex Finance Centre
Working Paper Series

"Testing for Episodic Predictability in Stock Returns"

"Matei Demetrescu, Iliyan Georgiev, Paulo M. M. Rodrigues, A.M. Robert Taylor"

Testing for Episodic Predictability in Stock Returns[*]

Matei Demetrescu[a], Iliyan Georgiev[b], Paulo M. M. Rodrigues[c] and A. M. Robert Taylor[d]

[a] Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel

[b] Department of Economics, University of Bologna

[c] Banco de Portugal and NOVA School of Business and Economics

[d] Essex Business School, University of Essex

December 30, 2019

Abstract

Standard tests based on predictive regressions estimated over the full available sample data have tended to find little evidence of predictability in stock returns. Recent approaches based on the analysis of subsamples of the data suggest in fact that predictability where it occurs might exist only within so-called "pockets of predictability" rather than across the entire sample. However, these methods are prone to the criticism that the subsample dates are endogenously determined such that the use of standard critical values appropriate for full sample tests will result in incorrectly sized tests leading to spurious findings of stock returns predictability. To avoid the problem of endogenously-determined sample splits, we propose new tests derived from sequences of predictability statistics systematically calculated over subsamples of the data. Specifically, we will base tests on the maximum of such statistics from sequences of forward and backward recursive, rolling, and double-recursive predictive subsample regressions. We develop our approach using the over-identified instrumental variable-based predictability test statistics of Breitung and Demetrescu (2015). This approach is based on partial-sum asymptotics and so, unlike many other popular approaches including, for example, those based on Bonferroni corrections, can be readily adapted to implementation over sequences of subsamples. We show that the limiting null distributions of our proposed test statistics depend in general on whether the putative predictor is strongly or weakly persistent and on any heteroskedasticity present (indeed on any time-variation present in the unconditional variance matrix of the innovations), the latter even if the subsample statistics are based on heteroskedasticity-robust standard errors. As a consequence, we develop fixed regressor wild bootstrap implementations of the tests which we demonstrate to be first-order asymptotically valid. Finite sample behaviour against a variety of temporarily predictable processes is considered. An empirical application to US stock returns illustrates the usefulness of the new predictability testing methods we propose.

Keywords: predictive regression; rolling and recursive IV estimation; persistence; endogeneity; conditional and unconditional heteroskedasticity.

JEL classification: C12, C22

1 Introduction

A large body of empirical research has been undertaken investigating whether stock returns can be predicted. Therein, a wide range of financial and macroeconomic variables have been considered as putative predictors for returns, including valuation ratios such as the dividend-price ratio, dividend yield, earnings-price ratio, book-to-market ratio, various interest rates and interest rate spreads, and macroeconomic variables including inflation and industrial production.

Early empirical studies, including Fama (1981), Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a,b), Fama and French (1988, 1989) and Fama (1990), often found significant evidence of in-sample predictability of U.S. stock index returns, at least over relatively long horizons. It has since been argued, however, that these findings could be spurious. Nelson and Kim (1993) and Stambaugh (1999) show that strongly persistent predictors lead to biased coefficients in predictive regressions if the innovations driving the predictors are correlated with returns, as is argued to be the case for many of the variables used as predictors; e.g., the stock price is a component of both the return and the dividend yield. Goyal and Welch (2003) show that the persistence of dividend-based valuation ratios increased significantly over the typical sample periods used in empirical studies, and argue that, as a consequence, out-of-sample predictions using these variables are no better than from a no-change strategy. Predictability tests which are asymptotically valid when the predictor is strongly persistent and driven by innovations which are correlated with returns have been proposed in Cavanagh *et al.* (1995), Campbell and Yogo (2006), Kostakis *et al.* (2015), Breitung and Demetrescu (2015), Elliott *et al.* (2015) and Jansson and Moreira (2006), *inter alia*. When such robust techniques are used the statistical evidence of predictability is considerably weaker and often disappears completely; see, among others, Ang and Bekaert (2007), Boudoukh *et al.* (2007), Welch and Goyal (2008) and Breitung and Demetrescu (2015).

The foregoing approaches are based on a maintained assumption that the coefficients of the predictive regression model are constant over time. However, there are several reasons to suspect that if returns are predictable, then it is likely to be a time-varying phenomenon. The business cycle, time-varying risk aversion, rare disasters, structural breaks, speculative bubbles, investor's market sentiment, and regime changes in monetary policy have all be cited as possible reasons; see, e.g., Pesaran and Timmermann (2002). For example, significant changes in monetary policy and financial regulations could lead to shifts in the relationship between macroeconomic variables and the fundamental value of stocks, via the impact of these changes on economic growth and the growth rates of earnings and dividends. Timmermann (2008) argues that for most time periods returns are not predictable but that there are 'pockets in time' where evidence of local predictability is seen. In particular, if predictability exists as a result of market inefficiency rather than because of time-varying risk premia, then rational investors will attempt to exploit its presence to earn abnormal profits. Assuming a large-enough proportion of investors are rational, this behaviour will eventually cause the predictive power of the relevant predictor to be eliminated. If a variable begins to have predictive power for returns then a window of predictability might exist before investors learn about that relationship, but it will eventually disappear; see, in particular, Paye and Timmermann (2006), Timmermann (2008) and Farmer *et al.* (2018). It therefore seems reasonable to consider the possibility that the predictive relationship might change over time, so that over a long span of data one may observe some windows of time during which predictability occurs.

A growing body of empirical evidence is supportive of the view that the slope parameter in

prediction models for returns varies over time. Henkel *et al.* (2011) find that return predictability in the stock market appears to be closely linked to economic recessions with dividend yield and term structure variables displaying predictive power only during recessions. Similarly, Gargano *et al.* (2017) find that commodity returns are predictable using macroeconomic information, but again only during recessions. Lettau and Ludvigsson (2001) find evidence of instability in the predictive ability of the dividend and earnings yield in the second half of the 1990s. Goyal and Welch (2003) and Ang and Bekaert (2007) find instability in prediction models for U.S. stock returns based on the dividend yield in the 1990s. Other studies which report evidence of time-varying behaviour in stock return predictability include Barberis (2000), Lettau and van Nieuwerburgh (2008), Welch and Goyal (2008), Pástor and Stambaugh (2009, 2012), Pettenuzzo and Timmermann (2011), Dangl and Halling (2012), Gonzalo and Pitarakis (2012), Rapach and Wohar (2006) and Giannetti (2007), *inter alia.* In the context of predicting the equity premium, Kolev and Karapandza (2017) find that, for a given set of predictors, alternative data splits often lead to strongly contradictory outcomes concerning return predictability. Paye and Timmermann (2006) undertake a comprehensive analysis of prediction model instability for international stock market indices using conventional Bai-Perron structural break tests and report statistically significant evidence of structural breaks for many of the countries considered, arguing that the "[e]mpirical evidence of predictability is not uniform over time and is concentrated in certain periods" (*op. cit.* p.312). Paye and Timmermann (2006) also cite a number of applied studies which find significant evidence of in-sample (ex post) predictability in returns data but yet find very weak evidence of out-of-sample (ex ante) predictability, and argue that a possible explanation is structural instability in the predictive relations involved.

A limitation of many of the statistical techniques used in previous research on the instability of return prediction models is that they are not designed for use with highly persistent, endogenous predictors. Paye and Timmermann (2006) investigate the effects of persistence and endogeneity of the regressors on the Bai-Perron tests for structural breaks using Monte Carlo simulations. Their simulations reveal that size distortions, whereby parameter change is falsely signalled when none is present, can be substantial. They also show that some of the tests lack power in this context because of the large amount of noise typically present in predictive regression models. Moreover, because tests from predictive regression models based on the full sample of available data will have relatively low power to detect short windows of predictability, a number of these studies applied such tests to separate subsamples of the data (data splits) with the timings of those subsamples either chosen by the practitioner or performed over a large set of possible subsamples of the data. In both cases the critical values which would apply to the test run on the full sample cannot validly be used. In the former case because the subsamples are endogenously determined. For the latter case the probability of spuriously signalling a predictive relationship when none is present will tend to one as the number of subsamples considered increases; see Inoue and Rossi (2005) for a detailed discussion of this problem in relation to the use of t-tests.

With these issues in mind, our goal is to provide size controlled tests designed to detect predictability regardless of whether it applies across the entire available sample or within pockets of predictability. Our proposed tests are based on predictability statistics obtained from sequences of predictive regressions computed over subsamples of the data. In particular we will consider forward and backward recursive sequences of predictive regressions as well as rolling and double-recursive sequences. For each of these sequences of statistics the proposed test will be given by the largest

(in absolute value) outcome. Because the range of subsamples considered in each sequence is set without reference to the particular data set involved this avoids the problem of endogenously chosen sample splits. Moreover, by considering the maximum of the sequences we avoid the spurious detection issues discussed in Inoue and Rossi (2005). As we will demonstrate in the Monte Carlo experiments considered in the paper, each of these sequences has particular patterns of local predictability that it is well designed to detect. For example, the test based on the (forward) reverse recursive sequence of statistics is suited to detecting (beginning-of-sample) end-of-sample pockets of predictability. As such, the reverse recursive based tests could usefully be employed in an on-going monitoring exercise for the emergence of predictive regimes. Because both the forward and reverse recursive sequences contain the usual full sample predictability test, they also deliver tests which have power to detect predictability which holds over the whole sample. For a given window width, tests based on a rolling sequence of statistics are designed to pick up a window of predictability, of roughly the same length, within the data. The double-recursive sequence amounts to considering all possible window width rolling sequences, subject to a minimum width. These then are useful for picking up multiple predictive regimes of potentially different lengths within the data.

The approach we take has implications for the type of predictability statistics that can be used, as it will be necessary to characterize the joint behaviour of the full sequence of statistics in order to conduct inference. Some commonly used testing approaches such as those based on the Bonferroni inequality (e.g. Campbell and Yogo, 2006) or on bias corrections (e.g. Amihud and Hurvich, 2004) are difficult to implement for sequences of statistics; the same holds for technically more involved procedures such as those proposed by Jansson and Moreira (2006) or Elliott *et al.* (2015). Rather we will use tests based on instrumental variable [IV] estimation which benefit from the fact that closed-form expressions for the test statistics exist, which can be characterised using familiar partial-sum-based asymptotics. Specifically we will adapt the full sample methods from Kostakis *et al.* (2015) who propose the use of the so-called extended IV, or IVX, approach, and Breitung and Demetrescu (2015) who propose the combination of several instruments with complementary properties. While the marginal null limiting distribution (at least when based on heteroskedasticity-robust standard errors) of any one of the subsample statistics in the sequences considered does not depend on either the degree of persistence or endogeneity of the regressors in the predictive regression, or on any heteroskedasticity present in the shocks, we show this is not the case for the limiting null distribution of the *maximum statistics* from these sequences. We therefore propose fixed regressor wild bootstrap implementations of the maximum tests and demonstrate that these are first-order asymptotically valid under heteroskedasticity, irrespective of whether the regressors are strongly or weakly persistent.

The paper is organised as follows. Section 2 introduces the time-varying predictive regression model we consider together with the assumptions needed for our analysis. Section 3 reviews the standard full sample IV-based predictability tests, while Section 4 details the subsample implementations of these statistics across the various sequence types discussed above and discusses relevant instruments that can validly be used in the context of our proposed approach. Representations for the limiting distributions of these statistics under both the null and local alternatives are provided and are shown to depend on any heteroskedasticity present, regardless of whether the putative predictor follows a strongly persistent process (modeled as near-integrated) or a weakly persistent process (modeled as a stable autoregression). Moreover, the form of these limiting distributions

depends on whether the predictor is near-integrated or weakly dependent, even under homoskedasticity. Section 5 discusses fixed regressor wild bootstrap implementations of our proposed tests and demonstrates the first-order asymptotic validity of these. Section 6 presents the results from a Monte Carlo analysis into the finite sample behaviour of the tests under both the null hypothesis of no predictability and alternatives where local predictability occurs within the data. An application to monthly U.S. stock returns data is presented in section 7. Section 8 concludes. An accompanying on-line supplementary appendix provides detailed proofs of the technical results given in the paper along with extensions to allow for multiple predictors and a general deterministic component, additional material relating to our empirical application and additional simulation results.

The notation $\mathcal{D}^k$ will be used to denote the space of càdlàg real functions on $[0,1]^k$ equipped with the Skorokhod topology, and we abbreviate $\mathcal{D}^1$ to $\mathcal{D}$. The weak convergence of probability measures on both function spaces (in particular, on $\mathcal{D}^k$) and on $\mathbb{R}^k$ is denoted by $\Rightarrow$. We reserve the notation P, E etc. for probability, expectation etc. with respect to the distribution of the original data and use P*, E* etc. for probability, expectation etc. induced by the data and the wild bootstrap multipliers (denoted $\{R_t\}$) conditionally on the data. The notation $\overset{w}{\Rightarrow}_p$ stands for weak convergence in probability; specifically, $\zeta_T^* \overset{w}{\Rightarrow}_p \zeta$ holds for random elements ζ_T^* and ζ, not necessarily defined on the same probability space, if $\mathrm{E}^* f(\zeta_T^*) \to \mathrm{E} f(\zeta)$ in P-probability for all bounded continuous real functions f with matching domain. In the special case $\zeta = 0 \in \mathbb{R}$, we recall that $\zeta_T \overset{w}{\Rightarrow}_p 0$ (equivalently, $\zeta_T = o_{p^*}(1)$ in P-probability) means that $\mathrm{P}^*(|\zeta_T| > \varepsilon) \to 0$ in P-probability for every $\varepsilon > 0$. Finally, $\zeta_T = O_{p^*}(1)$ in P-probability signifies that for every $\varepsilon > 0$ there exists a $K > 0$ such that $\mathrm{P}\{\mathrm{P}^*(|\zeta_T| > K) < \varepsilon\} > 1 - \varepsilon$ for all T. The o_p and O_p symbols retain their usual meaning.

2   The Episodic Predictive Regression Model

The basic predictive regression model for stock returns, y_t, allowing for time-variation in the slope coefficient on the predictor variable, is taken to be of the form

$$y_t = \beta_0 + \beta_{1,t} x_{t-1} + u_t, \qquad t = 1, \ldots, T \tag{2.1}$$

where x_t, $t = 0, \ldots, T$, is observed and satisfies the data generating process [DGP]

$$x_t = \mu_x + \xi_t, \qquad t = 0, \ldots, T \tag{2.2a}$$

$$\xi_t = \rho \xi_{t-1} + v_t, \qquad t = 1, \ldots, T \tag{2.2b}$$

with ξ_0 a mean zero $O_p(1)$ variate. The innovations u_t are martingale difference [MD] sequences, while v_t is allowed to exhibit weak serial dependence. For expositional simplicity we have only allowed for a single predictive regressor, x_{t-1}, and an intercept in (2.1). Generalisations to the case where the predictive regression contains multiple predictors and/or a general deterministic component of the form considered in section 3.2 of Breitung and Demetrescu (2015) are detailed in section S.2 of the supplementary appendix.

The DGP in (2.1) generalises the constant parameter predictive regression model by allowing the slope coefficient on x_{t-1} to vary over time, thereby allowing for changes over time in the predictive content of the regressor x_{t-1}. The constant parameter predictive regression model obtains by setting a constant slope parameter such that $\beta_{1,t} = \beta_1$, for all $t = 1, \ldots, T$. Our interest will focus in this

paper on testing the usual null hypothesis that $(y_t - \beta_0)$ is a MD sequence and, hence, that y_t is not predictable by x_{t-1}, which entails that $\beta_{1,t} = \beta_1 = 0$, for all $t = 1, \ldots, T$, in (2.1). In contrast to the extant literature which tests this null hypothesis against the alternative that y_t is predictable by x_{t-1} with a constant slope parameter holding across the whole sample, that is $\beta_1 \neq 0$, under the maintained hypothesis that $\beta_{1,t} = \beta_1$, for all $t = 1, \ldots, T$, we will test against alternatives such that $\beta_{1,t} \neq 0$ for some t but without imposing constancy on $\beta_{1,t}$. Some structure obviously needs to be placed on the class of alternative hypotheses we may consider and this will be formalised below.

As discussed in the Introduction it is important to allow for the possibility of high persistence in the predictor variable x_t and to allow the shocks driving the predictor, v_t in (2.2), to be correlated with the unpredictable component of stock returns, u_t in (2.1). As regards the latter, we will allow u_t and v_t to be contemporaneously correlated and heteroskedastic; exact conditions will be detailed in Assumption 3. For the former, we allow ρ in (2.2) to satisfy the following assumption.

Assumption 1 *Exactly one of the two following conditions holds true:*

1. **Weakly persistent predictors:** *The autoregressive parameter ρ in (2.2) is fixed and bounded away from unity, $|\rho| < 1$.*

2. **Strongly persistent predictors:** *The autoregressive parameter ρ in (2.2) is local-to-unity with $\rho := 1 - \frac{c}{T}$ where c is a fixed non-negative constant.*

Remark 1. Many predictors are strongly persistent, exhibiting sums of sample autoregressive coefficients which are close to unity. Near-integrated asymptotics has been found to provide better approximations for the behaviour of test statistics in such circumstances; see, *inter alia*, Elliott and Stock (1994). However, a large part of the literature works with models which take x_t to be generated from a stable autoregressive process; see, for example, Amihud and Hurvich (2004). Assumption 1 allows for either of these possibilities to hold on x_t. $\diamondsuit$

We will develop tests for the null hypothesis that y_t is not predictable by x_{t-1} in any subsample, which do not require the practitioner to know which of Assumption 1.1 or Assumption 1.2 holds in (2.2), nor indeed what the precise value of ρ is in either case. Moreover, we aim to develop tests which possess non-trivial asymptotic *local* power against DGPs where predictability is present. Predictive regressions for stock returns typically exhibit small R^2 and low signal-to-noise ratios (see, *inter alia*, Campbell, 2008, and Phillips, 2015) so departures from the null, should predictability be present, are small. We will therefore conduct our theoretical analysis of the large sample properties of the tests we discuss under local alternatives such that the slope parameter $\beta_{1,t}$ is local-to-zero for an asymptotically non-vanishing set of the sample observations. The localisation rate (or Pitman drift) will need to be such that $\beta_{1,t}$ is specified to lie in a neighbourhood of zero which shrinks with the sample size, T. The appropriate Pitman drift is dictated by which of Assumption 1.1 and Assumption 1.2 holds in (2.2). Where x_t is near-integrated the appropriate rate is T^{-1}, while for weakly dependent x_{t-1}, the rate is $T^{-1/2}$. The different localisation rates reflect the fact that near-integration implies a much stronger signal from the predictor x_{t-1}. Moreover, tests based on the maxima from sequences of subsample predictability test statistics can only deliver non-trivial asymptotic local power in cases where an asymptotically non-vanishing fraction of the data is such that $\beta_{1,t} \neq 0$ holds on the DGP. For example, if $\beta_{1,t} \neq 0$ at one time point only, then although this would formally violate the null that y_t is not predictable by x_{t-1}, this data point would, as

$T \to \infty$, however be dominated by the remaining $T-1$ data points where the null hypothesis holds. Formally, in our framework we specify $\beta_{1,t}$ to satisfy the following assumption.

Assumption 2 *In the context of (2.1) and (2.2), let $\beta_{1,t} := n_T^{-1} b(t/T)$, where $b(\cdot)$ is a piecewise Lipschitz-continuous real function on $[0,1]$, with $n_T = \sqrt{T}$ under Assumption 1.1, and $n_T = T$ under Assumption 1.2.*

Using the framework of Assumption 2 we can then equivalently write our null hypothesis that $\beta_{1,t} = 0$, for all $t = 1, \ldots, T$, as

$$H_0 : \text{ The function } b(\tau) \text{ is identically zero for all } \tau \in [0,1]. \tag{2.3}$$

We can now also formally specify the alternative hypothesis as,

$$H_{1,b(\cdot)} : \text{ The function } b(\cdot) \text{ is non-zero over at least one non-empty open interval contained in } [0,1]. \tag{2.4}$$

Remark 2. The alternative hypothesis specified by $H_{1,b(\cdot)}$ is very general but entails that at least one subset of the sample observations (this need not be a strict subset, so it could contain all of the sample observations) comprising contiguous observations exists for which $\beta_{1,t} \neq 0$, and where the size of this subset is proportional to the sample size T. Notice that under $H_{1,b(\cdot)}$ the integral of $|b(\cdot)|$ on $[0,1]$ is non-zero and it is this property which qualifies $H_{1,b(\cdot)}$ as a genuine (local) alternative. Moreover, as we will establish later, the form that $b(\cdot)$ takes under H_1 determines the local power offsets obtained in the limiting distributions of the statistics we propose. Notice also that, under $H_{1,b(\cdot)}$, $b(\cdot)$ may be zero in certain parts of its domain and it may also change magnitude and/or sign over its domain; the former corresponds to data points where $\beta_{1,t} = 0$, while the latter reflects observations for which $\beta_{1,t}$ does not have a fixed magnitude and/or sign across the full sample. $\diamond$

We conclude this section by detailing in Assumption 3 the conditions that we will place on the disturbances u_t and v_t in (2.1) and (2.2).

Assumption 3 *Let*

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & B(L) \end{pmatrix} \mathbf{H}(t/T) \begin{pmatrix} a_t \\ e_t \end{pmatrix}, \quad \text{with} \quad \begin{pmatrix} a_t \\ e_t \end{pmatrix} \sim \text{White Noise} (0, \mathbf{I}_2), \tag{2.5}$$

where $\mathbf{I}_k$ denotes the $k \times k$ identity matrix and:

1. *$\boldsymbol{\zeta}_t := (a_t, e_t)'$ is a uniformly L_4-bounded martingale difference sequence which is such that $\sup_t \mathrm{E} \left| \mathrm{E} \left(\boldsymbol{\zeta}_t \boldsymbol{\zeta}_t' - \mathbf{I}_2 | \boldsymbol{\zeta}_{t-m}, \boldsymbol{\zeta}_{t-m-1}, \ldots \right) \right| \to 0$ as $m \to \infty$;*

2. *$\mathbf{H}(\cdot) := \begin{pmatrix} h_{11}(\cdot) & h_{12}(\cdot) \\ h_{21}(\cdot) & h_{22}(\cdot) \end{pmatrix}$ is a matrix of piecewise Lipschitz-continuous bounded functions on $(-\infty, 1]$, which is of full rank at all but a finite number of points;*

3. *$B(L)$, where L denotes the usual lag operator, is an invertible lag polynomial with $b_0 = 1$ and 1-summable coefficients, $\sum_{j \geq 0} j |b_j| < \infty$, for which $\omega := \sum_{j \geq 0} b_j > 0$.*

6

Remark 3. The structure in (2.5) imposes that the disturbances u_t are uncorrelated with the increments of x_t at all (positive) lags. Where $\boldsymbol{\zeta}_t$ is independent and identically distributed [IID], this structure would entail that x_{t-1} is weakly exogenous with respect to u_t, and we will continue (with an abuse of language) to use the same term as a shorthand to describe this structure irrespective of whether $\boldsymbol{\zeta}_t$ is IID or not. Assumption 3.3 allows the increments to the predictor x_{t-1} to be serially correlated. These dynamics are not restricted beyond a 1-summability regularity condition on the moving average representation, as is typical in this literature; see, for example, Breitung and Demetrescu (2015) and Kostakis *et al.* (2015). $\diamond$

Remark 4. Assumption 3 allows for quite general forms of heteroskedasticity in $(\tilde{u}_t, \tilde{v}_t)' := \mathbf{H}(t/T)(a_t, e_t)'$ and hence in u_t and v_t. In particular, Assumption 3.1 imposes a MD structure on $\boldsymbol{\zeta}_t$ allowing for conditional heteroskedasticity which is natural for the empirical applications to financial data we have in mind. Assumption 3.1 also imposes finite fourth moments; while daily returns often display very fat tails (see, for example, Nicolau and Rodrigues, 2019) such that the assumption of finite fourth order moments might not be a suitable assumption for daily data, standard predictive regression models have tended to be run on lower frequency data (monthly, quarterly or even annual data) where infinite kurtosis does not appear to be a concern. Assumption 3.1 places summability conditions on the cross-product moments of the innovations which limits the degree of serial dependence allowed in the conditional variances; these conditions are satisfied, for example, by strictly stationary and ergodic MD sequences with finite variance. Assumption 3.2 allows for unconditional time heteroskedasticity in the innovations through the matrix $\mathbf{H}(\tau)$. Where $\mathbf{H}(\tau)$ is diagonal for all $\tau \in [0, 1]$ the innovations $(\tilde{u}_t, \tilde{v}_t)'$ can display time-varying variances but are contemporaneously uncorrelated, so in the case where $\boldsymbol{\zeta}_t$ is IID with independent components, this would entail that x_t is strictly exogenous with respect to u_t (again we will use this terminology, with an abuse of language, whether $\boldsymbol{\zeta}_t$ is IID or not). Importantly, the off-diagonal elements of $\mathbf{H}(\tau)\mathbf{H}(\tau)'$ (i.e., the covariance matrix of $(\tilde{u}_t, \tilde{v}_t)'$) are not imposed to be zero, thus allowing for contemporaneous and time-varying correlation among the innovations. The structure placed on $\mathbf{H}(\tau)$ by Assumption 3.2 allows for a wide class of models for the behaviour of the variance matrix of the innovations including single or multiple (co-) variance shifts, variances which follow a broken trend, and smooth transition variance shifts. As discussed in Breitung and Demetrescu (2015, p.360), such patterns are plausible with macro and financial data and it is therefore important to use tests which are robust to such behaviour to avoid the possibility of spurious rejection of the null because of non-constancy in the variance matrix rather than genuine predictability from x_{t-1}. $\diamond$

Under Assumption 1.1, x_t is a particular case of a *locally* stationary process which admits a time-varying variance when the series v_t displays time-varying volatility. In fact, the variance of the putative predictor is given as $\text{Var}\left(x_{\lfloor \tau T \rfloor}\right) \approx \bar{\sigma}_{\xi}^2(\rho)\left(h_{21}^2(\tau) + h_{22}^2(\tau)\right)$, where $\bar{\sigma}_{\xi}^2(\rho)$ denotes the sum of the squared coefficients of the lag polynomial $(1 - \rho L)^{-1}B(L)$ (which is finite in the stable autoregression case). This form of heteroskedasticity impacts on the inferential procedures based on subsample sequences of statistics discussed in this paper. Furthermore, time-varying volatility where present in the regression errors, u_t, and in the instrumental variables used in constructing the statistics can also affect the behaviour the sequences of statistics.

Heteroskedasticity has analogous effects under Assumption 1.2 (near-integration), though the transmission mechanism is somewhat different. In particular, under Assumption 3 we have that

$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} \mathbf{H}(t/T) \boldsymbol{\zeta}_t \Rightarrow \int_0^\tau \mathbf{H}(s) \mathrm{d}\boldsymbol{W}(s) =: (U(\tau), V(\tau))'$ on $\mathcal{D}^2$, where $\boldsymbol{W}$ is a two-dimensional standard Wiener process (see e.g. the invariance principle from Boswijk *et al.*, 2016), such that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} \begin{pmatrix} u_t \\ v_t \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 \\ 0 & \omega \end{pmatrix} \int_0^\tau \mathbf{H}(s) \mathrm{d}\boldsymbol{W}(s) =: \begin{pmatrix} U(\tau) \\ \omega V(\tau) \end{pmatrix} \tag{2.6}$$

on $\mathcal{D}^2$. The processes $U(\tau)$ and $\omega V(\tau)$ are individually time-transformed Brownian motions whose correlation may also vary over time; their covariance at time τ is given by $\omega \int_0^\tau \mathbf{H}(s)\mathbf{H}(s)'\mathrm{d}s$. Under Assumption 1.2 x_t satisfies the invariance principle, $\frac{1}{\sqrt{T}} x_{\lfloor \tau T \rfloor} \Rightarrow \omega J_{c,\mathrm{H}}(\tau)$, where $J_{c,\mathrm{H}}(\tau)$ is an Ornstein-Uhlenbeck-type process driven by $V(\tau)$, i.e., $J_{c,\mathrm{H}}(\tau) := \int_0^\tau e^{-c(\tau-s)} \mathrm{d}V(s)$. Notice that $J_{c,\mathrm{H}}(\tau)$ is a heteroskedastic process when $\mathbf{H}(\cdot)$ is not constant: the quadratic variation processes of $U(\tau)$ and $V(\tau)$, given by $[U](\tau) := \int_0^\tau \left(h_{11}^2(s) + h_{12}^2(s) \right) \mathrm{d}s$ and $[V](\tau) := \int_0^\tau \left(h_{21}^2(s) + h_{22}^2(s) \right) \mathrm{d}s$, respectively, are nonlinear in general, and their quadratic covariation process is given by $[UV](\tau) := \int_0^\tau \left(h_{11}(s)h_{21}(s) + h_{12}(s)h_{22}(s) \right) \mathrm{d}s$.

3   Full Sample Predictability Tests

Consider the maintained hypothesis that the slope parameter $\beta_{1,t}$ in (2.1) is constant, such that $\beta_{1,t} = \beta_1$, for all $t = 1, \ldots, T$. This yields the standard constant parameter predictive regression

$$y_t = \beta_0 + \beta_1 x_{t-1} + u_t, \qquad t = 1, \ldots, T. \tag{3.1}$$

A number of procedures have been developed for testing $H_0 : \beta_1 = 0$ in (3.1) against the local alternative $H_c : \beta_1 = n_T^{-1} b_1$, with b_1 a non-zero constant. Of these the simplest is the standard (full sample) ordinary least squares [OLS] t-test for the significance of x_{t-1} in (3.1). While standard normal asymptotic theory applies to the t-statistic under Assumption 1.1 provided the errors are homoskedastic (although this can be weakened by using heteroskedasticity-robust standard errors), it does not under Assumption 1.2 where the limiting null distribution of the t-statistic is nonstandard and depends on the local-to-unity parameter c unless x_t is strictly exogenous with respect to u_t.

Tests robust to c have been developed in Elliott and Stock (1994), who propose a Bayesian mixture procedure, and Cavanagh *et al.* (1995) and Campbell and Yogo (2006) who develop tests based on conservative bounds, and Jansson and Moreira (2006), who conduct inference on the basis of conditionally sufficient statistics. However, these procedures are all developed for the case where x_t is near-integrated, i.e. such that Assumption 1.2 holds, and for the case of homoskedastic disturbances. Variable addition [VA] techniques (see Breitung and Demetrescu (2015, p.359) for a literature review) can be used to develop predictability tests which can be validly used regardless of whether x_t is local-to-unity or stationary. However, these VA-based tests have only trivial asymptotic local power against the Pitman rate, T^{-1}, where x_t is near-integrated. Breitung and Demetrescu (2015) show that the finite sample power of the VA-based tests is indeed very low relative to the tests designed for the use with near-integrated x_t when the AR parameter ρ in (2.2) is close to unity. They also develop modifications of the VA approach but some loss of power still remains. Gorodnichenko *et al.* (2012) proposed tests based on quasi-differencing but like the original VA-based tests these only have power in $T^{-1/2}$ neighbourhoods of the null.

Breitung and Demetrescu (2015) also examine tests based on the instrumental variables [IV]

approach. They show that these can be validly implemented in the presence of endogeneity and uncertain regressor persistence and heteroskedasticity of the form specified in section 2. The basic idea underlying IV estimation of the predictive regression model is to use instruments such that the instrument has lower persistence than the regressor x_{t-1} (so-called type-I instruments), or is such that the instrument is strictly exogenous with respect to u_t (so-called type-II instruments). Formal conditions which must hold on these instruments are given in Breitung and Demetrescu (2015).

A range of possible type-I instruments is given in Breitung and Demetrescu (2015, p.361). These comprise: (i) a short memory instrument whereby we generate $z_{t-1} = (1 - \bar{\alpha}L)_+^{-1}\Delta x_{t-1}$ $:= \Delta x_{t-1} + \bar{\alpha}\Delta x_{t-2} + \cdots + \bar{\alpha}^{t-2}\Delta x_1$ with $|\bar{\alpha}| < 1$; (ii) a mildly integrated instrument, generated as $z_{t-1} = (1 - \alpha_T L)_+^{-1}\Delta x_{t-1}$, for $\alpha_T := 1 - aT^{-\gamma}$ with $a > 0$, $0 < \gamma < 1$; (iii) a fractionally integrated instrument, generated as $z_{t-1} = (1 - L)^{1-d^*}x_{t-1}\mathbb{I}(t > 0) := \Delta_+^{1-d^*}x_{t-1}$ for some $d^* \in (0, 1/2)$; (iv) a long differences instrument, generated as $z_{t-1} = x_{t-1} - x_{t-k_T}$ for $K_T := \min\{\lfloor KT^v\rfloor, t - 1\}$ for some $0 < v < 1$ and positive constant K. The use of the mildly integrated instrument in (ii) is an example of the so-called IVX approach of Phillips and Magdalinos (2009). In each case the generated instrument is, by design, free of a stochastic trend and hence less persistent than a near-integrated process, regardless of whether x_{t-1} is near-integrated or stationary. Being filtered versions of x_{t-1}, these instruments are driven by the same innovations and it is therefore expected that they provide valid instruments for x_{t-1}; at the same time, the reduced persistence leads to standard inference. Breitung and Demetrescu (2015, p.362) also discuss the following type-II instruments: (i) a generated random walk, $z_{t-1} = (1 - L)_+^{-1}w_{t-1}$ where $w_t \sim IID(0, \sigma_w^2)$ with w_t independent of u_t and v_t; (ii) deterministic functions of time, such as $z_{t-1} = (t - 1)$ or $z_{t-1} = \sin(\pi(t - 1)/2T)$, and (iii) Cauchy instruments, $z_{t-1} = \text{sign}(x_{t-1})$. Each of these is exogenous with respect to u_t by construction. However, they do not exploit any specific information about x_t, other than where x_t is near-integrated in which case they will be correlated with x_t; see Phillips (1998).

Simulation evidence in Breitung and Demetrescu (2015) shows that tests based on type-II instruments are significantly more powerful than those based on type-I instruments when x_t is strongly persistent. However, these instruments will be weak, in the sense that they will be almost uncorrelated with the regressor, where x_t is stationary. In such cases, Breitung and Demetrescu (2015) show that the resulting IV test for $\beta_1 = 0$ in (3.1) will have only trivial power. In order to simultaneously exploit the settings which result in superior power properties for the IV approach based on type-I and type-II instruments, Breitung and Demetrescu (2015) recommend the use of a test which combines two instruments for x_{t-1}, one of each type, which we denote by $z_{I,t-1}$ and $z_{II,t-1}$, collected into the vector $\boldsymbol{z}_{t-1} := (z_{I,t-1}, z_{II,t-1})'$ for $t = 1, \ldots, T$. The general form of the resulting full sample IV-combination test statistic of Breitung and Demetrescu (2015), implemented with Eicker-White standard errors to account for heteroskedasticity satisfying Assumption 3, is given by

$$t_{\beta_1} := \frac{\boldsymbol{A}_T'\mathbf{B}_T^{-1}\boldsymbol{C}_T}{\sqrt{\boldsymbol{A}_T'\mathbf{B}_T^{-1}\mathbf{D}_T\mathbf{B}_T^{-1}\boldsymbol{A}_T}} \tag{3.2}$$

where $\boldsymbol{A}_T := \sum_{t=1}^T \hat{x}_{t-1}\hat{\boldsymbol{z}}_{t-1}$, $\mathbf{B}_T := \sum_{t=1}^T \hat{\boldsymbol{z}}_{t-1}\hat{\boldsymbol{z}}_{t-1}'$, $\boldsymbol{C}_T := \sum_{t=1}^T \hat{\boldsymbol{z}}_{t-1}\hat{y}_t$ and $\mathbf{D}_T := \sum_{t=1}^T \hat{\boldsymbol{z}}_{t-1}\hat{\boldsymbol{z}}_{t-1}'\hat{u}_t^2$, with $\hat{y}_t$, $\hat{x}_{t-1}$ and $\hat{\boldsymbol{z}}_{t-1}$ denoting demeaned versions of y_t, x_{t-1} and $\boldsymbol{z}_{t-1}$, respectively, so that, for $\boldsymbol{w}_t$ generically denoting either y_t, x_{t-1} or $\boldsymbol{z}_{t-1}$, $\hat{\boldsymbol{w}}_t := \boldsymbol{w}_t - \frac{1}{T}\sum_{s=1}^T \boldsymbol{w}_s$, and where $\hat{u}_t$ denotes the regression residuals from estimating (3.1). For the reasons outlined in Remark 4 of Breitung and Demetrescu (2015), the IV-combination test must be run as two-sided and so we accordingly con-

sider tests based on the square of t_{β_1}; that is $t_{\beta_1}^2$. The limiting null distribution of $t_{\beta_1}^2$ is χ_1^2 under either Assumption 1.1 or 1.2; see Breitung and Demetrescu (2015) for details.

A variety of choices for the residuals $\hat{u}_t$ used in constructing $\mathbf{D}_T$ is possible. A natural choice is the IV regression residuals so that $\hat{u}_t := y_t - \hat{\beta}_0^{iv} - \hat{\beta}_1^{iv} x_{t-1}$, where $\hat{\beta}_j^{iv}$ denotes the two-stage least squares [2SLS] estimator of β_j, $j = 0, 1$. However, both Breitung and Demetrescu (2015) and Kostakis *et al.* (2015) recommend the use of OLS residuals on the grounds that they represent the best linear projection of y_t on x_{t-1} regardless of the persistence of the putative predictor, and that their finite-sample behaviour appears to be more stable than that of IV residuals. Finally, one could also use residuals computed under the null; i.e., $\hat{u}_t := y_t - \frac{1}{T} \sum_{s=1}^{T} y_s$. Under the local alternatives considered in Assumption 2, these three possible choices can be shown to be asymptotically equivalent to one another in so far as the behavior of (the suitably normalised) $\mathbf{D}_T$ is concerned.

As we will subsequently see, a special case of the large sample results which will be presented in section 4 is that the full-sample test based on $t_{\beta_1}^2$ has non-trivial asymptotic local power against $H_{1,b(\cdot)}$ for both weakly and strongly persistent regressors. This property of the full sample IV-based test statistic obtains through the limiting behaviour of the sample cross-product moment $\mathbf{A}_T$. In particular, its two components are not of the same order of magnitude; therefore, upon normalisation, one of these terms will converges to zero and so all weight is placed on the other instrument. Which instrument gets full weight depends on the persistence of x_{t-1}. The type-II instrument is selected for strongly persistent predictors (i.e., those satisfying Assumption 1.2), while the type-I instrument is selected for weakly persistent predictors (i.e., those satisfying Assumption 1.1); see the proof of Lemma S.6 in the supplementary appendix for details. As a result, regardless of the degree of persistence of the regressor, the appropriate instrument is chosen in the limit.

However, as the simulation results in section 6 demonstrate, the finite sample power of the full sample test can be quite low against such "pocket" alternatives. In the next section we therefore propose tests based on sequences of subsample implementations of the IV-combination test statistic. IV-based techniques are particularly useful to consider because the corresponding subsample-specific statistics may be expressed in terms of partial sums, whose behaviour may in turn be characterised in a tractable manner. This is not the case, for instance, with the test of Campbell and Yogo (2006) or those of Elliott and Müller (2006) and Elliott *et al.* (2015), where the analysis of the joint behavior of subsample-specific statistics is considerably more involved.

4   Subsample IV-Combination Tests for Predictability

Our aim is to develop predictability tests with good power to detect temporary periods of predictability irrespective of whether the putative predictor, x_{t-1} is stable or near-integrated, and which are robust to the presence of heteroskedasticity in the data. To that end, we will base our testing approach on the computation of the IV-combination predictability statistics outlined in the previous section in the context of (3.2) computed not over the full available sample but over various sequences of subsamples of the data. For each such sequence we consider, our proposed test will be based on the maximum (in absolute value) statistic within that sequence. By taking the maximum over these sequences, we therefore base our test on the particular subsample within the given sequence of subsamples where the predictability statistic gives the strongest signal of predictability.

4.1 Choice of Instruments

Before laying out our subsample IV-combination testing approach, we first need to state some regularity conditions which must hold on the type-I and type-II instruments such that we can validly use a testing strategy based on sequences of subsample IV-combination predictability statistics. We will then discuss the choice of instruments to use in practice which satisfy these conditions.

Assumption 4 details the conditions which need to hold on the type-I instrument used.

Assumption 4 *Let $z_{I,t}$ obey the following conditions:*

1. *Under either Assumption 1.1 or Assumption 1.2:* $\mathrm{E}\left(\zeta_t | \zeta_{t-1}, \zeta_{t-2}, \ldots, z_{I,t-1}, z_{I,t-2}, \ldots\right) = 0$, *there exists $\delta_I \geq 0$ such that $T^{-\delta_I} z_{I,t}$ is uniformly L_4-bounded,* $\sup_{\tau \in [0,1]} \left| \frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \right| \xrightarrow{p} 0$, *and* $\sup_{\tau \in [0,1]} \left| \frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t^2 \right| = O_p(1)$.

2. *Under Assumption 1.1, and jointly on $\mathcal{D}$*

   (a) $\frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \xi_{t-1} \Rightarrow K_{z_I x}(\tau)$, *where $K_{z_I x}(\tau)$ is a Hölder-continuous stochastic process of some order $\alpha > 0$ and nonzero w.p.1;*

   (b) $\frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 \xrightarrow{p} K_{z_I^2}(\tau)$, *where $K_{z_I^2}(\tau)$ is a deterministic Hölder-continuous function of some order $\alpha > 0$ and strictly increasing;*

   (c) $\frac{1}{T^{1/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t \Rightarrow G_I(\tau)$, *where $G_I(\tau)$ is a continuous process with independent increments (and therefore, Gaussian), with $G_I(0) = 0$ a.s., zero mean function, strictly increasing variance function $[G_I](\tau)$ and variance profile defined as $\eta_I(\tau) := \frac{[G_I](\tau)}{[G_I](1)}$;*

   (d) $\frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 u_t^2 \xrightarrow{p} [G_I](\tau)$.

3. *Under Assumption 1.2,*

   (a) $\sup_{\tau \in [0,1]} \left| \frac{1}{T^{3/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \xi_{t-1} \right| \xrightarrow{p} 0$;

   (b) $\frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 \xrightarrow{p} K_{z_I^2}(\tau)$ *on $\mathcal{D}$, where $K_{z_I^2}(\tau)$ is a deterministic Hölder-continuous function of some order $\alpha > 0$ and strictly increasing;*

   (c) $\sup_{\tau \in [0,1]} \left| \frac{1}{T^{1/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t \right| = O_p(1)$;

   (d) $\frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{T} z_{I,t-1}^2 u_t^2 = O_p(1)$.

Remark 5. The conditions placed on $z_{I,t-1}$ by Assumption 4 can differ depending on whether Assumption 1.1 or Assumption 1.2 holds. This distinction is germane in cases where $z_{I,t-1}$ is constructed from x_{t-1}; see the examples listed in section 3. In such cases δ_I may take different values for the same instrument, and, similarly, $K_{z_I^2}(\tau)$ may take different shapes under Assumptions 1.1 and 1.2. We do not, however, make this explicit to ease notation. Assumption 4.1 complements the condition in Assumption 3.1 to ensure that the innovations u_t are uncorrelated with the instruments. Assumption 4.2 is new compared to Breitung and Demetrescu (2015), and is required because we explicitly consider the behaviour of the IV-combination statistics under DGPs which can allow for either weak or strong persistence in the (putative) predictors; it requires the instruments to have stochastic properties similar to those of a stable autoregression driven by heteroskedastic innovations. Assumption 4.3 is the analogue of Assumption 3 of Breitung and Demetrescu (2015) but is

considerably less restrictive: rather than the weak convergence of suitably normalised cross-product sample moments required there, we only require uniform boundedness in probability. Assumption 4.3(b) regarding the partial sums of the squared instrument is new, but would appear fairly mild. Our conditions are weaker than those of Breitung and Demetrescu (2015) as we only consider the IV-combination statistic with two instruments, one of type-I and the other of type-II. $\diamond$

Remark 6. Although the weak convergence in Assumption 4.2 is joint, we do not specify the dependence structure between the limiting processes because our asymptotic results will hold irrespective of this structure. We note, however, that the variance profile, $\eta_I(\cdot)$, which turns out to play an important role in our asymptotics under stability (Assumption 1.1) depends on *both* the choice of type-I instrument and the DGP (specifically, on $\mathbf{H}(\cdot)$ and the unconditional variance of u_t); see Lemma 1 for an example. Similarly, the limiting processes $K_{z_I x}(\cdot)$ and $K_{z_I^2}(\cdot)$ also depend on both the DGP and the choice of instrument; again, see Lemma 1 for an example. $\diamond$

Assumption 5 details the corresponding regularity conditions on the type-II instrument.

Assumption 5 *The variable $z_{II,t}$ is deterministic and, for some function $Z(\tau)$, Hölder-continuous of order $\alpha > 1/2$, and some $\delta_{II} \geq 0$, satisfies $T^{-\delta_{II}} z_{II,\lfloor \tau T \rfloor} \Rightarrow Z(\tau)$ in $\mathcal{D}$ where $Z(\cdot)$ is such that, for all $0 \leq \tau_1 < \tau_2 \leq 1$, $\int_{\tau_1}^{\tau_2} \tilde{Z}_{\tau_1,\tau_2}^2(s)\mathrm{d}s \neq 0$ with $\tilde{Z}_{\tau_1,\tau_2}(s) := Z(s) - \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} Z(s)\mathrm{d}s$.*

Remark 7. Notice that the conditions stated in Assumption 5 do not involve the persistence of the regressor because the type-II instruments are exogenous. Assumption 5 essentially coincides with Assumption 4 of Breitung and Demetrescu (2015), up to minor differences. While Assumption 4 of Breitung and Demetrescu (2015) allows for stochastic $z_{II,t}$, it also requires the average cross-products of the instrument and the regression error to have a mixed Gaussian limiting distribution, such that it actually affords little additional flexibility in the choice of type-II instruments relative to Assumption 5. Indeed, under the above assumptions it holds, for example, that $\frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} u_t \Rightarrow \int_0^\tau Z(s)\mathrm{d}U(s)$ which is immediately seen to be a Gaussian process, given that Z is deterministic. However, the quadratic variation process of $\int_0^\tau Z(s)\mathrm{d}U(s)$, $\int_0^\tau Z^2(s)\left(h_{11}^2(s) + h_{12}^2(s)\right)\mathrm{d}s$, is in general nonlinear in τ and depends, analogously to the case of Assumption 1.1, on both the DGP and the choice of the instrument $z_{II,t-1}$. Finally, notice that $Z(\cdot)$ is not permitted to be constant for any of the subsamples over which the test statistics are computed, as this would entail perfect multicollinearity in those subsamples. $\diamond$

We also require further regularity conditions regarding the interaction of the type-I and type-II instruments used. These are now collected in Assumption 6.

Assumption 6 *For instruments $z_{I,t}$ and $z_{II,t}$ satisfying the conditions of Assumptions 4 and 5, respectively, it is also required that: 1. $\sup_{\tau \in [0,1]} \left| \frac{1}{T^{1+\delta_I+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} z_{II,t-1} \right| \overset{p}{\to} 0$; and 2. $\sup_{\tau \in [0,1]} \left| \frac{1}{T^{1+\delta_I+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} z_{II,t-1} u_t^2 \right| = O_p(1)$.*

Remark 8. Breitung and Demetrescu (2015) do not impose such conditions explicitly as they are implied by the stricter set of assumptions under which they work. For instance, Assumption 6.1 would be implied by the weak convergence of the partial sums of $z_{I,t-1}$ in Assumption 3 of Breitung and Demetrescu, but we do not require such weak convergence here because Assumption 4.1 on the

uniform boundedness of the partial sums of the type-I instrument, $\sup_{\tau \in [0,1]} \left| \frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \right| \overset{p}{\to} 0$, suffices for our purposes (and is, for example, implied by Assumption 3 of Breitung and Demetrescu under near-integration). Indeed, Assumption 6.1 only differs through the weights $T^{-\delta_{II}} z_{II,t-1}$, which are deterministic; Assumption 6.2 can be seen as a randomly weighted version thereof, with weights $T^{-\delta_{II}} z_{II,t-1} u_t^2$. Notice that Assumption 6.1 entails that the (appropriately scaled) type-I and type-II instruments are mutually asymptotically orthogonal for all subsamples of the data, $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$, such that $0 \le \tau_1 < \tau_2 \le 1$. $\diamond$

In the context of their full-sample predictability tests, Breitung and Demetrescu (2015) consider the following choice for the type-II instrument, $z_{II,t}$,

$$z_{II,t-1} = \sin\left(\frac{k\pi(t-1)}{2T}\right) \tag{4.1}$$

where k is a positive integer chosen by the practitioner. Breitung and Demetrescu (2015) find that the best performing IV-combination test obtains for $k = 1$ in (4.1). For the type-I instrument we use the IVX approach which has become popular in predictive regressions; see, among others, Gonzalo and Pitarakis (2012); Phillips and Lee (2013) and Kostakis et $al.$ (2015). This entails setting

$$z_{I,t-1} := \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-1-j} \quad \text{with} \quad \varrho := 1 - \frac{a}{T^\gamma} \tag{4.2}$$

for some $a > 0$ and $\gamma \in (0,1)$, with the convention that $\Delta x_0 = 0$. In Lemma 1 we show that these two instruments satisfy the set of conditions required by Assumptions 4–6.[1]

Lemma 1 Let $Assumptions$ 1 and 3 $hold$ $with$ $\boldsymbol{\zeta}_t$ $strictly$ $stationary$ and $ergodic$ $such$ $that,$ for $some$ $\vartheta > 0,$ $\sup_{t \in \mathbb{Z}} \left| \mathrm{E}\left(\left(\tilde{v}_t^2 - \mathrm{E}\left(\tilde{v}_t^2 \right) \right) \tilde{v}_{t-j} \tilde{v}_{t-k} \right) \right| \le C \left(jk \right)^{-1/2 - \vartheta/2}.$ $Then,$ $Assumptions$ 4–6 are $satisfied$ by $\boldsymbol{z}_{t-1} := (z_{I,t-1}, z_{II,t-1})'$ $when$ $z_{II,t-1}$ and $z_{I,t-1}$ are as $defined$ in $(4.1),$ for any $positive$ $integer$ $k,$ and $(4.2),$ $respectively.$ In $particular,$ we $have$

1. $Under$ $Assumption$ $1.1,$ $\delta_I = 0,$ $K_{z_I x}(\tau) = K_{z_I^2}(\tau) = \bar{\sigma}_\xi^2(\rho)[V](\tau)$ and $G_I(\tau)$ is a $time$-$transformed$ $Brownian$ $motion$ $given$ as $T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1} u_t \Rightarrow G_I(\tau),$ $where$ $this$ $weak$ $convergence$ $result$ $holds$ $jointly$ on $\mathcal{D}^3$ $with$ the $weak$ $convergence$ $given$ in $(2.6),$ and

2. $Under$ $Assumption$ $1.2,$ $\delta_I = \gamma/2$ and $K_{z_I^2}(\tau) = \frac{\omega^2}{a}[V](\tau).$

Remark 9. The additional assumptions required to ensure the validity of the IVX instrument are relatively mild. Strict stationarity and ergodicity restrict the weak stationarity of $\boldsymbol{\zeta}_t$ required in Assumption 3 such that the asymptotic behaviour of sample averages can be accounted for, as required for example in Assumption 4.2. The additional condition on the rate of decay of $\mathrm{E}\left(\left(\tilde{v}_t^2 - \mathrm{E}\left(\tilde{v}_t^2 \right) \right) \tilde{v}_{t-j} \tilde{v}_{t-k} \right)$ imposes a form of short memory on the conditional variances. This rate is obviously satisfied when $\mathrm{E}\left(\left(\tilde{v}_t^2 - \mathrm{E}\left(\tilde{v}_t^2 \right) \right) \tilde{v}_{t-j} \tilde{v}_{t-k} \right) = 0$, but is much weaker than that condition and, hence, still allows for asymmetric volatility clustering. $\diamond$

---

[1]We will formally establish this result for only these two instruments which will subsequently be used in both our Monte Carlo study and empirical application. We conjecture, however, that the other examples of type-I and type-II instruments considered in Breitung and Demetrescu (2015, pp.361-362) will also satisfy Assumptions 4–6.

Remark 10. Under Assumption 1.1, the processes $K_{z_I x}(\tau)$ and $K_{z_I^2}(\tau)$ are both proportional to the quadratic variation of $V(\tau)$, the limit process of the suitably normalised partial sums of ξ_t under stability. This demonstrates the usefulness of the IVX instrument in that, under stability, $z_{I,t-1}$ is approximately equal to the stochastic component ξ_{t-1} of the (putative) predictor, x_{t-1}, such that IVX effectively delivers the optimal instrument for x_{t-1} under Assumption 1.1. For a choice of type-I instrument other than IVX this is, in general, not true and one obtains different processes $K_{z_I x}(\tau)$ and $K_{z_I^2}(\tau)$ whose properties depend on the particular choice made; see also Corollary 2 of Breitung and Demetrescu (2015). Our large sample results will, however, be established under Assumptions 4 and 5 and, as such, will hold irrespective of the particular shape or properties of $K_{z_I x}(\tau)$ and $K_{z_I^2}(\tau)$. Furthermore, under Assumption 1.2 the IVX instrument will turn out to be dominated uniformly over all subsamples by the type-II instrument, such that the precise properties of $K_{z_I^2}$ will not be relevant under near-integration.[2] $\diamondsuit$

4.2 Subsample-Based Predictability Tests

For type-I and type-II instruments satisfying Assumptions 4-6, we can proceed to develop subsample implementations of the IV-combination predictability test discussed in section 3. To provide a unified notation for such subsample statistics it will prove useful to define the subsample-specific analogues $\boldsymbol{A}_T(\tau_1, \tau_2)$, $\mathbf{B}_T(\tau_1, \tau_2)$, $\boldsymbol{C}_T(\tau_1, \tau_2)$ and $\mathbf{D}_T(\tau_1, \tau_2)$ of the full-sample quantities $\boldsymbol{A}_T$, $\mathbf{B}_T$, $\boldsymbol{C}_T$ and $\mathbf{D}_T$, respectively, used to construct the standard IV-combination statistic, t_{β_1} of (3.2). These are defined analogously to their full-sample counterparts but for a sample consisting of observations $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$, so that, for example, $\boldsymbol{A}_T(\tau_1, \tau_2) := \sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} \tilde{x}_{t-1} \tilde{\boldsymbol{z}}_{t-1}$ where $\tilde{y}_t$, $\tilde{x}_{t-1}$ and $\tilde{\boldsymbol{z}}_{t-1}$ are now subsample-specific demeaned versions of y_t, x_{t-1} and $\boldsymbol{z}_{t-1}$, respectively, so that, for $\boldsymbol{w}_t$ generically denoting either y_t, x_{t-1} or $\boldsymbol{z}_{t-1}$, $\tilde{\boldsymbol{w}}_t := \boldsymbol{w}_t - \frac{1}{\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor} \sum_{s=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} \boldsymbol{w}_s$. The full-sample quantity is recovered on setting $\tau_1 = 0$ and $\tau_2 = 1$. Precise definitions of these quantities are provided (in partial sum notation) in section S.3.1 of the supplementary appendix.

If it was known that a pocket of predictability might occur over the particular subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$, then it would be logical to compute the subsample IV-combination statistic[3]

$$t_{\beta_1}(\tau_1, \tau_2) := \frac{\boldsymbol{A}_T'(\tau_1, \tau_2)\, \mathbf{B}_T^{-1}(\tau_1, \tau_2)\, \boldsymbol{C}_T(\tau_1, \tau_2)}{\sqrt{\boldsymbol{A}_T'(\tau_1, \tau_2)\, \mathbf{B}_T^{-1}(\tau_1, \tau_2)\, \mathbf{D}_T(\tau_1, \tau_2)\, \mathbf{B}_T^{-1}(\tau_1, \tau_2)\, \boldsymbol{A}_T(\tau_1, \tau_2)}} \tag{4.3}$$

and a test for predictability in this specific subsample could be obtained by comparing $(t_{\beta_1}(\tau_1, \tau_2))^2$ with the $\chi^2(1)$ distribution. Indeed, this would be nothing more than the approach of Breitung and Demetrescu (2015) applied to the particular subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$. Such a test would be expected to have considerably more power to detect a regime of predictability over the subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$ than would the full sample test based on t_{β_1} of (3.2) because the former would be calculated only for sample points where a predictive relationship holds.

In practice it is unlikely the practitioner will know which specific subsample(s) of the data might admit predictive regimes. While some previous applied studies in the literature have considered a

---

[2] It should be noted, however, that $K_{z_I^2}(\tau)$ is also proportional to $[V](\tau)$ under near-integration, albeit with a different constant of proportionality; this is a consequence of the fact that $z_{I,t}$ is mildly integrated in this case.

[3] In the context of $\mathbf{D}_T(\tau_1, \tau_2) := \sum_{t=1}^{T} \tilde{\boldsymbol{z}}_{t-1} \tilde{\boldsymbol{z}}_{t-1}' \tilde{u}_t^2$, the residuals, $\tilde{u}_t^2$, are now the subsample analogues of the full sample residuals, $\hat{u}_t^2$, used in the construction of the full-sample statistic t_{β_1} in (3.2). The three possible choices discussed there can also be used here for the subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$. As with the full sample statistic, these three are asymptotically equivalent in so far as the behaviour of (the suitably normalised) $\mathbf{D}_T(\tau_1, \tau_2)$ is concerned.

14

variety of sample splits and also looked at the evolution of predictive regression statistics over a sequence of subsamples, these studies have tended to signal the presence of a predictive episode based on comparing each of these subsample statistics with the critical value that would apply when running a test for predictability on a single *known* subsample. As discussed in section 1, this induces either multiple testing and/or endogenously determined breakdate problems and, hence, does not deliver size-controlled tests; see, *inter alia*, Inoue and Rossi (2005). In order to control for these issues, the critical value of the test needs to reflect the searching element involved. This can be done by basing one's test on certain functionals of the sequence of subsample predictability statistics considered. Given we are testing the null of no predictability against the alternative of predictability in at least one subsample of the data, an approach based on the maximum of the sequence of subsample predictability statistics considered would seem appropriate. The specific sequences of statistics that we take the maximum over must also be entirely agnostic of the data to avoid any endogenous selection bias; we could not, for example, validly choose to take the maximum statistic from the sequence of subsamples where previous studies had argued predictability holds.

There is an extensive literature on testing for fluctuations in the parameters of linear regression models; see, *inter alia*, Kuan and Hornik (1995). Common choices of agnostic sequences of statistics used include forward and reverse recursive sequences, rolling sequences, and double-recursive sequences. We will adopt these choices here and base our tests on the maximum statistic taken over each of these sequences of statistics. These can be formally defined as follows:

- The sequence of *forward recursive* statistics is given by $\{(t_{\beta_1}(0,\tau))^2\}_{\tau_L \leq \tau \leq 1}$, where the parameter $\tau_L \in (0,1)$ is chosen by the user. The forward recursive regression approach uses $\lfloor T\tau_L \rfloor$ start-up observations, where τ_L is the *warm-in* fraction, and then calculates the sequence of subsample predictive regression statistics $(t_{\beta_1}(0,\tau))^2$ for $t = 1, ..., \lfloor \tau T \rfloor$, with τ travelling across the interval $[\tau_L, 1]$. The statistic formed as the maximum taken across this sequence is then,

$$\mathcal{T}^f := \max_{\tau_L \leq \tau \leq 1} (t_{\beta_1}(0,\tau))^2. \tag{4.4}$$

- The sequence of *backward recursive* statistics is given by $\{(t_{\beta_1}(\tau,1))^2\}_{0 \leq \tau \leq \tau_U}$ with $\tau_U \in (0,1)$ again chosen by the user. In this case one calculates the sequence of subsample predictive regression statistics $(t_{\beta_1}(\tau,1))^2$ for $t = \lfloor \tau T \rfloor + 1, ..., T$, with τ travelling across the interval $[0, \tau_U]$. The maximum statistic from this backward recursive sequence is then,

$$\mathcal{T}^b := \max_{0 \leq \tau \leq \tau_U} (t_{\beta_1}(\tau,1))^2. \tag{4.5}$$

- The sequence of *rolling* statistics is given by $\{(t_{\beta_1}(\tau,\tau+\Delta\tau))^2\}_{0 \leq \tau \leq 1-\Delta\tau}$ where the user-defined parameter $\Delta\tau \in (0,1)$. The rolling regression approach calculates the sequence of subsample statistics $t_{\beta_1}(\tau,\tau+\Delta\tau))^2$ for $t = \lfloor \tau T \rfloor + 1, ..., \lfloor \tau T \rfloor + \lfloor T\Delta\tau \rfloor$, where $\Delta\tau$ is the window fraction with $\lfloor T\Delta\tau \rfloor$ the window width, with τ travelling across the interval $[0, 1-\Delta\tau]$. The maximum statistic from this rolling sequence is then,

$$\mathcal{T}^r := \max_{0 \leq \tau \leq 1-\Delta\tau} (t_{\beta_1}(\tau,\tau+\Delta\tau))^2. \tag{4.6}$$

- Finally, the *double-recursive* sequence of statistics is given by $\{(t_{\beta_1}(\tau_1, \tau_2))^2\}_{\substack{0 \leq \tau_1, \tau_2 \leq 1 \\ \tau_2 - \tau_1 \geq \Delta\tau}}$, where $\Delta\tau \in (0,1)$ is again a user-defined parameter. The double-recursive approach calculates a double indexed sequence of subsample statistics $(t_{\beta_1}(\tau_1, \tau_2))^2$ for $t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor$, for all subsamples such that $0 \leq \tau_1 < \tau_2 \leq 1$ and where $\tau_2 - \tau_1 \geq \Delta\tau$. Notice that this entails that the forward recursive sequence discussed above is calculated across all possible warm-in fractions such that $\tau_L \geq \Delta\tau$, which is why this sequence is referred to as double-recursive.[4] The maximum statistic from the double-recursive sequence is then,

$$\mathcal{T}^d := \max_{\substack{0 \leq \tau_1, \tau_2 \leq 1 \\ \tau_2 - \tau_1 \geq \Delta\tau}} (t_{\beta_1}(\tau_1, \tau_2))^2. \tag{4.7}$$

Remark 11. The full sample IV-combination statistic $t_{\beta_1}^2$ of (3.2) is contained within the forward recursive sequence of statistics and obtains by setting $\tau = 1$, and similarly is contained within the backward recursive sequence for $\tau = 0$. It is also contained within the double-recursive sequence for $\tau_1 = 0$ and $\tau_2 = 1$. Notice also that if we set $\Delta\tau = 1$ in the context of the rolling sequence then this would collapse to the single full sample statistic, $t_{\beta_1}^2$. ◇

Tests based on the maximum from each of the foregoing sequences of subsample statistics have particular patterns of local predictability that they will be well designed to detect. Tests based on the forward recursive sequence of statistics are designed to detect pockets of predictability which start at or near the start of the full sample period available to the practitioner. The longer the duration of such an episode the more powerful these tests will be, other things being equal, because they are based on a sequence of increasing subsamples all starting from the first data point. By analogy, tests based on the reverse recursive sequence of subsample statistics are designed to detect end-of-sample pockets of predictability. As such, reverse recursive based tests could therefore usefully be employed in an on-going monitoring exercise for the emergence of predictive regimes. Because both the forward and reverse recursive sequences, and indeed the double-recursive sequence, contain the usual full sample predictability statistic, regardless of the choice of the trimming parameters, they also deliver tests which have power to detect predictability which holds over the whole sample, although in this particular case they would not be expected to be as powerful as the standard full sample IV-combination test which is clearly designed for that specific alternative hypothesis.

For a given window width, tests based on a rolling sequence of statistics are designed to pick up a window of predictability, of (roughly) the same length, within the data. As discussed above, the double-recursive sequence amounts to considering all possible window width rolling sequences, subject to a minimum window width. These then are useful for picking up multiple predictive regimes, of potentially different lengths, within the data. However, because the double-recursive sequence considers such a large number of possible subsamples of the data a test based on the maximum from this sequence would necessarily be expected to be less powerful than the recursive or rolling-based tests in scenarios for which the latter are designed. This is because the more statistics one considers in a sequence over which the maximum is taken the stricter the critical value needs to be to maintain a correctly sized test. So, for example, in the case where a pocket of predictability existed in the middle of the sample data of length say m observations, a test based on the maximum

---

[4]Notice that this double sequence also obtains by calculating the rolling sequence discussed above for all possible rolling window widths between $\Delta\tau$ and 1 inclusive.

from the rolling sequence using a window width of m observations would be expected to be more powerful than a test based on the maximum of the double-recursive sequence because the critical value for the latter would be considerably larger than the former. However, a power advantage over the double-recursive test would not necessarily be expected to hold for the corresponding rolling tests where the window width was either smaller than m or greater than m. In the former case this would be because the maximum subsample length available over which predictability held (m observations) could never be utilised because the window width is less than m, while in the latter case all subsamples in the sequence will contain a mix of data points where predictability holds and where it does not. It is of course very hard to analytically predict what the relative finite sample power properties of the recursive, rolling and double recursive based tests tests will be in cases like these and so we will investigate these further using Monte Carlo experimentation in section 6.

Before establishing the asymptotic properties of the maximum subsample statistics, it is worth briefly commenting on estimation of the location of any predictive windows in cases where our proposed tests reject. Even for the simplest possible case where $H_{1,b(\cdot)}$ of (2.4) implies predictability over just a single subsample of the data, say $t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor$, with $\tau_1 < \tau_2$ and where either $\tau_1 > 0$ or $\tau_2 < 1$, consistent estimation of τ_1 and τ_2 is not possible because of the Pitman localisation to zero placed on $\beta_{1,t}$ in this interval by Assumption 2. In practice, however, if a given maximum statistic rejects then a sensible estimate of τ_1 and τ_2 would be given by the start and end points of the subsample corresponding to the maximum value from the sequence of statistics from which a rejection was obtained. If one was looking to date possibly multiple windows of predictability then one could reapply the procedures outlined above to the data set excluding those sample points for which a first stage rejection occurred, and do so repeatedly until no rejection was obtained.

4.3   Asymptotic Distributions

In Proposition 1 we now provide representations for the asymptotic distributions of the maximum subsample statistics defined in section 4.2 under the appropriate local alternative, $H_{1,b(\cdot)}$.

Proposition 1 *Consider the model in (2.1) and (2.2) and let Assumptions 2 – 6 hold. Then under the local alternative $H_{1,b(\cdot)}$ of (2.4):*

(i) *Under Assumption 1.1, as $T \to \infty$, it holds that,*

$$\mathcal{T}^f \Rightarrow \sup_{\tau \in [\tau_L, 1]} \frac{\left(G_I(\tau) + \int_0^\tau b(s)\mathrm{d}K_{z_I x}(s)\right)^2}{[G_I](1)\,\eta_I(\tau)}$$

$$\mathcal{T}^b \Rightarrow \sup_{\tau \in [0, \tau_U]} \frac{\left(G_I(1) - G_I(\tau) + \int_\tau^1 b(s)\mathrm{d}K_{z_I x}(s)\right)^2}{[G_I](1)(1 - \eta_I(\tau))}$$

$$\mathcal{T}^d \Rightarrow \sup_{\substack{0 \le \tau_1, \tau_2 \le 1 \\ \tau_2 - \tau_1 \ge \Delta\tau}} \frac{\left(G_I(\tau_2) - G_I(\tau_1) + \int_{\tau_1}^{\tau_2} b(s)\mathrm{d}K_{z_I x}(s)\right)^2}{[G_I](1)(\eta_I(\tau_2) - \eta_I(\tau_1))}$$

$$\mathcal{T}^r \Rightarrow \sup_{0 \le \tau \le 1 - \Delta\tau} \frac{\left(G_I(\tau + \Delta\tau) - G_I(\tau) + \int_\tau^{\tau+\Delta\tau} b(s)\mathrm{d}K_{z_I x}(s)\right)^2}{[G_I](1)(\eta_I(\tau + \Delta\tau) - \eta_I(\tau))}$$

where $G_I(\cdot)$, $[G_I](\cdot)$, $K_{z_I x}(\cdot)$ and $\eta_I(\cdot)$, are as defined in Assumption 4.2.

(ii) *Under Assumption 1.2, as $T \to \infty$, it holds that,*

$$\mathcal{T}^f \Rightarrow \sup_{\tau_L \leq \tau \leq 1} \frac{\left(\int_0^\tau \tilde{Z}_{0,\tau}(s) \mathrm{d}U(s) + \omega \int_0^\tau \tilde{Z}_{0,\tau}(s) b(s) J_{c,\mathrm{H}}(s) \mathrm{d}s \right)^2}{\int_0^\tau \tilde{Z}_{0,\tau}^2(s) \mathrm{d}[U](s)}$$

$$\mathcal{T}^b \Rightarrow \sup_{0 \leq \tau \leq \tau_U} \frac{\left(\int_\tau^1 \tilde{Z}_{\tau,1}(s) \mathrm{d}U(s) + \omega \int_\tau^1 \tilde{Z}_{\tau,1}(s) b(s) J_{c,\mathrm{H}}(s) \mathrm{d}s \right)^2}{\int_\tau^1 \tilde{Z}_{\tau,1}^2(s) \mathrm{d}[U](s)}$$

$$\mathcal{T}^d \Rightarrow \sup_{\substack{0 \leq \tau_1, \tau_2 \leq 1 \\ \tau_2 - \tau_1 \geq \Delta\tau}} \frac{\left(\int_{\tau_1}^{\tau_2} \tilde{Z}_{\tau_1,\tau_2}(s) \mathrm{d}U(s) + \omega \int_{\tau_1}^{\tau_2} \tilde{Z}_{\tau_1,\tau_2}(s) b(s) J_{c,\mathrm{H}}(s) \mathrm{d}s \right)^2}{\int_{\tau_1}^{\tau_2} \tilde{Z}_{\tau_1,\tau_2}^2(s) \mathrm{d}[U](s)}$$

$$\mathcal{T}^r \Rightarrow \sup_{0 \leq \tau \leq 1 - \Delta\tau} \frac{\left(\int_\tau^{\tau+\Delta\tau} \tilde{Z}_{\tau,\tau+\Delta\tau}(s) \mathrm{d}U(s) + \omega \int_\tau^{\tau+\Delta\tau} \tilde{Z}_{\tau,\tau+\Delta\tau}(s) b(s) J_{c,\mathrm{H}}(s) \mathrm{d}s \right)^2}{\int_\tau^{\tau+\Delta\tau} \tilde{Z}_{\tau,\tau+\Delta\tau}^2(s) \mathrm{d}[U](s)}$$

where $J_{c,\mathrm{H}}(\cdot)$, $U(\cdot)$ and $[U](\cdot)$ are defined in section 2, and $\tilde{Z}_{\cdot,\cdot}(\cdot)$ is defined in Assumption 5.

Remark 12. Expressions for the limiting null distributions of the statistics can be obtained by omitting those terms involving the function $b(\cdot)$ from the representations given in Proposition 1. In what follows we will denote the resulting limiting null distributions of the $\mathcal{T}^f$, $\mathcal{T}^b$, $\mathcal{T}^d$ and $\mathcal{T}^r$ statistics under Assumption 1.1 as $\mathcal{T}_\infty^{f,I}$, $\mathcal{T}_\infty^{b,I}$, $\mathcal{T}_\infty^{d,I}$ and $\mathcal{T}_\infty^{r,I}$, respectively, and under Assumption 1.2 as $\mathcal{T}_\infty^{f,II}$, $\mathcal{T}_\infty^{b,II}$, $\mathcal{T}_\infty^{d,II}$ and $\mathcal{T}_\infty^{r,II}$, respectively. For any given $s \in \{f, b, d, r\}$, the limiting distribution $\mathcal{T}_\infty^{s,I}$, appropriate for the case where the predictor is weakly persistent satisfying Assumption 1.1, and $\mathcal{T}_\infty^{s,II}$, the corresponding limiting null distribution where the predictor is strongly persistent satisfying Assumption 1.2, have different functional forms. For example, while $\mathcal{T}_\infty^{s,II}$, $s = f, b, d, r$, all depend on the choice of type-II instrument, $\mathcal{T}_\infty^{s,I}$, $s = f, b, d, r$, do not. The impact of non-constancy in $\mathbf{H}(\cdot)$ on these limiting null distributions also differs between the strongly and weakly persistent cases; see the discussion in Remarks 15 and 16 below. $\diamond$

Remark 13. All of the statistics in the sequences are exact invariant to both μ_x and β_0 by virtue of being based on subsample demeaned variables. Moreover, the vector of instruments used is, by construction, invariant to μ_x, because $z_{I,t}$ is based on *differences* of x_t for the instruments mentioned in section 3, and $z_{II,t}$ is a deterministic function of time chosen by the user *without* reference to μ_x. Consequently the limiting representations in Proposition 1 do not depend on either μ_x or β_0. $\diamond$

Remark 14. Under Assumption 1.1, local power depends indirectly on the persistence of the putative predictor, as measured by ρ and $B(L)$ through $K_{z_Ix}(\cdot)$; see Lemma 1 for the particular example of the IVX instrument. Under Assumption 1.2, while the mean-reversion parameter c does not affect the limiting null behaviour of the maximum statistics, the local power functions depend explicitly on c through $J_{c,\mathrm{H}}(\cdot)$. In each case, the rule-of-thumb that the stronger the mean reversion, the lower the local power, seems to hold; see the Monte Carlo results in section 6. $\diamond$

For weakly persistent regressors, a time transformation can shed further light on the influence of heteroskedasticity. Under Assumption 4.2(c), the process $W(\cdot) := G_I(\eta^{-1}(\cdot))/\sqrt{[G_I](1)}$ is continuous with stationary independent increments, $W(0) = 0$ a.s. and $\mathrm{Var}(W(\tau)) = \tau$, and therefore, $W(\cdot)$ is a standard Wiener process. It follows that $G_I(\cdot) = \sqrt{[G_I](1)} W(\eta_I(\cdot))$ is a time-transformed Wiener process. Consequently, taking the limiting functional associated with $\mathcal{T}^f$ as an example,

we have that $\sup_{\tau\in[\tau_L,1]} \frac{\left(G_I(\tau)+\int_0^\tau b(s)\mathrm{d}K_{z_Ix}(s)\right)^2}{[G_I](1)\,\eta_I(\tau)} \overset{d}{=} \sup_{\tau\in[\tau_L,1]} \frac{\left(W(\eta_I(\tau))+\int_0^\tau b(s)\mathrm{d}\frac{K_{z_Ix}(s)}{\sqrt{[G_I](1)}}\right)^2}{\eta_I(\tau)}$ with similar distributional identities holding for the remaining statistics. As the maximum of a function is invariant to monotonic transformations of the argument, we may set $r=\eta_I(\tau)$ and therefore obtain the following alternative representations of the limiting results in part (i) of Proposition 1.

Corollary 1 *Let the conditions of Proposition 1 hold. Then under Assumption 1.1, as $T\to\infty$,*

$$\mathcal{T}^f \Rightarrow \sup_{r\in[\eta_I(\tau_L),1]} \frac{\left(W(r)+\int_0^{\eta_I^{-1}(r)} b(s)\mathrm{d}\bar{K}_{z_Ix}(s)\right)^2}{r}$$

$$\mathcal{T}^b \Rightarrow \sup_{r\in[0,\eta_I(\tau_U)]} \frac{\left(W(1)-W(r)+\int_{\eta_I^{-1}(r)}^1 b(s)\mathrm{d}\bar{K}_{z_Ix}(s)\right)^2}{1-r}$$

$$\mathcal{T}^d \Rightarrow \sup_{\substack{0\le r_1,r_2\le 1 \\ \eta_I^{-1}(r_2)-\eta_I^{-1}(r_1)\ge\Delta\tau}} \frac{\left(W(r_2)-W(r_1)+\int_{\eta_I^{-1}(r_1)}^{\eta_I^{-1}(r_2)} b(s)\mathrm{d}\bar{K}_{z_Ix}(s)\right)^2}{r_2-r_1}$$

$$\mathcal{T}^r \Rightarrow \sup_{0\le r\le\eta_I(1-\Delta\tau)} \frac{\left(W\left(\eta_I\left(\eta_I^{-1}(r)+\Delta\tau\right)\right)-W(r)+\int_{\eta_I^{-1}(r)}^{\eta_I^{-1}(r)+\Delta\tau} b(s)\mathrm{d}\bar{K}_{z_Ix}(s)\right)^2}{\eta_I\left(\eta_I^{-1}(r)+\Delta\tau\right)-r}$$

where $W(\cdot):=G_I(\eta^{-1}(\cdot))/\sqrt{[G_I](1)}$ is a standard Wiener process and $\bar{K}_{z_Ix}(\cdot):=K_{z_Ix}(\cdot)/\sqrt{[G_I](1)}$. Moreover, the limiting null distributions discussed in Remark 12 are such that,

$$\mathcal{T}^{f,I}_\infty \overset{d}{=} \sup_{r\in[\eta_I(\tau_L),\,1]} \frac{(W(r))^2}{r}, \qquad \mathcal{T}^{b,I}_\infty \overset{d}{=} \sup_{r\in[0,\,\eta_I(\tau_U)]} \frac{(W(1)-W(r))^2}{1-r}$$

$$\mathcal{T}^{r,I}_\infty \overset{d}{=} \sup_{0\le r\le\eta_I(1-\Delta_\tau)} \frac{\left(W\left(\eta_I\left(\eta_I^{-1}(r)+\Delta_\tau\right)\right)-W(r)\right)^2}{\eta_I\left(\eta_I^{-1}(r)+\Delta_\tau\right)-r}$$

$$\mathcal{T}^{d,I}_\infty \overset{d}{=} \sup_{\substack{0\le r_1,r_2\le 1 \\ \eta_I^{-1}(r_2)-\eta_I^{-1}(r_1)\ge\Delta\tau}} \frac{(W(r_2)-W(r_1))^2}{r_2-r_1}.$$

Remark 15. The results in Proposition 1 and Corollary 1 highlight that both the limiting null distributions and the local power functions of all of the tests depend, in general, on any unconditional heteroskedasticity present through the resulting non-constancy of $\mathbf{H}(\cdot)$. This holds irrespective of the persistence of the regressor x_t; moreover, heteroskedasticity has differing effects on the limiting distributions depending on the degree of persistence of x_t. At least under the null this may seem surprising, as Eicker-White standard errors are designed to robustify any of the subsample statistics, $0\le\tau_1<\tau_2\le 1$, to heteroskedasticity (conditional or unconditional). However, this asymptotic invariance only holds *marginally* for a given statistic in the sequence; indeed, it can be shown for each of the sequences of statistics, and regardless of which of Assumption 1.1 and Assumption 1.2 holds, any given statistic in the sequence has a marginal χ_1^2 limiting null distribution. The representations in Corollary 1, for example, show that under Assumption 1.1 the suprema are taken over statistics computed for various intervals whose endpoints depend on the variance profile $\eta_I(\cdot)$ defined in Assumption 4.2, which depends in turn on both the DGP

and the choice of type-I instrument. Moreover, under Assumption 1.2, the same phenomenon explains part (ii) of Proposition 1, with the additional complication that one cannot represent the subsample statistics more tractably using a time transformation due to the presence of the *subsample-demeaned* process Z. Here, too, heteroskedasticity depends on the choice of instrument (now the type-II instrument) in addition to the DGP. Under local alternatives, heteroskedasticity additionally enters by means of $K_{z_I x}(\cdot)$ and $J_{c,\mathrm{H}}(\cdot)$, under Assumption 1.1 and Assumption 1.2, respectively. It is important to emphasise that the precise effect of non-constancy of $\mathbf{H}(\cdot)$ due to unconditional heteroskedasticity on the limiting distributions of our maximum statistics depends on which of Assumptions 1.1 or 1.2 holds. $\diamondsuit$

Remark 16. More generally, the impact of the DGP on the large sample behaviour of the statistics depends on the choice of instrument and on the persistence of the (putative) predictor. Consider first the results under Assumption 1.1. Here the limiting null distributions, $\mathcal{T}_\infty^{s,I}$, $s = f, b, d, r$, all depend on $\eta_I(\cdot)$ which in turn depends on the unconditional variance of u_t. In the case where $\eta_I(s) = s$, these limiting null distributions simplify to the suprema of squared standardised Wiener processes taken over the range of the subsamples. However, constancy of $\mathbf{H}(\cdot)$ is not sufficient to ensure linearity of $\eta_I(\cdot)$, because heteroskedasticity can still enter via the instrument $z_{I,t-1}$. Under the local alternative, the key quantity controlling power is $K_{z_I x}(\cdot)$ which can be deterministic under Assumption 1.1 (see, for example, Lemma 1 for the case of the IVX instrument), and (upon normalisation) characterises the strength of the instrument $z_{I,t-1}$. However, $K_{z_I x}(\cdot)$ also characterises the signal; other things equal, if x_t has a large marginal variance relative to u_t, then local power will increase. In the case of Assumption 1.2, local power depends on the process $J_{c,\mathrm{H}}(\cdot)$ in a more intricate way, due to the fact that $J_{c,\mathrm{H}}$ and $\int \tilde{Z} \mathrm{d}U$ may be dependent. Clearly, local power is influenced by all three factors c, ω and $\mathbf{H}(\cdot)$. The effect of the elements of $\mathbf{H}(\cdot)$ is not easy to disentangle, as can be seen from the expressions given for the quadratic variation processes of $U(\cdot)$ and $V(\cdot)$ at the end of section 2. $\diamondsuit$

In Corollary 2 we detail the limiting distributions of the full sample statistic $t_{\beta_1}^2$ of (3.2) under the local alternative, $H_{1,b(\cdot)}$ of (2.4).

Corollary 2 *Let the conditions of Proposition 1 hold. Then under $H_{1,b(\cdot)}$, as $T \to \infty$: (i) Under Assumption 1.1, $t_{\beta_1}^2 \Rightarrow \left(W(1) + \int_0^1 b(s)\mathrm{d}\bar{K}_{z_I x}(s)\right)^2$; (ii) under Assumption 1.2, $t_{\beta_1}^2 \Rightarrow \left(\int_0^1 \tilde{Z}^2(s)\mathrm{d}[U](s)\right)^{-1} \left(\int_0^1 \tilde{Z}(s)\mathrm{d}U(s) + \omega \int_0^1 \tilde{Z}(s)b(s)J_{c,\mathrm{H}}(s)\mathrm{d}s\right)^2$, where $\tilde{Z}(s) := Z(s) - \int_0^1 Z(s)$.*

Remark 17. From Corollary 2, under the null hypothesis H_0 of (2.3), $t_{\beta_1}^2 \Rightarrow W(1)^2$ under Assumption 1.1, while $t_{\beta_1}^2 \Rightarrow \left(\int_0^1 \tilde{Z}^2(s)\mathrm{d}[U](s)\right)^{-1} \left(\int_0^1 \tilde{Z}(s)\mathrm{d}U(s)\right)^2$ under Assumption 1.2 with $\int_0^1 \tilde{Z}(s)\mathrm{d}U(s) \sim N\left(0, \int_0^1 \tilde{Z}^2(s)\mathrm{d}[U](s)\right)$. Consequently, $t_{\beta_1}^2$ is seen to possess a standard χ_1^2 limiting null distribution regardless of whether x_t is stable or near-integrated. Moreover, the results in Corollary 2 show that the full sample IV-combination test exhibits non-trivial power against the class of local alternatives we consider in this paper; that is, it has power to detect predictive episodes. However, local power depends indirectly on heteroskedasticity which influences the stochastic properties of $K_{z_I x}(\cdot)$ and $J_{c,\mathrm{H}}(\cdot)$; see Remarks 15 and 16. $\diamondsuit$

Remark 18. Where $\beta_{1,t} = \beta_1 \neq 0$, for all $t = 1, ..., T$, the results in Corollary 2 specialise to the standard local power of the full sample IV-combination test based on $t_{\beta_1}^2$. For type-II instruments without demeaning, one recovers the result of Breitung and Demetrescu (2015, Theorem 2.2). $\diamondsuit$

To summarise, the limiting null distributions of the maximum subsample statistics all depend both on any heteroskedasticity present and on whether the putative predictor x_t is a near-integrated or weakly dependent process. This poses significant problems for conducting inference not encountered with tests based on the full sample statistic, $t_{\beta_1}^2$ of (3.2). We next demonstrate that these issues can be solved using fixed regressor wild bootstrap implementations of the subsample tests.

5 Bootstrap Implementation

As the results in the previous section show, implementing tests based on the $\mathcal{T}^s$, $s = f, b, d, r$, statistics from section 4.2 will require us to address the fact that their limiting null distributions depend on any unconditional heteroskedasticity present in u_t and v_t, and on whether the predictor x_{t-1} is weakly dependent or near-integrated. To account for the former we employ a wild bootstrap resampling scheme applied to the demeaned dependent variable $\hat{y}_t := y_t - \frac{1}{T} \sum_{t=1}^{T} y_t$, while for the latter we use the observed outcomes on $\mathbf{x} := [x_0, x_1, ..., x_T]'$ and $\mathbf{z} := [\mathbf{z}_0', \mathbf{z}_1', ..., \mathbf{z}_T']'$ as a fixed regressor and fixed instrument vector, respectively, when implementing the bootstrap procedure.

We now outline our fixed regressor wild bootstrap approach in Algorithm 1. We will then demonstrate the asymptotic validity of this approach in Proposition 2.

Algorithm 1

Step 1 *Construct the wild bootstrap innovations $y_t^* := \hat{y}_t R_t$, where $\hat{y}_t := y_t - \frac{1}{T} \sum_{t=1}^{T} y_t$ are the demeaned sample observations on y_t, and R_t, $t = 1, ..., T$, is an IID $N(0,1)$ sequence independent of the data.[5]*

Step 2 *Using the bootstrap sample data $\left(y_t^*, x_{t-1}, \mathbf{z}_{t-1}'\right)'$, in place of the original sample data $\left(y_t, x_{t-1}, \mathbf{z}_{t-1}'\right)'$, construct the bootstrap analogues of the statistics $\mathcal{T}^s$, $s = f, b, d, r$, from section 4.2. Denote these bootstrap statistics as $\mathcal{T}^{s*}$, $s = f, b, d, r$.*

Step 3 *Define the bootstrap p-values as $P_T^{s,*} := 1 - G_T^{s,*}(\mathcal{T}^s)$, $s = f, b, d, r$, with $G_T^{s,*}(\cdot)$ denoting the conditional (on the original data) cumulative distribution function (cdf) of $\mathcal{T}^{s*}$, $s = f, b, d, r$. In practice, the $G_T^{s,*}(\cdot)$, $s = f, b, d, r$, will be unknown, but can be simulated in the usual way by repeating Steps 1 and 2 a large number, say B, times to obtain empirical analogues of $G_T^{s,*}(\cdot)$, $s = f, b, d, r$. The $\{R_t\}_{t=1}^{T}$ variables used in Step 1 must also be independent across the B bootstrap replications.*

Step 4 *The wild bootstrap test of the null hypothesis H_0 of (2.3) at level α based on $\mathcal{T}^s$ rejects if $P_T^{s,*} \leq \alpha$, $s = f, b, d, r$.*

Remark 19. The bootstrap statistics $\mathcal{T}^{s*}$, $s = f, b, d, r$, are calculated treating both x_{t-1} and the vector of instruments, $\mathbf{z}_{t-1}$, as fixed; i.e., they are calculated using the *same* observed x_{t-1} and $\mathbf{z}_{t-1}$ as were used in the construction of $\mathcal{T}^s$, $s = f, b, d, r$. This aspect is crucial for delivering bootstrap tests that are asymptotically valid regardless of whether x_t satisfies Assumption 1.1 or 1.2 and without knowledge of which of these holds. In particular, the same instrument (either type-I or type-II, depending on the true regressor persistence) gets full asymptotic weight in both the original 2SLS and the bootstrap t-ratios (see section S.3 of the Supplement). $\diamond$

---

[5] The Gaussianity assumption on R_t is standard in the literature and simplifies the proof of Proposition 2 below. This can, however, be generalised such that R_t is any IID sequence with $\mathrm{E}(R_t) = 0$, $\mathrm{E}(R_t^2) = 1$ and $\mathrm{E}(R_t^4) < \infty$.

Remark 20. The wild bootstrap generating y_t^* in Step 1 of Algorithm 1 replicates the pattern of unconditional heteroskedasticity present in the original innovations, as conditionally on $\hat{y}_t$, y_t^* is independent over time with zero mean and variance $\hat{y}_t^2$. Moreover, any heteroskedasticity present in x_{t-1} and $\boldsymbol{z}_{t-1}$ is replicated through the fixed regressor/instrument aspect of the bootstrap statistics. In particular, as u_t is a MD sequence, it is anticipated that the bootstrap will replicate the variance properties of either $z_{I,t-1}u_t$ or $z_{II,t-1}u_t$, depending on the degree of persistence exhibited by x_t. Having fixed the regressor and the instruments when bootstrapping, the analogous terms in the bootstrap test statistics are given by $z_{I,t-1}\hat{y}_t R_t$ and $z_{II,t-1}\hat{y}_t R_t$ with variances $z_{I,t-1}^2\hat{y}_t^2$ and $z_{II,t-1}^2\hat{y}_t^2$, respectively. Using the result detailed in the last sentence of Remark 19, it is then seen that the correct variance profile is replicated in the limit. For full details see the proof of Proposition 2. $\diamond$

Remark 21. Step 1 of Algorithm 1 is based on residuals obtained under the null hypothesis. It is straightforward to show that the large sample properties of corresponding bootstrap tests based on either the OLS or IVX residuals from estimating the predictive regression over the full sample are unaltered from those given here. Moreover, albeit more computationally intensive, one could also use the analogous subsample implementations of any of these three full sample residuals. $\diamond$

In Proposition 2 we now demonstrate the large sample validity of the fixed regressor wild bootstrap implementation of the tests from section 4.2. In particular, we show that our proposed bootstrap in Algorithm 1 correctly replicates the first order asymptotic null distributions of the statistics given in Remark 12 under both the null hypothesis and local alternatives.

Proposition 2 *Let the conditions of Proposition 1 hold. Then, as $T \to \infty$, under either the null hypothesis H_0 of (2.3) or the local alternative $H_{1,b(\cdot)}$ of (2.4): (i) under Assumption 1.1, as $T \to \infty$, it holds that $\mathcal{T}^{f*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{f,I}$, $\mathcal{T}^{b*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{b,I}$, $\mathcal{T}^{r*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{r,I}$, and $\mathcal{T}^{d*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{d,I}$; (ii) under Assumption 1.2 as $T \to \infty$, it holds that, $\mathcal{T}^{f*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{f,II}$, $\mathcal{T}^{b*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{b,II}$, $\mathcal{T}^{r*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{r,II}$, and $\mathcal{T}^{d*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{d,II}$.*

A consequence of Proposition 2 is that we obtain asymptotically correctly sized tests when using bootstrap critical values obtained using Algorithm 1. We now formalise this result in Corollary 3.

Corollary 3 *As $T \to \infty$, under H_0, $P^*(\mathcal{T}^{s*} \leq \mathcal{T}^s) \Rightarrow Unif[0,1]$, for each of $s = f,b,d,r$, where P^* denotes probability conditional on the original sample $(y_t, x_{t-1}, \boldsymbol{z}_{t-1}')'$, $t = 1,\ldots,T$.*

Remark 22. Corollary 3 establishes the asymptotic validity of our proposed bootstrap tests. This result holds without knowledge of whether the (putative) predictor x_t satisfies Assumption 1.1 or 1.2, and holds regardless of any heteroskedasticity present in u_t and v_t satisfying Assumption 3. $\diamond$

Remark 23. Proposition 2 shows that each of the bootstrap statistics $\mathcal{T}^{s*}$, $s = f,b,d,r$, attains the same first order limiting distribution under both the null hypothesis and local alternatives as that attained under the null hypothesis by the corresponding original (non-bootstrap) statistic $\mathcal{T}^s$, $s = f,b,d,r$. An immediate consequence of this is that each of the wild bootstrap tests proposed in Algorithm 1 will admit the same asymptotic local power function as the (infeasible) size-adjusted test based on the corresponding original statistic $\mathcal{T}^s$, $s = f,b,d,r$. $\diamond$

Remark 24. The full sample IV-combination statistic, $t_{\beta_1}^2$, of Breitung and Demetrescu uses Eicker-White standard errors to correct the limiting null distribution of the statistic for non-constancy in $\mathbf{H}(\cdot)$ due to unconditional heteroskedasticity in the innovations. Because it is still

necessary to implement the subsample maximum tests using a wild bootstrap it would be feasible to replace the Eicker-White standard errors used in the computation of the subsample $\left(t_{\beta_1}(\tau_1, \tau_2)\right)^2$ statistic in (4.3) and its bootstrap equivalent, computed in Step 2 of Algorithm 1, with conventional standard errors. While this would alter the limiting representations given for the maximum statistics in Proposition 1 and Corollary 1, it can be shown that the resulting wild bootstrap tests would still be asymptotically valid with an analogous result to that in Corollary 3 holding. In this case the wild bootstrap tests would attain the same asymptotic local power functions as (infeasible) size-corrected implementations of the (non-bootstrap) maximum tests based on conventional standard errors. These asymptotic local power functions will not in general coincide with those obtained for the statistics based on Eicker-White standard errors, but they would where $\mathbf{H}(\cdot)$ is constant. $\quad\diamond$

Remark 25. The bootstrap validity results given in this section also apply to a fixed regressor wild bootstrap implementation of the full sample IV-combination test based on $t_{\beta_1}^2$. In particular, this will satisfy a result of the form given in Corollary 3 and will have the same asymptotic local power function as the test based on $t_{\beta_1}^2$ using χ_1^2 critical values, discussed in section 4.3. As with the discussion for the subsample maximum statistics in Remark 24, one could replace Eicker-White standard errors with conventional standard errors without losing asymptotic validity. $\quad\diamond$

6   Numerical Results

We use Monte Carlo simulation methods to investigate the finite sample performance of the bootstrap implementations of the subsample-based predictability tests $\mathcal{T}^f$, $\mathcal{T}^b$, $\mathcal{T}^r$ and $\mathcal{T}^d$ proposed in section 4 for testing the null hypothesis of no predictability in (2.3); i.e., $H_0 : \beta_{1,t} = 0$, for all $t = 1, \ldots, T$, against the alternative $H_{1,b(\cdot)}$ of (2.4) that predictability holds across some subset of the sample data. Data are generated from (2.1)-(2.2). In section 6.1 we explore the empirical size properties of these tests comparing with the corresponding full sample IV-combination test of Breitung and Demetrescu (2015), $t_{\beta_1}^2$ of (3.2). In section 6.2 we compare the finite sample local power properties of these tests against a variety of DGPs displaying temporary predictability.

Following the discussion in section 4.1, we base both the full sample t_{β_1} IV-combination statistic in (3.2) and the corresponding subsample $t_{\beta_1}(\tau_1, \tau_2)$ statistics in (4.3) on the instrument vector $\mathbf{z}_{t-1} := (z_{I,t-1}, z_{II,t-1})'$ with the type-II instrument, $z_{II,t-1}$, defined as in (4.1) with $k = 1$, and the type-I instrument, $z_{I,t-1}$, given by the IVX choice of Kostakis *et al.* (2015) defined as in (4.2), with $a = 1$ and $\gamma = 0.95$.[6] Excepting the IVX instrument, $z_{I,t-1}$, all variables and instruments entering the estimated predictive regressions are demeaned, as in the main text. As discussed in Kostakis *et al.* (2015, p.1514) the IVX instrument, $z_{I,t-1}$, does not need to be demeaned as the slope estimator in the predictive regression is invariant to whether $z_{I,t-1}$ is demeaned or not. In order to correct for the finite sample effects of estimating the intercept term in (2.1), which are most pronounced for highly persistent regressors which are strongly correlated with the predictive model's innovations, Kostakis *et al.* (2015, p.1516) recommend the use of a finite-sample correction factor. We also found that this correction factor led to significant improvements in the finite sample properties of our proposed tests and hence is implemented in all of the numerical and empirical results we report.

---

[6]We also considered tests based on using the fractionally integrated instrument suggested on page 363 of Breitung and Demetrescu (2015) for $z_{I,t-1}$. We do not report these results here as the IVX choice performed better in our results, but they can be obtained from the authors on request.

All simulations are preformed in MATLAB, versions R2018a and R2018b, using the Mersenne Twister random number generator. All results pertain to the nominal 5% level; qualitatively similar results were obtained for other conventional significance levels. All of the subsample tests are computed according to Algorithm 1 using 399 bootstrap replications; the bootstrap tests are denoted $\mathcal{T}^{s*}$, $s = f, b, d, r$. Following Banerjee *et al.* (1992), we set $\tau_L = 1/4$ and $\tau_U = 3/4$ in the context of the forward and backward recursive statistics, respectively, and $\Delta\tau = 1/3$ for the rolling and double recursive statistics. The empirical size simulations were based on 5000 Monte Carlo replications and the local power simulations on 1000 replications, with the exception of the double recursive tests where 1000 replications were used for size and 500 for power because of the much higher computing time required. For the full sample $t^2_{\beta_1}$ test, results for versions based on the asymptotic χ^2_1 critical value and on a fixed regressor wild bootstrap are reported, the latter using 399 bootstrap replications. For all of the bootstrap tests, two versions are reported. The first is based on statistics using Eicker-White standard errors while for the second, following the discussion in Remark 24, conventional standard errors are used. These two variants are distinguished apart by the additional "*NW*" nomenclature in the subscripts of the latter. Following the discussion in section 3 and footnote 3, all of the reported statistics use residuals, $\hat{u}_t$, computed under the null hypothesis.

6.1 Empirical Size

We first investigate the finite sample size properties of our proposed tests. To that end, we consider the simulation DGP given by (2.1)-(2.2) with $\beta_{1,t} = \beta_1 = 0$ for all $t = 1, ..., T$. Results are reported for $T = 250$ and $T = 500$. In generating the simulation data we set the intercepts β_0 and μ_x in (2.1) and (2.2), respectively, to zero with no loss of generality. The autoregressive process characterising the dynamics of the putative predictor, x_t, in (2.2) was initialised at $\xi_0 = 0$. Results are reported for a range of values of the autoregressive parameter ρ in (2.2) that cover both stationary and persistent predictors; in particular, for $\rho := 1 - c/T$ we consider $c \in \{0, 2.5, 5, 10, 20, 0.5T\}$. Notice that $c = 0.5T$ corresponds to $\rho = 0.5$, such that the autoregressive parameter is fixed and stable.

In our simulation DGP the innovation vector $(u_t, v_t)'$ is drawn from an i.i.d. bivariate Gaussian distribution with mean zero and covariance matrix $\mathbf{\Sigma}_t := \begin{bmatrix} \sigma^2_{ut} & \phi\sigma_{ut}\sigma_{vt} \\ \phi\sigma_{ut}\sigma_{vt} & \sigma^2_{vt} \end{bmatrix}$. Notice, therefore, that ϕ corresponds to the correlation between the innovations u_t and v_t. Results are reported in Table 1 for the case where $\phi = 0$, and in Table 2 for the case where $\phi = -0.90$.[7] We report results for the case where the innovations are homoskedastic, $\sigma^2_{ut} = \sigma^2_{vt} = 1$ (labelled DGP1 in the Tables 1 and 2), and for the case where there is a contemporaneous one-time break of equal magnitude in the variances of u_t and v_t. Following the simulations designs considered in Georgiev *et al.* (2018a), two such heteroskedastic cases are considered: (i) an upward change in variance (labelled DGP2 in Tables 1 and 2) such that $\sigma^2_{ut} = \sigma^2_{vt} = 1\mathbb{I}(t \leq \lfloor 0.5T \rfloor) + 4\mathbb{I}(t > \lfloor 0.5T \rfloor)$, and (ii) a downward change (labelled DGP3 in Tables 1 and 2) where $\sigma^2_{ut} = \sigma^2_{vt} = 1\mathbb{I}(t \leq \lfloor 0.5T \rfloor) + \frac{1}{4}\mathbb{I}(t > \lfloor 0.5T \rfloor)$, where in each case $\mathbb{I}(\cdot)$ denotes the indicator function, taking the value one when its argument is true and zero otherwise. DGP2 and DGP3 allow us to examine the impact of unconditional heteroscedasticity, both in isolation and in its interaction with ϕ, on the finite sample size of the tests. In each of DGP2 and DGP3 a fourfold change in variance is seen which is likely to be of

---

[7]In predictive regression models for the equity premium employing valuation ratios as predictors (e.g. the dividend-price ratio, earnings-price ratio), as we shall do in the empirical application in section 7, the relevant innovation terms are strongly negatively correlated, hence our choice of $\phi = -0.90$.

rather larger magnitude than we might expect to see in practice, but serves to illustrate how the tests behave in the presence of a large change in unconditional volatility. We also considered further DGPs allowing for stationary GARCH(1,1) with different degrees of persistence coupled with either Gaussian or t-distributed innovations, thereby allowing for unconditionally heteroskedastic and fat-tailed innovations. These results were qualitatively similar to those reported here for DGP1 and can be found in the Supplement.

Consider first the results pertaining to the homoskedastic DGP1. A comparison of the results in Table 1 for $\phi = 0$ and Table 2 for $\phi = -0.90$ shows that, in the homoskedastic case at least, the correlation parameter ϕ has relatively little impact on the size properties of the tests. For the full sample tests there is relatively little difference between the tests based on the asymptotic χ_1^2 critical value and the fixed regressor wild bootstrap. Similarly, as might be expected, there is little to choose between the versions of the full sample tests with Eicker-White standard errors and those with conventional standard errors. For the subsample tests, there is a general trend towards undersizing in the Eicker-White versions in cases where the putative predictor, x_{t-1}, displays persistence at or close to a unit root process. This is most pronounced in the rolling and double recursive statistics. However, this undersizing is not seen with the versions of the subsample tests based on conventional standard errors. It is well known that Eicker-White standard errors can be heavily downward biased in small samples leading to incorrectly sized tests; see, for example, MacKinnon and White (1985).

We next turn to the results for the two unconditionally heteroskedastic DGPs, DGP2 and DGP3. Consider first the full sample tests. As expected, the full sample test based on conventional standard errors and the asymptotic χ_1^2 critical value, $t_{\beta_1,NW}^2$, is unreliable in the presence of heteroskedasticity. These size distortions are considerably worse for $\phi = -0.90$ than for $\phi = 0$ when $c = 0$; for the other values of c considered the differences between $\phi = 0$ and $\phi = -0.90$ are much smaller. The size distortions observed with $t_{\beta_1,NW}^2$ are significantly ameliorated by the use of Eicker-White standard errors $(t_{\beta_1}^2)$ in all but the case of DGP2 with $c = 0$ where no apparent improvements are seen. The bootstrap implementations of the full sample tests do a much better job at controlling finite sample size, regardless of whether Eicker-White or conventional standard errors are used, although some over-sizing is still seen for $\phi = -0.90$ when $c = 0$. There appears to be no need to use Eicker-White standard errors with the fixed regressor bootstrap implementation of the full sample test.

Consider next the subsample predictability tests. Undersizing, in many cases substantial, is again seen with the subsample bootstrap tests based on Eicker-White standard errors. As with the full sample tests, these effects tend to be larger, other things equal, for $\phi = -0.90$ *vis-à-vis* $\phi = 0$. As with the results for DGP1, the subsample bootstrap tests based on conventional standard errors are much less prone to this undersizing phenomenon, albeit some undersizing is seen in the case of DGP2 with $\phi = -0.90$ for small values of c, most notably for the rolling and double recursive tests. Moreover, under DGP3 with $\phi = -0.90$ some oversizing is seen in the persistent x_{t-1} cases for the backward recursive, rolling and double recursive tests. For $\phi = 0$ all of the subsample bootstrap tests implemented with conventional standard errors appear to display good finite sample size control.

6.2 Finite Sample Local Power

We now turn to an investigation into the relative finite sample local power properties of the tests. We again generate simulation data from DGP (2.1)–(2.2) but now for a variety of local alternatives satisfying $H_{1,b(\cdot)}$ of (2.4). To keep the set of results to a manageable level we report results only for

$\phi = -0.90$, for the homoskedastic case, $\sigma_{ut}^2 = \sigma_{vt}^2 = 1$, for a sample of size $T = 250$ and for three values of the persistence parameter, c, associated with x_t; specifically, $c = \{0, 10, 0.5T\}$. In all of our experiments the slope parameter β_{1t} in (2.1) is set to be local-to-zero. As specified by Assumption 2, for $c = 0$ and $c = 10$, where x_t is strongly persistent, we parameterise the slope parameter in (2.1) as $\beta_{1t} = b_{1t}/T$, and here we consider the following values of the Pitman drift parameter, $b_{1t} \in \{0, 5, ..., 80\}$. For the case of a weakly dependent predictor, $c = 0.5T$, we parameterise the slope parameter as $\beta_{1t} = b_{1t}/\sqrt{T}$, and here we consider the Pitman drift values $b_{1t} \in \{0, 1, ..., 21\}$.

We report results for three distinct experimental cases, where episodes of predictability occur once in the sample either at the beginning, the end or within the sample. To that end, we consider the following three simulation DGPs, with the range of non-zero values of b_{1t} as outlined above,

Case 1:
$$\begin{cases} b_{1t} > 0 & \text{for} \quad t = 1, ..., \lfloor T/5 \rfloor \\ b_{1t} = 0 & \text{for} \quad t = \lfloor T/5 \rfloor + 1, ..., T \end{cases}$$
Case 2:
$$\begin{cases} b_{1t} = 0 & \text{for} \quad t = 1, ..., \lfloor 4T/5 \rfloor \\ b_{1t} > 0 & \text{for} \quad t = \lfloor 4T/5 \rfloor + 1, ..., T \end{cases}$$

Case 3:
$$\begin{cases} b_{1t} = 0 & \text{for} \quad t = 1, ..., \lfloor T/5 \rfloor \\ b_{1t} > 0 & \text{for} \quad t = \lfloor T/5 \rfloor + 1, ..., \lfloor 3T/5 \rfloor \\ b_{1t} = 0 & \text{for} \quad t = \lfloor 3T/5 \rfloor + 1, ..., T \end{cases}$$

All other aspects of the simulation design are as described previously.

Figures 1–3 graph the simulated finite sample local power curves for each of Cases 1–3, respectively. Each figure contains power curves for the fixed regressor wild bootstrap implementations of the full sample $t_{\beta_1, NW}^{2*}$ test along with the subsample-based predictability tests $\mathcal{T}_{NW}^{f*}$, $\mathcal{T}_{NW}^{b*}$, $\mathcal{T}_{NW}^{r*}$ and $\mathcal{T}_{NW}^{d*}$. To aid presentation of the graphs, we have chosen only to report the versions of the bootstrap tests implemented with conventional standard errors. Results with Eicker-White standard errors are available on request. In general the latter were less powerful (often considerably so) than the reported tests based on conventional standard errors.

Consider first the results pertaining to Case 1 in Figure 1. Recall from section 4.2 that the temporary predictability DGP in Case 1, with a pocket of predictability at the start of the sample, is one where we expect the forward recursive $\mathcal{T}_{NW}^{f*}$ test to perform best. Figure 1 bears out this prediction. Regardless of the value of c, $\mathcal{T}_{NW}^{f*}$ is significantly more powerful than the other tests considered. The double recursive test, $\mathcal{T}_{NW}^{d*}$, also displays significant power gains over the full sample $t_{\beta_1, NW}^{2\star}$ test, for all of the values of c considered. The rolling test, $\mathcal{T}_{NW}^{r*}$, displays a similar power profile to $\mathcal{T}_{NW}^{d*}$ for $c = 0$ and $c = 0.5T$, but is significantly less powerful than $\mathcal{T}_{NW}^{d*}$ for $c = 10$. The least powerful tests among those considered is the backward recursive test, as expected, and the full sample $t_{\beta_1, NW}^{2\star}$ test. To illustrate, the empirical power of $t_{\beta_1, NW}^{2\star}$ at $b_T = 50$ is approximately 50% for both $c = 0$ and $c = 10$ while for $\mathcal{T}_{NW}^{f*}$ it is around 75%. For $c = 0.5T$ and $b_T = 10$ the power of $t_{\beta_1, NW}^{2}$ is about 55% while that of $\mathcal{T}_{NW}^{f*}$ is in excess of 95%. In the latter example both the rolling ($\mathcal{T}_{NW}^{r*}$) and double recursive ($\mathcal{T}_{NW}^{d*}$) tests have power of approximately 80%.

Consider next the results for Case 2, given in Figure 2, where the pocket of predictability now occurs at the end of the sample. When x_t is weakly persistent the simulation DGP is approximately time-reversible and, as such, we would anticipate that all but the forward and backward recursive tests, whose relative behaviour would be expected to switch around, will behave similarly to how they behaved in Case 1 for the weakly dependent case. This is clearly seen to be the case in Figure 2(c), with the backward recursive test now clearly the most powerful, the forward recursive test

26

the least powerful, and the other tests all displaying almost identical power properties in Figures 1(c) and 2(c). These patterns are also seen, albeit not as clearly, in a comparison of Figures 1(b) and 2(b); the main difference being that the most of the tests (although not the double recursive test) tend to be slightly more powerful for $c = 10$ *vis-à-vis* $c = 0.5T$. The pattern of a general increase in power of the tests as c decreases for an end-of-sample pocket of predictability is very clearly continued in Figure 2(a) for the case where $c = 0$ and x_t follows a pure unit root. Here, comparing with Figure 1(a), we see that all of the tests display considerably higher local power against an end-of-sample pocket of predictability than against a pocket of predictability at the start of the sample, and, comparing with Figures 2(b) and 2(c), that the power of the tests is considerably higher than for $c = 10$ and $c = 0.5T$. A possible explanation for this improvement in power is the shape of the non-centrality term, $\int_{\tau_1}^{\tau_2} \tilde{Z}_{\tau_1, \tau_2}(s) b(s) J_{c,\mathrm{H}}(s) \mathrm{d}s$, entering the limiting distributions of the statistics under local alternatives in the case where x_t is strongly persistent. Clearly, end of sample predictability will be boosted from the larger magnitude of $J_{c,\mathrm{H}}(\tau)$ when τ is close to 1, and this will be most evident when $c = 0$. Interestingly, the full sample $t_{\beta_1, NW}^{2\star}$ test displays competitive power in Figure 2(a) although it should be recalled from Table 2 that $t_{\beta_1, NW}^{2\star}$ is significantly over-sized in this case while the subsample tests are not.

Finally, the results in Figure 3 pertain to Case 3, where the simulation DGP admits a window of predictability of size $\lfloor 2T/5 \rfloor$ within the sample. Here the double recursive test, $\mathcal{T}_{NW}^{d*}$, displays superior power to the other tests considered for both $c = 10$ and $c = 0.5T$ (Figures 3(b) and 3(c) respectively), and is jointly most powerful along with the forward recursive $\mathcal{T}_{NW}^{f*}$ test for $c = 0$ (Figure 3(a)). Notice also that for a given value of c, $\mathcal{T}_{NW}^{d*}$ displays considerably higher power under Case 3 than it does under both Cases 1 and 2. This is expected given that a larger window of predictive data is now present in the sample which the double recursive procedure is best able to exploit. Indeed, most of the tests considered display improved power performance compared to Figures 1 and 2. This is particularly evident for the rolling test, $\mathcal{T}_{NW}^{r*}$, and again is to be expected given that a greater number of the subsample predictability statistics in the rolling sequence will contain data from a predictive period relative to the DGPs in Cases 1 and 2. For Case 3, the $\mathcal{T}_{NW}^{b*}$ test (as expected, given that the window of predictability begins early in the sample) and the full sample $t_{\beta_1, NW}^{2}$ test display the lowest power among the tests considered.

7   Empirical Application

The dataset used consists of monthly observations on the equity premium for the S&P Composite index calculated using CRSP's month-end values together with 14 different putative predictors, generically denoted x_t, and are taken from the updated monthly data set on Amit Goyal's website (`www.hec.unil.ch/agoyal/`) which is an extended version of the data set used by Welch and Goyal (2008). The data cover the period 1950:01-2017:12 ($T = 817$). We define the equity premium as in Goyal and Welch (2003) as the log return on the value-weighted CRSP stock market index minus the log return on the risk-free Treasury bill: $y_t = ep_t = \log(1 + R_{m,t}) - \log(1 + R_{f,t})$ where $R_{m,t}$ is the CRSP return and $R_{f,t}$ is the Treasury bill return. The variables are in log form (as in Goyal and Welch, 2003) and each of the predictors is lagged one period. A full list of the predictors together with graphs of the excess returns and the predictors can be found in the supplementary appendix.

Table 3 reports the outcome of the conventional IV-combination test from bivariate predictive regression models applied to the full sample of data. We report versions of the statistic using

Eicker-White ($t^2_{\beta_1}$) and conventional ($t^2_{\beta_1,NW}$) standard errors. All of the IV-based test statistics computed in the empirical analysis follow the same specification as was used in the Monte Carlo experiments; that is, they are based on a combination of the IVX instrument, $z_{I,t-1}$ (as defined in (4.2), with $a = 1$ and $\gamma = 0.95$), and the sine instrument, $z_{II,t-1}$ (as defined in (4.1) with $k = 1$), with all of the observed variables and $z_{II,t-1}$ (but not $z_{I,t-1}$) entering the estimated predictive regressions demeaned, and with the finite-sample correction factor of Kostakis *et al.* (2015,p.1516) implemented. Fixed regressor wild bootstrap p-values computed according to Algorithm 1 with 999 bootstrap replications are reported in parentheses. For most of the putative predictors considered, the results in Table 3 yield no statistically significant evidence of predictability. Exceptions are seen for the treasury bill rate (tbl_{t-1}), the long term government bond yield and rate of return series (lty_{t-1} and ltr_{t-1}, respectively), and inflation ($infl_{t-1}$) all of which are significant at the 5% level. Rejections of the null of no predictability are also seen at the 10% level for the term spread (tms_{t-1}) and the equity premium volatility ($rvol_{t-1}$) series.

To provide an insight into how stable the full sample predictive regressions are, Table 3 also reports the tests proposed in Georgiev *et al.* (2018b) for the stability of the slope coefficient in the bivariate predictive regression of the equity premium on each (lagged) predictor. These tests are denoted LM_x and $supF_x$. The former is designed to test for the stability of the slope coefficient against a smoothly evolving slope change model and the latter against a one-time change in the slope. Bootstrap p-values calculated as outlined in Georgiev *et al.* (2018b) for 999 bootstrap replications are reported in parentheses. Significant rejections at the 5% level by at least one of these tests are observed for the predictive regressions involving the dividend price ratio (dp_{t-1}), dividend yield (dy_{t-1}), earnings price ratio (e/p_{t-1}), book to market ratio (bm_{t-1}), term spread (tms_{t-1}) and $infl_{t-1}$. The rejections seen for dp_{t-1} and e/p_{t-1} are particularly strong. A rejection at the 10% level is also seen for the net equity expansion ratio ($ntis_{t-1}$) predictor. Interestingly, for three of the four series (tbl_{t-1}, lty_{t-1} and ltr_{t-1}) for which the full sample IV-combination tests are significant at the 5% level these stability tests provide no evidence of structural instability in the slope coefficient.

To provide some additional insight into any time-varying behaviour present in the slope coefficients, Figures 4 and 5 plot forward recursive and rolling IV (using the same choice of instruments as detailed above for the full sample IV-combination tests) slope estimates from the predictive regression of y_t on x_{t-1} and associated approximate 95% marginal confidence bands.[8] The warm-in fraction for the recursive sequence, τ_L, and the rolling window fraction, $\Delta\tau$, were both set at 1/4. In each case the horizontal axis dates correspond to the end of a given subsample. Commensurate with the results of the stability tests of Georgiev *et al.* (2018b), these graphs highlight considerable time variation in the sequences of subsample slope estimates. A general pattern evident in Figure 4 is a decline over time in the absolute value of the estimated slope coefficient with the recursive slope estimates generally tending to move closer to zero over time. This pattern can also be seen, albeit less clearly, in the rolling estimates in Figure 5. This suggests that for some of these variables, any predictive ability they might have for the equity premium weakens over time. As a further heuristic device, rather than a formal statistical test, many of the graphs show some periods where the 95% marginal confidence intervals do not include zero, which is at least suggestive that pockets

---

[8]Denoting the IV slope estimate as $\hat{\beta}_1$, the confidence bands were computed as $\hat{\beta}_1 \pm 1.96se(\hat{\beta}_1)$, where $se(\hat{\beta}_1)$ are the associated IV Eicker-White standard errors. These confidence bands should, however, be treated with caution as they are not joint 95% confidence bands for the entire sequence of slope estimates, but rather represent the marginal 95% confidence band at each point in the sequences of estimated slope coefficients.

of predictability may be present in the data. Most of these episodes occur nearer the start of the data, such as, for example, with dy_{t-1}, but some are much longer lived as with, for example, the sequences of recursive estimates for tbl_{t-1}, ltr_{t-1}, tms_{t-1} and $infl_{t-1}$; recall that for tbl_{t-1}, ltr_{t-1} and $infl_{t-1}$, the full sample IV-combination tests gave significant rejections at the 5% level.

To pursue these findings further using statistically rigorous size-controlled methods, we next apply our proposed subsample-based predictability statistics. We report versions of the statistics using Eicker-White ($\mathcal{T}^f$, $\mathcal{T}^b$, $\mathcal{T}^r$ and $\mathcal{T}^d$) and conventional ($\mathcal{T}^f_{NW}$, $\mathcal{T}^b_{NW}$, $\mathcal{T}^r_{NW}$ and $\mathcal{T}^d_{NW}$) standard errors. Fixed regressor wild bootstrap p-values computed according to Algorithm 1 with 999 bootstrap replications are again reported in parentheses. In the computation of the forward and backward recursive statistics we set $\tau_L = 1/4$ and $\tau_U = 3/4$, respectively, while we set $\Delta\tau = 1/4$ for the rolling and double recursive statistics. The instruments used are as described above for the full sample statistics. Focusing on the forward recursive tests we see significant rejections at the 5% level (or stricter) of the null hypothesis of no predictability for each of dp_{t-1}, dy_{t-1}, e/p_{t-1}, de_{t-1}, tbl_{t-1}, lty_{t-1}, ltr_{t-1}, tms_{t-1} and $infl_{t-1}$; indeed, in many cases these rejections are also significant at the 1% level. While these rejections tally with those delivered by the full sample test for tbl_{t-1}, lty_{t-1}, ltr_{t-1} and $infl_{t-1}$, for the other series, all of which (other than de_{t-1}) fail the structural stability tests of Georgiev *et al.* (2018b), these are series for which the full sample tests delivered no significant evidence of predictability. With the exception of dp_{t-1} and e/p_{t-1}, those series for which $\mathcal{T}^f_{NW}$ delivers a rejection at the 5% significance level also show rejections at the 5% level for at least one of the other subsample maximum tests reported. Additional evidence of temporary predictability at the 5% level (or stricter) is provided for dfy_{t-1} by both $\mathcal{T}^r_{NW}$ and $\mathcal{T}^d_{NW}$ (notice that for this series the $supF_x$ test is in fact very close to giving a rejection at the 10% level). A significant rejection at the 10% level is also provided for $ntis_{t-1}$ by the $\mathcal{T}^r_{NW}$ test.

To gain further insight, Figure 6 graphs the forward recursive sequences of $t_{\beta_1}(\tau_1, \tau_2)$ subsample statistics for each case where a rejection at the 5% level is observed for the corresponding maximum test[9]. Also reported on these graphs are the 5% and 10% bootstrap critical values for the null distribution of the maximum statistic in the sequence, together with the 5% and 10% critical values from the χ^2_1 distribution (the marginal critical values which apply for any given subsample).

Consider first the graph in part (a) of Figure 6 for the dividend price ratio, dp_{t-1}. Looking at the time path of the forward recursive subsample statistic we can see that for much of the first half of the sequence (up until roughly the early 1980s) the statistic exceeds the χ^2_1 5% critical value, suggesting that running the IV-combination test on any subsample of the data selected up until this point would have delivered a significant rejection at the (marginal) 5% level. After this sample endpoint no significant evidence of predictability would have been found. We can also see that a large number of exceedances of the 10% bootstrap critical value for the maximum are seen in the early part of the data, with exceedances of the 5% bootstrap critical value also seen, most notably in the mid 1970s. These results are suggestive that a pocket of predictability for returns existed for the predictor dp_{t-1} in the 1970s with peak predictability seen in the middle of that decade, and that since the 1980s onwards predictability appears to have evaporated. For the dividend yield, dy_{t-1}, a pocket of predictability appears to be present again from the early 1970s but lasting much longer, and with apparently stronger magnitude, displaying many more contiguous exceedances of the bootstrap critical values for the maximum than were seen for dp_{t-1}; indeed, here predictability

---

[9]Where both the maximum tests based on Eicker-White and conventional standard errors reject we report the version with the smallest p-value; cf. Table 3. Corresponding graphs for the rolling sequences are available on request.

appears to run until the early to mid 1990s. From the mid to late 1990s onwards the evidence for predictability disappears. Evidence for both the earnings price ratio, e/p_{t-1}, in part (c) and the dividend pay out ratio, de_{t-1}, in part (d) is less strong than for the previous two series (reflected in the considerably larger p-values for the maximum statistics for those series in Table 3), but again the period of predictability appears to be concentrated in mid 1970s. For both the treasury bill rate, tbl_{t-1}, in part (e) and the long term bond yield, lty_{t-1}, in part (f) there appears to be evidence of predictability across a window from the early 1970s until the mid 1980s, albeit the strength of predictability appears to waver somewhat over this period, particularly so for lty_{t-1}. For both of these series, there is also evidence that predictability is re-emerging from around the period of the recent financial crisis onwards, most notably so for tbl_{t-1} where a number of exceedances of the bootstrap critical values occur. In the case of tbl_{t-1} running the IV-combination test on almost any subsample of the data would yield a rejection at the 5% using the marginal χ^2_1 critical value. This observation is also true for the long term rate, ltr_{t-1}, in part (g) and for inflation, $infl_{t-1}$, in part (i). Recall that these are the three series for which the full sample IV-combination tests gave significant rejections at the 5% level. Finally for the tms_{t-1} series in part (h) predictability appears evident and consistently strong up until the mid 1990s after which the magnitude of predictability starts to tail off and then falls markedly around the time of the financial crisis onwards. In contrast, the full sample tests reveal no significant evidence (at the 5% level) of predictability from tms_{t-1}.

These examples highlight the advantage of considering the recursive sequence of statistics and their evolution through time rather than just full sample IV-combination tests, with much stronger evidence for predictability earlier in the sample than later for a number of the predictors considered.

8   Conclusions

Recent research has suggested that should stock returns be predictable, then this is likely to be a temporary phenomenon. Our motivation has been to develop tests with good power to detect such episodes. To avoid the problem of endogenously-determined sample splits, our proposed tests are derived from sequences of predictability statistics calculated over systematic subsamples of the data. The tests are based on the maxima of the instrumental variable-based predictability statistics of Breitung and Demetrescu (2015) taken across sequences of forward and backward recursive, rolling, and double-recursive predictive regressions. The limiting distributions of these statistics were shown to depend both on any heteroskedasticity present and on whether the putative predictor follows a near-integrated or weakly dependent process. To account for these dependencies, fixed regressor wild bootstrap implementations of the tests were proposed and shown to be first-order asymptotically valid. Monte Carlo simulation demonstrated that the tests display decent finite sample size control, and can be considerably more powerful in detecting temporary predictability than full sample tests. An empirical application to a well-known US monthly stock returns data set highlighted the ability of the new tests to detect predictability within the data where full sample tests could not.

We conclude with two suggestions for further research. First, we have focussed on tests based on subsample implementations of the IV-combination statistics of Breitung and Demetrescu (2015) which use two instruments per predictor. It should be possible to apply the same approach to sub-sample implementations of statistics which use only one instrument, such as the statistics considered in section 2.2 of Breitung and Demetrescu (2015) or the IVX statistic of Kostakis *et al.* (2015). Second, our proposed tests are based on an approach which assumes a linear predictive regression

model with a constant slope parameter within the given subsample window and then bases a test on the fluctuations seen in the sequence of such statistics over a range of subsamples. As such, this approach is ambivalent about the true form of any time-variation present in the slope parameter and so would be expected to have reasonable power against a wide range of patterns of time-variation in the slope parameter, including those generated by threshold or other non-linear DGPs. LM-type tests could be developed based on an assumed non-linear model for the time-variation in the slope and would be expected to be more powerful than the tests developed here where this assumed model coincided with, or was at least a close approximation to, the true (unknown) DGP, but would likely have much lower power if the true DGP was not well approximated by the model.

References

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

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

Banerjee, A., R. L. Lumsdaine and J. H. Stock (1992). Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence. *Journal of Business and Economic Statistics* 10, 271–87.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance* 55, 225–264.

Boswijk, H. P., Cavaliere, G., Rahbek, A., and Taylor, A. M. R. (2016). Inference on co-integration parameters in heteroskedastic vector autoregressions. *Journal of Econometrics* 192, 64–85.

Boudoukh, J. R., M. Michaely, P. Richardson and M. R. Roberts (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *Journal of Finance* 62, 877–915.

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

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics* 18, 373–400.

Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics/Revue canadienne d'économique* 41, 1–21.

Campbell, J. Y. and R. J. Shiller (1988a). Stock prices, earnings, and expected dividends. *The Journal of Finance* 43, 661–676.

Campbell, J. and R. J. Shiller (1988b). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195–228.

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

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

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics* 106, 157–181.

Elliott, G. and J. H. Stock (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory* 10, 672–700.

Elliott, G. and U. K. Müller (2006). Efficient tests for general persistent time variation in regression coefficients. *The Review of Economic Studies* 73, 907-940.

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

Fama, E. F. (1981). Stock returns, real activity, inflation, and money. *American Economic Review* 71, 545–565.

Fama, E. F. (1990). Stock returns, expected returns, and real activity. *Journal of Finance* 45, 1089–1108.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3–24.

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25, 23–49.

Farmer, L., L. Schmidt and A. G. Timmermann (2018). Pockets of predictability. *CEPR Discussion Paper*, DP12885.

Gargano, A., D. Pettenuzzo and A. Timmermann (2017). Bond return predictability: economic value and links to the macroeconomy. *Management Science* forthcoming.

Georgiev, I., D. I. Harvey, S. J. Leybourne and A. M. R. Taylor (2018a). A bootstrap stationarity test for predictive regression invalidity. *Journal of Business and Economic Statistics* forthcoming.

Georgiev, I., D. I. Harvey, S. J. Leybourne and A. M. R. Taylor (2018b). Testing for parameter instability in predictive regression models. *Journal of Econometrics* 204, 101–118.

Giannetti, A., (2007). The short term predictive ability of earnings-price ratios: The recent evidence (1994-2003). *The Quarterly Review of Economics and Finance* 47, 26–39.

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

Gorodnichenko, Y., A. Mikusheva, and S. Ng (2012). Estimators for persistent and possibly non-stationary data with classical properties. *Econometric Theory* 28, 1003–1036.

Goyal, A. and I. Welch (2003). Predicting the equity premium with dividend ratios. *Management Science* 49, 639–654.

Henkel,S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics* 99, 560–580.

Inoue, A. and B. Rossi (2005). Recursive predictability tests for real-time data. *Journal of Business & Economic Statistics* 23, 336–345.

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

Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17, 357–390.

Kolev, G.I. and R. Karapandza (2017). Out-of-sample equity premium predictability and sample-split invariance. *Journal of Banking and Finance* 34, 188–201.

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

Kuan, C.-M. and K. Hornik (1995). The generalized fluctuation test: A unifying view. *Econometric Reviews* 14, 135–161.

Lettau, M. and S. Ludvigsson (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56, 815–850.

Lettau, M. and S. van Nieuwerburgh (2008). Reconciling the return predictability evidence. *The Review of Financial Studies* 21, 1607–1652.

MacKinnon, J. G. and H. White (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29, 305–325.

Nelson C. R. and M. J. Kim (1993). Predictable stock returns: The role of small sample bias. *Journal of Finance* 48, 641–661.

Nicolau, J. and P. M. M. Rodrigues (2019). A new regression-based tail index estimator. *The Review of Economics and Statistics* 101(4), 667–680.

Pástor, L. and R. F. Stambaugh (2009). Predictive systems: Living with imperfect predictors. *Journal of Finance* 64, 1583–1628.

Pástor, L. and R. F. Stambaugh (2012). Are stocks really less volatile in the long run? *Journal of Finance* 67, 431–478.

Paye, B. S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance* 13, 274–315.

Pesaran, M. H. and A. Timmermann (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance* 9, 495–510.

Pettenuzzo, D. and A. Timmermann (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics* 164, 60–78.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica* 66, 1299–1325.

Phillips, P. C. B. (2015). Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* 13, 521–555.
(1990).

Phillips, P. C. B. and J. H. Lee (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics* 177, 250–264.

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

Rapach, D. E. and M. Wohar (2006). Structural breaks and predictive regression models of aggregate U.S. stock returns. *Journal of Financial Econometrics* 4, 238–274.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics* 54, 375–421.

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting* 24, 1–18.

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

Table 1: Empirical Rejection Frequencies under the Null Hypothesis, H_0. Nominal 5% significance level. DGP1-DGP3 with $\phi = 0$.

c	$t^{2*}_{\beta_1}$	$t^{2*}_{\beta_1,NW}$	$t^2_{\beta_1}$	$t^2_{\beta_1,NW}$	$\mathcal{T}^{f*}$	$\mathcal{T}^{b*}$	$\mathcal{T}^{f*}_{NW}$	$\mathcal{T}^{b*}_{NW}$	$\mathcal{T}^{r*}$	$\mathcal{T}^{r*}_{NW}$	$\mathcal{T}^{d*}$	$\mathcal{T}^{d*}_{NW}$
DGP1: $T = 250$, $\phi = 0$ and $\sigma^2_{ut} = \sigma^2_{vt} = 1$
0.0	0.045	0.046	0.046	0.045	0.037	0.036	0.049	0.051	0.017	0.057	0.015	0.067
2.5	0.047	0.047	0.049	0.045	0.032	0.032	0.048	0.049	0.013	0.053	0.008	0.059
5.0	0.043	0.044	0.043	0.041	0.031	0.032	0.047	0.044	0.009	0.050	0.012	0.062
10.0	0.042	0.043	0.040	0.039	0.033	0.039	0.046	0.048	0.014	0.050	0.015	0.057
20.0	0.049	0.048	0.046	0.044	0.039	0.039	0.044	0.048	0.025	0.050	0.028	0.056
0.5T	0.049	0.047	0.046	0.047	0.058	0.053	0.052	0.047	0.061	0.045	0.070	0.044
DGP2: $T = 250$, $\phi = 0$ and $\sigma^2_{ut} = \sigma^2_{vt} = 1\mathbb{I}(t \le \lfloor 0.5T \rfloor) + 4\mathbb{I}(t > \lfloor 0.5T \rfloor)$
0.0	0.047	0.048	0.032	0.059	0.040	0.044	0.057	0.050	0.024	0.052	0.020	0.048
2.5	0.044	0.047	0.036	0.066	0.033	0.033	0.053	0.048	0.018	0.055	0.009	0.059
5.0	0.043	0.043	0.035	0.068	0.036	0.033	0.056	0.046	0.015	0.054	0.008	0.061
10.0	0.046	0.045	0.037	0.078	0.037	0.038	0.057	0.048	0.018	0.054	0.020	0.055
20.0	0.048	0.047	0.041	0.084	0.044	0.041	0.058	0.049	0.027	0.051	0.028	0.055
0.5T	0.052	0.051	0.047	0.090	0.062	0.058	0.051	0.050	0.066	0.050	0.070	0.057
DGP3: $T = 250$, $\phi = 0$ and $\sigma^2_{ut} = \sigma^2_{vt} = 1\mathbb{I}(t \le \lfloor 0.5T \rfloor) + \frac{1}{4}\mathbb{I}(t > \lfloor 0.5T \rfloor)$
0.0	0.045	0.047	0.076	0.061	0.036	0.043	0.047	0.057	0.029	0.059	0.037	0.072
2.5	0.045	0.047	0.058	0.072	0.031	0.033	0.046	0.061	0.022	0.058	0.017	0.063
5.0	0.044	0.045	0.045	0.067	0.029	0.036	0.043	0.056	0.016	0.060	0.012	0.062
10.0	0.044	0.045	0.042	0.069	0.033	0.038	0.040	0.051	0.018	0.057	0.020	0.059
20.0	0.045	0.043	0.040	0.071	0.041	0.044	0.046	0.049	0.027	0.053	0.031	0.057
0.5T	0.047	0.045	0.045	0.084	0.056	0.062	0.049	0.053	0.069	0.049	0.063	0.038
DGP1: $T = 500$, $\phi = 0$ and $\sigma^2_{ut} = \sigma^2_{vt} = 1$
0.0	0.046	0.046	0.043	0.045	0.043	0.039	0.049	0.051	0.017	0.052	0.011	0.050
2.5	0.048	0.049	0.047	0.047	0.042	0.039	0.052	0.052	0.016	0.054	0.014	0.048
5.0	0.047	0.048	0.047	0.046	0.044	0.045	0.053	0.056	0.015	0.054	0.013	0.050
10.0	0.050	0.051	0.049	0.048	0.047	0.042	0.056	0.049	0.027	0.055	0.027	0.056
20.0	0.053	0.052	0.053	0.053	0.051	0.043	0.052	0.052	0.041	0.051	0.043	0.062
0.5T	0.047	0.047	0.044	0.045	0.053	0.053	0.048	0.050	0.065	0.054	0.072	0.058
DGP2: $T = 500$, $\phi = 0$ and $\sigma^2_{ut} = \sigma^2_{vt} = 1\mathbb{I}(t \le \lfloor 0.5T \rfloor) + 4\mathbb{I}(t > \lfloor 0.5T \rfloor)$
0.0	0.052	0.053	0.036	0.061	0.050	0.039	0.063	0.055	0.025	0.061	0.033	0.062
2.5	0.051	0.052	0.036	0.065	0.046	0.040	0.062	0.053	0.023	0.060	0.022	0.060
5.0	0.050	0.050	0.037	0.073	0.046	0.044	0.056	0.052	0.026	0.061	0.026	0.055
10.0	0.050	0.050	0.041	0.083	0.050	0.045	0.055	0.054	0.033	0.058	0.035	0.052
20.0	0.050	0.051	0.045	0.088	0.053	0.050	0.051	0.056	0.046	0.055	0.041	0.047
0.5T	0.052	0.050	0.045	0.092	0.066	0.054	0.053	0.050	0.063	0.056	0.069	0.050
DGP3: $T = 500$, $\phi = 0$ and $\sigma^2_{ut} = \sigma^2_{vt} = 1\mathbb{I}(t \le \lfloor 0.5T \rfloor) + \frac{1}{4}\mathbb{I}(t > \lfloor 0.5T \rfloor)$
0.0	0.050	0.051	0.085	0.073	0.045	0.042	0.056	0.062	0.029	0.066	0.033	0.062
2.5	0.052	0.053	0.065	0.077	0.043	0.041	0.056	0.060	0.028	0.063	0.020	0.083
5.0	0.053	0.051	0.050	0.075	0.042	0.053	0.051	0.060	0.025	0.065	0.025	0.078
10.0	0.052	0.053	0.049	0.076	0.044	0.049	0.054	0.058	0.030	0.062	0.026	0.075
20.0	0.052	0.051	0.048	0.081	0.047	0.056	0.053	0.058	0.041	0.057	0.041	0.064
0.5T	0.051	0.051	0.048	0.092	0.056	0.061	0.050	0.053	0.066	0.056	0.055	0.050

Notes: A superscript $*$ denotes tests run using the fixed regressor wild bootstrap outlined in Algorithm 1; $t^2_{\beta_1}$ and $t^2_{\beta_1,NW}$ denote the full sample IV-combination predictability tests of Breitung and Demetrescu (2015) based on the 5% asymptotic critical value from the χ^2_1 distribution and computed with Eicker-White [EW] and conventional standard errors, respectively, and $t^{2*}_{\beta_1}$ and $t^{2*}_{\beta_1,NW}$ their bootstrap analogues; $\mathcal{T}^{f*}$, $\mathcal{T}^{b*}$ and $\mathcal{T}^{f*}_{NW}$, $\mathcal{T}^{b*}_{NW}$, denote the maximum forward and backward recursive tests computed with EW and conventional standard errors, respectively; $\mathcal{T}^{r*}$ and $\mathcal{T}^{r*}_{NW}$ denote the maximum rolling tests computed with EW and conventional standard errors, respectively; $\mathcal{T}^{d*}$ and $\mathcal{T}^{d*}_{ols}$ denote the maximum double recursive tests computed with EW and conventional standard errors, respectively.

Table 2: Empirical Rejection Frequencies under the Null Hypothesis, H_0. Nominal 5% significance level. DGP1-DGP3 with $\phi = -0.90$.

| c	$t_{\beta_1}^{2*}$	$t_{\beta_1,NW}^{2*}$	$t_{\beta_1}^{2}$	$t_{\beta_1,NW}^{2}$	$\mathcal{T}^{f*}$	$\mathcal{T}^{b*}$	$\mathcal{T}_{NW}^{f*}$	$\mathcal{T}_{NW}^{b*}$	$\mathcal{T}^{r*}$	$\mathcal{T}_{NW}^{r*}$	$\mathcal{T}^{d*}$	$\mathcal{T}_{NW}^{d*}$
DGP1: $T = 250$, $\phi = -0.90$ and $\sigma_{ut}^2 = \sigma_{vt}^2 = 1$												
0.0	0.069	0.073	0.069	0.074	0.037	0.035	0.047	0.055	0.011	0.053	0.018	0.055
2.5	0.055	0.056	0.055	0.057	0.030	0.040	0.037	0.058	0.011	0.052	0.020	0.062
5.0	0.053	0.052	0.051	0.050	0.034	0.045	0.039	0.055	0.012	0.051	0.021	0.064
10.0	0.057	0.055	0.056	0.058	0.038	0.048	0.044	0.059	0.017	0.055	0.029	0.065
20.0	0.060	0.060	0.057	0.056	0.046	0.051	0.050	0.062	0.029	0.061	0.034	0.067
0.5T	0.055	0.053	0.051	0.052	0.059	0.067	0.057	0.060	0.073	0.058	0.067	0.055
DGP2: $T = 250$, $\phi = -0.90$ and $\sigma_{ut}^2 = \sigma_{vt}^2 = 1\mathbb{I}(t \leq \lfloor 0.5T \rfloor) + 4\mathbb{I}(t > \lfloor 0.5T \rfloor)$												
0.0	0.057	0.057	0.044	0.071	0.032	0.031	0.026	0.046	0.009	0.030	0.018	0.021
2.5	0.052	0.053	0.044	0.076	0.035	0.032	0.027	0.051	0.011	0.028	0.012	0.031
5.0	0.054	0.053	0.046	0.076	0.037	0.037	0.033	0.054	0.013	0.032	0.014	0.038
10.0	0.054	0.051	0.046	0.077	0.042	0.043	0.037	0.055	0.018	0.039	0.019	0.051
20.0	0.054	0.052	0.049	0.080	0.047	0.046	0.042	0.054	0.032	0.045	0.036	0.055
0.5T	0.058	0.055	0.053	0.097	0.067	0.059	0.054	0.057	0.063	0.056	0.073	0.058
DGP3: $T = 250$, $\phi = -0.90$ and $\sigma_{ut}^2 = \sigma_{vt}^2 = 1\mathbb{I}(t \leq \lfloor 0.5T \rfloor) + \frac{1}{4}\mathbb{I}(t > \lfloor 0.5T \rfloor)$												
0.0	0.072	0.078	0.125	0.118	0.029	0.036	0.049	0.076	0.007	0.072	0.009	0.084
2.5	0.046	0.048	0.059	0.068	0.022	0.046	0.038	0.071	0.008	0.071	0.005	0.077
5.0	0.049	0.048	0.055	0.068	0.027	0.049	0.040	0.064	0.010	0.068	0.006	0.075
10.0	0.058	0.051	0.052	0.073	0.033	0.049	0.041	0.056	0.018	0.061	0.013	0.062
20.0	0.053	0.053	0.049	0.077	0.040	0.052	0.050	0.050	0.028	0.055	0.036	0.061
0.5T	0.053	0.052	0.047	0.091	0.058	0.067	0.056	0.056	0.064	0.056	0.070	0.061
DGP1: $T = 500$, $\phi = -0.90$ and $\sigma_{ut}^2 = \sigma_{vt}^2 = 1$												
0.0	0.076	0.079	0.077	0.078	0.041	0.054	0.045	0.072	0.015	0.055	0.023	0.060
2.5	0.059	0.060	0.056	0.058	0.034	0.060	0.037	0.075	0.020	0.053	0.016	0.045
5.0	0.061	0.060	0.058	0.061	0.038	0.064	0.041	0.075	0.023	0.057	0.023	0.056
10.0	0.062	0.063	0.061	0.063	0.045	0.065	0.049	0.076	0.034	0.059	0.035	0.053
20.0	0.063	0.062	0.060	0.061	0.050	0.065	0.054	0.072	0.047	0.063	0.046	0.064
0.5T	0.050	0.049	0.049	0.050	0.060	0.058	0.053	0.053	0.070	0.059	0.067	0.058
DGP2: $T = 500$, $\phi = -0.90$ and $\sigma_{ut}^2 = \sigma_{vt}^2 = 1\mathbb{I}(t \leq \lfloor 0.5T \rfloor) + 4\mathbb{I}(t > \lfloor 0.5T \rfloor)$												
0.0	0.072	0.067	0.055	0.082	0.041	0.047	0.029	0.066	0.022	0.037	0.021	0.033
2.5	0.060	0.057	0.052	0.079	0.041	0.049	0.030	0.063	0.021	0.036	0.030	0.039
5.0	0.061	0.060	0.052	0.085	0.045	0.057	0.033	0.068	0.024	0.040	0.039	0.043
10.0	0.058	0.058	0.053	0.088	0.049	0.055	0.040	0.064	0.034	0.043	0.049	0.045
20.0	0.061	0.058	0.052	0.095	0.056	0.060	0.044	0.066	0.044	0.045	0.056	0.051
0.5T	0.048	0.048	0.045	0.084	0.063	0.052	0.051	0.050	0.067	0.057	0.073	0.054
DGP3: $T = 500$, $\phi = -0.90$ and $\sigma_{ut}^2 = \sigma_{vt}^2 = 1\mathbb{I}(t \leq \lfloor 0.5T \rfloor) + \frac{1}{4}\mathbb{I}(t > \lfloor 0.5T \rfloor)$												
0.0	0.082	0.089	0.133	0.129	0.043	0.054	0.060	0.092	0.019	0.092	0.017	0.100
2.5	0.053	0.053	0.066	0.074	0.029	0.071	0.047	0.090	0.022	0.093	0.023	0.081
5.0	0.054	0.054	0.056	0.076	0.033	0.068	0.047	0.077	0.027	0.083	0.026	0.083
10.0	0.060	0.058	0.057	0.086	0.040	0.069	0.048	0.070	0.033	0.076	0.033	0.078
20.0	0.061	0.058	0.055	0.092	0.047	0.063	0.054	0.063	0.044	0.070	0.049	0.074
0.5T	0.053	0.052	0.051	0.096	0.057	0.068	0.053	0.058	0.069	0.054	0.071	0.068

Notes: See notes to Table 1

Table 3: Application to updated Welch and Goyal (2008) data: bivariate regressions - (1950:01 - 2017:12)

	$t^2_{\beta_1}$	$t^2_{\beta_1,NW}$	$\mathcal{T}^f$	$\mathcal{T}^f_{NW}$	$\mathcal{T}^b$	$\mathcal{T}^b_{NW}$	$\mathcal{T}^r$	$\mathcal{T}^r_{NW}$	$\mathcal{T}^d$	$\mathcal{T}^d_{NW}$	LM_x	$supF_x$
dp_{t-1}	0.472	0.457	10.088	**11.480**	5.608	6.885	6.882	8.280	10.284	12.519	**2.229**	**131.915**
	(0.481)	(0.492)	(0.121)	**(0.006)**	(0.340)	(0.182)	(0.729)	(0.287)	(0.998)	(0.194)	**(0.000)**	**(0.000)**
dy_{t-1}	0.581	0.568	**15.565**	12.616	6.241	7.849	11.143	9.318	10.252	**11.891**	0.295	11.178
	(0.414)	(0.408)	**(0.041)**	**(0.005)**	(0.252)	(0.133)	(0.507)	(0.182)	(0.072)	**(0.029)**	**(0.028)**	**(0.038)**
e/p_{t-1}	0.335	0.451	8.459	**9.189**	2.163	4.744	8.583	9.419	8.583	11.742	0.209	**38.229**
	(0.510)	(0.513)	(0.316)	**(0.043)**	(0.617)	(0.398)	(0.360)	(0.152)	(0.990)	(0.231)	(0.116)	**(0.000)**
de_{t-1}	0.291	0.490	**12.553**	**17.087**	0.291	0.490	13.399	**20.112**	15.145	**21.400**	0.192	5.031
	(0.594)	(0.575)	**(0.029)**	**(0.023)**	(0.852)	(0.851)	(0.058)	**(0.017)**	(0.099)	**(0.010)**	(0.210)	(0.355)
$rvol_{t-1}$	1.809	2.288	3.765	4.432	2.657	3.200	4.230	6.187	4.624	6.692	0.124	6.614
	(0.096)	(0.114)	(0.308)	(0.313)	(0.182)	(0.167)	(0.694)	(0.698)	(0.455)	(0.278)	(0.525)	(0.192)
bm_{t-1}	0.037	0.042	7.150	7.125	5.959	7.321	7.612	8.299	7.612	8.299	**0.342**	7.555
	(0.841)	(0.844)	(0.222)	(0.229)	(0.287)	(0.154)	(0.764)	(0.322)	(0.989)	(0.544)	**(0.012)**	(0.148)
$ntis_{t-1}$	0.041	0.059	5.634	5.235	1.648	2.600	9.383	10.543	9.679	10.874	0.375	8.102
	(0.830)	(0.821)	(0.180)	(0.312)	(0.623)	(0.546)	(0.074)	(0.059)	(0.101)	(0.114)	(0.070)	(0.180)
tbl_{t-1}	**7.001**	**9.408**	**11.763**	**15.989**	7.001	**9.408**	8.203	**11.618**	11.764	**16.588**	0.131	12.348
	(0.004)	**(0.004)**	**(0.003)**	**(0.001)**	(0.096)	**(0.040)**	(0.293)	**(0.035)**	(0.267)	**(0.029)**	(0.174)	(0.132)
lty_{t-1}	**4.178**	**5.524**	**11.036**	**13.410**	6.529	6.547	8.224	10.070	11.135	**14.308**	0.103	7.985
	(0.029)	**(0.026)**	**(0.013)**	**(0.009)**	(0.140)	(0.152)	(0.446)	(0.115)	(0.487)	**(0.044)**	(0.303)	(0.182)
ltr_{t-1}	**4.172**	**5.724**	**8.479**	10.313	4.438	6.021	8.145	9.702	8.902	10.709	0.163	6.407
	(0.045)	**(0.044)**	**(0.034)**	(0.055)	(0.136)	(0.167)	(0.082)	(0.115)	(0.061)	(0.094)	(0.341)	(0.325)
tms_{t-1}	2.726	3.075	**15.839**	**17.534**	4.769	5.133	12.142	**12.611**	**15.839**	**17.534**	0.350	**10.877**
	(0.084)	(0.084)	**(0.001)**	**(0.001)**	(0.145)	(0.150)	(0.061)	**(0.008)**	**(0.042)**	**(0.003)**	(0.060)	**(0.026)**
dfy_{t-1}	0.011	0.020	2.777	3.079	0.996	2.872	7.431	**19.169**	8.805	**19.473**	0.063	9.773
	(0.908)	(0.909)	(0.617)	(0.587)	(0.695)	(0.546)	(0.246)	**(0.044)**	(0.216)	**(0.046)**	(0.788)	(0.106)
dfr_{t-1}	0.768	1.535	4.487	5.475	2.407	5.398	11.218	7.685	11.218	13.141	0.156	6.134
	(0.477)	(0.434)	(0.475)	(0.416)	(0.336)	(0.346)	(0.228)	(0.521)	(0.295)	(0.336)	(0.537)	(0.377)
$infl_{t-1}$	3.042	**4.890**	9.083	**16.834**	3.999	6.138	11.705	11.516	12.195	**16.839**	0.326	**11.517**
	(0.070)	**(0.048)**	(0.276)	**(0.003)**	(0.269)	(0.155)	(0.477)	(0.060)	(0.722)	**(0.026)**	(0.102)	**(0.032)**

Notes: Numbers in parentheses are bootstrap p-values. Bold entries are those which are statistically significant at the 5% level (or stricter).
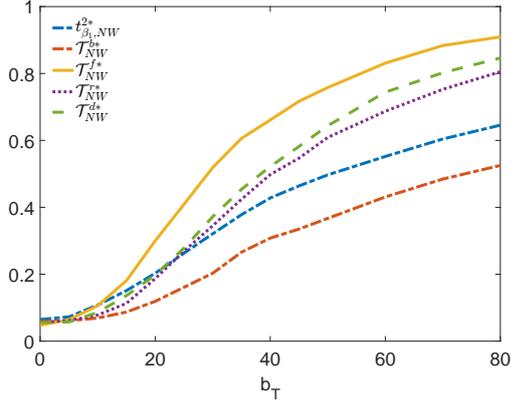
(a) $c = 0, \phi = -0.90$
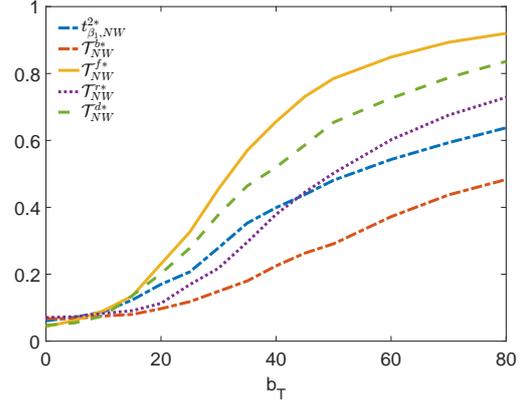
(b) $c = 10, \phi = -0.90$
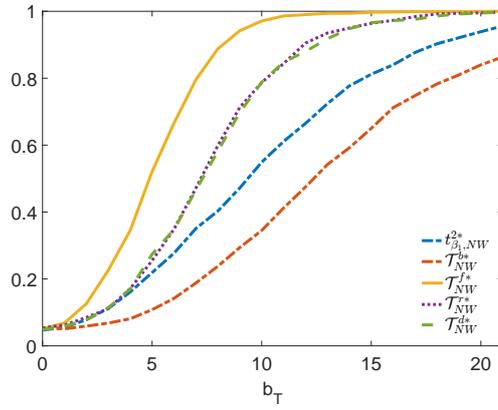
(c) $c = 0.5T, \phi = -0.90$

Figure 1: Finite sample local power: Case 1, $T = 250$.
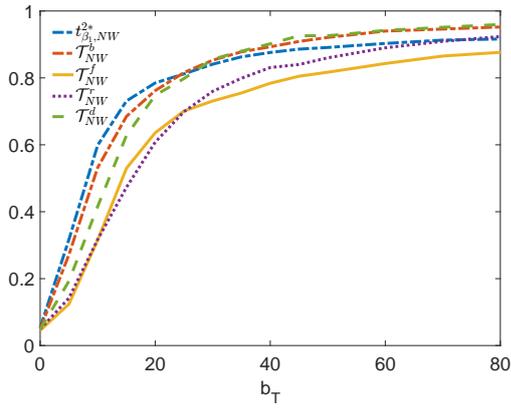
(a) $c = 0, \phi = -0.90$
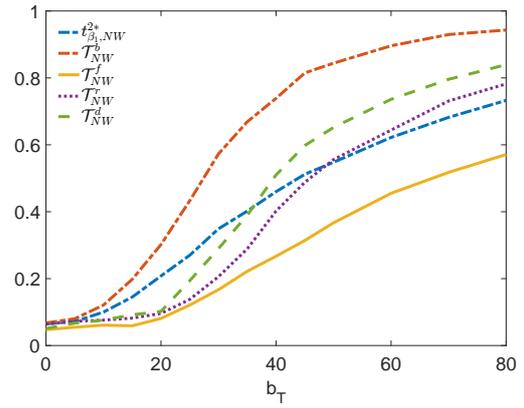
(b) $c = 10, \phi = -0.90$
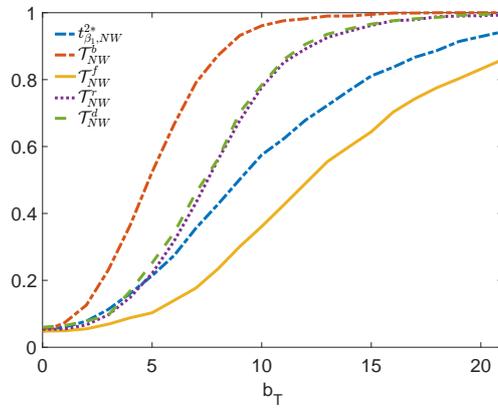
(c) $c = 0.5T, \phi = -0.90$

Figure 2: Finite sample local power: Case 2, $T = 250$.
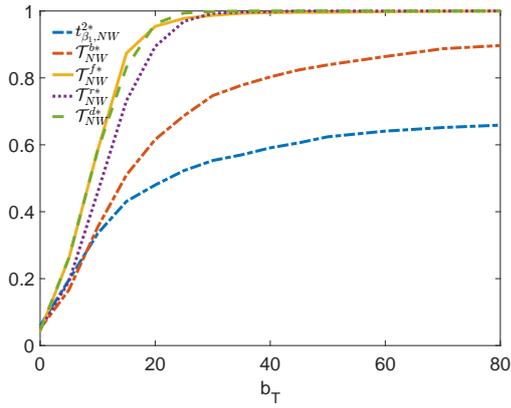
(a) $c = 0, \phi = -0.90$
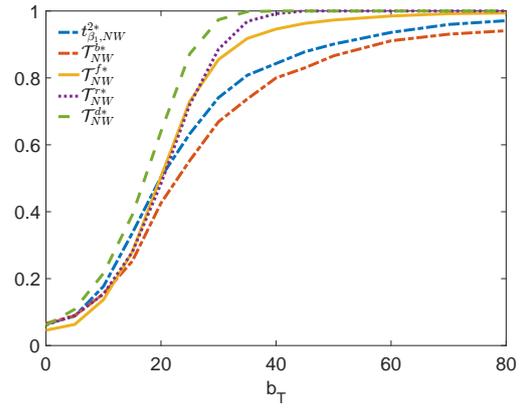
(b) $c = 10, \phi = -0.90$
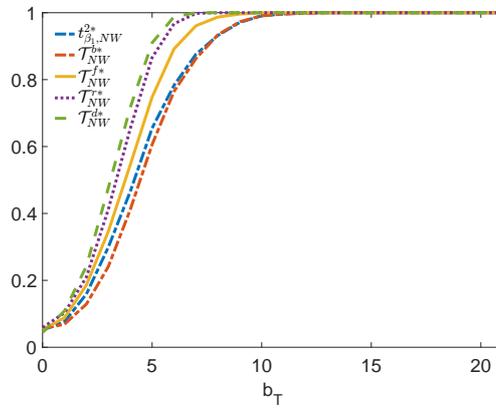
(c) $c = 0.5T, \phi = -0.90$

Figure 3: Finite sample local power: Case 3, $T = 250$.

(a) $y_t = ep_t, x_{t-1} = dp_{t-1}$

(b) $y_t = ep_t, x_{t-1} = dy_{t-1}$

(c) $y_t = ep_t, x_{t-1} = e/p_{t-1}$

(d) $y_t = ep_t, x_{t-1} = de_{t-1}$

(e) $y_t = ep_t, x_{t-1} = rvol_{t-1}$

(f) $y_t = ep_t, x_{t-1} = bm_{t-1}$

(g) $y_t = ep_t, x_{t-1} = ntis_{t-1}$

(h) $y_t = ep_t, x_{t-1} = tbl_{t-1}$

(i) $y_t = ep_t, x_{t-1} = lty_{t-1}$

(j) $y_t = ep_t, x_{t-1} = ltr_{t-1}$

(k) $y_t = ep_t, x_{t-1} = tms_{t-1}$

(l) $y_t = ep_t, x_{t-1} = dfy_{t-1}$

(m) $y_t = ep_t, x_{t-1} = df_{t-1}$

(n) $y_t = ep_t, x_{t-1} = infl_{t-1}$
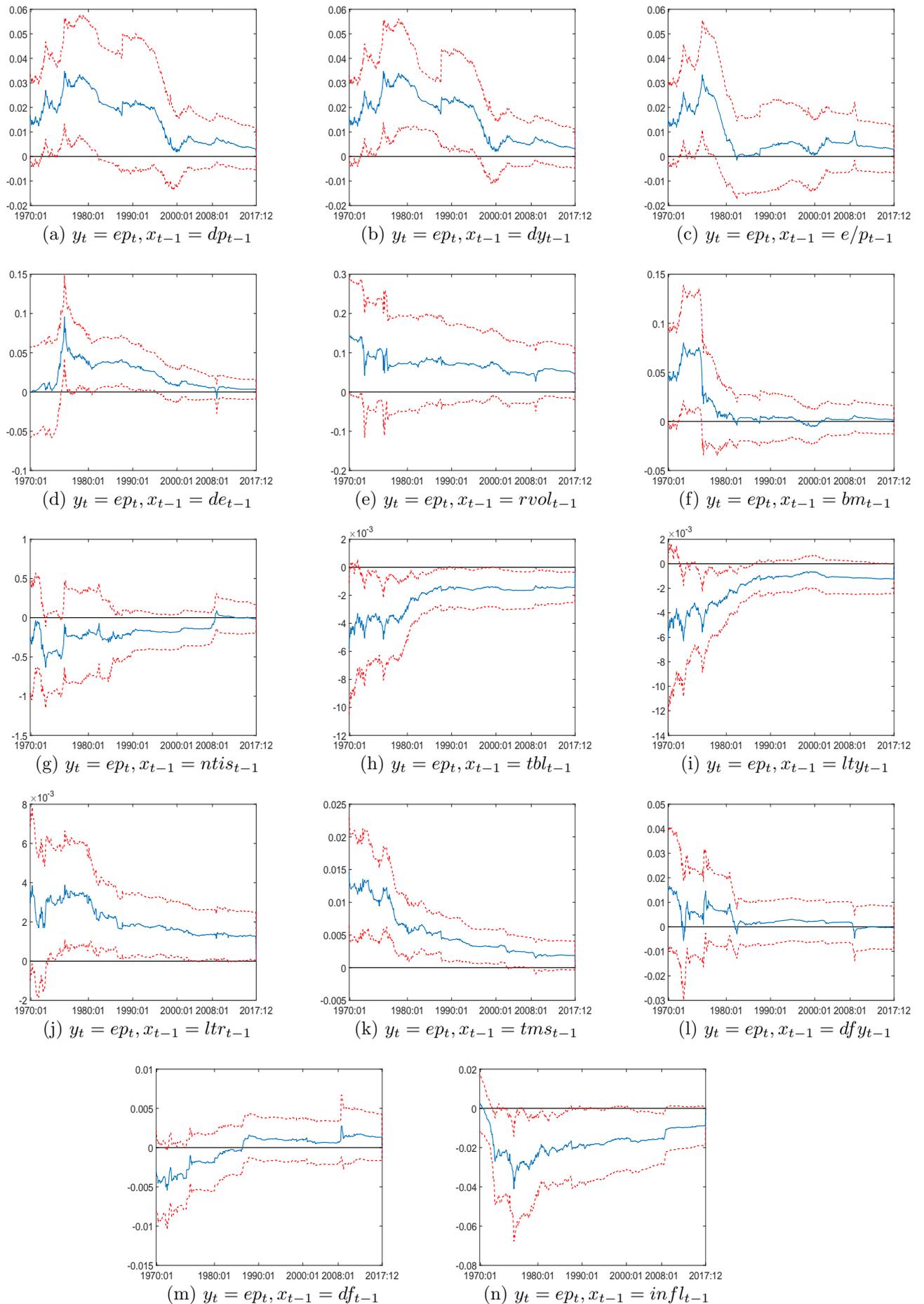
Figure 4: Forward recursive slope estimates (solid line) and 95% confidence bands (dotted lines). Sample period 1950:01 - 2017:12. The y_t and x_t variable labels are as defined in the main text.
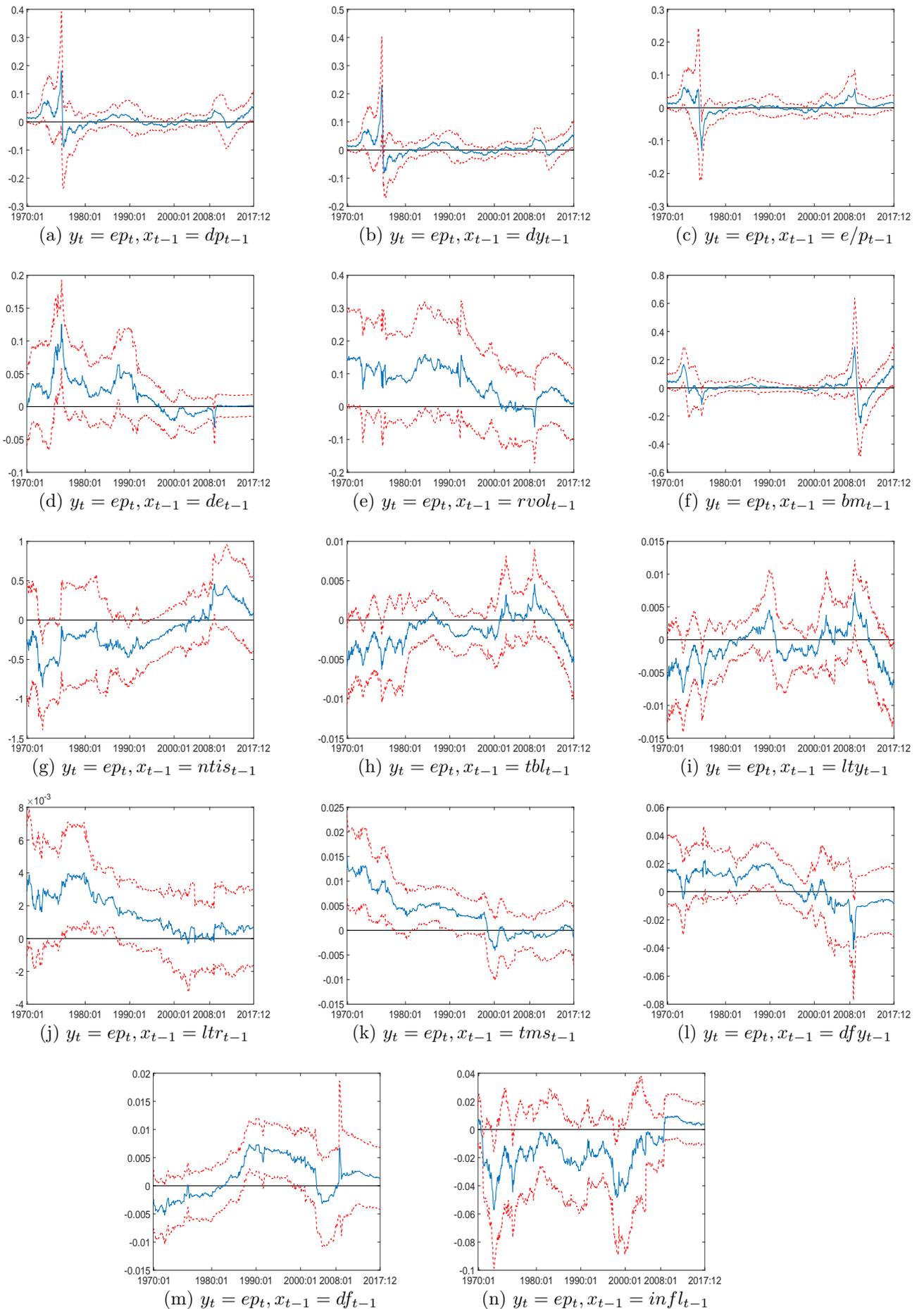
Figure 5: Rolling slope estimates (solid line) and 95% confidence bands (dotted lines). Sample period 1950:01 - 2017:12. The y_t and x_t variable labels are as defined in the main text.

(a) $y_t = ep_t, x_{t-1} = dp_{t-1}$

(b) $y_t = ep_t, x_{t-1} = dy_{t-1}$

(c) $y_t = ep_t, x_t = e/p_{t-1}$

(d) $y_t = ep_t, x_{t-1} = de_{t-1}$

(e) $y_t = ep_t, x_{t-1} = tbl_{t-1}$

(f) $y_t = ep_t, x_{t-1} = lty_{t-1}$

(g) $y_t = ep_t, x_{t-1} = ltr_{t-1}$

(h) $y_t = ep_t, x_{t-1} = tms_{t-1}$

(i) $y_t = ep_t, x_{t-1} = infl_{t-1}$

- - - BS 5% - - - BS 10% .... $\chi^2_{1,5\%}$ .... $\chi^2_{1,10\%}$

Figure 6: Plots of forward recursive subsample statistics with marginal and bootstrap 10% and 5% critical values.

On-Line Supplementary Appendix:

Testing for Episodic Predictability in Stock Returns

by

M. Demetrescu, I. Georgiev, P.M.M. Rodrigues and A.M.R. Taylor

S.1 Introduction

This supplement has four main sections. Section S.2 details how the methods outlined in the main paper can be extended to allow for multiple predictors and deterministic components beyond an intercept/mean. Section S.3 contains the proofs of the main theorems from the paper. Additional Monte Carlo results relating to GARCH and fat-tailed innovations are reported in Section S.4. Finally, additional material relating to the empirical example in section 7 of the main paper is provided in Section S.5.

Equation references (S.n) for $n \geq 1$ refer to equations in this supplementary appendix and other equation references are to the main paper. Additional references are included at the end of the supplement.

S.2 Extensions: Deterministic Components and Multiple Predictors

In this section we briefly outline how the subsample predictability tests developed in the context of (2.1) and (2.2) with a single predictive regressor, x_{t-1}, and an intercept can be generalised to the case where the predictive regression contains multiple predictors and/or a general deterministic component.

S.2.1 Deterministic Components

To allow for a more flexible deterministic component, we can generalise equations (2.1) and (2.2) to

$$y_t = \boldsymbol{\psi}_y' \boldsymbol{f}_t + \beta_{1,t} x_{t-1} + u_t \tag{S.1}$$

and

$$x_t = \boldsymbol{\psi}_x' \boldsymbol{f}_t + \xi_t \tag{S.2}$$

respectively, where $\boldsymbol{f}_t$ is a vector whose elements are deterministic satisfying typical conditions. In particular, $\boldsymbol{f}_t := \mathbf{L}_T^{-1} \boldsymbol{F}(t/T)$ for some vector $\boldsymbol{F}(\cdot)$ of (piecewise) smooth deterministic functions and $\mathbf{L}_T$ diagonal weighting matrix with 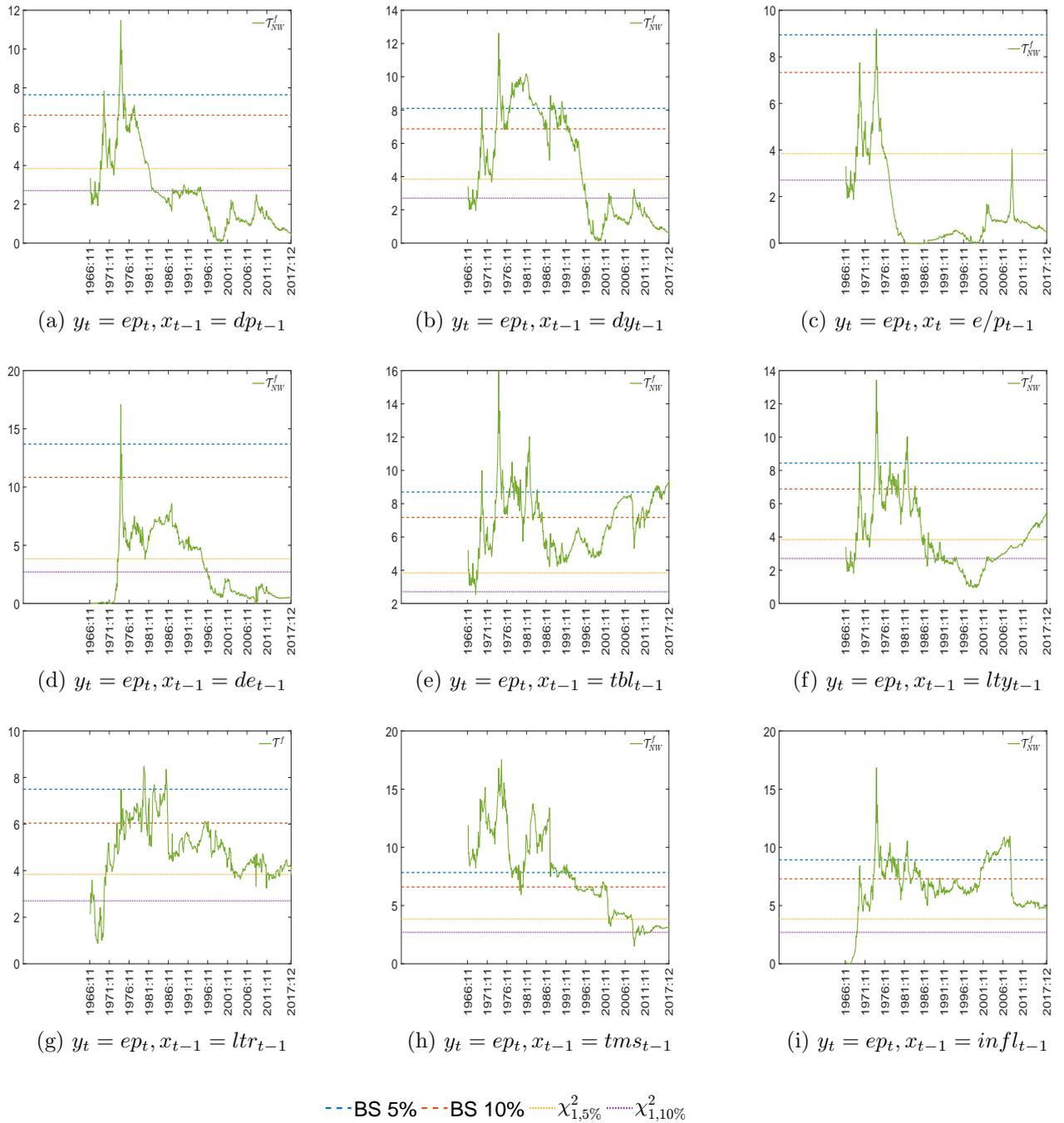diagonal elements $l_{ii} := T^{-\delta_l}$ for $\delta_l \geq 0$. An obvious example is the case of constant and trend, which obtains for $\boldsymbol{f}_t := (1, t)'$, where $\mathbf{L}_T := \text{diag}\{1, T^{-1}\}$ and $\boldsymbol{F}(s) = (1, s)'$. Following Breitung and Demetrescu (2015, p. 365), $\boldsymbol{f}_t$ can be regarded as collecting together, or "pooling", all of the deterministic variables that feature in either (S.1) or (S.2) or both, so that different deterministic components can feature in (S.1) and (S.2) by setting the appropriate elements of $\boldsymbol{\psi}_y$ and/or $\boldsymbol{\psi}_x$ to zero.

Under this extended deterministic form, the instruments used in calculating the subsample sequences of statistics must be such that they are not linear combinations of the pooled deterministic trend function on any of the subsamples considered in a given sequence; that is, we require that the condition

$$\left| \int_{\tau_1}^{\tau_2} \begin{pmatrix} \boldsymbol{F}(s) \\ Z(s) \end{pmatrix} \left(\boldsymbol{F}(s)', \ Z(s) \right) \mathrm{d}s \right| \neq 0 \tag{S.3}$$

holds for all possible τ_1 and τ_2 considered in the sequence of subsamples; for example, in the case of the forward recursive sequence it must hold for all τ_1, τ_2 such that $\tau_1 = 0$ and $\tau_L \leq \tau_2 \leq 1$.

Remark S.1. To gauge the impact of the condition in (S.3) on the validity of our test procedures, consider again the leading linear trend example, $\boldsymbol{f}_t := (1, t)'$. Here we may not choose as type-II instrument any linear combination of the intercept and linear trend terms. The choice of type-II instrument given in (4.1) therefore remains valid in this example. This family of instruments is motivated by the Karhunen-Loéve decomposition of Wiener and OU processes; see Phillips (1998). The Karhunen-Loéve representation of a deterministic linear trend is formed from a weighted infinite series of terms of the form given in (4.1) over $k = 1, 2, \ldots$, and therefore the linear trend cannot be collinear with just one of them. In the unlikely event that $\sin\left(k \frac{\pi t}{2T}\right)$ is an element of the deterministic

trend vector $\boldsymbol{f}_t$, one could simply pick $\sin\left(\ell\frac{\pi t}{2T}\right)$, for some $\ell \geq 1$ such that $\ell \neq k$, as the type-II instrument. We also note that any failure of condition (S.3) in practice would be immediately apparent as one would encounter singular matrices in the computation of the test statistics for those subsamples were the condition failed. $\diamond$

In order to obtain tests which are exact invariant to the vectors of parameters $\boldsymbol{\psi}_y$ and $\boldsymbol{\psi}_x$, the sequences of subsample statistics from which our maximum statistics are obtained need to be based on appropriately detrended data. To that end, let $\breve{y}_t$, $\breve{\boldsymbol{z}}_{t-1}$ and $\breve{x}_{t-1}$ denote the residuals from the projection of y_t, $\boldsymbol{z}_{t-1}$ and x_{t-1}, respectively, onto the deterministic component $\boldsymbol{f}_t$ for the subsample $t = \lfloor\tau_1 T\rfloor + 1, ..., \lfloor\tau_2 T\rfloor$; that is, for $\boldsymbol{w}_t$ generically denoting any of y_t, $\boldsymbol{z}_{t-1}$ and x_{t-1},

$$\breve{\boldsymbol{w}}_t := \boldsymbol{w}_t - \sum_{s=\lfloor\tau_1 T\rfloor+1}^{\lfloor\tau_2 T\rfloor} \boldsymbol{w}_s \boldsymbol{f}_s' \left(\sum_{s=\lfloor\tau_1 T\rfloor+1}^{\lfloor\tau_2 T\rfloor} \boldsymbol{f}_s \boldsymbol{f}_s'\right)^{-1} \boldsymbol{f}_t. \tag{S.4}$$

In order to obtain exact invariance one then simply replaces $\tilde{y}_t$ by $\breve{y}_t$, $\tilde{\boldsymbol{z}}_{t-1}$ by $\breve{\boldsymbol{z}}_{t-1}$ and $\tilde{x}_{t-1}$ by $\breve{x}_{t-1}$, respectively, in computing $\boldsymbol{A}_T(\tau_1,\tau_2)$, $\mathbf{B}_T(\tau_1,\tau_2)$, $\boldsymbol{C}_T(\tau_1,\tau_2)$ and $\mathbf{D}_T(\tau_1,\tau_2)$ in (4.3). The limiting distributions for the resulting maximum statistics can be shown to be the same as those given in part (i) of Proposition 1 for the case where Assumption 1.1 holds. However, the results under Assumption 1.2 do change relative to those given in Proposition 1. In particular, the results given in part (ii) of Proposition 1 hold for the extended deterministic form discussed here provided $\tilde{Z}_{\tau_1,\tau_2}$ is replaced by

$$\breve{Z}_{\tau_1,\tau_2}(s) := Z(s) - \int_{\tau_1}^{\tau_2} Z(s)\boldsymbol{F}(s)' \left(\int_{\tau_1}^{\tau_2} \boldsymbol{F}(r)\boldsymbol{F}(r)'\mathrm{d}r\right)^{-1} \boldsymbol{F}(s).$$

We note in passing that the detrending scheme will affect the local power in the case of near-integrated regressors, but not when the regressors admit stable autoregressions.[10]

The bootstrap Algorithms 1 and 2 will need to be modified accordingly. In particular, in Step 1 of Algorithms 1 and 2 the bootstrap sample data, y_t^*, should now be constructed from full sample detrended data; that is, $y_t^* := \hat{y}_t^{\boldsymbol{f}} R_t$, where

$$\hat{y}_t^{\boldsymbol{f}} := y_t - \sum_{s=1}^{T} y_s \boldsymbol{f}_s' \left(\sum_{s=1}^{T} \boldsymbol{f}_s \boldsymbol{f}_s'\right)^{-1} \boldsymbol{f}_t.$$

The bootstrapped statistics are then computed analogously to the original statistics, as outlined above, using the subsample detrended data, $\breve{y}_t^*$ (the subsample detrended y_t^*, obtained setting $\boldsymbol{w}_t = y_t^*$ in the generic detrending formula in (S.4)), $\breve{\boldsymbol{z}}_{t-1}$ and $\breve{x}_{t-1}$. The key result given in Proposition 2, that the fixed regressor wild bootstrap implementation of the subsample predictability tests are asymptotically valid, continues to hold.

S.2.2 Multiple Predictors

In empirical work one might wish to consider predictive regression models with several (putative) predictors. This can help avoid the problem of spurious predictive regression effects in the case where relevant strongly persistent predictors are omitted from the estimated predictive regression; cf. Georgiev *et al.* (2018a).

---

[10]It should also be noted that some care may be required when generating the type-I instrument, $z_{I,t}$, from the regressor x_t. Depending on the form of the deterministic component of x_t and on the mechanism used to generate the instrument, $z_{I,t}$ could exhibit some form of deterministic component itself (notice that this does not occur with the IVX instrument given in (4.2) in the case considered there where x_t contains a constant mean because the mean is purged in the construction of the IVX instrument). It would then be imperative to check that the instrument $z_{I,t}$ still obeys Assumptions 4-6 in spite of any such induced deterministic component.

To that end consider replacing (2.1) by its multivariate counterpart

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_{i,t} x_{i,t-1} + u_t, \qquad t = 1, \ldots, T \qquad (S.5)$$

where the $x_{i,t}$ are generated analogously to the form given in (2.2), or the extended form including additional deterministic variables given in (S.2). Precisely, denoting the $k \times 1$ vector of predictive regressors as $\boldsymbol{x}_{t-1} := (x_{1,t-1}, \ldots, x_{k,t-1})'$, these are assumed to satisfy the DGP

$$\boldsymbol{x}_t = \boldsymbol{\Psi_x} \boldsymbol{f}_t + \boldsymbol{\xi}_t$$

where the ith row of $\boldsymbol{\Psi_x}$ gives the trend coefficients of the ith putative predictor $x_{i,t}$, and

$$\boldsymbol{\xi}_t = \boldsymbol{\Gamma} \boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t$$

with $\boldsymbol{v}_t$ and u_t obeying a multivariate version of Assumption 3. Assume that there exists an ordering of the elements of $\boldsymbol{x}_t$ such that $\boldsymbol{\Gamma}$ is block-diagonal. The first block relates to any near-integrated variables in $\boldsymbol{x}_{t-1}$ and, as in Equation (16) of Breitung and Demetrescu (2015), is specified as $\boldsymbol{\Gamma}_1 = \mathbf{I}_r - \frac{1}{T}\mathbf{C}$ with $\mathbf{C}$ a (not necessarily diagonal) $r \times r$ matrix, for some $0 \le r \le k$. The second block relates to any stable regressors in $\boldsymbol{x}_{t-1}$, and is such that the $(k-r) \times (k-r)$ matrix $\boldsymbol{\Gamma}_0$ has all of its $k - r$ eigenvalues smaller than unity in absolute value. In this setup, any given element of $\boldsymbol{x}_{t-1}$ may be either stable or near-integrated; indeed, we could have only near-integrated regressors ($r = k$), or only regressors generated from stable autoregressive processes ($r = 0$). We will not require knowledge which of these any given element of $\boldsymbol{x}_{t-1}$ satisfies because when implementing 2SLS-based inference the same mechanism discussed for the single-regressor case will ensure that the suitable instrument gets full GMM weight in the limit. This is an important advantage of our approach, as correctly classifying the individual variables in $\boldsymbol{x}_{t-1}$ as either stable or near integrated would be infeasible in practice. For each predictor, $x_{i,t-1}$, $i = 1, \ldots, k$, the associated slope parameter $\beta_{i,t}$ needs to satisfy the relevant localisation given in Assumption 2, so that $n_T = \sqrt{T}$ where $x_{i,t-1}$ is stable and $n_T = T$ where $x_{i,t-1}$ is near-integrated. Here the piecewise function $b(\cdot)$ in Assumption 2, which defines local power, need not be the same for each of the $x_{i,t-1}$, $i = 1, \ldots, k$.

The extension of the approach proposed in this paper to multiple predictors is, in principle, straightforward. For all predictor variables $x_{i,t-1}$, $i = 1, \ldots, k$, we use type-I and type-II instruments as before. As discussed in Breitung and Demetrescu (2015, p. 366), we will require multivariate versions of Assumptions 4-6 to hold. While using exactly one type-I instrument for each regressor is not problematic, Breitung and Demetrescu (2015, p. 366) point out that one should not use more than one type-II instrument for each strongly persistent regressor. We must therefore ensure that the number k^* of included type-II instruments does not exceed the number r of strongly persistent regressors. This is of course trivial if r is known (again, we note that specifying *which* of the regressors are strongly persistent is not required). Should r be unknown, it is straightforward but tedious to show that setting $k^* = 1$ is a valid choice irrespective of the true value of r ($0 \le r \le k$) provided each regressor is additionally instrumented by a suitable type-I instrument.

Let the resulting vector of instruments be given as $\boldsymbol{z}_{t-1}$ which stacks k type-I and k^* type-II instruments. For computational reasons (specifically, to avoid perfect multicollinearity), it must also hold that, for all of the subsamples over which the statistics of interest are computed, the instruments are linearly independent and are linearly independent of the deterministic component.

For the type-I instruments, an obvious choice is again given by the IVX approach, setting

$$z_{i,I,t-1} := \sum_{j=0}^{t-1} \varrho_i^j \Delta x_{i,t-1-j} \quad \text{with} \quad \varrho_i := 1 - \frac{a_i}{T^{\gamma_i}}, \ i = 1, \ldots, k$$

where $a_i > 0$ and $\gamma_i \in (0,1)$, $i = 1, \ldots, k$, and with the convention that $\Delta x_{i,0} = 0$, $i = 1, \ldots, k$. For the type-II instruments we could simply adopt a different spectral frequency for each of the type-II

instruments; viz.,

$$z_{i,II,t-1} = \sin\left(\frac{\omega_i(t-1)}{2T}\right), \; i = 1, ..., k^*$$

where $\omega_i \in (0, \pi]$, $i = 1, ..., k^*$, are fixed and distinct spectral frequencies. Along the lines of the proof of Lemma 1 it can be shown that these choices of instruments satisfy the required regularity conditions on the instruments as long as the condition in (S.3) is fulfilled with $Z(s)$ replaced by $\boldsymbol{Z}(s)$, the vector of limit functions of the k^* type-II instruments.

Assuming that with a given subsample $t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor$, the slope parameters in (S.5) are constant, such that $\beta_{i,t} = \beta_i$, $i = 1, ..., k$, then the 2SLS estimator of $\boldsymbol{\beta} := (\beta_1, ..., \beta_k)'$ using the set of instruments, $\boldsymbol{z}_{t-1}$, defined above for the subsample $t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor$ is given by

$$\hat{\boldsymbol{\beta}}_1(\tau_1, \tau_2) := \mathbf{M}_T^{-1} \mathbf{A}_T'(\tau_1, \tau_2) \mathbf{B}_T^{-1}(\tau_1, \tau_2) \mathbf{C}_T(\tau_1, \tau_2)$$

where $\mathbf{M}_T := \mathbf{C}_T'(\tau_1, \tau_2) \mathbf{B}_T^{-1}(\tau_1, \tau_2) \mathbf{C}_T(\tau_1, \tau_2)$, with $\mathbf{A}_T(\tau_1, \tau_2) := \sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} \breve{\boldsymbol{z}}_{t-1} \breve{\boldsymbol{x}}_{t-1}'$, and where the matrices $\mathbf{B}_T$, $\mathbf{C}_T$, and $\mathbf{D}_T$ retain their definitions from the previous subsection but for the form of $\boldsymbol{z}_{t-1}$ defined above.

Based on $\hat{\boldsymbol{\beta}}_1(\tau_1, \tau_2)$, we can then form IV-combination predictability tests, for the given subsample $t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor$, along the same principles as outlined in section 4.2 for the single predictor statistic $t_{\beta_1}(\tau_1, \tau_2)$ in (4.3). To that end, define the $q \times k$ full row rank matrix $\mathbf{R}$ of constants defining q linearly independent restrictions on $\boldsymbol{\beta}$. The 2SLS-based Wald statistic for testing the null hypothesis that $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$ holds over the subsample $t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor$ is then given by

$$W_{\boldsymbol{\beta}_1}^{\mathbf{R}}(\tau_1, \tau_2) := \left(\mathbf{R}\hat{\boldsymbol{\beta}}_1(\tau_1, \tau_2)\right)' \left(\mathbf{R}\widehat{\mathrm{Cov}\left(\hat{\boldsymbol{\beta}}_1(\tau_1, \tau_2)\right)}\mathbf{R}'\right)^{-1} \mathbf{R}\hat{\boldsymbol{\beta}}_1(\tau_1, \tau_2),$$

where

$$\widehat{\mathrm{Cov}\left(\hat{\boldsymbol{\beta}}_1(\tau_1, \tau_2)\right)} := \mathbf{M}_T^{-1} \left(\mathbf{A}_T'(\tau_1, \tau_2) \mathbf{B}_T^{-1}(\tau_1, \tau_2) \mathbf{D}_T(\tau_1, \tau_2) \mathbf{B}_T^{-1}(\tau_1, \tau_2) \mathbf{A}_T(\tau_1, \tau_2)\right) \mathbf{M}_T^{-1}.$$

A test statistic for the significance of any given predictor, $x_{i,t-1}$, $i \in \{1, ..., k\}$, obtains by setting $\mathbf{R}$ equal to the $1 \times k$ selection (row) vector whose ith element is equal to unity and all other elements are equal to zero. Test statistics for the joint significance of any subset of q of the predictors can be formed by merging such selection vectors into a $q \times k$ matrix. For $q = k$, so that $\mathbf{R} = \mathbf{I}_k$, we obtain a test statistic for the joint significance of all of the predictors $x_{i,t-1}$, $i = 1, ..., k$. Notice also that the joint predictability statistic proposed in section 3.3 of Breitung and Demetrescu (2015) in the context of the full sample obtains for $\tau_1 = 0$ and $\tau_2 = 1$. It can be shown that $W_{\boldsymbol{\beta}_1}^{\mathbf{R}}(\tau_1, \tau_2)$ has a χ_q^2 limiting null distribution.

Along exactly the lines detailed in section 4.2, for a given choice of $\mathbf{R}$, we can form forward and backward recursive, rolling, and double-recursive sequences of the $W_{\boldsymbol{\beta}_1}^{\mathbf{R}}(\tau_1, \tau_2)$ statistic calculated over the relevant sequences of subsamples. Tests for the null hypothesis that the subset of the elements of $\boldsymbol{x}_{t-1}$ chosen by $\mathbf{R}$ has no predictive power anywhere in the available sample against the alternative that they have joint (local) predictive power within some subset of the sample can then be based, as with the single regressor case, on the maxima of these sequences of subsample Wald statistics. The limiting distributions of these maximum statistics are multivariate variants of Proposition 1 which we do not provide here to save space. Clearly, they are not invariant to time-varying volatility for the same reasons as were outlined previously in the discussion of Proposition 1. Fixed regressor wild bootstrap versions of these tests can be implemented with an obvious generalisation of Algorithms 1 and 2. The behaviour of these bootstrapped maximum statistics can be derived in a similar way to those established in Proposition 2 and we omit the details here; the important point is that the leading conclusion regarding the asymptotic validity of the bootstrap given in Corollary 3 also holds for these multivariate tests.

S.3 Proofs

For the purposes of this supplementary appendix we set $\mu_x = 0$ and $\beta_0 = 0$ in (2.1) and (2.2), respectively, throughout. This is done without loss of generality given that the statistics proposed in the paper are, by design, exact invariant to both μ_x and β_0. Moreover, we will establish the validity of the large sample results given in the paper for the case where the residuals used in computing the statistics are constructed under the null hypothesis; see again the discussion following Equation (3.2) and in Footnote 3. This simplifies the algebra in the proofs, but can straightforwardly be shown to be asymptotically equivalent to the two other possible choices (OLS and 2SLS) of residuals discussed. Throughout this supplementary appendix we will use 'null hypothesis' as shorthand for H_0 of (2.3), and 'local alternative' as shorthand for $H_{1,b(\cdot)}$ of (2.4).

S.3.1 Notation

With the convention stated above that we set $\mu_x = \beta_0 = 0$, we may define the following partial sums:

$$S_y(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} y_t, \qquad S_x(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} x_{t-1}, \qquad \boldsymbol{S_z}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1}$$

(and analogously $S_{z_I}(\tau)$ and $S_{z_{II}}(\tau)$ for the partial sums involving each of the two instruments $z_{I,t}$ and $z_{II,t}$ taken alone),

$$\boldsymbol{S_{xz}}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} x_{t-1} \boldsymbol{z}_{t-1}, \qquad \mathbf{S_{zz}}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}'$$

(and analogously $S_{xz_I}(\tau)$, $S_{xz_{II}}(\tau)$, $S_{z_I^2}(\tau)$, $S_{z_I z_{II}}(\tau)$, $S_{z_{II}^2}(\tau)$), as well as

$$\mathbf{S_{zzy^2}}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}' y_t^2, \quad \mathbf{S_{zzy}}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}' y_t, \quad \boldsymbol{S_{zy}}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} y_t,$$

$$\boldsymbol{S_{zy^2}}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} y_t^2, \qquad S_{y^2}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} y_t^2$$

(and analogously $S_{z_I y}(\tau)$ etc.). For $0 \leq \tau_1 < \tau_2 \leq 1$, we then have (with the standard convention that $\sum_{t=1}^{0} = 0$) the following alternative representations for the objects $\boldsymbol{A}_T(\tau_1, \tau_2)$, $\mathbf{B}_T(\tau_1, \tau_2)$, $\boldsymbol{C}_T(\tau_1, \tau_2)$ and $\mathbf{D}_T(\tau_1, \tau_2)$ used in the definition of $t_{\beta_1}(\tau_1, \tau_2)$ of (4.3):

$$\boldsymbol{A}_T(\tau_1, \tau_2) := \boldsymbol{S_{xz}}(\tau_2) - \boldsymbol{S_{xz}}(\tau_1) - \frac{1}{\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor}(S_x(\tau_2) - S_x(\tau_1))(\boldsymbol{S_z}(\tau_2) - \boldsymbol{S_z}(\tau_1)),$$

$$\mathbf{B}_T(\tau_1, \tau_2) := \mathbf{S_{zz}}(\tau_2) - \mathbf{S_{zz}}(\tau_1) - \frac{1}{\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor}(\boldsymbol{S_z}(\tau_2) - \boldsymbol{S_z}(\tau_1))(\boldsymbol{S_z}(\tau_2) - \boldsymbol{S_z}(\tau_1))',$$

$$\boldsymbol{C}_T(\tau_1, \tau_2) := \boldsymbol{S_{zy}}(\tau_2) - \boldsymbol{S_{zy}}(\tau_1) - \frac{1}{\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor}(\boldsymbol{S_z}(\tau_2) - \boldsymbol{S_z}(\tau_1))(S_y(\tau_2) - S_y(\tau_1))$$

and

$$\begin{aligned}
\mathbf{D}_T(\tau_1, \tau_2) \quad := \quad & \mathbf{S_{zzy^2}}(\tau_2) - \mathbf{S_{zzy^2}}(\tau_1) - \frac{2}{\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor}(S_y(\tau_2) - S_y(\tau_1))(\mathbf{S_{zzy}}(\tau_2) - \mathbf{S_{zzy}}(\tau_1)) \\
& + \frac{1}{(\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor)^2}(S_y(\tau_2) - S_y(\tau_1))^2(\mathbf{S_{zz}}(\tau_2) - \mathbf{S_{zz}}(\tau_1)) \\
& - \frac{1}{\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor}(\boldsymbol{S_{zy^2}}(\tau_2) - \boldsymbol{S_{zy^2}}(\tau_1))(\boldsymbol{S_z}(\tau_2) - \boldsymbol{S_z}(\tau_1))'
\end{aligned}$$

$$+ \frac{2}{(\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor)^2} (S_y(\tau_2) - S_y(\tau_1)) (\boldsymbol{S}_{zy}(\tau_2) - \boldsymbol{S}_{zy}(\tau_1)) (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1))'$$

$$- \frac{2}{(\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor)^3} (S_y(\tau_2) - S_y(\tau_1))^2 (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1)) (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1))'$$

$$- \frac{1}{\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor} (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1)) (\boldsymbol{S}_{zy^2}(\tau_2) - \boldsymbol{S}_{zy^2}(\tau_1))'$$

$$+ \frac{2}{(\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor)^2} (S_y(\tau_2) - S_y(\tau_1)) (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1)) (\boldsymbol{S}_{zy}(\tau_2) - \boldsymbol{S}_{zy}(\tau_1))'$$

$$+ \frac{1}{(\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor)^2} (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1)) (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1))' (S_{y^2}(\tau_2) - S_{y^2}(\tau_1))$$

$$- \frac{1}{(\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor)^3} (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1)) (\boldsymbol{S}_z(\tau_2) - \boldsymbol{S}_z(\tau_1))' (S_y(\tau_2) - S_y(\tau_1))^2.$$

Next the bootstrap analogue of the subsample $t_{\beta_1}(\tau_1, \tau_2)$ statistic in (4.3) can be written as

$$t^*_{\beta_1}(\tau_1, \tau_2) := \frac{\boldsymbol{A}'_T(\tau_1, \tau_2) \mathbf{B}_T^{-1}(\tau_1, \tau_2) \boldsymbol{C}^*_T(\tau_1, \tau_2)}{\sqrt{\boldsymbol{A}'_T(\tau_1, \tau_2) \mathbf{B}_T^{-1}(\tau_1, \tau_2) \mathbf{D}^*_T(\tau_1, \tau_2) \mathbf{B}_T^{-1}(\tau_1, \tau_2) \boldsymbol{A}_T(\tau_1, \tau_2)}}.$$

where $\boldsymbol{C}^*_T(\tau_1, \tau_2)$ and $\mathbf{D}^*_T(\tau_1, \tau_2)$ denote the bootstrap analogues of $\boldsymbol{C}_T(\tau_1, \tau_2)$ and $\mathbf{D}_T(\tau_1, \tau_2)$, respectively, which can be written in the same way as was done above for the original $t_{\beta_1}(\tau_1, \tau_2)$ statistic by replacing the original y_t with the bootstrap regressand y_t^* in the definitions of the partial sum and product moment terms $S_y(\tau)$, $S_{y^2}(\tau)$, $\boldsymbol{S}_{zy}(\tau)$, $\boldsymbol{S}_{zy^2}(\tau)$, $\boldsymbol{S}_{zzy}(\tau)$ and $\boldsymbol{S}_{zzy^2}(\tau)$ above to obtain the bootstrap analogues of these quantities, denoted $S_y^*(\tau)$, $S_{y^2}^*(\tau)$, $\boldsymbol{S}_{zy}^*(\tau)$, $\boldsymbol{S}_{zy^2}^*(\tau)$, $\boldsymbol{S}_{zzy}^*(\tau)$ and $\boldsymbol{S}_{zzy^2}^*(\tau)$, respectively.

Finally, we denote by $\mathbf{W}_T$ the diagonal normalisation matrix $\mathrm{diag}(T^{-1/2-\delta_I}, T^{-1/2-\delta_{II}})$ and adopt the convention that, throughout, for random processes indexed by τ weak convergence is always understood in the functional sense, with reference to a space equipped with the Skorokhod topology.

S.3.2 Preparatory Lemmas

Lemma S.1 *Let $\tilde{d}_t$ be mildly integrated in the sense of Phillips and Magdalinos (2009), i.e. $\tilde{d}_t = \sum_{j=0}^{t-1} \varrho^j \psi_{t-j}$ where $\psi_t = \sum_{j \geq 0} \tilde{b}_j \tilde{v}_{t-j}$ with $\tilde{b}_j$ a 1-summable sequence of coefficients such that $\sum_{j \geq 0} \tilde{b}_j = \tilde{\omega} > 0$, and $\tilde{v}_t$ the innovations of the process v_t from Assumption 3. Then,*

1. $\frac{1}{T^{1/2+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{d}_t \Rightarrow \frac{\tilde{\omega}}{a} V(\tau)$ *and*

2. $T^{-\gamma/2} \tilde{d}_t$ *is uniformly L_4 bounded if assuming that*

$$\sup_{t \in \mathbb{Z}} \left| \mathrm{E}\left((\tilde{v}_t^2 - \mathrm{E}(\tilde{v}_t^2)) \tilde{v}_{t-j} \tilde{v}_{t-k} \right) \right| \leq C (jk)^{-1/2 - \vartheta/2}$$

for some $\vartheta > 0$. ∎

Lemma S.2 *Under Assumptions 1.1 and 2 – 6, we have as $T \to \infty$ that, jointly:*

1. $\frac{1}{\sqrt{T}} S_x(\tau) \Rightarrow \frac{\omega}{1-\rho} V(\tau)$;

2. $\frac{1}{\sqrt{T}} S_y(\tau) \Rightarrow U(\tau)$;

3. $\frac{1}{\sqrt{T}} \mathbf{W}_T \boldsymbol{S}_z(\tau) \Rightarrow \begin{pmatrix} 0 \\ \int_0^\tau Z(s) \mathrm{d}s \end{pmatrix}$;

4. $\begin{pmatrix} \frac{1}{T^{1+\delta_I}} & 0 \\ 0 & \frac{1}{T^{1/2+\delta_{II}}} \end{pmatrix} \boldsymbol{S}_{xz}(\tau) \Rightarrow \begin{pmatrix} K_{z_I x}(\tau) \\ \frac{\omega}{1-\rho} \int_0^\tau Z(s)\mathrm{d}V(s) \end{pmatrix}$;

5. $\mathbf{W}_T \boldsymbol{S}_{zz}(\tau) \mathbf{W}_T \Rightarrow \begin{pmatrix} K_{z_I^2}(\tau) & 0 \\ 0 & \int_0^\tau Z^2(s)\mathrm{d}s \end{pmatrix}$ with $\mathbf{W}_T$ defined in section S.3.1;

6. $\sup_{\tau \in [0,1]} \|\mathbf{W}_T \boldsymbol{S}_{zzy^2}(\tau) \mathbf{W}_T\| = O_p(1)$ and, in particular, $T^{-1-2\delta_I} S_{z_I^2 y^2}(\tau) \Rightarrow [G_I](1) \cdot \eta_I(\tau)$ and $T^{-1-2\delta_{II}} S_{z_{II}^2 y^2}(\tau) \Rightarrow \int_0^\tau Z^2(s)\mathrm{d}[U](s)$;

7. $\mathbf{W}_T \boldsymbol{S}_{zzy}(\tau) \mathbf{W}_T \Rightarrow 0$;

8. $\mathbf{W}_T \boldsymbol{S}_{zy}(\tau) \Rightarrow \begin{pmatrix} G_I(\tau) + \int_0^\tau b(s)\mathrm{d}K_{z_I x}(s) \\ \int_0^\tau Z(s)\mathrm{d}U(s) \end{pmatrix}$;

9. $\sup_{\tau \in [0,1]} \|T^{-1/2}\mathbf{W}_T \boldsymbol{S}_{zy^2}(\tau)\| = O_p(1)$ and, in particular, $T^{-1-\delta_{II}} S_{z_{II} y^2}(\tau) \Rightarrow \int_0^\tau Z(s)\mathrm{d}[U](s)$;

10. $\frac{1}{T} S_{y^2}(\tau) \Rightarrow [U](\tau)$.

■

Lemma S.3 *Under Assumptions 1.2 and 2 – 6, we have as $T \to \infty$ that, jointly:*

1. $\frac{1}{T\sqrt{T}} S_x(\tau) \Rightarrow \omega \int_0^\tau J_{c,\mathrm{H}}(s)\mathrm{d}s$;

2. $\frac{1}{\sqrt{T}} S_y(\tau) \Rightarrow U(\tau) + \omega \int_0^\tau b(s) J_{c,\mathrm{H}}(s)\mathrm{d}s$;

3. $T^{-1/2}\mathbf{W}_T \boldsymbol{S}_z(\tau) \Rightarrow \begin{pmatrix} 0 \\ \int_0^\tau Z(s)\mathrm{d}s \end{pmatrix}$;

4. $T^{-1}\mathbf{W}_T \boldsymbol{S}_{xz}(\tau) \Rightarrow \begin{pmatrix} 0 \\ \omega \int_0^\tau Z(s)J_{c,\mathrm{H}}(s)\mathrm{d}s \end{pmatrix}$;

5. $\mathbf{W}_T \boldsymbol{S}_{zz}(\tau) \mathbf{W}_T \Rightarrow \begin{pmatrix} K_{z_I^2}(\tau) & 0 \\ 0 & \int_0^\tau Z^2(s)\mathrm{d}s \end{pmatrix}$;

6. $\|\mathbf{W}_T \boldsymbol{S}_{zzy^2}(\tau) \mathbf{W}_T\| = O_p(1)$ and, in particular, $T^{-1-2\delta_{II}} S_{z_{II}^2 y^2}(\tau) \Rightarrow \int_0^\tau Z^2(s)\mathrm{d}[U](s)$;

7. $\mathbf{W}_T \boldsymbol{S}_{zzy}(\tau) \mathbf{W}_T \Rightarrow 0$;

8. $\|\mathbf{W}_T \boldsymbol{S}_{zy}(\tau)\| = O_p(1)$ and, in particular, $T^{-1/2-\delta_{II}} S_{z_{II} y}(\tau) \Rightarrow \int_0^\tau Z(s)\mathrm{d}U(s) + \omega \int_0^\tau Z(s)b(s)J_{c,\mathrm{H}}(s)\mathrm{d}s$;

9. $\sup_{\tau \in [0,1]} \|T^{-1/2}\mathbf{W}_T \boldsymbol{S}_{zy^2}(\tau)\| = O_p(1)$ and, in particular, $T^{-1-\delta_{II}} S_{z_{II} y^2}(\tau) \Rightarrow \int_0^\tau Z(s)\mathrm{d}[U](s)$;

10. $\frac{1}{T} S_{y^2}(\tau) \Rightarrow [U](\tau)$.

■

Lemma S.4 *Consider a two-dimensional zero-mean Gaussian process $(U^*(\tau), G_I^*(\tau))$, $\tau \in [0,1]$, with independent increments, independent components U^* and G_I^* and component variance functions $[U^*](\tau) := \int_0^\tau \left(h_{11}^2(s) + h_{12}^2(s)\right)\mathrm{d}s = [U](\tau)$ and $[G_I^*](\tau) := [G_I](\tau)$. Under Assumption 1.1, the bootstrap partial sum processes converge jointly as follows, under the null and under local alternatives:*

1. $T^{-1/2} S_y^*(\tau) \overset{w}{\Rightarrow}_p U^*(\tau)$;

2. $T^{-1/2-\delta_I} S_{z_I y}^*(\tau) \overset{w}{\Rightarrow}_p G_I^*(\tau)$ and $T^{-1/2-\delta_{II}} \sup_{\tau \in [0,1]} |S_{z_{II} y}^*(\tau)| = O_{p^*}(1)$ in P-probability.

Moreover, it holds that:

7

3. $\sup_{\tau \in [0,1]} \|\mathbf{W}_T(\mathbf{S}^*_{\boldsymbol{zz}y^2}(\tau) - \mathbf{S}_{\boldsymbol{zz}y^2}(\tau)]\mathbf{W}_T\| \overset{w}{\Rightarrow}_p 0;$

4. $\sup_{\tau \in [0,1]} \|\mathbf{W}_T\mathbf{S}^*_{\boldsymbol{zz}y}(\tau)\,\mathbf{W}_T\| \overset{w}{\Rightarrow}_p 0;$

5. $T^{-1/2}\sup_{\tau \in [0,1]} \|\mathbf{W}_T(\boldsymbol{S}^*_{\boldsymbol{z}y^2}(\tau) - \boldsymbol{S}_{\boldsymbol{z}y^2}(\tau))\| \overset{w}{\Rightarrow}_p 0;$

6. $T^{-1}\sup_{\tau \in [0,1]} |S^*_{y^2}(\tau) - S_{y^2}(\tau)| \overset{w}{\Rightarrow}_p 0.$

$\blacksquare$

Lemma S.5 *Let U^* be defined as in Lemma S.4. Under Assumptions 1.2 and 2 – 6, under the null hypothesis as well as under local alternatives, we have as $T \to \infty$ that, jointly:*

1. $T^{-1/2}S^*_y(\tau) \overset{w}{\Rightarrow}_p U^*(\tau);$

2. $\sup_{\tau \in [0,1]} \|\mathbf{W}_T\boldsymbol{S}^*_{\boldsymbol{z}y}(\tau)\| = O_{p^*}(1)$ *in* P-*probability and, in particular,* $T^{-1/2-\delta_{II}}S^*_{z_{II}y} \overset{w}{\Rightarrow}_p \int_0^\tau Z(s)\mathrm{d}U^*(s).$

Moreover, the convergence statements in parts 3, 4, 5 and 6 of Lemma S.4 hold.

$\blacksquare$

Lemma S.6 *Under the conditions of Proposition 1, it holds,*

(i) *under Assumption 1.1 that,*

$$\sup_{\substack{0 \leq \tau_1, \tau_2 \leq 1 \\ \tau_2 - \tau_1 \geq \Delta\tau}} \left| t^2_{\beta_1}(\tau_1, \tau_2) - \frac{Q_I^2(\tau_1, \tau_2)}{P_I(\tau_1, \tau_2)} \right| \overset{p}{\to} 0$$

where $t_{\beta_1}(\tau_1, \tau_2)$ is defined in (4.3),

$$Q_I(\tau_1, \tau_2) := \frac{1}{T^{1/2+\delta_I}}S_{z_I y}(\tau_2) - \frac{1}{T^{1/2+\delta_I}}S_{z_I y}(\tau_1)$$

and

$$P_I(\tau_1, \tau_2) := \frac{1}{T^{1+2\delta_I}}S_{z_I^2 y^2}(\tau_2) - \frac{1}{T^{1+2\delta_I}}S_{z_I^2 y^2}(\tau_1);$$

(ii) *under Assumption 1.2 that*

$$\sup_{\substack{0 \leq \tau_1, \tau_2 \leq 1 \\ \tau_2 - \tau_1 \geq \Delta\tau}} \left| t^2_{\beta_1}(\tau_1, \tau_2) - \frac{Q_{II}^2(\tau_1, \tau_2)}{P_{II}(\tau_1, \tau_2)} \right| \overset{p}{\to} 0$$

where

$$Q_{II}(\tau_1, \tau_2) := \frac{1}{T^{1/2+\delta_{II}}}S_{z_{II}y}(\tau_2) - \frac{1}{T^{1/2+\delta_{II}}}S_{z_{II}y}(\tau_1)$$
$$- \frac{1}{\tau_2 - \tau_1}\left(\frac{1}{T^{1+\delta_{II}}}S_{z_{II}}(\tau_2) - \frac{1}{T^{1+\delta_{II}}}S_{z_{II}}(\tau_1)\right)\left(\frac{1}{\sqrt{T}}S_y(\tau_2) - \frac{1}{\sqrt{T}}S_y(\tau_1)\right)$$

and

$$P_{II}(\tau_1, \tau_2) := \frac{1}{T^{1+2\delta_{II}}}S_{z_{II}^2 y^2}(\tau_2) - \frac{1}{T^{1+2\delta_{II}}}S_{z_{II}^2 y^2}(\tau_1)$$
$$- \frac{2}{\tau_2 - \tau_1}\left(\frac{1}{T^{1+\delta_{II}}}S_{z_{II}y^2}(\tau_2) - \frac{1}{T^{1+\delta_{II}}}S_{z_{II}y^2}(\tau_1)\right) \times$$
$$\times \left(\frac{1}{T^{1+\delta_{II}}}S_{z_{II}}(\tau_2) - \frac{1}{T^{1+\delta_{II}}}S_{z_{II}}(\tau_1)\right)$$
$$+ \frac{1}{(\tau_2 - \tau_1)^2}\left(\frac{1}{T^{1+\delta_{II}}}S_{z_{II}}(\tau_2) - \frac{1}{T^{1+\delta_{II}}}S_{z_{II}}(\tau_1)\right)^2\left(\frac{1}{T}S_{y^2}(\tau_2) - \frac{1}{T}S_{y^2}(\tau_1)\right).$$

■

Lemma S.7 *Under the conditions of Proposition 2 and under local alternatives, it holds,*

(i) *under Assumption 1.1 that*

$$\sup_{\substack{0 \le \tau_1, \tau_2 \le 1 \\ \tau_2 - \tau_1 \ge \bar{\Delta}\tau}} \left| t_{\beta_1}^{*2}(\tau_1, \tau_2) - \frac{Q_I^{*2}(\tau_1, \tau_2)}{P_I(\tau_1, \tau_2)} \right| \overset{w}{\Rightarrow}_p 0,$$

where $Q_I^*(\tau_1, \tau_2) := S_{z_I y}^*(\tau_2) - S_{z_I y}^*(\tau_1)$ and $P_I(\tau_1, \tau_2)$ is defined in Lemma S.6;

(ii) *under Assumption 1.2 that*

$$\sup_{\substack{0 \le \tau_1, \tau_2 \le 1 \\ \tau_2 - \tau_1 \ge \bar{\Delta}\tau}} \left| t_{\beta_1}^{*2}(\tau_1, \tau_2) - \frac{Q_{II}^{*2}(\tau_1, \tau_2)}{P_{II}(\tau_1, \tau_2)} \right| \overset{w}{\Rightarrow}_p 0,$$

where

$$
\begin{aligned}
Q_{II,b}^{*2}(\tau_1, \tau_2) \quad : \quad &= \frac{1}{T^{1/2+\delta_{II}}} S_{z_{II} y}^*(\tau_2) - \frac{1}{T^{1/2+\delta_{II}}} S_{z_{II} y}^*(\tau_1) \\
&\quad - \frac{1}{\tau_2 - \tau_1} \left(\frac{1}{T^{1+\delta_{II}}} S_{z_{II}}(\tau_2) - \frac{1}{T^{1+\delta_{II}}} S_{z_{II}}(\tau_1) \right) \left(\frac{1}{\sqrt{T}} S_y^*(\tau_2) - \frac{1}{\sqrt{T}} S_y^*(\tau_1) \right)
\end{aligned}
$$

and $P_{II}(\tau_1, \tau_2)$ is defined in Lemma S.6.

■

S.3.3   Proofs

Proof of Lemma S.1

Begin by using the Phillips-Solo decomposition to conclude that $\psi_t = \tilde{\omega}\tilde{v}_t + \Delta\bar{\bar{v}}_t$ with $\bar{\bar{v}}_t$ a linear process in $\tilde{v}_t$ with absolutely summable coefficients, where, recall, $(\tilde{u}_t, \tilde{v}_t)' := \mathbf{H}(t/T)\boldsymbol{\zeta}_t$.

To establish the first result, write then

$$\frac{1}{T^{1/2+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{d}_t = \tilde{\omega} \frac{1}{T^{1/2+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\sum_{j=0}^{t-1} \varrho^j \tilde{v}_{t-j} \right) + \frac{1}{T^{1/2+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\sum_{j=0}^{t-1} \varrho^j \Delta\bar{\bar{v}}_{t-j} \right). \tag{S.6}$$

To discuss the first term on the r.h.s. of (S.6), let $\tilde{V}_t = \sum_{j=1}^t \tilde{v}_t$ (with the usual convention that $\tilde{V}_0 = \tilde{V}_{-1} = \ldots = 0$) and note that, upon re-arranging sum elements, we have

$$
\begin{aligned}
\sum_{t=1}^{\lfloor \tau T \rfloor} \left(\sum_{j=0}^{t-1} \varrho^j \tilde{v}_{t-j} \right) &= \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\sum_{j=0}^{t-1} \varrho^j \Delta\tilde{V}_{t-j} \right) = \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \left(\sum_{j=0}^{t-1} \varrho^j \tilde{V}_{t-j} \right) = \sum_{t=0}^{\lceil \tau T \rceil - 1} \varrho^t \tilde{V}_{\lceil \tau T \rceil - t} \\
&= \left(\sum_{t=0}^{\lfloor \tau T \rfloor - 1} \varrho^t \right) \tilde{V}_{\lceil \tau T \rceil} - \sum_{t=1}^{\lfloor \tau T \rfloor - 1} \varrho^t \left(\sum_{j=\lfloor \tau T \rfloor - t}^{\lfloor \tau T \rfloor} \tilde{v}_j \right).
\end{aligned}
$$

It is easily shown that $\text{Var}\left(\left| \sum_{j=\lfloor \tau T \rfloor - t}^{\lfloor \tau T \rfloor} \tilde{v}_j \right| \right) \le Ct$, $\forall t \in \{1, \ldots, \lfloor \tau T \rfloor - 1\}$. Therefore, $\text{E}\left(\left| \sum_{j=\lfloor \tau T \rfloor - t}^{\lfloor \tau T \rfloor} \tilde{v}_j \right| \right) \le C\sqrt{t}$, and the arguments of Breitung and Demetrescu (2015, proof of Corollary 1.2) may be applied

9

to show the second summand to be dominated uniformly in t. We may therefore focus on

$$\frac{1}{T^{1/2+\gamma}}\left(\sum_{t=0}^{\lfloor\tau T\rfloor-1}\varrho^t\right)\tilde{V}_{\lfloor\tau T\rfloor} = \frac{1}{a\sqrt{T}}\tilde{V}_{\lfloor\tau T\rfloor} - \frac{1}{a\sqrt{T}}\varrho^{\lfloor\tau T\rfloor}\tilde{V}_{\lfloor\tau T\rfloor}$$

The first summand on the r.h.s.ï¿œleads to the desired weak convergence, but the second must be shown to vanish uniformly in $\tau\in[0,1]$. While this is trivial at $\tau>0$, it requires more care at 0, where we examine $\limsup_{\tau\to 0}\left|\frac{1}{a\sqrt{T}}\varrho^{\lfloor\tau T\rfloor}\tilde{V}_{\lfloor\tau T\rfloor}\right| \le \limsup_{\tau\to 0}\left|\frac{1}{a\sqrt{T}}\tilde{V}_{\lfloor\tau T\rfloor}\right|$, which, in turn, is easily shown to vanish as required and we may write

$$\frac{1}{T^{1/2+\gamma}}\sum_{t=1}^{\lfloor\tau T\rfloor}\left(\sum_{j=0}^{t-1}\varrho^j\tilde{v}_{t-j}\right) = \frac{1}{a\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}\tilde{v}_t + o_p(1)$$

with the o_p term uniform in $\tau\in[0,1]$ as required. To show the second term on the r.h.s. of (S.6) to vanish uniformly in τ, notice that

$$\sum_{t=1}^{\lfloor\tau T\rfloor}\left(\sum_{j=0}^{t-1}\varrho^j\Delta\bar{\bar{v}}_{t-j}\right) = \sum_{t=1}^{\lfloor\tau T\rfloor}\left(\bar{\bar{v}}_t - \varrho^{t-1}\bar{\bar{v}}_0 - (1-\varrho)\sum_{j=0}^{t-2}\varrho^j\bar{v}_{t-1-j}\right)$$

where $\bar{d}_t = \sum_{j=0}^{t-2}\varrho^j\bar{v}_{t-1-j}$ is mildly integrated. Now, $\sup_\tau\left|\sum_{t=1}^{\lfloor\tau T\rfloor}\bar{\bar{v}}_t\right| = O_p\left(\sqrt{T}\right)$, $\bar{\bar{v}}_0\sum_{t=1}^{\lfloor\tau T\rfloor}\varrho^{t-1} = O_p\left(T^\gamma\right)$ and, like above, it can be shown that $\frac{1}{T^{1/2+\gamma}}\sum_{t=1}^{\lfloor\tau T\rfloor}\bar{d}_t = C\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}\tilde{v}_t + o_p(1)$ such that $(1-\varrho)\sum_{t=1}^{\lfloor\tau T\rfloor}\sum_{j=0}^{t-2}\varrho^j\bar{v}_{t-1-j} = O_p\left(\sqrt{T}\right)$ uniformly as required.

To show the second result, notice that $\bar{\bar{v}}_t$ in the the Phillips-Solo decomposition of ψ_t is uniformly L_4-bounded, just like $\tilde{v}_t$. Examining

$$\tilde{d}_t = \tilde{\omega}\sum_{j=0}^{t-1}\varrho^j\tilde{v}_{t-j} + \sum_{j=0}^{t-1}\varrho^j\Delta\bar{\bar{v}}_t,$$

it is easily seen that $\sum_{j=0}^{t-1}\varrho^j\Delta\bar{\bar{v}}_t$ is uniformly L_4-bounded, so it suffices to show that the first summand on the r.h.s. is uniformly L_4-bounded upon division by $T^{\gamma/2}$. Write to this end

$$\mathrm{E}\left(\left(\sum_{j=0}^{t-1}\varrho^j\tilde{v}_{t-j}\right)^4\right) = \sum_{j=0}^{t-3}\sum_{k=0}^{t-3}\sum_{l=0}^{t-3}\sum_{m=0}^{t-3}\varrho^j\varrho^k\varrho^l\varrho^m\,\mathrm{E}\left(\tilde{v}_{t-1-j}\tilde{v}_{t-1-k}\tilde{v}_{t-1-l}\tilde{v}_{t-1-m}\right)$$

which, upon exploiting the MD property of $\tilde{v}_t$, gives

$$\begin{aligned}\mathrm{E}\left(\left(\sum_{j=0}^{t-1}\varrho^j\tilde{v}_{t-j}\right)^4\right) &= \sum_{j=0}^{t-3}\varrho^{4j}\,\mathrm{E}\left(\tilde{v}_{t-1-j}^4\right) + 3\sum_{\substack{j=0\\j\ne k}}^{t-3}\sum_{k=0}^{t-3}\varrho^{3j}\varrho^k\,\mathrm{E}(\tilde{v}_{t-j}^3\tilde{v}_{t-k}) \\ &\quad + 3\sum_{\substack{j=0\\j\ne k}}^{t-3}\sum_{k=0}^{t-3}\varrho^{2j}\varrho^{2k}\,\mathrm{E}\left(\tilde{v}_{t-1-j}^2\tilde{v}_{t-1-k}^2\right) + 6\sum_{\substack{j=0\\j\ne k\ne l}}^{t-3}\sum_{k=0}^{t-3}\sum_{l=0}^{t-3}\varrho^{2j}\varrho^k\varrho^l\,\mathrm{E}\left(\tilde{v}_{t-1-j}^2\tilde{v}_{t-1-k}\tilde{v}_{t-1-l}\right).\end{aligned}$$

The Hï¿œœlder inequality leads to $\left|\mathrm{E}(\tilde{v}_{t-j}^3\tilde{v}_{t-k})\right| \le \mathrm{E}\left(\left|\tilde{v}_{t-j}^3\tilde{v}_{t-k}\right|\right) \le \left\|\tilde{v}_{t-j}^3\right\|_{4/3}\|\tilde{v}_{t-k}\|_4$ where the

latter L norms are uniformly bounded, so

$$\left| \sum_{\substack{j=0 \\ j\neq k}}^{t-3} \sum_{k=0}^{t-3} \varrho^{3j}\varrho^{k}\, \mathrm{E}(\tilde{v}_{t-j}^{3}\tilde{v}_{t-k}) \right| \leq C \sum_{\substack{j=0 \\ j\neq k}}^{t-3}\sum_{k=0}^{t-3} \varrho^{3j}\varrho^{k} \leq C\left(\sum_{j=0}^{t-3}\varrho^{3j}\right)\left(\sum_{k=0}^{t-3}\varrho^{k}\right) \leq CT^{2\gamma},$$

and, analogously, $3\sum_{\substack{j=0 \\ j\neq k}}^{t-3}\sum_{k=0}^{t-3}\varrho^{2j}\varrho^{2k}\,\mathrm{E}\left(\tilde{v}_{t-1-j}^{2}\tilde{v}_{t-1-k}^{2}\right) = O\left(T^{2\gamma}\right)$. Furthermore, since $\mathrm{E}\left(\tilde{v}_{t-1-k}\tilde{v}_{t-1-l}\right) = 0$ and $\mathrm{E}\left(\tilde{v}_{t-1-j}^{2}\tilde{v}_{t-1-k}\tilde{v}_{t-1-l}\right) = 0$ for $j\neq k\neq l$ whenever $j\geq l$ or $j\geq k$,

$$
\begin{aligned}
\left|\mathrm{E}\left(\tilde{v}_{t-1-j}^{2}\tilde{v}_{t-1-k}\tilde{v}_{t-1-l}\right)\right| &\leq \mathrm{Var}\left(\tilde{v}_{t-1-j}^{2}\right)\left|\mathrm{E}\left(\tilde{v}_{t-1-k}\tilde{v}_{t-1-l}\right)\right| \\
&\quad + \left|\mathrm{E}\left(\left(\tilde{v}_{t-1-j}^{2} - \mathrm{Var}\left(\tilde{v}_{t-1-j}^{2}\right)\right)\tilde{v}_{t-1-k}\tilde{v}_{t-1-l}\right)\right| \\
&\leq \frac{C}{\sqrt{(k-j)^{1+\vartheta}(l-j)^{1+\vartheta}}}
\end{aligned}
$$

for all t, and thus, with the variance of $\tilde{v}_{t}$ uniformly bounded under our assumptions,

$$
\left| \sum_{\substack{j=0 \\ j\neq k\neq l}}^{t-3}\sum_{k=0}^{t-3}\sum_{l=0}^{t-3} \varrho^{2j}\varrho^{k}\varrho^{l}\,\mathrm{E}\left(\tilde{v}_{t-1-j}^{2}\tilde{v}_{t-1-k}\tilde{v}_{t-1-l}\right) \right| \leq
$$

$$
\leq C\sum_{j=0}^{t-3}\varrho^{4j}\sum_{k=j+1}^{t-3}\sum_{\substack{l=j+1 \\ k\neq l}}^{t-3}\frac{\varrho^{k-j}\varrho^{l-j}}{\sqrt{(k-j)^{1+\vartheta}(l-j)^{1+\vartheta}}} \leq C\sum_{j=0}^{t-3}\varrho^{4j}\sum_{k=j+1}^{t-3}\sum_{\substack{l=j+1 \\ k\neq l}}^{t-3}\frac{\varrho^{k-j}\varrho^{l-j}}{\sqrt{(k-j)^{1+\vartheta}(l-j)^{1+\vartheta}}}
$$

$$
\leq C\sum_{j=0}^{t-3}\varrho^{4j}\sum_{k=j+1}^{t-3}\sum_{l=j+1}^{t-3}\frac{\varrho^{k-j}\varrho^{l-j}}{\sqrt{(k-j)^{1+\vartheta}(l-j)^{1+\vartheta}}} \leq C\sum_{j=0}^{t-3}\varrho^{4j}\left(\sqrt{\sum_{k=1}^{t-j-3}\varrho^{2k}}\sqrt{\sum_{k=1}^{t-j-3}\frac{1}{k^{1+\vartheta}}}\right)^{2}
$$

$$
\leq C\left(\sum_{j=0}^{T-1}\varrho^{4j}\right)\left(\sum_{k=0}^{T-1}\varrho^{2k}\right)\left(\sum_{k=1}^{T-1}\frac{1}{k^{1+\vartheta}}\right).
$$

Summing up, we have for all $t = 2,\ldots,T$ that $\mathrm{E}\left(\left(\sum_{j=0}^{t-1}\varrho^{j}\tilde{v}_{t-j}\right)^{4}\right) \leq CT^{2\gamma}$ as required. $\blacksquare$

Proof of Lemma 1

Proof of part 1 We work under Assumption 1.1 such that ξ_{t} is a stable autoregression, and begin by showing that Assumption 4.1 is fulfilled. Notice that

$$z_{I,t-1} = \xi_{t-1} - \varrho^{t-2}\xi_{0} + (\varrho-1)\sum_{j=0}^{t-3}\varrho^{j}\xi_{t-2-j},$$

where

$$(\varrho-1)\sum_{j=0}^{t-3}\varrho^{j}\xi_{t-2-j} = -\frac{a}{T^{\gamma}}\sum_{j=0}^{t-3}\varrho^{j}\xi_{t-2-j} = -\frac{a}{T^{\gamma}}d_{t-2}$$

with d_{t-2} zero-mean mildly integrated.

Then, $\mathrm{E}\left(\zeta_{t}|\zeta_{t-1},\zeta_{t-2},\ldots,z_{I,t-1},z_{I,t-2},\ldots\right) = 0$ obviously holds; ξ_{t-1} and $\varrho^{t-2}\xi_{0}$ are uniformly L_{4}-bounded, and so is $\frac{1}{T^{\gamma/2}}d_{t-2}$ (see Lemma S.1), such that $T^{-\delta_{I}}z_{I,t-1}$, with $\delta_{I} = 0$, is itself uniformly L_{4}-bounded due to Minkowski's norm inequality.

Furthermore,

$$\sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \right| \leq \sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1} \right| + \sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \varrho^{t-2} \xi_0 \right| + \sup_{\tau \in [0,1]} \left| \frac{a}{T^{1+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} d_{t-2} \right| \xrightarrow{p} 0$$

since $\sum_{t=1}^{\lfloor \tau T \rfloor} \varrho^{t-2} = O(T^\gamma) = o(T)$ and $\frac{a}{T^{1/2+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} d_{t-2} \Rightarrow \frac{\omega}{1-\rho} V(\tau)$, see Lemma S.1.

Similarly,

$$\sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t^2 \right| \leq \frac{1}{T} \sup_{\tau \in [0,1]} \sum_{t=1}^{\lfloor \tau T \rfloor} \left(\left| \xi_{t-1} u_t^2 \right| + \left| \xi_0 \right| \varrho^{t-2} u_t^2 + \left| \frac{1}{T^\gamma} d_{t-2} u_t^2 \right| \right)$$

$$\leq \frac{1}{T} \sum_{t=1}^{T} \left(\left| \xi_{t-1} u_t^2 \right| + \left| \xi_0 \right| \varrho^{t-2} u_t^2 + \left| \frac{1}{T^\gamma} d_{t-2} u_t^2 \right| \right).$$

But all summands on the r.h.s. are easily shown to have bounded expectation, and Markov's inequality then indicates that $\sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t^2 \right| = O_p(1)$ as required.

Moving on to assumption 4.2 (a), write

$$\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \xi_{t-1} = \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1}^2 - \frac{\xi_0}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \varrho^{t-2} \xi_{t-1} - \frac{1}{T^{1+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} d_{t-2} \xi_{t-1}$$

where it is easily shown using standard arguments that $\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1}^2 \Rightarrow \bar{\sigma}_\xi^2(\rho)[V](\tau)$ with $\bar{\sigma}_\xi^2(\rho)$ the sum of the squared coefficients of $(1-\rho L)^{-1} B(L)$.

Moreover,

$$\sup_{\tau \in [0,1]} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \varrho^{t-2} \xi_{t-1} \right| \leq \sum_{t=1}^{T} \varrho^{t-2} \left| \xi_{t-1} \right|$$

where the expectation of the r.h.s. is bounded by $C \sum_{t=1}^{T} \varrho^{t-2}$ since ξ_{t-1} is uniformly L_4-bounded, so $\sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \varrho^{t-2} \xi_{t-1} \right| = O_p \left(\frac{1}{T} \sum_{t=1}^{T} \varrho^{t-2} \right) = O_p \left(T^{\gamma-1} \right) = o_p(1)$ as required, and the term $\frac{1}{T^{1+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} d_{t-2} \xi_{t-1}$ is shown to vanish analogously given that $T^{-\gamma/2} d_{t-2}$ is uniformly L_4-bounded.

The term $\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2$ in (b) can be shown to have the same limit behaviour using similar derivations.

For (c), consider

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1} u_t - \frac{\xi_0}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} \varrho^{t-2} u_t - \frac{1}{T^{1/2+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} d_{t-2} u_t.$$

The second and third term on the r.h.s. are easily shown to vanish using Doob's martingale inequality. To discuss the convergence of the first, note that $(u_t, v_t, \xi_t u_t)'$ form an MD array. Since we are dealing with partial sums of positive semi-definite matrices, uniform convergence of

$$\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \begin{pmatrix} u_t^2 & v_t u_t & \xi_{t-1} u_t^2 \\ v_t u_t & v_t^2 & \xi_{t-1} v_t u_t \\ \xi_{t-1} u_t^2 & \xi_{t-1} v_t u_t & \xi_{t-1}^2 u_t^2 \end{pmatrix}$$

can be established, and a MD invariance principle (e.g. Boswijk *et al.*, 2016, Lemma 1) leads as required to

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} \begin{pmatrix} u_t \\ v_t \\ \xi_{t-1} u_t \end{pmatrix} \Rightarrow \begin{pmatrix} U(\tau) \\ V(\tau) \\ G_I(\tau) \end{pmatrix}.$$

Assumption 4.2 part (d) is established along the lines of the proof of (b) by first showing that

$$\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 u_t^2 = \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1}^2 u_t^2 + o_p(1)$$

where the o_p term is uniform in τ. This is straightforward and, hence, we omit the details. Here, $\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1}^2 u_t^2$ clearly converges uniformly to the quadratic variation of G_I from (c).

To complete the proof of part 1 of this lemma, notice that Assumption 4.3 refers to the near-integrated case, while Assumption 5 is obviously fulfilled with $\delta_{II} = 0$ and $Z(\tau) = \sin(k\frac{\pi}{2}\tau)$ irrespective of which assumption, 1.1 or 1.2, holds true. Finally, Assumption 6 follows along the lines of the derivations above; the key observation is that the type-II instrument only adds a deterministic weight component to $\frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}$ and $\frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t^2$ and the order of magnitude of the maxima do not change.

Proof of part 2 We now work under Assumption 1.2, and use the Phillips-Solo decomposition to write $v_t = \omega \tilde{v}_t + \Delta \bar{v}_t$ where $\bar{v}_t$ is a linear process with absolutely summable coefficients driven by $\tilde{v}_t$. It holds that

$$z_{I,t-1} = \sum_{j=0}^{t-2} \varrho^j \Delta \xi_{t-1-j} = \sum_{j=0}^{t-2} \varrho^j \left(v_{t-1-j} - \frac{c}{T} \xi_{t-2-j} \right) = \omega \tilde{z}_{t-1} + \sum_{j=0}^{t-2} \varrho^j \left(\Delta \bar{v}_{t-1-j} - \frac{c}{T} \xi_{t-2-j} \right)$$

with $\tilde{z}_{t-1} = \sum_{j=0}^{t-2} \varrho^j \tilde{v}_{t-j}$ and $\tilde{z}_0 = \tilde{z}_{-1} = \ldots = 0$.

We now deal with assumption 4.1. With $\delta_I = \gamma/2$, we have

$$\sup_{\tau \in [0,1]} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \right| \le \omega \sup_{\tau \in [0,1]} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1} \right| + \sup_{\tau \in [0,1]} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \sum_{j=0}^{t-2} \varrho^j \Delta \bar{v}_{t-1-j} \right| + \frac{c}{T} \sup_{\tau \in [0,1]} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \sum_{j=0}^{t-2} \varrho^j \xi_{t-2-j} \right|,$$

where $\frac{1}{T^{1/2+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1} \Rightarrow CV(\tau)$ (see Lemma S.1), and $\sup_{\tau \in [0,1]} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \sum_{j=0}^{t-2} \varrho^j \Delta \bar{v}_{t-1-j} \right| = O_p \left(\max \left\{ \sqrt{T}, T^{\gamma} \right\} \right)$ like in the proof of Lemma S.1; moreover, $\frac{c}{T} \sup_{\tau \in [0,1]} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} \sum_{j=0}^{t-2} \varrho^j \xi_{t-2-j} \right|$ is easily shown to be of order $O_p \left(T^{1/2+\gamma} \right)$ itself, such that

$$\frac{1}{T^{1+\gamma/2}} \sup_{\tau \in [0,1]} \left| \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \right| = O_p \left(T^{\gamma/2-1/2} \right) = o_p(1)$$

as required. Using similar arguments, it follows that $\sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t^2 \right| = O_p(1)$, while $\mathrm{E} \left(\zeta_t | \zeta_{t-1}, \zeta_{t-2}, \ldots, z_{I,t-1}, z_{I,t-2}, \ldots \right) = 0$ holds as in the proof of part 1.

Assumption 4.2 refers to the stable autoregression case, so we move on to dealing with assumption 4.3. We have for (a) that

$$\sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} \xi_{t-1} = \omega \sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1} \xi_{t-1} + \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1} \sum_{j=0}^{t-2} \varrho^j \Delta \bar{v}_{t-1-j} - \frac{c}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1} \sum_{j=0}^{t-2} \varrho^j \xi_{t-2-j}$$

where, since $\sup_{1 \le t \le T} |\xi_t| = O_p \left(\sqrt{T} \right)$, it can be shown as above that $\sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1} \sum_{j=0}^{t-2} \varrho^j \Delta \bar{v}_{t-1-j} = O_p \left(\max \left\{ T, T^{1/2+\gamma} \right\} \right)$ and $\frac{c}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \xi_{t-1} \sum_{j=0}^{t-2} \varrho^j \xi_{t-2-j} = O_p \left(T^{1+\gamma} \right)$ uniformly, such that they van-

13

ish upon division by $T^{3/2+\gamma/2}$. Furthermore, with $\tilde{S}_{t-1} = \sum_{j=1}^{t-1} \tilde{z}_j$ (and $S_0 = 0$),

$$\sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1}\xi_{t-1} = \sum_{t=2}^{\lfloor \tau T \rfloor} \left(\tilde{S}_{t-1} - \tilde{S}_{t-2}\right)\xi_{t-1} = \tilde{S}_{[\tau T]-1}\xi_{t-1} - \sum_{t=2}^{\lfloor \tau T \rfloor - 1} \tilde{S}_{t-1}\Delta\xi_t.$$

But we know from the proof of Lemma S.1 that $\frac{1}{T^{1/2+\gamma}}\tilde{S}_{[\tau T]} \Rightarrow CV(\tau)$ and it immediately follows that $\sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1}\xi_{t-1} = O_p\left(T^{1+\gamma}\right)$ uniformly, as required.

In establishing (b), it is not difficult to show that, given the Phillips-Solo decomposition of v_t,

$$\frac{1}{T^{1+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 = \frac{\omega^2}{T^{1+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1}^2 + o_p(1)$$

uniformly in τ. The leading term may be written with $\tilde{z}_0 = 0$ as

$$\frac{\omega^2}{T^{1+\gamma}} \sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1}^2 = \frac{\omega^2}{T^{1+\gamma}} \sum_{t=2}^{\lfloor \tau T \rfloor}\sum_{j=0}^{t-2} \varrho^{2j}\tilde{v}_{t-1-j}^2 + \frac{2\omega^2}{T^{1+\gamma}} \sum_{t=3}^{\lfloor \tau T \rfloor}\sum_{j=0}^{t-3}\sum_{k=j+1}^{t-2} \varrho^j\varrho^k\tilde{v}_{t-1-j}\tilde{v}_{t-1-k}, \qquad \text{(S.7)}$$

and, after, re-arranging the summands of the first term on the r.h.s. we have

$$\frac{\omega^2}{T^{1+\gamma}} \sum_{t=2}^{\lfloor \tau T \rfloor}\sum_{j=0}^{t-2} \varrho^{2j}\tilde{v}_{t-1-j}^2 = \frac{1}{1-\varrho}\frac{\omega^2}{T^{1+\gamma}} \sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{v}_{t-1}^2 + \frac{1}{1-\varrho}\frac{\omega^2}{T^{1+\gamma}} \sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{v}_{t-1}^2\varrho^{2t-2}$$

where $T^{-1}\sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{v}_{t-1}^2 \xrightarrow{p} \int_0^\tau \left(h_{21}^2(s) + h_{22}^2(s)\right)\mathrm{d}s$ and

$$0 \le \frac{1}{T}\sum_{t=2}^{\lfloor \tau T \rfloor} \tilde{v}_{t-1}^2\varrho^{2t-2} \le \frac{1}{T}\sum_{t=2}^{T} \tilde{v}_{t-1}^2\varrho^{2t-2} = o_p(1)$$

since $\mathrm{E}\left(\frac{1}{T}\sum_{t=2}^{T}\tilde{v}_{t-1}^2\varrho^{2t-2}\right) \le \sup_{t=1,\ldots,T}\mathrm{E}\left(\tilde{v}_{t-1}^2\right)\frac{1}{T}\sum_{t=2}^{T}\varrho^{2t-2} = O\left(T^{\gamma-1}\right)$, so the first term on the r.h.s. of (S.7) delivers the desired limit (point wise in τ). The second term on the r.h.s. of (S.7) may be expressed as

$$\frac{2\omega^2}{T^{1+\gamma}} \sum_{t=3}^{\lfloor \tau T \rfloor}\sum_{j=0}^{t-3} \tilde{v}_{t-1-j}\varrho^{2j}\sum_{k=1}^{t-j-2} \varrho^k\tilde{v}_{t-j-1-k} = \frac{2\omega^2}{T^{1+\gamma}} \sum_{t=3}^{\lfloor \tau T \rfloor} \tilde{v}_{t-1}\sum_{s=t-1}^{[\tau T]-1} \varrho^{2(s-t+1)}\sum_{k=1}^{t-2} \varrho^k\tilde{v}_{t-1-k}$$

after re-arranging terms. Notice that $q_t := \tilde{v}_{t-1}\sum_{s=t-1}^{[\tau T]-1} \varrho^{2(s-t+1)}\sum_{k=1}^{t-2} \varrho^k\tilde{v}_{t-1-k}$ has the MD property, where, after some algebra as in the proof of Lemma S.1, $T^{-\gamma/2}\sum_{k=1}^{t-2} \varrho^k\tilde{v}_{t-1-k}$ is seen to be uniformly L_4-bounded. Then, $T^{-3\gamma/2}\sum_{s=t-1}^{[\tau T]-1} \varrho^{2(s-t+1)}\sum_{k=1}^{t-2} \varrho^k\tilde{v}_{t-1-k}$ is uniformly L_4-bounded due to the Minkowski's norm inequality, so $T^{-3\gamma/2}q_t$ is uniformly L_2 bounded. Therefore, the variance of the sum of the q_t is of order $O\left(T^{1+3\gamma}\right)$ due to the MD property, and the entire term is then $O_p\left(T^{\gamma/2-1/2}\right)$, also pointwise in τ.

Uniform convergence of $\frac{\omega^2}{T^{1+\gamma}}\sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1}^2$ follows from the fact that both the partial sums and the limit are nondecreasing functions of τ.

For (c), we may similarly show that

$$\sup_{\tau \in [0,1]}\left|\frac{1}{T^{1/2+\gamma/2}}\sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}u_t\right| = \sup_{\tau \in [0,1]}\left|\frac{\omega}{T^{1/2+\gamma/2}}\sum_{t=1}^{\lfloor \tau T \rfloor} \tilde{z}_{t-1}u_t\right| + O_p(1),$$

where the $\tilde{z}_{t-1}u_t$ have the MD property; since $\mathrm{E}\left(\frac{1}{T^{1+\gamma}}\left(\sum_{t=1}^{T}\tilde{z}_{t-1}u_t\right)^2\right) = \frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\mathrm{E}\left(\tilde{z}_{t-1}^2 u_t^2\right) \leq$

$\frac{1}{T}\sum_{t=1}^{T}\sqrt{\left\|\frac{1}{T^{\gamma/2}}\tilde{z}_{t-1}\right\|_4^4 \left\|u_t^2\right\|_4^4}$ is bounded, we may apply Doob's martingale inequality to conclude that $\sup_{\tau\in[0,1]}\left|\frac{1}{T^{1/2+\gamma/2}}\sum_{t=1}^{\lfloor\tau T\rfloor}z_{I,t-1}u_t\right| = O_p(1)$.

In dealing with (d), it can be shown that

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{\lfloor\tau T\rfloor}z_{I,t-1}^2 u_t^2 = \frac{\omega^2}{T^{1+\gamma}}\sum_{t=1}^{T}\tilde{z}_{t-1}^2 u_t^2 + O_p(1)$$

where $\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\tilde{z}_{t-1}^2 u_t^2 = O_p(1)$ due to the Markov inequality.

Assumption 5 obviously holds, and Assumption 6 can again be reduced to the behaviour of $\frac{1}{T^{1+\delta_I}}\sum_{t=1}^{\lfloor\tau T\rfloor}z_{I,t-1}$ and $\frac{1}{T^{1+\delta_I}}\sum_{t=1}^{\lfloor\tau T\rfloor}z_{I,t-1}u_t^2$ like in the proof of part 1, and so we omit the details. ∎

Proof of Lemma S.2

1. Under Assumption 1.1, we have that $x_t = \sum_{j=0}^{t-1}\rho^j v_{t-j} + \rho^t x_0$. Using the Phillips-Solo decomposition and the 1-summability of the coefficients of $B(L)$ (and thus of $(1-\rho L)^{-1}B(L)$ when $|\rho| < 1$), it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}x_{t-1} = \frac{1}{1-\rho}\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}v_{t-1} + \frac{1}{\sqrt{T}}\bar{v}_{\lfloor\tau T\rfloor} - \frac{1}{\sqrt{T}}\bar{v}_0 + o_p(1)$$

uniformly in $\tau \in [0,1]$, where $\bar{v}_t$ is a linear process driven by v_t with absolutely summable coefficients. Then, the uniform L_4-boundedness of ζ_t and the boundedness of the function $\mathbf{H}(\cdot)$ imply uniform L_4-boundedness of $\bar{v}_t$ such that $\max_{\tau\in[0,1]}\left|\bar{v}_{\lfloor\tau T\rfloor}\right| = o_p\left(\sqrt{T}\right)$ and the result then follows from (2.6).

2. Notice that

$$\frac{1}{\sqrt{T}}S_y(\tau) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}u_t + \frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}b\left(\frac{t}{T}\right)x_{t-1}\right)$$

where we recall that $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}u_t \Rightarrow U(\tau)$. Because of the piecewise Lipschitz continuity of $b(\cdot)$ and the well-known Hölder continuity of any order $\alpha \in (0,1/2)$ of Gaussian processes with independent increments, and thus of $V(\cdot)$, the Stjeltjes integral $\frac{\omega}{1-\rho}\int_0^\tau b(s)\mathrm{d}V(s)$ exists (pathwise) and, with item 1 of this Lemma, it follows that $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor\tau T\rfloor}b\left(\frac{t}{T}\right)x_{t-1} \Rightarrow \frac{\omega}{1-\rho}\int_0^\tau b(s)\mathrm{d}V(s)$ as required for the result.

3. Follows directly from Assumptions 4 and 5 and the CMT.

4. Follows directly from Assumptions 4 and 5 and item 1 of this Lemma.

5. Follows directly from Assumptions 4-6 and the CMT.

6. We have elementwise

$$\frac{1}{T^{1+2\delta_I}}\sum_{t=1}^{\lfloor\tau T\rfloor}z_{I,t-1}^2 y_t^2 = \frac{1}{T^{1+2\delta_I}}\sum_{t=1}^{\lfloor\tau T\rfloor}z_{I,t-1}^2 u_t^2 + \frac{2}{\sqrt{T}}\left(\frac{1}{T^{1+2\delta_I}}\sum_{t=1}^{\lfloor\tau T\rfloor}z_{I,t-1}^2 u_t b\left(\frac{t}{T}\right)x_{t-1}\right)$$

$$+\frac{1}{T}\left(\frac{1}{T^{1+2\delta_I}}\sum_{t=1}^{\lfloor\tau T\rfloor}z_{I,t-1}^2 b^2\left(\frac{t}{T}\right)x_{t-1}^2\right). \tag{S.8}$$

15

The first summand on the r.h.s. gives the desired limit directly from Assumption 4. For the second summand, an application of Markov's inequality shows it to be $o_p(1)$ uniformly in τ, as

$$\max_{\tau \in [0,1]} \left| \frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{[\tau T]} z_{I,t-1}^2 u_t b\left(\frac{t}{T}\right) x_{t-1} \right| \leq \max_{\tau \in [0,1]} |b(\tau)| \max_{\tau \in [0,1]} \frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{[\tau T]} \left| z_{I,t-1}^2 u_t x_{t-1} \right|$$

$$\leq C \frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{T} \left| z_{I,t-1}^2 u_t x_{t-1} \right|$$

which is independent of τ and has expectation bounded by $C \frac{1}{T} \sum_{t=1}^{T} \sqrt[4]{\mathrm{E}\left(\frac{z_{I,t-1}^4}{T^{4\delta_I}}\right) \mathrm{E}\left(u_t^4\right) \mathrm{E}\left(x_{t-1}^4\right)} = O(1)$, given the uniform L_4-boundedness of $T^{-\delta_I} z_{I,t-1}$, of u_t and, under Assumption 1.1, of x_t. For the third summand, write

$$\max_{\tau \in [0,1]} \left| \frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 b^2\left(\frac{t}{T}\right) x_{t-1}^2 \right| \leq \max_{\tau \in [0,1]} b^2(\tau) \frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 x_{t-1}^2 \leq C \frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{T} z_{I,t-1}^2 x_{t-1}^2,$$

where again the uniform upper bound is $O_p(1)$ since $\mathrm{E}\left(z_{I,t-1}^2 x_{t-1}^2\right) \leq \sqrt{\mathrm{E}\left(z_{I,t-1}^4\right) \mathrm{E}\left(x_{t-1}^4\right)}$ is uniformly bounded. Next,

$$\frac{1}{T^{1+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 y_t^2 = \frac{1}{T^{1+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 u_t^2 + \frac{2}{T^{3/2+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 u_t b\left(\frac{t}{T}\right) x_{t-1}$$

$$+ \frac{1}{T} \left(\frac{1}{T^{1+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 b^2\left(\frac{t}{T}\right) x_{t-1}^2 \right).$$

The first summand delivers the desired limit since,

$$\frac{1}{T^{1+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 u_t^2 = \frac{1}{T^{1+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 \mathrm{E}\left(u_t^2\right) + \frac{1}{T^{1+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 \left(u_t^2 - \mathrm{E}\left(u_t^2\right)\right)$$

where $\sup_{\tau \in [0,1]} \left| \frac{1}{T^{1+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 \left(u_t^2 - \mathrm{E}\left(u_t^2\right)\right) \right|$ vanishes, since the assumption that we have $\sup_t \mathrm{E}\left|\mathrm{E}\left(\zeta_t \zeta_t' - \mathbf{I}_2 | \zeta_{t-m}, \zeta_{t-m-1}, \ldots\right)\right| \to 0$ as $m \to \infty$ allows us to use the arguments of Hansen (1992, Theorem 3.3); the second and the third summands are dealt with analogously to those in (S.8) and we omit the details here. Finally, the result that

$$\max_{\tau \in [0,1]} \left| \frac{1}{T^{1+\delta_I+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} z_{II,t-1} y_t^2 \right| = O_p(1)$$

follows directly from the positive definiteness of $\mathbf{W}_T \mathbf{S}_{zzy^2} \mathbf{W}_T$ given the previous conclusions regarding its diagonal elements.

7. We have elementwise

$$\frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 y_t = \frac{1}{\sqrt{T}} \left(\frac{1}{T^{1/2+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 u_t \right) + \frac{1}{T^{3/2+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 b\left(\frac{t}{T}\right) x_{t-1}.$$

We note that $T^{-2\delta_I} z_{I,t-1}^2 u_t$ is a uniformly L_2-bounded MD sequence; therefore, by means of

16

Doob's martingale inequality, $T^{-1/2-2\delta_I} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 u_t$ is shown to be uniformly bounded in probability, and the first summand on the r.h.s. is $o_p(1)$. For the second summand, write for any $\tau \in [0,1]$

$$\left| \frac{1}{T^{3/2+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 b\left(\frac{t}{T}\right) x_{t-1} \right| \leq \max_{\tau \in [0,1]} |b(\tau)| \max_{t=1,\dots,T} \frac{|x_{t-1}|}{\sqrt{T}} \frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{T} z_{I,t-1}^2$$

where $\frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{T} z_{I,t-1}^2 \Rightarrow K_{z_I^2}(1)$ due to Assumption 4 and the fact that $\max_{t=1,\dots,T} |x_{t-1}| = o_p\left(\sqrt{T}\right)$ with x_{t-1} being uniformly L_4-bounded. Then,

$$\frac{1}{T^{1+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 y_t = \frac{1}{\sqrt{T}} \frac{1}{T^{1/2+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 u_t + \frac{1}{T} \frac{1}{T^{1/2+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 b\left(\frac{t}{T}\right) x_{t-1},$$

where we have $\frac{1}{T^{1/2+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 u_t \Rightarrow \int_0^\tau Z^2(s) \mathrm{d}U(s)$ and $\frac{1}{T^{1/2+2\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}^2 b\left(\frac{t}{T}\right) x_{t-1} \Rightarrow \frac{\omega}{1-\rho} \int_0^\tau Z^2(s) b(s) \mathrm{d}V(s)$. Finally,

$$\begin{aligned}
\frac{1}{T^{1+\delta_I+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} z_{II,t-1} y_t &= \frac{1}{\sqrt{T}} \frac{1}{T^{1/2+\delta_I+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} z_{II,t-1} u_t \\
&\quad + \frac{1}{T} \frac{1}{T^{1/2+\delta_I+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} z_{II,t-1} b\left(\frac{t}{T}\right) x_{t-1}
\end{aligned}$$

where the weak convergence of $\frac{1}{T^{1/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t$ and the continuity requirements from Assumptions 4 and 5 lead to

$$\frac{1}{T^{1/2+\delta_I+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} \left(z_{I,t-1} u_t \right) \Rightarrow \int_0^\tau Z(s) \mathrm{d}G_I(s)$$

and, correspondingly, to

$$\frac{1}{T^{1/2+\delta_I+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} b\left(\frac{t}{T}\right) \left(z_{I,t-1} x_{t-1} \right) \Rightarrow \int_0^\tau Z(s) b(s) \mathrm{d}K_{z_I x}(s)$$

as required.

8. We have elementwise that

$$\frac{1}{T^{1/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} y_t = \frac{1}{T^{1/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t + \frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} b\left(\frac{t}{T}\right) x_{t-1}$$

where the behaviour of the first summand on the r.h.s. follows directly from Assumption 4 and, for the second,

$$\frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} b\left(\frac{t}{T}\right) \left(z_{I,t-1} x_{t-1} \right) \Rightarrow \int_0^\tau b(s) \mathrm{d}K_{z_I x}(s)$$

as in the proof of item 7. Then,

$$\frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} y_t = \frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} u_t + \frac{1}{\sqrt{T}} \left(\frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} b\left(\frac{t}{T}\right) x_{t-1} \right)$$

17

where $\frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} u_t \Rightarrow \int_0^\tau Z(s) dU(s)$ and

$$\frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} b\left(\frac{t}{T}\right) x_{t-1} \Rightarrow \frac{\omega}{1-\rho} \int_0^\tau Z(s) b(s) dV(s).$$

9. We have elementwise

$$\frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} y_t^2 = \frac{1}{T^{1+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t^2 + \frac{2}{T^{3/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} u_t b\left(\frac{t}{T}\right) x_{t-1}$$
$$+ \frac{1}{T^{2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1} b^2\left(\frac{t}{T}\right) x_{t-1}^2.$$

The first summand is uniformly bounded in probability, see Assumption 4, while the second and third summands vanish analogously to the proof of item 6. Then,

$$\frac{1}{T^{1+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} y_t^2 = \frac{1}{T^{1+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} u_t^2 + \frac{2}{T^{3/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} u_t b\left(\frac{t}{T}\right) x_{t-1}$$
$$+ \frac{1}{T^{2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} b^2\left(\frac{t}{T}\right) x_{t-1}^2.$$

The first summand delivers the desired limit like in the proof of item 6, while the second and the third can be shown to vanish uniformly in τ as follows. First, since $u_t x_{t-1}$ is a uniformly L_2-bounded MD sequence, so is $T^{-\delta_{II}} z_{II,t-1} b\left(\frac{t}{T}\right) x_{t-1} u_t$, such that $\frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} u_t b\left(\frac{t}{T}\right) x_{t-1}$ is uniformly $O_p(1)$ by Doob's martingale inequality and we therefore have that

$$\max_{\tau \in [0,1]} \left| \frac{2}{T^{3/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} u_t b\left(\frac{t}{T}\right) x_{t-1} \right| = O_p\left(T^{-1}\right)$$

and, second,

$$\frac{1}{T^{1+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} b^2\left(\frac{t}{T}\right) x_{t-1}^2 \leq \max_{\tau \in [0,1]} b^2(\tau) \max_{\tau \in [0,1]} \frac{z_{II,\lfloor \tau T \rfloor - 1}}{T^{\delta_{II}}} \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} x_{t-1}^2 \leq C \frac{1}{T} \sum_{t=1}^{T} x_{t-1}^2 = O_p(1)$$

uniformly in τ, as required.

10. Write

$$\frac{1}{T} S_{y^2}(\tau) = \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} u_t^2 + \frac{2}{T} \left(\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor \tau T \rfloor} b\left(\frac{t}{T}\right) x_{t-1} u_t \right) + \frac{1}{T} \left(\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} b^2\left(\frac{t}{T}\right) x_{t-1}^2 \right).$$

Using standard methods, the first summand on the r.h.s. delivers the desired limit,

$$\sup_{\tau \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} u_t^2 - [U](\tau) \right| \xrightarrow{p} 0,$$

while, for the second summand, $\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor \tau T \rfloor} b\left(\frac{t}{T}\right) x_{t-1} u_t$ is uniformly $O_p(1)$ by Doob's mar-

tingale inequality, since $b\left(\frac{t}{T}\right) x_{t-1}u_t$ is a uniformly L_2-bounded MD sequence, and thus

$$\max_{\tau \in [0,1]} \left| \frac{2}{T} \left(\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor \tau T \rfloor} b\left(\frac{t}{T}\right) x_{t-1}u_t \right) \right| = O_p\left(T^{-1}\right).$$

The third summand is easily analyzed, with

$$\frac{1}{T} \left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} b^2\left(\frac{t}{T}\right) x_{t-1}^2 \right| \leq \max_{\tau \in [0,1]} b^2(\tau) \frac{1}{T^2} \sum_{t=1}^{\lfloor \tau T \rfloor} x_{t-1}^2 \leq C \frac{1}{T^2} \sum_{t=1}^{T} x_{t-1}^2 = O_p\left(\frac{1}{T}\right)$$

uniformly in τ as required. ∎

Proof of Lemma S.3

1. Follows directly from $\frac{1}{\sqrt{T}} x_{\lfloor \tau T \rfloor} \Rightarrow \omega J_{c,\mathrm{H}}(\tau)$ using the CMT.

2. Write $\frac{1}{\sqrt{T}} S_y(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} u_t + \frac{1}{T\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} b\left(\frac{t}{T}\right) x_{t-1}$ where, recall, $\frac{1}{\sqrt{T}} x_{\lfloor \tau T \rfloor} \Rightarrow \omega J_{c,\mathrm{H}}(\tau)$ such that $\frac{1}{T\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} b\left(\frac{t}{T}\right) x_{t-1} \Rightarrow \omega \int_0^\tau b(s) J_{c,\mathrm{H}}(s)\,\mathrm{d}s$ as required.

3. Follows directly from Assumptions 4 and 5 and the CMT.

4. Follows directly from Assumptions 4 and 5 and the CMT.

5. Follows directly from Assumptions 4-6 and the CMT.

6. Note that, under Assumption 1.2, $\max_{t=1,\dots,T} |x_{t-1}| = O_p\left(\sqrt{T}\right)$ while $\beta_{1,t} = O\left(\frac{1}{T}\right)$, so it is not difficult to show that the local alternative does not influence the limit and, hence, we omit the details here. The result then follows along the lines of the proof of Lemma S.2 item 6 and we omit the details.

7. The proof of this result is straightforward but tedious and, hence, is omitted in the interests of brevity.

8. We have elementwise that

$$\frac{1}{T^{1/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}y_t = \frac{1}{T^{1/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}u_t + \frac{1}{T^{3/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}b\left(\frac{t}{T}\right) x_{t-1}$$

where the behaviour of the first summand on the r.h.s. follows directly from Assumption 4 and, for the second, the Cauchy-Schwarz inequality implies that

$$\left| \frac{1}{T^{3/2+\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}b\left(\frac{t}{T}\right) x_{t-1} \right| \leq \sqrt{\frac{1}{T^{1+2\delta_I}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{I,t-1}^2 \frac{1}{T^2} \sum_{t=1}^{\lfloor \tau T \rfloor} b^2\left(\frac{t}{T}\right) x_{t-1}^2}$$

where the r.h.s. is easily seen to be uniformly bounded in probability as required. Then,

$$\frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}y_t = \frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}u_t + \frac{1}{T^{3/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}b\left(\frac{t}{T}\right) x_{t-1}$$

where $\frac{1}{T^{1/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}u_t \Rightarrow \int_0^\tau Z(s)\mathrm{d}U(s)$ and

$$\frac{1}{T^{3/2+\delta_{II}}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1}b\left(\frac{t}{T}\right) x_{t-1} \Rightarrow \omega \int_0^\tau Z(s)b(s)J_{c,\mathrm{H}}(s)\mathrm{d}s.$$

19

9. The proof of this result is straightforward and so again is omitted in the interests of brevity.

10. Analogous to the proof of Lemma S.2 item 10 and, hence, omitted here.

<div style="text-align: right;">■</div>

Proof of Lemma S.4

Consider a special probability space where the original data samples indexed by T are redefined as a triangular array (with a slight abuse, we maintain the original notation) such that the convergence results in Lemma S.2 hold a.s.; this is standard in proofs of bootstrap convergence and is possible due to Corollary 5.12 of Kallenberg (1997). Let the special probability space be extended by a product construction in order to support an IID standard normal sequence $\{R_t\}$ independent of the redefined data. We show that on this probability space the convergence relations asserted in Lemma S.4 hold weakly a.s. ($\theta_T^* \overset{w}{\Rightarrow}_{a.s.} \theta$ defined as $\mathrm{E}^* f(\theta_T^*) \to \mathrm{E} f(\theta)$ P-a.s. for all bounded continuous real functions f with matching domain, where θ_T^*, θ are random elements defined on the special probability space). Then on general probability spaces these relations hold weakly in probability (since the limit measures concerned, or equivalently, the $\mathrm{E} f(\theta)$ concerned, are non-random). For the purposes of unified exposition, '$O_{p^*}(1)$ in P-probability' statements are also established on the special probability space, and transfer to general probability spaces automatically, depending exclusively on the distributions involved.

1 & 2. Let $\mathbf{M}_T := \mathrm{diag}(T^{-1/2}, T^{-1/2-\delta_I}, T^{-1/2-\delta_{II}})$ and $\boldsymbol{S}^*_{\cdot y}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor}(1, \boldsymbol{z}'_{t-1})' y_t R_t$. It holds that $(S^*_y(\tau), \boldsymbol{S}^*_{\boldsymbol{z}y}(\tau)')' - \boldsymbol{S}^*_{\cdot y}(\tau) = T^{-1} S_y(1) \sum_{t=1}^{\lfloor \tau T \rfloor}(1, \boldsymbol{z}'_{t-1})' R_t$, where (i) $T^{-1} S_y(1) \to 0$ P-a.s. by Lemma S.2(2) recast on the special probability space, and (ii), conditionally on the data, $\sum_{t=1}^{\lfloor \tau T \rfloor}(1, \boldsymbol{z}'_{t-1})' R_t$ is a zero-mean process with independent increments and

$$\mathbf{M}_T \operatorname{Var}^* \left\{ \sum_{t=2}^{T}(1, \boldsymbol{z}'_{t-1})' R_t \right\} \mathbf{M}_T = \mathbf{M}_T \begin{pmatrix} T & \boldsymbol{S}_{\boldsymbol{z}}(1)' \\ \boldsymbol{S}_{\boldsymbol{z}}(1) & \mathbf{S}_{\boldsymbol{zz}}(1) \end{pmatrix} \mathbf{M}_T$$

is P-a.s. convergent to a P-a.s. finite limit, by Lemma S.2(3,5) recast on the special probability space. Therefore, by Kolmogorov's inequality applied conditionally,

$$\sup_{\tau \in [0,1]} \left\| \mathbf{M}_T(S^*_y(\tau), \boldsymbol{S}^*_{\boldsymbol{z}y}(\tau)')' - \mathbf{M}_T \boldsymbol{S}^*_{\cdot y}(\tau) \right\| \overset{w}{\Rightarrow}_{a.s.} 0,$$

or equivalently, $\mathbf{M}_T(S^*_y(\tau), \boldsymbol{S}^*_{\boldsymbol{z}y}(\tau)')' = \mathbf{M}_T \boldsymbol{S}^*_{\cdot y}(\tau) + o_{p^*}(1)$ P-a.s. uniformly over τ. On the other hand, $\mathbf{M}_T \boldsymbol{S}^*_{\cdot y}(\tau)$ conditionally on the data is a zero-mean Gaussian process with independent increments and $\mathbf{M}_T \boldsymbol{S}^*_{\cdot y}(0) = 0$. For such processes weak convergence a.s. follows from the a.s. convergence of the conditional variance function uniformly over τ; cf. the proof of Lemma A.5 in Cavaliere *et al.* (2010). The process $\mathbf{M}_T \boldsymbol{S}^*_{\cdot y}(\tau)$ conditionally on the data has variance function

$$\tau \mapsto \mathbf{M}_T \begin{pmatrix} S_{y^2}(\tau) & \boldsymbol{S}'_{\boldsymbol{z}y^2}(\tau) \\ \boldsymbol{S}_{\boldsymbol{z}y^2}(\tau) & \mathbf{S}_{\boldsymbol{zz}y^2}(\tau) \end{pmatrix} \mathbf{M}_T = \begin{pmatrix} T^{-1} S_{y^2}(\tau) & T^{-1/2} \boldsymbol{S}'_{\boldsymbol{z}y^2}(\tau) \mathbf{W}_T \\ T^{-1/2} \mathbf{W}_T \boldsymbol{S}_{\boldsymbol{z}y^2}(\tau) & \mathbf{W}_T \mathbf{S}_{\boldsymbol{zz}y^2}(\tau) \mathbf{W}_T \end{pmatrix}. \quad \text{(S.9)}$$

By Lemma S.2(6,9,10) recast on the special probability space and applied to the respective blocks of the variance function in (S.9), it follows that $T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor}(1, T^{-\delta_I} z_{I,t-1})' y_t R_t$ and $T^{-1/2-\delta_{II}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} y_t R_t$, which are components of $\mathbf{M}_T \boldsymbol{S}^*_{\cdot y}(\tau)$, each converge $\overset{w}{\Rightarrow}_{a.s.}$ to Gaussian processes starting at 0, with independent increments and variance functions given by $\mathrm{diag}([U](\tau), [G_I](\tau))$ and $\int_0^\tau Z^2(s) \mathrm{d}[U](s)$, respectively. Note that there is no claim about the two $\overset{w}{\Rightarrow}_{a.s.}$ convergence facts being joint. Indeed, we only need to establish the precise limit of $T^{-1/2}(S^*_y(\tau), T^{-\delta_I} S^*_{z_I y}(\tau))' = T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor}(1, T^{-\delta_I} z_{I,t-1})' y_t R_t + o_{p^*}(1)$ P-a.s., whereas for

<div style="text-align: center;">20</div>

$T^{-1/2-\delta_{II}} \sum_{t=1}^{\lfloor \tau T \rfloor} z_{II,t-1} y_t R_t$ it suffices to note that it is $O_{p^*}(1)$ P-a.s. as a result of its $\overset{w}{\Rightarrow}_{a.s.}$ convergence.

3. We decompose $\Delta(\tau) := \mathbf{S}^*_{zzy^2}(\tau) - \mathbf{S}_{zzy^2}(\tau)$ into

$$\Delta(\tau) = \Delta^{(1)}(\tau) - 2\left\{T^{-1} S_y(1)\right\} \Delta^{(2)}(\tau) + \left\{T^{-1} S_y(1)\right\}^2 \Delta^{(3)}(\tau)$$

with $\Delta^{(1)}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} \boldsymbol{z}'_{t-1} y_t^2 (R_t^2 - 1)$, $\Delta^{(2)}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} \boldsymbol{z}'_{t-1} y_t R_t^2$ and $\Delta^{(3)}(\tau) := \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} \boldsymbol{z}'_{t-1} R_t^2$. The term $\Delta^{(1)}(\tau)$ is, conditionally on the data, a zero-mean process with independent increments and

$$\text{Var}^*\{\text{vec}(\Delta^{(1)}(1))\} = 2 \sum_{t=1}^{T} \text{vec}(\boldsymbol{z}_{t-1} \boldsymbol{z}'_{t-1})\{\text{vec}(\boldsymbol{z}_{t-1} \boldsymbol{z}'_{t-1})\}' y_t^4,$$

where the factor 2 arises because $\text{Var}(R_t^2 - 1) = 2$, such that

$$
\begin{aligned}
\| \text{Var}^*\{\text{vec}(\mathbf{W}_T \Delta^{(1)}(1)\mathbf{W}_T)\}\| &\leq 2\sum_{t=1}^{T} \|\mathbf{W}_T \boldsymbol{z}_{t-1}\|^4 y_t^4 \\
&\leq 4T^{-2} \sum_{t=1}^{T} (|T^{-\delta_I} z_{I,t-1}|^4 + |T^{-\delta_{II}} z_{II,t-1}|^4) y_t^4 \\
&\leq 4 \sum_{r \in \{I,II\}} T^{-1-2\delta_r} S_{z_r^2 y^2}(1) \max_{\tau \in (0,1]} |T^{-1-2\delta_r} \Delta S_{z_r^2 y^2}(\tau)| \to 0
\end{aligned}
$$

P-a.s. because each of $T^{-1-2\delta_r} S_{z_r^2 y^2}(\tau)$, $r \in \{I, II\}$, converges P-a.s. to a pathwise continuous limit process, by Lemma S.2(6) recast on the special probability space. A conditional application of Kolmogorov's inequality shows that $\sup_{\tau \in [0,1]} \|\mathbf{W}_T \Delta^{(1)}(\tau)\mathbf{W}_T\| \overset{w}{\Longrightarrow}_{a.s.} 0$. The term $\Delta^{(2)}(\tau)$ equals $\mathbf{S}_{zzy}(\tau) + \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} \boldsymbol{z}'_{t-1} y_t (R_t^2 - 1)$, with $\sup_{\tau \in [0,1]} \|\mathbf{W}_T \mathbf{S}_{zzy}(\tau)\mathbf{W}_T\| \to 0$ P-a.s. by Lemma S.2(7) recast on the special probability space and

$$\sup_{\tau \in [0,1]} \left\| \mathbf{W}_T \sum_{t=1}^{\lfloor \tau T \rfloor} \boldsymbol{z}_{t-1} \boldsymbol{z}'_{t-1} y_t (R_t^2 - 1) \mathbf{W}_T \right\| \overset{w}{\Longrightarrow}_{a.s.} 0$$

similarly to $\sup_{\tau \in [0,1]} \|\mathbf{W}_T \Delta^{(1)}(\tau)\mathbf{W}_T\|$, so also $\sup_{\tau \in [0,1]} \|\mathbf{W}_T \Delta^{(2)}(\tau)\mathbf{W}_T\| \overset{w}{\Longrightarrow}_{a.s.} 0$. Finally, it holds that $\sup_{\tau \in [0,1]} \|T^{-1} \mathbf{W}_T \Delta^{(3)}(\tau)\mathbf{W}_T\| \leq \max_{t=1,\ldots,T}(T^{-1} R_t^2) \sup_{\tau \in [0,1]} \|\mathbf{W}_T \mathbf{S}_{zz}(\tau)\mathbf{W}_T\| \overset{w}{\Longrightarrow}_p 0$ by the distributional assumption on R_t and Lemma S.2(5) recast on the special probability space. As $T^{-1/2} S_y(1)$ converges P-a.s. to a P-a.s. finite r.v. on the special probability space, item 3 follows.

4, 5 & 6 These results follow similarly and, hence, we omit the details here.

∎

Proof of Lemma S.5

Similarly to the proof of Lemma S.4, consider a special probability space where the convergence results established in Lemma S.3 hold P-a.s. and on whose extension an IID standard Gaussian sequence $\{R_t\}$ independent of the redefined data is available. We show that on this extended special probability space the convergence relations asserted in Lemma S.5 hold P*-weakly P-a.s.

1 & 2. By the same argument as in the proof of Lemma S.4, it holds that

$$\sup_{\tau\in[0,1]} \|\mathbf{M}_T(S_y^*(\tau),\boldsymbol{S}_{zy}^*(\tau)')' - \mathbf{M}_T\boldsymbol{S}_{\cdot y}^*(\tau)\| \overset{w}{\Rightarrow}_{a.s.} 0.$$

The component $T^{-1/2}\sum_{t=1}^{\lfloor\tau T\rfloor}(1,T^{-\delta_{II}}z_{II,t-1})'y_tR_t$ of $\mathbf{M}_T\boldsymbol{S}_{\cdot y}^*(\tau)$ is, conditionally on the data, a zero-mean Gaussian process starting at 0, with independent increments and variance function

$$\begin{pmatrix} T^{-1}S_{y^2} & T^{-1-\delta_{II}}S_{z_{II}y^2} \\ T^{-1-\delta_{II}}S_{z_{II}y^2} & T^{-1-2\delta_{II}}S_{z_{II}^2y^2} \end{pmatrix} \to \int_0^\tau \begin{pmatrix} 1 & Z(s) \\ Z(s) & Z^2(s) \end{pmatrix} \mathrm{d}[U](s) =: C(\tau) \quad \text{P-a.s.,}$$

the convergence by Lemma S.3(6,9,10) recast on the special probability space. This proves that the process $T^{-1/2}\sum_{t=1}^{\lfloor\tau T\rfloor}(1,T^{-\delta_{II}}z_{II,t-1})'y_tR_t$ converges $\overset{w}{\Longrightarrow}_{a.s.}$ to a zero-mean Gaussian process starting at 0, with independent increments and variance function $C(\tau)$, and as required, the same convergence holds for the process $T^{-1/2}(S_y^*(\tau),T^{-\delta_{II}}S_{z_{II}y}^*(\tau))' = T^{-1/2}\sum_{t=1}^{\lfloor\tau T\rfloor}(1,T^{-\delta_{II}}z_{II,t-1})'y_tR_t + o_{p^*}(1)$ P-a.s. The component $\sum_{t=1}^{\lfloor\tau T\rfloor}T^{-1/2-\delta_I}z_{I,t-1}y_tR_t$ of $\mathbf{M}_T\boldsymbol{S}_{\cdot y}^*(\tau)$ is, conditionally on the data, a process with zero-mean independent increments and

$$\mathrm{Var}^*\left(\sum_{t=1}^{T}T^{-1/2-\delta_I}z_{I,t-1}y_tR_t\right) = T^{-1-2\delta_I}S_{z_I^2y^2}(1) = O_p(1)$$

by Lemma S.3(6), such that by Kolmogorov's inequality applied conditionally on the data it follows that $\sup_{\tau\in[0,1]}\left|\sum_{t=1}^{\lfloor\tau T\rfloor}T^{-1/2-\delta_I}z_{I,t-1}y_tR_t\right| = O_{p^*}(1)$ in P-probability.

3, 4, 5 & 6. Identical to Lemma S.4.

∎

Before progressing to the proof of Lemma S.6 we require some additional definitions and notation.

Let $T_\Delta := [0,1]^2 \cap \{(\tau_1,\tau_2)\in\mathbb{R}^2 : \tau_2-\tau_1\geq\Delta_\tau\}$ for some $\Delta_\tau\in(0,1)$. Let $\mathcal{D}(T_\Delta)$ be the set of real functions on T_Δ which are continuous from the 'right' (i.e., $f(\tau_1^{(n)},\tau_2^{(n)})\to f(\tau_1,\tau_2)$ when $\tau_i^{(n)}\downarrow\tau_i$, $i=1,2$, for $(\tau_1^{(n)},\tau_2^{(n)})$, $(\tau_1,\tau_2)\in T_\Delta$ and $f\in\mathcal{D}(T_\Delta)$) and have limits from within each of the four right angles $[A_1\times A_2]\cap T_\Delta$, $A_i\in\{[0,\tau_i),[\tau_i,1]\}$, $i=1,2$, when the angles are non-empty. For clarity, note that all bivariate cdf's with domain restricted to T_Δ belong to $\mathcal{D}(T_\Delta)$. It is well-known (e.g., Bickel and Wichura, 1971, p. 1662) that $\mathcal{D}(T_\Delta)$ can be equipped with a Skorokhod-like metric which makes it a separable and complete metric space such that stochastic process with values in $\mathcal{D}(T_\Delta)$ are measurable w.r.t. the resulting Borel σ-algebra. Moreover, the resulting topology relativised to $\mathcal{C}(T_\Delta)\subset\mathcal{D}(T_\Delta)$, the subspace of continuous real functions on T_Δ, coincides with the uniform topology. As we will only be interested in convergence to limits in $\mathcal{C}(T_\Delta)$, in what follows convergence and continuity issues involving elements of $\mathcal{D}(T_\Delta)$ are always discussed w.r.t. the uniform metric on $\mathcal{D}(T_\Delta)$. It is then straightforward to see that the function from $\mathcal{D}^2$ to $\mathcal{D}(T_\Delta)$ which associates to every $(f_1,f_2)\in\mathcal{D}^2$ the element $(\tau_1,\tau_2)\mapsto f_2(\tau_2)-f_1(\tau_1)$ of $\mathcal{D}(T_\Delta)$ is continuous on the subspace of continuous functions $\mathcal{C}^2$ of $\mathcal{D}^2$. Moreover, linearly combining functions in $\mathcal{D}(T_\Delta)$, multiplication of functions in $\mathcal{D}(T_\Delta)$ and division of functions in $\mathcal{D}(T_\Delta)$ (for denominators bounded away from zero) are continuous transformations of the product subspace $\mathcal{C}(T_\Delta)\times\mathcal{C}(T_\Delta)$ of $\mathcal{D}(T_\Delta)\times\mathcal{D}(T_\Delta)$. Finally, the functional from $\mathcal{D}(T_\Delta)$ to $\mathbb{R}$ defined by $f\mapsto\sup_{T_\Delta}|f|$ is continuous on $\mathcal{C}(T_\Delta)$; it arises in the discussion of the statistic $\mathcal{T}^d$. Although the statistics $\mathcal{T}^s$, $s\in\{f,b,r\}$, can be most naturally discussed by considering maxima of single-parameter processes, to unify the exposition we note that the functionals $\sup_{A^s}|f|$, $s\in\{f,b,r\}$, are also continuous on $\mathcal{C}(T_\Delta)$, where $A^f := \{0\}\times[\tau_L,1]$, $A^b := [0,\tau_U]\times\{1\}$ and $A^r := \{(\tau,\tau+\Delta_\tau):\tau\in[0,1-\Delta_\tau]\}$ with $\tau_L\geq\Delta_\tau$ and $1-\tau_U\geq\Delta_\tau$.

Proof of Lemma S.6

Proof of (i). Using $\mathbf{W}_T = \mathrm{diag}(T^{-1/2-\delta_I}, T^{-1/2-\delta_{II}})$, the expression for $t_{\beta_1}^2(\tau_1, \tau_2)$ in (4.3) can be equivalently written as

$$
\frac{\left(\frac{1}{\sqrt{T}}\left(\mathbf{W}_T \boldsymbol{A}_T(\tau_1, \tau_2)\right)' \left(\mathbf{W}_T \mathbf{B}_T(\tau_1, \tau_2)\mathbf{W}_T\right)^{-1} \mathbf{W}_T \boldsymbol{C}_T(\tau_1, \tau_2)\right)^2}{\left(\frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{A}_T(\tau_1, \tau_2)\right)' \left(\mathbf{W}_T \mathbf{B}_T(\tau_1, \tau_2)\mathbf{W}_T\right)^{-1} \mathbf{W}_T \mathbf{D}_T(\tau_1, \tau_2)\mathbf{W}_T \left(\mathbf{W}_T \mathbf{B}_T(\tau_1, \tau_2)\mathbf{W}_T\right)^{-1} \frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{A}_T(\tau_1, \tau_2)}.
$$

Consider first

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{A}_T(\tau_1, \tau_2) &= \frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{S}_{xz}(\tau_2) - \frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{S}_{xz}(\tau_1) \\
&\quad - \frac{\sqrt{T}}{\lfloor\tau_2 T\rfloor - \lfloor\tau_1 T\rfloor}\left(\frac{1}{\sqrt{T}}S_x(\tau_2) - \frac{1}{\sqrt{T}}S_x(\tau_1)\right)\left(\frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{S}_z(\tau_2) - \frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{S}_z(\tau_1)\right).
\end{aligned}
$$

Recalling the discussion of the space $\mathcal{D}(T_\Delta)$, by Lemma S.2 and the CMT we obtain that

$$
\frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{A}_T(\tau_1, \tau_2) \Rightarrow \begin{pmatrix} K_{z_I x}(\tau_2) - K_{z_I x}(\tau_1) \\ 0 \end{pmatrix}
$$

on $\mathcal{D}(T_\Delta)^2$. Next,

$$
\begin{aligned}
\mathbf{W}_T \mathbf{B}_T(\tau_1, \tau_2)\mathbf{W}_T &= \mathbf{W}_T \mathbf{S}_{zz}(\tau_2)\mathbf{W}_T - \mathbf{W}_T \mathbf{S}_{zz}(\tau_1)\mathbf{W}_T \\
&\quad - \frac{T}{\lfloor\tau_2 T\rfloor - \lfloor\tau_1 T\rfloor}\frac{1}{T}\left(\mathbf{W}_T \boldsymbol{S}_z(\tau_2) - \mathbf{W}_T \boldsymbol{S}_z(\tau_1)\right)\left(\mathbf{W}_T \boldsymbol{S}_z(\tau_2) - \mathbf{W}_T \boldsymbol{S}_z(\tau_1)\right)'
\end{aligned}
$$

such that, using Lemma S.2 and the CMT again, we have

$$
\left(\mathbf{W}_T \mathbf{B}_T(\tau_1, \tau_2)\mathbf{W}_T\right)^{-1} \Rightarrow \begin{pmatrix} K_{z_I^2}(\tau_2) - K_{z_I^2}(\tau_1) & 0 \\ 0 & \int_{\tau_1}^{\tau_2} Z^2(s)\mathrm{d}s - \frac{1}{\tau_2-\tau_1}\left(\int_{\tau_1}^{\tau_2} Z(s)\mathrm{d}s\right)^2 \end{pmatrix}^{-1}
$$

on $\mathcal{D}(T_\Delta)^4$, since $K_{z_I^2}(\tau_2) - K_{z_I^2}(\tau_1) > 0 \; \forall \tau_2 > \tau_1$. Notice furthermore, using continuity of the limiting processes involved, that

$$
\frac{1}{\sqrt{T}}\left(\mathbf{W}_T \boldsymbol{A}_T(\tau_1, \tau_2)\right)' \left(\mathbf{W}_T \mathbf{B}_T(\tau_1, \tau_2)\mathbf{W}_T\right)^{-1} \Rightarrow \left(\frac{K_{z_I x}(\tau_2) - K_{z_I x}(\tau_1)}{K_{z_I^2}(\tau_2) - K_{z_I^2}(\tau_1)}; 0\right) \tag{S.10}
$$

such that, in the limit, only the first element of $\mathbf{W}_T \boldsymbol{C}_T(\tau_1, \tau_2)$ and the first diagonal element of $\mathbf{W}_T \mathbf{D}_T(\tau_1, \tau_2)\mathbf{W}_T$ play a role (the other elements are all bounded in probability uniformly on T_Δ and so vanish upon multiplication by the zero limit on the right hand side of S.10). For the first of these asymptotically relevant terms, write

$$
\begin{aligned}
\mathbf{W}_T \boldsymbol{C}_T(\tau_1, \tau_2) &= \mathbf{W}_T \boldsymbol{S}_{zy}(\tau_2) - \mathbf{W}_T \boldsymbol{S}_{zy}(\tau_1) \\
&\quad - \frac{T}{\lfloor\tau_2 T\rfloor - \lfloor\tau_1 T\rfloor}\left(\frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{S}_z(\tau_2) - \frac{1}{\sqrt{T}}\mathbf{W}_T \boldsymbol{S}_z(\tau_1)\right)\left(\frac{1}{\sqrt{T}}S_y(\tau_2) - \frac{1}{\sqrt{T}}S_y(\tau_1)\right) \\
&= \begin{pmatrix} \frac{1}{T^{1/2+\delta_I}}\left(S_{z_I y}(\tau_2) - S_{z_I y}(\tau_2)\right) \\ \frac{1}{T^{1/2+\delta_{II}}}\left(S_{z_{II} y}(\tau_2) - S_{z_{II} y}(\tau_2)\right) \end{pmatrix} \\
&\quad - \begin{pmatrix} 0 \\ \frac{1}{\tau_2-\tau_1}\frac{1}{T^{1+\delta_{II}}}\left(S_{z_{II}}(\tau_2) - S_{z_{II}}(\tau_2)\right)\frac{1}{\sqrt{T}}\left(S_y(\tau_2) - S_y(\tau_1)\right) \end{pmatrix} + o_p(1)
\end{aligned}
$$

23

with the $o_p(1)$ remainder term applying uniformly on T_Δ. For the second, a tedious but straightforward application of Lemma S.2 yields that

$$
\begin{aligned}
\mathbf{D}_T\left(\tau_1, \tau_2\right) =\ & \mathbf{R}_T + \mathbf{S}_{\boldsymbol{zzy}^2}\left(\tau_2\right) - \mathbf{S}_{\boldsymbol{zzy}^2}\left(\tau_1\right) \\
& - \frac{1}{\lfloor\tau_2 T\rfloor - \lfloor\tau_1 T\rfloor}\left(\boldsymbol{S}_{\boldsymbol{zy}^2}\left(\tau_2\right) - \boldsymbol{S}_{\boldsymbol{zy}^2}\left(\tau_1\right)\right)\left(\boldsymbol{S}_{\boldsymbol{z}}\left(\tau_2\right) - \boldsymbol{S}_{\boldsymbol{z}}\left(\tau_1\right)\right)' \\
& - \frac{1}{\lfloor\tau_2 T\rfloor - \lfloor\tau_1 T\rfloor}\left(\boldsymbol{S}_{\boldsymbol{z}}\left(\tau_2\right) - \boldsymbol{S}_{\boldsymbol{z}}\left(\tau_1\right)\right)\left(\boldsymbol{S}_{\boldsymbol{zy}^2}\left(\tau_2\right) - \boldsymbol{S}_{\boldsymbol{zy}^2}\left(\tau_1\right)\right)' \\
& + \frac{1}{\left(\lfloor\tau_2 T\rfloor - \lfloor\tau_1 T\rfloor\right)^2}\left(\boldsymbol{S}_{\boldsymbol{z}}\left(\tau_2\right) - \boldsymbol{S}_{\boldsymbol{z}}\left(\tau_1\right)\right)\left(\boldsymbol{S}_{\boldsymbol{z}}\left(\tau_2\right) - \boldsymbol{S}_{\boldsymbol{z}}\left(\tau_1\right)\right)'\left(S_{y^2}\left(\tau_2\right) - S_{y^2}\left(\tau_1\right)\right)
\end{aligned}
$$

where $\mathbf{R}_T$ is uniformly dominated by the other terms on T_Δ due to Lemma S.2, such that

$$
\mathbf{W}_T \mathbf{D}_T\left(\tau_1, \tau_2\right) \mathbf{W}_T = \mathbf{W}_T \mathbf{S}_{\boldsymbol{zzy}^2}\left(\tau_2\right) \mathbf{W}_T - \mathbf{W}_T \mathbf{S}_{\boldsymbol{zzy}^2}\left(\tau_1\right) \mathbf{W}_T
$$

$$
-2 \frac{1}{\tau_2 - \tau_1}\left(\begin{array}{cc} 0 & R_{1T}(\tau_1, \tau_2) \\ R_{1T}(\tau_1, \tau_2) & \frac{1}{T^{1+\delta_{II}}}\left(S_{z_{II}}\left(\tau_2\right) - S_{z_{II}}\left(\tau_1\right)\right) \frac{1}{T^{1+\delta_{II}}}\left(S_{z_{II}y^2}\left(\tau_2\right) - S_{z_{II}y^2}\left(\tau_1\right)\right) \end{array}\right)
$$

$$
+ \frac{1}{\left(\tau_2 - \tau_1\right)^2}\left(\begin{array}{cc} 0 & 0 \\ 0 & \frac{1}{T^{2+2\delta_{II}}}\left(S_{z_{II}}\left(\tau_2\right) - S_{z_{II}}\left(\tau_1\right)\right)^2 \frac{1}{T}\left(S_{y^2}\left(\tau_2\right) - S_{y^2}\left(\tau_1\right)\right) \end{array}\right) + o_p(1)
$$

with both $R_{1T}(\tau_1, \tau_2) = R_{2T}(\tau_1, \tau_2) = O_p(1)$ and the $o_p(1)$ remainder term applying uniformly on T_Δ. After some further algebra we then obtain that

$$
t_{\beta_1}^2\left(\tau_1, \tau_2\right) = \frac{Q_I^2\left(\tau_1, \tau_2\right)}{P_I\left(\tau_1, \tau_2\right)} + o_p(1)
$$

with the $o_p(1)$ remainder term applying uniformly on T_Δ, as required.

Proof of (ii). Consider now the equivalent expression

$$
\frac{\left(\frac{1}{T}\left(\mathbf{W}_T \boldsymbol{A}_T\left(\tau_1, \tau_2\right)\right)'\left(\mathbf{W}_T \mathbf{B}_T\left(\tau_1, \tau_2\right) \mathbf{W}_T\right)^{-1} \mathbf{W}_T \boldsymbol{C}_T\left(\tau_1, \tau_2\right)\right)^2}{\left(\frac{1}{T} \mathbf{W}_T \boldsymbol{A}_T\left(\tau_1, \tau_2\right)\right)'\left(\mathbf{W}_T \mathbf{B}_T\left(\tau_1, \tau_2\right) \mathbf{W}_T\right)^{-1} \mathbf{W}_T \mathbf{D}_T\left(\tau_1, \tau_2\right) \mathbf{W}_T\left(\mathbf{W}_T \mathbf{B}_T\left(\tau_1, \tau_2\right) \mathbf{W}_T\right)^{-1} \frac{1}{T} \mathbf{W}_T \boldsymbol{A}_T\left(\tau_1, \tau_2\right)}.
$$

Let us first examine $\boldsymbol{A}_T\left(\tau_1, \tau_2\right)$, for which we have using Lemma S.3 and the CMT,

$$
\begin{aligned}
\frac{1}{T} \mathbf{W}_T \boldsymbol{A}_T\left(\tau_1, \tau_2\right) =\ & \frac{1}{T} \mathbf{W}_T \boldsymbol{S}_{xz}\left(\tau_2\right) - \frac{1}{T} \mathbf{W}_T \boldsymbol{S}_{xz}\left(\tau_1\right) \\
& - \frac{T}{\lfloor\tau_2 T\rfloor - \lfloor\tau_1 T\rfloor}\left(\frac{1}{T\sqrt{T}} S_x\left(\tau_2\right) - \frac{1}{T\sqrt{T}} S_x\left(\tau_1\right)\right)\left(\frac{1}{\sqrt{T}} \mathbf{W}_T \boldsymbol{S}_{\boldsymbol{z}}\left(\tau_2\right) - \frac{1}{\sqrt{T}} \mathbf{W}_T \boldsymbol{S}_{\boldsymbol{z}}\left(\tau_1\right)\right) \\
\Rightarrow\ & \left(\begin{array}{c} 0 \\ \omega \int_{\tau_1}^{\tau_2} Z(s) J_{c,\mathrm{H}}(s)\, \mathrm{d}s - \frac{\omega}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} J_{c,\mathrm{H}}(s)\, \mathrm{d}s \int_{\tau_1}^{\tau_2} Z(s)\, \mathrm{d}s \end{array}\right)
\end{aligned}
$$

on $\mathcal{D}(T_\Delta)$. Then, the weak limit of $\left(\mathbf{W}_T \mathbf{B}_T\left(\tau_1, \tau_2\right) \mathbf{W}_T\right)^{-1}$ has the same expression as given in item **(i)** (recall that the actual shape of $K_{z_I^2}$ may be different under the two cases spelled out in Assumption 1), such that

$$
\frac{1}{T}\left(\mathbf{W}_T \boldsymbol{A}_T\left(\tau_1, \tau_2\right)\right)'\left(\mathbf{W}_T \mathbf{B}_T\left(\tau_1, \tau_2\right) \mathbf{W}_T\right)^{-1} \Rightarrow \left(\begin{array}{cc} 0, & \frac{\omega \int_{\tau_1}^{\tau_2} Z(s) J_{c,\mathrm{H}}(s)\, \mathrm{d}s - \frac{\omega}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} J_{c,\mathrm{H}}(s)\, \mathrm{d}s \int_{\tau_1}^{\tau_2} Z(s)\, \mathrm{d}s}{\int_{\tau_1}^{\tau_2} Z^2(s)\, \mathrm{d}s - \frac{1}{\tau_2 - \tau_1}\left(\int_{\tau_1}^{\tau_2} Z(s)\, \mathrm{d}s\right)^2} \end{array}\right)
$$

on $\mathcal{D}(T_\Delta)^2$ and, in the limit, it is now only the second element of $\mathbf{W}_T \boldsymbol{C}_T\left(\tau_1, \tau_2\right)$ and the second diagonal element of $\mathbf{W}_T \mathbf{D}_T\left(\tau_1, \tau_2\right) \mathbf{W}_T$ that play a role compared to item **(i)** of this lemma. The result follows with a tedious, yet straightforward application of Lemma S.3 and the CMT. ∎

Proof of Lemma S.7

On the (extended) special probability spaces where Lemmas S.4 and S.5 were established to hold weakly a.s., the proof is analogous to that of Lemma S.6. First, $\mathbf{W}_T \boldsymbol{C}_T^* \left(\tau_1, \tau_2 \right)$ can be expanded similarly to $\mathbf{W}_T \boldsymbol{C}_T \left(\tau_1, \tau_2 \right)$ into

$$
\mathbf{W}_T \left(\begin{array}{c} S_{z_I y}^* \left(\tau_2 \right) - S_{z_I y}^* \left(\tau_2 \right) \\ S_{z_{II} y}^* \left(\tau_2 \right) - S_{z_{II} y}^* \left(\tau_2 \right) - \frac{1}{T(\tau_2 - \tau_1)} \left(S_{z_{II}} \left(\tau_2 \right) - S_{z_{II}} \left(\tau_2 \right) \right) \left(S_y^* \left(\tau_2 \right) - S_y^* \left(\tau_1 \right) \right) \end{array} \right) + o_{p^*} \left(1 \right)
$$

P-a.s. and uniformly on T_Δ, by using Lemmas S.2(3), S.3(3), S.4(1,2) and S.5(1,2) recast on the special probability space. Here we have also used the fact that $f_T \to f$ a.s. for measurable $\mathbb{R}^{m \times k}$-valued transformations f_T of the (redefined) data and $g_T^* \overset{w}{\Rightarrow}_{a.s.} g$ for $\mathbb{R}^{k \times n}$ random matrices defined on the special probability space imply that $f_T g_T^* \overset{w}{\Rightarrow}_{a.s.} fg$. Second, Lemmas S.4 and S.5 can be used to see that $\sup_{T_\Delta} \| \mathbf{W}_T \{ \mathbf{D}_T^* \left(\tau_1, \tau_2 \right) - \mathbf{D}_T \left(\tau_1, \tau_2 \right) \} \mathbf{W}_T \| = o_p^*(1)$ P-a.s. in both the stationary and the (near-)unit root case, such that the effect of replacing $\mathbf{D}_T^* \left(\tau_1, \tau_2 \right)$ by $\mathbf{D}_T \left(\tau_1, \tau_2 \right)$ in the expression

$$
\frac{\left(\boldsymbol{A}_T' \left(\tau_1, \tau_2 \right) \mathbf{B}_T^{-1} \left(\tau_1, \tau_2 \right) \boldsymbol{C}_T^* \left(\tau_1, \tau_2 \right) \right)^2}{\boldsymbol{A}_T' \left(\tau_1, \tau_2 \right) \mathbf{B}_T^{-1} \left(\tau_1, \tau_2 \right) \mathbf{D}_T^* \left(\tau_1, \tau_2 \right) \mathbf{B}_T \left(\tau_1, \tau_2 \right) \boldsymbol{A}_T \left(\tau_1, \tau_2 \right)}
$$

is $o_p^*(1)$ P-a.s., uniformly on T_Δ. The rest of the proof replicates the proof of Lemma S.6.

On general probability spaces the $o_p^*(1)$ in P-probability approximations of Lemma S.7 remain valid because weak convergence in probability to zero is equivalent to the $o_p^*(1)$ in P-probability property. ∎

Proof of Proposition 1

Proof of (i). We prove the result for $\mathcal{T}^d$, the other cases being similar. From Lemma S.6, we have that, with a uniform remainder term on T_Δ,

$$
t_{\beta_1}^2 (\tau_1, \tau_2) = \frac{T^{-1-2\delta_I} \{ S_{z_I y} \left(\tau_2 \right) - S_{z_I y} \left(\tau_1 \right) \}^2}{T^{-1-2\delta_I} \{ S_{z_I^2 y^2} \left(\tau_2 \right) - S_{z_I^2 y^2} \left(\tau_1 \right) \}} + o_p(1) \Rightarrow \frac{\left(G_I \left(\tau_2 \right) - G_I \left(\tau_1 \right) + \int_{\tau_1}^{\tau_2} b(s) \mathrm{d} K_{z_I x}(s) \right)^2}{[G_I] \left(1 \right) \left(\eta_I \left(\tau_2 \right) - \eta_I \left(\tau_1 \right) \right)}
$$

on $\mathcal{D}(T_\Delta)$, the convergence by Lemma S.2(6,8) for $T^{-1-2\delta_I} S_{z_I^2 y^2}(\tau)$ and $T^{-1/2-\delta_I} S_{z_I y}(\tau)$, and the CMT (regarding the continuity of the involved transformations of $T^{-1-2\delta_I} S_{z_I^2 y^2}(\tau)$ and $T^{-1/2-\delta_I} S_{z_I y}(\tau)$, see the introductory discussion of $\mathcal{D}(T_\Delta)$ and $\mathcal{C}(T_\Delta)$). Since the limit process of $t_{\beta_1}^2 (\tau_1, \tau_2)$ is a random element of $\mathcal{C}(T_\Delta)$ and the functional $f \mapsto \sup_{T_\Delta} |f|$ is continuous on $\mathcal{C}(T_\Delta)$, by the CMT we obtain that

$$
\mathcal{T}^d \Rightarrow \sup_{(\tau_1, \tau_2) \in T_\Delta} \frac{\left(G_I \left(\tau_2 \right) - G_I \left(\tau_1 \right) + \int_{\tau_1}^{\tau_2} b(s) \mathrm{d} K_{z_I x}(s) \right)^2}{[G_I] \left(1 \right) \left(\eta_I \left(\tau_2 \right) - \eta_I \left(\tau_1 \right) \right)}
$$

as required.

Proof of (ii). The proof is entirely analogous under Assumption 1.2; applying Lemma S.6 item **(ii)** where

$$
\begin{aligned}
Q_{II} \left(\tau_1, \tau_2 \right) \Rightarrow & \int_{\tau_1}^{\tau_2} Z(s) \mathrm{d} U(s) + \omega \int_{\tau_1}^{\tau_2} Z(s) b(s) J_{c,\mathrm{H}}(s) \mathrm{d} s \\
& - \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} Z(s) \mathrm{d} s \left(U \left(\tau_2 \right) - U \left(\tau_2 \right) + \omega \int_{\tau_1}^{\tau_2} b(s) J_{c,\mathrm{H}}(s) \mathrm{d} s \right)
\end{aligned}
$$

and

$$
P_{II} \left(\tau_1, \tau_2 \right) \Rightarrow \int_{\tau_1}^{\tau_2} Z^2(s) \mathrm{d} [U](s)
$$

$$-\frac{2}{\tau_2 - \tau_1} \left(\int_{\tau_1}^{\tau_2} Z(s)\mathrm{d}[U](s) \right) \left(\int_{\tau_1}^{\tau_2} Z(s)\mathrm{d}s \right)$$

$$+\frac{1}{(\tau_2 - \tau_1)^2} \left(\int_{\tau_1}^{\tau_2} Z(s)\mathrm{d}s \right)^2 ([U](\tau_2) - [U](\tau_1))$$

give the desired result after defining $\tilde{Z}_{\tau_1,\tau_2}(s) := Z(s) - \frac{1}{\tau_2-\tau_1} \int_{\tau_1}^{\tau_2} Z(s)$. ∎

Proof of Proposition 2

On the special probability spaces where the convergence results in Lemmas S.4 and S.5 were established to hold weakly a.s., an argument similar to that for Proposition 1 establishes the weak a.s. convergence of the distributions of the bootstrap statistics, under the null and under local alternatives, to the limiting null distributions of the original statistics. On general probability spaces this convergence remains valid weakly in probability.

For instance, on the special probability space from the proof of Lemma S.4 it holds, by the proof of Lemma S.7 that, with a uniform remainder term on T_Δ P-a.s.,

$$t_{\beta_1}^{*2}(\tau_1,\tau_2) = \frac{T^{-1-2\delta_I}\{S_{z_Iy}^*(\tau_2) - S_{z_Iy}^*(\tau_1)\}^2}{T^{-1-2\delta_I}\{S_{z_I^2y^2}^*(\tau_2) - S_{z_I^2y^2}^*(\tau_1)\}} + o_{p^*}(1) \overset{w}{\Rightarrow}_{a.s.} \frac{(G_I^*(\tau_2) - G_I^*(\tau_1))^2}{[G_I](1)(\eta_I(\tau_2) - \eta_I(\tau_1))} \quad \text{(S.11)}$$

on $\mathcal{D}(T_\Delta)$, the convergence by Lemmas S.2(6) for $T^{-1-2\delta_I}S_{z_I^2y^2}^*(\tau)$ (in a.s. sense), Lemma S.4(2) for $T^{-1/2-\delta_I}S_{z_Iy}^*(\tau)$ (in $\overset{w}{\Rightarrow}_{a.s.}$ sense), and the CMT applied conditionally on the (redefined) data. As the limit process of $t_{\beta_1}^{*2}(\tau_1,\tau_2)$ takes values in $\mathcal{C}(T_\Delta)$ a.s. and the functional $f \mapsto \sup_{T_\Delta}|f|$ is continuous on $\mathcal{C}(T_\Delta)$, by applying the CMT again it follows that

$$\mathcal{T}^{d*} \overset{w}{\Rightarrow}_{a.s.} \sup_{(\tau_1,\tau_2)\in T_\Delta} \frac{(G_I^*(\tau_2) - G_I^*(\tau_1))^2}{[G_I](1)(\eta_I(\tau_2) - \eta_I(\tau_1))} \overset{d}{=} \sup_{\substack{0\leq r_1,r_2\leq 1 \\ \eta_I^{-1}(r_2)-\eta_I^{-1}(r_1)\geq\Delta\tau}} \frac{(W(r_2) - W(r_1))^2}{r_2 - r_1} \equiv \mathcal{T}_\infty^{d,I},$$

the latter being the limiting null distribution of the original statistic $\mathcal{T}^d$ under stable predictors. On a general probability space, therefore, $\mathcal{T}^{d*} \overset{w}{\Rightarrow}_p \mathcal{T}_\infty^{d,I}$. Convergence of the remaining bootstrap distributions is established similarly.

Returning to the pair $\mathcal{T}^d, \mathcal{T}^{d*}$ under the conditions of Proposition 1(i), the weak convergence

$$\begin{pmatrix} \mathrm{P}(\mathcal{T}^d \leq \cdot) \\ \mathrm{P}^*(\mathcal{T}^{d*} \leq \cdot) \end{pmatrix} \Longrightarrow \mathrm{P}(\mathcal{T}_\infty^{d,I} \leq \cdot)\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

on $\mathcal{D}(\mathbb{R})^2$ and the continuity of the cdf $F(\cdot) = \mathrm{P}(\mathcal{T}_\infty^{d,I} \leq \cdot)$ lead, by means of a standard argument, to $\mathrm{P}^*(\mathcal{T}^{d*} \leq \mathcal{T}^d) \Longrightarrow F(\mathcal{T}_\infty^{d,I}) \overset{d}{=} U[0,1]$. The proof of bootstrap validity in the near-unit root case and for the remaining test statistics is analogous. ∎

S.4 Additional Monte Carlo Results

To illustrate the impact of conditional heteroskedasticity and heavy tails on our proposed tests we now consider an additional DGP (DGP4), in which the innovation vector $(u_t, v_t)'$ is generated to exhibit time-varying conditional second-order moments according to the design

$$
\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{bmatrix} \sigma_{1t} & 0 \\ 0 & \sigma_{2t} \end{bmatrix} \boldsymbol{\eta}_t; \quad E(\boldsymbol{\eta}_t) = 0, \quad E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') =: \boldsymbol{\Omega}_\phi = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}
$$

where $\boldsymbol{\eta}_t := (\eta_{1t}, \eta_{2t})'$ is an i.i.d. vector drawn from either a multivariate Gaussian distribution or a (fat-tailed) multivariate Student-t distribution with 5 degrees of freedom. The covariance matrix $\boldsymbol{\Omega}_\phi$ depends on the contemporaneous correlation coefficient ϕ which, as in section 6 of the main paper, was set to $\phi = -0.90$. The conditional variances σ_{it}^2, $i = 1, 2$, are driven by (normalised) stationary GARCH(1,1) processes characterised by:

$$
\sigma_{it}^2 = (1 - \theta_1 - \theta_2) + \theta_1 e_{i,t-1}^2 + \theta_2 \sigma_{i,t-1}^2, \ \ i = 1, 2
$$

with $\theta_1, \theta_2 \geq 0$ and $\theta_1 + \theta_2 < 1$, $e_{1,t-1} = u_{t-1}$, $e_{2,t-1} = v_{t-1}$ and $E(e_{it}^2) = 1$. For simplicity, we impose the same GARCH dynamics on the two series, focusing on GARCH parameter configurations that allow for varying degrees of persistence in the conditional variances as measured by $\theta_1 + \theta_2$, namely, $(\theta_1, \theta_2) \in \{(0.15, 0.5), (0.1, 0.85)\}$.

For each of the parameter configurations, the two sample lengths ($T = 250$ and $T = 500$), and the two conditional distributions (multivariate Gaussian distribution and multivariate Student-t distribution with 5 degrees of freedom), we compute the test statistics and determine empirical rejection frequencies at the 5% nominal (asymptotic) level, using the same number of Monte Carlo and bootstrap replications as were used for the simulation results relating to empirical size reported in section 6.1 of the main paper. The results from these experiments are reported in Tables S.1 and S.2.

Table S.1: Empirical Rejection Frequencies under the Null Hypothesis, H_0. 5% Nominal Significance Level. DGP4 with $\phi = -0.90$ and Gaussian Innovations.

c	$t^{2*}_{\beta_1}$	$t^{2*}_{\beta_1,NW}$	$t^2_{\beta_1}$	$t^2_{\beta_1,NW}$	$\mathcal{T}^{f*}$	$\mathcal{T}^{b*}$	$\mathcal{T}^{f*}_{NW}$	$\mathcal{T}^{b*}_{NW}$	$\mathcal{T}^{r*}$	$\mathcal{T}^{r*}_{NW}$	$\mathcal{T}^{d*}$	$\mathcal{T}^{d*}_{NW}$
				$\theta_1 = 0.15$, $\theta_2 = 0.5$ and $T = 250$								
0	0.069	0.073	0.071	0.070	0.033	0.033	0.046	0.057	0.011	0.058	0.009	0.067
2.5	0.057	0.059	0.058	0.059	0.028	0.044	0.040	0.065	0.012	0.053	0.008	0.048
5.0	0.057	0.060	0.057	0.059	0.030	0.048	0.039	0.062	0.011	0.053	0.011	0.051
10.0	0.062	0.062	0.060	0.061	0.040	0.049	0.046	0.059	0.016	0.053	0.020	0.062
20.0	0.058	0.057	0.056	0.058	0.046	0.054	0.049	0.063	0.031	0.056	0.030	0.061
$0.5T$	0.054	0.052	0.051	0.057	0.062	0.062	0.055	0.056	0.071	0.058	0.069	0.060
				$\theta_1 = 0.15$, $\theta_2 = 0.5$ and $T = 500$								
0	0.069	0.071	0.068	0.071	0.040	0.051	0.048	0.070	0.016	0.063	0.014	0.055
2.5	0.051	0.054	0.054	0.054	0.029	0.053	0.035	0.070	0.020	0.059	0.023	0.058
5.0	0.058	0.060	0.056	0.058	0.031	0.059	0.036	0.072	0.022	0.056	0.027	0.057
10.0	0.059	0.061	0.058	0.059	0.036	0.063	0.043	0.066	0.034	0.057	0.036	0.055
20.0	0.057	0.059	0.057	0.058	0.040	0.059	0.045	0.061	0.043	0.055	0.048	0.061
$0.5T$	0.053	0.051	0.050	0.054	0.056	0.057	0.052	0.054	0.062	0.056	0.055	0.047
				$\theta_1 = 0.1$, $\theta_2 = 0.85$ and $T = 250$								
0	0.073	0.076	0.073	0.077	0.035	0.034	0.043	0.056	0.010	0.061	0.012	0.070
2.5	0.056	0.058	0.057	0.062	0.028	0.039	0.035	0.060	0.012	0.054	0.010	0.051
5.0	0.058	0.058	0.055	0.063	0.031	0.047	0.038	0.062	0.014	0.054	0.011	0.051
10.0	0.057	0.057	0.056	0.065	0.038	0.046	0.042	0.062	0.018	0.051	0.022	0.057
20.0	0.057	0.055	0.053	0.064	0.047	0.053	0.048	0.060	0.029	0.055	0.036	0.061
$0.5T$	0.057	0.055	0.052	0.066	0.062	0.060	0.054	0.053	0.069	0.059	0.071	0.060
				$\theta_1 = 0.1$, $\theta_2 = 0.85$ and $T = 500$								
0	0.072	0.073	0.070	0.070	0.036	0.049	0.043	0.071	0.014	0.057	0.035	0.063
2.5	0.054	0.053	0.053	0.058	0.028	0.052	0.032	0.070	0.016	0.059	0.021	0.054
5	0.055	0.055	0.055	0.061	0.031	0.061	0.037	0.070	0.024	0.056	0.026	0.052
10	0.055	0.056	0.052	0.062	0.036	0.063	0.040	0.066	0.034	0.059	0.037	0.054
20	0.055	0.054	0.052	0.063	0.039	0.061	0.042	0.061	0.042	0.055	0.055	0.051
$0.5T$	0.053	0.051	0.051	0.073	0.055	0.054	0.051	0.049	0.066	0.052	0.066	0.056

Notes: A superscript $*$ denotes tests run using the fixed regressor wild bootstrap outlined in Algorithm 1; $t^2_{\beta_1}$ and $t^2_{\beta_1,NW}$ denote the full sample IV-combination predictability tests of Breitung and Demetrescu (2015) based on the 5% asymptotic critical value from the χ^2_1 distribution and computed with Eicker-White [EW] and conventional standard errors, respectively, and $t^{2*}_{\beta_1}$ and $t^{2*}_{\beta_1,NW}$ their bootstrap analogues; $\mathcal{T}^{f*}$, $\mathcal{T}^{b*}$ and $\mathcal{T}^{f*}_{NW}$, $\mathcal{T}^{b*}_{NW}$, denote the maximum forward and backward recursive tests computed with EW and conventional standard errors, respectively; $\mathcal{T}^{r*}$ and $\mathcal{T}^{r*}_{NW}$ denote the maximum rolling tests computed with EW and conventional standard errors, respectively; $\mathcal{T}^{d*}$ and $\mathcal{T}^{d*}_{ols}$ denote the maximum double recursive tests computed with EW and conventional standard errors, respectively.

Table S.2: Empirical Rejection Frequencies under the Null Hypothesis, H_0. 5% Nominal Significance Level. DGP4 with $\phi = -0.90$ and Student-t (with 5 degrees of freedom) Distributed Innovations.

c	$t^{2*}_{\beta_1}$	$t^{2*}_{\beta_1,NW}$	$t^2_{\beta_1}$	$t^2_{\beta_1,NW}$	$\mathcal{T}^{f*}$	$\mathcal{T}^{b*}$	$\mathcal{T}^{f*}_{NW}$	$\mathcal{T}^{b*}_{NW}$	$\mathcal{T}^{r*}$	$\mathcal{T}^{r*}_{NW}$	$\mathcal{T}^{d*}$	$\mathcal{T}^{d*}_{NW}$
				$\theta_1 = 0.15$, $\theta_2 = 0.5$ and $T = 250$								
0	0.071	0.074	0.071	0.072	0.029	0.035	0.042	0.060	0.009	0.056	0.008	0.066
2.5	0.049	0.051	0.050	0.052	0.026	0.043	0.036	0.061	0.009	0.051	0.011	0.063
5.0	0.050	0.052	0.052	0.055	0.026	0.043	0.032	0.061	0.013	0.051	0.011	0.060
10.0	0.053	0.053	0.053	0.056	0.032	0.048	0.040	0.060	0.019	0.054	0.025	0.067
20.0	0.052	0.051	0.049	0.056	0.040	0.049	0.043	0.057	0.031	0.057	0.041	0.072
0.5T	0.053	0.052	0.050	0.061	0.062	0.060	0.055	0.054	0.063	0.053	0.072	0.055
				$\theta_1 = 0.15$, $\theta_2 = 0.5$ and $T = 500$								
0	0.068	0.072	0.067	0.070	0.036	0.048	0.045	0.0696	0.015	0.053	0.015	0.062
2.5	0.051	0.050	0.052	0.054	0.031	0.055	0.039	0.068	0.017	0.053	0.022	0.055
5.0	0.054	0.053	0.053	0.056	0.034	0.055	0.042	0.067	0.024	0.058	0.026	0.066
10.0	0.054	0.055	0.055	0.058	0.039	0.056	0.045	0.064	0.033	0.059	0.037	0.070
20.0	0.056	0.056	0.055	0.059	0.046	0.055	0.050	0.064	0.043	0.059	0.046	0.065
0.5T	0.050	0.049	0.049	0.064	0.060	0.056	0.057	0.049	0.066	0.057	0.076	0.067
				$\theta_1 = 0.10$, $\theta_2 = 0.85$ and $T = 250$								
0	0.069	0.073	0.074	0.079	0.033	0.036	0.045	0.060	0.010	0.058	0.010	0.062
2.5	0.054	0.055	0.055	0.066	0.026	0.043	0.036	0.060	0.009	0.049	0.010	0.050
5.0	0.053	0.054	0.054	0.068	0.027	0.046	0.035	0.061	0.014	0.052	0.014	0.057
10.0	0.054	0.054	0.053	0.067	0.035	0.051	0.041	0.061	0.021	0.052	0.026	0.071
20.0	0.054	0.053	0.051	0.073	0.040	0.051	0.044	0.059	0.031	0.054	0.047	0.072
0.5T	0.057	0.053	0.052	0.082	0.063	0.061	0.055	0.057	0.061	0.050	0.073	0.051
				$\theta_1 = 0.10$, $\theta_2 = 0.85$ and $T = 500$								
0	0.068	0.071	0.072	0.078	0.034	0.045	0.045	0.066	0.013	0.058	0.013	0.048
2.5	0.050	0.051	0.052	0.063	0.031	0.054	0.036	0.066	0.015	0.051	0.020	0.058
5.0	0.055	0.052	0.053	0.069	0.032	0.055	0.039	0.064	0.023	0.053	0.023	0.060
10.0	0.053	0.053	0.052	0.070	0.041	0.058	0.046	0.063	0.036	0.057	0.037	0.066
20.0	0.056	0.053	0.051	0.079	0.048	0.058	0.050	0.059	0.048	0.057	0.039	0.054
0.5T	0.051	0.049	0.048	0.098	0.058	0.059	0.057	0.053	0.066	0.062	0.076	0.055

Notes: See notes to Table S.1

S.5 Additional Material for Section 7

Table S.3: List of predictors used

1.	the log dividend price ratio (dp_t).
2.	the log dividend yield (dy_t).
3.	the log earnings price ratio (e/p_t).
4.	the log dividend payout ratio (de_t).
5.	the equity risk premium volatility, $rvol_t$.
6.	the book to market ratio (bm_t).
7.	the net equity expansion ($ntis_t$).
8.	the treasury bill rate (tbl_t).
9.	the long-term government bond yield (lty_t).
10.	the long-term government bond rate of return (ltr_t).
11.	the term spread (tms_t).
12.	the default yield spread (dfy_t).
13.	the default return spread (dfr_t).
14.	inflation ($infl_t$).

Notes: Detailed description of the variables used can be found on Amit Goyal's web page (see http://www.hec.unil.ch/agoyal/docs/AllTables.pdf). The equity risk premium volatility ($rvol_t$) is computed as in Neely *et al.* (2014).
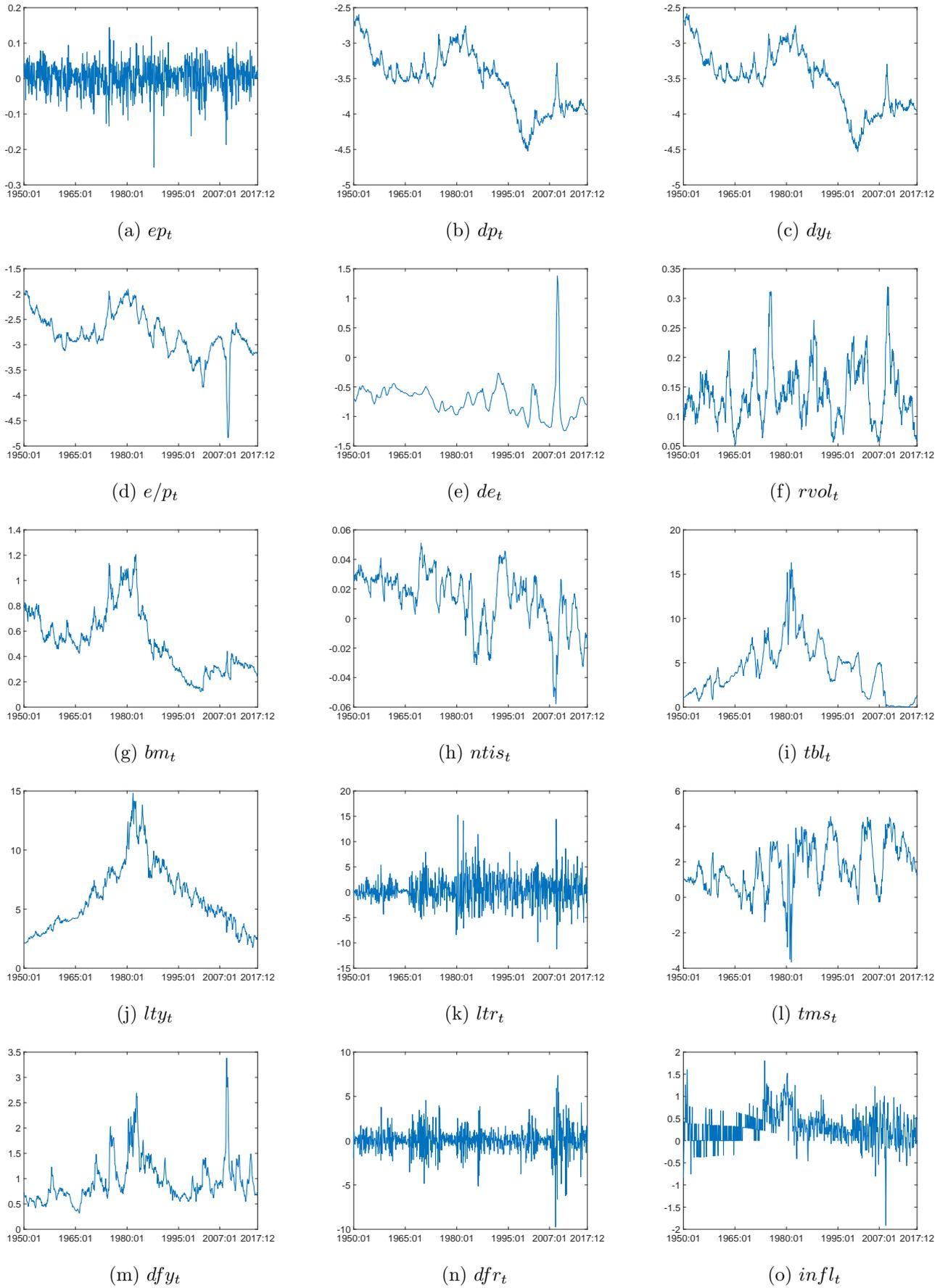
Figure S.1: Graphs of excess returns and predictors

Additional References

References

Bickel, P. and M. Wichura (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics* 42, 1656–1670.

Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2010). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics* 158, 7–24.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

Kallenberg O. (1997). *Fundamentals of Modern Probability*. Springer, NY.

Neely C. J., D. E. Rapach, J. Tu and G. Zhou (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science* 60.