Econometric Theory, **0**, 2023, 1–44.
doi:10.1017/S0266466623000154

A NOVEL APPROACH TO PREDICTIVE ACCURACY TESTING IN NESTED ENVIRONMENTS

JEAN-YVES PITARAKIS
University of Southampton

We introduce a new approach for comparing the predictive accuracy of two nested models that bypasses the difficulties caused by the degeneracy of the asymptotic variance of forecast error loss differentials used in the construction of commonly used predictive comparison statistics. Our approach continues to rely on the out of sample mean squared error loss differentials between the two competing models, leads to nuisance parameter-free Gaussian asymptotics, and is shown to remain valid under flexible assumptions that can accommodate heteroskedasticity and the presence of mixed predictors (e.g., stationary and local to unit root). A local power analysis also establishes their ability to detect departures from the null in both stationary and persistent settings. Simulations calibrated to common economic and financial applications indicate that our methods have strong power with good size control across commonly encountered sample sizes.

1. INTRODUCTION

This paper is concerned with comparing the forecasting performance of two nested models through tests that rely on out of sample mean squared error (MSE) loss differentials. Our proposed approach bypasses the widely documented complications caused by the degenerate asymptotic variances of these differentials that occur in nested environments while also leading to nuisance parameter-free standard normal asymptotics. Our approach remains valid under both stationary and persistent predictors thus also greatly expanding its practical relevance in economics and finance.

Since the early work of Diebold and Mariano (1995) and West (1996), a vast body of theoretical research has been concerned with developing new methods for comparing the out of sample predictive ability of competing models. Such tests typically compare the out of sample forecast errors generated from two models under a variety of loss functions and forecasting schemes (e.g., recursive, rolling, or fixed updating of model parameter estimates) with the aim of testing the null hypothesis of equal predictive accuracy. Most of the test statistics introduced

I wish to thank the Editor, Co-Editor, and three anonymous referees for the quality of the reports I have received and their in-depth review of an earlier version of this paper. I also wish to thank the ESRC for its financial support via grant ES/W000989/1. Any errors are my own responsibility. Address correspondence to Jean-Yves Pitarakis, Department of Economics, University of Southampton, Southampton SO17 1BJ, United Kingdom; e-mail: j.pitarakis@soton.ac.uk.

in this literature are based on estimated out of sample MSE loss differentials associated with the two competing forecast error series and have been shown to be asymptotically normally distributed provided that the models being compared are non-nested and a set of standard regularity conditions hold.

The fundamental difficulties that arise as one moves from a non-nested to a nested environment have also generated a vast and growing literature aiming to operationalize and adapt the above approach to nested models. In a nested modeling context, a key complication comes from the fact that under the null of interest the population errors of the two models are identical thus leading to sample MSE loss differentials that are identically zero in the limit with null asymptotic variances. These in turn result in test statistics that are not well-defined asymptotically and in the failure of normal approximations for popular test statistics such as the Diebold–Mariano statistic (henceforth referred to as DM).

Alternative normalizations applied to the MSE loss differentials in nested contexts have subsequently been shown to lead to test statistics with well-defined but no longer Gaussian limiting null distributions expressed as functionals of stochastic integrals in Brownian Motions (Clark and McCracken, 2001, 2005; McCracken, 2007; Hansen and Timmermann, 2015). With the exception of restrictive frameworks that rule out heteroskedasticity or allow only a single additional predictor in the nesting model these distributions typically depend on a variety of model specific parameters that cannot be eliminated via standard HAC-type corrections, requiring simulation-based approaches for their implementation (see West, 2006, pp. 126–127; Clark and McCracken, 2013, pp. 10–15). The asymptotics of these test statistics are further influenced by how the in-sample observations are allowed to grow relative to the out of sample observations and the particular choice of the forecasting scheme used to generate forecasts.

Rather than relying on these nonstandard and non-Gaussian distributions this same literature has also proposed to bypass the difficulties underlying nested model comparisons by continuing to use normal approximations for adjusted versions of DM-type statistics. In Clark and West (2007), for instance, the authors introduced an adjustment to the spread of the out of sample MSEs of the two competing models and argued that although asymptotic normality cannot be established per se such an approach results in reasonably accurate inferences with acceptable size distortions. The adjustment essentially corrects for the fact that under the null hypothesis of equal predictive accuracy the MSE of the larger model is contaminated with estimation noise. This adjusted DM-type statistic proposed in Clark and West (2007) has become the norm in economic applications involving out of sample forecast comparisons with recent examples found in Molodotsova and Papell (2009), Ince, Molodotsova, and Papell (2016), Engel and Wu (2021) amongst numerous others.

In this paper, we introduce an alternative formulation of the out of sample MSE loss differential between two models that is not subject to the variance degeneracy problem of existing procedures. This subsequently allows us to construct novel test statistics for testing the null hypothesis of equal out of sample population MSEs which are shown to have simple nuisance parameter-free normal distributions. The main idea underlying our proposed approach is based on the observation that MSE comparisons across two competing models need not be performed within the same out of sample span of available forecast error observations. These can be performed over partially overlapping segments instead, leading to test statistics that accumulate MSE spreads over all possible such segments. This new setting can trivially accommodate desirable features such as conditional heteroskedasticity and persistent predictors and is also shown to lead to both consistent and locally powerful test statistics. As we discuss further below, our approach can also be adapted to broader contexts where the nestedness of models is an important consideration for inferences such as model selection testing.

Besides conventional forecasting objectives, nested models are commonly encountered environments when it comes to testing economic hypotheses and validating theories. Notable examples include forecast accuracy comparisons against random walk models in the exchange rate literature spurred by the early work of Meese and Rogoff (1983) and more recently reconsidered in Rossi (2005), Molodotsova and Papell (2009) amongst others, equity premium predictability issues as recently investigated in Ferson, Nallareddy, and Biqin (2013), Avdis and Wachter (2017), and numerous others. Our key aim here is to propose a way of addressing and resolving a long-standing issue that has generated a vast agenda on the formal comparison of such models via their out-of-sample predictive accuracy. The important auxiliary debate on the advantages or disadvantages of using out-of-sample versus in-sample approaches is not part of our focus. It is also important to emphasize that our interest here is on testing population-level predictive ability when forecasts are generated recursively as opposed to finite sample-based predictive ability as considered, for instance, in Giacomini and White (2006). This latter approach is able to avoid the complications induced by the nestedness of models being compared by proceeding via a rolling-fixed window-based forecasting scheme so that the issue of competing models becoming identical in the limit can be bypassed.

Throughout this paper, we also followed the common practice of referring to statistics based on MSE differentials obtained from competing estimated models as Diebold–Mariano-type statistics. We must acknowledge, however, that the specific testing approach initially developed by these authors was not concerned with model specific considerations or specification testing motives as its underlying theory was developed for *given* sequences of forecast errors assumed to satisfy certain regularity conditions (see Diebold, 2015). Nevertheless, the forecasting literature of the past decade has generally amalgamated the notion of forecast evaluation with the evaluation of models on the basis of their forecasting abilities.

The paper is organized as follows: Section 2 introduces the nested forecasting environment and establishes the limiting null distributions of two novel test statistics. Section 3 concentrates on their asymptotic power properties, establishing their consistency and ability to detect local departures from the null. Section 4 introduces a simple adjustment to the same statistics shown to further enhance their power properties without affecting their null distributions. Section 5 provides a comprehensive finite sample evaluation of our methods based on two data generating processes (DGPs) calibrated to commonly encountered applications. Section 6 illustrates the use of our proposed methods via an application to exchange rate models. Section 7 overviews our key results and discusses extensions. Proofs are given in the Appendix. Further simulation results are provided in the Supplementary Material.

2. MODELS AND TEST STATISTICS: THEORY

We consider the following predictive regressions:

$$y_{t+1} = \boldsymbol{x}_{1t}'\boldsymbol{\delta}_1 + v_{t+1}, \tag{1}$$

$$y_{t+1} = \boldsymbol{x}_{1t}'\boldsymbol{\beta}_1 + \boldsymbol{x}_{2t}'\boldsymbol{\beta}_2 + u_{t+1}, \tag{2}$$

where the $\boldsymbol{x}_{it}$'s are the $(p_i \times 1)$ vectors of predictors, $\boldsymbol{\delta}_1$ and $\boldsymbol{\beta}_i$ the $(p_1 \times 1)$ and $(p_i \times 1)$ parameter vectors, and v_t and u_t the random disturbance terms. We let $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}', \boldsymbol{x}_{2t}')'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ and set $p = p_1 + p_2$. Here, model (1) is nested within the larger model in (2) and under $\boldsymbol{\beta}_2 = 0$, we have $\boldsymbol{\delta}_1 \equiv \boldsymbol{\beta}_1$ and $v_{t+1} \equiv u_{t+1}$. The formulation of the above two nested models is standard and parallels closely the most commonly encountered setting considered in the predictive accuracy testing literature as, for instance, in Hansen and Timmermann (2015).

One step ahead forecasts of y_{t+1} from (1) and (2) are generated recursively as $\hat{y}_{1,t+1|t} = \boldsymbol{x}_{1t}'\hat{\boldsymbol{\delta}}_{1t}$ and $\hat{y}_{2,t+1|t} = \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_t$ for $t = k_0, \ldots, T-1$, where $\hat{\boldsymbol{\delta}}_{1t} = (\sum_{j=1}^t \boldsymbol{x}_{1j-1}\boldsymbol{x}_{1j-1}')^{-1}\sum_{j=1}^t \boldsymbol{x}_{1j-1}'y_j$, $\hat{\boldsymbol{\beta}}_t = (\sum_{j=1}^t \boldsymbol{x}_{j-1}\boldsymbol{x}_{j-1}')^{-1}\sum_{j=1}^t \boldsymbol{x}_{j-1}'y_j$ and the resulting pseudo out of sample forecast errors are then obtained as $\hat{e}_{1,t+1} = y_{t+1} - \boldsymbol{x}_{1t}'\hat{\boldsymbol{\delta}}_{1t}$ and $\hat{e}_{2,t+1} = y_{t+1} - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_t$. Here, k_0 is the sample location used to initiate the first recursive forecasts that lead to the first out of sample forecast errors $\hat{e}_{1,k_0+1}$ and $\hat{e}_{2,k_0+1}$ and subsequently resulting in $(T-k_0)$ out of sample forecast error observations. Throughout this paper, we take k_0 to be a given fraction of the sample size, setting $k_0 = [T\pi_0]$ for some $\pi_0 \in (0,1)$.

Following the early work of Diebold and Mariano (1995), West (1996), and others, a common approach for comparing the predictive accuracy of the two models under MSE loss involves testing the null hypothesis

$$H_0\colon E[y_{t+1} - \hat{y}_{1,t+1}(\boldsymbol{\delta}_1)]^2 = E[y_{t+1} - \hat{y}_{2,t+1}(\boldsymbol{\beta})]^2 \tag{3}$$

using a test statistic based on suitably normalized versions of the sample average MSE loss differentials

$$\overline{D}_T = \frac{1}{T-k_0}\left(\sum_{t=k_0}^{T-1}\hat{e}_{1,t+1}^2 - \sum_{t=k_0}^{T-1}\hat{e}_{2,t+1}^2\right). \tag{4}$$

Within a non-nested setting and a strictly stationary and ergodic environment, Diebold and Mariano (1995) and West (1996) established a standard normal limit

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

theory for this class of test statistics (e.g., $\sqrt{T-k_0}\,\overline{D}_T/\hat{\sigma}_{\overline{D}_T}$ with $\hat{\sigma}^2_{\overline{D}_T}$ denoting some suitable long-run variance estimator) leading to their systematic use in applied work and a voluminous literature on their refinements. Within a nested context, where $u_{t+1} \equiv v_{t+1}$, however, it is straightforward to observe that $\sqrt{T-k_0}\,\overline{D}_T \xrightarrow{p} 0$ and $\hat{\sigma}_{\overline{D}_T} \xrightarrow{p} 0$ invalidating the limiting standard normal approximation and the use of these test statistics for inference purposes. This degeneracy problem is not solely confined to the Diebold—Mariano-type statistics but universally affects all existing methods that compare forecast errors (or models) in nested settings with recursively generated forecasts.

These observations have led to a vast body of research on out of sample predictive accuracy testing in nested models due to their importance in empirical applications in areas such as asset pricing and the modeling of expected returns in particular. For their validity, inferences in nested contexts such as (1) and (2) must rely on the observation that under H_0 it is $(T-k_0)\overline{D}_T$ rather than $\sqrt{T-k_0}\,\overline{D}_T$ that turns out to have a nondegenerate limit which could be used for developing suitable inferences (see Clark and McCracken (2001, 2005); McCracken (2007); Hansen and Timmermann (2015)). This, however, is also problematic due to the nonstandard and non-pivotal nature of the resulting asymptotic distributions. These take the form of functionals of stochastic integrals in Brownian Motions and with the exception of some special cases contain nuisance parameters that are difficult to remove via standard HAC-type normalizations. Even under special instances such as conditional homoskedasticity these distributions continue to depend on the number of extra predictors included in the nesting models and the fraction of the sample used to build the first recursive forecasts. Equally importantly these results have been obtained under stationarity and ergodicity assumptions ruling out the important and frequently encountered case of predictors having roots near unity in their autoregressive representations.

Instead of evaluating the two sequences of squared forecast errors $\{\hat{e}^2_{1,t+1}\}$ and $\{\hat{e}^2_{2,t+1}\}$ over the entire *and* same interval $[k_0+1,T]$ as it is done in the formulation of all commonly used test statistics based on $\overline{D}_T$ we here propose to compare the two out of sample MSEs over partially overlapping segments of the $[k_0+1,T]$ interval instead. For this purpose, we introduce the following generalized MSE spread

$$\widetilde{D}_T(\ell_1,\ell_2) = \frac{\sum_{t=k_0}^{k_0+\ell_1-1} \hat{e}^2_{1,t+1}}{\ell_1} - \frac{\sum_{t=k_0}^{k_0+\ell_2-1} \hat{e}^2_{2,t+1}}{\ell_2}, \tag{5}$$

where ℓ_1 and ℓ_2 control the range over which the two squared forecast error sequences are evaluated. Note that setting $\ell_1 = \ell_2 = T-k_0$ in (5) reduces it to $\overline{D}_T$ which can be viewed as a special case of $\widetilde{D}_T(\ell_1,\ell_2)$. In line with the analysis based on (4), we take $\ell_1 = [(T-k_0)\lambda_1]$ and $\ell_2 = [(T-k_0)\lambda_2]$ with λ_1 and λ_2 referring to the fraction of the $(T-k_0)$ squared forecast errors associated with models (1) and (2), respectively.

From a theoretical standpoint, proceeding with the use of $\widetilde{D}_T(\ell_1,\ell_2)$ instead of $\overline{D}_T$ has no bearing on the null hypothesis being tested in the sense that when $u_{t+1} \equiv v_{t+1}$ the population counterpart of $\widetilde{D}_T(\ell_1,\ell_2)$ also equals zero. A key feature of (5) that distinguishes it from $\overline{D}_T$, however, is that the variance of its suitably normalized version will no longer be degenerate provided that $\ell_1 \neq \ell_2$ (equivalently $\lambda_1 \neq \lambda_2$). This normalized version of (5) which forms the basis of our proposed test statistics is given by $Z_T(\ell_1,\ell_2) = \sqrt{T-k_0}\,\widetilde{D}_T(\ell_1,\ell_2)$,

$$Z_T(\ell_1,\ell_2) = \frac{T-k_0}{\ell_1}\left[\frac{\sum_{t=k_0}^{k_0+\ell_1-1}\hat{e}^2_{1,t+1}}{\sqrt{T-k_0}} - \frac{\ell_1}{\ell_2}\frac{\sum_{t=k_0}^{k_0+\ell_2-1}\hat{e}^2_{2,t+1}}{\sqrt{T-k_0}}\right]. \tag{6}$$

Note that (6) is simply the normalized difference in the means of the two sample MSEs evaluated over the two relevant segments of the effective sample size.

Remark 1. A key point to observe here is that the variance of (6) is well-defined and no longer collapses to zero in the limit provided that λ_1 and λ_2 are bounded away from zero and bounded away from each other. To illustrate and motivate this point heuristically, let us replace both $\hat{e}^2_{1t+1}$ and $\hat{e}^2_{2t+1}$ in (6) with $(u^2_{t+1}-\sigma^2_u)$ for $\sigma^2_u \equiv E[u^2_t]$. Taking the $u_t's$ to be IID(0,1) with $E[u^4_{t+1}] < \infty$ it follows that

$$V[Z_T(\ell_1,\ell_2)] \to V[u^2_{t+1}]\,\frac{|\lambda_2-\lambda_1|}{\lambda_1\lambda_2} \tag{7}$$

suggesting that a test statistic based on $\widetilde{D}_T(\ell_1,\ell_2)$ will not have a degenerate distribution as it was the case with the use of $\overline{D}_T$ in nested contexts. We may also wish to point out that having λ_1 and λ_2 bounded away from zero is merely a technical requirement in the asymptotics that follow as in practice these parameters will naturally be set at or near their maximum boundary of one.

The quantity in (6) forms the building block of our proposed test statistics for testing the null hypothesis in (3) against one-sided right tail-based alternatives as it is the norm in this literature. We consider two types of test statistics that operationalize (6). Our choice is guided by the simplicity of the ensuing asymptotics and their intuitive interpretation while recognizing that alternative constructions/normalizations of $\widetilde{D}_T(\ell_1,\ell_2)$ may also be considered.

The first test statistic that we consider is denoted $Z^0_T(\lambda^0_1,\lambda^0_2)$ and is based on implementing inferences for given magnitudes $\ell^0_1 = [(T-k_0)\lambda^0_1]$ and $\ell^0_2 = [(T-k_0)\lambda^0_2]$. We write

$$Z^0_T(\lambda^0_1,\lambda^0_2) = \frac{1}{\hat{\sigma}}\,Z_T([(T-k_0)\lambda^0_1],[(T-k_0)\lambda^0_2]) \tag{8}$$

with $\hat{\sigma}^2$ denoting a consistent estimator of $V[u^2_{t+1}]$.

Our second test statistic is based on averaging (6) across the ℓ_j's. The averaging can be implemented over $\ell_1 \in [1, T-k_0]$ for a given ℓ^0_2 (e.g., $\ell^0_2 = T-k_0$) so that

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

the MSE of the smaller model accumulates progressively as ℓ_1 increases. More generally, this averaging can be performed over any desired and feasible range of ℓ_1. To allow such level of generality, we introduce the fractional parameter τ_0 and write

$$\overline{Z}_T(\tau_0;\lambda_2^0) = \frac{1}{\hat{\sigma}} \frac{1}{[(T-k_0)(1-\tau_0)]} \sum_{\ell_1=[(T-k_0)\tau_0]+1}^{T-k_0} Z_T(\ell_1, [(T-k_0)\lambda_2^0]), \tag{9}$$

where the choice of τ_0 determines the user-chosen range of ℓ_1 over which the average of $Z_T(\ell_1, \ell_2^0)$ is taken (given $\ell_2^0 = [(T-k_0)\lambda_2^0]$). Rather than imposing a fixed and given λ_1^0 as in (8), this average-based statistic essentially considers a range of such magnitudes and subsequently aggregates outcomes via averaging. Given the role played by ℓ_1 and ℓ_2 in our inferences, we can expect that choosing the averaging range in a way that excludes low magnitudes of ℓ_1 (so that the estimated MSEs associated with model 1 remain sufficiently accurate) will result in more reliable inferences. The issue of how best to select these user inputs is postponed until Section 3 where we provide precise guidelines informed by a theoretical local power analysis. One motivation behind this average-based statistic when compared with (8) is that one can remain partly more agnostic about the specific magnitude to use for one of the two required user inputs in $Z_T([(T-k_0)\lambda_1^0], [(T-k_0)\lambda_2^0])$ while setting the other one (e.g., λ_2^0) at or near its maximum boundary of one. Although our context is different, this is also reminiscent of the various approaches used in the structural break testing literature when one does not wish to take a stance on the location of a potential change-point. More importantly, and borrowing from the same literature, we may also conjecture that the averaging process may result in tests with more favorable size–power tradeoffs.

At this stage, it is also important to point out that there are numerous alternative possibilities for designing test statistics in the spirit of (8) and (9) (e.g., double averaging across ℓ_1 and ℓ_2, alternative functional forms, etc.). An interesting avenue for future research could be the design of a class of test statistics based on $Z_T(\ell_1, \ell_2)$ and having desirable optimality properties as it has been attempted in the structural break literature.

Although both (8) and (9) allow for a broad range of theoretically feasible magnitudes for $(\lambda_1^0, \lambda_2^0)$ in $Z_T^0(\lambda_1^0, \lambda_2^0)$ and (τ_0, λ_2^0) in $\overline{Z}_T(\tau_0; \lambda_2^0)$ one naturally expects that choosing $(\lambda_1^0, \lambda_2^0)$ and (τ_0, λ_2^0) to lie in the vicinity of unity would capture the greatest amount of information from the two competing models. As we show further below such a choice does indeed lead to remarkably powerful tests with excellent size control. Given a sequence of forecast errors available to the investigator, the practical implementation of either (8) or (9) is also as straightforward as calculating standard DM-type test statistics.

To establish the limiting properties of our test statistics under the null hypothesis in (3), we introduce a set of high-level assumptions ensuring a flexible environment that encompasses the vast majority of settings considered in the literature while also allowing for a richer temporal structure. As we wish to highlight the generality and usefulness of our methods based on the use of (5) and (6), we abstain from primitive conditions that may unnecessarily suggest a restrictive scope for their use. More importantly, our use of high-level assumptions is motivated by the fact that our proposed methods can be immediately seen to be robust to a very rich dynamic structure of predictors including highly persistent processes, strictly stationary and ergodic processes, long memory processes, etc.

Assumption A.

(i) $$\sup_{\lambda\in(0,1]}\left|\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}\hat{e}_{j,t+1}^2}{\sqrt{T-k_0}}-\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}u_{t+1}^2}{\sqrt{T-k_0}}\right|\overset{H_0}{=}o_p(1)\text{ for }j=1,2.$$

(ii) The sequence of demeaned squared errors $\eta_t=u_{t+1}^2-\sigma_u^2$ has autocovariances γ_j^η that satisfy $\sum_{j=0}^{\infty}|\gamma_j^\eta|<\infty$ and fulfills a functional central limit theorem, that is, $T^{-\frac{1}{2}}\sum_{t=1}^{[Ts]}(u_{t+1}^2-\sigma_u^2)\overset{\mathcal{D}}{\to}\sigma W_\eta(s)$ on $D_{\mathbb{R}}([0,1])$ the space of cadlag functions on $[0,1]$ with $W_\eta(.)$ denoting a standard Brownian Motion and $\sigma^2=\gamma_0^\eta+2\sum_{j=1}^{\infty}\gamma_j^\eta>0$.

(iii) A consistent estimator $\hat{\sigma}^2$ of σ^2 exists, that is, $\hat{\sigma}^2\overset{p}{\to}\sigma^2\in(0,\infty)$.

We note that condition A(i) holds for both $j=1$ and $j=2$ highlighting the fact that we operate within a nested environment with (1) being the true model. A(i) is trivially satisfied under a very broad range of settings used to obtain the large sample properties of DM-type statistics in nested models. An important feature to also highlight here is the fact that A(i) does not restrict the persistence properties of the predictors which could be highly persistent in the sense of following local to unit-root processes for instance. This greatly expands and enriches the environment in which predictive accuracy inferences have commonly been introduced. The robustness of A(i) to the persistence properties of the predictors is an important and useful feature of the squared forecast errors as opposed to their level for which a result such as A(i) would not hold. For more primitive conditions illustrating specialized environments under which A(i) holds, see Deng and Perron (2008a) and Hansen and Timmermann (2015) for strictly stationary and ergodic/mixing settings and Berenguer-Rico and Nielsen (2020) for environments where A(i) is shown to hold under both stationary and unit-root or near unit-root regressors.

Assumption A(ii) requires the centered squared errors driving (1) and (2) to satisfy a functional central limit theorem with σ^2 referring to their long-run variance. The absolute summability of the autocovariances of η_t ensures that σ^2 the limit of $V[\sum_{t=k_0}^{T-1}\eta_{t+1}/\sqrt{T-k_0}]$ exists. Examples of processes which satisfy Assumption A(ii) include a broad range of conditionally heteroskedastic ARCH/GARCH processes under suitable existence of moments restrictions. For a detailed set of primitive assumptions ensuring that the stated FCLT holds, see Giraitis, Kokoszka, and Leipus (2000, Thm. 5.1), Giraitis et al. (2001, Ex. 2.2

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

and Thm. 2.1), Berkes, Hörmann, and Horvath (2008), Linder (2009), and the references therein.

Assumption A(iii) requires that the long-run variance of η_t be estimated consistently. Such an estimator could be trivially constructed using least squares residuals from either the null or alternative models. Under conditional homoskedasticity, an obvious candidate would be $\hat{\sigma}^2_{hom} = \sum_{t=k_0}^{T-1} \hat{\eta}^2_{t+1}/(T-k_0)$ while under dependent errors (e.g., if the $u_t's$ follow a GARCH-type process) a Newey–West-type formulation as in Deng and Perron (2008b) would be suitable and ensure that A(iii) holds.

Remark 2. As pointed out in Remark 1, our asymptotic theory for $Z^0_T(\lambda^0_1, \lambda^0_2)$ in (8) imposes λ^0_1 and λ^0_2 to be bounded away from zero and to be bounded away from each other, say $0 < \underline{\lambda} \leq \lambda^0_i \leq 1$ for $i = 1, 2$ and $|\lambda^0_1 - \lambda^0_2| \geq \epsilon$ for some positive fraction ϵ. In what follows, we refer to such a set from which these two user inputs can be selected as Λ^0. The implementation of the average-based statistic $\overline{Z}_T(\tau_0; \lambda^0_2)$ in (9) requires setting λ^0_2 as above and averaging $Z_T(\ell_1, [(T-k_0)\lambda^0_2])$ across $\ell_1 = [(T-k_0)\tau_0]+1, \ldots, (T-k_0)$ for some τ_0 bounded away from zero and one. We refer to this set as $\overline{\Lambda}^0$.

The following two propositions summarize the large sample behavior of our two test statistics under the null hypothesis stated in (3).

PROPOSITION 1. *Under Assumption A(i)–(iii), the null hypothesis in (3), and for given* $(\lambda^0_1, \lambda^0_2) \in \Lambda^0$, *we have as* $T \to \infty$

$$Z^0_T(\lambda^0_1, \lambda^0_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, v^0(\lambda^0_1, \lambda^0_2)), \quad \textbf{(10)}$$

where

$$v^0(\lambda^0_1, \lambda^0_2) = \frac{|\lambda^0_1 - \lambda^0_2|}{\lambda^0_1 \lambda^0_2}. \quad \textbf{(11)}$$

PROPOSITION 2. *Under Assumption A(i)–(iii), the null hypothesis in (3), and for given* $(\tau_0, \lambda^0_2) \in \overline{\Lambda}^0$, *we have as* $T \to \infty$

$$\overline{Z}_T(\tau_0; \lambda^0_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \bar{v}(\tau_0; \lambda^0_2)), \quad \textbf{(12)}$$

where

$$\bar{v}(\tau_0; {\lambda_2}^0) = \begin{cases} \dfrac{(1-\tau_0)^2 + 2\lambda^0_2(1-\tau_0+\ln\tau_0)}{\lambda^0_2(1-\tau_0)^2}, & \lambda^0_2 \leq \tau_0, \quad \textbf{(13)} \\ \dfrac{1-\tau_0^2 + 2\lambda^0_2((1-\tau_0)\ln\lambda^0_2 + \tau_0\ln\tau_0)}{\lambda^0_2(1-\tau_0)^2}, & \lambda^0_2 > \tau_0. \quad \textbf{(14)} \end{cases}$$

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

The variance components of the distributional outcomes in (10) and (12) are of course known to the investigator so that both $Z_T^0(\lambda_1^0, \lambda_2^0)$ and $\overline{Z}_T(\tau_0; \lambda_2^0)$ can be trivially standardized as

$$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) \equiv \frac{Z_T^0(\lambda_1^0, \lambda_2^0)}{\sqrt{\nu^0(\lambda_1^0, \lambda_2^0)}} \tag{15}$$

and

$$\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0) \equiv \frac{\overline{Z}_T(\tau_0; \lambda_2^0)}{\sqrt{\bar{\nu}(\tau_0; \lambda_2^0)}} \tag{16}$$

to proceed with standard normal inferences for testing H_0.

The results in (10)–(14) highlight the simplicity and practicality of our proposed inferences while at the same time offering a solution to an important problem that has not been satisfactorily resolved in this literature. The test statistics in (15) and (16) allow us to generalize the widely used DM style forecast accuracy testing approach to a broad class of empirically relevant models including specifications with highly persistent predictors with or without conditional heteroskedasticity.

We naturally expect the quality of inferences (e.g., power, size vs. power tradeoffs) to be influenced by the specific choices of $(\lambda_1^0, \lambda_2^0)$ in (15) and (τ_0, λ_2^0) in (16). Although the above null asymptotics hold under a very broad range of parameterizations for those user inputs, a formal analysis of their local asymptotic power allows us to provide precise and tight guidelines ensuring excellent power properties with good size control.

3. ASYMPTOTIC POWER AND TEST PARAMETERIZATIONS

We here deviate from Assumption A(i) in order to evaluate the large sample behavior of $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ when the DGP is given by (2). Assumption A(i) continues to hold for $j = 2$ but no longer for $j = 1$ since model (1) is misspecified due to the omitted $\boldsymbol{x}_{2,t}$ predictors. As $\hat{e}_{1,t+1}^2$ will now be contaminated by those omitted predictors we expect the stochastic properties of the latter (e.g., the variance of the $\boldsymbol{x}_{2,t}$'s and their correlation with the $\boldsymbol{x}_{1,t}$'s) to influence the power properties of both test statistics. Unlike their null distributions, we thus also expect the test statistics to diverge at different rates depending on whether the predictors are stationary or highly persistent. Our analysis of the consistency and local power properties of $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ is guided by these two distinct scenarios which we consider separately.

3.1. Consistency and Local Power under Stationarity

We initially concentrate on the case where the predictors driving both (1) and (2) are stationary and ergodic. Specifically, we operate under the following set

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

of high-level assumptions that mirror closely the most common environments considered in the predictive accuracy testing literature.

Assumption B1.

(i) $\sup_{\lambda \in [0,1]} \left\| \frac{\sum_{t=1}^{[T\lambda]} \boldsymbol{x}_t \boldsymbol{x}_t'}{T} - \lambda\, \boldsymbol{Q} \right\| = o_p(1)$ with $\boldsymbol{Q}$ a $p \times p$ nonrandom positive definite matrix,

(ii) $\frac{\sum_{t=1}^{[T\lambda]} \boldsymbol{x}_t u_{t+1}}{\sqrt{T}} \xrightarrow{\mathcal{D}} \boldsymbol{\Omega}^{1/2}\ \boldsymbol{W}(\lambda)$ with $\boldsymbol{W}(.)$ denoting a p-dimensional standard Brownian Motion and $\boldsymbol{\Omega} = E[\boldsymbol{x}_t \boldsymbol{x}_t' u_{t+1}^2] > 0$,

(iii) Assumption A(ii) and (iii) holds.

(iv) The user inputs in $\mathcal{S}_T(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ are such that $(\lambda_1^0, \lambda_2^0) \in \Lambda^0$ and $(\tau_0, \lambda_2^0) \in \overline{\Lambda}^0$, respectively.

Assumption B1(i)–(iii) mirrors closely the environment of Hansen and Timmermann (2015) and can be viewed as more primitive conditions ensuring that Assumption A(i) holds. 3.1(i) requires that the predictors satisfy a uniform law of large numbers and rules out trending or local to unit-root predictors while B1(ii) ensures that $\{\boldsymbol{x}_t u_{t+1}\}$ satisfies a multivariate functional central limit theorem. Our main result regarding the asymptotic power properties of the two tests within such a stationary environment is now summarized in Proposition 3.

PROPOSITION 3. *(i) Suppose model (2) holds with* $\boldsymbol{\beta}_2 \neq 0$ *and fixed, then under Assumption B1 and as* $T \to \infty$*, we have* $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) \xrightarrow{p} \infty$ *and* $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0) \xrightarrow{p} \infty$. *(ii) Suppose model (2) holds with* $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{1/4}$ *for* $\boldsymbol{\gamma} \neq \boldsymbol{0}$. *Under Assumption B1,* $\lim_{||\boldsymbol{\gamma}|| \to \infty} \lim_{T \to \infty} \mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) = \infty$ *and* $\lim_{||\boldsymbol{\gamma}|| \to \infty} \lim_{T \to \infty} \overline{\mathcal{S}}_T(\tau_0; \lambda_2^0) = \infty$ *in probability.*

The above results highlight the consistency of both test statistics as well as their ability to detect local departures from the null hypothesis under stationary settings. It is here also important to point out that the local to the null parameterization of $\boldsymbol{\beta}_2$ based on $T^{1/4}$ rather than the usual $T^{1/2}$ rate commonly encountered in stationary settings is not in any way due to our specific test statistics or assumptions. The same scenario would also occur in a conventional regression-based testing environment and is due to the fact that we are dealing with inferences about the behavior of *squared* errors rather than their level.

To gain further insights into the specific role played by key factors influencing power it is useful to also present the explicit asymptotic local power functions of the two tests for a given size $\alpha \in (0,1)$. These will in turn be used to provide explicit guidance on selecting suitable parameterizations of our two test statistics (i.e., $(\lambda_1^0, \lambda_2^0)$ in (15) and $(\tau_0; \lambda_2^0)$ in (16)). In what follows, it is useful to also recall that π_0 refers to the given fraction of the sample size used to initiate the recursive computation of forecasts.

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

COROLLARY 1. *Suppose model (2) holds with* $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{1/4}$ *for* $\boldsymbol{\gamma} \neq 0$. *Under Assumption B1 and letting* q_α *denote the upper* α*-quantile of the standard normal distribution with cdf* $\Phi(.)$*, the asymptotic local power functions of the tests based on* $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ *and* $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ *are given by* $1 - \Phi(q_\alpha - \psi^0)$ *and* $1 - \Phi(q_\alpha - \overline{\psi})$*, respectively, where*

$$\psi^0 = \left[\frac{\sqrt{1-\pi_0}}{\sigma\sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \boldsymbol{\gamma}'(\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\gamma} \right], \tag{17}$$

$$\overline{\psi} = \left[\frac{\sqrt{1-\pi_0}}{\sigma\sqrt{\bar{v}(\tau_0; \lambda_2^0)}} \boldsymbol{\gamma}'(\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\gamma} \right], \tag{18}$$

with $v^0(\lambda_1^0, \lambda_2^0)$ *and* $\bar{v}(\tau_0; \lambda_2^0)$ *as in (11) and (13), (14), and the* $\boldsymbol{Q}_{ij}$*'s referring to the components of the population moment matrix* $\boldsymbol{Q}$ *in Assumption B1(i).*

We note that power is monotonic in the sense that both ψ^0 and $\overline{\psi}$ are nondecreasing as $||\boldsymbol{\gamma}||$ gets large. For a given significance level, the larger the two noncentrality parameters are the greater the associated probabilities of rejecting the null hypothesis.

The expressions in (17) and (18) are particularly useful for highlighting the factors that influence power by shifting the center of the null asymptotic standard normal distributions away from zero. Viewing the asymptotic local power functions $\Phi(\psi^0 - q_\alpha)$ and $\Phi(\overline{\psi} - q_\alpha)$ in Corollary 1 as providing approximations to the correct decision frequencies of the two test statistics under a sufficiently large T and specific alternatives, we note that for a given size α both test statistics are expected to exhibit a stronger ability to detect departures from the null when the variances of the omitted predictors are large and their correlation with the included predictors small. This feature is particularly important since it hints at the fact that the presence of nearly integrated predictors may help enhance power, a scenario we formally consider further below.

To highlight these points with greater clarity, it is useful to focus on the simplified case of two centered predictors $\boldsymbol{x}_t = (x_{1,t}, x_{2.t})$ so that (17) and (18) simplify as

$$\psi^0 = \left[\frac{\sqrt{1-\pi_0}}{\sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \frac{\gamma^2}{\sigma}(1-\rho_{12}^2)E[x_{2,t}^2] \right], \tag{19}$$

$$\overline{\psi} = \left[\frac{\sqrt{1-\pi_0}}{\sqrt{\bar{v}(\tau_0; \lambda_2^0)}} \frac{\gamma^2}{\sigma}(1-\rho_{12}^2)E[x_{2,t}^2] \right], \tag{20}$$

with $\rho_{12} = Corr[x_{1,t}, x_{2,t}]$. All other things being equal, power is expected to deteriorate under a noisy omitted predictor that has low variance (low $E[x_{2,t}^2]$) and/or that is highly correlated with the included predictor (e.g., $|\rho_{12}| \approx 1$).

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Interestingly, this also suggests that an ideal setting in terms of power implications is one where omitted predictors are highly persistent while included predictors are stationary so that $\rho_{12} \approx 0$ with $E[x_{2,t}^2]$ large. Another important factor affecting power is the variance of the u_t^2's which impacts the magnitudes of ψ^0 and $\overline{\psi}$ via $\sigma \equiv \sqrt{V[u_{t+1}^2]}$. Within an NID errors setting, for instance, we have $V[u_{t+1}^2] = E[u_{t+1}^4] - \sigma_u^4 = 2\sigma_u^4$ so that *all other things being equal*, an environment with high kurtosis will have a detrimental impact on the power properties of both test statistics.

Power enhancing choices for $(\lambda_1^0, \lambda_2^0)$ ***and*** $(\tau_0; \lambda_2^0)$

The non-centrality parameters in (17) and (18) are also useful for assessing the impact of $(\lambda_1^0, \lambda_2^0)$ and (τ_0, λ_2^0) on both the absolute and relative local powers of the two tests and for providing useful guidance on suitable choices for those user inputs. From Corollary 1, since the mapping $m \mapsto P[Z > q_\alpha - m]$ is increasing in m on $[0, \infty)$, a test of size α based on $S_T^0(\lambda_1^0, \lambda_2^0)$ will be preferable, in terms of its local power, to a test of the same size based on $S_T^0(\lambda_1^{0'}, \lambda_2^{0'})$ whenever $\psi^0(\lambda_1^0, \lambda_2^0) > \psi^0(\lambda_1^{0'}, \lambda_2^{0'})$, holding all other parameters entering ψ^0 constant. Given ψ^0 in (17) with $v^0(\lambda_1^0, \lambda_2^0)$ defined as in (11) it follows that those two parameters should be set near their boundary of 1 and in close vicinity of one another (e.g., $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ for $\lambda_2^0 \approx 0.9$ as a possibility).

Regarding the average-based statistic $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$, we note from (13) and (14) that $\overline{\psi}$ in (18) viewed as a function of λ_2^0 and τ_0 (holding all other parameters constant) reaches its unique maximum for

$$\lambda_2^0 = 0.5\ \tau_0 + 0.5 \qquad \textbf{(21)}$$

supporting the use of $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0 = 0.5\tau_0 + 0.5)$ in its practical implementation. If $\tau_0 = 0.5$, for instance, which corresponds to a test statistic that averages across the largest half of the ℓ_1 magnitudes, this approximate asymptotic power-based metric points to an implementation-based on $\overline{\mathcal{S}}_T(\tau_0 = 0.5; \lambda_2^0 = 0.75)$. Since $\overline{\psi}$ is also a monotonically increasing function of τ_0, however, it also follows that the same average-based statistic should be operationalized with a choice of τ_0 that is in the vicinity of 1 (e.g., $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0 = 0.5(0.8) + 0.5)$ or $\overline{\mathcal{S}}_T(\tau_0 = 0.9; \lambda_2^0 = 0.5(0.9) + 0.5)$ as possibilities). A practical side to this power enhancing choice of λ_2^0 is that the implementation of $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ essentially requires only a single user input.

Given these preferred parameterizations of the two test statistics it is also useful to evaluate whether either of the two statistics is expected to dominate the other in the sense of ψ^0 being greater or smaller than $\overline{\psi}$ over particular regions of the pairs $(\lambda_1^0, \lambda_2^0)$ and $(\tau_0, \lambda_2^0 = 0.5\tau_0 + 0.5)$, holding all other parameters constant. Given the standard normal asymptotics of both test statistics a useful metric for comparing their local powers is Pitman's Asymptotic Relative Efficiency (ARE) which here takes particularly simple forms, following directly from Corollary 1. To avoid

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

confusion between the λ_2^0 parameter used in $\mathcal{S}_T^0(\lambda_1^0,\lambda_2^0)$ and λ_2^0 used in $\overline{\mathcal{S}}_T(\tau_0,\lambda_2^0)$, we write the two statistics as $\mathcal{S}_T^0(\lambda_1^0,\lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0;\overline{\lambda}_2^0)$ with $\overline{\lambda}_2^0 = 0.5\tau_0 + 0.5$ as in (21). Their ARE is now given by

$$\text{ARE}(\mathcal{S}^0,\overline{\mathcal{S}}) = \left[\frac{\overline{v}(\tau_0;0.5\tau_0+0.5)}{v^0(\lambda_1^0;\lambda_2^0)}\right] \tag{22}$$

and more specifically

$$\text{ARE}(\mathcal{S}^0,\overline{\mathcal{S}}) = \frac{\lambda_1^0\lambda_2^0}{|\lambda_1^0-\lambda_2^0|}\frac{2(1-\tau_0)(1+\ln((1+\tau_0)/2))+2\tau_0\ln\tau_0}{(1-\tau_0)^2}. \tag{23}$$

From (23), we can observe a clear trade-off between τ_0 and the magnitudes of λ_1^0 and λ_2^0 used in $\mathcal{S}_T^0(\lambda_1^0,\lambda_2^0)$. If we focus on $\lambda_1^0 = 1$ it follows from (23) that ARE ≥ 1 for

$$\lambda_2^0 \geq \frac{(1-\tau_0)^2}{(1-\tau_0)(3-\tau_0)+2[\ln 0.5(1+\tau_0)-\tau_0\ln((1+\tau_0)/2\tau_0)]}, \tag{24}$$

which is a monotonically increasing function of τ_0 and highlights the fact that the average based statistic will dominate $\mathcal{S}_T^0(\lambda_1^0=1,\lambda_2^0)$ in terms of its local power (i.e., ARE < 1) unless impractically large magnitudes of λ_2^0 are used in its implementation. If the average based statistic is implemented with $\tau_0 = 0.8$, for instance, its power properties will dominate $\mathcal{S}_T^0(\lambda_1^0=1,\lambda_2^0)$ unless $\lambda_2^0 > 0.9798$. If it is implemented with $\tau_0 = 0.9$, the average-based statistic will again dominate $\mathcal{S}_T^0(\lambda_1^0=1,\lambda_2^0)$ unless $\lambda_2^0 > 0.9908$. These values suggest that the average-based statistic with τ_0 set in the vicinity of unity (e.g., $\overline{\mathcal{S}}(\tau_0 = 0.8;\overline{\lambda}_2^0 = 0.9)$) will dominate $\mathcal{S}_T^0(\lambda_1^0=1,\lambda_2^0)$ in terms of its local power unless impractically large magnitudes of λ_2^0 are used in $\mathcal{S}_T^0(\lambda_1^0=1,\lambda_2^0)$.

3.2. Consistency and Local Power under Persistence

We now consider an environment where the p predictors $\boldsymbol{x}_t$ entering (1) and (2) are modeled as local to unit-root processes specified as

$$\boldsymbol{x}_t = \left(\boldsymbol{I}_p - \frac{\boldsymbol{C}}{T}\right)\boldsymbol{x}_{t-1} + \epsilon_t, \tag{25}$$

where $\boldsymbol{C} = diag(c_1,\ldots,c_p)$ for $c_i > 0$, $i = 1,\ldots,p$ and ϵ_t some stationary and ergodic random disturbance process. The new set of assumptions under which we establish our results are now summarized in Assumption B2, where $\boldsymbol{J}_C(s) = (\boldsymbol{J}_{1C}(s),\boldsymbol{J}_{2C}(s))'$ denotes a p-dimensional Ornstein–Uhlenbeck process whose two components $\boldsymbol{J}_{1C}(s)$ and $\boldsymbol{J}_{2C}(s)$ are associated with the dynamics of $\boldsymbol{x}_{1,t}$ and $\boldsymbol{x}_{2,t}$, respectively.

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Assumption B2.

(i) $\left(\frac{\boldsymbol{x}_{[Ts]}}{\sqrt{T}}, \frac{\sum_{t=1}^{[Ts]} u_t}{\sqrt{T}}, \frac{\sum_{t=1}^{[Ts]}(u_t^2-\sigma_u^2)}{\sqrt{T}}\right) \xrightarrow{\mathcal{D}} (\boldsymbol{J}_C(s), \sigma_u W_u(s), \sigma W(s))$, $s \in [0,1]$.

(ii) Assumption A(iii) holds.

(iii) The user inputs in $\mathcal{S}_T(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ are such that $(\lambda_1^0, \lambda_2^0) \in \Lambda^0$ and $(\tau_0, \lambda_2^0) \in \overline{\Lambda}^0$, respectively.

The asymptotic power properties of $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2)$ are now summarized in Proposition 4.

Proposition 4. *(i) Suppose model (2) holds with* $\boldsymbol{\beta}_2 \neq 0$ *and fixed, then under Assumption B2 and as* $T \to \infty$*, we have* $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) \xrightarrow{p} \infty$ *and* $\overline{\mathcal{S}}_T(\tau_0; \lambda_2) \xrightarrow{p} \infty$*. (ii) Suppose model (2) holds with* $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{3/4}$ *for* $\boldsymbol{\gamma} \neq \mathbf{0}$*. Under Assumption B2,* $\lim_{||\boldsymbol{\gamma}||\to\infty} \lim_{T\to\infty} \mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) = \infty$ *and* $\lim_{||\boldsymbol{\gamma}||\to\infty} \lim_{T\to\infty} \overline{\mathcal{S}}_T(\tau_0; \lambda_2) = \infty$ *in probability.*

A key message that is conveyed by Proposition 4 when contrasted with Proposition 3 is the important impact of persistence on the power properties the test statistics. The presence of persistent predictors leads to a faster divergence rate for both statistics as reflected in the faster convergence rate towards zero of $\boldsymbol{\beta}_2$ that can be accommodated. With highly persistent predictors, both test statistics diverge at the same $T^{3/2}$ rate compared with a rate of $T^{1/2}$ when predictors were stationary.

A more explicit formulation of the departure from the null distribution in this local to unit-root context can also be highlighted through the following formulations of the limiting distributions of the two test statistics under the local alternative of interest.

Corollary 2. *Suppose model (2) holds with* $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{3/4}$ *for* $\boldsymbol{\gamma} \neq \mathbf{0}$*. Under assumption B2 and as* $T \to \infty$*, we have*

$$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) + \xi^0, \tag{26}$$

$$\overline{\mathcal{S}}_T(\tau_0; \lambda_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) + \overline{\xi}, \tag{27}$$

with

$$\xi^0 = \frac{\sqrt{1-\pi_0}}{\sigma\sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \boldsymbol{\gamma}' \left(\frac{1}{(1-\pi_0)\lambda_1^0} \int_{\pi_0}^{\pi_0+(1-\pi_0)\lambda_1^0} \boldsymbol{J}_C^*(s) \boldsymbol{J}_C^*(s)' \right) \boldsymbol{\gamma}, \tag{28}$$

$$\overline{\xi} = \frac{\sqrt{1-\pi_0}}{\sigma\sqrt{\overline{v}(\tau_0; \lambda_2^0)}} \boldsymbol{\gamma}' \left(\frac{1}{(1-\tau_0)} \int_{\tau_0}^{1} \frac{1}{(1-\pi_0)\lambda_1} \left(\int_{\pi_0}^{\pi_0+(1-\pi_0)\lambda_1} \boldsymbol{J}_C^*(s) \boldsymbol{J}_C^*(s)' \right) d\lambda_1 \right) \boldsymbol{\gamma}, \tag{29}$$

where $\boldsymbol{J}_C^*(s) = \boldsymbol{J}_{2C}(s) - \boldsymbol{M}(s)\boldsymbol{J}_{1C}(s)$ *and* $\boldsymbol{M}(s) = (\int_0^s \boldsymbol{J}_{1C}\boldsymbol{J}_{1C}')^{-1}(\int_0^s \boldsymbol{J}_{1C}\boldsymbol{J}_{2C}')$.

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

It is here interesting to compare (28) and (29) with the non-centrality parameters (17) and (18) obtained in the stationary context. The two pairs are essentially analogous in the sense that the constant population moments of predictors (i.e., the $\boldsymbol{Q}_{i,j}$'s) are now replaced by stochastic integrals in Ornstein–Uhlenbeck processes (i.e., $\boldsymbol{J}_{i,C}$). The homogeneity throughout the sample of the limit moment matrix in Assumption B1(i) is of course no longer valid in the context of the stochastic integrals in (28) and (29). The above results also imply that the role played by the pairs $(\lambda_1^0, \lambda_2^0)$ and (τ_0, λ_2^0) in this local to unit-root context will mirror our earlier analysis based on a stationary setting, supporting the same practical implementation of both test statistics in terms of their parameterizations, i.e., the power enhancing choices for $(\lambda_1^0, \lambda_2^0)$ and (τ_0, λ_2^0) discussed above continue to hold in the current context.

4. POWER ENHANCEMENTS

Here, we explore a particular adjustment that can be applied to our two test statistics $\mathcal{S}_T(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ with the purpose of boosting their asymptotic local power properties without affecting their limiting null distributions. The theoretical principle underlying our proposed approach mirrors the idea in Fan, Liao, and Yao (2015) where the authors proposed to augment Wald-type statistics with a component that vanishes asymptotically under the null while diverging under alternatives of interest. More formally, we seek to augment our proposed two test statistics as

$$\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0) \equiv \mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) + h_T^0(\lambda_1^0, \lambda_2^0), \tag{30}$$

$$\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda_2^0) \equiv \overline{\mathcal{S}}_T(\tau_0; \lambda_2^0) + \overline{h}_T(\tau_0; \lambda_2^0), \tag{31}$$

for some suitably chosen $h_T^0(\lambda_1^0, \lambda_2^0)$ and $\overline{h}_T(\tau_0; \lambda_2^0)$ terms which are such that these adjusted versions of our two test statistics maintain the same limiting null distributions as in Proposition 1 while at the same time displaying more favorable power properties.

In what follows, we show that a particular transformation of the forecast errors $\hat{e}_{2,t+1}$ associated with the larger forecasting model can be used to design such augmentation terms in a way that fulfills the requirement that $h_T^0(\lambda_1^0, \lambda_2^0)$ and $\overline{h}_T(\tau_0; \lambda_2^0)$ vanish asymptotically under the null while diverging at a desirable rate under the alternative. The augmentation we propose to consider is motivated by the well-known Clark and West adjustment to DM-type statistics introduced in Clark and West (2007). The original motivation behind Clark and West's adjustment relied on the intuition that under the null hypothesis estimation noise contaminates the $\hat{e}_{2,t+1}$'s due to the estimation of parameters that are zero in the population. This in turn translates into an inflated MSE_2 resulting in test statistics that are severely undersized. Clark and West proposed to correct for such distortions by suitably adjusting the magnitudes of the forecast errors estimated from the larger model.

To lay down the context and with no loss of generality, it is useful to operate within a simplified version of (1) and (2), setting $\boldsymbol{\delta}_1 = 0$ and $\boldsymbol{\beta}_1 = 0$ so that $\hat{e}_{1,t+1} = u_{t+1}$ and $\hat{e}_{2,t+1} = u_{t+1} - \boldsymbol{x}'_{2,t}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$. We can now write $Z_T(\ell_1, \ell_2)$ in (6) as

$$\frac{\sum_{t=k_0}^{k_0+\ell_1-1} \hat{e}^2_{1,t+1}}{\sqrt{T-k_0}} - \frac{\ell_1}{\ell_2}\frac{\sum_{t=k_0}^{k_0+\ell_2-1} \hat{e}^2_{2,t+1}}{\sqrt{T-k_0}} = \frac{\sum_{t=k_0}^{k_0+\ell_1-1} u^2_{t+1}}{\sqrt{T-k_0}} - \frac{\ell_1}{\ell_2}\frac{\sum_{t=k_0}^{k_0+\ell_2-1} u^2_{t+1}}{\sqrt{T-k_0}}$$
$$+2\frac{\ell_1}{\ell_2}\frac{\sum_{t=k_0}^{k_0+\ell_2-1} (\hat{\boldsymbol{\beta}}_{2,t} - \boldsymbol{\beta}_2)'\boldsymbol{x}_{2,t}u_{t+1}}{\sqrt{T-k_0}}$$
$$-\frac{\ell_1}{\ell_2}\frac{\sum_{t=k_0}^{k_0+\ell_2-1} (\hat{\boldsymbol{\beta}}_{2,t} - \boldsymbol{\beta}_2)'\boldsymbol{x}_{2,t}\boldsymbol{x}'_{2,t}(\hat{\boldsymbol{\beta}}_{2,t} - \boldsymbol{\beta}_2)}{\sqrt{T-k_0}}. \quad \textbf{(32)}$$

Although it is implicit in our Assumption A(i) that the last two terms in the right-hand side of (32) vanish asymptotically under the null hypothesis, in finite samples, the rightmost quadratic form is likely to pull down the spread in MSEs causing their null distribution to be mis-centered. Noting that $(\hat{\boldsymbol{\beta}}_{2,t} - \boldsymbol{\beta}_2)'\boldsymbol{x}_{2,t}\boldsymbol{x}'_{2,t}(\hat{\boldsymbol{\beta}}_{2,t} - \boldsymbol{\beta}_2) \equiv (\hat{e}_{1,t+1} - \hat{e}_{2,t+1})^2$, Clark and West's (2007) proposal was to reformulate the sample MSE spreads between Models 1 and 2 with an adjusted version of $\hat{e}^2_{2,t+1}$, say $\widetilde{e}^2_{2,t+1}$, given by

$$\widetilde{e}^2_{2,t+1} = \hat{e}^2_{2,t+1} - (\hat{e}_{1,t+1} - \hat{e}_{2,t+1})^2. \quad \textbf{(33)}$$

It turns out that implementing the adjustment in (33) within our two test statistics (i.e., using $\widetilde{e}^2_{2,t+1}$ instead of $\hat{e}^2_{2,t+1}$ in $\mathcal{S}^0_T(\lambda^0_1, \lambda^0_2)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda^0_2)$) allows us to reformulate them as in (30) and (31) with $h^0_T(\lambda^0_1, \lambda^0_2)$ and $\overline{h}_T(\tau_0; \lambda^2_0)$ fulfilling the desirable requirements in Fan et al. (2015) in the sense that the adjustments do not alter the asymptotic null distributions of the test statistics while at the same time leading to an increase in the associated non-centrality parameters under the local alternatives of interest.

The expression in (32) is also useful for highlighting what distinguishes our framework that operates under $\ell_1 \neq \ell_2$ with a standard approach that sets $\ell_1 = \ell_2 = T - k_0$ as, for instance, in all Diebold–Mariano-type statistics. Under $\ell_1 = \ell_2$, we note that the first two terms in the right-hand side of (32) cancel out so that the asymptotic behavior of the expression is determined by the two rightmost quadratic forms whose *non-normalized* versions have been shown to be $O_p(1)$ with nonstandard limits (see Clark and McCracken, 2001, 2005). Allowing $\ell_1 \neq \ell_2$ essentially forces the asymptotics of the MSE spreads to be driven solely by the first two components in the right-hand side of (32).

Letting $\mathcal{S}^0_{T,adj}(\lambda^0_1, \lambda^0_2)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda^0_2)$ denote the adjusted versions of our two test statistics, it immediately follows from (33) and standard algebra that

$$\mathcal{S}^0_{T,adj}(\lambda^0_1,\lambda^0_2) = \mathcal{S}^0_T(\lambda^0_1,\lambda^0_2) + \frac{1}{\hat{\sigma}} \frac{1}{\lambda^0_2\sqrt{v^0(\lambda^0_1,\lambda^0_2)}} \frac{\sum_{t=k_0}^{k_0+[(T-k_0)\lambda^0_2]}(\hat{e}_{1,t+1}-\hat{e}_{2,t+1})^2}{\sqrt{T-k_0}}$$
$$\equiv \mathcal{S}^0_T(\lambda^0_1,\lambda^0_2) + h^0_T(\lambda^0_1,\lambda^0_2) \tag{34}$$

and

$$\overline{\mathcal{S}}_{T,adj}(\tau_0;\lambda^0_2) = \overline{\mathcal{S}}_T(\tau_0;\lambda^0_2) + \frac{1}{\hat{\sigma}} \frac{1}{\lambda^0_2\sqrt{\overline{v}(\tau_0;\lambda^0_2)}} \frac{\sum_{t=k_0}^{k_0+[(T-k_0)\lambda^0_2]}(\hat{e}_{1,t+1}-\hat{e}_{2,t+1})^2}{\sqrt{T-k_0}}$$
$$\equiv \overline{\mathcal{S}}_T(\tau_0;\lambda^0_2) + \overline{h}_T(\tau_0;\lambda^0_2). \tag{35}$$

The expressions in (34) and (35) highlight the fact that the adjustment to the MSE of the larger model results in test statistics that are augmented versions of $\mathcal{S}_T(\lambda^0_1,\lambda^0_2)$ and $\overline{\mathcal{S}}_T(\tau_0;\lambda^0_2)$. As we establish formally below, the presence of the additional terms $h^0_T(\lambda^0_1,\lambda^0_2)$ and $\overline{h}_T(\tau_0;\lambda^0_2)$ leaves the limiting null distributions unchanged as both quantities vanish asymptotically. Under the alternative, both $h^0_T(\lambda^0_1,\lambda^0_2)$ and $\overline{h}_T(\tau_0;\lambda^0_2)$ diverge to infinity at the same rate as $\mathcal{S}_T(\lambda^0_1,\lambda^0_2)$ and $\overline{\mathcal{S}}_T(\tau_0;\lambda^0_2)$ implying that $\mathcal{S}_{T,adj}(\lambda^0_1,\lambda^0_2)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0;\lambda^0_2)$ will also share the consistency and local power characteristics of their unadjusted counterparts in the sense of diverging to infinity as $||\boldsymbol{\gamma}|| \to \infty$. More importantly, however, the presence of $h^0_T(\lambda^0_1,\lambda^0_2)$ and $\overline{h}_T(\tau_0;\lambda^0_2)$ does result in different (strictly larger) noncentrality parameters that make these adjusted statistics have more favorable power properties. These features are formalized in Proposition 5 and Corollary 3.

PROPOSITION 5. *The results in Propositions 1–4 continue to hold when* $\mathcal{S}^0_T(\lambda^0_1,\lambda^0_2)$ *and* $\overline{\mathcal{S}}_T(\tau_0;\lambda^0_2)$ *are replaced with* $\mathcal{S}^0_{T,adj}(\lambda^0_1,\lambda^0_2)$ *and* $\overline{\mathcal{S}}_{T,adj}(\tau_0;\lambda^0_2)$*, respectively.*

COROLLARY 3. *(i) Under the assumptions of Corollary 1 (stationary predictors), the asymptotic local power functions of the tests based on* $\mathcal{S}^0_{T,adj}(\lambda^0_1,\lambda^0_2)$ *and* $\overline{\mathcal{S}}_{T,adj}(\tau_0;\lambda^0_2)$ *are given by* $1-\Phi(q_\alpha - 2\psi^0)$ *and* $1-\Phi(q_\alpha - 2\overline{\psi})$ *with* ψ^0 *and* $\overline{\psi}$ *as in (17) and (18). (ii) Under the assumptions of Corollary 2 (persistent predictors), we have* $\boldsymbol{\mathcal{S}}^0_{T,adj}(\lambda^0_1,\lambda^0_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)+\xi^0_{adj}$ *and* $\overline{\boldsymbol{\mathcal{S}}}_{T,adj}(\tau_0;\lambda^0_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)+\overline{\xi}_{adj}$*, where*

$$\xi^0_{adj} = \xi^0 + \frac{\sqrt{1-\pi_0}}{\sigma\sqrt{v^0(\lambda^0_1,\lambda^0_2)}}\boldsymbol{\gamma}'\left(\frac{1}{(1-\pi_0)\lambda^0_2}\int_{\pi_0}^{\pi_0+(1-\pi_0)\lambda^0_2}\boldsymbol{J}^*_C(s)\boldsymbol{J}^*_C(s)'\right)\boldsymbol{\gamma}, \tag{36}$$

$$\overline{\xi}_{adj} = \overline{\xi} + \frac{\sqrt{1-\pi_0}}{\sigma\sqrt{\overline{v}(\tau_0;\lambda^0_2)}}\boldsymbol{\gamma}'\left(\frac{1}{(1-\pi_0)\lambda^0_2}\int_{\pi_0}^{\pi_0+(1-\pi_0)\lambda^0_2}\boldsymbol{J}^*_C(s)\boldsymbol{J}^*_C(s)'\right)\boldsymbol{\gamma}, \tag{37}$$

with ξ^0 *and* $\overline{\xi}$ *as in (28) and (29).*

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Proposition 5 essentially implies that all our results regarding the null limiting distributions of $\mathcal{S}_T(\lambda_1^0, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ and their general power properties (consistency and detectability of local departure from the null) continue to hold for their adjusted counterparts while Corollary 3 documents important differences in their specific non-centrality terms.

Indeed, the results in Corollary 3 are particularly interesting and useful for the practical assessment of the power properties of the adjusted versus unadjusted statistics. We have an environment whereby the limiting distributions of the two types of test statistics are the same under the null hypothesis while their non-centrality parameters differ under the alternative, pointing to a more favorable behavior for the adjusted statistics when it comes to detecting local departures from the null.

Letting ψ_{adj}^0 and $\overline{\psi}_{adj}$ denote the non-centrality parameters associated with the adjusted statistics, Corollary 3(i) establishes that in a stationary context, we have $\psi_{adj}^0 = 2\psi^0$ and $\overline{\psi}_{adj} = 2\overline{\psi}$ so that $\psi_{adj}^0/\psi_0 = 2$ and $\overline{\psi}_{adj}/\overline{\psi} = 2$. In the case of persistent predictors, the comparison between ξ^0 and ξ_{adj}^0 and between $\overline{\xi}$ and $\overline{\xi}_{adj}$ also indicates that the adjusted quantities will stochastically dominate their non-adjusted counterparts in the sense that $P[\xi_{adj}^0 > q] \geq P[\xi^0 > q]$ and $P[\overline{\xi}_{adj} > q] \geq P[\overline{\xi} > q]$ for some given critical value q and this is again expected to translate into more favorable power outcomes for the adjusted statistics under persistent predictors as well.

5. EMPIRICAL SIZE AND POWER

In this section, we investigate the size and power properties of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ together with their adjusted versions across two DGPs calibrated to commonly encountered applications and sample sizes in macroeconomics and finance. The experiments are designed to emphasize the role of the pairs $(\lambda_1^0, \lambda_2^0)$ and $(\tau_0; \lambda_2^0)$ on inferences with the choice of their magnitudes guided by the analysis surrounding our results in Corollaries 1 and 2.

More specifically, the implementation of $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ is restricted to $\lambda_1^0 = 1$ across $\lambda_2^0 \in \{0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$ and similarly for $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$, thus providing a very broad coverage across a range of user inputs. The average-based statistic $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ and its adjusted version $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda_2^0)$ are implemented for $\tau_0 \in \{0.5, 0.8\}$ across $\lambda_2^0 \in \{0.50, 0.60, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00\}$.

All our size and power simulations below set $\pi_0 = 0.25$ (i.e., $k_0 = [T\ 0.25]$) to initiate the recursively generated forecasts.

5.1. DGP1

A specification that mimics a frequently encountered setting in the asset pricing literature is one where the null model is the martingale difference sequence,

$y_{t+1} = u_{t+1}$, and the larger model the single predictor-based predictive regression, $y_{t+1} = \beta x_t + u_{t+1}$ with $x_t = \phi_1 x_{t-1} + v_t$. Letting $\Sigma = \{\{\sigma_u^2, \rho_{uv}\sigma_u\sigma_v\}, \{\rho_{uv}\sigma_u\sigma_v, \sigma_v^2\}\}$ denote the covariance of (u_t, v_t), in line with commonly encountered magnitudes from the equity premium predictability literature we set $\sigma_u^2 = 3$, $\sigma_v^2 = 0.01$, $\rho_{uv} = -0.8$ and experiment with $\phi_1 \in \{0.75, 0.95, 0.98\}$. The conditionally homoskedastic setting takes $(u_t, v_t) \sim NID(0, \Sigma)$ while conditional heteroskedasticity is modeled via an ARCH(1) specification, writing $u_t = \epsilon_t \sqrt{h_t}$ with $h_t = \alpha_0 + \alpha_1 u_{t-1}^2$ and $\epsilon_t \sim NID(0,1)$. This latter choice naturally influences the magnitudes of σ_u^2 and ρ_{uv} chosen above and we parameterize $\{\alpha_0, \alpha_1\}$ in a way that maintains the same magnitude for σ_u^2 as in the conditionally homoskedastic case, i.e., $\alpha_0/(1-\alpha_1) = \sigma_u^2$. For this purpose, we set $(\alpha_0, \alpha_1) = (1.8, 0.4)$ throughout.

Size experiments set $\beta = 0$ while for the power properties of the tests we fix the sample size at $T = 500$ and evaluate correct decision frequencies as β moves away from the null with $\beta \in \{0, -1.5, -1.75, -2.0, -2.25, -2.5, -3, -3.5\}$. For $\beta = \gamma/T^{1/4}$, this is equivalent to $|\gamma|$ increasing with $\gamma \in \{0, -7.1, -8.3, -9.5, -10.6, -11.8, -14.2, -16.6\}$. Lastly, all of the above experiments are conducted using two alternative estimators for σ. The first one denoted $\hat{\sigma}_{hom}^2$ is suitable under conditional homoskedasticity while the second one denoted $\hat{\sigma}_{nw}^2$ is its robustified version à la Newey–West. Both estimators are based on the residuals from the model estimated under the alternative.

As our Monte-Carlo simulations encompass a very broad range of scenarios and test statistic parameterizations, we provide an extensive selection of outcomes in the Supplementary Material accompanying this paper. Our focus below is on a selection of key size/power results under conditional homoskedasticity and test statistic parameterizations that mainly rely on our recommendations based on our theoretical local power analysis above.

Empirical Size

Table 1 presents size estimates for $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ across a broad range of parameterizations in a conditionally homoskedastic setting. For both test statistics, we note good to excellent matches of the nominal size of 10% across almost all choices of λ_2^0 for $T \geq 500$. The adjusted statistic $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ in particular has empirical sizes that almost perfectly match 10% for virtually all magnitudes of λ_2^0. Under $\phi_1 = 0.75$, for instance, $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0 = 0.8)$ has resulted in an empirical size of 10.4% for $T = 1{,}000$ and 11.0% for $T = 500$. The corresponding figures for $\phi_1 = 0.95$ were 10.6% and 11.2%, respectively, thus also highlighting the robustness of the test statistics to the degree of persistence of the predictors as expected from our results in Propositions 1 and 2. Similar outcomes also characterize $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0 = 0.9)$ suggesting that these test statistics maintain good size control for $T \geq 500$ even when λ_2^0 is as large as 0.90 or 0.95 and λ_1^0 is set equal to one.

Regarding the size properties of the unadjusted $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ statistic, we note a mild undersizeness for magnitudes of λ_2^0 that are in the vicinity of 1, with its empirical sizes clustered around 7–8%. Overall the outcomes in Table 1 have

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Table 1. DGP1 empirical size of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity and 10% nominal size

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	
						$\phi = 0.75$						DM
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.080	0.086	0.086	0.084	0.086	0.084	0.077	0.080	0.080	0.072	0.006
	T=500	0.088	0.086	0.089	0.087	0.092	0.090	0.093	0.093	0.088	0.083	0.006
	T=1,000	0.098	0.098	0.099	0.089	0.088	0.089	0.090	0.093	0.096	0.086	0.007
												CW
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.099	0.106	0.106	0.107	0.110	0.110	0.109	0.116	0.120	0.126	0.055
	T=500	0.102	0.100	0.104	0.104	0.108	0.108	0.110	0.114	0.117	0.124	0.055
	T=1,000	0.106	0.107	0.108	0.102	0.101	0.103	0.104	0.109	0.117	0.111	0.055
						$\phi = 0.95$						DM
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.076	0.077	0.077	0.074	0.074	0.072	0.069	0.069	0.069	0.062	0.006
	T=500	0.083	0.084	0.084	0.083	0.086	0.085	0.081	0.078	0.075	0.075	0.008
	T=1,000	0.088	0.089	0.085	0.088	0.089	0.086	0.090	0.087	0.089	0.084	0.007
												CW
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.102	0.107	0.107	0.109	0.108	0.115	0.112	0.118	0.124	0.140	0.064
	T=500	0.098	0.102	0.103	0.104	0.108	0.108	0.108	0.108	0.110	0.124	0.059
	T=1,000	0.098	0.099	0.096	0.098	0.101	0.101	0.106	0.105	0.112	0.114	0.056
						$\phi = 0.98$						DM
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.078	0.079	0.078	0.077	0.075	0.073	0.073	0.071	0.074	0.068	0.011
	T=500	0.082	0.081	0.081	0.081	0.080	0.078	0.080	0.075	0.072	0.062	0.010
	T=1,000	0.085	0.086	0.090	0.088	0.087	0.090	0.088	0.086	0.084	0.073	0.007
												CW
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.115	0.117	0.119	0.122	0.121	0.128	0.128	0.133	0.150	0.170	0.084
	T=500	0.104	0.103	0.106	0.106	0.107	0.105	0.113	0.119	0.122	0.129	0.072
	T=1,000	0.098	0.100	0.104	0.103	0.102	0.104	0.108	0.110	0.113	0.109	0.060

highlighted remarkably stable size properties for both the unadjusted and adjusted statistics across the different magnitudes of λ_2^0 including when it is set at 0.90 or 0.95. This is particularly reassuring given our earlier theoretical power analysis which pointed at desirable parameterizations that satisfy $\lambda_1^0 \approx \lambda_2^0$ with both λ_1^0 and λ_2^0 set in the vicinity of 1 in the practical implementation of $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0)$.

Before proceeding further it is also useful to briefly rationalize the size behavior of these two statistics when λ_2^0 is chosen to lie almost at its boundary as when we set $\lambda_2^0 = 0.95$. In such instances we noted the mild undersizeness of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and mild oversizeness of $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ when operating with small to moderately sized samples. A magnitude of λ_2^0 that is close to 1 essentially translates into more "MSE content" from the larger model and hence a greater exposure to

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

estimation noise when the null model holds true. This results in the unadjusted $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ test statistic's distribution being pushed leftward with fewer than expected rejections of the null. On the other hand, the adjusted statistic which aims to correct for estimation error via (32) sees its correction factor's contribution increase as $\lambda_2^0 \to 1$, a correction factor that is overly inflated in small samples. At this stage, it is also useful to point out that the effective sample size is given by $T - k_0$ so that under $\pi_0 = 0.25$ and $T = 250$, we have only about 188 data points when implementing the tests. A highly persistent predictor combined with such a small sample size can be seen to result in some degree of oversizeness for $S_{T,adj}(\lambda_1^0, \lambda_2^0)$ based inferences when $|\lambda_1^0 - \lambda_2^0|$ is particularly small (e.g., for $(\lambda_1^0, \lambda_2^0) = (1, 0.95)$). Nevertheless, these finite sample distortions quickly fade away as we increase the sample size to $T = 500$.

For comparison purposes, the last column of Table 1 also includes the corresponding size estimates for the DM and CW statistics. These conform with the consensus view that the DM statistic is severely undersized under such nested settings while the CW statistics' empirical sizes are clustered around 5% for a nominal size of 10%, in line with the simulation results in Clark and West (2007).

We next consider the finite sample size properties of the average-based statistics $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda_2^0)$. Recall that the averaging is performed across a portion of the null model's MSE as captured by τ_0 and for a given fraction of the second model's MSE λ_2^0. Here, we present outcomes obtained under $\tau_0 = 0.8$ which only sum across the larger magnitudes of λ_1. Such a choice is theoretically justified by our earlier power analysis with further scenarios presented in the Supplementary Material. Results are presented in Table 2 from which we note that the adjusted statistic $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ displays good to excellent size control (e.g., empirical size estimates near 10% under $\lambda_2^0 = 1$) under moderate to large sample size choices.

An exception to this is when $\lambda_2^0 \approx 0.5\,\tau_0 + 0.5(\approx 0.9\ \textit{here})$ under which it shows a tendency to overreject the null hypothesis in smaller samples. This is in complete agreement with our earlier theoretical power analysis where we showed that holding all else constant the power of the test statistic must peak under $\lambda_2^0 = 0.5\tau_0 + 0.5$. Thus the empirical sizes peaking for λ_2^0 in the vicinity of $0.5(0.8) + 0.5 = 0.90$ highlight the size versus power trade-off that will characterize this average-based test statistic.

Regarding the unadjusted statistic $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$, we can note a tendency to underreject (e.g., empirical sizes in the vicinity of 7% under $\tau_0 = 0.8$) and this undersizeness deteriorating as $\lambda_2^0 \to 1$ and ϕ_1 gets closer to 1. This behavior conforms with the intuition that estimation noise caused by the estimation of parameters that are zero in the population pushes the test statistic too much to the left, a feature that was the key motivation behind Clark and West's adjustment to the DM statistic. Note, for instance, that these distortions are substantially dampened when the test statistic is implemented with smaller magnitudes of λ_2^0 for which it shows good to excellent size control.

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

TABLE 2. DGP1 empirical size of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional homoskedasticity and 10% nominal size

	λ_2^0	0.500	0.600	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
						$\phi = 0.75$					DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.085	0.088	0.086	0.082	0.075	0.076	0.061	0.042	0.047	0.006
	T=500	0.090	0.092	0.088	0.084	0.085	0.080	0.066	0.056	0.065	0.006
	T=1,000	0.097	0.097	0.089	0.089	0.086	0.083	0.071	0.073	0.073	0.007
											CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.104	0.111	0.117	0.120	0.124	0.151	0.154	0.121	0.110	0.055
	T=500	0.103	0.110	0.108	0.107	0.119	0.130	0.139	0.116	0.106	0.055
	T=1,000	0.108	0.108	0.103	0.106	0.108	0.118	0.121	0.113	0.099	0.055
						$\phi = 0.95$					DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.078	0.077	0.078	0.072	0.068	0.067	0.053	0.038	0.043	0.006
	T=500	0.083	0.085	0.087	0.085	0.079	0.072	0.055	0.046	0.054	0.008
	T=1,000	0.086	0.086	0.088	0.081	0.084	0.077	0.067	0.058	0.069	0.007
											CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.108	0.112	0.121	0.126	0.138	0.161	0.169	0.140	0.115	0.064
	T=500	0.102	0.106	0.114	0.118	0.122	0.134	0.139	0.118	0.104	0.059
	T=1,000	0.097	0.097	0.104	0.102	0.108	0.116	0.126	0.104	0.098	0.056
						$\phi = 0.98$					DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.079	0.082	0.074	0.072	0.070	0.067	0.057	0.043	0.039	0.011
	T=500	0.080	0.079	0.078	0.075	0.077	0.065	0.053	0.043	0.053	0.010
	T=1,000	0.083	0.088	0.088	0.082	0.084	0.074	0.061	0.059	0.067	0.007
											CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.119	0.127	0.131	0.144	0.157	0.189	0.206	0.174	0.143	0.084
	T=500	0.104	0.109	0.110	0.116	0.134	0.148	0.159	0.129	0.119	0.072
	T=1,000	0.097	0.102	0.107	0.104	0.115	0.127	0.130	0.108	0.101	0.060

Empirical Power

Table 3 presents empirical power estimates for $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ across $\lambda_2^0 \in \{0.80, 0.85, 0.90, 0.95\}$. The sample size is fixed at $T = 500$ and power is evaluated as the DGP moves further away from the null hypothesis. The choices of λ_2^0 are dictated by our theoretical results in Corollaries 1 and 2 which pointed to magnitudes satisfying $\lambda_1^0 \approx \lambda_2^0 \approx 1$. For both test statistics, we note the tendency of their empirical power to converge to 1 as $|\gamma|$ is allowed to increase. We can also clearly observe the particularly favorable impact that the degree of persistence of predictors has on power. As expected from our findings in Propositions 3 and 4 and their corollaries, power improves as $\lambda_2^0 \to 1$ and as $\phi_1 \to 1$.

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

TABLE 3. **DGP1** empirical power of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity

β	−1.500	−1.750	−2.000	−2.250	−2.500	−3.000	−3.500
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$				$\phi_1 = 0.75$			
$\lambda_2^0 = 0.80$	0.194	0.241	0.308	0.371	0.442	0.615	0.770
$\lambda_2^0 = 0.85$	0.213	0.272	0.349	0.427	0.519	0.685	0.831
$\lambda_2^0 = 0.90$	0.245	0.328	0.412	0.512	0.605	0.779	0.895
$\lambda_2^0 = 0.95$	0.317	0.425	0.542	0.653	0.741	0.877	0.956
DM	0.312	0.440	0.566	0.666	0.752	0.877	0.946
$\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.350	0.447	0.565	0.671	0.758	0.895	0.966
$\lambda_2^0 = 0.85$	0.399	0.511	0.635	0.732	0.815	0.933	0.983
$\lambda_2^0 = 0.90$	0.468	0.590	0.712	0.812	0.879	0.960	0.992
$\lambda_2^0 = 0.95$	0.579	0.708	0.819	0.897	0.941	0.986	0.998
CW	0.740	0.853	0.924	0.965	0.984	0.997	1.000
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$				$\phi_1 = 0.95$			
$\lambda_2^0 = 0.80$	0.615	0.747	0.843	0.917	0.952	0.990	0.997
$\lambda_2^0 = 0.85$	0.681	0.801	0.885	0.945	0.968	0.994	0.999
$\lambda_2^0 = 0.90$	0.753	0.860	0.924	0.967	0.984	0.998	0.999
$\lambda_2^0 = 0.95$	0.845	0.924	0.962	0.984	0.993	0.999	1.000
DM	0.857	0.925	0.959	0.979	0.989	0.998	0.999
$\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.850	0.925	0.965	0.988	0.995	1.000	1.000
$\lambda_2^0 = 0.85$	0.885	0.949	0.980	0.992	0.997	1.000	1.000
$\lambda_2^0 = 0.90$	0.919	0.966	0.986	0.996	0.999	1.000	1.000
$\lambda_2^0 = 0.95$	0.955	0.986	0.993	0.999	1.000	1.000	1.000
CW	0.985	0.996	0.998	1.000	1.000	1.000	1.000
$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$				$\phi_1 = 0.98$			
$\lambda_2^0 = 0.80$	0.853	0.927	0.964	0.984	0.992	0.999	1.000
$\lambda_2^0 = 0.85$	0.886	0.949	0.977	0.990	0.996	0.999	1.000
$\lambda_2^0 = 0.90$	0.921	0.966	0.986	0.994	0.998	1.000	1.000
$\lambda_2^0 = 0.95$	0.956	0.981	0.994	0.997	0.999	1.000	1.000
DM	0.956	0.982	0.992	0.997	0.998	1.000	1.000
$\mathcal{S}_{T,adj}^0(\lambda_1^0, \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.957	0.982	0.995	0.998	0.999	1.000	1.000
$\lambda_2^0 = 0.85$	0.968	0.988	0.997	0.999	1.000	1.000	1.000
$\lambda_2^0 = 0.90$	0.981	0.992	0.998	0.999	1.000	1.000	1.000
$\lambda_2^0 = 0.95$	0.991	0.996	0.999	1.000	1.000	1.000	1.000
CW	0.997	0.999	1.000	1.000	1.000	1.000	1.000

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Given the good overall size control displayed by $\mathcal{S}^0_{T,adj}(\lambda^0_1 = 1, \lambda^0_2)$, Table 3 clearly highlights the benefits of basing inferences on this adjusted version of our first test statistic calibrated to $\lambda^0_1 = 1$ and $\lambda^0_2 \approx 0.9$, also noting that its performance improves considerably as $\phi_1 \to 1$. For $\beta_2 = -2$, for instance, it displays power in the vicinity of 70%–80% under $\phi_1 = 0.75$ and 100% under $\phi_1 = 0.95$ or $\phi_1 = 0.98$. It is here important to relate our simulation outcomes with the theoretical results of Corollary 3 where we established an ARE of 2 for the adjusted statistic $\mathcal{S}^0_{T,adj}(\lambda^0_1, \lambda^0_2)$ relative to its unadjusted counterpart. This theoretical power enhancement is clearly supported by the empirical power estimates of Table 3. Compare for instance the empirical power of 51.2% for the unadjusted statistic under $\phi_1 = 0.75$ and $\beta = -2.25$ with 81.2% for its adjusted version, a power gain of 30 percentage points.

At this stage, it is also important to recall that our theoretical analysis based on the asymptotic relative efficiency of $\mathcal{S}^0_{T,adj}(\lambda^0_1, \lambda^0_2)$ versus $\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda^0_2)$ clearly pointed to the potentially superior power performance of the average-based statistic, for larger magnitudes of τ_0 in particular. This is clearly corroborated by the comparison between the power outcomes in Tables 3 and 4 with the latter presenting power outcomes for the $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda^0_2)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda^0_2)$ statistics. Focusing on the "optimal" choice of $\lambda^0_2 = 0.5\tau_0 + 0.5 = 0.90$ when $\tau_0 = 0.8$ we note from Table 4 that $\overline{\mathcal{S}}_{T,adj}(\boldsymbol{\tau}_0 = 0.8; \lambda^0_2 = 0.9)$ clearly dominates all configurations of $\mathcal{S}^0_{T,adj}(\lambda^0_1, \lambda^0_2)$ in terms of its power properties, typically resulting in relative power gains in excess of 10 percentage points.

Before proceeding further, it is also useful to comment on the power behavior of the DM and CW statistics in comparison to $\overline{\mathcal{S}}_T(\boldsymbol{\tau}_0 = 0.8; \lambda^0_2 = 0.90)$. Despite being severely undersized and theoretically unsuitable in the present nested context, we note that the DM statistic does show a reasonable ability to detect departures from the null. However, we can also observe that it is uniformly dominated by $\overline{\mathcal{S}}_{T,adj}(\boldsymbol{\tau}_0 = 0.8; \lambda^0_2 = 0.90)$ which under $\phi_1 = 0.75$, for instance, exceeds its power by about 10 percentage points. Comparing the power performance of $\overline{\mathcal{S}}_{T,adj}(\boldsymbol{\tau}_0 = 0.8; \lambda^0_2 = 0.90)$ with that of the CW statistic, we note that these two test statistics display very similar power outcomes across most scenarios. Although the CW statistic does not have a well-defined limiting distribution due to the nestedness of the competing models it appears to display reasonably good power properties across the DGPs we have considered, despite being far-off the standard normal distribution under the null (as implied by its size properties).

5.2. DGP2

The second DGP allows for multiple predictors and is calibrated to mimic US inflation-based predictive regressions as considered in Stock and Watson (2010) and Granziera, Hubrich, and Moon (2014). We use the same setting as in Granziera et al. (2014) and consider a DGP given by $y_{t+1} = \mu + \rho y_t + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + u_{t+1}$ with $\mu = 1$ and $\rho = 0.25$. The predictors $\boldsymbol{x}_t = (x_{1,t}, x_{2,t}, x_{3,t})'$ follow the VAR(1) process $\boldsymbol{x}_t = \Phi\ \boldsymbol{x}_{t-1} + \boldsymbol{v}_t$ with $\Phi = \{\{0.6, 0.1, 0\}, \{0.6, 0.25, 0\}, \{0, 0, 0.9\}\}$ thus

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

TABLE 4. DGP1 empirical power of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional homoskedasticity

β	−1.500	−1.750	−2.000	−2.250	−2.500	−3.000	−3.500
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$				$\phi_1 = 0.75$			
$\lambda_2^0 = 0.80$	0.272	0.345	0.450	0.548	0.635	0.800	0.911
$\lambda_2^0 = 0.85$	0.363	0.475	0.600	0.695	0.785	0.904	0.965
$\lambda_2^0 = 0.90$	0.449	0.567	0.685	0.782	0.853	0.942	0.980
$\lambda_2^0 = 0.95$	0.383	0.495	0.615	0.723	0.803	0.911	0.967
$\lambda_2^0 = 1$	0.305	0.397	0.498	0.602	0.695	0.840	0.926
DM	0.312	0.440	0.566	0.666	0.752	0.877	0.946
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.501	0.613	0.735	0.824	0.889	0.969	0.992
$\lambda_2^0 = 0.85$	0.632	0.747	0.849	0.914	0.955	0.990	0.998
$\lambda_2^0 = 0.90$	0.704	0.815	0.892	0.945	0.973	0.994	0.999
$\lambda_2^0 = 0.95$	0.642	0.763	0.853	0.920	0.955	0.990	0.998
$\lambda_2^0 = 1$	0.541	0.659	0.768	0.853	0.915	0.974	0.993
CW	0.740	0.853	0.924	0.965	0.984	0.997	1.000
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$				$\phi_1 = 0.95$			
$\lambda_2^0 = 0.80$	0.771	0.870	0.925	0.967	0.982	0.997	0.999
$\lambda_2^0 = 0.85$	0.865	0.934	0.969	0.987	0.994	0.999	1.000
$\lambda_2^0 = 0.90$	0.904	0.956	0.976	0.992	0.996	1.000	1.000
$\lambda_2^0 = 0.95$	0.876	0.936	0.966	0.986	0.993	0.999	1.000
$\lambda_2^0 = 1$	0.802	0.889	0.939	0.970	0.985	0.997	0.999
DM	0.857	0.925	0.959	0.979	0.989	0.998	0.999
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.922	0.970	0.989	0.996	0.999	1.000	1.000
$\lambda_2^0 = 0.85$	0.960	0.987	0.995	0.999	1.000	1.000	1.000
$\lambda_2^0 = 0.90$	0.974	0.990	0.997	0.999	1.000	1.000	1.000
$\lambda_2^0 = 0.95$	0.964	0.989	0.994	0.999	1.000	1.000	1.000
$\lambda_2^0 = 1$	0.941	0.975	0.989	0.997	0.999	1.000	1.000
CW	0.985	0.996	0.998	1.000	1.000	1.000	1.000
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$				$\phi_1 = 0.98$			
$\lambda_2^0 = 0.80$	0.925	0.966	0.985	0.994	0.997	1.000	1.000
$\lambda_2^0 = 0.85$	0.960	0.983	0.994	0.998	0.999	1.000	1.000
$\lambda_2^0 = 0.90$	0.972	0.990	0.996	0.999	0.999	1.000	1.000
$\lambda_2^0 = 0.95$	0.960	0.982	0.994	0.998	0.999	1.000	1.000
$\lambda_2^0 = 1$	0.933	0.970	0.986	0.995	0.998	1.000	1.000
DM	0.956	0.982	0.992	0.997	0.998	1.000	1.000

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

TABLE 4. (Continued)

β	−1.500	−1.750	−2.000	−2.250	−2.500	−3.000	−3.500
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$							
$\lambda_2^0 = 0.80$	0.978	0.992	0.997	1.000	1.000	1.000	1.000
$\lambda_2^0 = 0.85$	0.990	0.997	0.999	1.000	1.000	1.000	1.000
$\lambda_2^0 = 0.90$	0.994	0.998	1.000	1.000	1.000	1.000	1.000
$\lambda_2^0 = 0.95$	0.991	0.997	0.999	1.000	1.000	1.000	1.000
$\lambda_2^0 = 1$	0.984	0.995	0.998	1.000	1.000	1.000	1.000
CW	0.997	0.999	1.000	1.000	1.000	1.000	1.000

TABLE 5. DGP2 empirical size of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	
												DM
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.055	0.055	0.055	0.054	0.054	0.050	0.047	0.046	0.043	0.037	0.002
	T=500	0.066	0.065	0.065	0.063	0.065	0.061	0.060	0.055	0.051	0.044	0.002
	T=1,000	0.078	0.079	0.078	0.073	0.073	0.071	0.073	0.070	0.064	0.053	0.001
												CW
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.110	0.110	0.114	0.117	0.119	0.120	0.124	0.134	0.151	0.172	0.074
	T=500	0.099	0.099	0.103	0.105	0.109	0.110	0.114	0.118	0.123	0.139	0.065
	T=1,000	0.105	0.104	0.104	0.102	0.103	0.102	0.110	0.115	0.120	0.127	0.066

encompassing both persistent and much noisier processes while also being interdependent. The conditionally homoskedastic scenario takes $(u_t, v_{1,t}, v_{2,t}, v_{3,t})' \sim NID(0, I_4)$ while conditional heteroskedasticity is captured via an ARCH(1) process for u_t as in DGP1 with $\alpha_0 = 0.6$ and $\alpha_1 = 0.4$ so that its unconditional variance matches unity as in the homoskedastic scenario.

For our size experiments, we set $\beta_1 = \beta_2 = \beta_3 = 0$ and our power analysis focuses on alternatives to $\beta_1 = \beta_2 = \beta_3 = 0$ by fixing $(\beta_1, \beta_2, \beta_3) = (0.15, 0.15, -0.15)$ and evaluating rejection rates of the null hypothesis for $T = 250, 500, 1,000$.

Empirical Size

Tables 5 and 6 present empirical size estimates corresponding to the null DGP under $\beta_1 = \beta_2 = \beta_3 = 0$ for $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\overline{\mathcal{S}}_T(\tau_0 = 0.80; \lambda_2^0)$, respectively, together with their adjusted versions. As DGP2 contains a larger number of predictors than DGP1, we expect the impact of estimation error on the MSE of the second/larger model to be more pronounced under the null. This is indeed corroborated by the size estimates in Table 5 where we note the undersizeness of the unadjusted $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)$ statistic which is biased downward and thus results

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

TABLE 6. DGP2 empirical size of $\overline{\mathcal{S}}_T(\tau_0=0.8;\lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0=0.8;\lambda_2^0)$ under conditional homoskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
													DM
$\overline{\mathcal{S}}_T(\tau_0=0.8;\lambda_2^0)$	T=250	0.054	0.054	0.054	0.052	0.049	0.046	0.039	0.034	0.028	0.022	0.019	0.002
	T=500	0.066	0.066	0.066	0.061	0.064	0.058	0.052	0.039	0.028	0.023	0.027	0.002
	T=1,000	0.074	0.074	0.070	0.070	0.069	0.064	0.061	0.049	0.037	0.034	0.043	0.001
													CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0=0.8;\lambda_2^0)$	T=250	0.112	0.115	0.123	0.129	0.132	0.144	0.165	0.208	0.233	0.198	0.157	0.074
	T=500	0.104	0.107	0.112	0.110	0.118	0.121	0.133	0.158	0.180	0.154	0.124	0.065
	T=1,000	0.100	0.104	0.101	0.105	0.109	0.111	0.117	0.137	0.151	0.129	0.111	0.066

in too few rejections of the null (e.g., 6.4% under $T=1{,}000$ and $\lambda_2^0=0.9$ vs. a nominal size of 10%). Furthermore, its undersizeness tends to deteriorate for larger magnitudes of λ_2^0 as this translates into an increased influence of the larger model's MSE.

The adjusted version of the test statistic $\mathcal{S}^0_{T,adj}(\lambda_1^0,\lambda_2^0)$ on the other hand is quite effective in adjusting for estimation noise (e.g., the earlier empirical size of 6.4% is now pushed up to 12%), while also showing a tendency to "over-adjust" in small to moderately size samples, in particular, for larger magnitudes of λ_2^0. Table 6 presents the corresponding size estimates for the average-based statistic $\overline{\mathcal{S}}_{T,adj}(\boldsymbol{\tau}_0=0.8;\lambda_2^0)$. We note that this latter statistic maintains good to excellent size control for moderate sample sizes across all magnitudes of λ_2^0 but requires larger samples when λ_2^0 is set near one.

Empirical Power

For this DGP, our power experiments focus on documenting the rejection frequencies of the null hypothesis for a fixed alternative as the sample size is allowed to increase. Results are presented in Tables 7 and 8 for $\mathcal{S}^0_T(\lambda_1^0,\lambda_2^0)$ and $\overline{\mathcal{S}}_T(\boldsymbol{\tau}_0=0.8;\lambda_2^0)$ and their adjusted versions. Under either $T=500$ or $T=1{,}000$, all test statistics have powers at or near 100%.

Tables 7 and 8 also clearly corroborate our theoretical power analysis that highlighted peaking powers under $\lambda_2^0=0.5\tau_0+0.5$. Focusing on $\overline{\mathcal{S}}_T(\boldsymbol{\tau}_0=0.8;\lambda_2^0)$, we can clearly observe the empirical powers to be largest under $\lambda_2^0=0.9$ for all sample sizes (e.g., 98.2% vs. 95.7% when $\lambda_2^0=1$ and for $T=250$).

The main findings from our simulation experiments can be summarized as follows: (i) The adjusted versions of the two test statistics $\mathcal{S}^0_{T,adj}(\lambda_1^0,\lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\boldsymbol{\tau}_0;\lambda_2^0)$ displayed good to excellent size control across most parameterizations of their respective inputs $(\lambda_1^0,\lambda_2^0)$ and (τ_0,λ_2^0). (ii) Both test statistics are consistent and have nontrivial local asymptotic power while their finite sample power properties are strongly influenced by the respective magnitudes of

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

TABLE 7. DGP2 empirical power of $\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$ and $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$ under conditional homoskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	
												DM
$\mathcal{S}_T^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.520	0.572	0.620	0.668	0.724	0.778	0.835	0.885	0.929	0.971	0.884
	T=500	0.784	0.836	0.877	0.914	0.945	0.967	0.986	0.994	0.998	1.000	0.998
	T=1,000	0.960	0.976	0.987	0.995	0.998	0.999	1.000	1.000	1.000	1.000	1.000
												CW
$\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0)$	T=250	0.912	0.936	0.956	0.969	0.979	0.987	0.993	0.997	0.999	1.000	1.000
	T=500	0.991	0.996	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	T=1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 8. DGP2 empirical power of $\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$ under conditional homoskedasticity

	λ_2^0	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000	
													DM
$\overline{\mathcal{S}}_T(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.557	0.619	0.677	0.739	0.805	0.871	0.933	0.970	0.982	0.975	0.957	0.884
	T=500	0.821	0.874	0.918	0.951	0.974	0.991	0.998	1.000	1.000	1.000	0.999	0.998
	T=1,000	0.974	0.988	0.994	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
													CW
$\overline{\mathcal{S}}_{T,adj}(\tau_0 = 0.8; \lambda_2^0)$	T=250	0.930	0.953	0.970	0.982	0.991	0.995	0.998	1.000	1.000	1.000	1.000	1.000
	T=500	0.995	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	T=1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

those same inputs. The guidelines provided by our theoretical power analysis do however lead to highly favorable power outcomes with implementations such as $\mathcal{S}_{T,adj}^0(\lambda_1^0 = 1, \lambda_2^0 \approx 0.9)$ and $\overline{\mathcal{S}}_{T,adj}(\boldsymbol{\tau}_0 \approx 0.8; \lambda_2^0 \approx 0.9)$ standing out in terms of their size/power tradeoffs, especially for moderately sized samples such as $T \geq 500$. (iii) The proposed methods are valid irrespective of the degree of persistence of the predictors as also corroborated by our finite sample simulations. (iv) Our Monte-Carlo analysis did show that Clark and West's CW statistic which although not grounded on formal standard normal asymptotics also performed particularly well in terms of its power, despite relatively important size distortions. Strictly speaking the CW statistic has been introduced for handling nested models estimated via a rolling as opposed to a recursive approach since from a theoretical standpoint it continues to suffer from the variance degeneracy problem characterizing DM-type constructions.

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

5.3. Summary and Tuning-Parameter Guidelines

The above simulation-based outcomes combined with our earlier local power analysis point to precise guidelines for the choice of tuning parameters required in the implementation of our test statistics. For the $\mathcal{S}^0_{T,adj}(\lambda^0_1,\lambda^0_2)$ statistic, we argued that λ^0_1 and λ^0_2 should be set near their boundary of one and in close vicinity of one another. Our simulations based on $\mathcal{S}^0_{T,adj}(\lambda^0_1=1,\lambda^0_2)$ for λ^0_2 set in the 0.80–0.95 range have indeed resulted in good to excellent size–power tradeoffs and good to excellent size control. The robustness of the empirical size outcomes to a much broader range of the λ^0_2 magnitudes is also noteworthy as illustrated by the outcomes in Tables 1 and 5.

Regarding the $\overline{\mathcal{S}}_{T,adj}(\tau_0;\lambda^0_2)$ statistic, our theoretical local power analysis led us to argue for τ_0 to be set in the vicinity of unity and λ^0_2 as $\lambda^0_2=0.5\ \tau_0+0.5$. Simulations based on $\tau_0=0.8$ did indeed result in good to excellent size and power properties. As the above choice for λ^0_2 is a maximizer of local power it is perhaps natural to expect some size distortions in smaller samples when λ^0_2 is set in this way and, in particular, when this is combined with the presence of highly persistent predictors. This is indeed confirmed by our size experiments implemented under $(\tau_0;\lambda^0_2)=(0.8,0.9)$ and a local-to-unity-type predictor. Nevertheless, these specific finite sample distortions can also be seen to progressively vanish as the sample size is allowed to grow.

6. APPLICATION

We illustrate the implementation of our proposed methods by revisiting a widely considered puzzle in the international economics literature, namely, the random walk like behavior of exchange rates. Our goal is to use our new test statistics in order to evaluate whether past exchange rate levels have any predictive power for subsequent exchange rate changes. Letting s_t denote the log of the spot exchange rate, we compare the out of sample predictive accuracy of the larger model $\Delta s_{t+1}=\alpha+\beta\ s_t+u_{t+1}$ (model 2) with the random walk with drift specification $\Delta s_{t+1}=\alpha+u_{t+1}$ (model 1).

We consider six major currencies (EUR, YEN, GBP, CHF, AUD, and CAD) and implement our tests on daily spot rates spanning the period between 4 January 1999 and 16 July 2021, sourced from the Saint-Louis Fred database. An important advantage of the methods developed in this paper is their robustness to the persistence properties of predictors which is particularly relevant when considering exchange rate series. Indeed, for all three daily series, we have considered the first-order autocorrelation coefficient from an AR(1) fit is 0.99.

Predictive accuracy testing outcomes (p-values) based on $\mathcal{S}^0_{T,adj}(\lambda^0_1,\lambda^0_2)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0;\lambda^0_2)$ are presented in Table 9 where we used $\pi_0=0.5$ to initiate the expanding window estimation (i.e., starting from the middle of the sample). For robustness considerations, inferences based on $\mathcal{S}_{T,adj}(\lambda^0_1=1,\lambda^0_2)$ are implemented across $\lambda^0_2=\{0.80,0.85,0.90,0.95\}$ while for $\overline{\mathcal{S}}_{T,adj}(\tau_0;\lambda^0_2=1)$ we consider $\tau_0=$

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Table 9. Exchange rate predictability

	EURUSD	YENUSD	GBPUSD	CHFUSD	AUDUSD	CADUSD
$\mathcal{S}^0_{T,adj}(\lambda^0_1 = 1, \lambda^0_2)_{nw}$						
$\lambda^0_2 = 0.80$	1.000	1.000	0.360	0.924	0.764	0.997
$\lambda^0_2 = 0.85$	1.000	0.998	0.219	0.874	0.130	0.862
$\lambda^0_2 = 0.90$	1.000	1.000	0.731	0.838	0.846	0.989
$\lambda^0_2 = 0.95$	0.997	0.997	0.814	0.755	0.711	0.945
$\overline{\mathcal{S}}_{T,adj}(\tau_0; \lambda^0_2)_{nw}$						
$\tau_0 = 0.80$	0.640	0.928	0.957	0.508	0.993	0.954
$\tau_0 = 0.90$	0.471	0.607	0.695	0.500	0.216	0.374
DM_{nw}	0.091	0.438	0.530	0.643	0.276	0.096
CW_{nw}	0.048	0.248	0.299	0.532	0.155	0.044

0.80 and $\tau_0 = 0.90$. Looking at the top and middle panels of Table 9, we note that results unanimously corroborate the fact that the level of exchange rates does not have any meaningful forecasting power for future currency returns over the period considered and across all major currencies. This result based on the use of daily data also corroborates the recent findings in Engel and Wu (2021) based on monthly frequencies. The bottom panel of Table 9 displays the p-values associated with the standard DM and the CW statistics. It is here interesting to point out that inferences based on the standard CW statistic lead to a rejection of the random walk specification for the EURUSD and CADUSD series when implementing the test at a 5% level or above which is in sharp contrast with the large p-values obtained using our two test statistics.

7. CONCLUDING REMARKS

The main motivation of this paper was to provide a way of bypassing the variance degeneracy problem that arises in the context of out-of-sample nested model comparisons. We did so by developing two new test statistics shown to have nuisance parameter-free standard normal asymptotics and good power properties including in the close vicinity of the null hypothesis. Our proposed inferences can trivially accommodate conditional heteroskedasticity and are also shown to be robust to the presence of highly persistent predictors. Although our power analysis has ruled out the case of deterministic trends via Assumption B1, for instance, these can also be easily accommodated within our framework without any changes to the implementation of the tests provided that the trends are formulated in a scaled form as (t/T) and its powers. Nested comparisons in purely deterministic environments would be particularly relevant in areas such as temperature modeling (e.g., Wu and Zhao, 2007; Gades-Riva and Gonzalo, 2020) or the recent literature on modeling pandemic dynamics (e.g., Jiang, Zhao, and Shao, 2020; Li and Linton, 2021).

Although our proposed test statistics require two user inputs each, both have been shown to display good to excellent size control across a very broad range of such parameterizations in a multitude of empirically relevant settings. Although these user inputs do have considerable influence on the finite sample power properties of both test statistics their choices can be accurately guided by examining the power functions associated with each test statistic, as demonstrated by our simulations and their theoretical backing.

It is important to recognize that our proposed test statistics do involve the discarding of some information albeit very limited and with its amount under the control of the user. As a result, some power loss is of course unavoidable but the absence of any alternative approach that uses more information while achieving the same purpose in an environment that can accommodate both stationary and persistent predictors as well as conditional heteroskedasticity makes such power losses only notional. Our simulation results have indeed shown that very little needs to be discarded for our methods to work well and to provide reliable inferences. In this sense, they are not subject to the disadvantages of sample splitting-based techniques used for instance in the goodness of fit literature (e.g., half-sample methods).

The principles underlying our proposed inferences based on (6) should also be portable beyond out of sample forecasting considerations to areas involving model selection testing à la Vuong (1989) where nestedness versus non-nestedness or the overlapping nature of models being compared influences test procedures due to variance degeneracy problems (see also Shi, 2015). In Schennah and Wilhelm (2017), for instance, the authors developed a model selection test for choosing between two parametric likelihoods based on sample splitting principles which although different from our approach based on MSE comparisons on overlapping intervals was driven by similar concerns. Adapting the analysis of this paper to such model selection testing contexts is a promising avenue currently being explored.

Appendix: Proofs

Proof of Proposition 1. We consider the asymptotic behavior of $Z_T(\ell_1, \ell_2)$ in (6). Rescaling the time axis, we write $Z_T(\lambda_1, \lambda_2) \equiv Z_T([(T-k_0)\lambda_1], [(T-k_0)\lambda_2])$ and focus on $Z_T(\lambda_1, \lambda_2)$. Using $\hat{e}^2_{j,t+1} = u^2_{t+1} + (\hat{e}^2_{j,t+1} - u^2_{t+1})$ $(j = 1, 2)$ in (6) yields

$$\begin{aligned} Z_T(\lambda_1, \lambda_2) &= \frac{T-k_0}{[(T-k_0)\lambda_1]} \left(\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1]} u^2_{t+1}}{\sqrt{T-k_0}} - \frac{[(T-k_0)\lambda_1]}{[(T-k_0)\lambda_2]} \frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2]} u^2_{t+1}}{\sqrt{T-k_0}} \right) \\ &+ \frac{T-k_0}{[(T-k_0)\lambda_1]} \frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1]} (\hat{e}^2_{1,t+1} - u^2_{t+1})}{\sqrt{T-k_0}} \\ &- \frac{T-k_0}{[(T-k_0)\lambda_2]} \frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2]} (\hat{e}^2_{2,t+1} - u^2_{t+1})}{\sqrt{T-k_0}} \end{aligned}$$

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

$$\equiv \frac{T-k_0}{[(T-k_0)\lambda_1]}\,\mathcal{N}_{1T}(\lambda_1,\lambda_2)+\frac{T-k_0}{[(T-k_0)\lambda_1]}\,\mathcal{N}_{2T}(\lambda_1)-\frac{T-k_0}{[(T-k_0)\lambda_2]}\,\mathcal{N}_{3T}(\lambda_2). \tag{A.1}$$

From Assumption A(i), we have $\sup_{\lambda_1}|\mathcal{N}_{2T}(\lambda_1)|=o_p(1)$ and $\sup_{\lambda_2}|\mathcal{N}_{3T}(\lambda_2)|=o_p(1)$. Combining with

$$\sup_{\lambda_1,\lambda_2}\left|\frac{[(T-k_0)\lambda_1]}{[(T-k_0)\lambda_2]}-\frac{\lambda_1}{\lambda_2}\right|=O(1/(T-k_0)) \tag{A.2}$$

and

$$\sup_{\lambda_j}\left|\frac{(T-k_0)}{[(T-k_0)\lambda_j]}-\frac{1}{\lambda_j}\right|=O(1/(T-k_0)) \quad j=1,2 \tag{A.3}$$

gives

$$Z_T(\lambda_1,\lambda_2)=\frac{1}{\lambda_1}\left(\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1]}u_{t+1}^2}{\sqrt{T-k_0}}-\frac{\lambda_1}{\lambda_2}\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2]}u_{t+1}^2}{\sqrt{T-k_0}}\right)+o_p(1). \tag{A.4}$$

It is now convenient to reformulate (A.4) as

$$\begin{aligned}Z_T(\lambda_1,\lambda_2)&=\frac{1}{\lambda_1}\left(\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1]}(u_{t+1}^2-\sigma_u^2)}{\sqrt{T-k_0}}-\frac{\lambda_1}{\lambda_2}\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2]}(u_{t+1}^2-\sigma_u^2)}{\sqrt{T-k_0}}\right)\\&\quad+\sigma_u^2\sqrt{T-k_0}\left(\frac{[(T-k_0)\lambda_1]}{(T-k_0)\lambda_1}-\frac{[(T-k_0)\lambda_2]}{(T-k_0)\lambda_2}\right)+o_p(1),\end{aligned} \tag{A.5}$$

and note that the second component in the right-hand side of (A.5) is $O(1/\sqrt{T-k_0})$. We now recall that our setting operates under fixed and given magnitudes of (λ_1,λ_2), say $(\lambda_1^0,\lambda_2^0)$ chosen such that $(\lambda_1^0,\lambda_2^0)\in\Lambda^0$. We have

$$\begin{aligned}&Z_T(\lambda_1^0,\lambda_2^0)\\&=\frac{1}{\lambda_1^0}\left(\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1^0]}(u_{t+1}^2-\sigma_u^2)}{\sqrt{T-k_0}}-\frac{\lambda_1^0}{\lambda_2^0}\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2^0]}(u_{t+1}^2-\sigma_u^2)}{\sqrt{T-k_0}}\right)+o_p(1).\end{aligned} \tag{A.6}$$

It follows from Assumption A(ii) and (iii), the continuous mapping theorem and Slutsky's theorem that

$$Z_T^0(\lambda_1^0,\lambda_2^0)\equiv\frac{Z_T(\lambda_1^0,\lambda_2^0)}{\hat{\sigma}}\xrightarrow{\mathcal{D}}\frac{1}{\lambda_1^0}\left(W_\eta(\lambda_1^0)-\frac{\lambda_1^0}{\lambda_2^0}\,W_\eta(\lambda_2^0)\right). \tag{A.7}$$

The right-hand side of (A.7) is a centered Gaussian random variable with variance $v^0(\lambda_1^0,\lambda_2^0)=|\lambda_1^0-\lambda_2^0|/\lambda_1^0\lambda_2^0$ as stated in (11). Specifically, the statement in (A.7) is equivalent to $Z_T^0(\lambda_1^0,\lambda_2^0)\xrightarrow{\mathcal{D}}N(0,|\lambda_1^0-\lambda_2^0|/\lambda_1^0\lambda_2^0)$. This also establishes that $\mathcal{S}_T^0(\lambda_1^0,\lambda_2^0)\equiv Z_T^0(\lambda_1^0,\lambda_2^0)/\sqrt{v^0(\lambda_1^0,\lambda_2^0)}\xrightarrow{\mathcal{D}}N(0,1)$. □

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Proof of Proposition 2. We view $Z_T(\lambda_1, \lambda_2)$ in (A.5) as a functional of λ_1 whose range is determined by the choice of τ_0, and for a given $\lambda_2 = \lambda_2^0$, satisfying $(\tau_0, \lambda_2^0) \in \overline{\Lambda}^0$. Assumption A(ii) combined with standard continuous mapping arguments applied to (A.5) yields

$$Z_T(\lambda_1; \lambda_2^0) \xrightarrow{\mathcal{D}} \sigma \frac{1}{\lambda_1} \left(W_\eta(\lambda_1) - \frac{\lambda_1}{\lambda_2^0} W_\eta(\lambda_2^0) \right). \tag{A.8}$$

The asymptotic behavior of $\overline{Z}_T(\tau_0; \lambda_2^0)$ in (9) now follows by appealing to Assumption A(i)–(iii), the continuity of the average operation and (A.8). Specifically,

$$\overline{Z}_T(\tau_0; \lambda_2^0) \xrightarrow{\mathcal{D}} \frac{1}{1-\tau_0} \int_{\tau_0}^1 \left[\frac{W_\eta(\lambda_1)}{\lambda_1} - \frac{W_\eta(\lambda_2^0)}{\lambda_2^0} \right] d\lambda_1. \tag{A.9}$$

Note that

$$E\left| \int_{\tau_0}^1 \frac{W_\eta(\lambda_1)}{\lambda_1} d\lambda_1 \right| \leq \int_{\tau_0}^1 E\left| \frac{W_\eta(\lambda_1)}{\lambda_1} \right| d\lambda_1 = \int_{\tau_0}^1 \frac{E|W_\eta(1)|}{\sqrt{\lambda_1}} d\lambda_1 < \infty, \tag{A.10}$$

so that (A.9) is well-defined almost surely and by construction centered Gaussian. It now suffices to obtain its variance. We have

$$\begin{aligned}
&\frac{1}{(1-\tau_0)^2} Cov\left[\int_{\tau_0}^1 \left(\frac{W_\eta(s_1)}{s_1} - \frac{W_\eta(\lambda_2^0)}{\lambda_2^0} \right) ds_1, \int_{\tau_0}^1 \left(\frac{W_\eta(s_2)}{s_2} - \frac{W_\eta(\lambda_2^0)}{\lambda_2^0} \right) ds_2 \right] \\
&= \frac{1}{(1-\tau_0)^2} \int_{\tau_0}^1 \left[\int_{\tau_0}^1 Cov\left[\left(\frac{W_\eta(s_1)}{s_1} - \frac{W_\eta(\lambda_2^0)}{\lambda_2^0} \right), \left(\frac{W_\eta(s_2)}{s_2} - \frac{W_\eta(\lambda_2^0)}{\lambda_2^0} \right) \right] ds_2 \right] ds_1 \\
&= \frac{1}{(1-\tau_0)^2} \int_{\tau_0}^1 \int_{\tau_0}^1 \left[\frac{s_1 \wedge s_2}{s_1 s_2} - \frac{s_1 \wedge \lambda_2^0}{s_1 \lambda_2^0} - \frac{\lambda_2^0 \wedge s_2}{s_2 \lambda_2^0} + \frac{1}{\lambda_2^0} \right] ds_1\, ds_2,
\end{aligned} \tag{A.11}$$

where we appealed to Fubini's theorem for interchanging expectations with integration in the second row of (A.11). Standard integral calculus now leads to (13) and (14). □

The following lemma collects some key results used in the proofs of Proposition 3 and Corollary 1 on the power properties of the proposed tests under stationarity.

LEMMA A1. *Suppose model (2) holds with $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{1/4}$. Under Assumption B1 and as $T \to \infty$, we have*

(i)

$$\sup_{\lambda} \left| \frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]} (\hat{e}_{1,t+1}^2 - u_{t+1}^2)}{\sqrt{T-k_0}} - \lambda\sqrt{1-\pi_0}\, \boldsymbol{\gamma}'(\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\gamma} \right| = o_p(1), \tag{A.12}$$

(ii)

$$\sup_{\lambda} \left| \frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]} (\hat{e}_{2,t+1}^2 - u_{t+1}^2)}{\sqrt{T-k_0}} \right| = o_p(1). \tag{A.13}$$

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Proof of Lemma A1. (i) As we operate under model (2) with $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{1/4}$, we have $\hat{e}_{1,t+1} - u_{t+1} = \boldsymbol{x'}_{2,t}\boldsymbol{\beta}_2 - \boldsymbol{x'}_{1,t}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1)$ so that the following identity holds:

$$\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{1,t+1}^2 - u_{t+1}^2)}{\sqrt{T-k_0}} = A_{1T}(\lambda) + A_{2T}(\lambda) - 2A_{3T}(\lambda) + 2A_{4T}(\lambda) - 2A_{5T}(\lambda), \tag{A.14}$$

where

$$A_{1T}(\lambda) = T^{-\frac{1}{2}}(T-k_0)^{-\frac{1}{2}}\boldsymbol{\gamma}'\sum_t \boldsymbol{x}_{2,t}\boldsymbol{x}_{2,t}'\boldsymbol{\gamma},$$

$$A_{2T}(\lambda) = (T-k_0)^{-\frac{1}{2}}\sum_t(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1)'\boldsymbol{x}_{1,t}\boldsymbol{x}_{1,t}'(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1),$$

$$A_{3T}(\lambda) = T^{-\frac{1}{4}}(T-k_0)^{-\frac{1}{2}}\sum_t(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1)'\boldsymbol{x}_{1,t}\boldsymbol{x}_{2,t}'\boldsymbol{\gamma},$$

$$A_{4T}(\lambda) = T^{-\frac{1}{4}}(T-k_0)^{-\frac{1}{2}}\boldsymbol{\gamma'}\sum_t \boldsymbol{x}_{2,t}\boldsymbol{u}_{t+1},$$

$$A_{5T}(\lambda) = (T-k_0)^{-\frac{1}{2}}\sum_t(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1)'\boldsymbol{x}_{1,t}u_{t+1},$$

with $t = k_0, \ldots, k_0 - 1 + [(T-k_0)\lambda]$ in all of the above summations and below, unless otherwise indicated. □

For $A_{1T}(\lambda)$, we write

$$A_{1T}(\lambda) = \sqrt{\frac{T-k_0}{T}}\boldsymbol{\gamma}'\left(\frac{\sum_t \boldsymbol{x}_{2,t}\boldsymbol{x}_{2,t}'}{T-k_0} - \lambda \boldsymbol{Q}_{22}\right)\boldsymbol{\gamma} + \sqrt{\frac{T-k_0}{T}}\lambda\boldsymbol{\gamma}'\boldsymbol{Q}_{22}\boldsymbol{\gamma} \tag{A.15}$$

and, as $|\sqrt{(T-[T\pi_0])/T} - \sqrt{1-\pi_0}| = o(1)$, we have

$$\left|A_{1T}(\lambda) - \lambda\sqrt{1-\pi_0}\,\boldsymbol{\gamma}'\boldsymbol{Q}_{22}\boldsymbol{\gamma}\right| \leq \sqrt{1-\pi_0}\,\|\boldsymbol{\gamma}\|^2\left\|\frac{\sum_t \boldsymbol{x}_{2,t}\boldsymbol{x}_{2,t}'}{T-k_0} - \lambda\,\boldsymbol{Q}_{22}\right\|, \tag{A.16}$$

so that Assumption B1(i) directly implies

$$\sup_\lambda\left|A_{1T}(\lambda) - \lambda\sqrt{1-\pi_0}\,\boldsymbol{\gamma}'\boldsymbol{Q}_{22}\boldsymbol{\gamma}\right| = o_p(1). \tag{A.17}$$

Before focusing on the remainder quantities, we consider the limiting behavior of $(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1)$. Setting $t = [Ts]$, we write

$$T^{1/4}(\hat{\boldsymbol{\delta}}_{1,[Ts]} - \boldsymbol{\beta}_1) = \left(\frac{\sum_{j=1}^{[Ts]}\boldsymbol{x}_{1,j-1}\boldsymbol{x}_{1,j-1}'}{T}\right)^{-1}\left(\frac{\sum_{j=1}^{[Ts]}\boldsymbol{x}_{1,j-1}\boldsymbol{x}_{2,j-1}'}{T}\right)\boldsymbol{\gamma}$$
$$+T^{-1/4}\left(\frac{\sum_{j=1}^{[Ts]}\boldsymbol{x}_{1,j-1}\boldsymbol{x}_{1,j-1}'}{T}\right)^{-1}\left(\frac{\sum_{j=1}^{[Ts]}\boldsymbol{x}_{1,j-1}u_j}{\sqrt{T}}\right). \tag{A.18}$$

For the second term in the right-hand side of (A.18), we have

$$\sup_s \frac{1}{T^{1/4}} \left\| \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{1,j-1}}{T} \right)^{-1} \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}u_j}{\sqrt{T}} \right) \right\| = o_p(1) \tag{A.19}$$

due to Assumption B1(i) and (ii). For the first term in the right-hand side of (A.18), we can write

$$\begin{aligned} &\left\| \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{1,j-1}}{T} \right)^{-1} \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{2,j-1}}{T} \right) \boldsymbol{\gamma} - \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma} \right\| \\ \leq &\left\| \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{1,j-1}}{T} \right)^{-1} - \boldsymbol{Q}_{11}^{-1} \right\| \left(\left\| \frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{2,j-1}}{T} \boldsymbol{\gamma} - \boldsymbol{Q}_{12}\boldsymbol{\gamma} \right\| + \| \boldsymbol{Q}_{12}\boldsymbol{\gamma} \| \right) \\ &+ \left\| \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{1,j-1}}{T} \right)^{-1} \right\| \left\| \frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{2,j-1}}{T} \boldsymbol{\gamma} - \boldsymbol{Q}_{12}\boldsymbol{\gamma} \right\|, \end{aligned} \tag{A.20}$$

so that Assumption B1(i) and (ii) also ensure that

$$\sup_s \left\| \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{1,j-1}}{T} \right)^{-1} \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1}\boldsymbol{x}'_{2,j-1}}{T} \right) \boldsymbol{\gamma} - \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma} \right\| = o_p(1). \tag{A.21}$$

Combining (A.19) and (A.21) and using the triangle inequality in (A.18) yields

$$\sup_s \left\| T^{1/4}(\hat{\boldsymbol{\delta}}_{1,[Ts]} - \boldsymbol{\beta}_1) - \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma} \right\| = o_p(1). \tag{A.22}$$

We now focus on $A_{2T}(\lambda)$. Using suitable normalizations and appealing to (A.22), we can express $A_{2T}(\lambda)$ as

$$\begin{aligned} A_{2T}(\lambda) &= \sqrt{\frac{T-k_0}{T}}(T-k_0)^{-1} \sum_t T^{1/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1)'\boldsymbol{x}_{1,t}\boldsymbol{x}'_{1,t} T^{1/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1) \\ &= \sqrt{\frac{T-k_0}{T}}(T-k_0)^{-1} \sum_t (T^{1/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1) - \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma})'\boldsymbol{x}_{1,t}\boldsymbol{x}'_{1,t} T^{1/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1) \\ &\quad + \sqrt{\frac{T-k_0}{T}}(T-k_0)^{-1} \boldsymbol{\gamma}'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} \sum_t \boldsymbol{x}_{1,t}\boldsymbol{x}'_{1,t}(T^{1/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1) - \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma}) \\ &\quad + \sqrt{\frac{T-k_0}{T}} \boldsymbol{\gamma}'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1} \left(\frac{\sum_t \boldsymbol{x}_{1,t}\boldsymbol{x}'_{1,t}}{T-k_0} - \lambda \boldsymbol{Q}_{11} \right) \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma} \\ &\quad + \lambda \sqrt{\frac{T-k_0}{T}} \boldsymbol{\gamma}'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma}. \end{aligned} \tag{A.23}$$

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Assumption B1(i) combined with the result in (A.22) give

$$\sup_{\lambda}\left|A_{2T}(\lambda)-\lambda\sqrt{1-\pi_0}\boldsymbol{\gamma}'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma}\right|=o_p(1). \tag{A.24}$$

For $A_{3T}(\lambda)$, we write

$$\begin{aligned}A_{3T}(\lambda)=&\sqrt{\frac{T-k_0}{T}}(T-k_0)^{-1}\sum(T^{1/4}(\hat{\boldsymbol{\delta}}_{1,t}-\boldsymbol{\beta}_1)-\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma})'\boldsymbol{x}_{1,t}\boldsymbol{x}_{2,t}'\boldsymbol{\gamma}\\&+\sqrt{\frac{T-k_0}{T}}\boldsymbol{\gamma}'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\left(\frac{\sum\boldsymbol{x}_{1,t}\boldsymbol{x}_{2,t}'}{T-k_0}-\lambda\boldsymbol{Q}_{12}\right)\boldsymbol{\gamma}+\sqrt{\frac{T-k_0}{T}}\lambda\boldsymbol{\gamma}'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma},\end{aligned} \tag{A.25}$$

so that using (A.22), Assumption B1(i) and the triangle inequality yields

$$\sup_{\lambda}\left|A_{3T}(\lambda)-\lambda\sqrt{1-\pi_0}\boldsymbol{\gamma}'\boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}\boldsymbol{\gamma}\right|=o_p(1). \tag{A.26}$$

Next, as an immediate consequence of Assumption B1(ii), we have

$$\sup_{\lambda}|A_{4T}(\lambda)|=o_p(1). \tag{A.27}$$

Finally, using (A.22) together with Assumption B1(ii) yields

$$\sup_{\lambda}|A_{5T}(\lambda)|=o_p(1). \tag{A.28}$$

Combining (A.17), (A.24), and (A.26)–(A.28) with successive uses of the triangle inequality yields the stated result in Lemma A1(i). The statement in (A.13) follows an identical line of argument as above and details are therefore omitted from the exposition here.

Proof of Proposition 3. (i) We initially consider the case of a fixed and nonzero $\boldsymbol{\beta}_2$ and establish that $\mathcal{S}_T^0(\lambda_1^0,\lambda_2^0)\xrightarrow{p}\infty$. Using (A.1) and appealing to Lemma A1(ii), we have

$$\frac{Z_T(\lambda_1^0,\lambda_2^0)}{\sqrt{T-k_0}}=\frac{1}{\lambda_1^0}\frac{\mathcal{N}_{1T}(\lambda_1^0,\lambda_2^0)}{\sqrt{T-k_0}}+\frac{1}{\lambda_1^0}\frac{\mathcal{N}_{2T}(\lambda_1^0)}{\sqrt{T-k_0}}+o_p(1). \tag{A.29}$$

We can now note from (A.7) and (A.8) that the first term in the right-hand side of (A.29) is $O_p(T^{-1/2})$ so that

$$\frac{Z_T(\lambda_1^0,\lambda_2^0)}{\sqrt{T-k_0}}=\frac{1}{\lambda_1^0}\left[\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1^0]}(\hat{e}_{1,t+1}^2-u_{t+1}^2)}{T-k_0}\right]+o_p(1). \tag{A.30}$$

It is now straightforward to adapt the result in Lemma A1(i) to a fixed $\boldsymbol{\beta}_2$ setting and infer that

$$\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1^0]}(\hat{e}_{1,t+1}^2-u_{t+1}^2)}{T-k_0}\xrightarrow{p}\lambda_1^0\,\boldsymbol{\beta}_2'(\boldsymbol{Q}_{22}-\boldsymbol{Q}_{12}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\beta}_2, \tag{A.31}$$

yielding (for fixed $\boldsymbol{\beta}_2$)

$$\frac{Z_T^0(\lambda_1^0,\lambda_2^0)}{\sqrt{T-k_0}}\xrightarrow{p}\frac{1}{\sigma}\boldsymbol{\beta}_2'(\boldsymbol{Q}_{22}-\boldsymbol{Q}_{12}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\beta}_2, \tag{A.32}$$

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

where we also made use of Assumption B1(iii) ensuring that $\hat{\sigma} \xrightarrow{p} \sigma \in (0, \infty)$. It now follows that

$$\frac{\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0)}{\sqrt{T-k_0}} \equiv \frac{1}{\sqrt{T-k_0}} \frac{Z_T^0(\lambda_1^0, \lambda_2^0)}{\sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \xrightarrow{p} \frac{1}{\sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \frac{1}{\sigma} \boldsymbol{\beta_2}'(\boldsymbol{Q}_{22} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\beta_2} \tag{A.33}$$

with $v^0(\lambda_1^0, \lambda_2^0)$ given by (11), thus leading to $\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) \xrightarrow{p} \infty$ as stated. Proceeding similarly for $\overline{Z}_T(\tau_0; \lambda_2^0)$, we have

$$\frac{\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0)}{\sqrt{T-k_0}} \xrightarrow{p} \frac{1}{\sqrt{\overline{v}(\tau_0; \lambda_2^0)}} \frac{1}{\sigma} \boldsymbol{\beta}_2'(\boldsymbol{Q}_{22} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\beta}_2 \tag{A.34}$$

with $\overline{v}(\tau_0; \lambda_2^0)$ as in (13) and (14), thus also establishing that $\overline{\mathcal{S}}_T(\tau_0; \lambda_2^0) \xrightarrow{p} \infty$.

(ii) We next focus on the local asymptotic behavior of the two test statistics with $\boldsymbol{\beta}_2$ parameterized as $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{1/4}$. Using (A.1) in conjunction with Lemma A1(i) and (ii), we have

$$\begin{aligned}\frac{Z_T^0(\lambda_1^0, \lambda_2^0)}{\sqrt{v^0(\lambda_1^0, \lambda_2^0)}} &= \frac{1}{\hat{\sigma}} \frac{1}{\lambda_1^0 \sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \left[\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1^0]} (\hat{e}_{1,t+1}^2 - u_{t+1}^2)}{\sqrt{T-k_0}} \right] \\ &+ \frac{1}{\hat{\sigma}} \frac{1}{\lambda_1^0 \sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \left[\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1^0]} u_{t+1}^2}{\sqrt{T-k_0}} - \frac{\lambda_1^0}{\lambda_2^0} \frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2^0]} u_{t+1}^2}{\sqrt{T-k_0}} \right] \\ &+ o_p(1). \end{aligned} \tag{A.35}$$

It now follows directly from (A.12) in Lemma A1, Assumption B1(iii), and Slutsky's theorem that

$$\mathcal{S}_T^0(\lambda_1^0, \lambda_2^0) \equiv \frac{Z_T^0(\lambda_1^0, \lambda_2^0)}{\sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \xrightarrow{\mathcal{D}} \frac{\sqrt{1-\pi_0}}{\sigma\sqrt{v^0(\lambda_1^0, \lambda_2^0)}} \boldsymbol{\gamma}'(\boldsymbol{Q}_{22} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\gamma} + N(0,1) \tag{A.36}$$

as required. The result for $\overline{Z}_T(\tau_0; \lambda_2^0)$ follows identical arguments and is therefore omitted. □

Proof of Corollary 1. Follows directly from (A.34) to (A.36) in the proof of Proposition 3. □

LEMMA A2.

(i) *Under Assumption B2 and as $T \to \infty$, we have for $\lambda \in [0,1]$*

$$\frac{1}{(T-k_0)^2} \sum_{t=k_0}^{k_0+[(T-k_0)\lambda]} \boldsymbol{x}_t \boldsymbol{x}_t' \xrightarrow{\mathcal{D}} \frac{1}{(1-\pi_0)^2} \int_{\pi_0}^{\pi_0+(1-\pi_0)\lambda} \boldsymbol{J}_C \boldsymbol{J}_C' dr. \tag{A.37}$$

(ii) *Under Assumption B2 and as $T \to \infty$, we have for $\lambda \in [0,1]$*

$$\sup_{\lambda} \left\| \frac{1}{T-k_0} \sum_{t=k_0}^{k_0+[(T-k_0)\lambda]} \boldsymbol{x}_t u_{t+1} \right\| = O_p(1). \tag{A.38}$$

(iii) *Suppose model (2) holds with $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{3/4}$. Under Assumption B2 and as $T \to \infty$, we have*

$$T^{3/4}(\boldsymbol{\delta}_{1,[Ts]} - \boldsymbol{\beta}_1) \overset{\mathcal{D}}{\to} \left(\int_0^s \boldsymbol{J}_{1C} \boldsymbol{J}'_{1C} dr \right)^{-1} \left(\int_0^s \boldsymbol{J}_{1C} \boldsymbol{J}'_{2C} dr \right) \boldsymbol{\gamma} \equiv \boldsymbol{M}(s)\, \boldsymbol{\gamma}. \tag{A.39}$$

Proof of Lemma A2. For (A.37), we have

$$\begin{aligned}
\frac{1}{(T-k_0)^2} \sum_{t=k_0}^{k_0+[(T-k_0)\lambda]} \boldsymbol{x}_t \boldsymbol{x}'_t &= \left(\frac{T}{T-k_0} \right)^2 \frac{1}{T^2} \sum_{t=k_0}^{k_0+[(T-k_0)\lambda]} \boldsymbol{x}_t \boldsymbol{x}'_t \\
&= \left(\frac{T}{T-k_0} \right)^2 \sum_{t=k_0}^{k_0+[(T-k_0)\lambda]} \int_{\frac{t-1}{T}}^{\frac{t}{T}} \left(\frac{\boldsymbol{x}_{[Tr]}}{\sqrt{T}} \right) \left(\frac{\boldsymbol{x}_{[Tr]}}{\sqrt{T}} \right)' dr \\
&\overset{\mathcal{D}}{\to} \frac{1}{(1-\pi_0)^2} \int_{\pi_0}^{\pi_0+(1-\pi_0)\lambda} \boldsymbol{J}_C \boldsymbol{J}'_C dr
\end{aligned} \tag{A.40}$$

due to Assumption B2(i). For (A.38), we have

$$\begin{aligned}
\sup_{\lambda} \left\| \frac{1}{T-k_0} \sum_{t=k_0}^{k_0+[(T-k_0)\lambda]} \boldsymbol{x}_t u_{t+1} \right\| &\le \sup_{\lambda} \left| \frac{\sum u_{t+1}}{\sqrt{T-k_0}} \right| \sup_{\lambda} \left\| \frac{\boldsymbol{x}_{[(T-k_0)\lambda]}}{\sqrt{T-k_0}} \right\| \\
&= O_p(1),
\end{aligned} \tag{A.41}$$

which also follows from Assumption B2(i). For (A.39), we write

$$\begin{aligned}
T^{3/4}(\boldsymbol{\delta}_{1,[Ts]} - \boldsymbol{\beta}_1) &= \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1} \boldsymbol{x}'_{1,j-1}}{T^2} \right)^{-1} \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1} \boldsymbol{x}'_{2,j-1}}{T^2} \right) \boldsymbol{\gamma} \\
&\quad + \frac{1}{T^{1/4}} \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1} \boldsymbol{x}'_{1,j-1}}{T^2} \right)^{-1} \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1} u_j}{T} \right).
\end{aligned} \tag{A.42}$$

From (A.38), it also follows that

$$T^{3/4}(\boldsymbol{\delta}_{1,[Ts]} - \boldsymbol{\beta}_1) = \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1} \boldsymbol{x}'_{1,j-1}}{T^2} \right)^{-1} \left(\frac{\sum_{j=1}^{[Ts]} \boldsymbol{x}_{1,j-1} \boldsymbol{x}'_{2,j-1}}{T^2} \right) \boldsymbol{\gamma} + o_p(1) \tag{A.43}$$

and the statement in (A.39) follows directly using (A.37) in (A.43) and appealing to the continuous mapping theorem. □

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

LEMMA A3.

(i) *Suppose model (2) holds with* $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{3/4}$. *Under Assumption B2 and as* $T \to \infty$, *we have*

$$\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{1,t+1}^2 - u_{t+1}^2)}{\sqrt{T-k_0}} \xrightarrow{\mathcal{D}} \frac{1}{\sqrt{1-\pi_0}}\boldsymbol{\gamma}'\left(\int_{\pi_0}^{\pi_0+(1-\pi_0)\lambda} \boldsymbol{J}_C{}^*(s)\boldsymbol{J}_C{}^*(s)'ds\right)\boldsymbol{\gamma}, \quad \textbf{(A.44)}$$

where $\boldsymbol{J}_C^*(s) = \boldsymbol{J}_{2C}(s) - \boldsymbol{M}(s)\boldsymbol{J}_{1C}(s)$ *for* $\boldsymbol{M}(s) = (\int_0^s \boldsymbol{J}_{1C}\boldsymbol{J}_{1C}'dr)^{-1}(\int_0^s \boldsymbol{J}_{1C}\boldsymbol{J}_{2C}'dr)$.

(ii) *Suppose model (2) holds with* $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{3/4}$. *Under Assumption B2 and as* $T \to \infty$, *we have*

$$\sup_{\lambda\in(0,1]}\left|\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{2,t+1}^2 - u_{t+1}^2)}{\sqrt{T-k_0}}\right| \xrightarrow{p} 0. \quad \textbf{(A.45)}$$

Proof of Lemma A3. We consider (A.44) first. We operate under $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{3/4}$. Recalling that $\hat{e}_{1,t+1} - u_{t+1} = \boldsymbol{x}_{2,t}'\boldsymbol{\beta}_2 - \boldsymbol{x}_{1,t}'(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1)$ and using $\lim_{T\to\infty}((T-k_0)/T)^j \to (1-\pi_0)^j$, we write

$$\begin{aligned}
&\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{1,t+1}^2 - u_{t+1}^2)}{\sqrt{T-k_0}} \\
&= \frac{(1-\pi_0)^{3/2}}{(T-k_0)^2}\boldsymbol{\gamma}'\left(\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]} \boldsymbol{x}_{2,t}\boldsymbol{x}_{2,t}'\right)\boldsymbol{\gamma} \\
&\quad + \frac{(1-\pi_0)^{3/2}}{(T-k_0)^2}\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]} (T^{3/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1))'\boldsymbol{x}_{1,t}\boldsymbol{x}_{1,t}'(T^{3/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1)) \\
&\quad - 2\frac{(1-\pi_0)^{3/2}}{(T-k_0)^2}\left(\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]} (T^{3/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1))'\boldsymbol{x}_{1,t}\boldsymbol{x}_{2,t}'\right)\boldsymbol{\gamma} \\
&\quad + 2\frac{(1-\pi_0)^{1/2}}{T^{1/4}}\left(\frac{1}{T-k_0}\boldsymbol{\gamma}'\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]} \boldsymbol{x}_{2,t}u_{t+1}\right) \\
&\quad - 2\frac{(1-\pi_0)^{1/2}}{T^{1/4}}\left(\frac{1}{T-k_0}\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]} (T^{3/4}(\hat{\boldsymbol{\delta}}_{1,t} - \boldsymbol{\beta}_1))'\boldsymbol{x}_{1,t}u_{t+1}\right) + o(1). \quad \textbf{(A.46)}
\end{aligned}$$

It next follows from (A.38) and (A.39) that the last two terms in the right-hand side of (A.46) are $O_p(T^{-1/4})$ so that we also have

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

$$
\begin{aligned}
&\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{1,t+1}^2-u_{t+1}^2)}{\sqrt{T-k_0}} \\
&\quad = \frac{(1-\pi_0)^{3/2}}{(T-k_0)^2}\boldsymbol{\gamma}'\left(\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}\boldsymbol{x}_{2,t}\boldsymbol{x}_{2,t}'\right)\boldsymbol{\gamma} \\
&\quad\quad + \frac{(1-\pi_0)^{3/2}}{(T-k_0)^2}\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(T^{3/4}(\hat{\boldsymbol{\delta}}_{1,t}-\boldsymbol{\beta}_1))'\boldsymbol{x}_{1,t}\boldsymbol{x}_{1,t}'(T^{3/4}(\hat{\boldsymbol{\delta}}_{1,t}-\boldsymbol{\beta}_1)) \\
&\quad\quad - 2\frac{(1-\pi_0)^{3/2}}{(T-k_0)^2}\left(\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(T^{3/4}(\hat{\boldsymbol{\delta}}_{1,t}-\boldsymbol{\beta}_1))'\boldsymbol{x}_{1,t}\boldsymbol{x}_{2,t}'\right)\boldsymbol{\gamma}+o_p(1). \qquad \textbf{(A.47)}
\end{aligned}
$$

Using (A.37) and (A.39) from Lemma A2 together with the continuous mapping theorem, (A.47) leads to the required result in (A.44). The result in (A.45) is established following similar arguments and details are omitted. □

Proof of Proposition 4 and Corollary 2. We focus on part (ii) of the Proposition as the test consistency property stated in part (i) follows as its direct consequence. From (A.1) and Lemma A2, under the local alternative $\boldsymbol{\beta}_2=\boldsymbol{\gamma}/T^{3/4}$, we have

$$
\begin{aligned}
&\frac{Z_T^0(\lambda_1^0,\lambda_2^0)}{\sqrt{v^0(\lambda_1^0,\lambda_2^0)}} \\
&\quad = \frac{1}{\hat{\sigma}}\frac{1}{\lambda_1^0\sqrt{v^0(\lambda_1^0,\lambda_2^0)}}\left[\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1^0]}u_{t+1}^2}{\sqrt{T-k_0}}-\frac{\lambda_1}{\lambda_2^0}\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2^0]}u_{t+1}^2}{\sqrt{T-k_0}}\right] \\
&\quad\quad + \frac{1}{\hat{\sigma}}\frac{1}{\lambda_1^0\sqrt{v^0(\lambda_1^0,\lambda_2^0)}}\left[\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_1^0]}(\hat{e}_{1,t+1}^2-u_{t+1}^2)}{\sqrt{T-k_0}}\right]+o_p(1) \\
&\quad \xrightarrow{\mathcal{D}} N(0,1)+\frac{1}{\sigma}\frac{1}{\lambda_1^0\sqrt{v^0(\lambda_1^0,\lambda_2^0)}}\frac{1}{\sqrt{1-\pi_0}}\boldsymbol{\gamma}'\left(\int_{\pi_0}^{\pi_0+(1-\pi_0)\lambda_1^0}\boldsymbol{J_C}^*(s)\boldsymbol{J_C}^*(s)'ds\right)\boldsymbol{\gamma} \qquad \textbf{(A.48)}
\end{aligned}
$$

using (A.44), Slutsky and the continuous mapping theorems (note that the standard normality of the first component in the right-hand side of (A.48) has been established in Proposition 1). It now follows directly from (A.48) that $\lim_{||\gamma||\to\infty}\lim_{T\to\infty}\mathcal{S}_T(\lambda_1^0,\lambda_2^0)$ as required. The result for $\overline{\mathcal{S}}_T(\tau_0;\lambda_2^0)$ follows identical lines and its details are omitted. □

Proof of Proposition 5 and Corollary 3. We have $\hat{e}_{1,t+1}^2-\tilde{e}_{2,t+1}^2=(\hat{e}_{1,t+1}^2-\hat{e}_{2,t+1}^2)+(\hat{e}_{1,t+1}-\hat{e}_{2,t+1})^2$ which leads to the formulations of $\mathcal{S}_{T,adj}^0(\lambda_1^0,\lambda_2^0)$ and $\overline{\mathcal{S}}_{T,adj}(\tau_0;\lambda_2^0)$ in (34) and (35), respectively. Under the null hypothesis and for both test statistics, the result follows by verifying that

$$
\sup_{\lambda}\left|\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{1,t+1}-\hat{e}_{2,t+1})^2}{\sqrt{T-k_0}}\right|\xrightarrow{p}0. \qquad \textbf{(A.49)}
$$

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

Noting that

$$\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{1,t+1}-\hat{e}_{2,t+1})^2}{\sqrt{T-k_0}} = \frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{1,t+1}^2-u_{t+1}^2)}{\sqrt{T-k_0}} + \frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{2,t+1}^2-u_{t+1}^2)}{\sqrt{T-k_0}} - 2\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda]}(\hat{e}_{1,t+1}\hat{e}_{2,t+1}-u_{t+1}^2)}{\sqrt{T-k_0}}, \tag{A.50}$$

the statement in (A.49) follows directly from Assumption A(i) since we operate under the null hypothesis noting also that $(\hat{e}_{1,t+1}\hat{e}_{2,t+1}-u_{t+1}^2) = (\hat{e}_{1,t+1}-\hat{e}_{2,t+1})\hat{e}_{2,t+1} + (\hat{e}_{2,t+1}^2-u_{t+1}^2)$ from which we infer the $o_p(1)$'ness of the third component in the right-hand side of (A.50). It now follows that Propositions 1 and 2 continue to hold for the two adjusted statistics.

For the behavior of the adjusted statistics under the alternative, we initially consider the case of stationary predictors and operate under $\boldsymbol{\beta}_2 = \boldsymbol{\gamma}/T^{1/4}$ as in the setting of Corollary 1. Using (A.50) with Lemma A1, we can write

$$\begin{aligned} h_T^0(\lambda_1^0,\lambda_2^0) &= \frac{1}{\hat{\sigma}}\frac{1}{\sqrt{v^0(\lambda_1^0,\lambda_2^0)\lambda_2^0}}\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2^0]}(\hat{e}_{1,t+1}^2-u_{t+1}^2)}{\sqrt{T-k_0}} + o_p(1) \\ &\stackrel{p}{\to} \frac{1}{\sigma}\frac{1}{\sqrt{v^0(\lambda_1^0,\lambda_2^0)}}\sqrt{1-\pi_0}\,\boldsymbol{\gamma}'(\boldsymbol{Q}_{22}-\boldsymbol{Q}_{12}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\gamma} \equiv \psi^0 \end{aligned} \tag{A.51}$$

and

$$\begin{aligned} \overline{h}_T(\tau_0;\lambda_2^0) &= \frac{1}{\hat{\sigma}}\frac{1}{\sqrt{\overline{v}(\tau_0;\lambda_2^0)\lambda_2^0}}\frac{\sum_{t=k_0}^{k_0-1+[(T-k_0)\lambda_2^0]}(\hat{e}_{1,t+1}^2-u_{t+1}^2)}{\sqrt{T-k_0}} + o_p(1) \\ &\stackrel{p}{\to} \frac{1}{\sigma}\frac{1}{\sqrt{\overline{v}(\tau_0;\lambda_2^0)}}\sqrt{1-\pi_0}\,\boldsymbol{\gamma}'(\boldsymbol{Q}_{22}-\boldsymbol{Q}_{12}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\boldsymbol{\gamma} \equiv \overline{\psi}. \end{aligned} \tag{A.52}$$

Using (A.51) and (A.52), it follows that $\mathcal{S}_{T,adj}^0(\lambda_1^0,\lambda_2^0) \stackrel{\mathcal{D}}{\to} N(2\psi^0,1)$ and similarly for $\overline{\mathcal{S}}_{T,adj}^0(\lambda_1^0,\lambda_2^0) \stackrel{\mathcal{D}}{\to} N(2\overline{\psi},1)$ which establishes the fact that Proposition 3 continues to hold for the two adjusted statistics in addition to part (i) of Corollary 3. The result for the case of persistent predictors follows identical lines, making use of Lemma A3 and Proposition 3 which in turn establishes part (ii) of Corollary 3 and that Proposition 4 also holds for the two adjusted statistics. □

DATA AVAILABILITY STATEMENT

MATLAB codes and replication files are available for download from the author's webpage at https://sites.google.com/view/jpitarakis.

https://doi.org/10.1017/S0266466623000154 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

Pitarakis, Jean-Yves (2023): Supplement to "A novel approach to predictive accuracy testing in nested environments," Econometric Theory Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466623000154.

REFERENCES

Avdis, E. & J.A. Wachter (2017) Maximum likelihood estimation of the equity premium. *Journal of Financial Economics* 125, 589–609.

Berenguer-Rico, V. & B. Nielsen (2020) Cumulated sum of squares statistics for nonlinear and nonstationary regressions. *Econometric Theory* 36, 1–47.

Berkes, I., S. Hörmann, & L. Horvath (2008) The functional central limit theorem for a family of GARCH observations with applications. *Statistics and Probability Letters* 78, 2725–2730.

Clark, T.E. & M.W. McCracken (2001) Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* 105, 85–110.

Clark, T.E. & M.W. McCracken (2005) Evaluating direct multistep forecasts. *Econometric Reviews* 24, 369–404.

Clark, T.E. & M.W. McCracken (2013) Advances in forecast evaluation. In G. Elliott & A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 2, Part B, pp. 1107–1201. Elsevier.

Clark, T.E. & K.D. West (2007) Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138, 291–311.

Deng, A. & P. Perron (2008a) The limit distribution of the Cusum of squares test under general mixing conditions. *Econometric Theory* 24, 809–822.

Deng, A. & P. Perron (2008b) A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change. *Journal of Econometrics* 142, 212–240.

Diebold, F.X. (2015) Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business and Economic Statistics* 33, 1–24.

Diebold, F.X. & R. Mariano (1995) Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–265.

Engel, C. & S. Wu (2021) Forecasting the U.S. Dollar in the 21st Century. NBER Working Paper no. 28447.

Fan, J., Y. Liao, & J. Yao (2015) Power enhancement in high dimensional cross-sectional tests. *Econometrica* 83, 1497–1541.

Ferson, W., S. Nallareddy, & X. Biqin (2013) The out-of-sample performance of long run risk models. *Journal of Financial Economics* 107, 537–556.

Gades-Riva, M.D. & J. Gonzalo (2020) Trends in distributional characteristics: Existence of global warming. *Journal of Econometrics* 214, 153–174.

Giacomini, R. & H. White (2006) Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.

Giraitis, L., P. Kokoszka, & R. Leipus (2000) Stationary ARCH models: Dependence structure and central limit theorems. *Econometric Theory* 16, 3–22.

Giraitis, L., P. Kokoszka, & R. Leipus (2001) Testing for long memory in the presence of a general trend. *Journal of Applied Probability* 38, 1033–1054.

Granziera, E., K. Hubrich, & H. Moon (2014) Predictability tests for a small number of nested models. *Journal of Econometrics* 182, 174–185.

Hansen, P.R. & A. Timmermann (2015) Equivalence between out-of-sample forecast comparisons and Wald statistics. *Econometrica* 83, 2485–2505.

Ince, O., T. Molodotsova, & D.H. Papell (2016) Taylor rule deviations and out-of-sample exchange rate predictability. *Journal of International Money and Finance* 69, 22–44.

Jiang, F., Z. Zhao, & X. Shao (2020) Modelling the COVID-19 infection trajectory: A piecewise linear quantile trend model. *Journal of the Royal Statistical Society: Series B* 84, 1589–1607.

Li, S. & O. Linton (2021) When will the COVID-19 pandemic peak? *Journal of Econometrics* 220, 130–157.
Linder, A.M. (2009) Stationarity, mixing, distributional properties and moments of GARCH(p,q)-processes. In T. Mikosch, J. Kreiss, A.D. Richard, & T.G. Andersen (eds.), *Handbook of Financial Time Series*, pp. 43–69. Springer.
McCracken, M. (2007) Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics* 140, 719–752.
Meese, R.A. & K. Rogoff (1983) Empirical exchange rate models of the seventies. *Journal of International Economics* 14, 3–24.
Molodotsova, T. & D.H. Papell (2009) Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics* 77, 167–180.
Rossi, B. (2005) Testing long-horizon predictive ability with high persistence, and the Meese–Rogoff puzzle. *International Economic Review* 46, 61–92.
Schennah, S.M. & D. Wilhelm (2017) A simple parametric model selection test. *Journal of the American Statistical Association* 112, 1663–1674.
Shi, X. (2015) A nondegenerate Vuong test. *Quantitative Economics* 6, 85–121.
Stock, J. & M. Watson (2010) Modeling Inflation after the Crisis. *Macroeconomic Policy: Post-Crisis and Risks Ahead*, Proceedings of the Federal Reserve Bank of Kansas City 2010 Jackson Hole Symposium.
Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.
West, K. (1996) Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.
West, K. (2006) Forecast evaluation. In G. Elliott, C.W.J. Granger, & A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1, pp. 99–134. Elsevier.
Wu, W. & Z. Zhao (2007) Inference of trends in time series. *Journal of the Royal Statistical Society: Series B* 69, 391–410.