UNIVERSITY *of York*

Discussion Papers in Economics

No. 21/03

Testing for equal predictive accuracy with strong dependence

Laura Coroneo and Fabrizio Iacone

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Testing for equal predictive accuracy with strong dependence

Laura Coroneo[1] and Fabrizio Iacone[*2,1]

[1]University of York

[2]Università degli Studi di Milano

21st May 2021

Abstract

We revisit the Diebold and Mariano (1995) test, investigating the consequences of having autocorrelation in the loss differential. This situation can arise not only when a forecast is sub-optimal under MSE loss, but also when it is optimal under an alternative loss, or it is evaluated on a short sample, or when a forecast with weakly dependent forecast errors is compared to a naive benchmark. We show that the power of the Diebold and Mariano (1995) test decreases as the dependence increases, making it more difficult to obtain statistically significant evidence of superior predictive ability against less accurate benchmarks. Moreover, we find that after a certain threshold the test has no power and the correct null hypothesis is spuriously rejected. Taken together, these results caution to seriously consider the dependence properties of the selected forecast and of the loss differential before the application of the Diebold and Mariano (1995) test, especially when naive benchmarks are considered.

JEL classification codes: C12; C32; C53.

Keywords: strong autocorrelation, Forecast evaluation, Diebold and Mariano Test, Long Run Variance Estimation.

[*]Corresponding author. Contact: Department of Economics, Management and Quantitative Methods,Università degli Studi di Milano Via Conservatorio 7, 20122 Milano. Email: fabrizio.iacone@unimi.it

1 Introduction

Accurate forecasts are extremely important for forward-looking decision making. Weather forecasts often have a dedicated section even on the daily news, and recently predictions of the diffusion of the COVID-19 pandemic have had a critical impact on the life of most people. In economics, decisions over individual savings, firm-level investments, government fiscal policies and central bank monetary policies rely on forecasts of, among others, future economic activity and price levels.

To discriminate between good and bad forecasts, Diebold and Mariano (1995) [DM hereafter] suggested comparing alternative forecasts using a test for equal predictive ability. The DM test is based on a loss function associated with the forecast errors of each forecast, and allows to test the null hypothesis of zero expected loss differential between two competing forecasts. This approach takes forecast errors as model-free, and the test is valid also when the forecasts are produced from unknown models, as for example from forecast survey data. In addition, if the objective is to compare forecasting methods as opposed to forecasting models, then Giacomini and White (2006) showed that in an environment with asymptotically non-vanishing estimation uncertainty the DM test can still be applied.

The DM test allows us to test for equal predictive accuracy using any loss function, and the test statistic is valid for contemporaneously correlated, serially correlated, and non-normal forecast errors. The test relies on the assumption that the loss differential is weakly dependent. The rationale for this assumption is that under mean squared error (MSE) loss optimal q-step ahead forecasts should generate at most MA$(q-1)$ errors, so if the considered forecast approximates the optimal forecast, then its forecast errors should not be too correlated over time, although dependence beyond the MA$(q-1)$ boundary may take place. In practice, forecasts with errors that are fairly correlated can occur not only when the considered forecast fails to approximate the optimal forecast under MSE loss, but also when the forecast is optimal under an alternative loss function. Patton and Timmermann (2007) show that an optimal forecast under asymmetric loss can have forecast errors that are serially correlated of arbitrarily high order. Another situation in which we can have dependency in the loss differential is when a long-run forecast is evaluated on a relatively short sample.

Still, we can encounter situations in which the loss differential is highly autocorrelated even in the presence of a prediction with weakly dependent forecast errors and in samples of moderate size. This can happen when the DM test is used to compare the predictive ability of a selected forecast against a naive benchmark. This is a common practice, as naive benchmarks are cost-effective and readily available at any time, so they provide a reference standard for comparisons. Using simple benchmarks allows us to understand the added value of a specific forecasting technique, as it is desirable that predictions from sophisticated forecasting methods (for example complex models or expensive surveys) are more accurate than naive benchmarks. However, naive forecasts may in some cases generate relevant autocorrelation in the loss differential.

In this paper, we study the performance of the DM test when the assumption of weak autocorrelation does not hold. We characterise strong dependence as local to unity as in Phillips (1987) and Phillips and Magdalinos (2007b). This definition is at odds with the more popular characterisation in the literature that treats strong autocorrelation and long memory as synonyms. Local to unity, however, seems well suited to derive reliable guidance when the sample is not very large, as it is the case in many applications. With this definition the strength of the dependence is determined also by the sample size: a stationary AR(1) process with root close to unity may be treated as weakly dependent in a very large sample, but standard asymptotics may be a poor guidance for cases with smaller samples and local to unity asymptotics may be more informative. We show that the power of the DM test decreases as the dependence increases, making it more difficult to obtain statistically significant evidence of superior predictive ability against less accurate benchmarks. We also find that after a certain threshold the test has no power and the correct null hypothesis is spuriously rejected. These results caution us to seriously consider the dependence properties not only of the selected forecast but also of the loss differential before the application of the DM test, especially when naive benchmarks are considered.

In the literature, there has been some attention to the issue of forecast evaluation in presence of persistence. Corradi, Swanson and Olivetti (2001) examines the DM statistic in the presence of cointegration, whereas Rossi (2005) examines the effect of high persistence on the loss differential. McCracken (2020) provides an example to show that using a fixed and finite estimation window can result in loss differentials that depend on the first observations, so that the time series of the loss differential in the DM test is not ergodic for the mean. These works considered a framework with parameter estimation error; our study is rather closer to Kruse, Leschinski and Will (2019), who take forecast errors as primitives and derive the properties of the DM test in the presence of long memory using standard asymptotics, and memory and autocorrelation consistent standardisation. Kruse et al. (2019) had a sample of 4883 observations, so long memory seems a reasonable modelling strategy in their case.

To illustrate the problems associated with the DM test when there is dependence in the loss differential, we consider the case in which an AR(1) forecast for inflation in the Euro Area is compared to two naive benchmarks: a constant 2% prediction (that represents the inflation target in the Euro Area) and a moving average prediction. These benchmark predictions have highly dependent forecast errors. As a consequence, the loss differential is dependent, and the DM test fails to reject the null of equal predictive accuracy, even if the benchmarks are less accurate than the AR(1) forecast for short forecasting horizons.

The paper is organised as follows. We formally introduce the DM test in Section 2, and derive the limit properties of the DM statistics in presence of dependence in Section 3. We investigate the practical implication of our theoretical findings in a Monte Carlo exercise (Section 4) and in the empirical application (Section 5). Details on the assumptions of the DGP and formal derivations are in the Appendix.

2 DM test

The DM test was introduced to compare two forecasts of a time series, according to a user chosen loss metric. For $t = \{1, \ldots, T\}$, denoting the forecast errors as $e_{1,t}$ and $e_{2,t}$, respectively, and the loss function $L(.)$, Diebold and Mariano (1995) consider the loss differential

$$d_t = L(e_{1,t}) - L(e_{2,t}) \tag{1}$$

and test the null hypothesis of equal predictive ability $H_0 : E(d_t) = 0$ against the alternative $H_1 : E(d_t) \neq 0$. The key assumptions by Diebold and Mariano (1995) and Diebold (2015) are that d_t is stationary and weakly dependent, and that the average loss, $\overline{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$, follows a Central Limit Theorem. In particular, denoting $\mu = E(d_t)$, it is assumed that $\sqrt{T}(\overline{d} - \mu) \to_d N(0, \sigma^2)$ as $T \to \infty$, where $0 < \sigma^2 < \infty$ is the long run variance of d_t.

Thus, inference on $E(d_t)$ can be based on the normalised limit

$$\sqrt{T}\frac{(\overline{d} - \mu)}{\sigma} \to_d N(0, 1). \tag{2}$$

Denoting $\widehat{\sigma}^2$ an estimate of σ^2, $\widehat{\sigma} = \sqrt{\widehat{\sigma}^2}$, then the classical DM test uses the statistic

$$DM = \sqrt{T}\frac{\overline{d}}{\widehat{\sigma}}$$

where the null hypothesis of equal predictive ability is rejected at 5% significance level against a two-sided alternative if the realization of $|DM|$ is above the 1.96 threshold.

The original DM test exploits the consistency of $\widehat{\sigma}^2$ to justify the standard normal as the limit distribution under the null. This may generate rather poor size performance in finite sample, see DM and also Clark (1999). Coroneo and Iacone (2020) use fixed smoothing asymptotics instead, see for example Kiefer and Vogelsang (2005), Sun (2014c), Sun (2014b), Lazarus, Lewis, Stock and Watson (2018). With this approach, the limit for $\widehat{\sigma}^2$ is derived under alternative asymptotics: the DM statistic does not have limit standard normal distribution, but the alternative limit provides a better approximation of the distribution of the DM statistic in finite samples.

As the alternative distribution depends on the way σ^2 is estimated, we focus on the Weighted Periodogram estimate using the Daniell kernel. Denoting $w(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} d_t e^{i\lambda t}$ the Fourier transform of d_t at frequency λ, and $I(\lambda) = |w(\lambda)|^2$ as the periodogram, then we consider in particular the estimate

$$\widehat{\sigma}^2 = \frac{2\pi}{m} \sum_{j=1}^{m} I(\lambda_j) \tag{3}$$

where $\lambda_j = \frac{2\pi j}{T}$ are the Fourier frequencies and m is a user-chosen bandwidth parameter: this is treated as $m \to \infty$ as $T \to \infty$ (and such that $m/T \to 0$) for standard asymptotics but as fixed for fixed smoothing asymptotics. Under this alternative asymptotics, for $\widehat{\sigma}^2$ defined as the Weighted periodogram estimate with Daniell kernel as in (3), it holds that, under H_0, as $T \to \infty$,

$$DM \to_d t_{2m}, \tag{4}$$

see Coroneo and Iacone (2020) for more details.

3 The DM statistic with dependence

The key assumption in the construction of the DM test is the conjecture that d_t is weakly dependent. This assumption seems reasonable in the context of forecasts as it is well known that, under MSE loss, optimal $q-$step ahead forecasts should be at most $MA(q-1)$. For this very reason Diebold and Mariano (1995) considered estimating σ^2 using only the first $q-1$ autocovariances of d_t, and verified that this assumption was met in the data in the empirical application that they presented.

However, it is common practice to apply the DM test to test for equal predictive accuracy of a selected forecast against a naive benchmark, resulting in dependent forecast errors and loss differential, possibly even strongly autocorrelated. Strong autocorrelation in the loss differential might also arise even in presence of forecasts that are optimal under alternative loss functions, or when the forecast evaluation sample T is short.

Denoting $y_t = d_t - \mu$, so that

$$d_t = \mu + y_t, \tag{5}$$

we assume that

$$y_t = \rho_T y_{t-1} + u_t \tag{6}$$

where u_t is a zero mean, weakly dependent process with long run variance ω. We consider two different models for ρ_T: in subsection 3.1 we discuss the local-to-unity AR(1) approximation as in Phillips (1987) (alongside with the standard unit root model); in subsection 3.2 we consider the moderate deviations from a unit root as in Phillips and Magdalinos (2007b). These models are a convenient representation of dependence for y_t when the dimension in T is relatively short, as it is indeed the case in many empirical studies. In both cases, we refer to the Appendix for a detailed presentation and discussion of the assumptions, and for the derivation of the results.

3.1 Local to unity autocorrelation

In this case we assume for ρ_t in (6)

$$\rho_T = e^{c/T} \quad \text{with } c \leq 0. \tag{7}$$

When c is in the neighbourhood of 0, ρ_T is approximated as $1 + c/T$, i.e. $\rho_T \sim 1 + c/T$ as $c \to 0$. When $c = 0$ then the process y_t has a unit root, and it is initialised setting the initial condition $y_0 = O_p(1)$.

Define

$$J_c(r) = \int_0^r e^{(r-s)c} dW(s)$$

where $W(r)$ is a standard Brownian motion. The process $J_c(r)$ is a Ornstein-Uhlenbeck process: for given r it is normally distributed (when $c = 0$ the Ornstein-Uhlenbeck process is the standard Brownian motion). We refer to Phillips (1987) for a detailed discussion, but we state some important results from Lemma 1 of Phillips (1987):

$$T^{-1/2} y_{\lfloor rT \rfloor} \Rightarrow \omega J_c(r) \tag{8}$$

$$T^{-1/2} \overline{y} \Rightarrow \omega \int_0^1 J_c(r) dr \tag{9}$$

$$T^{-2} \sum_{t=1}^{T} y_t^2 \Rightarrow \omega^2 \int_0^1 J_c(r)^2 dr \tag{10}$$

where $J_c(r) = W(r)$ when $c = 0$. The result for $\overline{y}$ in particular means that the sample average is not consistent in the neighbourhood of a unit root. Denoting

$$\overline{J_c} = \int_0^1 J_c(r) dr,$$

from these results it is also possible to establish that, under H_0, for $m \to \infty$, $m/T \to 0$ as $T \to \infty$

$$\frac{1}{\sqrt{m}} DM \Rightarrow \frac{\overline{J_c}}{\sqrt{\frac{1}{2} \int_0^1 (J_c(r) - \overline{J_c})^2 dr}}, \tag{11}$$

we refer to the Appendix for a more detailed derivation of this and other results. In view of (11), as $m \to \infty$ the DM test statistic diverges even when the null hypothesis is correct, thus giving spurious evidence of superior predictive ability. As we interpret the local to unit root as an approximation of an AR(1) in finite sample, this result suggests that the DM test diverges in presence of a root that is stationary but close to 1. Notice that this result is not affected by whether the standard normal or the t_{2m} is used to draw the critical values.

Next, we present the limit of the DM statistic using fixed smoothing asymptotic. For the generic function of $r \in (0,1)$ $a(r)$, denote $A(r) = \int_0^r a(s)ds$, $\widehat{A(r)} = A(r) - r\,A(1)$ and

$$Q_A(j) = (2\pi j)^2 \left\{ \left(\int_0^1 sin(2\pi jr)\widehat{A(r)}dr \right)^2 + \left(\int_0^1 cos(2\pi jr)\widehat{A(r)}dr \right)^2 \right\}$$

then under $c \leq 0$, for m fixed as $T \to \infty$,

$$DM \Rightarrow \frac{\overline{J_c}}{\sqrt{\frac{1}{m}\sum_{j=1}^m Q_{Jc}(j)}}. \tag{12}$$

When $c = 0$, so that the process y_t is characterised by a unit root, then the limit distribution in (11) is $\frac{\overline{W}}{\sqrt{\frac{1}{2}\int_0^1 (W(r)-\overline{W})^2 dr}}$, where we used $\overline{W} = \int_0^1 W(r)dr$; also, in this case $\widehat{A(r)}$ is $\int_0^r W(s)ds - r\int_0^1 W(s)ds$ and the limit is distribution in (12) is $\frac{\overline{W}}{\sqrt{\frac{1}{m}\sum_{j=1}^m Q_W(j)}}$.

Notice that the limit (12) holds regardless of whether $\mu = 0$ or $\mu \neq 0$. This means that the DM test is not consistent under this alternative asymptotic.

The limit in (12) exibits the self-normalization property as in Shao (2015). It is interesting to compare the limits in (11) and in (12) with results in the literature. It is well known that the standardised mean is diverging in presence of strongly autocorrelated series when $m \to \infty$ is assumed, but not when m is assumed fixed, see for example McElroy and Politis (2012) and Hualde and Iacone (2017). In the context of local to unity process, this result was established, for example, in Sun (2014a). The advantages of series variance estimators of the long run variance in presence of autocorrelation is also discussed in Müller (2007). Our interest in (11) and (12) is, however, slightly different, as we do not see these as alternative limits under different asymptotics, but rather as guidance that show properties of the DM test for relatively small and large m.

Remark 1. *Results in (11) and in (12) indicate that:*

a. With relatively large values for m, the DM statistics diverges even under H_0 (spurious significance).

b. With small values for m, the DM statistics does not diverge even under H_1, so the test is not consistent.

However, as the distribution in (12) has much thicker tails than a t_{2m} distribution, then it is still possible (and indeed it may be frequent) to have many spurious rejections of the null hypothesis.

3.2 Moderate deviations from unit root

As c in (7) varies between $-\infty$ and 0, it is possible to use the theory from subsection 3.1 for any AR(1) model with positive autocorrelation. However, the limits (11) - (12) may not provide a valuable guideline when ρ_T is not in fact in the very close neighbourhood of unity. For this class of models, Phillips and Magdalinos (2007b) generalise ρ_T to moderate deviations from the unit root. We simplify the model slightly, and consider

$$\rho_T = 1 + c/T^{\alpha} \text{ for } \alpha \in (0,1), \text{ and } c < 0. \tag{13}$$

Moderate deviations from the unit root following (13) are also discussed in Phillips and Magdalinos (2007a). Giraitis and Phillips (2012) provide a generalisation of some results under a weaker condition, similar to $(1-\rho_T)T \to \infty$. Under assumption (13), $\overline{d}$ is a consistent estimate of μ only when $\alpha \in (0, 1/2)$, but the CLT in (2) still holds for any $\alpha \in (0,1)$, see Theorem 2.1 and the discussion on page 168 of Giraitis and Phillips (2012). Recalling that, for any T, $\sigma^2 = (1-\rho_T)^{-2}\omega^2$ and noticing that this is proportional to $T^{2\alpha}$ in large sample, the rate of convergence of the CLT is reduced to $\sqrt{T^{1-2\alpha}}$ (the theory does not cover the $\alpha = 1$ case but notice that $\sqrt{T^{1-2\alpha}} \to T^{-1/2}$ as $\alpha \to 1$ and this is the rate in (8), suggesting a proximity of the two representations; the extension of (8) under (13) is explored more in Phillips and Magdalinos (2007a)).

We can establish that, for $m \to \infty$, $m/T \to 0$ as $T \to \infty$,

$$\sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}} \to_d N(0,1), \text{ for } mT^{\alpha-1} \to 0 \tag{14}$$

$$(mT^{\alpha-1})^{-1/2} 1/2(-c)^{1/2} \left\{ \sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}} \right\} \to_d N(0,1), \text{ for } mT^{\alpha-1} \to \infty \tag{15}$$

for any $\alpha \in (0,1)$ where as usual $\widehat{\sigma}^2$ is the Daniell estimate of σ^2 as in (3).

Remark 2. *Results in (14) and in (15) indicate that:*

a. *Rewriting $DM = \sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}} + \sqrt{T}\frac{\mu}{\widehat{\sigma}}$ for m as in (14), the power depends on the drift $\sqrt{T}\frac{\mu}{\widehat{\sigma}} = O(T^{1/2-\alpha})$. Therefore, the DM test still has power in cases of moderate deviations from the unit root, when $\alpha < 1/2$. However, from this representation it is immediate to see that as the drift is of order $T^{1/2-\alpha}$, the power decreases as $\alpha \to 1/2$. This result has an important practical implication, as it means that it is progressively more difficult to detect forecast inaccuracy as the dependence increases, even well within the weak dependence region.*

b. *The test is not consistent for $\alpha \in (1/2, 1)$.*

c. *When $mT^{\alpha-1} \to \infty$, the DM statistics diverges even under H_0, thus resulting again in spurious significance: condition $mT^{\alpha-1} \to 0$ in (14) is not binding when $\alpha = 0$ but it is very strong as $\alpha \to 1$. Thus, results in (14) and (15) are intermediate between the weakly dependent $|\rho_T| = |\rho| < 1$ case and the unit root case. Taken together, they suggest that for large values of m the DM test will give spurious significance in finite sample as ρ is close to 1, and this problem is more relevant the closer ρ is to unity, relative to the sample size, and the larger the bandwidth m is.*

4 Monte Carlo results

In this section, we investigate the properties of the DM statistic in the neighbourhood of unity in a Monte Carlo exercise. In the first part of the Monte Carlo exercise we simulate the d_t statistics and verify the predictions that we formulated in Remarks 1 and 2. We derived the results in Section 3 under assumptions that have been formulated for d_t, in the spirit of Diebold and Mariano (1995) and Diebold (2015). This is very convenient in our situation, as it gave us the possibility to characterise results in (11), (12), (14), and in (15), with a smooth increase of dependence. Giacomini and Rossi (2010), on the other hand, derived the properties of the DM test from assumptions on a specific forecasting model. This seems rather difficult under assumptions on dependence as in (7) or (13), although it is possible to derive for example the properties of e_{1t}^2 using (8) and the continuous mapping theorem when e_{1t} is a local to unity process, see for example Rossi (2005). Proceeding as in Clark (1999), in the second part of the Monte Carlo exercise, we explore the properties of the DM test imposing autocorrelation in e_{1t}, e_{2t} in (1), and we verify that the conclusions that we drew in Remarks 1 and 2 are still a valid guidance.

In both parts, we consider two sample sizes, $T = 50$ and $T = 100$, and a range of bandwidths spanning $m = 1$ to $m = \lfloor T^{2/3} \rfloor$. For each experiment we run $10,000$ repetitions, and we compute the empirical frequencies of rejections of the two sided version of the test, i.e. we compare the $|DM|$ statistic against the appropriate 5% critical value from the t_{2m} distribution. We always use the t_{2m} statistic as this has better size properties, as for example discussed in Lazarus et al. (2018) or in Coroneo and Iacone (2020).

In the first experiment, we simulate

$$d_t = \mu + y_t, \; y_t = \rho y_{t-1} + \varepsilon_t \tag{16}$$

where ε_t is $NIID(0, 1)$, for a range of values of ρ from 0 to 1. For $\rho = 1$, to facilitate comparison, we initialise the process setting the same initial condition that we use for $\rho = 0.99$. We consider two values of μ, $\mu = 0$ to

Table 1: Empirical null rejection frequencies

$\mu = 0$, $T = 50$

m \ ρ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99	1
1	0.046	0.048	0.056	0.060	0.068	0.086	0.150	0.255	0.427	0.498
$\lfloor T^{1/4} \rfloor$	0.051	0.055	0.076	0.091	0.119	0.178	0.314	0.467	0.647	0.708
$\lfloor T^{1/3} \rfloor$	0.049	0.061	0.104	0.131	0.175	0.254	0.419	0.566	0.727	0.776
$\lfloor T^{1/2} \rfloor$	0.048	0.096	0.213	0.265	0.339	0.441	0.599	0.720	0.827	0.861
$\lfloor T^{2/3} \rfloor$	0.049	0.157	0.336	0.398	0.471	0.569	0.695	0.795	0.874	0.897

$\mu = 1$, $T = 50$

m \ ρ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99	1
1	0.914	0.497	0.214	0.175	0.148	0.137	0.175	0.262	0.429	0.497
$\lfloor T^{1/4} \rfloor$	0.999	0.785	0.386	0.323	0.279	0.270	0.357	0.486	0.651	0.712
$\lfloor T^{1/3} \rfloor$	1.000	0.874	0.504	0.436	0.385	0.376	0.463	0.583	0.731	0.777
$\lfloor T^{1/2} \rfloor$	1.000	0.957	0.706	0.635	0.581	0.570	0.639	0.730	0.830	0.864
$\lfloor T^{2/3} \rfloor$	1.000	0.982	0.808	0.746	0.695	0.680	0.730	0.804	0.876	0.899

$\mu = 0$, $T = 100$

m \ ρ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99	1
1	0.052	0.052	0.054	0.055	0.057	0.062	0.089	0.148	0.295	0.403
$\lfloor T^{1/4} \rfloor$	0.052	0.055	0.067	0.077	0.096	0.139	0.264	0.422	0.620	0.729
$\lfloor T^{1/3} \rfloor$	0.051	0.057	0.079	0.095	0.122	0.182	0.330	0.494	0.673	0.768
$\lfloor T^{1/2} \rfloor$	0.051	0.079	0.166	0.205	0.266	0.365	0.526	0.666	0.794	0.856
$\lfloor T^{2/3} \rfloor$	0.052	0.138	0.293	0.349	0.416	0.515	0.657	0.765	0.860	0.901

$\mu = 1$, $T = 100$

m \ ρ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99	1
1	0.991	0.726	0.319	0.241	0.172	0.124	0.116	0.165	0.300	0.405
$\lfloor T^{1/4} \rfloor$	1.000	0.986	0.627	0.495	0.375	0.302	0.325	0.442	0.623	0.724
$\lfloor T^{1/3} \rfloor$	1.000	0.993	0.694	0.565	0.445	0.367	0.389	0.510	0.678	0.769
$\lfloor T^{1/2} \rfloor$	1.000	0.998	0.854	0.763	0.659	0.571	0.588	0.682	0.798	0.858
$\lfloor T^{2/3} \rfloor$	1.000	1.000	0.926	0.860	0.780	0.707	0.710	0.779	0.859	0.904

Note: this table reports empirical null rejection frequencies for the DM test with fixed-m asymptotics. The data generating process is in (16). The sample size is 50 (two top panels) and 100 (two bottom panels). The bandwidth values m of $\lfloor T^{1/4} \rfloor$, $\lfloor T^{1/3} \rfloor$, $\lfloor T^{1/2} \rfloor$ and $\lfloor T^{2/3} \rfloor$ are 2, 3, 7 and 13 when $T = 50$, and 3, 4, 10 and 21 for $T = 100$. $\mu = 0$ refers to the empirical size (first and third panels), and $\mu = 1$ refers to the empirical power (second and last panels).

observe the effects on the empirical size, and $\mu = 1$ to observe the effects on power.

Results in Table 1 confirm the findings in Section 3. In particular:

a. The empirical size is correct for $\rho = 0$, but it deteriorates as we move closer to $\rho = 1$ and as m is larger.

b. The empirical power drops as we move closer to $\rho = 1$ and as m is larger, in the sense that the presence

Table 2: Empirical null rejection frequencies

$\mu = 0$, $T = 50$

m \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	0.046	0.044	0.063	0.069	0.081	0.112	0.185	0.275	0.424
$\lfloor T^{1/4} \rfloor$	0.047	0.047	0.070	0.087	0.117	0.168	0.276	0.417	0.597
$\lfloor T^{1/3} \rfloor$	0.045	0.048	0.081	0.102	0.137	0.196	0.335	0.495	0.669
$\lfloor T^{1/2} \rfloor$	0.046	0.051	0.110	0.145	0.199	0.293	0.478	0.644	0.780
$\lfloor T^{2/3} \rfloor$	0.047	0.064	0.150	0.200	0.275	0.392	0.579	0.724	0.835

$\mu = 1$, $T = 50$

m \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	0.295	0.228	0.160	0.151	0.150	0.164	0.210	0.291	0.428
$\lfloor T^{1/4} \rfloor$	0.478	0.366	0.250	0.237	0.233	0.244	0.313	0.434	0.604
$\lfloor T^{1/3} \rfloor$	0.566	0.424	0.294	0.276	0.266	0.285	0.372	0.513	0.671
$\lfloor T^{1/2} \rfloor$	0.670	0.503	0.361	0.345	0.347	0.283	0.512	0.656	0.783
$\lfloor T^{2/3} \rfloor$	0.704	0.545	0.408	0.402	0.419	0.475	0.607	0.732	0.839

$\mu = 0$, $T = 100$

m \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	0.050	0.054	0.057	0.066	0.075	0.093	0.144	0.213	0.322
$\lfloor T^{1/4} \rfloor$	0.050	0.048	0.075	0.088	0.113	0.153	0.242	0.360	0.555
$\lfloor T^{1/3} \rfloor$	0.051	0.050	0.079	0.094	0.120	0.167	0.274	0.407	0.608
$\lfloor T^{1/2} \rfloor$	0.049	0.055	0.095	0.120	0.166	0.241	0.405	0.575	0.740
$\lfloor T^{2/3} \rfloor$	0.048	0.063	0.140	0.180	0.248	0.357	0.539	0.689	0.815

$\mu = 1$, $T = 100$

m \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	0.481	0.356	0.201	0.172	0.157	0.151	0.176	0.225	0.326
$\lfloor T^{1/4} \rfloor$	0.856	0.650	0.354	0.303	0.260	0.245	0.286	0.377	0.560
$\lfloor T^{1/3} \rfloor$	0.892	0.686	0.379	0.324	0.280	0.266	0.315	0.423	0.613
$\lfloor T^{1/2} \rfloor$	0.939	0.751	0.444	0.390	0.353	0.351	0.445	0.589	0.743
$\lfloor T^{2/3} \rfloor$	0.949	0.789	0.509	0.459	0.433	0.454	0.573	0.695	0.818

Note: this table reports empirical null rejection frequencies for the DM test with fixed-m asymptotics. The data generating process is in equations (17)–(19). The sample size is 50. The bandwidth values m of $\lfloor T^{1/4} \rfloor$, $\lfloor T^{1/3} \rfloor$, $\lfloor T^{1/2} \rfloor$ and $\lfloor T^{2/3} \rfloor$ are 2, 3, 7 and 13. $\mu = 0$ refers to the empirical size, and $\mu = 1$ refers to the empirical power.

of $\mu = 1$ does not affect much the number of rejections of the null hypothesis, in those cases.

In the second part of the Monte Carlo exercise, we simulate

$$e_{1t} = \varepsilon_{1t} + k(\phi, \mu), \quad k^2(\phi, \mu) = \frac{\phi^2}{1-\phi^2} + \mu, \quad \varepsilon_{1t} \sim N(0,1) \tag{17}$$

$$e_{2t} = \phi e_{2,t-1} + \varepsilon_{2t}, \quad \varepsilon_{2t} \sim N(0,1) \tag{18}$$

$$d_t = e_{1t}^2 - e_{2t}^2 \tag{19}$$

As $E(d_t) = \mu$, we set $\mu = 0$ to study the empirical size and $\mu = 1$ to study the empirical power. Notice that the variance of d_t depends on μ, but the covariances of d_t and d_s $(s \neq t)$ do not. Results in Table 2 support the key conclusions that we derived in Section 3. In particular, in the size exercise, the distortion increases with ϕ and with the bandwidth m, and in the power exercise the power drops as ϕ increases from 0 to 0.85. Although it is not possible to directly compare the outcomes for $\mu = 0$ and $\mu = 1$, because the long run variance is slightly different, it seems fair to conclude that any apparent increase in power for larger values of ϕ is in fact due to the size distortion.

5 Empirical application

To illustrate the problems associated with the DM test when there is dependence in the loss differential, in this section we present the case in which a forecast for inflation with weakly dependent forecast errors is compared to two naive benchmarks. In particular, we consider quarterly predictions for the inflation rate in the Euro Area from a standard AR(1) model: for applications in the Euro Area that consider this model see Forni, Hallin, Lippi and Reichlin (2003) and Marcellino, Stock and Watson (2003). As for the benchmarks, we consider a constant 2% prediction (that represents the inflation target in the Euro Area) and a moving average prediction.

We use data on the Harmonized Index of Consumer Prices from the FRED database, and we compute quarterly year on year inflation rate from 2000.Q1 to 2020.Q4. All coefficients are estimated using a rolling window of 10 years. We compute predictions for horizons from 1 quarter to 8 quarters-ahead, and we evaluate them on the period from 2010.Q1 to 2020.Q4 (44 observations) using a quadratic loss function.

In Table 3, we report summary statistics for the forecast errors (defined as the realised value minus the prediction) for the AR(1) and the two benchmark predictions. The forecast errors are all negative on average, implying that in this period the inflation in the Euro Area has been lower than predicted by the AR(1) and the benchmarks. This result is not generated by a few large negative errors, as also the median forecast errors are negative for all the forecasts and the forecasting horizons.

The average and median forecast errors for the AR(1) increase (in absolute value) with the forecast horizon,

Table 3: Summary statistics of forecast errors

AR(1) forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4
1	-0.059	-0.077	0.367	0.284	0.264	0.182	0.179
2	-0.100	-0.092	0.590	0.690	0.384	0.350	0.313
3	-0.148	-0.135	0.724	0.817	0.627	0.433	0.397
4	-0.204	-0.134	0.827	0.860	0.697	0.569	0.424
5	-0.252	-0.314	0.904	0.882	0.733	0.603	0.456
6	-0.317	-0.182	0.983	0.898	0.760	0.577	0.390
7	-0.378	-0.288	1.046	0.907	0.738	0.551	0.354
8	-0.419	-0.420	1.078	0.881	0.717	0.535	0.366

MA forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4
1	-0.505	-0.411	0.855	0.912	0.787	0.639	0.491
2	-0.525	-0.428	0.871	0.914	0.789	0.641	0.493
3	-0.546	-0.413	0.882	0.916	0.792	0.644	0.497
4	-0.566	-0.409	0.890	0.917	0.794	0.648	0.503
5	-0.585	-0.427	0.895	0.917	0.795	0.653	0.509
6	-0.605	-0.438	0.896	0.919	0.799	0.658	0.517
7	-0.624	-0.454	0.895	0.919	0.802	0.665	0.526
8	-0.642	-0.475	0.893	0.920	0.806	0.673	0.536

2% forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4
1-8	-0.765	-0.673	0.923	0.926	0.821	0.695	0.569

Note: this table reports summary statistics for forecast errors from AR(1) predictions (top panel), moving average predictions (middle panel) and constant 2% predictions. Forecast errors are defined as the realised value minus the prediction.

but they remain lower than the ones of the two benchmarks for all the forecasting horizons. The standard deviations of the AR(1) forecast errors also increase with the forecasting horizon, and they are smaller than the ones of the two benchmarks for forecasting horizons up to 4 quarters. Another difference between the AR(1) predictions and the two benchmarks comes from the sample autocorrelations of the forecast errors. The AR(1) predictions have forecast errors that are only mildly correlated for short forecasting horizons; as we increase the forecasting horizon the autocorrelations increase, but they still decay quite quickly. On the other hand, the two benchmarks display large and persistent autocorrelations in their forecast errors at all forecasting horizons. This indicates that while the AR(1) predictions approximate optimal forecasts under MSE loss, clearly the two benchmarks fail to do so as their forecast errors are serially correlated, and thus predictable. Overall, looking at these summary statistics, we can see how AR(1) predictions for horizons up to 4 quarters ahead are more accurate, with lowers absolute means, medians, and standard deviations, and

Table 4: Summary statistics loss differential

Benchmark: MA

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1	0.835	0.261	1.152	0.852	0.652	0.499	0.385	-1.015
2	0.667	0.067	1.181	0.826	0.647	0.504	0.382	-1.480
3	0.525	-0.001	1.127	0.848	0.675	0.501	0.348	-1.490
4	0.384	-0.003	1.093	0.837	0.675	0.479	0.283	-1.819*
5	0.263	-0.020	1.046	0.843	0.644	0.440	0.233	-1.802*
6	0.105	-0.068	1.015	0.770	0.583	0.400	0.283	-2.616**
7	-0.041	-0.074	0.954	0.797	0.511	0.370	0.299	-3.258**
8	-0.121	-0.076	0.865	0.651	0.386	0.230	0.195	-3.032**

Benchmark: 2%

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1	1.283	0.449	1.573	0.862	0.635	0.437	0.320	-0.084
2	1.068	0.433	1.530	0.834	0.626	0.443	0.305	-0.529
3	0.885	0.335	1.389	0.851	0.643	0.440	0.269	-0.531
4	0.708	0.305	1.306	0.834	0.640	0.410	0.225	-1.263
5	0.556	0.121	1.255	0.846	0.634	0.414	0.191	-0.939
6	0.373	0.039	1.208	0.794	0.602	0.416	0.265	-1.943*
7	0.206	0.026	1.192	0.819	0.561	0.395	0.275	-2.295**
8	0.106	0.009	1.161	0.736	0.474	0.285	0.185	-2.046**

Note: this table reports summary statistics for the AR(1) loss differential with respect to moving average predictions (top panel) and constant 2% predictions (bottom panel). The loss function is quadratic and the loss differential is computed as the loss of the benchmark minus the loss of the AR(1). ADF refers to the augmented Dickey–Fuller test and * and ** denote significance at 10% and 5% level.

less persistent forecast errors than the two benchmarks.

Summary statistics of the loss differentials reported in Table 4 show that at short horizons loss differential are positive, indicating that AR(1) predictions may be more accurate than the benchmarks. As the forecasting horizon increases, the average loss differential decreases, and for the MA benchmark it becomes negative, indicating that at 7 and 8 quarters ahead MA predictions might be more accurate than AR(1) predictions.

The table also shows that the loss differentials display strong autocorrelations, even at short forecasting horizons. This indicates that the properties of the loss differentials are heavily affected by the benchmark considered, and even with a forecast with weakly dependent forecast errors it is possible to have strong autocorrelation of the loss differential. In the last column of the table, we report the augmented Dickey–Fuller test statistic, which clearly indicates that at short forecasting horizons the null of unit root of the loss differential cannot be rejected. With these levels of dependence, the DM test statistic is going to be subject to the drawbacks described in Section 3.

Table 5 reports the outcome of the DM tests for the null of equal predictive ability of AR(1) predictions

Table 5: Forecast evaluation

Benchmark: MA

m \ Horizon	1	2	3	4	5	6	7	8
1	1.881	1.389	1.123	0.859	0.628	0.249	-0.105	-0.403
$\lfloor T^{1/4} \rfloor$	1.736	1.397	1.141	0.896	0.671	0.300	-0.135	-0.500
$\lfloor T^{1/3} \rfloor$	2.091*	1.641	1.300	0.981	0.711	0.297	-0.127	-0.437
$\lfloor T^{1/2} \rfloor$	2.752**	2.210**	1.744	1.300	0.929	0.395	-0.171	-0.595
$\lfloor T^{2/3} \rfloor$	3.665**	2.941**	2.360**	1.783*	1.269	0.530	-0.220	-0.725

Benchmark: 2%

m \ Horizon	1	2	3	4	5	6	7	8
1	3.524*	2.829	2.527	2.191	1.841	1.228	0.762	0.508
$\lfloor T^{1/4} \rfloor$	2.138*	1.913	1.709	1.518	1.335	1.083	0.750	0.452
$\lfloor T^{1/3} \rfloor$	2.564**	2.238*	1.925	1.592	1.294	0.916	0.546	0.303
$\lfloor T^{1/2} \rfloor$	3.240**	2.886**	2.496**	2.063*	1.687	1.232	0.743	0.407
$\lfloor T^{2/3} \rfloor$	4.150**	3.668**	3.256**	2.748**	2.271**	1.608	0.890	0.469

Note: this table reports the DM test statistic for the null of equal predictive ability the AR(1) predictions with respect to a moving average (top panel) and a constant 2% (bottom panel) benchmarks. A positive value of the test statistics denotes a larger loss for the benchmark. Long-run variances are computed using (3). The sample size is 44 and bandwidth values m of $\lfloor T^{1/4} \rfloor$, $\lfloor T^{1/3} \rfloor$, $\lfloor T^{1/2} \rfloor$ and $\lfloor T^{2/3} \rfloor$ are respectively 2, 3, 6 and 12. Critical value are obtained from (4). * and ** denote significance at 10% and 5% level.

with respect to a moving average and a constant 2% benchmarks. We consider tests in which the DM statistics are computed estimating the long-run variances as in (3) using the bandwidth values m of 1, $\lfloor T^{1/4} \rfloor$, $\lfloor T^{1/3} \rfloor$, $\lfloor T^{1/2} \rfloor$ and $\lfloor T^{2/3} \rfloor$, that for a sample of 44 observations are respectively 2, 3, 6 and 12. In all cases we used critical values from the corresponding t_{2m} distribution. Results in Table 5 are in line with our theory. In particular, we can see that for small m the DM test does not reject the null of equal predictive accuracy of AR(1) and MA predictions for any forecasting horizon; when we use a constant 2% benchmark, we reject the null of equal predictive accuracy only for the 1 quarter-ahead forecasting horizon at 10% level. These results are in contrast with the evidence reported in Table 3, and may be due to the fact that, as we saw in Table 4, the loss differential is highly autocorrelated at these forecasting horizons, which implies that the test has no power for small values of m. As we increase m, we see that we reject the null more often, possibly due to spurious significance in some cases.

These results highlight how applying the DM test when the loss differential is not weakly dependent may generate undesirable results. For example, if we choose the bandwidth recommended in Coroneo and Iacone (2020), which is $\lfloor T^{1/3} \rfloor$, at 5% level we are not able to reject the null of equal predictive ability of AR(1) and MA predictions for any forecasting horizons, despite the summary statistics in Table 3 clearly indicate that

AR(1) predictions are more accurate at short forecasting horizons.

6 Conclusion

In this paper, we have verified that the DM test may be seriously misleading in presence of strong correlation in the loss differential. Diebold (2015) mentions that "[o]f course forecasters may not achieve optimality, resulting in serially correlated, and indeed forecastable, forecast errors. But I(1) nonstationarity of forecast errors takes serial correlation to the extreme". This is certainly true. However, the DM test is often used against naive benchmarks, for which an I(1) forecast error may not be impossible (or, as we showed, an error with root close enough to 1, given the sample size). While this may be seen as an "abuse" of the DM test, it seems desirable that a test is robust to such abuse. Our results warn that this is not the case.

For comparing forecasts, the DM test is "the only game in town", as noted in Diebold (2015). However, one should also be aware that the game has its rules. In the empirical application, we used the DM test to compare AR(1) inflation forecasts to two naive benchmarks. Results indicate that using a quadratic loss the test fails exactly because the benchmark forecasts are not optimal under MSE loss, which is not a nice feature. This does not mean that one should not use the DM test. Rather, our work suggests that one should take the recommendation in Diebold (2015) to use diagnostic procedures to assess the validity of the assumption of weak dependence of the loss differential very seriously.

Appendix

We provide here a more detailed derivation of some of the results that we claimed in the paper, with accompanying regularity conditions when needed. When we establish bounds we occasionally use $C < \infty$ as a finite bound, not necessarily the same one in every case.

Assumption A1. *Let ε_t be independent and identically distributed (iid) random variables, with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$. Then, assume that $u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ is such that*

$$\left(\sum_{j=0}^{\infty} \psi_j\right)^2 > 0, \; \sum_{j=0}^{\infty} j^{1/2}|\psi_j| < \infty.$$

Denoting $g(\lambda)$ as the spectral density of u_t, Assumption A1 implies that $g(0) > 0$ and that $g(\lambda) < \infty$ uniformly in λ. Assumption A1 is sufficient to establish the functional central limit theorem (FCLT) for a stationary, weakly dependent linear process as in Phillips and Solo (1992). This in turn is sufficient to establish the limit in (8), as in Phillips (1987), and then (9) and (10) by the continuous mapping theorem.

Assumption A2. *Assume that*

$$\sum_{s=j}^{\infty} |\psi_s| < Cj^{-1-a} \text{ for } j \geq 1$$

for $a > 2$.

Notice that the condition in Assumption A2 implies the condition in A1. By summation by parts, for any n,

$$\sum_{j=0}^{n} j^{1/2}|\psi_j| = \sum_{j=1}^{n} j^{1/2}|\psi_j| = 1\sum_{j=1}^{n} |\psi_j| - (\sum_{j=1}^{n-1} ((j+1)^{1/2} - j^{1/2}) \sum_{s=j+1}^{n} |\psi_s|$$
$$\sum_{j=0}^{n} j^{1/2}|\psi_j| \leq \sum_{j=1}^{n} |\psi_j| + C\sum_{j=1}^{n-1} j^{-1/2} \sum_{s=j+1}^{n} |\psi_s| \leq \sum_{j=1}^{\infty} |\psi_j| + C\sum_{j=1}^{\infty} j^{-1/2} \sum_{s=j+1}^{\infty} |\psi_s|$$

so

$$\sum_{j=1}^{\infty} j^{1/2}|\psi_j| \leq \sum_{j=1}^{\infty} |\psi_j| + C\sum_{j=1}^{\infty} j^{-1/2} \sum_{s=j+1}^{\infty} |\psi_s| \leq C + C\sum_{j=1}^{\infty} j^{-1/2} j^{-1-\alpha} < C$$

To see that the reverse is not true, which means that Assumption A2 really strengthens A1, notice that $\psi_j = (j+1)^{-3/2-\eta}$ for $\eta > 0$ and suitably small meets Assumption A1 but not A2.

Moreover, denoting $\gamma_u(k) = Cov(u_t, u_{t+j})$ Assumption A2 also implies that $\sum_{k=0}^{\infty} k|\gamma_u(k)| < \infty$. To verify this, first notice that $\sum_{k=0}^{\infty} k|\gamma_u(k)| = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} k|\psi_j \psi_{j+k}|$. Another application of summation by parts gives

$$\sum_{k=1}^{n} k|\psi_{j+k}| = 1 \times \sum_{k=1}^{n} |\psi_{j+k}| - \{\sum_{k=1}^{n-1} ((k+1)-k) \sum_{s=k+1}^{n} |\psi_{j+s}|\} = \sum_{k=1}^{n} |\psi_{j+k}| - \sum_{k=1}^{n-1} \sum_{s=k+1}^{n} |\psi_{j+s}|$$
$$\sum_{k=1}^{\infty} k|\psi_{j+k}| \leq \sum_{k=1}^{\infty} |\psi_{j+k}| + \sum_{k=1}^{\infty} \sum_{s=k+1}^{\infty} |\psi_{j+k}| \leq Cj^{-1-a} + \sum_{k=1}^{\infty} (k+j)^{-1-a} \leq Cj^{-1-a} + Cj^{-a}$$

then

$$\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} k|\psi_j \psi_{j+k}| \leq C \sum_{j=0}^{\infty} (j^{-1-a} + j^{-a}) < C$$

Denoting $g(\lambda)$ the spectral density of u_t, $\sum_{k=0}^{\infty} k|\gamma_u(k)| < \infty$ is sufficient to establish that $g(\lambda)$ is differentiable in the neighbourhood of $\lambda = 0$ and, for $g(\lambda)' = \frac{\partial g(\lambda)}{\partial \lambda}$ then $g(\lambda)' = O(\lambda^{-1})$ as $\lambda \to 0$.

A Results for Subsection 3.1

All the results in this subsection are for d_t generated as in (5)-(7), with $c \leq 0$, under Assumption A1. Recall that $I(\lambda_j)$ is the periodogram if d_t.

Lemma 1.

$$I(\lambda_j) = O\left(j^{-2} T^2\right)$$

Proof. We first present the proof for $c < 0$.

Denoting $f(\lambda)$ as the spectral density of d_t, $g(\lambda)$ as the density of u_t, for

$$f^{\star}(\lambda) = \left|1 - \rho\, e^{-i\lambda}\right|^{-2} = \frac{1}{\upsilon^2 + 2\rho(1 - cos(\lambda))},$$

so that $f(\lambda) = f^{\star}(\lambda) g(\lambda)$, we use for $f^{\star}(\lambda)$ the same bound as in Giraitis and Phillips (2012): for $|\rho| < 1$, $\lambda \leq \pi$

$$f^{\star}(\lambda) \leq \frac{1}{\upsilon^2 + \rho\lambda^2/3} \tag{20}$$

where $\upsilon = 1 - \rho$ and notice that we dropped the reference to T in ρ_T to simplify the notation and to align it to Giraitis and Phillips (2012).

We follow closely the proof in Robinson (1995b), but our proof is easier as we only need to establish an upper bound. Then

$$E\left(I(\lambda_j)\right) = \int_{-\pi}^{\pi} f(\lambda) K(\lambda - \lambda_j) d\lambda$$

where $K(\lambda)$ is proportional to the Fejér's kernel, $K(\lambda) = (2\pi T)^{-1} \left|\sum \sum_{t,s=1}^{T} e^{i(t-s)\lambda}\right|^2$.

Proceeding as in Robinson (1995b) we then partition the integral as

$$\int_{-\pi}^{\pi} = \int_{-\pi}^{-\lambda_j/2} + \int_{-\lambda_j/2}^{\lambda_j/2} + \int_{\lambda_j/2}^{2\lambda_j} + \int_{2\lambda_j}^{\pi}$$

and discuss them separately.

$$\int_{-\pi}^{-\lambda_j/2} f(\lambda)K(\lambda-\lambda_j)d\lambda \leq C\ \{sup_{\lambda\in[\lambda_j/2,\pi]} f^{\star}(\lambda)\} \int_{\lambda_j/2}^{\pi} K(\lambda+\lambda_j)d\lambda \leq C\ \lambda_j^{-2}\ j^{-1} = O(j^{-3}T^2)$$

where we used the bounds $g(\lambda) \leq C$, $sup_{\lambda\in[\lambda_j/2,\pi]} f^{\star}(\lambda) \leq C\lambda_j^{-2}$ from (20) and $\int_{\lambda_j/2}^{\pi} K(\lambda+\lambda_j)d\lambda = O(j^{-1})$ as in Robinson (1995b), page 1061 in the text above (4.6); the bound $\int_{2\lambda_j}^{\pi} = O(j^{-3}T^2)$ can be established in the same way. Next,

$$\int_{-\lambda_j/2}^{\lambda_j/2} f(\lambda)K(\lambda-\lambda_j)d\lambda \leq \int_{-\lambda_j/2}^{\lambda_j/2} f(\lambda)d\lambda\ \{sup_{\lambda\in[-\lambda_j/2,\lambda_j/2]} K(\lambda-\lambda_j)\} = O(T\times T^{-1}\lambda_j^{-2}) = O(T^2 j^{-2})$$

where we bounded $\{sup_{\lambda\in[-\lambda_j/2,\lambda_j/2]} K(\lambda-\lambda_j)\} = O(T^{-1}\lambda_j^{-2})$ as in Robinson (1995b) and $\int_{-\lambda_j/2}^{\lambda_j/2} f(\lambda)d\lambda \leq Var(y_t) = O(T)$. Finally,

$$\int_{\lambda_j/2}^{2\lambda_j} f(\lambda)K(\lambda-\lambda_j)d\lambda \leq C\{\sup_{\lambda\in[\lambda_j/2,2\lambda_j]} f^{\star}(\lambda)\} \int_{\lambda_j/2}^{2\lambda_j} K(\lambda-\lambda_j)d\lambda\ = O(\lambda_j^{-2})$$

where we bounded $\int_{\lambda_j/2}^{2\lambda_j} K(\lambda-\lambda_j)d\lambda = O(1)$. This completes the proof for $c < 0$.

When $c = 0$ we rewrite, as in Lemma A.1 of Phillips and Shimotsu (2004) for the unit root case,

$$(1-e^{i\lambda})w(\lambda) = w_u(\lambda) - \frac{e^{i\lambda}}{\sqrt{2\pi T}}(e^{iT\lambda}y_T - y_0)$$

where $w_u(\lambda)$ is the Fourier transform of u_t. Thus, bounding $E|w_u(\lambda)| = O((|w_u(\lambda)|^2)^{1/2}) = O(1)$, $E|y_T| = O((y_T^2)^{1/2}) = O(T^{1/2})$, $|(1-e^{i\lambda})|^{-2} < C\lambda^{-2}$, the result follows immediately. □

Lemma 2.

$$\frac{1}{T^2}2\pi\sum_{j=1}^{m} I(\lambda_j) \Rightarrow \omega^2\frac{1}{2}\int_0^1 (J_c(r)-\overline{J_c})^2 dr\ \ as\ m\to\infty,\ m/T\to 0$$

Proof. We rewrite

$$\frac{1}{T^2}2\pi\sum_{j=1}^{m} I(\lambda_j) = \frac{1}{T^2}2\pi\sum_{j=1}^{T} I(\lambda_j) - \frac{1}{T^2}2\pi\sum_{j=m+1}^{T} I(\lambda_j)$$

and notice that

$$\frac{1}{T^2}2\pi\sum_{j=1}^{T}I(\lambda_j)=\frac{1}{T^2}\frac{1}{2}\sum_{j=m+1}^{T}(y_t-\overline{y})^2\Rightarrow\omega^2\frac{1}{2}\int_0^1(J_c(r)-\overline{J_c})^2dr$$

using (9) and (10), while

$$\frac{1}{T^2}2\pi\sum_{j=m+1}^{T}I(\lambda_j)=O_p(T^{-2}m^{-1}T^2)=O_p(m^{-1})=o_p(1)$$

using Lemma 1 (notice that that result is not restricted to a band of frequencies degenerating to 0). The result when $c=0$ can be deduced from Marinucci and Robinson (2001): their moments condition is stronger and their proof is more complex than the argument given here because they established more results. □

Theorem 1. *For d_t generated as in (5)-(7), with $\mu=E(d_t)=0$ and under Assumption A1,*

$$\frac{1}{\sqrt{m}}DM\Rightarrow\frac{\int_0^1J_c(r)dr}{\sqrt{\frac{1}{2}\int_0^1(J_c(r)-\overline{J_c})^2dr}}\ \textit{as } m\to\infty,\ m/T\to0 \tag{21}$$

Proof. This follows from the convergence in (8), Lemma 2 and the continuous mapping theorem. □

Theorem 2. *For d_t generated as in (5)-(7), with $\mu=E(d_t)$ and under Assumption A1, for m fixed as $T\to\infty$ then (12) holds.*

Proof. We discuss the case $c=0$ as this is easier to follow and the notation is lighter, the generic $c<0$ proceeding in the same way but with different notation. Then,

$$DM=\sqrt{T}\frac{\overline{y}+\mu}{\sqrt{\frac{1}{m}2\pi\sum_{j=1}^{m}I(\lambda_j)}}=\frac{\frac{\sqrt{T}}{T}(\overline{y}+\mu)}{\sqrt{\frac{1}{T^2}\frac{1}{m}2\pi\sum_{j=1}^{m}I(\lambda_j)}}$$

where in particular notice that $\frac{\sqrt{T}}{T}\mu\to0$ so $\frac{\sqrt{T}}{T}(\overline{y}+\mu)\Rightarrow\omega\int_0^1W(r)dr$ by a standard FCLT even when $\mu\neq0$ and $\frac{1}{T^2}\frac{1}{m}2\pi\sum_{j=1}^{m}I(\lambda_j)\Rightarrow\omega^2\frac{1}{m}2\pi\sum_{j=1}^{m}Q_W(j)$ using the same argument as in Lemma 1 of Hualde and Iacone (2017). □

B Results for Subsection 3.2

All the results in this subsection are for d_t generated as in (5), (6), (13), with $c<0$, under Assumption A2.

Lemma 3. *As* $T \to \infty$

$$\frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} (d_t - \overline{d})^2 = \frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} (d_t - \mu)^2 + o_p(1) \to_p \frac{\omega^2}{-2c} \tag{22}$$

Proof. The limit

$$\frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} (d_t - \mu)^2 \to_p \frac{\omega^2}{-2c}$$

is already in Phillips and Magdalinos (2007a); it can also be derived from (2.13) and (2.16) of Giraitis and Phillips (2012), setting $\upsilon = 1 - \rho_T = cT^{-\alpha}$ and taking the limit for $T \to \infty$.
Next, rewriting $\sum_{t=1}^{T} (d_t - \overline{d})^2 = \sum_{t=1}^{T} (d_t - \mu)^2 - T(\overline{d} - \mu)^2$, from the CLT on Theorem 2.1 of Giraitis and Phillips (2012), $(\overline{d} - \mu)^2 = O_p\left(T^{2\alpha - 1}\right)$, so

$$\frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} (d_t - \overline{d})^2 - \frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} (d_t - \mu)^2 = O_p(T^{-1-\alpha}\ T\ T^{2\alpha-1}) = O_p(T^{\alpha-1}).$$

The lemma is established as we combine these results. □

Lemma 4.

$$I\left(\lambda_j\right) = O\left(j^{-2}T^2\right)$$

Proof. The proof follows as in Lemma 1, but using the bound $\int_{-\pi}^{\pi} f(\lambda)d\lambda = O(T^{\alpha})$. □

Lemma 5. *For frequencies* λ_j *such that* $j < m$, $m/T \to 0$, $mT^{\alpha-1} \to 0$ *as* $T \to \infty$,

$$f(\lambda_j)^{-1} I\left(\lambda_j\right) = 1 + O(j^{-1} ln(j+1))$$

Proof. We consider

$$f(\lambda_j)^{-1} E\left(I\left(\lambda_j\right)\right) - 1 = f(\lambda_j)^{-1} \int_{-\pi}^{\pi} \left(f(\lambda) - f(\lambda_j)\right) K(\lambda - \lambda_j) d\lambda$$

and again we evaluate this integral over subsets of $(-\pi, \pi)$ as in Lemma 1. Then,

$$f(\lambda_j)^{-1} \int_{-\pi}^{-\lambda_j/2} \left(f(\lambda) - f(\lambda_j)\right) K(\lambda - \lambda_j) d\lambda \leq C\ f^{\star}(\lambda_j)^{-1} \{sup_{\lambda \in [\lambda_j/2, \pi]} f^{\star}(\lambda)\} \int_{\lambda_j/2}^{\pi} K(\lambda + \lambda_j) d\lambda \tag{23}$$

where again we used $\{sup_{\lambda \in [\lambda_j/2,\pi]} f(\lambda)\} \leq \{sup_{\lambda \in [\lambda_j/2,\pi]} g(\lambda)\}\{sup_{\lambda \in [\lambda_j/2,\pi]} f^{\star}(\lambda)\}$ and $g(\lambda) < C$ uniformly in λ and $g(\lambda_j)^{-1} < C$. Recalling that $f^{\star}(\lambda_j)^{-1} = \upsilon^2 + 2\rho(1 - cos(\lambda_j))$, we bound $f^{\star}(\lambda_j)^{-1} \leq \upsilon^2 +$

$C(sin(\lambda_j/2))^2 \leq \upsilon^2 + C(\lambda_j/2)^2 \leq \upsilon^2 + C\lambda_j^2$ where we used $sin(\lambda_j/2) \sim \lambda_j/2$ as $j/T \to 0$. Then,

$$f^\star(\lambda_j)^{-1}\{sup_{\lambda\in[\lambda_j/2,\pi]}f^\star(\lambda)\} \leq (\upsilon^2 + C\lambda_j^2)\{sup_{\lambda\in[\lambda_j/2,\pi]}f^\star(\lambda)\}$$
$$\leq \upsilon^2\{sup_{\lambda\in[\lambda_j/2,\pi]}f^\star(\lambda)\} + C\lambda_j^2\{sup_{\lambda\in[\lambda_j/2,\pi]}f^\star(\lambda)\} \leq \upsilon^2\upsilon^{-2} + C\lambda_j^2\lambda_j^{-2} \leq C$$

using (20). Thus, recalling $\int_{\lambda_j/2}^{\pi} K(\lambda + \lambda_j)d\lambda = O(j^{-1})$, then (23) $= O(j^{-1})$. The bound $f(\lambda_j)^{-1}\int_{2\lambda_j}^{\pi} =$ $O(j^{-1})$ can be established in the same way.

Next,

$$f(\lambda_j)^{-1}\int_{-\lambda_j/2}^{\lambda_j/2}(f(\lambda) - f(\lambda_j))\,K(\lambda-\lambda_j)d\lambda \leq C\,f^\star(\lambda_j)^{-1}\{sup_{\lambda\in[-\lambda_j/2,\lambda_j/2]}K(\lambda-\lambda_j)\}\int_{-\lambda_j/2}^{\lambda_j/2}(f(\lambda) - f(\lambda_j))\,d\lambda \tag{24}$$

where again we bound $sup_{\lambda\in[-\lambda_j/2,\lambda_j/2]}K(\lambda - \lambda_j)\} = O(T^{-1}\lambda_j^{-2})$; moreover,

$$\int_{-\lambda_j/2}^{\lambda_j/2}(f(\lambda) - f(\lambda_j))\,d\lambda = \int_{-\lambda_j/2}^{\lambda_j/2}(f(\lambda) - f(0) + f(0) - f(\lambda_j))\,d\lambda \leq 2\int_0^{\lambda_j/2}|f(\lambda) - f(0)|\,d\lambda + |f(0) - f(\lambda_j)|\,\lambda_j$$

From Giraitis and Phillips (2012), page 175,

$$|f(\lambda) - f(0)| \leq \lambda^2 f^\star(\lambda)\upsilon^{-2} + \lambda^2\upsilon^{-2}$$

so

$$2\int_0^{\lambda_j/2}|f(\lambda) - f(0)|\,d\lambda + |f(0) - f(\lambda_j)|\,\lambda_j \leq C\lambda_j\upsilon^{-2}$$

and (24) is bounded by

$$CT^{-1}\lambda_j^{-2}(\upsilon^2 + \lambda_j^2)\lambda_j\upsilon^{-2} \leq C(T^{-1}\lambda_j^{-1} + T^{-1}\lambda_j\upsilon^{-2}) = O(j^{-1} + j^{-1}(jT^{\alpha-1})^2)$$

where the last bound is $o(j^{-1})$ because $j \leq m$ and $mT^{\alpha-1} \to 0$.

For the next integral we introduce $f^\star(\lambda)' = \frac{\partial f^\star(\lambda)}{\partial\lambda}$, where $f^\star(\lambda)' = -(f^\star(\lambda))^2 2\rho\sin(\lambda)$ and rewrite

$$f(\lambda) - f(\lambda_j) = f^\star(\lambda)g(\lambda) + f^\star(\lambda_j)g(\lambda) - f^\star(\lambda_j)g(\lambda) - f^\star(\lambda_j)g(\lambda_1)$$

then

$$
\begin{aligned}
&f(\lambda_j)^{-1}\int_{\lambda_j/2}^{2\lambda_j}\left(f(\lambda)-f(\lambda_j)\right)K(\lambda-\lambda_j)d\lambda \\
&\leq Cf^{\star}(\lambda_j)^{-1}\{\sup_{\lambda\in[\lambda_j/2,2\lambda_j]}|f^{\star}(\lambda)'|\}\int_{\lambda_j/2}^{2\lambda_j}|\lambda-\lambda_j|\;K(\lambda-\lambda_j)d\lambda \qquad (25)\\
&+Cf^{\star}(\lambda_j)^{-1}f^{\star}(\lambda_j)\{\sup_{\lambda\in[\lambda_j/2,2\lambda_j]}|g(\lambda)'|\}\int_{\lambda_j/2}^{2\lambda_j}|\lambda-\lambda_j|\;K(\lambda-\lambda_j)d\lambda \qquad (26)
\end{aligned}
$$

where notice that

$$
\begin{aligned}
&\{\sup_{\lambda\in[\lambda_j/2,2\lambda_j]}|f^{\star}(\lambda)'|\}\leq Cf^{\star}(\lambda_j)^2\lambda_j,\\
&\{\sup_{\lambda\in[\lambda_j/2,2\lambda_j]}|g(\lambda)'|\}\leq C\lambda_j^{-1}
\end{aligned}
$$

The bound in (25) is

$$
\begin{aligned}
&Cf^{\star}(\lambda_j)\lambda_j\int_{\lambda_j/2}^{2\lambda_j}|\lambda-\lambda_j|\;K(\lambda-\lambda_j)d\lambda\;\leq C\lambda_j^{-1}\int_{\lambda_j/2}^{2\lambda_j}|\lambda-\lambda_j|\;K(\lambda-\lambda_j)d\lambda\\
&=O\left((j/T)^{-1}T^{-1}\ln(j+1)\right)=O(j^{-1}\ln(j+1))
\end{aligned}
$$

where we used

$$
\int_{\lambda_j/2}^{2\lambda_j}|\lambda-\lambda_j|\;K(\lambda-\lambda_j)d\lambda=O\left(T^{-1}\ln(j+1)\right)
$$

as in Robinson (1995b); proceeding in the same way, the bound in (26) is

$$
C\lambda_j^{-1}\int_{\lambda_j/2}^{2\lambda_j}|\lambda-\lambda_j|\;K(\lambda-\lambda_j)d\lambda\;=O(j^{-1}\ln(j+1)).
$$

□

Lemma 6. *For frequencies λ_j such that $j<m$, $m/T\to 0$, $mT^{\alpha-1}\to 0$ as $T\to\infty$,*

$$
\frac{1}{m}\sum_{j=1}^{m}(f\left(\lambda_j\right)^{-1}I\left(\lambda_j\right)-1)=o_p(1)
$$

Proof. The proof follows as on pages 1636-1638 of Robinson (1995a) using the bound from Lemma 5. The other bounds in (3.17) of Robinson (1995a) can be computed adapting arguments in Theorem 2 of Robinson (1995b) as we did for Lemma 5. □

Theorem 3. *For $m \to \infty$, $m/T \to 0$ as $T \to \infty$,*

$$\sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}} \to_d N(0,1) \quad mT^{\alpha-1} \to 0$$
$$(mT^{\alpha-1})^{-1/2} 1/2 (-c)^{1/2} \left\{ \sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}} \right\} \to_d N(0,1) \quad mT^{\alpha-1} \to \infty$$

for any $\alpha \in (0,1)$ where as usual $\widehat{\sigma}^2$ is the Daniell estimate of σ^2 as in (3) .

Proof. We rewrite $\sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}}$ as

$$\sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}} = \frac{\sqrt{(2\pi f(0))^{-1}}\sqrt{T}(\overline{d}-\mu)}{\sqrt{(2\pi f(0))^{-1}}\sqrt{\frac{2\pi}{m}\sum_{j=1}^{m} I(\lambda_j)}}$$

and recall $2\pi f(0) = \upsilon^2\omega^2 = (-c)^{-2}T^{2\alpha}\omega^2$.

From Theorem 2.1 of Giraitis and Phillips (2012), $\sqrt{(2\pi f(0))^{-1}}\sqrt{T}(\overline{d}-\mu) \to_d N(0,1)$.

As for the denominator, we discuss the cases $mT^{\alpha-1} \to 0$ and $mT^{\alpha-1} \to \infty$ separately.

We begin with $mT^{\alpha-1} \to 0$.

We rewrite the argument of the square root of the denominator as

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m} I(\lambda_j) = f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m} f(\lambda_j)^{-1}I(\lambda_j)f(\lambda_j) = f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}(f(\lambda_j)^{-1}I(\lambda_j) - 1 + 1)f(\lambda_j)$$
$$= f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}(f(\lambda_j)^{-1}I(\lambda_j) - 1)f(\lambda_j) + f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m} f(\lambda_j)$$

and

$$|f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}(f(\lambda_j)^{-1}I(\lambda_j) - 1)f(\lambda_j)|$$
$$\leq f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1}|f(\lambda_j) - f(\lambda_{j+1})||\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k) - 1)| + f(0)^{-1}f(\lambda_m)\frac{1}{m}|\sum_{j=1}^{m}(f(\lambda_j)^{-1}I(\lambda_j) - 1)| \quad (27)$$

The first term of (27) has the same order as

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1}|f^{\star}(\lambda_j)'|\frac{1}{T}|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k) - 1)| + f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1} f^{\star}(\lambda_j)|g(\lambda_j)'|\frac{1}{T}|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k) - 1)| \quad (28)$$

using the differentiability of $g(\lambda)$, and the first element in (28) has order

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1}|(f^{\star}(\lambda_j))'|\frac{1}{T}j\ j^{-1}|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k)-1)|$$
$$\leq Cf^{\star}(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1}f^{\star}(0)\lambda_j^{-2}\lambda_j\frac{1}{T}j\ j^{-1}|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k)-1)| = o_p(\frac{1}{m}\sum_{j=1}^{m-1}\lambda_j^{-1}\frac{1}{T}j) = o_p(1)$$

where we used the bound $|f'^{\star}(\lambda_j)'| \leq Cf^{\star}(\lambda_j)^2\lambda_j \leq Cf^{\star}(0)\lambda_j^{-2}\lambda_j$ and Lemma 6; the second element in (28) has order

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1}f^{\star}(\lambda_j)|g(\lambda_j)'|\ j\ j^{-1}\frac{1}{T}|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k)-1)|$$
$$\leq C\frac{1}{m}\sum_{j=1}^{m-1}\lambda_j^{-1}\frac{1}{T}j\ j^{-1}|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k)-1)| = o_p(\frac{1}{m}\sum_{j=1}^{m-1}\lambda_j^{-1}\frac{1}{T}j) = o_p(1).$$

The last term in (27) is also $o_p(1)$.

To complete the discussion of the denominator when $m\,T^{\alpha-1}\to 0$, we need to consider the term $f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}f(\lambda_j)$. This is

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}f(\lambda_j) = f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}(f(\lambda_j)-f(0))+1$$

and

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}|f(\lambda_j)-f(0)| \leq Cf(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}|\lambda_j^2 f^{\star}(\lambda_j)\upsilon^{-2}| + Cf(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}|\lambda_j^2\upsilon^{-2}| = O(T^{2\alpha-2}m^2) = o(1),$$

Combining these results,

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}I(\lambda_j)\to_p 1,$$

completing the discussion for the case $T^{\alpha-1}\ m\to 0$.

For the $T^{\alpha-1}\ m\to\infty$, again we only need to consider the denominator. The argument of the square root in this case, is

$$\{(-c)^2T^{-2\alpha}\omega^2\}^{-1}\frac{2\pi}{m}\sum_{j=1}^{m}I(\lambda_j)$$

and notice that

$$\frac{2\pi}{T\ T^{\alpha}}\sum_{j=1}^{m}I(\lambda_j) = \frac{1}{2}\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}(d_t-\overline{d})^2 - \frac{2\pi}{T\ T^{\alpha}}\sum_{j=m+1}^{T/2}I(\lambda_j)$$

From Lemma 4,

$$\frac{2\pi}{T\ T^{\alpha}} \sum_{j=m+1}^{T/2} I(\lambda_j) = O_p\left(\frac{1}{T\ T^{\alpha}} \sum_{j=m+1}^{T/2} \lambda_j^{-2}\right) = O_p\left(T^{1-\alpha} m^{-1}\right) = o_p(1) \tag{29}$$

The application of Lemma 3 for $\frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} (d_t - \overline{d})^2$ completes the proof. $\square$

References

Clark, Todd E (1999) 'Finite-sample properties of tests for equal forecast accuracy.' *Journal of Forecasting* 18(7), 489–504

Coroneo, Laura, and Fabrizio Iacone (2020) 'Comparing predictive accuracy in small samples using fixed-smoothing asymptotics.' *Journal of Applied Econometrics* 35(4), 391–409

Corradi, Valentina, Norman R Swanson, and Claudia Olivetti (2001) 'Predictive ability with cointegrated variables.' *Journal of Econometrics* 104(2), 315–358

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

Diebold, Francis X, and Roberto S Mariano (1995) 'Comparing predictive accuracy.' *Journal of Business & Economic Statistics* pp. 253–263

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin (2003) 'Do financial variables help forecasting inflation and real activity in the euro area?' *Journal of Monetary Economics* 50(6), 1243–1255

Giacomini, Raffaella, and Barbara Rossi (2010) 'Forecast comparisons in unstable environments.' *Journal of Applied Econometrics* 25(4), 595–620

Giacomini, Raffaella, and Halbert White (2006) 'Tests of conditional predictive ability.' *Econometrica* 74(6), 1545–1578

Giraitis, Liudas, and Peter C.B. Phillips (2012) 'Mean and autocovariance function estimation near the boundary of stationarity.' *Journal of Econometrics* 169(2), 166–178. Recent Advances in Nonstationary Time Series: A Festschrift in honor of Peter C.B. Phillips

Hualde, Javier, and Fabrizio Iacone (2017) 'Fixed bandwidth asymptotics for the studentized mean of fractionally integrated processes.' *Economics Letters* 150, 39–43

Kiefer, Nicholas M, and Timothy J Vogelsang (2005) 'A new asymptotic theory for heteroskedasticity-autocorrelation robust tests.' *Econometric Theory* 21(6), 1130–1164

Kruse, Robinson, Christian Leschinski, and Michael Will (2019) 'Comparing predictive accuracy under long memory, with an application to volatility forecasting.' *Journal of Financial Econometrics* 17(2), 180–228

Lazarus, Eben, Daniel J Lewis, James H Stock, and Mark W Watson (2018) 'Har inference: Recommendations for practice.' *Journal of Business & Economic Statistics* 36(4), 541–559

Marcellino, Massimiliano, James H Stock, and Mark W Watson (2003) 'Macroeconomic forecasting in the euro area: Country specific versus area-wide information.' *European Economic Review* 47(1), 1–18

Marinucci, D., and P. M. Robinson (2001) 'Narrow-band analysis of nonstationary processes.' *The Annals of Statistics* 29(4), 947 – 986

McCracken, Michael W (2020) 'Diverging tests of equal predictive ability.' *Econometrica* 88(4), 1753–1754

McElroy, T., and D.N. Politis (2012) 'Fixed-b asymptotics for the studentized mean from time series with short, long, or negative memory.' *Econometric Theory* 28(2), 471–481. cited By 13

Müller, Ulrich K. (2007) 'A theory of robust long-run variance estimation.' *Journal of Econometrics* 141(2), 1331–1352

Patton, Andrew J, and Allan Timmermann (2007) 'Properties of optimal forecasts under asymmetric loss and nonlinearity.' *Journal of Econometrics* 140(2), 884–918

Phillips, P. C. B. (1987) 'Towards a unified asymptotic theory for autoregression.' *Biometrika* 74(3), 535–547

Phillips, P.C.B., and T. Magdalinos (2007a) 'Limit theory for moderate deviations from a unit root under weak dependence.' *In: Phillips, G.D.A., Tzavalis, E. (Eds.), The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Cambridge University Press, Cambridge*

Phillips, Peter C. B., and Katsumi Shimotsu (2004) 'Local Whittle estimation in nonstationary and unit root cases.' *The Annals of Statistics* 32(2), 656 – 692

Phillips, Peter C. B., and Victor Solo (1992) 'Asymptotics for Linear Processes.' *The Annals of Statistics* 20(2), 971 – 1001

Phillips, Peter C.B., and Tassos Magdalinos (2007b) 'Limit theory for moderate deviations from a unit root.' *Journal of Econometrics* 136(1), 115–130

Robinson, P. M. (1995a) 'Gaussian Semiparametric Estimation of Long Range Dependence.' *The Annals of Statistics* 23(5), 1630 – 1661

—— (1995b) 'Log-Periodogram Regression of Time Series with Long Range Dependence.' *The Annals of Statistics* 23(3), 1048 – 1072

Rossi, Barbara (2005) 'Testing long-horizon predictive ability with high persistence, and the meese–rogoff puzzle.' *International Economic Review* 46(1), 61–92

Shao, Xiaofeng (2015) 'Self-normalization for time series: A review of recent developments.' *Journal of the American Statistical Association* 110(512), 1797–1817

Sun, Yixiao (2014a) 'Fixed-smoothing asymptotics and asymptotic: F: and: t: Tests in the presence of strong autocorrelation.' In 'Essays in Honor of Peter C. B. Phillips,' vol. 33 (Emerald Publishing Ltd) pp. 23–63

—— (2014b) 'Fixed-smoothing asymptotics in a two-step generalized method of moments framework.' *Econometrica* 82(6), 2327–2370

—— (2014c) 'Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference.' *Journal of Econometrics* 178, 659–677