Article:
Coroneo, Laura and Iacone, Fabrizio (Accepted: 2024) Testing for equal predictive accuracy with strong dependence. International journal of forecasting. ISSN: 0169-2070 (In Press)

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

20th September 2024

Abstract

We analyse the properties of the Diebold and Mariano (1995) test in the presence of autocorrelation in the loss differential. 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. We also 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 loss differential before the application of the Diebold and Mariano (1995) test.

JEL classification codes: C12; C32; C53.

Keywords: strong autocorrelation, forecast evaluation, Diebold and Mariano test.

Total Words: 5,744

*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. The authors thank the Editor, Esther Ruiz, the associate editor and two anonymous referees for their constructive and insightful comments, which helped improve the paper substantially. We also thank Roberto Casarin, Gergley Gánics, Raffaella Giacomini, Liudas Giraitis, Anders Kock, Andrey Vasnev, and participants to the 7th RCEA Time Series Workshop (University of Milan Bicocca, 2021), to the 42nd International Symposium on Forecasting (ISF 2022 - University of Oxford), and the Bergamo Workshop of Econometrics and Statistics for useful comments.

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 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 asymptotically 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. Thus if the considered forecast approximates the optimal forecast, 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 (see Patton and Timmermann 2007) or it 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 standard reference 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 of

the loss differential does not hold. We characterise strong dependence as local to unity as in Phillips (1987) and Phillips and Magdalinos (2007a). 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 in economics. 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 of the loss differential before the application of the DM test, especially when naive benchmarks are considered. In this respect, a unit root test could be a valuable diagnostic for the preliminary detection of critical situations.

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 rolling 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.

In the literature, there has been some attention to the issue of forecast evaluation in presence of persistence. Corradi, Swanson and Olivetti (2001) examined the DM statistic in the presence of cointegration, whereas Rossi (2005) examined the effect of high persistence on the loss differential. McCracken (2020) provided 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; instead Clark (1999), Khalaf and Saunders (2017) and Kruse, Leschinski and Will (2019) took forecast errors as primitives. Clark (1999) and Khalaf and Saunders (2017), in particular, provided convincing simulation evidence that the DM test is incorrectly sized in the presence of dependence. Giacomini and White (2006) and Coroneo and Iacone (2020) showed that under benign forms of weak dependence, the correct size may be restored using bootstrap or fixed smoothing asymptotics, respectively, but results in Khalaf and Saunders (2017) suggest that even bootstrap is only a useful and reliable guidance when the depedence is not too strong, relative to the sample size. On the other hand Kruse et al. (2019) derived 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. However, this approach cannot be applied to moderate sample size, such as the ones typically encountered in macroeconomic forecasting.

Of course, the DM test can be seen as a particular application of the standard t-test on the mean in presence of dependence. A certain level of persistence can be accommodated for the t-test using bootstrap or alternative asymptotics, see for example Gonçalves and Vogelsang (2011) for the former, and Kiefer and Vogelsang (2005), Sun (2014c), Sun (2014b), Lazarus, Lewis, Stock and Watson (2018) for the latter. Following Müller (2014) or Giraitis and Phillips (2012), it is clear that the even the limit distribution in Kiefer and Vogelsang (2005) and related works does not provide a useful guidance when the persistence is too strong in relation to the sample size. Müller (2014) does provide an algorithm to perform a reliable test in that case, although the sample size required is larger than the samples that are usually available in forecast evaluation exercises. As a promising alternative option, Henzi and Ziegel (2021) and Choe and Ramdas (2023) recently proposed procedures based on sequential testing that do not require assumptions on the dependence of the process. These new procedures may therefore be robust even in situations of dependence, however the assumption of bounded loss differentials may limit their applicability.

As it is clear from this review of the literature, the core of the statistical results discussed here should not come as a surprise. However, their implication in the context of testing for equal predictive ability remains relevant, in particular when good forecasts are compared against poor benchmarks, with relevant autocorrelations, as our results suggest that in these cases the blind application of the DM test leads to incorrect conclusions. Even more importantly, it may be more difficult to reject the null hypothesis when good forecasts are compared to poor competitors, than when the same good forecasts are compared to competitors that are nearly as good. This perverse and undesirable feature of the DM test should be kept in mind in forecast evaluation exercises.

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 $t = \{1, \dots, T\}$, 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 Diebold and Mariano (1995) and also Clark (1999). With fixed smoothing asymptotics, the limit for $\widehat{\sigma}^2$ is derived under alternative asymptotics, accounting for the distribution of $\widehat{\sigma}^2$ and, as a consequence, 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 consider two cases: the weighted autocovariance estimate using the Bartlett kernel and the weighted periodogram estimate using the Daniell kernel.

Denoting by $\widehat{\gamma}_l$ the sample autocovariance of lag l, the weighted autocovariance estimate of the long run variance using the Bartlett kernel is

$$\widehat{\sigma}_A^2 = \widehat{\gamma}_0 + 2\sum_{l=1}^{M}\frac{M - l}{M}\widehat{\gamma}_l. \tag{3}$$

where M is a user-chosen bandwidth parameter, and

$$DM_A = \sqrt{T}\frac{\overline{d}}{\widehat{\sigma}_A}.$$

Under H_0,

$$DM_A \to_d N(0,1), \text{ if } 1/M + M/T \to 0 \text{ as } T \to \infty,$$
$$DM_A \to_d \Phi_A(b), \text{ if } M/T \to b \in (0,1] \text{ as } T \to \infty$$

where the distribution of $\Phi_A(b)$ depends on b; this is characterised in Kiefer and Vogelsang (2005), where relevant quantiles are also provided.

For the weighted periodogram estimate, 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, the Daniell weighted periodogram estimate is

$$\widehat{\sigma}_P^2 = \frac{2\pi}{m} \sum_{j=1}^{m} I(\lambda_j) \tag{4}$$

where, for integer j, $\lambda_j = \frac{2\pi j}{T}$ are the Fourier frequencies and m is a user-chosen bandwidth parameter. The test statistic is then given by

$$DM_P = \sqrt{T} \frac{\overline{d}}{\widehat{\sigma}_P}. \tag{5}$$

Under H_0,

$$DM_P \to_d N(0,1), \text{ if } 1/m + m/T \to 0 \text{ as } T \to \infty,$$
$$DM_P \to_d t_{2m}, \text{ if } m \text{ is fixed as } T \to \infty$$

where t_{2m} is the Student's t-distribution with $2m$ degrees of freedom, 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 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, in practice, it is not uncommon to have strong autocorrelation in the loss differential. This can happen in presence of forecasts that are optimal under alternative loss functions, or when the forecast evaluation sample T is short. In addition, 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. Section 4 contains an example illustrating how stochastically trending behavior may appear in the loss differentials.

Denoting $\mu = E(d_t)$ and $y_t = d_t - \mu$, so that

$$d_t = \mu + y_t, \tag{6}$$

we assume that

$$y_t = \rho_T y_{t-1} + u_t \tag{7}$$

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 (2007a). 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 Appendix A 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 (7)

$$\rho_T = e^{c/T} \quad \text{with } c \leq 0. \tag{8}$$

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 $y_0 = O_p(1)$.

Our model is completed by formalising the assumptions on u_t.

Assumption A1. *Let ε_t be independent and identically distributed (iid) random variables, with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \varsigma^2$. 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), see remark 3.5 of Phillips and Solo

(1992), and notice that condition $\sum_{j=0}^{\infty} j^{1/2}|\psi_j| < \infty$ implies (16) of Phillips and Solo (1992), as discussed on page 973.

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):

$$T^{-1/2} y_{\lfloor rT \rfloor} \to \omega J_c(r) \tag{9}$$

$$T^{-1/2} \overline{y} \to \omega \int_0^1 J_c(r) dr \tag{10}$$

$$T^{-2} \sum_{t=1}^{T} y_t^2 \to \omega^2 \int_0^1 J_c(r)^2 dr \tag{11}$$

where $J_c(r) = W(r)$ when $c = 0$. The limit in (9) follows proceeding as in Phillips (1987) but using the FCLT for linear processes instead of for mixing sequencies (also see Chan and Wei (1987)); (10) and (11) are then due to the continuous mapping theorem. The result for $\overline{y}$ in particular means that the sample average is not consistent in the neighbourhood of a unit root.

Denoting

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

we can now establish the limit properties of the DM_A and DM_P statistics.

Theorem 1. *For d_t generated as in (6)-(8), and under Assumption A1,*

Case A: For $1/M + M/T \to 0$ as $T \to \infty$

$$\frac{\sqrt{M}}{\sqrt{T}} DM_A \to_d \frac{\overline{J}_c}{\sqrt{\int_0^1 \left(J_c(r)^2 - \overline{J}_c^2 \right) dr}} \tag{12}$$

Case P: For $1/m + m/T \to 0$ as $T \to \infty$

$$\frac{1}{\sqrt{m}} DM_P \to_d \frac{\overline{J}_c}{\sqrt{\frac{1}{2}\int_0^1 \left(J_c(r)^2 - \overline{J_c}^2 \right) dr}} \tag{13}$$

We refer to Appendix A for a more detailed derivation of this and other results in this section. In view

of (12) or (13), as $M/T \to 0$ or $m \to \infty$ 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.

Next, we present the limit of the DM_A and DM_P statistics using fixed smoothing asymptotic. Denoting

$$\widetilde{J}_c(r) = J_c(r) - rJ_c(1)$$
$$Q_A(b) = \frac{2}{b}\int_0^1 \widetilde{J}_c(r)^2 dr - \frac{2}{b}\int_0^{1-b} \widetilde{J}_c(r)\widetilde{J}_c(r+b)dr$$
$$Q_P(j) = (2\pi j)^2 \left\{ \left(\int_0^1 sin(2\pi jr)\widetilde{J}_c(r)dr \right)^2 + \left(\int_0^1 cos(2\pi jr)\widetilde{J}_c(r)dr \right)^2 \right\}$$

Theorem 2. *For d_t generated as in (6)-(8), and under Assumption A1,*

Case A: For $M/T \to b \in (0,1]$ as $T \to \infty$,

$$DM_A \to_d \frac{\overline{J}_c}{\sqrt{Q_A(b)}} \tag{14}$$

Case P: For m fixed as $T \to \infty$ then

$$DM_P \to_d \frac{\overline{J_c}}{\sqrt{\frac{1}{m}\sum_{j=1}^m Q_P(j)}}. \tag{15}$$

When $c = 0$, so that the process y_t is characterised by a unit root, the limit distribution $\widetilde{J}_c(r)$ is replaced by $\widetilde{W}(r) = W(r) - rW(1)$ where $W(r)$ is a standard Brownian motion, and $\overline{J_c}$ is replaced by $\overline{W} = \int_0^1 W(r)dr$.

The limits in (14) and (15) exhibit the self-normalization property as in Shao (2015). It is interesting to compare the limits in Theorem 1 and in Theorem 2 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 (and an analogue result would hold when $M/T \to 0$), 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 the limits in Theorem 1 and in Theorem 2 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 (large M/T and small M/T). The limits in Theorem 1 and in Theorem 2 are only relevant in the sense that we can establish whether the test statistic is divergent or not.

Remark 1. *Results in in Theorem 1 and in Theorem 2 indicate that:*

a. *With relatively large values for m (or equivalently small M/T), the DM statistics diverges even under H_0 (spurious significance).*

b. *With small values for m (or equivalently large M/T), the DM statistics does not diverge even under H_1, so the test is not consistent. However, as the distribution in (15) has much thicker tails than a t_{2m} distribution (and similarly for the distribution in (14) with respect to the $\Phi_A(b)$ distribution), then it is still possible (and indeed it may be frequent) to have many spurious rejections of the null hypothesis.*

c. *The limits in Theorem 1 and in Theorem 2 hold regardless of whether $\mu = 0$ or $\mu \neq 0$, so they are not affected by whether the null hypothesis is correct or not. This means that the DM test cannot discriminate between null and alternative hypotheses in this case.*

3.2 Moderate deviations from unit root

As c in (8) 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 in Theorem 1 and in Theorem 2 may not provide a valuable guideline when ρ_T is not in fact in the very close neighbourhood of unity. For this situation, Phillips and Magdalinos (2007a) generalise ρ_T to moderate deviations from the unit root. We simplify the model slightly and consider

$$\rho_T = 1 + c/T^{\alpha} \text{ for } \alpha \in (0,1), \text{ and } c < 0. \tag{16}$$

Moderate deviations from the unit root following (16) are also discussed in Phillips and Magdalinos (2007b). Giraitis and Phillips (2012) provide a generalisation of some results under a weaker condition, similar to $(1-\rho_T)T \to \infty$. We strengthen Assumption A1 slightly, as

Assumption A2. *Assume that*

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

for $a > 2$.

Under (16), $\overline{d}$ 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 samples, 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 (10), suggesting a proximity of the two representations; the extension of (9) under (16) is explored more in Phillips and Magdalinos (2007b)).

Theorem 3. *For d_t generated as in (6)-(7) and (16), under Assumption A2, , for any $\alpha \in (0,1)$:*

Case A: For $1/M + M/T^{\alpha} \to 0$ *as* $T \to \infty$*:*

$$(-c)^{1/2}(M/T^{\alpha})^{1/2}\frac{\overline{d}-\mu}{\widehat{\sigma}_A} \to_d N(0,1) \tag{17}$$

Case P: For $1/m + m/T \to 0$ *as* $T \to \infty$*,*

$$\text{if } mT^{\alpha-1} \to 0: \ \sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}_P} \to_d N(0,1) \tag{18}$$

$$\text{if } mT^{\alpha-1} \to \infty: \ (mT^{\alpha-1})^{-1/2}1/2(-c)^{1/2}\left\{\sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}_P}\right\} \to_d N(0,1) \tag{19}$$

Remark 2. *a. Results in* (17) *and* (19) *are analogue; we conjecture that a result analogue to* (18) *also exists when* $\widehat{\sigma}_A$ *is used and* $M/T^{\alpha} \to \infty$*.*

b. Rewriting $DM_P = \sqrt{T}\frac{\overline{d}-\mu}{\widehat{\sigma}_P} + \sqrt{T}\frac{\mu}{\widehat{\sigma}_P}$ *for* m *as in* (18)*, the power depends on the drift* $\sqrt{T}\frac{\mu}{\widehat{\sigma}_P} = O(T^{1/2-\alpha})$*. 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 means that it is progressively more difficult to detect forecast inaccuracy as the dependence increases, even well within the weak dependence region.*

c. When $mT^{\alpha-1} \to \infty$ *(or* $M/T^{\alpha} \to 0$*), the DM statistics diverges even under* H_0*, thus resulting again in spurious significance.*

d. Condition $mT^{\alpha-1} \to 0$ *in* (18) *is not binding when* $\alpha = 0$ *but it is very strong as* $\alpha \to 1$*. Thus, results in* (18) *and* (19) *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 samples since* 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*.*

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. We consider the DGP

$$y_t = \alpha + \beta x_{t-1} + u_t \tag{20}$$

where

$$x_t = \phi x_{t-1} + \varepsilon_t,\ |\phi| < 1,\ \varepsilon_t \sim N.i.d.(0, \sigma_\varepsilon^2)$$
$$u_t \sim N.i.d.(0, \sigma_u^2)$$

and u_t independent from ε_s for all s, t.

Notice that we have assumed $E(x_t) = 0$, this is without loss of generality because in (20) we could use deviations from the mean, $y_t = (\alpha + \beta E(x_{t-1})) + \beta(x_{t-1} - E(x_{t-1})) + u_t$. Finally, we also set $\beta = 1$, again without loss of generality.

We consider two forecasting strategies:

$$y_{1,t} = \widehat{\beta} x_{t-1} \quad \text{where } \widehat{\beta} = Plim_{R\to\infty} \frac{\sum_{s=t-R}^{t-1} y_s x_{s-1}}{\sum_{s=t-R}^{t-1} x_{s-1}^2} \tag{21}$$

$$y_{2,t} = \widetilde{y} \quad \text{where } \widetilde{y} = Plim_{R\to\infty} \frac{1}{R} \sum_{s=t-R}^{t-1} y_s \tag{22}$$

so $\widehat{\beta} = 1$ (in general it would be β) and $\widetilde{y} = \alpha$, and

$$e_{1,t} = y_t - y_{1,t} = \alpha + u_t$$
$$e_{2,t} = y_t - y_{2,t} = x_{t-1} + u_t$$
$$d_t = (e_{1,t}^2 - e_{2,t}^2) = \alpha^2 + 2\alpha u_t - x_{t-1}^2 - 2x_{t-1}u_t$$

and

$$E(d_t) = (e_{1,t}^2 - e_{2,t}^2) = \alpha^2 - \frac{1}{1-\phi^2}\sigma_\varepsilon^2$$

so, for

$$\alpha = \sqrt{\frac{1}{1-\phi^2}\sigma_\varepsilon^2} + \delta \tag{23}$$

then $E(d_t) = 0$ when $\delta = 0$. Notice that as $|\phi| < 1$ then both x_t, y_t and d_t are mixing with sufficient rate, $E(d_s^2) < \infty$ (in view of the Gaussianity) and the long run variance exists.

Remark *From Lemma 1 of Dittmann and Granger (2002), x_{t-1}^2 is AR(1) with coefficient ϕ^2; αu_t is an independent process and $x_{t-1}u_t$ is Martingale difference. Thus d_t is like AR(1) plus noise.*

As in any realistic situation the sample size T is given, whether the standard normal limit or Theorem 1 is a better approximation depends on the relative interplay between ϕ and T (by the same token, the Lemma in Dittmann and Granger (2002) should also be seen as an approximation, when ϕ is close to 1, relative to

T). For values of ϕ close to 1 (relative to T) the limit (10) would be a better approximation for the sample average, and Theorem 1 is a more appropriate guideline; conversely, (2) and the standard normal limit for the DM statistic should be a more reliable guideline when ϕ is not close to 1, relative to T. Thus, the same value of ϕ could generate spurious rejections or not depending on the sample size. In this Monte Carlo study, we thus consider a range of values for ϕ and T to assess the interplay between these two key elements.

We consider two sample sizes, $T = 50$ and $T = 100$, and a range of bandwidths spanning $m = 1$ to $m = \lfloor T^{2/3} \rfloor$ when the average periodogram is used, and a range of bandwidths spanning from $M = \lfloor T^{1/4} \rfloor$ to $M = T$ when the weighted autocovariance estimator is used. 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} or the $\Phi_A(b)$ distribution, respectively, the latter as in Kiefer and Vogelsang (2005). We always use these critical values since they yield better size properties, as discussed, for example, in Lazarus et al. (2018) or in Coroneo and Iacone (2020).

We consider $\sigma^2_\varepsilon = 1$, $\sigma^2_u = 1$, and a range of values for ϕ; α is as in (23) for two values of δ. We set $\delta = 0$ to observe the effects on the empirical size; to observe the effects on power we set $\delta = -\sqrt{\sigma^2_\varepsilon/(1-\phi^2)} + \sqrt{\sigma^2_\varepsilon/(1-\phi^2) - 1)}$ as this yields $E(d_t) = -1$: with this choice we can observe how the power changes as the persistence increases, for the same deviation from the null hypothesis.

We report in Tables 1-2 the simulation results using the weighted autocovariance and the weighted periodogram estimates of the long-run variance, respectively. Results confirm the findings in Section 3. In particular:

a. The empirical size is correct for $\phi = 0$, but (for given T) it deteriorates as we move closer to $\phi = 1$ and as M is smaller or m is larger.

b. The empirical power drops as we move closer to $\phi = 1$ and as M is smaller or m is larger, in the sense that the presence of $E(d_t) \neq 0$ does not affect much the number of rejections of the null hypothesis in those cases.

These results 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 decreases with M). In the power exercise, the power drops as ϕ increases from 0 to 0.85, but notice that for large values of ϕ this power is in fact fictitious, in the sense that it rather reflects the spurious size distortion that we observed under the null. Automatic procedures to select the bandwidth, as in Delgado and Robinson (1996), Robinson (1983) or in Newey and West (1994) would not solve these problems, although the fact that smaller ms (larger Ms) are automatically selected as the autocorrelation in the loss function increases, would at least avoid the most adverse effects.

Table 1: Empirical null rejection frequencies - weighted autocovariances

$E(d_t) = 0$, $T = 50$

M \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
$\lfloor T^{1/4} \rfloor$	0.055	0.069	0.134	0.176	0.237	0.353	0.556	0.706	0.827
$\lfloor T^{2/9} \rfloor$	0.055	0.069	0.134	0.176	0.237	0.353	0.556	0.706	0.827
$\lfloor T^{1/3} \rfloor$	0.054	0.062	0.118	0.152	0.204	0.307	0.500	0.658	0.798
$\lfloor T^{1/2} \rfloor$	0.052	0.056	0.092	0.118	0.155	0.229	0.380	0.539	0.710
T	0.051	0.056	0.083	0.100	0.128	0.178	0.284	0.401	0.545

$E(d_t) = -1$, $T = 50$

M \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
$\lfloor T^{1/4} \rfloor$	0.885	0.583	0.244	0.208	0.217	0.309	0.530	0.696	0.826
$\lfloor T^{2/9} \rfloor$	0.885	0.583	0.244	0.208	0.217	0.309	0.530	0.696	0.826
$\lfloor T^{1/3} \rfloor$	0.873	0.554	0.197	0.157	0.169	0.255	0.471	0.645	0.794
$\lfloor T^{1/2} \rfloor$	0.817	0.482	0.120	0.085	0.093	0.157	0.343	0.525	0.706
T	0.654	0.379	0.099	0.070	0.072	0.117	0.247	0.386	0.541

$E(d_t) = 0$, $T = 100$

M \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
$\lfloor T^{1/4} \rfloor$	0.047	0.056	0.108	0.140	0.194	0.294	0.489	0.643	0.788
$\lfloor T^{2/9} \rfloor$	0.048	0.060	0.125	0.165	0.232	0.350	0.547	0.694	0.823
$\lfloor T^{1/3} \rfloor$	0.047	0.053	0.098	0.126	0.171	0.259	0.447	0.603	0.760
$\lfloor T^{1/2} \rfloor$	0.044	0.047	0.078	0.095	0.126	0.180	0.309	0.458	0.645
T	0.046	0.047	0.066	0.079	0.100	0.136	0.214	0.310	0.441

$E(d_t) = -1$, $T = 100$

M \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
$\lfloor T^{1/4} \rfloor$	0.997	0.880	0.376	0.275	0.224	0.263	0.460	0.634	0.787
$\lfloor T^{2/9} \rfloor$	0.997	0.892	0.430	0.328	0.280	0.322	0.523	0.680	0.820
$\lfloor T^{1/3} \rfloor$	0.995	0.871	0.336	0.233	0.183	0.221	0.412	0.591	0.755
$\lfloor T^{1/2} \rfloor$	0.990	0.826	0.237	0.135	0.087	0.117	0.266	0.441	0.641
T	0.884	0.629	0.180	0.104	0.067	0.084	0.176	0.293	0.439

Note: empirical null rejection frequencies for the DM test with the DM_A statistic and fixed-b critical values. The data generating process is in equations (20) and (23), with $E(d_t) = 0$ for the size exercise, and $E(d_t) = -1$ for the power study. The sample size is 50 and 100; 10,000 replications.

Table 2: Empirical null rejection frequencies - weighted periodogram

$E(d_t) = 0$, $T = 50$

m \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	0.050	0.048	0.058	0.068	0.082	0.112	0.184	0.279	0.420
$\lfloor T^{1/4} \rfloor$	0.052	0.054	0.075	0.088	0.114	0.160	0.270	0.407	0.594
$\lfloor T^{1/3} \rfloor$	0.054	0.055	0.077	0.098	0.130	0.191	0.324	0.484	0.669
$\lfloor T^{1/2} \rfloor$	0.053	0.060	0.103	0.133	0.181	0.274	0.463	0.628	0.779
$\lfloor T^{2/3} \rfloor$	0.056	0.067	0.133	0.173	0.234	0.351	0.558	0.709	0.830

$E(d_t) = -1$, $T = 50$

m \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	0.356	0.201	0.059	0.047	0.048	0.073	0.157	0.264	0.413
$\lfloor T^{1/4} \rfloor$	0.610	0.321	0.077	0.052	0.056	0.100	0.233	0.392	0.590
$\lfloor T^{1/3} \rfloor$	0.728	0.401	0.085	0.057	0.066	0.122	0.286	0.471	0.666
$\lfloor T^{1/2} \rfloor$	0.847	0.506	0.153	0.123	0.132	0.215	0.432	0.616	0.776
$\lfloor T^{2/3} \rfloor$	0.878	0.571	0.238	0.204	0.215	0.310	0.532	0.698	0.827

$E(d_t) = 0$, $T = 100$

m \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	0.049	0.048	0.053	0.059	0.071	0.092	0.135	0.203	0.320
$\lfloor T^{1/4} \rfloor$	0.045	0.045	0.064	0.078	0.099	0.140	0.226	0.350	0.544
$\lfloor T^{1/3} \rfloor$	0.046	0.049	0.067	0.086	0.108	0.151	0.255	0.399	0.597
$\lfloor T^{1/2} \rfloor$	0.048	0.051	0.084	0.105	0.146	0.215	0.397	0.560	0.730
$\lfloor T^{2/3} \rfloor$	0.048	0.056	0.111	0.146	0.209	0.320	0.519	0.671	0.810

$E(d_t) = -1$, $T = 100$

m \ ϕ	0	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	0.582	0.349	0.109	0.075	0.053	0.058	0.107	0.188	0.315
$\lfloor T^{1/4} \rfloor$	0.959	0.698	0.161	0.085	0.052	0.073	0.182	0.332	0.539
$\lfloor T^{1/3} \rfloor$	0.977	0.758	0.181	0.097	0.057	0.085	0.213	0.382	0.593
$\lfloor T^{1/2} \rfloor$	0.993	0.848	0.274	0.178	0.134	0.175	0.359	0.547	0.727
$\lfloor T^{2/3} \rfloor$	0.996	0.878	0.387	0.293	0.245	0.288	0.490	0.660	0.807

Note: empirical null rejection frequencies for the DM test with the DM_P statistic and fixed-m critical values. The data generating process is in equations(20) and (23), with $E(d_t) = 0$ for the size exercise, and $E(d_t) = -1$ for the power study. The sample size is 50 and 100; 10,000 replications.

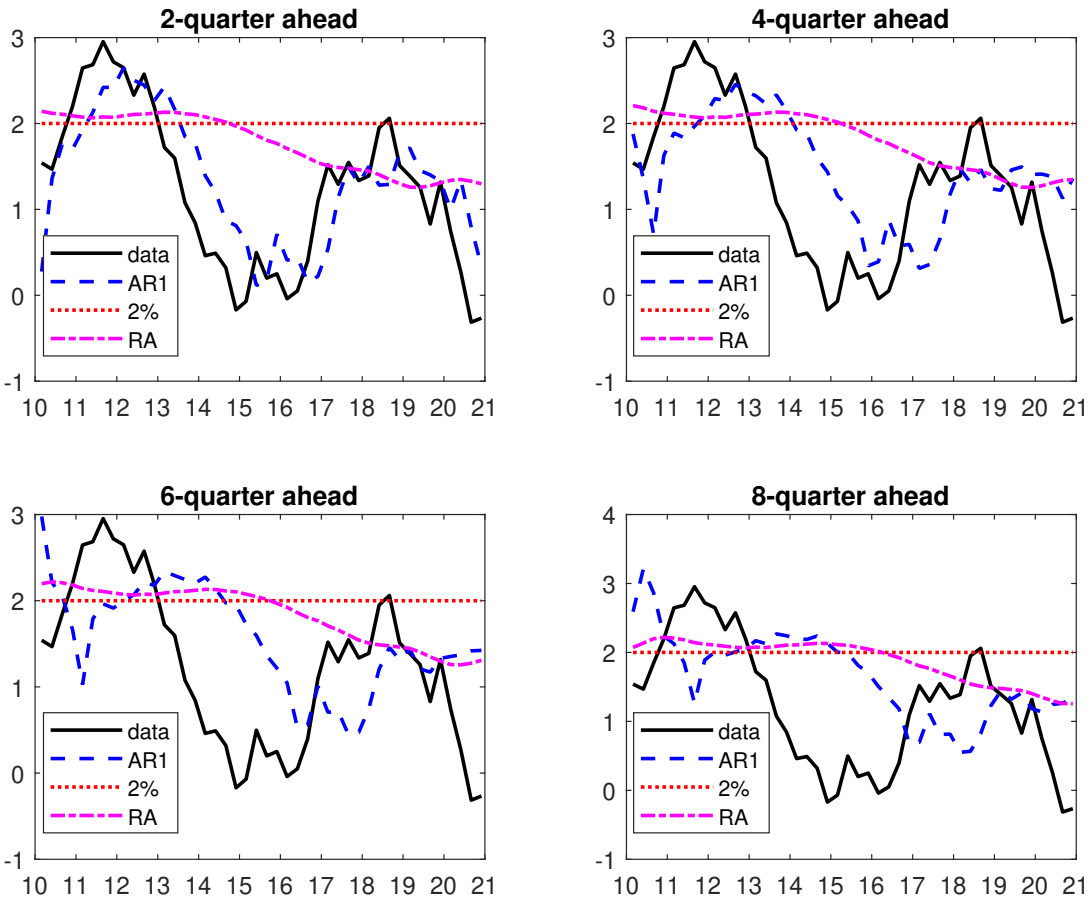
5 Empirical application

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 strongly dependent naive benchmarks. In particular, we consider quarterly predictions for the inflation rate in the Euro Area from a standard AR(1) model, as in 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 rolling average (RA) prediction.

We use data on the Harmonized Index of Consumer Prices from the FRED database, and we compute quarterly year-on-year inflation rates from 1996.Q1 to 2020.Q4. Given that our objective is to compare forecasting methods as opposed to forecasting models, we consider the case of non-vanishing estimation uncertainty, as in Giacomini and White (2006), and estimate all coefficients and rolling averages 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.

The series of inflation and the forecasts for selected forecast horizons for the AR(1) model are shown in Figure 1, along with the 2% and the rolling average benchmarks. A visual inspection of the plots immediately suggests that the forecast from the AR(1) model is clearly superior to the 2% benchmark for the 2-quarter horizon, but, not surprisingly, the superior performance of the forecast from the AR(1) model is eroded as longer forecasting horizons are considered. Additional plots for the realised forecast errors, the realised losses and the realised loss differentials are reported in Appendix B.

Figure 1: Realised inflation and forecasts



Note: realised inflation (data), AR(1) forecasts, along with the 2% (2%) and the rolling average (RA) benchmark forecasts for forecasting horizons 2, 4, 6 and 8 quarters.

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 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 all the median forecast errors are also negative.

The average and median forecast errors for the AR(1) increase (in absolute value) with the forecast horizon, 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. Finally, we also present the autocorrelation structure and the ADF tests for the forecast errors (we estimated the model with the intercept, with lags selected by means of the BIC). Especially at the lowest horizons, the autocorrelations

Table 3: Summary statistics of forecast errors

AR(1) forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1	-0.059	-0.077	0.367	0.284	0.264	0.182	0.179	-4.893**
2	-0.100	-0.092	0.590	0.690	0.384	0.350	0.313	-1.823
3	-0.148	-0.135	0.724	0.817	0.627	0.433	0.397	-1.892
4	-0.204	-0.134	0.827	0.860	0.697	0.569	0.424	-1.587
5	-0.252	-0.314	0.904	0.882	0.733	0.603	0.456	-2.222
6	-0.317	-0.182	0.983	0.898	0.760	0.577	0.390	-2.417
7	-0.378	-0.288	1.046	0.907	0.738	0.551	0.354	-2.009
8	-0.419	-0.420	1.078	0.881	0.717	0.535	0.366	-2.014

RA forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1	-0.505	-0.411	0.855	0.912	0.787	0.639	0.491	-1.116
2	-0.525	-0.428	0.871	0.914	0.789	0.641	0.493	-1.085
3	-0.546	-0.413	0.882	0.916	0.792	0.644	0.497	-1.074
4	-0.566	-0.409	0.890	0.917	0.794	0.648	0.503	-1.065
5	-0.585	-0.427	0.895	0.917	0.795	0.653	0.509	-1.083
6	-0.605	-0.438	0.896	0.919	0.799	0.658	0.517	-1.091
7	-0.624	-0.454	0.895	0.919	0.802	0.665	0.526	-1.108
8	-0.642	-0.475	0.893	0.920	0.806	0.673	0.536	-1.123

2% forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1-8	-0.765	-0.673	0.923	0.926	0.821	0.695	0.569	-0.732

Note: summary statistics for forecast errors from AR(1) predictions (top panel), rolling average (RA) predictions (middle panel) and constant 2% predictions. Forecast horizons are in quarters and forecast errors are defined as the realised value minus the prediction. ADF refers to the augmented Dickey–Fuller test (with intercept and lag order selected using the BIC criterion). * and ** denote significance at 10% and 5% level.

of the errors from the AR(1) forecasts declines fairly quickly, in comparison with the autocorrelations of the benchmarks. Despite this fact, the ADF test fails to reject the null hypothesis in all the cases, except for the one period horizon. We interpret this as a situation of low power of the ADF test, and therefore as evidence of persistence, but possibly not a unit root. We verify this interpretation by analysing the properties of the realised forecast losses reported in Table 4. The average realised losses associated to the AR(1) forecast are lower, at least for forecasts up to six quarters, and less dispersed than the ones of the two benchmarks, so they are, in this sense, more precise. Moreover, the losses from the AR(1) predictions are not very correlated for short forecasting horizons. As we increase the forecasting horizon the dependence increases, but the autocorrelations still decay reasonably quickly. On the other hand, the two benchmarks display large and persistent autocorrelations in their realised forecast losses at all forecasting horizons. We further investigate

Table 4: Summary statistics of realised losses

AR(1) forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1	0.135	0.092	0.128	-0.092	0.121	-0.030	-0.250	-6.938**
2	0.350	0.230	0.394	0.301	-0.145	-0.073	0.023	-4.735**
3	0.533	0.317	0.596	0.462	0.181	0.028	0.004	-3.820**
4	0.710	0.404	0.764	0.693	0.416	0.213	-0.016	-2.466
5	0.862	0.486	0.927	0.738	0.470	0.292	0.112	-2.429
6	1.045	0.600	1.133	0.706	0.459	0.303	0.182	-2.208
7	1.213	0.732	1.240	0.776	0.472	0.308	0.200	-2.094
8	1.312	0.700	1.323	0.728	0.486	0.329	0.247	-2.163

RA forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1	0.970	0.363	1.172	0.849	0.663	0.519	0.400	-1.688
2	1.017	0.357	1.241	0.858	0.672	0.527	0.410	-1.610
3	1.058	0.353	1.304	0.863	0.682	0.538	0.420	-1.579
4	1.094	0.341	1.357	0.869	0.693	0.553	0.433	-1.541
5	1.125	0.349	1.401	0.874	0.705	0.567	0.449	-1.526
6	1.150	0.336	1.443	0.877	0.713	0.579	0.459	-1.546
7	1.171	0.329	1.478	0.881	0.721	0.587	0.469	-1.555
8	1.191	0.332	1.510	0.883	0.727	0.592	0.475	-1.564

2% forecast

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1-8	1.418	0.509	1.593	0.862	0.654	0.461	0.338	-1.581

Note: summary statistics for realised losses from AR(1) predictions (top panel), rolling average (RA) predictions (middle panel) and constant 2% predictions. Forecast horizons are in quarters and forecast errors are defined as the realised value minus the prediction. ADF refers to the augmented Dickey–Fuller test (with intercept and lag order selected using the BIC criterion). * and ** denote significance at 10% and 5% level.

the dependence in the realised losses using the ADF test: the difference in the persistence that we observed in the sample autocorrelations of the realised losses is confirmed by the outcome of the ADF test, where the unit root hypothesis is rejected only for the forecasts from the AR(1) model (and only for short horizons).

Overall, these results suggest that the AR(1) model should be more precise for short-term forecasts, but this superiority could be masked empirically by the excessive dependence in the benchmarks. On the other hand, the AR(1) does not seem to produce more precise forecasts than the benchmarks at longer horizons, and the outcomes of the unit roots tests should be interpreted as a warning that any potential statistical significant difference may be spurious.

Summary statistics of the loss differentials, computed as the loss of the benchmark minus the loss of the AR(1), reported in Table 5, show that at short horizons loss differentials are positive, indicating that AR(1)

Table 5: Summary statistics loss differential

Benchmark: RA

Horizon	Mean	Median	Std	AC1	AC2	AC3	AC4	ADF
1	0.835	0.261	1.152	0.852	0.652	0.499	0.385	-1.624
2	0.667	0.067	1.181	0.826	0.647	0.504	0.382	-1.855
3	0.525	-0.001	1.127	0.848	0.675	0.501	0.348	-2.172
4	0.384	-0.003	1.093	0.837	0.675	0.479	0.283	-2.581
5	0.263	-0.020	1.046	0.843	0.644	0.440	0.233	-3.047**
6	0.105	-0.068	1.015	0.770	0.583	0.400	0.283	-2.128
7	-0.041	-0.074	0.954	0.797	0.511	0.370	0.299	-1.999
8	-0.121	-0.076	0.865	0.651	0.386	0.230	0.195	-3.678**

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	-1.664
2	1.068	0.433	1.530	0.834	0.626	0.443	0.305	-1.353
3	0.885	0.335	1.389	0.851	0.643	0.440	0.269	-1.775
4	0.708	0.305	1.306	0.834	0.640	0.410	0.225	-2.601
5	0.556	0.121	1.255	0.846	0.634	0.414	0.191	-2.556
6	0.373	0.039	1.208	0.794	0.602	0.416	0.265	-2.290
7	0.206	0.026	1.192	0.819	0.561	0.395	0.275	-2.517
8	0.106	0.009	1.161	0.736	0.474	0.285	0.185	-2.626*

Note: summary statistics for the AR(1) loss differential with respect to rolling average (RA) 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 (with intercept and lag order selected using the BIC). * and ** denote significance at 10% and 5% level.

Table 6: Forecast evaluation - weighted autocovariances

Benchmark: RA

M \ Horizon	1	2	3	4	5	6	7	8
$\lfloor T^{2/9} \rfloor$	3.594**	2.837**	2.304**	1.741	1.241	0.526	-0.222	-0.733
$\lfloor T^{1/3} \rfloor$	3.061**	2.420**	1.953*	1.473	1.055	0.451	-0.194	-0.650
$\lfloor T^{1/2} \rfloor$	2.430**	1.917	1.551	1.182	0.857	0.368	-0.159	-0.551
T	4.976**	3.772	3.098	2.378	1.712	0.665	-0.251	-0.870

Benchmark: 2%

M \ Horizon	1	2	3	4	5	6	7	8
$\lfloor T^{2/9} \rfloor$	4.087**	3.543**	3.190**	2.702**	2.215**	1.572	0.881	0.473
$\lfloor T^{1/3} \rfloor$	3.526**	3.059**	2.734**	2.312**	1.898*	1.355	0.770	0.418
$\lfloor T^{1/2} \rfloor$	2.885**	2.497**	2.231*	1.904	1.569	1.117	0.642	0.358
T	4.818**	3.964*	3.596	2.994	2.230	1.378	0.702	0.387

Note: DM test statistic for the null of equal predictive ability of AR(1) predictions with respect to a rolling average (RA) (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 the weighted autocovariances in (3). The sample size is 44 and bandwidth values M of $\lfloor T^{2/9} \rfloor$, $\lfloor T^{1/3} \rfloor$, $\lfloor T^{1/2} \rfloor$ and T are respectively 2, 3, 6 and 44. * and ** denote significance at 10% and 5% level using fixed-b critical values.

predictions may be more accurate than the benchmarks. As the forecasting horizon increases, the average loss differential decreases. In particular, for the RA benchmark, it becomes negative, so that at 7 and 8 quarters ahead RA predictions might be more accurate than AR(1) predictions.

However, the table also shows that the loss differentials are characterised by relevant autocorrelations, even at short forecasting horizons: the properties of the loss differentials are therefore heavily affected by the benchmark considered, and even with a forecast with weakly dependent loss 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 even 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.

We report the outcome of the DM tests for the null of equal predictive ability of AR(1) predictions with respect to a rolling average and a constant 2% benchmarks in Tables 6 and7.

We consider tests in which the DM statistics are computed estimating the long-run variances as in (3) or in (4). For $\widehat{\sigma}_A$ we used bandwidths M as $\lfloor T^{2/9} \rfloor$, $\lfloor T^{1/3} \rfloor$, $\lfloor T^{1/2} \rfloor$ and T, taking values 2, 3, 6 and 44 (the case $\lfloor T^{1/4} \rfloor$ is not present as this is again 2); for $\widehat{\sigma}_P$ 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 use critical

Table 7: Forecast evaluation - weighted periodogram

Benchmark: RA

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: DM test statistic for the null of equal predictive ability of AR(1) predictions with respect to a rolling average (RA) (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 the weighted autocovariance in (4). 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 (5). * and ** denote significance at 10% and 5% level using fixed-m critical values.

values from the corresponding fixed smoothing ($\Phi_A(b)$ or t_{2m}) distribution.

Results in Tables 6 and 7 are in line with our theory, as the outcome of the DM test reflects the autocorrelation documented in Table 5. In view of the high autocorrelation of d_t, the test may be affected by size distortion, especially with the larger bandwidths m, $m = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{2/3} \rfloor$, or for short M. We therefore discard results for these bandwidths. Even results with $m = \lfloor T^{1/3} \rfloor$ or $M = \lfloor T^{1/2} \rfloor$ should be considered with caution in this case, especially when the accuracy of the forecasts is compared against the 2% benchmark, as the autocorrelation seems to be particularly strong in that case. Remarkably, this is also the only case in which the equal predictive accuracy null between the AR(1) and the 2% forecast is significant at 5% level using $\widehat{\sigma}_P$ and the $m = \lfloor T^{1/3} \rfloor$ bandwidth; with shorter bandwidths, on the other hand, the null hypothesis of equal predictive accuracy is never rejected at 5% level. This seems to be a disappointing outcome, given the apparent superior performance of forecasts from the AR(1) model at short horizons (as documented in Figure 1 and in Table 5), and we suspect that it is due to the lack of power of the DM test in the presence of autocorrelation. The results when $\widehat{\sigma}_A$ and $M = T$ are used are perhaps slightly more convincing, at least when the one period ahead forecasts are evaluated. Overall, these results highlight

how applying the DM test when the loss differential is not weakly dependent may generate unrealiable results.

6 Conclusion

In this paper, we have verified that the DM test may be seriously misleading in presence of strong autocorrelation 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 (or with root close enough to 1, given the sample size) may not be impossible. 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, and that the DM test may perform poorly, generating size distortion or low power, also in the presence weakly dependent processes with autocorrelation close to unity.

For comparing forecasts, the DM test is "the only game in town", as noted in Diebold (2015). However, one should 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.

References

Chan, N. H., and C. Z. Wei (1987) 'Asymptotic Inference for Nearly Nonstationary AR(1) Processes.' *The Annals of Statistics* 15(3), 1050 – 1063

Choe, Yo Joong, and Aaditya Ramdas (2023) 'Comparing sequential forecasters.' *Operations Research*

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

Delgado, Miguel A, and Peter M Robinson (1996) 'Optimal spectral bandwidth for long memory.' *Statistica Sinica* 6, 97–112

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* 20(1), 253–263

Dittmann, Ingolf, and Clive WJ Granger (2002) 'Properties of nonlinear transformations of fractionally integrated processes.' *Journal of Econometrics* 110(2), 113–133

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 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

Gonçalves, Sílvia, and Timothy J Vogelsang (2011) 'Block bootstrap hac robust tests: The sophistication of the naive bootstrap.' *Econometric Theory* 27(4), 745–791

Henzi, Alexander, and Johanna F Ziegel (2021) 'Valid sequential inference on probability forecast performance.' *Biometrika* 109(3), 647–663

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

Khalaf, Lynda, and Charles J. Saunders (2017) 'Monte carlo forecast evaluation with persistent data.' *International Journal of Forecasting* 33(1), 1–10

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

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

McElroy, Tucker S., and Dimitris 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

Móricz, Ferenc (2006) 'Absolutely convergent fourier series and function classes.' *Journal of Mathematical Analysis and Applications* 324(2), 1168–1177

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

___ (2014) 'HAC corrections for strongly autocorrelated time series.' *Journal of Business & Economic Statistics* 32(3), 311–322

Newey, Whitney K, and Kenneth D West (1994) 'Automatic lag selection in covariance matrix estimation.' *The Review of Economic Studies* 61(4), 631–653

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

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

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 (2007a) 'Limit theory for moderate deviations from a unit root.' *Journal of Econometrics* 136(1), 115–130

Phillips, Peter C.B., and Tassos Magdalinos (2007b) '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*

Robinson, Peter 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

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

Robinson, PM (1983) 'Review of various approaches to power spectrum estimation.' *Handbook of Statistics* 3, 343–368

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

A Derivations

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. Recall that $I(\lambda_j)$ is the periodogram if d_t.

We discuss the results for DM_A and DM_P separately, starting from DM_P.

A.1 Results for Subsection 3.1

All the results in this subsection are for d_t generated as in (6)-(8), with $c \leq 0$, under Assumption A1.

Lemma 1.

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

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

Denoting $f(\lambda)$ as the spectral density of y_t, $g(\lambda)$ as the density of u_t, and

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

where $\upsilon = 1 - \rho$, then $f(\lambda) = f^\star(\lambda)g(\lambda)$.

We use for $f^\star(\lambda)$ the same bound as in Giraitis and Phillips (2012): for $|\rho| < 1$, $\lambda \leq \pi$

$$f^\star(\lambda) \leq \frac{1}{\upsilon^2 + \rho\lambda^2/3}. \tag{24}$$

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 (24) 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,

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

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,

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

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

Lemma 2.

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

Proof. We rewrite

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

and notice that

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

using (10) and (11), while

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

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 Robinson and Marinucci (2001): their moments condition is stronger and their proof is more complex than the argument given here because they established more results. □

Proof of Theorem 1 - *Case P*

Proof.

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

where in particular notice that $\frac{\sqrt{T}}{T}\mu \to 0$ so $\frac{\sqrt{T}}{T}(\overline{y}+\mu) \Rightarrow \omega \int_0^1 J_c(r)dr$ by a standard FCLT even when $\mu \neq 0$. The result thus follows from the convergence in (10), Lemma 2 and the continuous mapping theorem. □

Proof of Theorem 2 - *Case P*

Proof. The proof proceeds in the same way as in Theorem 1, but in this case we consider m as finite.

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

where the numerator is discussed as in as in Theorem 1, and $\frac{1}{T^2}\frac{1}{m}2\pi\sum_{j=1}^{m} I(\lambda_j) \Rightarrow \omega^2 \frac{1}{m}2\pi\sum_{j=1}^{m} Q_P(j)$ using the same argument as in Lemma 1 of Hualde and Iacone (2017). □

Lemma 3.

$$As\ M \to \infty,\ M/T \to 0, \quad \frac{1}{TM}\widehat{\sigma}_A^2 \to_d \omega^2 \int_0^1 \left(J_c(r)^2 - \overline{J}_c^2\right) dr$$

Proof. Notice that

$$\widehat{\gamma}_l = \frac{1}{T}\sum_{t=l+1}^{T}(d_t - \overline{d})(d_{t-l} - \overline{d}) = \frac{1}{T}\sum_{t=l+1}^{T}(y_t - \overline{y})(y_{t-l} - \overline{y})$$

Using recursive substitutions,

$$y_t = \rho^l y_{t-l} + \sum_{j=0}^{l-1}\rho^j u_{t-j} = y_{t-l} - (1-\rho^l)y_{t-l} + \sum_{j=0}^{l-1}\rho^j u_{t-j}$$

and

$$y_t - \overline{y} = (y_{t-l} - \overline{y}) - (1-\rho^l)y_{t-l} + \sum_{j=0}^{l-1} \rho^j u_{t-j}$$

so that

$$\widehat{\gamma}_l = \frac{1}{T}\sum_{t=l+1}^{T}(y_{t-l}-\overline{y})^2 - (1-\rho^l)\frac{1}{T}\sum_{t=l+1}^{T} y_{t-l}(y_{t-l}-\overline{y}) + \frac{1}{T}\sum_{t=l+1}^{T}\left(\sum_{j=0}^{l-1}\rho^j u_{t-j}\right)(y_{t-l}-\overline{y}). \tag{25}$$

We discuss the three terms in (25) separately. For the first one,

$$\frac{1}{T}\sum_{t=l+1}^{T}(y_{t-l}-\overline{y})^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t-\overline{y})^2 - \frac{1}{T}\sum_{t=1}^{l}(y_t-\overline{y})^2$$

and notice that

$$\frac{1}{T^2}\sum_{t=1}^{T}(y_t-\overline{y})^2 \to_d \omega^2 \int_0^1 \left(J_c(r)^2 - \overline{J}_c^2\right) dr$$

and, in view of Lemma 1 of Phillips (1987)

$$\frac{1}{T^2}\sum_{t=1}^{l}(y_t-\overline{y})^2 = O_p\left(\frac{1}{T^2} l (T^{1/2})^2\right) = O_p\left(\frac{M}{T}\right).$$

For the second term in (25), using the mean value theorem expansion

$$1-\rho^l = 1 - e^{lc/T} = lc_m/T \text{ for } c \leq \ c_m \leq 0$$

and bounds from Lemma 1 in Phillips (1987) then

$$(1-\rho^l)\frac{1}{T^2}\sum_{t=l+1}^{T} y_{t-l}(y_{t-l}-\overline{y}) = O_p\left(\frac{M}{T}\right)$$

Finally, as $E\left((\sum_{j=0}^{l-1}\rho^j u_{t-j})^2\right) = O(l)$, then with an application of the Cauchy-Schwarz inequality,

$$\frac{1}{T^2}\sum_{t=l+1}^{T}\left(\sum_{j=0}^{l-1}\rho^j u_{t-j}\right)(y_{t-l}-\overline{y}). = O_p\left(\frac{M^{1/2}}{T^{1/2}}\right)$$

so

$$\frac{1}{T}\widehat{\gamma}_l \to_d \omega^2 \int_0^1 \left(J_c(r)^2 - \overline{J}_c^2\right) dr$$

Thus, recalling $\sum_{l=1}^{M} \frac{M-l}{M} = 1/2\frac{M(M+1)}{M} = 1/2(M+1)$,

$$\frac{1}{TM}\widehat{\sigma}_A^2 \to_d \omega^2 \int_0^1 \left(J_c(r)^2 - \overline{J}_c^2\right) dr$$

□

Proof of Theorem 1 - *Case A*

Proof. The proof proceeds in the same way as for Case A, but using the limit in Lemma 3 instead. □

Proof of Theorem 2 - *Case A*

Proof. Proceeding as in Kiefer and Vogelsang (2005)

$$\frac{1}{T}\widehat{\sigma}_A^2 \to_d \omega^2 Q_A(b)^2$$

The proof then proceeds in the same way as for Case A. □

A.2 Results for Subsection 3.2

All the results in this subsection are for d_t generated as in (6), (7), (16) with $c < 0$, under Assumption A2.

We first state some properties that can be derived from Assumption A2

Lemma 4.

i. The condition in Assumption A2 implies the condition in A1.

ii. The condition in Assumption A2 is stronger than Assumption A1.

iii. $|g(x+h) - g(x)| < C|h|$ *for all* x, h.

Proof. 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

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

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, another application of summation by parts gives

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

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

so Assumption A2 also implies that

$$\sum_{k=0}^{\infty} k|\psi_k| < \infty$$

which is a sufficient condition for Theorem 2.1 of Móricz (2006), see page 1169. Denoting $g(\lambda)$ the spectral density of u_t, Theorem 2.1 of implies Móricz (2006) that $g(\lambda)$ belongs either to the Lipschitz class Lip(1), for which

$$|g(x+h) - g(x)| < C|h|$$

for all x, h. □

Lemma 5. *As $T \to \infty$*

$$\frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} (d_t - \overline{d})^2 = \frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} (d_t - \mu)^2 + o_p(1) \to_p \frac{\omega^2}{-2c} \tag{26}$$

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 (2007b); 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 6.

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

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

Lemma 7. *For frequencies λ_j such that $j < m$, $mT^{\alpha-1} \to 0$ as $T \to \infty$,*

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

Proof. We consider

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

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

$$f(\lambda_j)^{-1} \int_{-\pi}^{-\lambda_j/2} (f(\lambda) - f(\lambda_j)) K(\lambda - \lambda_j) d\lambda \leq C\ f^{\star}(\lambda_j)^{-1} \{sup_{\lambda \in [\lambda_j/2, \pi]} f^{\star}(\lambda)\} \int_{\lambda_j/2}^{\pi} K(\lambda + \lambda_j) d\lambda \quad (27)$$

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,

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

using (24). Thus, recalling $\int_{\lambda_j/2}^{\pi} K(\lambda + \lambda_j) d\lambda = O(j^{-1})$, then (27) $= O(j^{-1})$. The bound $f(\lambda_j)^{-1} \int_{2\lambda_j}^{\pi} = O(j^{-1})$ can be established in the same way.

Next,

$$f(\lambda_j)^{-1} \int_{-\lambda_j/2}^{\lambda_j/2} (f(\lambda) - f(\lambda_j)) K(\lambda - \lambda_j) d\lambda \leq C\ f^{\star}(\lambda_j)^{-1} \{sup_{\lambda \in [-\lambda_j/2, \lambda_j/2]} K(\lambda - \lambda_j)\} \int_{-\lambda_j/2}^{\lambda_j/2} (f(\lambda) - f(\lambda_j)) d\lambda \quad (28)$$

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 (28) is bounded by

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

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

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

then

$$\begin{aligned}
&f(\lambda_j)^{-1} \int_{\lambda_j/2}^{2\lambda_j} (f(\lambda) - f(\lambda_j))\, K(\lambda - \lambda_j) d\lambda \\
&\leq C f^\star(\lambda_j)^{-1} \left\{ \sup_{\lambda \in [\lambda_j/2, 2\lambda_j]} |f^\star(\lambda)'| \right\} \int_{\lambda_j/2}^{2\lambda_j} |\lambda - \lambda_j|\ K(\lambda - \lambda_j) d\lambda \qquad (29) \\
&+ C f^\star(\lambda_j)^{-1} f^\star(\lambda_j) \int_{\lambda_j/2}^{2\lambda_j} |\lambda - \lambda_j|\ K(\lambda - \lambda_j) d\lambda \qquad (30)
\end{aligned}$$

where notice that

$$\left\{ \sup_{\lambda \in [\lambda_j/2, 2\lambda_j]} |f^\star(\lambda)'| \right\} \leq C f^\star(\lambda_j)^2 \lambda_j.$$

The bound in (29) is

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

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 (30) is

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

□

Lemma 8. *For $m \to \infty$, $m\ T^{\alpha-1} \to 0$ 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 7. 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 7. □

Proof of Theorem 3 - *Case P*

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

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

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

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

The first term is bounded as

$$\left|f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m}(f(\lambda_j)^{-1}I(\lambda_j) - 1)f(\lambda_j)\right|$$
$$\leq f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1}|f(\lambda_j) - f(\lambda_{j+1})|\left|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k) - 1)\right| + f(0)^{-1}f(\lambda_m)\frac{1}{m}\left|\sum_{j=1}^{m}(f(\lambda_j)^{-1}I(\lambda_j) - 1)\right| \quad (31)$$

The first term of (31) has the same order as

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1}|f^{\star}(\lambda_j)'|\frac{1}{T}\left|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k) - 1)\right| + f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1} f^{\star}(\lambda_j)\frac{1}{T}\left|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k) - 1)\right| \quad (32)$$

the first element in (32) has order

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

where we used the bound $|f'^{\star}(\lambda_j)'|\leq Cf^{\star}(\lambda_j)^2\lambda_j\leq Cf^{\star}(0)f^{\star}(\lambda_j)\lambda_j\leq Cf^{\star}(0)\lambda_j^{-2}\lambda_j$ and Lemma 8; the second element in (32) has order

$$f(0)^{-1}\frac{1}{m}\sum_{j=1}^{m-1}f^{\star}(\lambda_j)\ j\ j^{-1}\frac{1}{T}\left|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k)-1)\right|$$
$$\leq C\frac{1}{m}\sum_{j=1}^{m-1}\frac{1}{T}j\ j^{-1}\left|\sum_{k=1}^{j}(f(\lambda_k)^{-1}I(\lambda_k)-1)\right|=o_p\left(\frac{1}{m}\sum_{j=1}^{m-1}\frac{1}{T}j\right)=o_p\left(\frac{m}{T}\right)=o_p(1).$$

so the bound in (32) is $o_p(1)$. The second term in (31) can be discussed in the same way, thus establishing that the bound in (31) is $o_p(1)$.

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 notice that

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

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

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

From Lemma 6,

$$\frac{2\pi}{T\,T^{\alpha}}\sum_{j=m+1}^{T/2} I(\lambda_j) = O_p\left(\frac{1}{T\,T^{\alpha}}\sum_{j=m+1}^{T/2}\lambda_j^{-2}\right) = O_p\left(T^{1-\alpha}m^{-1}\right) = o_p(1) \tag{33}$$

The application of Lemma 5 for $\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}(d_t-\overline{d})^2$ completes the proof. □

Proof of Theorem 3 - *Case A*

Proof. The proof uses the same decomposition (25) and the limit in Lemma 5; then, for all $l \leq M$

$$\frac{1}{T^{\alpha}}\widehat{\gamma}_l = \frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}(y_t-\overline{y}) + O_p(M/T^{\alpha}) + O_p(M^{1/2}/T^{\alpha/2})$$

and therefore

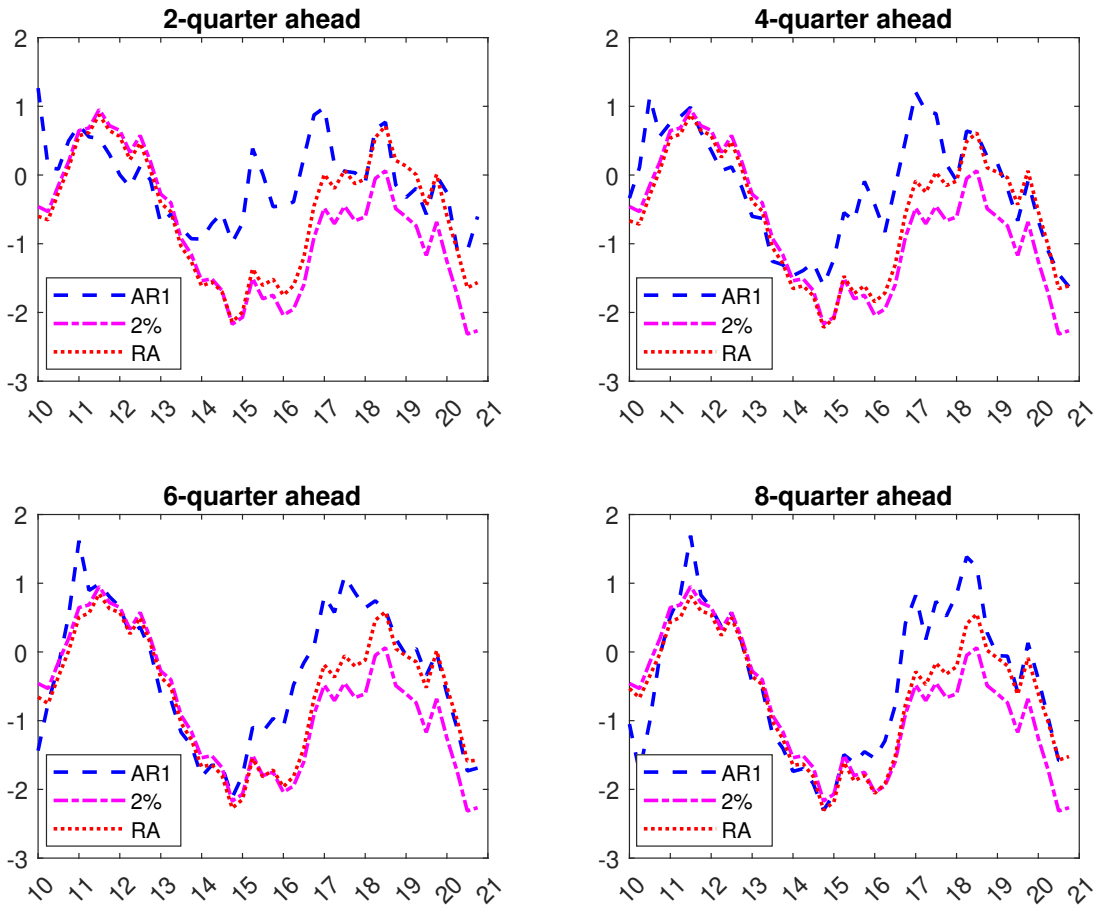
$$\frac{1}{MT^{\alpha}}\widehat{\sigma}_A^2 \to_p \frac{\omega^2}{-c}$$

The statement of the theorem follows from the CLT for $\overline{d}$ as in the proof for Case P and the continuous mapping theorem. □

B Additional plots

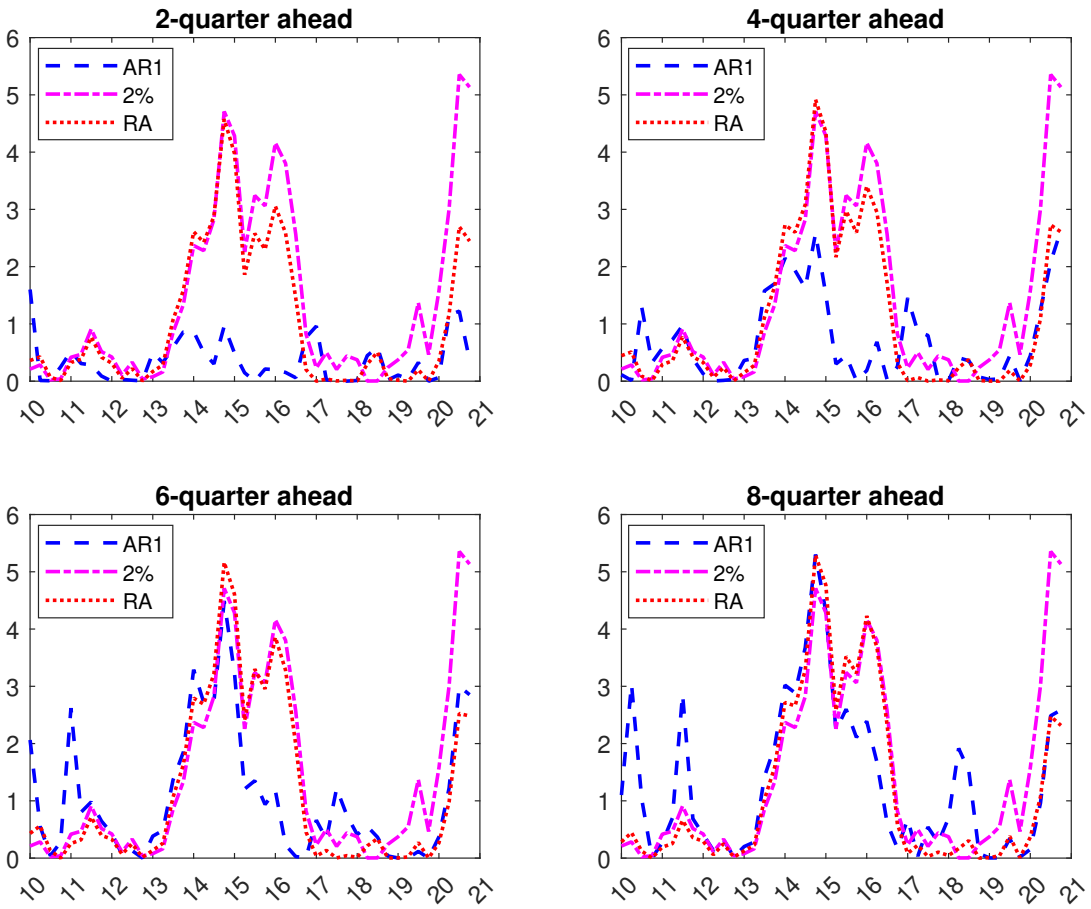
In this appendix, we report additional plots for the empirical application. In particular, Figure 2 plots the realised forecast errors from AR(1) forecasts along with the 2% and the rolling average benchmarks for forecasting horizons 2, 4, 6 and 8 quarters. As in the main part of the paper, forecast errors are defined as the realised value minus the prediction. The figure indicates that, for short forecasting horizons, AR(1) forecasts are more accurate and have less persistent forecast errors than the benchmarks. The realised losses of the three forecasts using a quadratic loss function reported in Figure 3 confirm that, for short forecasting horizons, AR(1) forecasts are more precise than the benchmarks and have losses that are not too correlated. However, when we look at the loss differentials reported in Figure 4, we see that they inherit the dependence properties of the benchmarks and display relevant autocorrelations, even at short forecasting horizons.

Figure 2: Realised forecast errors



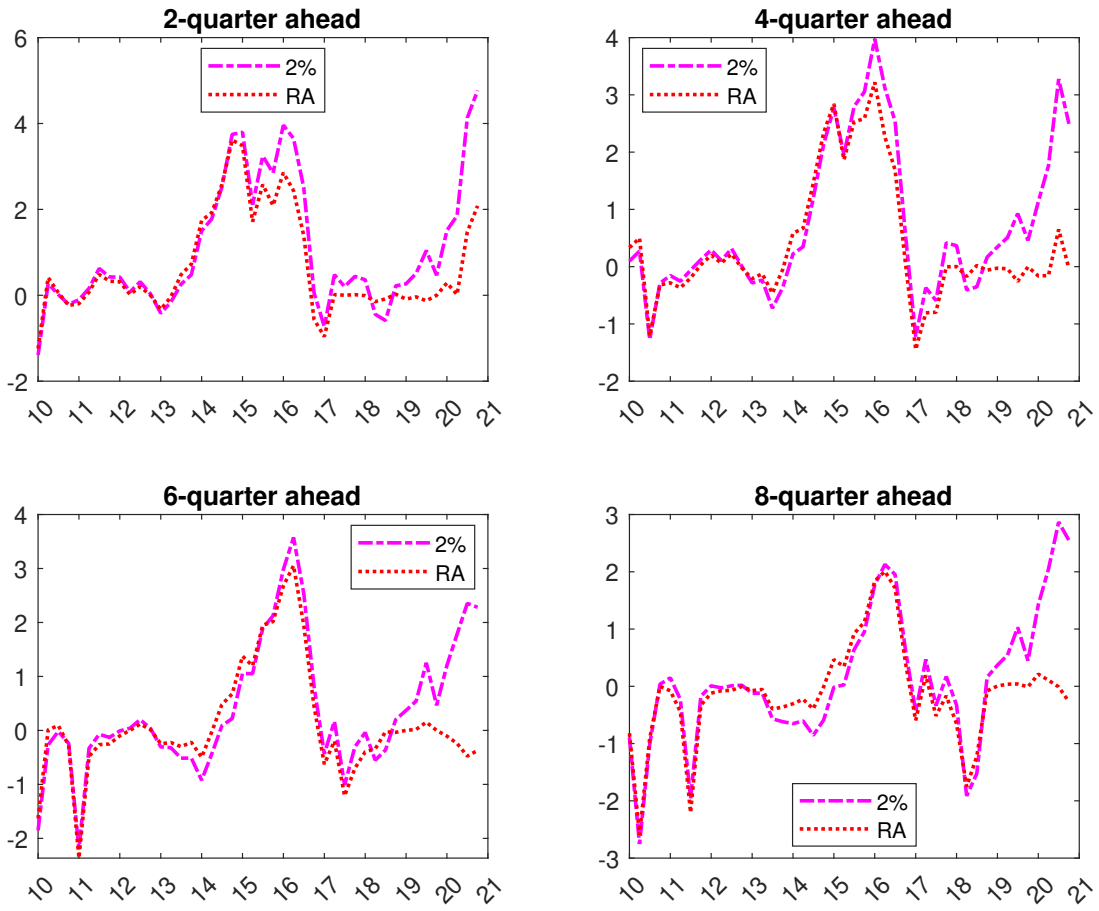
Note: realised forecast errors for AR(1) forecasts(AR1), along with the 2% (2%) and the rolling average (RA) benchmarks for forecasting horizons 2, 4, 6 and 8 quarters. Forecast errors are defined as the realised value minus the prediction.

Figure 3: Realised forecast losses



Note: realised forecast losses for AR(1) forecasts (AR1), along with the 2% (2%) and the rolling average (RA) benchmarks for forecasting horizons 2, 4, 6 and 8 quarters. The loss function is quadratic.

Figure 4: Realised loss differentials



Note: realised loss differentials for AR(1) forecasts (AR1) with respect to the 2% (2%) and the rolling average (RA) benchmarks for forecasting horizons 2, 4, 6 and 8 quarters. The loss function is quadratic, and the loss differential is computed as the loss of the benchmark minus the loss of the AR(1) forecast.