

Article

“A Bias Recognized Is a Bias Sterilized”: The Effects of a Bias in Forecast Evaluation

Nicolas Hardy ^{1,2}¹ Facultad de Economía y Empresa, Universidad Diego Portales, Santiago 8370179, Chile; nicolas.hardy@udp.cl² School of Business, Universidad Adolfo Ibáñez, Santiago 8380629, Chile

Abstract: Are traditional tests of forecast evaluation well behaved when the competing (nested) model is biased? No, they are not. In this paper, we show analytically and via simulations that, under the null hypothesis of no encompassing, a bias in the nested model may severely distort the size properties of traditional out-of-sample tests in economic forecasting. Not surprisingly, these size distortions depend on the magnitude of the bias and the persistency of the additional predictors. We consider two different cases: (i) There is both in-sample and out-of-sample bias in the nested model. (ii) The bias is present exclusively out-of-sample. To address the former case, we propose a modified encompassing test (MENC-NEW) robust to a bias in the null model. Akin to the ENC-NEW statistic, the asymptotic distribution of our test is a functional of stochastic integrals of quadratic Brownian motions. While this distribution is not pivotal, we can easily estimate the nuisance parameters. To address the second case, we derive the new asymptotic distribution of the ENC-NEW, showing that critical values may differ remarkably. Our Monte Carlo simulations reveal that the MENC-NEW (and the ENC-NEW with adjusted critical values) is reasonably well-sized even when the ENC-NEW (with standard critical values) exhibits rejections rates three times higher than the nominal size.

Keywords: forecasting; random walk; out-of-sample; bias; prediction; mean square prediction error



Citation: Hardy, N. “A Bias Recognized Is a Bias Sterilized”: The Effects of a Bias in Forecast Evaluation. *Mathematics* **2022**, *10*, 171. <https://doi.org/10.3390/math10020171>

Academic Editors: Javier Perote, Andrés Mora-Valencia and Trino-Manuel Níguez

Received: 16 November 2021

Accepted: 1 January 2022

Published: 6 January 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

“Fortunately for serious minds, a bias recognized is a bias sterilized” Benjamin Haydon. Diebold and Mariano (1995) [1] and West’s (1996) [2] seminal papers are typically pinpointed as the *Big Bang* of the forecast evaluation literature in economics and finance. Even though both papers propose asymptotically normal tests for forecast evaluation, the proper environment of each paper is different. The former considers the case of comparing forecasts (i.e., the forecasts are assumed to be given), while the latter focuses on the case of comparing models (i.e., the forecasts are constructed through estimated parametric models). Put simply, the key contribution of West’s asymptotic theory is that it accounts for parameter uncertainty.

While the asymptotic theory of [2] is quite general and allows a variety of estimation techniques and loss functions, it is not universal. One of the key assumptions in West’s theory is a full rank condition over the long-run variance of the loss function when parameters are set at their true values. One of the most iconic cases in which this condition is not fulfilled is the comparison of two competing nested models: Under the null hypothesis of no encompassing, both models are identical, and standard tests of forecast evaluation become degenerate. As pointed out by Clark and West (2006) [3] and West (2006) [4], for the case of MSPE, this degeneracy is not only important in a theoretical sense, but also in an empirical one: “[. . .] use of standard critical values usually results in very poorly sized tests, with far too few rejections. As well, the usual statistic has very poor power.” [4] p. 119 (A note of caution here. We are not arguing that this degeneracy necessarily imply that tests become undersize. Some of the simulations in [5] suggest both types of size distortions: sometimes tests are undersized, sometimes they are oversized.).

Not surprisingly, an important strand of this literature focuses on the case of nested models. Some of the most influential empirical papers in economic forecasting (such as Meese and Rogoff (1983, 1987) [6,7] and Goyal and Welsh (2003, 2008) [8,9], among many others) consider the random walk as one of the most relevant benchmarks. Anecdotally, and just to emphasize the relevance and popularity of nested models comparisons, to date, the number of citations of [2] is slightly lower than Clark and West (2007) [10] (1552 and 1615, respectively), even though the latter is exclusively useful for nested models (and published a decade later).

An interesting approach dealing with nested models is the one of Clark and McCracken (2001, 2005) [5,11] and McCracken (2007) [12] (henceforth, CM). Let P be the number of forecasts, R the initial number of observations used to estimate our parameters, $T = P + R$, and $f_{t+1}(\hat{\beta}_t)$ the loss function for one-step-ahead forecasts, constructed with a vector of estimated parameters $\hat{\beta}_t$. The intuition of CM asymptotics is as follows: when models are nested, traditional tests of encompassing/accuracy become degenerate under the null hypothesis in the sense that $P^{-0.5} \sum_{t=R}^T f_{t+1}(\hat{\beta}_t) \xrightarrow{plim} 0$. However, CM notice that $\sum_{t=R}^T f_{t+1}(\hat{\beta}_t)$ need not be degenerate. Based on this intuition, CM derive the asymptotic distribution of some traditional tests, such as the encompassing test of Harvey, Leybourne, and Newbold (1998) [13] (ENC-t), and a simple modification of the ENC-t (label as ENC-NEW). CM show that the asymptotic distribution of these tests is non-standard, but functions of integrals of quadratic Brownian motions. Under specific conditions, however, these statistics are free of nuisance parameters (See Clark and McCracken (2013) [14] for a detailed discussion on these conditions.). For those cases, CM tabulate their critical values, which depend on how we update our parameters (either rolling, recursive, or fixed), the number of excess parameters in the nesting model (k_2), and the asymptotic limit of the ratio P/R .

Even though the asymptotic theory of CM is useful in many contexts, it has some important caveats. In particular, Assumption 3 in [5,12] requires that the generalized forecast errors form a martingale difference sequence. This condition may be, of course, violated in the case of misspecification in the null model. As commented by [12]: *“The assumption has the side effect of imposing a type of correct model specification [. . .] and thus rules out applications where the predictive models are misspecified”* [12], p. 725. In our opinion, this is a somewhat overlooked aspect of CM results (Even though they take a completely different approach, the most related papers addressing misspecification in predictive models are those of Chao, Corradi and Swanson (2001) [15], Armah and Swanson (2008) [16], and Corradi and Swanson (2004, 2006, 2007) [17–19]. These authors propose bootstrap procedures that are robust to different types of misspecifications. In contrast, we derive the new asymptotics distributions in the presence of bias and propose a modification to the ENCNEW.). We show that the use of CM critical values ignoring the effects of these misspecifications may lead to severe size distortions; in specific, we address the effects of a bias in the predictive models over the asymptotic distributions (The implication of a bias in the nested model is that, under the null hypothesis, the forecasting errors of both models are biased. In essence, we show that a bias in the forecast errors distorts the size properties of traditional tests.).

Why not simply use a correctly specified model, then? First, our view about econometric forecasting models coincides with Box (1976) [20] in the sense that *“All models are wrong, but some of them are useful.”* In this context, as forecasting models are just parsimonious representations of complex phenomena, it is reasonable to expect some misspecification in the predictive models. Moreover, the econometric framework of CM considers exclusively linear models estimated by OLS; thus, any nonlinearity in the true model may introduce some type of misspecification. Second, if we knew in advance the “correct” model, then *“there may be no compelling reason for performing an out-of-sample version of the particular test (. . .) in the first place.”* [17], p. 188. As noticed by Inoue and Killian (2002) [21], in-sample analyses tend to outperform their out-of-sample counterparts in terms of power (when models are correctly specified); as a consequence, if we knew for sure that our models are “correct”, then the framework of CM may not be relevant to begin with: *“After all, under the assumption*

of correct specification under the null, why not simply carry out in-sample inference, for the sake of efficiency?" [16], p. 196.

If there is indeed a bias in the null model, then by construction, the critical values provided by CM are no longer valid as they rule out this possibility. In this context, we distinguish between two relevant cases: (i) there is a bias both in-sample and out-of-sample, and (ii) we find the bias exclusively out-of-sample. A leading case in which (i) is relevant is the following: If the expected value of the target variable is not zero, and the benchmark is the zero-forecast from a driftless random walk (DRW), then by construction, both models are biased under the null hypothesis, and inference may be severely incorrect when using CM critical values. As the DRW is one of the most important benchmarks in the economic/financial forecasting literature, we think this is an important extension for CM work, and we focus on this leading case (Some empirical papers using the DRW as a benchmark and the ENC-NEW with CM critical values are Pincheira and Hardy (2019, 2021) [22,23], Pincheira et al. (2021) [24,25], to name a few. We emphasize, however, that these papers consider multiple benchmarks as the nested model (not exclusively the DRW). In this sense, we are not questioning their main conclusions by any means.). Regarding (ii), an important case is a shift in the expected value of the target variable: because out-of-sample analyses require multiple in-sample estimations of the parameters, a structural break in the expected value of the target variable (outside the estimation window) may introduce an out-of-sample bias.

Based on CM, we relax Assumption 3 in [5] and derive the new distribution for the ENC-NEW imposing a bias in the null model (either out-of-sample or both in-sample and out-of-sample). While this is a subtle change in the set of assumptions, it has important implications over the asymptotic theory. In particular, the quadratic Brownian motions in CM arise because of the martingale difference terms: put simply, the orthogonality condition is assumed to hold both in-sample and out-of-sample. In contrast, in our case, the quadratic Brownian motions arise because of both the predictors and the martingale difference terms. Moreover, if the bias appears both in-sample and out-of-sample, the persistency of the predictors shift the expected value of the integrals of quadratic Brownian motions, thus a re-centering is required: our MENC-NEW is simply a re-centered version of the ENC-NEW. Of course, in the absence of a bias, our MENC-NEW reduces to the ENC-NEW. As expected, we show that this new asymptotic distribution depends on the magnitude of the bias and the persistency of the predictors. Even though the asymptotic distribution of the MENC-NEW (and the ENC-NEW with adjusted critical values) is not pivotal, the nuisance parameters can be easily estimated (one of these nuisance parameters being, of course, the magnitude of the bias).

To assess the adequacy of our approach, we provide five different sets of Monte Carlo simulations. The first two sets are designed to evaluate the size properties of our MENC-NEW with both an in-sample and out-of-sample bias. The third and fourth sets are designed to evaluate the size properties of the ENC-NEW with our adjusted critical values when there is only an out-of-sample bias (e.g., a structural break). Finally, the last set of simulations evaluates the power properties of our test. We compare the performance of our approach against the ENC-NEW (with CM critical values), the ENC-t ([5]), and the Wild Clark and West (WCW-t) proposed by Pincheira, Hardy, and Muñoz (2021) [26]. Our results are crystal clear in terms of size: (i) A bias in the null model may severely distort the size properties of the ENC-NEW and the ENC-t. For instance, under the null hypothesis, some of our simulations show that both tests may exhibit an empirical size of over 30% when the nominal size is just 10%. (ii) In sharp contrast, our approach is reasonably well-sized in all our simulations, and it seems to outperform the other tests in most of our simulations. As expected, our approach dominates the rest of the tests when the data-generating process considers a greater bias. (iii) Next to our approach, the WCW-t exhibits the best size properties. However, as Pincheira, Hardy, and Muñoz (2021) [26] noticed, this advantage comes with a cost: the WCW-t displays significantly less power

than the competitors. (iv) Even though the ENC-NEW is significantly more oversized than our test, there are no significant differences between both approaches in terms of power.

The rest of the paper is organized as follows. Section 2 provides a brief literature review on forecast evaluation in economics and finance. Section 3 considers the case of the DRW being the benchmark model (thus, a presence of a bias both in-sample and out-of-sample). In this case, we propose the MENC-NEW as the proper approach. Section 4 considers that the in-sample bias may be different than the out-of-sample bias. The interesting case here is an in-sample bias of zero but an out-of-sample bias different from zero (e.g., a shift in the drift of the target variable). In this case, there is no need for a re-centering term, although we derive the new asymptotic distribution of the ENC-NEW (thus, it is necessary to simulate new critical values). Section 5 considers four different Monte Carlo simulations to evaluate the size properties of our approach compared to traditional tests. At the end of Section 4 we evaluate the power of our test. Finally, Section 6 concludes.

2. Literature Review

Forecasting is a crucial area of study in financial econometrics: “It is obvious that forecasts are of great importance and widely used in economics and finance. Quite simply, good forecasts lead to good decisions.” Diebold and Lopez (1996) [27], page 241. Of course, a forecast has to be good in order to be useful, but how do we assess the quality of the forecast? This is the *raison d’être* of the literature on forecast evaluation. See West (2006) [4] and Clark and McCracken (2013) [28] for great reviews on forecast evaluation.

Forecast evaluation has a long tradition in economics and finance. For instance, West (2006) [4] recognizes Wilson (1934) [29] as one of the earlier examples. Other famous works are Fair and Shiller (1989, 1990) [30,31] and Meese and Rogoff (1983, 1988) [6,7]. The main intuition in these empirical papers is the following: if a model is a reasonably good representation of the target variable, then this model should accurately forecast this target variable. In this sense, forecast evaluation is at the core of empirical time series works.

Diebold and Mariano (1995) [1] (DM) and West (1996) [2] (DM) are two famous seminal papers addressing formal procedures to evaluate forecasts. In short, DM propose a direct application of the Central Limit Theorem for stationary series. In some sense, the DM approach is appealing since it relies on weak assumptions on the loss functions and the forecasting errors. Moreover, the Central Limit Theorem ensures asymptotic normality, and consequently, conducting inference in this framework is straightforward. Nevertheless, an important caveat of DM approach is that they consider forecasts as primitives: they do not consider the effects of parameter uncertainty. Intuitively, the DM approach is valid for evaluating forecasts but not for inference about models.

The main contribution of [2] is considering the case of forecasting models. The author provides a formal procedure for testing predictive ability at the population level. In essence, West (1996) [2] addresses how estimation error on the model’s parameters may affect the proper inference about predictive ability. West notices that, under specific conditions, estimation error vanishes asymptotically; if that is the case, then the approaches of DM and West are equivalent. Notably, the asymptotic theory of West rules out the comparison of nested models: West requires a full rank condition over the variance of the loss function. Intuitively, in nested models, the null hypothesis of equal predictive ability means that both models are identical; thus the asymptotic distribution of the test becomes degenerate.

The comparisons of nested models are extremely relevant in economics and finance. A common practice is to compare the predictive performance of a model with a simple benchmark. One of the most commonly considered benchmarks in the empirical literature is the random walk, which is typically nested on more complex predictive models. Even though the random walk is very simple, it is difficult to outperform empirically. As famously noticed by Meese and Rogoff (1983) [6]: “We find that a random walk model would have predicted major-country exchange rates during the recent floating-rate period as well as any of our candidate models” [6], page 3. See Goyal and Welch (2008) [9], Rossi

(2013) [32], and Meese and Rogoff (1988) [7] for more examples of how difficult it is to outperform a simple random walk with more complex nesting models.

Due to its relevance, an important strand of the literature focuses on formal procedures to compare the predictive performance of nested models. [3,10] (CW) propose an adjusted mean squared prediction error test and show via simulations that this test works reasonably well with standard normal critical values. Recently, Pincheira, Hardy, and Muñoz (2021) [26] (PHM) consider a simple modification of the CW test. This modification prevents the core statistic from becoming degenerate under the null hypothesis. Using the asymptotic theory of [2], PHM show their test is asymptotically standard normal. The main caveat of their approach is that their modification erodes some of the power properties of the CW test.

A different approach to deal with nested models is Clark and McCracken (2001, 2005) [5,11] and McCracken (2007) [12]. Using a completely different asymptotic theory, the authors consider some of the most popular tests on forecast evaluation and derive its asymptotic distributions for the special case of nested models. The authors show that the asymptotic distribution of these statistics are no longer standard, but they are functionals of quadratic Brownian motions. These distributions depend on the estimation scheme considered to update the parameters of the model (either rolling or recursive), on the number of excess parameters in the nesting model, and the asymptotic limit of the ratio P/R. Under some specific conditions, CM show that these distributions are pivotal, and they provide tables with the relevant critical values. Additionally, CM propose a different regularization of the encompassing test proposed by [13]: they label this new encompassing test as ENC-NEW. Notably, the simulations of CM suggest that the ENC-NEW has an edge in terms of power compared to other traditional tests.

One problem with the approach of CM is that it requires the strong assumption that the nested model is correctly specified. In essence, this assumption rules out the possibility of a bias in the benchmark model: CM critical values are no longer valid if that is the case. To the best of our knowledge, this is the first paper that addresses how this type of misspecification affects the asymptotic distribution of these tests. First, using the same asymptotic theory as CM, we analyze how a bias in the null model modifies the asymptotic distribution of the ENC-NEW. This asymptotic distribution is not pivotal and relies on the magnitude of the bias and the persistency of the additional predictor in the nesting model. Second, we show via simulations that ignoring this bias, and using CM critical values, may lead to severe size distortions. Not surprisingly, these size distortions depend on the magnitude of the bias and the persistency of the additional predictor.

3. The DRW as the Null Model and the MENC-NEW Test

The theory of CM imposes that the benchmark model is correctly specified. This section illustrates the effects of failing this condition because of a bias both in-sample and out-of-sample. The leading case is the DRW as the benchmark model: If the expected value of the target variable is not zero, then both models are biased under the null hypothesis of no encompassing. Considering the relevance of the DRW in the forecasting literature, we focus exclusively on this leading case. With some technical subtleties, the generalization to a linear parametric benchmark is straightforward and considered in Section 4. In fact, the results of Section 3 are special cases of our decompositions in Section 4.

Following CM notation, let k_2 be the number of parameters in the nesting model, and $W(s)$ be a $(k_2 \times 1)$ vector Brownian Motion with covariance kernel equal to the Identity matrix. Let Y_{t+1} be our scalar target variable, such that $Y_{t+1} = \Delta + u_{t+1}$, where u_{t+1} is simply white noise with variance σ_u^2 , and Δ captures the expected value of Y_{t+1} . Consider two competing models for one-step-ahead forecasts. The first one is simply the zero-forecast from a DRW, with forecast errors denoted by $u_{1,t+1}$

$$Y_{t+1} = u_{1,t+1} \text{ (model 1 : null model)}$$

The second one is a linear parametric model

$$Y_{t+1} = X_t' \beta^* + u_{2,t+1} \text{ (model 2 : alternative model)}$$

Notice that β^* denotes the $(k_2 \times 1)$ vector of true (population) parameters, and X_t is assumed to be a zero-mean covariance stationary vector with k_2 predictors. Let $\hat{\beta}_t$ be the OLS estimator of β^* constructed with either a rolling, recursive or fixed scheme.

As noticed by [2], under the null hypothesis of no encompassing, $\beta^* = 0$ and both models are identical $\forall t$. Therefore, traditional tests of encompassing/accuracy become degenerate under West's asymptotic theory. To address this issue, CM proposes a new statistic (ENC-NEW) and derives its distribution using a different set of asymptotics:

$$ENC - NEW = P \frac{P^{-1} \sum_t \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1})}{P^{-1} \sum_t \hat{u}_{2,t+1}^2}$$

The limit distribution of the ENC-NEW depends on how parameters are updated: either $\int_{R/T}^1 s^{-1} W'(s) dW(s)$ (recursive), or $T/R \int_{R/T}^1 (W(s) - W(s - \frac{R}{T})) dW(s)$ (rolling) (see [5], page 93). One of the key assumptions in CM is that model 1 is correctly specified under the null hypothesis. Notice, however, that the DRW is free of parameter uncertainty ($\hat{u}_{1,t+1} = u_{1,t+1}$), and

$$Y_{t+1} = \hat{u}_{1,t+1} = u_{1,t+1} = \Delta + u_{t+1}$$

In contrast, for the case of model 2

$$Y_{t+1} = X_t' \hat{\beta}_t + \hat{u}_{2,t+1}$$

where $\hat{u}_{2,t+1} = X_t' (\beta^* - \hat{\beta}_t) + u_{2,t+1}$.

Under the null of no encompassing $\beta^* = 0$, and $u_{1,t+1} = u_{2,t+1} = \Delta + u_{t+1}$. This is the key difference with CM asymptotic theory: Under the null hypothesis, both models are misspecified (biased); the reason is that none of them capture the mean of our target variable. *By construction, we will observe this type of misspecification whenever $\Delta = EY_{t+1} \neq 0$ and the nested model is the zero-forecast from a driftless random walk.* We focus on this leading case as the DRW is one of the most frequently used benchmarks in financial forecasting.

Let us now illustrate the effects of this bias over the asymptotic distribution of the ENC-NEW. For clarity of exposition, we focus our discussion and most of our proofs on the recursive scheme (since they are all similar). However, we do provide the main results for both the rolling and recursive schemes in Theorem 2.

From West (1996), under the null hypothesis, we have $P^{-1} \sum_t \hat{u}_{2,t+1}^2 \rightarrow \sigma_2^2 = Eu_{t+1}^2 + \Delta^2 = \sigma_u^2 + \Delta^2$ as $P, R \rightarrow \infty$. In what follows, we assume that $X_t u_{t+1}$ (rather than $X_t u_{2,t+1}$) is a martingale difference sequence. Now consider the following decomposition of the core statistic

$$\begin{aligned} \sum_{t=R}^{T-1} \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) &= \sum_{t=R}^{T-1} u_{1,t+1} (u_{1,t+1} - \hat{u}_{2,t+1}) = \sum_{t=R}^{T-1} u_{1,t+1} X_t' (\hat{\beta}_t - \beta^*) = \sum_{t=R}^{T-1} (\Delta + u_{t+1}) X_t' (\hat{\beta}_t - \beta^*) \\ &= \sum_{t=R}^{T-1} (\Delta + u_{t+1}) X_t' \left(\sum_{j=1}^{t-1} X_j X_j' \right)^{-1} \left(\sum_{j=1}^{t-1} X_j \hat{u}_{2,j+1} \right) = \sum_{t=R}^{T-1} (\Delta + u_{t+1}) X_t' \left(\sum_{j=1}^{t-1} X_j X_j' \right)^{-1} \left(\sum_{j=1}^{t-1} X_j (\Delta + u_{j+1}) \right) \\ &= \sum_{t=R}^{T-1} u_{t+1} X_t' \left(\sum_{j=1}^{t-1} X_j X_j' \right)^{-1} \left(\sum_{j=1}^{t-1} X_j u_{j+1} \right) \end{aligned} \tag{1}$$

$$+ \Delta \sum_{t=R}^{T-1} X_t' \left(\sum_{j=1}^{t-1} X_j X_j' \right)^{-1} \left(\sum_{j=1}^{t-1} X_j u_{j+1} \right) \tag{2}$$

$$+ \Delta \sum_{t=R}^{T-1} u_{t+1} X_t' \left(\sum_{j=1}^{t-1} X_j X_j' \right)^{-1} \left(\sum_{j=1}^{t-1} X_j \right) \tag{3}$$

$$+ \Delta^2 \sum_{t=R}^{T-1} X_t' \left(\sum_{j=1}^{t-1} X_j X_j' \right)^{-1} \left(\sum_{j=1}^{t-1} X_j \right) \tag{4}$$

This decomposition explicitly shows the effects of parameter uncertainty on the core statistic. This decomposition shows that the asymptotic distribution of the ENC-t and the ENC-NEW depend on how parameters are updated (either rolling or recursive). Notice that (1) is precisely the same term as Lemma A6 in [33] for the numerator of the ENC-NEW in a recursive scheme. We may interpret (1) as the core statistic when there is no bias; if that is the case, then $\Delta = 0$, and consequently (2) = 0, (3) = 0, (4) = 0, and the distribution of the ENC-NEW is given by Theorem 3.3 of [5]. Nevertheless, (2–4) in the ENC-NEW arise because Δ may not be zero. In some sense, we can view Clark and McCracken (2001) [5] results as the special case in which $\Delta = 0$. To establish the new asymptotic distribution of the ENC-NEW, we use the following results.

Lemma 1.

- (a) $\sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}) = \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}) + o_p(1)$
- (b) $\Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}) = \Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}) + o_p(1)$
- (c) $\Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j) = \Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j) + o_p(1)$
- (d) $\Delta^2 \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j) = \Delta^2 \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j) + o_p(1)$
- (e) $\sum_{t=R}^{T-1} \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) = \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}) + \Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}) + \Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j) + \Delta^2 \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j) + o_p(1)$

Proof of Lemma 1. See Appendix A. \square

Theorem 1 next establishes the asymptotic distribution of $\sum_t \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1})$

Theorem 1.

- (a) $\sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}) \rightarrow \int_{\lambda}^1 s^{-1} W_1'(s) \sigma_u^2 dW_1(s)$
- (b) $\Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}) \rightarrow \Delta \int_{\lambda}^1 s^{-1} W_1'(s) \sqrt{(E X_t X_t')^{-1} \sigma_u \sqrt{V_{LP}(X_t)}} dW_2(s)$
- (c) $\Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j) \rightarrow \Delta \int_{\lambda}^1 s^{-1} W_2'(s) \sqrt{V_{LP}(X_t)} \sigma_u \sqrt{(E X_t X_t')^{-1}} dW_1(s)$
- (d) $\Delta^2 \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' (E X_t X_t')^{-1} (T^{-0.5} \sum_{j=1}^{t-1} X_j) \rightarrow \Delta^2 \int_{\lambda}^1 s^{-1} W_2'(s) V_{LP}(X_t) (E X_t X_t')^{-1} dW_2(s) + \Lambda_{\lambda}$
- (e) $\sum_{t=R}^{T-1} \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) \rightarrow \int_{\lambda}^1 s^{-1} W_1'(s) \sigma_u^2 dW_1(s) + \Delta \int_{\lambda}^1 s^{-1} W_1'(s) \sqrt{(E X_t X_t')^{-1} \sigma_u \sqrt{V_{LP}(X_t)}} dW_2(s) +$

$$\Delta \int_{\lambda}^1 s^{-1} W_2'(s) \sqrt{V_{LP}(X_t)} \sigma_u \sqrt{(EX_t X_t')^{-1}} dW_1(s) + \Delta^2 \int_{\lambda}^1 s^{-1} W_2'(s) V_{LP}(X_t) (EX_t X_t')^{-1} dW_2(s) + \Lambda_{\lambda}$$

Proof of Theorem 1. See Appendix B. \square

Where $V_{LP}(X_t)$ is the long-run variance of X_t and $\lambda = R/T$. As expected, if $\Delta = 0$ (no bias), the asymptotic distribution of $\sum_t \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})$ is simply Theorem 1a $\int_{\lambda}^1 s^{-1} W_1'(s) \sigma_u^2 dW_1(s)$ (the same as CM). In contrast, if $\Delta \neq 0$, the limit distribution of $\sum_t \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})$ is given by Theorem 1e. Compared to CM, this distribution depends on three additional nuisance parameters: the magnitude of the bias Δ , the persistency of the additional predictors $V_{LP}(X_t)$, and Λ_{λ} (which is also related to the persistency of the predictors). Consequently, our Monte Carlo simulations in Section 5 reveal that the size distortions of the ENC-NEW with CM critical values are more severe with greater bias and persistency. Finally, Theorem 1e establishes that the asymptotic distribution of $\sum_t \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})$ depend not only on W_1 , but also on W_2 ; while the former appears because of $X_t u_{t+1}$ (as in CM), the latter arises because of X_t . Both standard Brownian motions are assumed to be independent.

Notice that the term Λ_{λ} in Theorem 1d,e arises because of the correlation between X_t and X_j (i.e., the correlation between the predictor in the estimation window and the evaluation window). Put simply, Λ_{λ} depends on the covariances $E(X_t X_j)$. As W_1 and W_2 are assumed to be independent, we must re-center the term $\Delta^2 \int_{\lambda}^1 s^{-1} W_2'(s) V_{LP}(X_t) (EX_t X_t')^{-1} dW_2(s)$ whenever $E(X_t X_j) \neq 0$. See White (2014) [34], p.198 for a similar discussion. Notably, our proposed MENC-NEW is simply a re-centered version of the ENC-NEW, corrected by the autocovariances that shift the distribution. The re-centering term is simply the expected value of $\Delta^2 \int_{\lambda}^1 s^{-1} W_2'(s) V_{LP}(X_t) (EX_t X_t')^{-1} dW_2(s)$. For clarity of exposition, consider the following example: suppose that $k_2 = 1$, then, the re-centering term is simply given by

$$\Lambda_{\lambda} = \Delta^2 (EX_t^2)^{-1} \sum_{t=R}^{T-1} \left(\frac{1}{t}\right) \sum_{j=1}^{t-1} E(X_t X_j)$$

While the sum of the covariances $E(X_t X_j)$ are unknown, they can be estimated. Our proposed MENC-NEW is simply

$$MENC - NEW = P \frac{P^{-1} \sum_t \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) - \hat{\Lambda}_{\lambda}}{P^{-1} \sum_t \hat{u}_{2,t+1}^2}$$

As $\hat{\Lambda}_{\lambda}$ depends on the sum of the covariances, its estimation depends on λ (i.e., how we split the data). Notice that this re-centering term will be particularly important for both the recursive and the rolling scheme. Theorem 2 next establishes the asymptotic distribution of our MENC-NEW for each updating scheme

Theorem 2. The asymptotic distribution of the MENC-NEW $= P \frac{P^{-1} \sum_t \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) - \hat{\Lambda}_{\lambda}}{P^{-1} \sum_t \hat{u}_{2,t+1}^2}$ is given by

$$\begin{aligned} MENC - NEW_{P,R \rightarrow \infty}(\text{recursive}) &\rightarrow \frac{1}{\sigma_u^2 + \Delta^2} \{ \sigma_u^2 \int_{\lambda}^1 s^{-1} W_1'(s) dW_1(s) \\ &+ \Delta \sigma_u \int_{\lambda}^1 s^{-1} W_1'(s) (EX_t X_t')^{-0.5} \sqrt{V_{LP}(X_t)} dW_2(s) \\ &+ \Delta \sigma_u \int_{\lambda}^1 s^{-1} W_1'(s) (EX_t X_t')^{-0.5} \sqrt{V_{LP}(X_t)} dW_2(s) \\ &+ \Delta^2 \int_{\lambda}^1 s^{-1} W_2'(s) V_{LP}(X_t) (EX_t X_t')^{-1} dW_2(s) \} \end{aligned}$$

$$\begin{aligned}
 &MENC - NEW_{P,R \rightarrow \infty}(\text{rolling}) \\
 &\rightarrow \frac{\lambda^{-1}}{\sigma_u^2 + \Delta^2} \{ \sigma_u^2 \int_{\lambda}^1 [W_1(s) - W_1(s - \lambda)]' dW_1(s) \\
 &+ \Delta \sigma_u \int_{\lambda}^1 [W_1(s) - W_1(s - \lambda)]' (EX_t X_t')^{-0.5} \sqrt{V_{LP}(X_t)} dW_2(s) \\
 &+ \Delta \sigma_u \int_{\lambda}^1 [W_2(s) - W_2(s - \lambda)]' \sqrt{V_{LP}(X_t)} (EX_t X_t')^{-0.5} dW_1(s) \\
 &+ \Delta^2 \int_{\lambda}^1 [W_2(s) - W_2(s - \lambda)]' V_{LP}(X_t) (EX_t X_t')^{-1} dW_2(s) \}
 \end{aligned}$$

where $R/T_{T,R \rightarrow \infty} \rightarrow \lambda$.

Proof of Theorem 2. This is a direct application of Theorem 1 and the definition of the MENC-NEW. □

Not surprisingly, if we set $\Delta = 0$ (i.e., no bias), the asymptotic distribution of the MENC-NEW is either $\int_{\lambda}^1 s^{-1} W_1'(s) dW_1(s)$ (recursive) or $\lambda^{-1} \int_{\lambda}^1 [W_1(s) - W_1(s - \lambda)]' dW_1(s)$, which is precisely Theorem 3c in [5]. In other words, Theorem 2 encompasses the case of unbiasedness.

The main caveat of our approach is that these distributions are not pivotal: to obtain our critical values, we must simulate the quadratic Brownian motions using estimates of these nuisance parameters. Notably, relaxing the assumption of $\Delta = 0$ has the implication of significantly complicating the asymptotic distribution of our test. Notice that the distribution of the MENC-NEW depends on (i) The number of excess parameters k_2 . (ii) How we update our parameters (either recursive or rolling). (iii) The persistency of the additional predictors (through $V_{LP}(X_t)$). (iv) The magnitude and sign of the bias (through Δ), (v) The variances $(EX_t X_t')^{-1}$ and σ_u^2 . (vi) How we split our dataset (through λ). In contrast, the distribution of the ENC-NEW simply depends on (i), (ii), and (vi); for this reason, CM are able to provide a set of critical values. We emphasize, however, that ignoring these additional parameters may distort the empirical size of these tests, as we show in Section 5.

Let us illustrate numerically how the critical values of the ENC-NEW might change with a bias. Suppose we are interested in a recursive scheme, and we use fifty percent of the sample to estimate our parameters and fifty percent to evaluate our forecasts ($\frac{R}{T} = 0.5$, or $\frac{P}{R} = 1$). Suppose the alternative model consider just one predictor X_t (i.e., $k_2 = 1$). Under the assumption of no bias, CM establishes the asymptotic distribution of the ENC-NEW in a recursive scheme as $\int_{R/T}^1 s^{-1} W(s) dW(s)$, and tabulate the corresponding critical values. In this case, the critical values for 10%, 5%, and 1% significance levels are 0.984, 1.584, and 3.209, respectively.

Nevertheless, Theorem 1 establishes that the proper asymptotic distribution of the ENC-NEW with a bias under the null model should be

$$\begin{aligned}
 &\frac{1}{\Delta + \sigma_u^2} (\int_{\lambda}^1 s^{-1} W_1'(s) \sigma_u^2 dW_1(s) \\
 &\quad + \Delta \int_{\lambda}^1 s^{-1} W_1'(s) \sqrt{(EX_t X_t')^{-1}} \sigma_u \sqrt{V_{LP}(X_t)} dW_2(s) \\
 &+ \Delta \int_{\lambda}^1 s^{-1} W_2'(s) \sqrt{V_{LP}(X_t)} \sigma_u \sqrt{(EX_t X_t')^{-1}} dW_1(s) \\
 &\quad + \Delta^2 \int_{\lambda}^1 s^{-1} W_2'(s) V_{LP}(X_t) (EX_t X_t')^{-1} dW_2(s) + \Lambda_{\lambda})
 \end{aligned}$$

Suppose the predictor in the alternative model follows an AR(1) process: $X_t = 0.5X_{t-1} + u_{X,t}$, with $V(u_{X,t}) = 0.1$. If that is the case, $V_{LP}(X_t) = \frac{0.1}{(1-0.5)^2} = 0.4$ and $EX_t^2 = 0.1333$. Set $\Delta = 0.5$ and $\sigma_u^2 = 0.1$. The critical values for 10%, 5%, and 1% significance levels in this new scenario are 2.380, 3.760, and 7.676, respectively. In this case, the bias in the null model shifts the critical values significantly to the right: in this illustration, the critical values are twice those tabulated by CM. Consequently, using CM critical values may lead to severely oversized tests. This is supported by our Monte Carlo simulations.

4. The General Case of a Parametric Nested Benchmark and the Asymptotic Distribution of the ENC-NEW

This section generalizes our previous result considering a parametric nested benchmark (possibly different from the DRW). Here, we allow the possibility of having different in-sample and out-of-sample biases. As we estimate our parameters by OLS, it is reasonable to think that the bias is present exclusively out-of-sample. For this reason, we derive the asymptotic distribution of the ENC-NEW considering an in-sample and an out-of-sample bias (Δ_{IS} and Δ_{OOS} respectively and possibly $\Delta_{IS} \neq \Delta_{OOS}$), and then we focus on the particular (but more relevant) case of $\Delta_{IS} = 0$.

Let X_1 and X_2 be the $(k_1 \times 1)$ and $(k \times 1)$ vectors of predictors of the nested and nesting models, respectively. In contrast to the previous section, $k = k_1 + k_2$; in other words, k_2 indicates the additional predictors considered in the nesting model. Let J be a $k_1 \times k_2$ indication matrix, constructed as a $(k_1 \times k_1)$ identity matrix, and zeroes elsewhere. In contrast to the previous section, additional terms arise in our decomposition because model 1 is also affected by parameter uncertainty. We focus our discussion and most of our proofs on the recursive scheme. We do provide, however, the main results for both the rolling and recursive methods. First, notice that

$$\begin{aligned} \hat{u}_{1,t+1} &= u_{1,t+1} - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*) = (u_{t+1} + \Delta_{OOS}) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*) \\ \hat{u}_{2,t+1} &= u_{2,t+1} - X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) = (u_{t+1} + \Delta_{OOS}) - X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) \end{aligned}$$

Then, the core statistic from the ENC-NEW is simply given by

$$\begin{aligned} &\sum_{t=R}^{T-1} \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) \\ &= \sum_{t=R}^{T-1} [(u_{t+1} + \Delta_{OOS}) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)][X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)] \\ &= \sum_{t=R}^{T-1} (u_{t+1} + \Delta_{OOS})[X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)] \end{aligned} \tag{5}$$

$$- \sum_{t=R}^{T-1} X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)[X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)] \tag{6}$$

To establish the distribution of (5) and (6), we need additional results, provided by Lemmas 2 through 5. While these Lemmas are similar to those of CM, they are not immediate from their results.

Lemma 2. $\sup_t T^{0.5} |\frac{1}{t} \sum_{j=1}^{t-1} X_{lj}| = O_p(1)$, with $l = 1, 2$.

Proof Lemma 2. See Appendix C. \square

Lemma 3.

$$\begin{aligned} &\sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X'_{1j} \right\}^{-1} X_{1t} X'_{2t} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X'_{2j} \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} \\ (a) \quad &= \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ E X_1 X'_1 \right\}^{-1} E X_{1t} X'_{2t} \left\{ E X_2 X'_2 \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} + o_p(1) \\ &\sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X'_{1j} \right\}^{-1} X_{1t} X'_{1t} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X'_{1j} \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \\ (b) \quad &= \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ E X_1 X'_1 \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1) \end{aligned}$$

$$\begin{aligned}
 & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} X_{1t} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\
 (c) \quad &= \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ EX_1 X_1' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \\
 & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} X_{1t} X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} \\
 (d) \quad &= \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ EX_1 X_1' \right\}^{-1} EX_{1t} X_{2t}' \left\{ EX_2 X_2' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} + o_p(1) \\
 & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} X_{1t} X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} \\
 (e) \quad &= \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \left\{ EX_1 X_1' \right\}^{-1} EX_{1t} X_{2t}' \left\{ EX_2 X_2' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} + o_p(1) \\
 & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} X_{1t} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\
 (f) \quad &= \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \left\{ EX_1 X_1' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \\
 & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} X_{1t} X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} \\
 (g) \quad &= \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \left\{ EX_1 X_1' \right\}^{-1} EX_{1t} X_{2t}' \left\{ EX_2 X_2' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} + o_p(1)
 \end{aligned}$$

Proof of Lemma 3. The proofs for Lemma 3a,b follow from Lemma A2 and A3b in CM. While the proofs for Lemmas 3c through 3g are not immediate from CM, they follow the same arguments and strategies. Since the proofs are very similar, Appendix D shows our proofs exclusively for Lemma 3c. □

Lemma 4.

$$\sum_{t=R}^{T-1} X_{1t}' (\hat{\beta}_{1t} - \beta_1^*) [X_{2t}' (\hat{\beta}_{2t} - \beta_2^*) - X_{1t}' (\hat{\beta}_{1t} - \beta_1^*)] \text{ is } o_p(1)$$

Proof Lemma 4. See Appendices E and F. □

Lemma 5.

$$\begin{aligned}
 & \sum_{t=R}^{T-1} \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) \\
 &= \sum_{t=R}^{T-1} u_{t+1} X_{2t}' [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sum_{j=1}^{t-1} X_{2j} u_{t+1} \\
 &+ \Delta_{IS} \sum_{t=R}^{T-1} u_{t+1} X_{2t}' [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sum_{j=1}^{t-1} X_{2j} \\
 &+ \Delta_{OOS} \sum_{t=R}^{T-1} X_{2t}' [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sum_{j=1}^{t-1} X_{2j} u_{t+1} \\
 &+ \Delta_{IS} \Delta_{OOS} \sum_{t=R}^{T-1} X_{2t}' [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sum_{j=1}^{t-1} X_{2j} + o_p(1)
 \end{aligned}$$

Proof of Lemma 5. See Appendix G. □

Theorem 3. Let $S_{uX} = \sigma_u \sqrt{(EX_{2t} X_{2t}')^{-1}}$,

$$(a) \quad \sum_{t=R}^{T-1} u_{t+1} X_{2t}' [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sum_{j=1}^{t-1} X_{2j} u_{t+1} \rightarrow \int_{\lambda}^1 s^{-1} W_1'(s) S_{uX} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] S_{uX} dW_1(s)$$

- (b) $\Delta_{IS} \sum_{t=R}^{T-1} u_{t+1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} X_{2j} \rightarrow$
 $\Delta_{IS} \int_{\lambda}^1 s^{-1} W'_2(s) \sqrt{V_{LP}(X_{2t})} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] S_{uX} dW_1(s)$
- (c) $\Delta_{OOS} \sum_{t=R}^{T-1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} X_{2j} u_{t+1} \rightarrow$
 $\Delta_{OOS} \int_{\lambda}^1 s^{-1} W'_1(s) S_{uX} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s)$
- (d) $\Delta_{OOS} \Delta_{IS} \sum_{t=R}^{T-1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} X_{2j} \rightarrow$
 $\Delta_{OOS} \Delta_{IS} \int_{\lambda}^1 s^{-1} W'_2(s) \sqrt{V_{LP}(X_{2t})} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) + \Lambda_{\lambda}$
- (e) $\sum_{t=R}^{T-1} \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) \rightarrow \int_{\lambda}^1 s^{-1} W'_1(s) S_{uX} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J']$
 $S_{uX} dW_1(s) + \Delta_{IS} \int_{\lambda}^1 s^{-1} W'_2(s) \sqrt{V_{LP}(X_{2t})} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] S_{uX} dW_1(s) +$
 $\Delta_{OOS} \int_{\lambda}^1 s^{-1} W'_1(s) S_{uX} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) +$
 $\Delta_{OOS} \Delta_{IS} \int_{\lambda}^1 s^{-1} W'_2(s) \sqrt{V_{LP}(X_{2t})} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) + \Lambda_{\lambda} + o_p(1)$
- (f) *If there is no in-sample bias (i.e., $\Delta_{IS} = 0$), then*

$$\sum_{t=R}^{T-1} \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) \int_{\lambda}^1 s^{-1} W'_1(s) S'_{uX} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] S_{uX} dW_1(s) + \Delta_{OOS} \int_{\lambda}^1 s^{-1} W'_1(s) S'_{uX} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s)$$

Proof of Theorem 3. Theorem 3e is simply the combination of Theorems 3a through 3d. Theorem 3f is equivalent to Theorem 3e imposing $\Delta_{IS} = 0$. The proof for Theorems 3a–3d is direct from Lemmas 3 through 5. □

Three things from Theorem 3 are worth mentioning. First, the re-centering term of the MENC-NEW is relevant in the distribution because of the term $\Delta_{OOS} \Delta_{IS} \sum_{t=R}^{T-1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} X_{2j}$. The reason is that the correlation between X_{2j} and X_{2t} may not be zero; in other words, we may find a correlation between the predictor in the estimation and evaluation window. Second, if there is no in-sample bias (i.e., $\Delta_{IS} = 0$) then the re-centering term is not required (thus, we could simply use the ENC-NEW with different critical values). In essence, the term $\Delta_{OOS} \sum_{t=R}^{T-1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} u_{t+1} X'_{2j}$ does not require a re-centering term since $X_{2j} u_{t+1}$ (the in-sample term) and X_{2t} (out-of-sample predictor) are assumed to be uncorrelated. Finally, the first term in the distribution of Theorem 3f is akin to CM if we are willing to assume that errors are not autocorrelated and homoscedastic. Notice, however, that even if that is the case, our distribution is different because of the out-of-sample bias. In contrast to CM, these additional terms depend on the long-run variance of X_{2t} ($V_{LP}(X_{2t})$), on an additional Brownian motion W_2 and, of course, the out-of-sample bias Δ_{OOS} . Theorem 4a next establishes the asymptotic distribution of the MENC-NEW in the general case in which $\Delta_{IS}, \Delta_{OOS} \neq 0$. In Theorem 4b, we consider the special (although more interesting) case in which $\Delta_{IS} = 0, \Delta_{OOS} \neq 0$; in this case, we establish the new asymptotic distribution of the ENC-NEW, as the re-centering term is not required.

Theorem 4.

- (a) *The asymptotic distribution of our MENC-NEW = $P \frac{P^{-1} \sum_t \hat{u}_{1,t+1} (\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) - \hat{\Lambda}_{\lambda}}{P^{-1} \sum_t \hat{u}_{2,t+1}^2}$ is simply given by*

$$MENC - NEW_{P,T \rightarrow \infty} (Recursive)$$

$$\frac{1}{\sigma_{\hat{u}}^2 + \Delta_{OOS}^2} \left[\int_{\lambda}^1 s^{-1} W'_1(s) S_{uX} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] S_{uX} dW_1(s) + \Delta_{IS} \int_{\lambda}^1 s^{-1} W'_2(s) \sqrt{V_{LP}(X_{2t})} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] S_{uX} dW_1(s) + \Delta_{OOS} \int_{\lambda}^1 s^{-1} W'_1(s) S_{uX} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) + \Delta_{OOS} \Delta_{IS} \int_{\lambda}^1 s^{-1} W'_2(s) \sqrt{V_{LP}(X_{2t})} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) \right]$$

$$MENC - NEW_{P,T \rightarrow \infty} (Rolling) \rightarrow$$

$$\begin{aligned} & \frac{\lambda^{-1}}{\sigma_u^2 + \Delta_{\text{oos}}^2} \left[\int_{\lambda}^1 (W_1(s) - W_1(s - \lambda))' S_{uX} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] S_{uX} dW_1(s) \right. \\ & + \Delta_{IS} \int_{\lambda}^1 (W_2(s) - W_2(s - \lambda))' \sqrt{V_{LP}(X_{2t})} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] S_{uX} dW_1(s) \\ & + \Delta_{\text{oos}} \int_{\lambda}^1 (W_1(s) - W_1(s - \lambda))' S_{uX} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) \\ & \left. + \Delta_{\text{oos}} \Delta_{IS} \int_{\lambda}^1 (W_2(s) - W_2(s - \lambda))' \sqrt{V_{LP}(X_{2t})} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) \right] \end{aligned}$$

(b) If there is no in-sample bias (i.e., $\Delta_{IS} = 0$), then the asymptotic distribution of the ENC-NEW is simply

$$ENC - NEW_{P,T \rightarrow \infty}(\text{recursive})$$

$$\begin{aligned} & \frac{1}{\sigma_u^2 + \Delta_{\text{oos}}^2} \left[\int_{\lambda}^1 s^{-1} W_1'(s) S_{uX} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] S_{uX} dW_1(s) \right. \\ & \left. + \Delta_{\text{oos}} \int_{\lambda}^1 s^{-1} W_1'(s) S_{uX} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) \right] \end{aligned}$$

$$ENC - NEW_{P,T \rightarrow \infty}(\text{rolling}) \rightarrow$$

$$\begin{aligned} & \frac{\lambda^{-1}}{\sigma_u^2 + \Delta_{\text{oos}}^2} \left[\int_{\lambda}^1 (W_1(s) - W_1(s - \lambda))' S_{uX} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] S_{uX} dW_1(s) \right] \\ & + \frac{\lambda^{-1} \Delta_{\text{oos}}}{\sigma_u^2 + \Delta_{\text{oos}}^2} \left[\int_{\lambda}^1 (W_1(s) - W_1(s - \lambda))' S_{uX} [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sqrt{V_{LP}(X_{2t})} dW_2(s) \right] \end{aligned}$$

Proof of Theorem 4. The proofs of Theorem 4a,b are immediate from Theorem 3e,f and the definition of the MENC-NEW and the ENC-NEW. \square

As expected from Theorem 3, even in the case of $\Delta_{IS} = 0$, we will see that the out of sample bias shift the asymptotic distribution of the ENC-NEW because of the term $\Delta_{\text{oos}} \sum_{t=R}^{T-1} X_{2t}' [(EX_2 X_2')^{-1} - J(EX_1 X_1')^{-1} J'] \sum_{j=1}^{t-1} X_{2j} u_{t+1}$. Akin to CM, the distribution of the ENC-NEW depends on how parameters are estimated (either rolling or recursive), on the number of excess parameters k_2 , and how we split our database (through $\lambda = \lim_{P,T \rightarrow \infty} R/T$). Nevertheless, the new asymptotic distribution of the ENC-NEW depend on more nuisance parameters than CM: it depends on the magnitude of the out-of-sample bias Δ_{oos} , on the persistency of the predictors in the nesting model ($V_{LP}(X_{2t})$), and it is necessary to simulate a second Brownian motion because of W_2 . In Section 5 next, we show that ignoring these additional terms may severely distort the empirical size of the ENC-NEW.

5. Monte Carlo Simulations

In this section, we consider four different sets of simulations to study the size properties of our test compared to some traditional out-of-sample tests: the ENC-NEW and the ENC-t with CM critical values, and the “Wild Clark and West” (WCW) of [26] (henceforth PHM).

The data generating process (DGP) in Section 5.1 is designed to study the effects of a bias both in-sample and out of sample (e.g., the benchmark model is the DRW). The experiment in Section 5.2 is akin to Section 5.1, but this time we set our parameters to match the empirical specifications of [22–24] using the Chilean peso and lead prices. In these two experiments, we consider the MENC-NEW according to Theorem 2.

Sections 5.3 and 5.4 consider two different DGPs that introduce an out-of-sample bias. In these cases, the benchmark model the random walk (with drift). As a consequence, there is no in-sample bias, and no re-centering term is required (according to Theorem 3). We use the ENC-NEW with corrected critical values in these two cases, following Theorem 4 (ENC-NEW*).

Finally, Section 5.5 studies the power properties of the MENC-NEW. To this end, we impose the alternative hypothesis over the DGP of Section 5.2.

5.1. Size Properties of the MENC-NEW: The Benchmark Model Is the DRW and $EY_t \neq 0$

To illustrate the effects of a bias in the nested model, we consider a simple DGP for both the target variable Y_t and the predictor X_t :

$$Y_{t+1} = \alpha + \beta X_t + \varepsilon_{t+1} \tag{7}$$

$$X_{t+1} = \rho X_t + u_{t+1} \tag{8}$$

where ε_{t+1} and u_{t+1} are drawn simply by independent standard normal distributions. The initial value Y_1 for each simulation is set at the unconditional expected value of Y_{t+1} ($EY_t = \alpha$). We consider two competing one-step-ahead forecasts. The first model is simply a zero-forecast (e.g., the DRW), and the second model uses exclusively X_t as a predictor with no drift (i.e., $Y_t^f(1) = \gamma X_t$).

To evaluate the empirical size of each test, we set $\beta = 0$. In this simple DGP, we introduce the bias through the parameter α , and we modify the persistency of the predictor X_{t+1} through ρ . If $\alpha = 0$, then the proper critical values for this setup are those of CM. We argue, however, that severe distortions may appear because of $\alpha \neq 0$, especially for highly persistent predictors. To illustrate the severity of these size distortions, we consider different sets of values for these two parameters α and ρ . We set α to either 0.3, 0.5, 0.7 or 1, and ρ to either 0.6 or 0.8. Additional simulations with different values of these parameters are available upon request. While this is an extremely simple DGP, it clearly illustrates the size distortions over traditional tests of out-of-sample evaluation.

We consider sample sizes of $T = 100, 150,$ and 200 . As the asymptotic distribution of most of these tests depends on how we split this sample, we consider $P/R = 2$ and 4 , using a recursive scheme of estimation (Additional results considering $P/R = 0.4, 1,$ and 3 are available upon request with similar conclusions.). We consider a total of 5000 Monte Carlo simulations for each exercise.

To obtain the critical values suggested by Theorem 2, we closely follow the simulations in CM. In each of the 5000 simulations, we generate 10,000 Brownian motions through simple random walks, each using an independent sequence of 10,000 i.i.d increments, drawn from a normal distribution, with a mean of 0 and a variance of $1/10,000$. Then, we obtain the integrals by summing the weighted quadratic of the random walks for the entire evaluation window (from $R + 1/T$ through T/T). In every Monte Carlo simulation, we estimate the nuisance parameters of the MENC-NEW with the synthetic sample data (such as the bias and the long-run variance of X_t), according to Theorem 2. Finally, the critical values are simply the 90th and 95th percentiles of our 10,000 simulated stochastic integrals for 10% and 5% significance levels, respectively.

To evaluate the empirical size of the ENC-t and ENC-NEW, we consider the critical values available in CM for the recursive scheme with the corresponding P/R and $k_2 = 1$. In a recent paper, PHM propose a new asymptotically normal test label as the Wild Clark and West (WCW). The authors show that parameter uncertainty in their framework is asymptotically irrelevant. The strategy of PHM is to introduce an independent random variable θ_t in the ENC-t core statistic that prevents degeneracy under the null hypothesis of no encompassing. PHM propose a gaussian distribution for θ_t ($\theta_t \sim N(1, \phi^2)$) such that

$$WCW - t = \sqrt{P-1} \frac{P^{-1} \sum_{t=R+1}^T \hat{e}_{1,t+1} (\hat{e}_{1,t+1} - \theta_t \hat{e}_{2,t+1})}{\sqrt{\widehat{S}_{ff}}}$$

where $\hat{e}_{1,t+1}$ and $\hat{e}_{2,t+1}$ are the forecasting errors of the nested and nesting models, respectively, and \widehat{S}_{ff} is a consistent estimate of the long-run variance of $\hat{e}_{1,t+1} (\hat{e}_{1,t+1} - \theta_t \hat{e}_{2,t+1})$. In this case, we consider the HAC estimator proposed by Newey and West (1987,1994) [35,36]. While PHM do not provide an optimal value for the tuning parameter ϕ , they give some recommendations. In particular, we set $\phi = \sqrt{V(\hat{e}_{2,t})} * 0.1$, as suggested by the authors (see PHM, page 6).

Tables 1 and 2 display our results for nominal sizes 10% and 5%, using $P/R = 2$, and 4 respectively. Each panel considers different parametrizations of our DGP. Three things about Tables 1 and 2 are worth mentioning. First, consistent with Theorem 2, the ENC-t and the ENC-NEW become severely oversized with an increasing bias and higher persistency on the predictor. Notably, the last panel of Table 2 reports an empirical size about *three times higher than the nominal size* for both statistics. The average size of the ENC-NEW across all exercises goes from 18.2% (11.5%) to 30.8% (21.6%) with a nominal size of 10% (5%). Second, the MENC-NEW is reasonably well-sized in most of our exercises. Notice that the size properties of our test improve significantly with larger samples; this is reasonable since the asymptotic distribution of our test relies on nuisance parameters estimated in each simulation. The average size of our test goes from 10.3% (5.4%) to 12.9% (7.5%) across all exercises. Moreover, if we focus on the simulations with the larger sample $T = 200$, our results range from 9.5% (4.9%) to 12.3% (6.9%). Finally, while the WCW-t is generally well-sized in these exercises, we make a note of caution. While we see improvements in size compared to the ENC-t and ENC-NEW, PHM show that these improvements come with a cost: the WCW tends to exhibit less power than other tests. Our results in Section 5.5 generally support this idea: the power of the WCW is sometimes less than a half than our test.

Table 1. Empirical Size with $P/R = 2$.

	Nominal Size 10%				Nominal Size 5%			
	ENC-NEW	ENC-t	WCW-t	MENC-NEW	ENC-NEW	ENC-t	WCW-t	MENC-NEW
				$\alpha = 0.3$	$\rho = 0.8$			
T = 100	0.192	0.195	0.134	0.136	0.119	0.122	0.069	0.074
T = 150	0.188	0.188	0.133	0.128	0.119	0.121	0.069	0.072
T = 200	0.187	0.184	0.114	0.123	0.119	0.116	0.061	0.065
Ave	0.189	0.189	0.127	0.129	0.119	0.120	0.066	0.070
				$\alpha = 0.5$	$\rho = 0.6$			
T = 100	0.183	0.187	0.116	0.107	0.114	0.117	0.060	0.057
T = 150	0.182	0.185	0.113	0.103	0.115	0.122	0.057	0.056
T = 200	0.182	0.181	0.105	0.100	0.116	0.112	0.053	0.050
Ave	0.182	0.184	0.111	0.103	0.115	0.117	0.057	0.054
				$\alpha = 0.7$	$\rho = 0.6$			
T = 100	0.233	0.227	0.112	0.109	0.157	0.156	0.064	0.064
T = 150	0.230	0.227	0.104	0.105	0.154	0.162	0.055	0.062
T = 200	0.230	0.220	0.101	0.097	0.154	0.148	0.047	0.051
Ave	0.231	0.225	0.106	0.104	0.155	0.155	0.055	0.059
				$\alpha = 1.0$	$\rho = 0.6$			
T = 100	0.291	0.280	0.113	0.114	0.208	0.198	0.062	0.067
T = 150	0.279	0.270	0.106	0.105	0.204	0.198	0.055	0.067
T = 200	0.276	0.261	0.098	0.096	0.201	0.186	0.050	0.053
Ave	0.282	0.270	0.106	0.105	0.204	0.194	0.056	0.062

Notes: Each entry report the percentage of rejections under the null hypothesis $\beta = 0$. The table’s left (right) side reports our results with a nominal size of 10% (5%). ρ is related to the persistency of the predictor and α with the magnitude of the bias. For each exercise we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the critical values (CV) for the MENC-NEW according to Theorem 2. CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1$, $P/R = 2$ and a recursive method.

Table 2. Empirical Size for P/R = 4.

	Nominal Size 10%				Nominal Size 5%			
	ENC-NEW	ENC-t	WCW-t	MENC-NEW	ENC-NEW	ENC-t	WCW-t	MENC-NEW
				$\alpha = 0.3$	$\rho = 0.8$			
T = 100	0.205	0.199	0.124	0.137	0.122	0.118	0.061	0.079
T = 150	0.201	0.195	0.120	0.131	0.121	0.122	0.064	0.078
T = 200	0.192	0.187	0.104	0.119	0.116	0.114	0.054	0.069
Ave	0.199	0.194	0.116	0.129	0.120	0.118	0.060	0.075
				$\alpha = 0.5$	$\rho = 0.6$			
T = 100	0.198	0.192	0.114	0.111	0.120	0.114	0.056	0.061
T = 150	0.193	0.193	0.106	0.105	0.121	0.121	0.054	0.059
T = 200	0.191	0.187	0.096	0.098	0.117	0.109	0.044	0.049
Ave	0.194	0.191	0.105	0.105	0.119	0.115	0.051	0.056
				$\alpha = 0.7$	$\rho = 0.6$			
T = 100	0.278	0.270	0.113	0.134	0.187	0.176	0.057	0.081
T = 150	0.276	0.261	0.107	0.126	0.193	0.183	0.056	0.077
T = 200	0.274	0.256	0.095	0.110	0.180	0.165	0.045	0.062
Ave	0.276	0.262	0.105	0.123	0.187	0.175	0.053	0.073
				$\alpha = 1.0$	$\rho = 0.6$			
T = 100	0.321	0.301	0.103	0.120	0.218	0.209	0.052	0.067
T = 150	0.302	0.288	0.095	0.114	0.220	0.206	0.052	0.064
T = 200	0.301	0.284	0.089	0.095	0.209	0.189	0.043	0.052
Ave	0.308	0.291	0.096	0.110	0.216	0.201	0.049	0.061

Notes: Each entry report the percentage of rejections under the null hypothesis $\beta = 0$. The table’s left (right) side reports our results with a nominal size of 10% (5%). ρ is related to the persistency of the predictor and α with the magnitude of the bias. For each exercise we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the critical values (CV) for the MENC-NEW according to Theorem 2. CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1$, $P/R = 4$ and a recursive method.

5.2. Size Properties of the MENC-NEW with a DGP

In this section, we use the same DGP as Section 5.1, but we set the parameters in our simulation to match the econometric setup of [22–24] (PH). Based on the present-value model for exchange rate determination (Campbell and Shiller (1987) [37] and Engel and West (2005) [38]), PH show that the Chilean exchange rate has the ability to predict base-metal prices. Table 2 in PH suggest the following econometric specification:

$$\Delta \ln(CP_t) = \beta[\Delta \ln(ER_{t-1}) + \Delta \ln(ER_{t-2})] + \varepsilon_t$$

where ER_t stands for the Chilean peso and CP_t for a generic commodity price (in this case, we use lead prices). PH use this econometric specification to determine whether this model can outperform the zero-forecast model (e.g., they evaluate the null hypothesis $H_0 : \beta = 0$). We argue that some distortions may appear if $E[\Delta \ln(CP_t)] \neq 0$, as both models are biased under the null hypothesis.

Let $Y_t = \Delta \ln(CP_t)$ and $X_t = [\Delta \ln(ER_{t-1}) + \Delta \ln(ER_{t-2})]$. Using the annualized monthly returns of the Chilean peso and lead prices from September 1999 through August 2020, and using the same DGP of Section 5.1 (under the null hypothesis $\beta = 0$), we estimate $\rho, \sigma_\varepsilon^2 = V(\varepsilon_{t+1}), \sigma_u^2 = V(u_{t+1})$, and $\delta = \text{Corr}(\varepsilon_{t+1}, u_{t+1})$:

$$Y_t = \alpha + \beta X_t + \varepsilon_{t+1} \tag{9}$$

$$X_t = 0.53X_{t-1} + u_{t+1} \tag{10}$$

where $\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} \rightarrow N\left(\begin{matrix} \sigma_\varepsilon^2 & \delta\sigma_u\sigma_\varepsilon \\ \delta\sigma_u\sigma_\varepsilon & \sigma_u^2 \end{matrix}\right)$ with $\delta = -0.36, \sigma_\varepsilon = 1.05$, and $\sigma_u = 0.50$. To introduce the bias, we consider two different values for α (0.3 and 0.6). Akin to Section 5.1,

we compare out-of-sample the zero-forecast against a model that uses exclusively X_t ($Y_t^f(1) = \gamma X_t$). We estimate recursively by OLS the parameter γ . Tables 3 and 4 exhibit our results for $P/R = 1$ and 2, respectively. Additional exercises are available upon request. Notice that the only difference in each panel is the magnitude of the bias.

Table 3. Empirical Size for $P/R = 1$.

	Nominal Size 10%				Nominal Size 5%			
	ENC-NEW	ENC-t	WCW-t	MENC-NEW	ENC-NEW	ENC-t	WCW-t	MENC-NEW
	Bias = 0.3							
T = 100	0.127	0.136	0.142	0.105	0.075	0.082	0.075	0.059
T = 150	0.120	0.134	0.132	0.100	0.074	0.081	0.073	0.054
T = 200	0.121	0.125	0.121	0.097	0.064	0.073	0.059	0.043
Ave	0.123	0.132	0.132	0.101	0.071	0.079	0.069	0.052
	Bias = 0.6							
T = 100	0.177	0.186	0.133	0.105	0.113	0.120	0.065	0.061
T = 150	0.175	0.183	0.123	0.103	0.112	0.119	0.058	0.052
T = 200	0.169	0.172	0.113	0.091	0.106	0.108	0.056	0.044
Ave	0.174	0.180	0.123	0.100	0.110	0.116	0.060	0.052

Notes: Each entry report the percentage of rejections under the null hypothesis. The table’s left (right) side reports our results with a nominal size of 10% (5%). We introduce the bias in this GDP through the parameter α . For each exercise, we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the critical values (CV) for the MENC-NEW according to Theorem 2. CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1$, $P/R = 1$ and a recursive method.

The main conclusions of Tables 3 and 4 are similar to those in Tables 1 and 2. First, consistent with Theorem 2, the ENC-t and ENC-NEW exhibit substantial size distortions when the bias increases. The average size of both tests is ~12–13% (7%) when $\alpha = 0.3$. Nevertheless, when the bias increases ($\alpha = 0.6$), the average size of both tests is ~17–18% (11%). Notably, the MENC-NEW seems to be reasonably well-sized in all our exercises. Specifically, the empirical size of our test ranges between 8.5 and 10.5% (4.3–6.1%), with an average empirical size across all our exercises of 9.8% (5.1%). Akin to the previous section, the size of our test seems to improve with the sample size; this is reasonable since we need to estimate some nuisance parameters to determine our critical values. Finally, the WCW exhibit mixed results compared to the ENC-t and ENC-NEW. In particular, the WCW is less oversized for $\alpha = 0.6$, but it tends to underperform for $\alpha = 0.3$.

5.3. Size Properties of the ENC-NEW with Adjusted Critical Values: The Effects of an Out-of-Sample Bias

In this DGP we introduce an out-of-sample bias by shifting the expected value of the target variable Y_t . The key point is that this shift happens in the evaluation window $t > R$. As the parameters are recursively estimated in each period, the OLS estimator will not fully accommodate this shift in small samples. As a consequence, we will observe an out-of-sample bias in both models under the null hypothesis. In contrast to Section 5.1, there is no need to re-center the ENC-NEW, as suggested by Theorem 3. Nevertheless, the asymptotic distribution of the ENC-NEW is now given by Theorem 4b, and it is necessary to obtain new critical values.

Table 4. Empirical Size for P/R = 2.

	Nominal Size 10%				Nominal Size 5%			
	ENC-NEW	ENC-t	WCW-t	MENC-NEW	ENC-NEW	ENC-t	WCW-t	MENC-NEW
				Bias = 0.3				
T = 100	0.120	0.127	0.128	0.098	0.069	0.073	0.070	0.056
T = 150	0.127	0.134	0.129	0.101	0.069	0.079	0.060	0.051
T = 200	0.113	0.121	0.121	0.090	0.062	0.069	0.061	0.043
Ave	0.120	0.127	0.126	0.096	0.067	0.074	0.064	0.050
				Bias = 0.6				
T = 100	0.174	0.175	0.120	0.097	0.105	0.114	0.058	0.057
T = 150	0.180	0.183	0.112	0.100	0.115	0.124	0.058	0.054
T = 200	0.176	0.177	0.101	0.085	0.101	0.104	0.049	0.045
Ave	0.177	0.178	0.111	0.094	0.107	0.114	0.055	0.052

Notes: Each entry report the percentage of rejections under the null hypothesis $\beta = 0$. The table’s left (right) side reports our results with a nominal size of 10% (5%). We introduce the bias in this GDP through the parameter α . For each exercise, we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the critical values (CV) for the MENC-NEW according to Theorem 2. CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1$, P/R = 2 and a recursive method.

To illustrate the effects of this bias, we introduce this shift immediately after the estimation window at $t = R + 1$ through a change in the drift. Let α_1 and α_2 be the drifts before and after the shift, respectively:

$$Y_t = \alpha_1 + \beta X_{t-1} + \varepsilon_t, t \leq R$$

$$Y_t = \alpha_2 + \beta X_{t-1} + \varepsilon_t, t > R + 1$$

where ε_t is an independent standard normal random variable and X_t is a stationary AR(1) process

$$X_t = \alpha_x + \rho X_{t-1} + v_t, \forall t$$

We consider two competing models for the target variable Y_t

$$Y_t = c_1 + v_{1,t} (M1)$$

$$Y_t = c_2 + \gamma X_{t-1} + v_{2,t} (M2)$$

where c_1, c_2 , and γ are regression parameters, and $v_{1,t}, v_{2,t}$ stand for the forecast errors. Notice that M1 forecasts Y_t using its historical sample mean, and it is identical to M2 whenever $\gamma = 0$. As the parameters in M1 and M2 are estimated in-sample in each recursive window by OLS, we should expect no in-sample bias whatsoever. Nevertheless, the OLS estimator will not fully capture the out-of-sample change on the drift in small samples. According to Theorem 3, if there is no in-sample bias, then the re-centering term of the MENC-NEW is unnecessary. In this case, we may simply consider the ENC-NEW with corrected critical values (as suggested by the asymptotic distribution in Theorem 4b). We consider different parametrizations for this DGP. In particular, we allow different levels of persistency on the predictor X_t ($\rho = 0.4$ and 0.8), and different changes on the drift parameter ($\alpha_1 = 0 \alpha_2 = 0.3, \alpha_1 = 0 \alpha_2 = 0.3$, and $\alpha_1 = 0 \alpha_2 = 0.3$). The idea of using different changes on the drift is to manage the magnitude of the out-of-sample bias.

Similar to Sections 5.1 and 5.2, we consider 5000 Monte Carlo simulations using a recursive scheme with P/R = 1 and T = 100, 150, and 200. To evaluate the empirical size of each test, we set $\beta = 0$. To simulate the corrected critical values of the ENC-NEW, we consider the asymptotic distribution suggested by Theorem 4b. Akin to the previous section, in each of the 5000 simulations, we generate 10,000 Brownian motions, each using an independent sequence of 10,000 i.i.d increments, drawn from a normal

distribution, with mean 0 and variance 1/10,000. Then, we obtain the integrals by summing the weighted quadratic of the random walks for the entire evaluation window (from $R + 1/T$ through T/T). We estimate the respective nuisance parameters in each simulation using the simulated data, according to Theorem 4b. The critical values are the 90th and 95th percentiles of our 10,000 simulated stochastic integrals for 10% and 5% significance levels, respectively.

Each panel in Table 5 exhibits our results using different parameterizations of our DGP. Three features in Table 5 are worth mentioning. First, the ENC-NEW with corrected critical values (ENC-NEW*) is generally well-sized in all exercises. In particular, the average size of the ENC-NEW* goes from 8.8% (4.4%) to 11.5% (6.1%), where the most oversized exercise exhibits a rejection of 11.6% (6.6%), and the average size across all exercises is 10.0% (5.16%). Second, the ENC-NEW and the ENC-t are again significantly more oversized. For instance, the average size of the ENC-t goes from 13.0% (7.3%) to 14.9% (9.0%), with an average size across all exercises of 13.7% (7.87%). Finally, consistent with Theorem 4, the worsts results for the ENC-NEW and ENC-t are those in simulations with a stronger persistency in the predictor and a bigger positive shift in the drift (first panel). In these exercises, the ENC-t ranges from 14.5% (8.8%) through 15.6% (9.3%), with an average size of 14.9% (9.0%). In sharp contrast, the ENC-NEW with corrected critical values has an empirical size ranging from 11.3% (5.0%) to 11.6% (6.6%), with an average size of 11.5% (6.1%).

Table 5. Empirical Size with an out-of-sample bias and $P/R = 1$.

	Nominal Size 10%			Nominal Size 5%				
	ENC-NEW	ENC-t	WCW-t	ENC-NEW*	ENC-NEW	ENC-t	WCW-t	ENC-NEW*
$\rho = 0.8 \alpha_1 = 0 \alpha_2 = 0.3 \alpha_X = 0.2$								
T = 100	0.142	0.145	0.141	0.116	0.086	0.088	0.077	0.066
T = 150	0.146	0.156	0.136	0.115	0.086	0.093	0.074	0.060
T = 200	0.141	0.145	0.121	0.113	0.088	0.088	0.067	0.058
Ave	0.143	0.149	0.133	0.115	0.087	0.090	0.073	0.061
$\rho = 0.8 \alpha_1 = 0.1 \alpha_2 = 0.3 \alpha_X = 0.2$								
T = 100	0.130	0.126	0.141	0.104	0.073	0.072	0.078	0.054
T = 150	0.128	0.135	0.135	0.099	0.073	0.076	0.072	0.048
T = 200	0.125	0.129	0.127	0.098	0.073	0.072	0.072	0.049
Ave	0.128	0.130	0.134	0.100	0.073	0.073	0.074	0.050
$\rho = 0.4 \alpha_1 = 0.2 \alpha_2 = -0.3 \alpha_X = 0.4$								
T = 100	0.120	0.136	0.137	0.094	0.069	0.077	0.074	0.048
T = 150	0.112	0.126	0.132	0.080	0.060	0.071	0.071	0.040
T = 200	0.121	0.131	0.128	0.090	0.068	0.071	0.069	0.045
Ave	0.118	0.131	0.132	0.088	0.066	0.073	0.071	0.044

Notes: Each entry report the percentage of rejections under the null hypothesis $\beta = 0$. The table's left (right) side reports our results with a nominal size of 10% (5%). We introduce the out-of-sample bias in this GDP through the parameters α_1 (the initial drift of Y_t) and α_2 (the drift of Y_t after the change). The parameter ρ is related to the persistency of the predictor X_t . α_X denotes the drift of the predictor. For each exercise, we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the corrected critical values (CV) for the ENC-NEW according to Theorem 2 (ENC-NEW*). CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1$, $P/R = 1$ and a recursive method.

5.4. Size Properties of the ENC-NEW with Adjusted Critical Values: Persistency in the Drift

In this DGP, we introduce an out-of-sample bias by drifting the expected value of the target variable Y_t . The key point is that the drift in Y_t follows a persistent, although stationary, AR(1) process. As the parameters are recursively estimated in each period, the OLS estimator will not be able to accommodate this shift in small samples, and as a consequence, we will observe an out-of-sample bias in both models under the null hypothesis. In contrast to Section 5.1, there is no in-sample bias, and there is no need

to re-center the ENC-NEW, as suggested by Theorem 3. Nevertheless, the asymptotic distribution of the ENC-NEW is now given by Theorem 4b.

Consider the following DGP for the target variable Y_t and the predictor X_t

$$Y_t = \alpha_{y,t} + \beta X_{t-1} + u_t$$

$$X_t = \alpha_x + \rho X_{t-1} + \varepsilon_t$$

where u_t and ε_t are independent standard normal random variables. We consider two competing models

$$Y_t = c_1 + v_{1,t} \quad (M1)$$

$$Y_t = c_2 + \gamma X_{t-1} + v_{2,t} \quad (M2)$$

where $c_1, c_2,$ and γ are simply regression parameters, and $v_{1,t}, v_{2,t}$ stand for the forecast errors. Notice that M1 forecasts Y_t using its historical sample mean, and it is identical to M2 whenever $\gamma = 0$. The interesting thing about this DGP is that we allow the drift of Y_t to follow a persistent (although stationary) AR(1) process

$$\alpha_{y,t} = \alpha_\alpha + \rho_\alpha \alpha_{y,t-1} + e_t$$

where e_t is an independent standard normal random variable. As the parameters in M1 and M2 are estimated in-sample in each recursive window by OLS, we should expect no in-sample bias whatsoever. Nevertheless, the OLS estimator will not capture the out-of-sample change on the drift. As the drift is persistently changing in each period, it may introduce an out-of-sample bias. According to Theorem 4b, if there is no in-sample bias, then the re-centering term of the MENC-NEW is unnecessary. In this case, we may simply consider the ENC-NEW with corrected critical values (as suggested by the asymptotic distribution in Theorem 4b).

Tables 6 and 7 report our results for $P/R = 1$ and 2, respectively (Additional results are available upon request.). Consistent with our previous simulations, the ENC-NEW with corrected critical values is reasonably well-sized, especially with large sample sizes. The size of the ENC-NEW* ranges between 7.3% (3.0%) to 13.1% (7.2%), with an average size across all exercises of 10.5% (5.2%). In sharp contrast, the ENC-NEW and the ENC-t are severely oversized in all our simulations: some exercises exhibit an empirical size four times bigger than its nominal size. For instance, the last panel of Table 7 shows an average empirical size of 31.2% (22.8%) and 31.2% (23.3%) for the ENC-NEW and ENC-t, respectively. Moreover, the average size of the ENC-NEW across all exercises is 26.3% (18.6%), more than two (three) times its nominal size. Finally, the WCW-t seems to be correctly sized for most of our exercises, with an empirical size ranging from 9.1% (4.6%) to 13.1% (6.8%).

5.5. Power Properties of the MENC-NEW

In this section, we provide some simulations evaluating the power properties of the MENC-NEW compared to the ENC-NEW, ENC-t, and WCW-t. We use the same DGP of Section 5.2, but this time we impose the alternative hypothesis setting $\beta \neq 0$. Tables 8 and 9 next exhibit our power results introducing a bias of $\alpha = 0.15, 0.30,$ and 0.60 , considering $P/R = 1$ and 2, respectively. In both tables, we impose the alternative by setting $\beta = 0.6$. Additional simulations using different values for β deliver a similar message, and they are available upon request.

Table 6. Empirical size for P/R = 1.

	Nominal Size 10%				Nominal Size 5%			
	ENC-NEW	ENC-t	WCW-t	ENC-NEW*	ENC-NEW	ENC-t	WCW-t	ENC-NEW*
	$\rho = 0.4 \rho_{\alpha} = 0.9 \alpha_{\alpha} = 0.01 \alpha_X = 0.3$							
T = 100	0.219	0.235	0.120	0.125	0.145	0.157	0.060	0.065
T = 150	0.207	0.220	0.112	0.106	0.137	0.141	0.056	0.049
T = 200	0.215	0.218	0.107	0.104	0.149	0.151	0.058	0.044
Ave	0.214	0.224	0.113	0.112	0.144	0.150	0.058	0.053
	$\rho = 0.5 \rho_{\alpha} = 0.95 \alpha_{\alpha} = 0.01 \alpha_X = 0.3$							
T = 100	0.264	0.276	0.128	0.131	0.189	0.204	0.063	0.072
T = 150	0.260	0.271	0.108	0.110	0.186	0.193	0.054	0.053
T = 200	0.270	0.274	0.115	0.105	0.197	0.201	0.058	0.052
Ave	0.265	0.274	0.117	0.115	0.191	0.199	0.058	0.059
	$\rho = 0.6 \rho_{\alpha} = 0.9 \alpha_{\alpha} = 0.05 \alpha_X = 0.4$							
T = 100	0.303	0.302	0.131	0.115	0.222	0.229	0.068	0.064
T = 150	0.289	0.292	0.113	0.090	0.215	0.214	0.060	0.042
T = 200	0.300	0.294	0.108	0.085	0.225	0.221	0.057	0.040
Ave	0.297	0.296	0.117	0.097	0.221	0.221	0.062	0.049

Notes: Each entry report the percentage of rejections under the null hypothesis $\beta = 0$. The table’s left (right) side reports our results with a nominal size of 10% (5%). ρ_{α} and α_{α} denote the persistency and the drift of $\alpha_{y,t}$, respectively. The parameter ρ is related to the persistency of the predictor X_t . α_X denotes the drift of the predictor. For each exercise, we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the corrected critical values (CV) for the ENC-NEW according to Theorem 2 (ENC-NEW*). CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1, P/R = 1$ and a recursive method.

Table 7. Empirical Size for P/R = 2.

	Nominal Size 10%				Nominal Size 5%			
	ENC-NEW	ENC-t	WCW-t	ENC-NEW*	ENC-NEW	ENC-t	WCW-t	ENC-NEW*
	$\rho = 0.4 \rho_{\alpha} = 0.9 \alpha_{\alpha} = 0.01 \alpha_X = 0.3$							
T = 100	0.216	0.236	0.111	0.121	0.145	0.156	0.053	0.063
T = 150	0.210	0.219	0.096	0.104	0.140	0.145	0.047	0.050
T = 200	0.220	0.218	0.096	0.099	0.148	0.155	0.048	0.042
Ave	0.215	0.224	0.101	0.108	0.144	0.152	0.049	0.052
	$\rho = 0.5 \rho_{\alpha} = 0.95 \alpha_{\alpha} = 0.01 \alpha_X = 0.3$							
T = 100	0.273	0.291	0.107	0.126	0.192	0.208	0.056	0.064
T = 150	0.272	0.285	0.092	0.103	0.186	0.203	0.046	0.050
T = 200	0.278	0.286	0.096	0.099	0.197	0.207	0.047	0.047
Ave	0.274	0.287	0.098	0.109	0.192	0.206	0.050	0.054
	$\rho = 0.6 \rho_{\alpha} = 0.9 \alpha_{\alpha} = 0.05 \alpha_X = 0.4$							
T = 100	0.316	0.319	0.116	0.115	0.226	0.239	0.062	0.059
T = 150	0.309	0.314	0.101	0.086	0.225	0.227	0.052	0.040
T = 200	0.310	0.303	0.091	0.073	0.234	0.232	0.049	0.030
Ave	0.312	0.312	0.103	0.091	0.228	0.233	0.054	0.043

Notes: Each entry report the percentage of rejections under the null hypothesis $\beta = 0$. The table’s left (right) side reports our results with a nominal size of 10% (5%). ρ_{α} and α_{α} denote the persistency and the drift of $\alpha_{y,t}$, respectively. The parameter ρ is related to the persistency of the predictor X_t . α_X denotes the drift of the predictor. For each exercise, we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the corrected critical values (CV) for the ENC-NEW according to Theorem 2 (ENC-NEW*). CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1, P/R = 2$ and a recursive method.

Table 8. Raw power with nominal sizes 10% and 5%, and P/R = 1.

	Nominal Size 10%				Nominal Size 5%			
	ENC-NEW	ENC-t	WCW-t	MENC-NEW	ENC-NEW	ENC-t	WCW-t	MENC-NEW
Bias = 0.15								
T = 100	0.925	0.883	0.419	0.923	0.878	0.801	0.364	0.871
T = 150	0.985	0.963	0.467	0.984	0.969	0.919	0.440	0.967
T = 200	0.997	0.991	0.501	0.997	0.994	0.976	0.489	0.993
Ave	0.969	0.946	0.462	0.968	0.947	0.899	0.431	0.944
Bias = 0.3								
T = 100	0.905	0.856	0.395	0.885	0.848	0.765	0.338	0.823
T = 150	0.970	0.945	0.453	0.963	0.947	0.890	0.417	0.933
T = 200	0.993	0.984	0.493	0.991	0.988	0.955	0.471	0.982
Ave	0.956	0.928	0.447	0.947	0.928	0.870	0.409	0.913
Bias = 0.6								
T = 100	0.807	0.751	0.330	0.738	0.739	0.659	0.267	0.630
T = 150	0.909	0.873	0.395	0.861	0.868	0.802	0.343	0.792
T = 200	0.962	0.931	0.450	0.928	0.936	0.885	0.407	0.892
Ave	0.892	0.851	0.392	0.842	0.847	0.782	0.339	0.771

Notes: Each entry report the percentage of rejections under the alternative hypothesis $\beta = 0.6$. The table's left (right) side reports our results with a nominal size of 10% (5%). We introduce the bias in this GDP through the parameter α . For each exercise, we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the critical values (CV) for the MENC-NEW according to Theorem 2. CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1$, P/R = 1 and a recursive method.

Table 9. Raw power with nominal sizes 10% and 5%, and P/R = 2.

	Nominal Size 10%				Nominal Size 5%			
	ENC-NEW	ENC-t	WCW-t	MENC-NEW	ENC-NEW	ENC-t	WCW-t	MENC-NEW
Bias = 0.15								
T = 100	0.938	0.929	0.450	0.934	0.896	0.870	0.409	0.893
T = 150	0.985	0.980	0.485	0.985	0.973	0.961	0.465	0.971
T = 200	0.997	0.997	0.507	0.996	0.994	0.989	0.501	0.994
Ave	0.973	0.969	0.481	0.972	0.954	0.940	0.458	0.953
Bias = 0.3								
T = 100	0.914	0.906	0.432	0.900	0.865	0.835	0.385	0.837
T = 150	0.976	0.968	0.473	0.969	0.954	0.937	0.440	0.941
T = 200	0.995	0.992	0.502	0.994	0.989	0.978	0.488	0.982
Ave	0.962	0.955	0.469	0.954	0.936	0.917	0.438	0.920
Bias = 0.6								
T = 100	0.830	0.814	0.364	0.748	0.757	0.728	0.303	0.655
T = 150	0.918	0.905	0.416	0.864	0.875	0.847	0.373	0.795
T = 200	0.969	0.958	0.468	0.938	0.946	0.925	0.438	0.902
Ave	0.906	0.892	0.416	0.850	0.859	0.833	0.372	0.784

Notes: Each entry report the percentage of rejections under the alternative hypothesis $\beta = 0.6$. The table's left (right) side reports our results with a nominal size of 10% (5%). We introduce the bias in this GDP through the parameter α . For each exercise, we consider 5000 Monte Carlo simulations. T stands for the sample size. We use a recursive scheme to update our parameters. P and R stand for the number of observations in the evaluation and estimation windows, respectively. We determine the critical values (CV) for the MENC-NEW according to Theorem 2. CV for the WCW are standard normal, while CV for the ENC-t and the ENC-NEW are those tabulated in [5] for $k_2 = 1$, P/R = 2 and a recursive method.

First, consistent with [26] (PHM), the size improvements of the WCW-t come with a cost: a deterioration in power. As commented by PHM: "In terms of power, results have been rather mixed, although CW has frequently exhibited some more power. All in all, our simulations reveal that asymptotic normality and size corrections come with a cost: the introduction of a random variable erodes some of the power of WCW." [26], page 3. In particular, the power of WCW-t

ranges from 33.0% (26.7%) to 50.7% (50.1%), with an average power of 44.5% (40.7%). Second, the ENC-NEW, ENC-t, and MENC-NEW exhibit significantly more power than WCW-t, generally more than double. For instance, the average power of the MENC-NEW is 92.2% (88%). Finally, while the power of the ENC-NEW and the MENC-NEW are very similar, we do observe a small edge on the ENC-NEW. This is, of course, reasonable as the ENC-NEW and ENC-t are extremely oversized in this environment, according to the results of Section 5.2. Not surprisingly, the differences in power become more important with a higher bias (e.g., last panel): the advantage of the ENC-NEW in terms of power coincides with the ENC-NEW being particularly more oversized. We see the results of Tables 8 and 9 encouraging since the differences in power are, most of the time, neglectable, even though the size of the ENC-NEW is sometimes more than double than the nominal size.

6. Concluding Remarks

In this paper, we show analytically and via simulations that a bias in the null model may severely distort the asymptotic distribution of some traditional tests of forecast evaluation in nested models comparisons. These distortions are more severe with an increasing bias and a higher persistency on the additional predictors. We address two relevant cases: (i) The presence of a bias in-sample and out-of-sample (e.g., the driftless random walk as the benchmark model). (ii) The presence of a bias exclusively out-of-sample (e.g., a shift in the expected value of the target variable). To deal with the former case, we consider a simple modification of the ENC-NEW (MENC-NEW) robust to a bias in the null model. In essence, the MENC-NEW introduces a re-centering term in the ENC-NEW that appears because of the additional predictors' bias and persistency. In both cases, the relevant asymptotic distribution is not pivotal; akin to CM, they are functionals of stochastic integrals, but they also depend on the magnitude of the bias and the persistency of the predictors.

Based on CM, we derive the new distribution for the ENC-NEW imposing a bias in the null model (either out-of-sample or both in-sample and out-of-sample). While this is a subtle change in the set of assumptions, it has important implications over the asymptotic theory. In particular, the quadratic Brownian motions in CM arise because of the martingale difference terms: put simply, the orthogonality condition is assumed to hold both in-sample and out-of-sample. In contrast, in our case, the quadratic Brownian motions arise because of both the predictors and the martingale difference terms. Moreover, if the bias appears both in-sample and out-of-sample, the persistency of the predictors shift the expected value of the integrals of quadratic Brownian motions, thus a re-centering is required: our MENC-NEW is simply a re-centered version of the ENC-NEW. Of course, in the absence of a bias, our MENC-NEW reduces to the ENC-NEW. As expected, we show that this new asymptotic distribution depends on the magnitude of the bias and the persistency of the predictors. Even though the asymptotic distribution of the MENC-NEW (and the ENC-NEW with adjusted critical values) is not pivotal, the nuisance parameters can be easily estimated (one of these nuisance parameters being, of course, the magnitude of the bias).

Our Monte Carlo simulations reveal that our MENC-NEW (and the ENC-NEW with adjusted critical values) is reasonably well-sized even when the ENC-NEW and ENC-t (with CM critical values) exhibits rejections three to four times higher than the nominal size. The severity of the size distortion is, of course, related to the magnitude of the bias, as demonstrated by our decompositions.

Our results are important since the "correct specification" assumption in CM is often overlooked in empirical works. We show that ignoring this type of misspecification (bias) may severely change the asymptotic distribution of traditional tests. We suggest four interesting avenues for future research: First, it is important to study the effects of other cases of misspecification. For instance, a usual finding in the empirical literature is that forecasts are often auto-inefficient, in the sense that $\text{Corr}(X_t, u_{t+1}) \neq 0$. This type of misspecification may also affect the asymptotic distribution of out-of-sample tests and, consequently, distort the size and power properties. Second, in our view, the main caveat of our approach is that the asymptotic distribution of our test is not pivotal. In this sense,

the empirical researcher must simulate its critical values to conduct inference correctly. An interesting contribution may be the development of asymptotically normal tests, or at least an approach free of nuisance parameters, in order to simplify the forecast evaluation. Third, the extension of our approach to a direct multi-step-ahead framework, along the lines of [11]. Fourth, akin to CM, all our derivations are based on non-linear least squares (where the least square estimator is, of course, a special case). It may be interesting to study how the asymptotic distribution of these tests is affected by different estimation methods (e.g., ridge regressions).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proof Lemma 1

Lemma 1e is simply the combination of Lemma 1a through 1d. We show first our result for Lemma 1(a).

Proof Lemma 1(a). From (1):

$$\sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right)$$

If we add and subtract $(EX_t X_t')^{-1}$

$$\begin{aligned} & \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right) \\ &= \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' (EX_t X_t')^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right) \\ &+ T^{-0.5} \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' T^{0.5} \left\{ \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} - (EX_t X_t')^{-1} \right\} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right) \end{aligned}$$

It suffices then to show that the last term on the RHS is $o_p(1)$. From Lemma A1(b) in [33] $\sup_t T^{0.5} |vec \left[\left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} \right] - vec[(EX_t X_t')^{-1}]|$ is $O_p(1)$.

Then, that $\sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' T^{0.5} \left\{ \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} - (EX_t X_t')^{-1} \right\} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right)$ is $O_p(1)$ follows from the fact that $X_j u_{j+1}$ is a martingale difference, Theorem 7.19 in [34], from Corollary 29.19 of Davidson (1994) [39] and Theorem 3.1 of Hansen (1992) [40]. Then, the proof is complete since $T^{-0.5}$ is $o_p(1)$. □

As expected, the previous proof is akin to Lemma A2 from [33]. The remainder of the proof is similar with one crucial difference: (2–4) depend on X_t rather than the martingale difference sequence $X_t u_{t+1}$. In other words, these terms depend not only on the generalized forecast errors, but also on the additional predictor X_t . The proofs for Lemma 1b through 1d are very similar and follow the same arguments, hence we only show our proofs for Lemma 1b.

Proof Lemma 1(b). From the proof of Lemma 1a, it follows that

$$= \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' (EX_t X_t')^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right)$$

$$+T^{-0.5} \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' T^{0.5} \left\{ \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} - (EX_t X_t')^{-1} \right\} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right)$$

Moreover, again, we only need to prove that the last term on the RHS is $o_p(1)$. From Lemma A1(b) in [33] $\sup_t T^{0.5} |vec \left[\left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} \right] - vec[(EX_t X_t')^{-1}]|$ is $O_p(1)$.

Then, that $\sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' T^{0.5} \left\{ \left(\frac{1}{t} \sum_{j=1}^{t-1} X_j X_j'\right)^{-1} - (EX_t X_t')^{-1} \right\} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right)$ is $O_p(1)$ follows from the fact that X_t is covariance stationary, Theorem 7.17 in [34], Corollary 29.19 of [39], and Equation (3) in [40] (mixing sequences). Then, the proof is complete since $T^{-0.5}$ is $o_p(1)$. \square

Notice that this result is similar to Lemma A2 in [33], with one crucial difference: we apply the FCLT for stationary ergodic series (to X_t), rather than the FCLT for a martingale difference sequence.

Appendix B. Proof Theorem 1

Proof Theorem 1(a). Let $\frac{R}{T} \rightarrow \lambda$ when $T, R \rightarrow \infty$. Since $X_j u_{j+1}$ is a martingale difference, from Theorem 7.19 in White (2014), the Invariance Principle (FCLT) for martingale differences applies (See McLeish (1974) [41] and Hall (1977) [42]). From Corollary 29.19 of [39] it follows that

$$T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1} \rightarrow \sqrt{EX_t X_t'} \sigma_u W_1(s)$$

From the Continuous Mapping theorem

$$(EX_t X_t')^{-1} \left(\frac{T}{t}\right) \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right) \rightarrow s^{-1} (EX_t X_t')^{-0.5} \sigma_u W_1'(s)$$

Finally, from Theorem 3.1 of Hansen (1992), it follows that

$$\sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} u_{t+1} X_t' (EX_t X_t')^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right) \rightarrow \int_{\lambda}^1 s^{-1} W_1'(s) \sigma_u^2 dW_1(s)$$

\square

This result is akin to Lemma A6 in [33] for the numerator of the ENCNEW with a recursive scheme. We may interpret this term as the asymptotic distribution of the ENCNEW when no bias is present.

Proof Theorem 1(b). That $T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1} \rightarrow \sqrt{EX_t X_t'} \sigma_u W_1(s)$ follows immediately from Proof of Theorem 1a. From the continuous mapping theorem

$$\Delta \left(\frac{T}{t}\right) (EX_t X_t')^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right) \rightarrow \Delta s^{-1} (EX_t X_t')^{-0.5} \sigma_u W_1(s)$$

As X_t is covariance stationary, it is also global covariance stationary (See Definition 7.14 in [34]). According to Theorem 7.17 in [34], we may apply the FCLT for stationary ergodic series. Then, from Corollary 29.19 of [39] and Equation (3) in [40] (mixing sequences), it follows that

$$\begin{aligned} \Delta \sum_{t=R}^{T-1} \left(\frac{T}{t}\right) T^{-0.5} X_t' (EX_t X_t')^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j u_{j+1}\right) \\ \rightarrow \Delta \int_{\lambda}^1 s^{-1} W_1'(s) (EX_t X_t')^{-0.5} \sigma_u V_{LP}(X)^{0.5} dW_2(s) \end{aligned}$$

\square

Proof Theorem 1(c). This proof is similar to that of Theorem 1b. We use the Donsker principle for stationary ergodic series (Theorem 7.17 in [34]) so that

$$T^{-0.5} \sum_{j=1}^{t-1} X_j \rightarrow V_{LP}(X)^{0.5} \sigma_u W_2(s)$$

From the continuous mapping theorem

$$\Delta \left(\frac{T}{t} \right) (EX_t X'_t)^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j \right) \rightarrow \Delta s^{-1} (EX_t X'_t)^{-1} V_{LP}(X)^{0.5} \sigma_u W_2(s)$$

Then, the desired result follows from Corollary 29.19 of [39] and Equation (3) in [40]

$$\begin{aligned} & \Delta \sum_{t=R}^{T-1} \left(\frac{T}{t} \right) T^{-0.5} u_{t+1} X'_t (EX_t X'_t)^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j \right) \\ & \rightarrow \Delta \int_{\lambda}^1 s^{-1} W'_2(s) \sqrt{V_{LP}(X)} \sigma_u (EX_t X'_t)^{-0.5} dW_1(s) \end{aligned}$$

□

Proof Theorem 1(d). That $\Delta^2 \left(\frac{T}{t} \right) (EX_t X'_t)^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j \right) \Delta^2 s^{-1} (EX_t X'_t)^{-1} V_{LP}(X)^{0.5} \sigma_u W_2(s)$

Follows directly from the proof of Theorem 1c. From Corollary 29.19 of [39] and Equation (3) in [40]

$$\begin{aligned} & \Delta^2 \sum_{t=R}^{T-1} \left(\frac{T}{t} \right) T^{-0.5} X'_t (EX_t X'_t)^{-1} \left(T^{-0.5} \sum_{j=1}^{t-1} X_j \right) \\ & \rightarrow \Delta^2 \int_{\lambda}^1 s^{-1} W'_2(s) V_{LP}(X) (EX_t X'_t)^{-1} dW_2(s) + \Lambda_{\lambda} \end{aligned}$$

□

Proof Theorem 1(e). Theorem 1e is simply the combination of Theorem 1a through 1d. □

Appendix C

Proof of Lemma 2. First notice that $\sup_t T^{0.5} \left| \frac{1}{t} \sum_{j=1}^{t-1} X_{lj} \right| \leq \frac{T}{R} \sup_t \left| T^{-0.5} \sum_{j=1}^{t-1} X_{lj} \right|$

As T/R is bounded, it is sufficient to show that $\sup_t \left| T^{-0.5} \sum_{j=1}^{t-1} X_{lj} \right|$ is $o_p(1)$. Then, by the FCLT it follows that $T^{-0.5} \sum_{j=1}^{t-1} X_{lj} \rightarrow \sqrt{V_{LP}(X_l)} W(s)$.

Then, as noticed by CM, from Lemma 2.1 in Corradi, Swanson and Olivetti (2001) [43] it follows that

$$\sup_t \left| T^{-0.5} \sum_{j=1}^{t-1} X_{lj} \right| \rightarrow \sup_{\lambda \leq \frac{t}{T} \leq 1} \left| \sqrt{V_{LP}(X_l)} W(s) \right| = O_p(1)$$

The proof is complete. □

Appendix D

Proof of Lemma 3. With minor technical subtleties, this proof is similar to Lemma A2 in CM. The first step is to show that.

$$\begin{aligned} & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X'_{1j} \right\}^{-1} X_{1t} X'_{1t} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X'_{1j} \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ & = \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_1 X'_1\}^{-1} X_{1t} X'_{1t} \{EX_1 X'_1\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \end{aligned}$$

First, we add and subtract twice $E(X_{1t}X'_{1t})^{-1}$

$$\begin{aligned} & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} X_{1t}X'_{1t} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} = \\ & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \{EX_1X'_1\}^{-1} X_{1t}X'_{1t} \{EX_1X'_1\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ & \quad + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] X_{1t}X'_{1t} \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} \right. \\ & \quad \left. - \{EX_1X'_1\}^{-1} \right] \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ & \quad + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \{EX_1X'_1\}^{-1} X_{1t}X'_{1t} \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ & \quad + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] X_{1t}X'_{1t} \{EX_1X'_1\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \end{aligned}$$

Then, we only need to show that the following term is $o_p(1)$

$$\begin{aligned} & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] X_{1t}X'_{1t} \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ & \quad + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \{EX_1X'_1\}^{-1} X_{1t}X'_{1t} \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ & \quad + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] X_{1t}X'_{1t} \{EX_1X'_1\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \end{aligned}$$

Moreover, each of the three terms are $o_p(1)$. The proof for each term follows the same argument; thus, we show its exclusively for the first term. Taking absolute value

$$\begin{aligned} & \left| \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] X_{1t}X'_{1t} \left[\left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1} \right] \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \right| \\ & \leq k^4 \frac{p}{T} \left(\frac{1}{p} \sum_{t=R}^{T-1} |X_{1t}X'_{1t}| \right) (\sup_t |T^{0.25} \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1}|) (\sup_t |T^{0.25} \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}|) (\sup_t |T^{0.25} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} \\ & \quad - \{EX_1X'_1\}^{-1}|) (\sup_t |T^{0.25} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1}|) \end{aligned}$$

Given that X_{1t} is assumed to be stationary, $\frac{1}{p} \sum_{t=R}^{T-1} |X_{1t}X'_{1t}|$ is bounded. That $\sup_t |T^{0.25} \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1}|$, $\sup_t |T^{0.25} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right\}^{-1} - \{EX_1X'_1\}^{-1}|$ are $o_p(1)$ follow directly from Lemmas A1 (a) and (b) in CM. Finally, from Lemma 4, $\sup_t |T^{0.5} \frac{1}{t} \sum_{j=1}^{t-1} X_j| = O_p(1)$, then $\sup_t |T^{0.25} \frac{1}{t} \sum_{j=1}^{t-1} X_j| = o_p(1)$ and the first part of the proof is complete.

Now we show that

$$\begin{aligned} & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \{EX_1X'_1\}^{-1} X_{1t}X'_{1t} \{EX_1X'_1\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ & = \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{j+1} \right\}' \{EX_1X'_1\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \end{aligned}$$

First, we add and subtract $EX_{1t}X'_{1t}$

$$\begin{aligned} & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_1 X_1'\}^{-1} X_{1t} X_{1t}' \{EX_1 X_1'\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ &= \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_1 X_1'\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_1 X_1'\}^{-1} (X_{1t} X_{1t}' \\ & \quad - EX_{1t} X_{1t}') \{EX_1 X_1'\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \end{aligned}$$

Then, we show that

$$\sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_1 X_1'\}^{-1} (X_{1t} X_{1t}' - EX_{1t} X_{1t}') \{EX_1 X_1'\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \text{ is } o_p(1):$$

$$\begin{aligned} & \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_1 X_1'\}^{-1} (X_{1t} X_{1t}' - EX_{1t} X_{1t}') \{EX_1 X_1'\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \\ &= T^{-0.5} \sum_{t=R}^{T-1} \left\{ \left(\frac{T}{t} \right)^2 \left[T^{-0.5} \sum_{j=1}^{t-1} X_{1j}' \{EX_1 X_1'\}^{-1} \right. \right. \\ & \quad \left. \left. \otimes \sum_{j=1}^{t-1} X_{1j}' u_{j+1} \{EX_1 X_1'\}^{-1} \text{vec} [T^{-0.5} (X_{1t} X_{1t}' - EX_{1t} X_{1t}')] \right] \right\} \end{aligned}$$

As $\sum_{t=R}^{T-1} \left\{ \left(\frac{T}{t} \right)^2 [T^{-0.5} \sum_{j=1}^{t-1} X_{1j}' \{EX_1 X_1'\}^{-1} \otimes \sum_{j=1}^{t-1} X_{1j}' u_{j+1} \{EX_1 X_1'\}^{-1} \text{vec} [T^{-0.5} (X_{1t} X_{1t}' - EX_{1t} X_{1t}')]] \right\}$ is bounded and $T^{-0.5}$ is $o_p(1)$, the proof is complete. \square

Appendix E

Proof of Lemma 4. While this result is similar to Lemma A3 in [33], it is not immediate from their paper. Notice that $\sum_{t=R}^{T-1} X_{1t}' (\hat{\beta}_{1t} - \beta_1^*) [X_{2t}' (\hat{\beta}_{2t} - \beta_2^*) - X_{1t}' (\hat{\beta}_{1t} - \beta_1^*)]$ is simply

$$-\sum_{t=R}^{T-1} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} (u_{j+1} + \Delta_{IS}) \right\} \left(X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} (u_{j+1} + \Delta_{IS}) \right\} - X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} (u_{j+1} + \Delta_{IS}) \right\} \right)$$

Then, we separate into eight terms

$$-\sum_{t=R}^{T-1} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \left(X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} - X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \right) \tag{A1}$$

$$-\sum_{t=R}^{T-1} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \left(X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} - \Delta_{IS} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \right) \tag{A2}$$

$$-\sum_{t=R}^{T-1} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \left(\Delta_{IS} X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} - X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \right) \tag{A3}$$

$$-\sum_{t=R}^{T-1} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \left(\Delta_{IS} X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} - \Delta_{IS} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \right) \tag{A4}$$

$$-\Delta_{IS} \sum_{t=R}^{T-1} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \left(X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} - X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \right) \tag{A5}$$

$$-\Delta_{IS} \sum_{t=R}^{T-1} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \left(X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} - \Delta_{IS} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \right) \tag{A6}$$

$$-\Delta_{IS} \sum_{t=R}^{T-1} X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \left(\Delta_{IS} X_{2t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X_{2j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} - X_{1t}' \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X_{1j}' \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} \right) \tag{A7}$$

$$-\Delta_{IS}^2 \sum_{t=R}^{T-1} X'_{1t} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X'_{1j} \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \left(X_{2t} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} X'_{2j} \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} - X'_{1t} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} X'_{1j} \right\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} \right) \tag{A8}$$

In Appendix F, we show that (A1), (A2+A3), (A4), (A5), (A6+A7), and (A8) are $o_p(1)$ and the proof is complete. \square

Appendix F

Proof that (A1), (A2+A3), (A4), (A5), (A6+A7), and (A8) in Appendix E are $o_p(1)$.

Proof that (A1) is $o_p(1)$. This is immediate from Lemma A3 in [33]. Using Lemma 4, notice that (A1) equals to

$$\begin{aligned} & - \sum_{t=R}^{T-1} X'_{1t} \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} X'_{2t} \{EX_{2t} X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} \\ & \quad + \sum_{t=R}^{T-1} X'_{1t} \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} X'_{1t} \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1) \\ & = - \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} X_{1t} X'_{2t} \{EX_{2t} X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} \\ & \quad + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} X_{1t} X'_{1t} \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1) \\ & = - \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{2t} \{EX_{2t} X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} \\ & \quad + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{1t} \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1) \end{aligned}$$

Notice that $EX_{1t} X'_{2t} = JEX_{2t} X'_{2t}$

$$= - \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} J \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1)$$

As $X_{1t} = JX_{2t}$, the proof is complete. \square

Proof that (A2) and (A3) are $o_p(1)$. Using Lemma 4, rewrite (A2) as

$$\begin{aligned} & - \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{2t} \{EX_{2t} X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} \\ & \quad + \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \\ & = - \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} J \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} + \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \end{aligned}$$

Again, from Lemma 4, we write (A3) as

$$\begin{aligned} & - \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{2t} \{EX_{2t} X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} \\ & \quad + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{1t} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1) \end{aligned}$$

$$= -\Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} J \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} + \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1)$$

As $X_{1t} = JX_{2t}$, and adding (A2) and (A3), the proof is complete. \square

Proof that (A4) is $o_p(1)$. Notice from Lemma 4 that (A4) is equal to

$$\begin{aligned} & -\Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{2t} \{EX_{2t} X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} \\ & \quad + \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{1t} \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \\ = & -\Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} J \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} + \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\}' \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \end{aligned}$$

As $X_{1t} = JX_{2t}$, (A4) is $o_p(1)$ and the proof is complete. \square

Proof that (A5) is $o_p(1)$. From (A5), using Lemma 4,

$$\begin{aligned} & -\Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{2t} \{EX_{2t} X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} \\ & \quad + \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{1t} \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1) \\ = & -\Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} J \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} + \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1) \end{aligned}$$

As $X_{1t} = JX_{2t}$, it follows that (A5) is $o_p(1)$ and the proof is complete. \square

Proof that (A6) and (A7) are $o_p(1)$. From Lemma 4, notice that (A6) is equal to

$$\begin{aligned} & -\Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{2t} \{EX_{2t} X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} \\ & \quad + \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} EX_{1t} X'_{1t} \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \\ = & -\Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} J \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} u_{j+1} \right\} + \Delta^2 \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \end{aligned}$$

Now, based on Lemma 4, write (A7) as

$$= -\Delta^2 \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} J \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} + \Delta \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t} X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} u_{j+1} \right\} + o_p(1)$$

As $X_{1t} = JX_{2t}$, we add (A6) and (A7), and the proof is complete. \square

Proof that (A8) is $o_p(1)$. Notice from Lemma 4 that (A8) is equal to

$$\begin{aligned}
 & -\Delta^2 \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t}X'_{1t}\}^{-1} EX_{1t}X'_{2t} \{EX_{2t}X'_{2t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} \\
 & \quad + \Delta^2 \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t}X'_{1t}\}^{-1} EX_{1t}X'_{1t} \{EX_{1t}X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1) \\
 & -\Delta^2 \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t}X'_{1t}\}^{-1} J \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{2j} \right\} + \Delta^2 \sum_{t=R}^{T-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\}' \{EX_{1t}X'_{1t}\}^{-1} \left\{ \frac{1}{t} \sum_{j=1}^{t-1} X_{1j} \right\} + o_p(1)
 \end{aligned}$$

As $X_{1t} = JX_{2t}$, the proof is complete. \square

Appendix G

Proof of Lemma 5. From (5) $\sum_{t=R}^{T-1} \hat{u}_{1t+1}(\hat{u}_{1t+1} - \hat{u}_{2t+1})$

$$\begin{aligned}
 & = \sum_{t=R}^{T-1} u_{t+1} [X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)] \\
 & \quad + \sum_{t=R}^{T-1} \Delta_{00s} [X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)] \\
 & \quad \quad \quad + o_p(1)
 \end{aligned}$$

First, notice that $X_{1j}u_{j+1} = JX_{2j}u_{j+1}$. Then, use the fact that, under the null hypothesis $u_{2j+1} = u_{1j+1} = u_{t+1} + \Delta_{00s}$. Finally, combining these results with Lemma A10 in CM it follows that

$$\begin{aligned}
 & \sum_{t=R}^{T-1} u_{t+1} [X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)] \\
 & \quad = \sum_{t=R}^{T-1} u_{t+1} X'_{2t} \left(\frac{1}{t} \sum_{j=1}^{t-1} X_{2j}X'_{2j} \right)^{-1} \left(\frac{1}{t} \sum_{j=1}^{t-1} X_{2j}u_{2j+1} \right) \\
 & \quad \quad - J \sum_{t=R}^{T-1} u_{t+1} X'_{2t} \left(\frac{1}{t} \sum_{j=1}^{t-1} X_{1j}X'_{1j} \right)^{-1} \left(\frac{1}{t} \sum_{j=1}^{t-1} X_{1j}u_{1j+1} \right) \\
 & = \sum_{t=R}^{T-1} u_{t+1} X'_{2t} [(EX_2X'_2)^{-1} - J(EX_1X'_1)^{-1}J'] \sum_{j=1}^{t-1} (u_{t+1} + \Delta_{IS})X_{2j} + o_p(1)
 \end{aligned}$$

Then,

$$\begin{aligned}
 & \sum_{t=R}^{T-1} u_{t+1} X'_{2t} [(EX_2X'_2)^{-1} - J(EX_1X'_1)^{-1}J'] \sum_{j=1}^{t-1} (u_{j+1} + \Delta_{IS})X_{2j} \\
 & \quad = \sum_{t=R}^{T-1} u_{t+1} X'_{2t} [(EX_2X'_2)^{-1} - J(EX_1X'_1)^{-1}J'] \sum_{j=1}^{t-1} X_{2j}u_{t+1} \\
 & \quad \quad + \Delta_{IS} \sum_{t=R}^{T-1} u_{t+1} X'_{2t} [(EX_2X'_2)^{-1} - J(EX_1X'_1)^{-1}J'] \sum_{j=1}^{t-1} X_{2j}
 \end{aligned}$$

Second, following similar arguments

$$\sum_{t=R}^{T-1} \Delta_{\text{oos}} [X'_{2t}(\hat{\beta}_{2t} - \beta_2^*) - X'_{1t}(\hat{\beta}_{1t} - \beta_1^*)] = \Delta_{\text{oos}} \sum_{t=R}^{T-1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} (u_{t+1} + \Delta_{\text{IS}}) X_{2j} + o_p(1)$$

Then,

$$\begin{aligned} & \Delta_{\text{oos}} \sum_{t=R}^{T-1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} (u_{t+1} + \Delta_{\text{IS}}) X_{2j} \\ &= \Delta_{\text{oos}} \sum_{t=R}^{T-1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} X_{2j} u_{t+1} \\ &+ \Delta_{\text{oos}} \Delta_{\text{IS}} \sum_{t=R}^{T-1} X'_{2t} [(EX_2 X'_2)^{-1} - J(EX_1 X'_1)^{-1} J'] \sum_{j=1}^{t-1} X_{2j} \end{aligned}$$

The proof is complete. \square

References

- Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Statics* **1995**, *13*, 253–263.
- West, K.D. Asymptotic Inference about Predictive Ability. *Econometrica* **1996**, *64*, 1067. [[CrossRef](#)]
- Clark, T.E.; West, K.D. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *J. Econom.* **2006**, *135*, 155–186. [[CrossRef](#)]
- West, K.D. Chapter 3 Forecast Evaluation. *Handb. Econ. Forecast.* **2006**, *1*, 99–134.
- Clark, T.E.; McCracken, M.W. Tests of equal forecast accuracy and encompassing for nested models. *J. Econom.* **2001**, *105*, 85–110. [[CrossRef](#)]
- Meese, R.A.; Rogoff, K. Empirical exchange rate models of the seventies: Do they fit out of sample? *J. Int. Econ.* **1983**, *14*, 3–24. [[CrossRef](#)]
- Meese, R.; Rogoff, K. Was it real? The exchange rate-interest differential relation over the modern floating-rate period. *J. Financ.* **1988**, *43*, 933–948. [[CrossRef](#)]
- Goyal, A.; Welch, I. Predicting the equity premium with dividend ratios. *Manag. Sci.* **2003**, *49*, 639–654. [[CrossRef](#)]
- Welch, I.; Goyal, A. A comprehensive look at the empirical performance of equity premium prediction. *Rev. Financ. Stud.* **2008**, *21*, 1455–1508. [[CrossRef](#)]
- Clark, T.E.; West, K.D. Approximately normal tests for equal predictive accuracy in nested models. *J. Econom.* **2007**, *138*, 291–311. [[CrossRef](#)]
- Clark, T.E.; McCracken, M.W. Evaluating direct multistep forecasts. *Econom. Rev.* **2005**, *24*, 369–404. [[CrossRef](#)]
- McCracken, M.W. Asymptotics for out of sample tests of Granger causality. *J. Econom.* **2007**, *140*, 719–752. [[CrossRef](#)]
- Harvey, D.S.; Leybourne, S.J.; Newbold, P. Tests for forecast encompassing. *J. Bus. Econ. Stat.* **1998**, *16*, 254–259.
- Clark, T.; McCracken, M. Evaluating the accuracy of forecasts from vector autoregressions. In *Vector Autoregressive Modeling—New Developments and Applications: Essays in Honor of Christopher A. Sims*; Fomby, T., Kilian, L., Murphy, A., Eds.; Emerald Group Publishing: Bingley, UK, 2013.
- Chao, J.; Corradi, V.; Swanson, N. Out-of-sample tests for Granger causality. *Macroecon. Dyn.* **2001**, *5*, 598–620. [[CrossRef](#)]
- Armah, N.; Swanson, N.R. Predictive inference under model misspecification. In *Forecasting in the Presence of Structural Breaks and Model Uncertainty*; Emerald Group Publishing Limited: Bingley, UK, 2008.
- Corradi, V.; Swanson, N. Some recent developments in predictive accuracy testing with nested models and (generic) nonlinear alternatives. *Int. J. Forecast.* **2004**, *20*, 185–199. [[CrossRef](#)]
- Corradi, V.; Swanson, N. Bootstrap conditional distribution tests in the presence of dynamic misspecification. *J. Econom.* **2006**, *133*, 779–806. [[CrossRef](#)]
- Corradi, V.; Swanson, N. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. *Int. Econ. Rev.* **2007**, *48*, 67–109. [[CrossRef](#)]
- Box, G. Science and Statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799. [[CrossRef](#)]
- Inoue, A.; Killian, L. In-sample or out-of-sample tests of predictability: Which one should we use? *Econom. Rev.* **2002**, *23*, 371–402. [[CrossRef](#)]
- Pincheira, P.; Hardy, N. Forecasting base metal prices with the Chilean exchange rate. *Resour. Policy* **2019**, *62*, 256–281. [[CrossRef](#)]
- Pincheira, P.; Hardy, N. Forecasting aluminum prices with commodity currencies. *Resour. Policy* **2021**, *73*, 102066. [[CrossRef](#)]
- Pincheira, P.; Bentancor, A.; Hardy, N.; Jarsun, N. Forecasting fuel prices with the Chilean exchange rate: Going beyond the commodity currency hypothesis. *Energy Econ.* **2021**, 105802. [[CrossRef](#)]

25. Pincheira, P.; Hardy, N.; Henriquez, C.; Tapia, I.; Bentancor, A. Forecasting Base Metal Prices with an International Stock Index. 2021. Available online: <https://ssrn.com/abstract=3849161> (accessed on 1 September 2021).
26. Pincheira, P.; Hardy, N.; Muñoz, F. “Go Wild for a While!”: A New Test for Forecast Evaluation in Nested Models. *Mathematics* **2021**, *9*, 2254. [[CrossRef](#)]
27. Diebold, F.X.; Lopez, J.A. Forecast evaluation and combination. *Handb. Stat.* **1996**, *14*, 241–268.
28. Clark, T.; McCracken, M. Advances in forecast evaluation. *Handb. Econ. Forecast.* **2013**, *2*, 1107–1201.
29. Wilson, E.B. The periodogram of American business activity. *Q. J. Econ.* **1934**, *48*, 375–417. [[CrossRef](#)]
30. Fair, R.C.; Shiller, R.J. The informational content of ex ante forecasts. *Rev. Econ. Stat.* **1989**, *71*, 325–331. [[CrossRef](#)]
31. Fair, R.C.; Shiller, R.J. Comparing information in forecasts from econometric models. *Am. Econ. Rev.* **1990**, *80*, 375–389.
32. Rossi, B. Exchange rate predictability. *J. Econ. Lit.* **2013**, *51*, 1063–1119. [[CrossRef](#)]
33. Clark, T.; McCracken, M. Not-for-Publication Appendix to “Tests of Equal Forecast Accuracy and Encompassing for Nested Models”. 2000. Available online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.161.4290> (accessed on 1 January 2021).
34. White, H. *Asymptotic Theory for Econometricians*; Academic Press: Cambridge, MA, USA, 2014.
35. Newey, W.; West, K. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **1987**, *703*–708. Available online: https://www.nber.org/system/files/working_papers/t0055/t0055.pdf (accessed on 1 November 2021). [[CrossRef](#)]
36. Newey, W.; West, K. Automatic lag selection in covariance matrix estimation. *Rev. Econ. Stud.* **1994**, *61*, 631–653. [[CrossRef](#)]
37. Campbell, J.Y.; Shiller, R.J. Cointegration and Tests of Present Value Models. *J. Political Econ.* **1987**, *95*, 1062–1088. [[CrossRef](#)]
38. Engel, C.; West, K.D. Exchange Rates and Fundamentals. *J. Political Econ.* **2005**, *113*, 485–517. [[CrossRef](#)]
39. Davidson, J. *Stochastic Limit Theory: An Introduction for Econometricians*; OUP Oxford: Oxford, UK, 1994.
40. Hansen, B. Convergence to stochastic integrals for dependent heterogeneous processes. *Econom. Theory* **1992**, *8*, 489–500. [[CrossRef](#)]
41. McLeish, D. Dependent central limit theorems and invariance principles. *Ann. Probab.* **1974**, *2*, 620–628. [[CrossRef](#)]
42. Hall, P. Martingale invariance principles. *Ann. Probab.* **1977**, *5*, 875–887. [[CrossRef](#)]
43. Corradi, V.; Swanson, N.; Olivetti, C. Predictive ability with cointegrated variables. *J. Econom.* **2001**, *104*, 315–358. [[CrossRef](#)]