

Article

Multistep Forecast Averaging with Stochastic and Deterministic Trends

Mohitosh Kejriwal ^{1,*} , Linh Nguyen ¹  and Xuewen Yu ²

¹ Daniels School of Business, Purdue University, 403 Mitch Daniels Blvd., West Lafayette, IN 47907, USA; nguye535@purdue.edu

² Department of Applied Economics, School of Management, Fudan University, 670 Guoshun Road, Shanghai 200433, China; xuewenyu@fudan.edu.cn

* Correspondence: mkejriwa@purdue.edu

Abstract: This paper presents a new approach to constructing multistep combination forecasts in a nonstationary framework with stochastic and deterministic trends. Existing forecast combination approaches in the stationary setup typically target the in-sample asymptotic mean squared error (AMSE), relying on its approximate equivalence with the asymptotic forecast risk (AFR). Such equivalence, however, breaks down in a nonstationary setup. This paper develops combination forecasts based on minimizing an accumulated prediction errors (APE) criterion that directly targets the AFR and remains valid whether the time series is stationary or not. We show that the performance of APE-weighted forecasts is close to that of the optimal, infeasible combination forecasts. Simulation experiments are used to demonstrate the finite sample efficacy of the proposed procedure relative to Mallows/Cross-Validation weighting that target the AMSE as well as underscore the importance of accounting for both persistence and lag order uncertainty. An application to forecasting US macroeconomic time series confirms the simulation findings and illustrates the benefits of employing the APE criterion for real as well as nominal variables at both short and long horizons. A practical implication of our analysis is that the degree of persistence can play an important role in the choice of combination weights.

Keywords: averaging; combination; cross-validation; Mallows; unit root; accumulated prediction errors



Citation: Kejriwal, Mohitosh, Linh Nguyen, and Xuewen Yu. 2023. Multistep Forecast Averaging with Stochastic and Deterministic Trends. *Econometrics* 11: 28. <https://doi.org/10.3390/econometrics11040028>

Received: 24 August 2023
Revised: 27 November 2023
Accepted: 7 December 2023
Published: 15 December 2023



Copyright: © 2023 by the authors. 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/>).

JEL Classification: C22; C53

1. Introduction

The pioneering work of Granger (1966) demonstrated that a large number of macroeconomic time series have a typical spectral shape dominated by a peak at low frequencies. This finding suggests the presence of relatively long run information in the current level of the variables, which should be taken into account when modeling their time series evolution and can potentially be exploited to yield improved forecasts. One way to incorporate this long-run information in econometric modeling is through stochastic trends (unit roots) and/or deterministic trends. However, given that trends are slowly evolving, there is only limited information in any data set about how best to specify the trend or distinguish between alternative models of the trend. For instance, unit root tests often fail to reject a unit root despite the fact that theory does not postulate the presence of a unit root for many macroeconomic variables [see Elliott (2006), for further discussion of this issue]. Therefore, it appears prudent to incorporate the uncertainty arising from the presence of a stochastic trend when constructing macroeconomic forecasts. Moreover, this uncertainty is likely to be particularly important for longer horizons.¹

A second source of uncertainty involved in the construction of forecasts relates to the specification of short-run dynamics driving the time series. Within an autoregressive

modeling framework, this form of uncertainty can be expressed in terms of the lags of first differences of the time series being analyzed. Since the number of lags that ought to be included in the model is unknown in practice, there is a bias–variance trade-off facing the forecaster: underspecifying the number of lags would lead to biased forecasts, while including irrelevant lags would induce a higher forecast variance. The challenge therefore lies in incorporating lag order uncertainty in a manner that best addresses this trade-off.

Motivated by these considerations, this paper proposes a new multistep forecast combination approach designed for forecasting a highly persistent time series that simultaneously addresses uncertainty about the presence of a stochastic trend and uncertainty about the nature of short-run dynamics within a unified autoregressive modeling framework. Unlike extant forecast combination approaches, we develop combination forecasts based on minimizing the so-called accumulated prediction errors (APE) criterion that directly targets the asymptotic forecast risk (AFR) instead of the in-sample asymptotic mean squared error (AMSE). This is particularly relevant, since the equivalence between AFR and AMSE breaks down in a nonstationary setup. Our analysis generalizes existing results by establishing the asymptotic validity of the APE for multistep forecasts in the unit root and (fixed) stationary cases, both for models with and without deterministic trends. We further show that, regardless of the presence of a unit root, the performance of APE-weighted forecasts remains close to that of the infeasible combination forecasts which assume that the optimal (i.e., AFR minimizing) weights are known. Monte Carlo experiments are used to (i) demonstrate the finite sample efficacy of the proposed procedure relative to Mallows/Cross-Validation weighting that target the AMSE; (ii) underscore the importance of accounting for uncertainty about the stochastic trend and/or the lag order. In a pseudo out-of-sample forecasting exercise applied to US monthly macroeconomic time series, we evaluate the performance of a variety of selection/combination-based approaches at horizons of one, three, six, and twelve months. Consistent with the simulation results, the empirical analysis provides strong evidence in favor of a version of the advocated approach that simultaneously addresses stochastic trend and lag order uncertainty regardless of the forecast horizon considered.

The present study builds on previous work by Hansen (2010a) and Kejriwal and Yu (2021) who analyzed one-step ahead combination forecasts allowing for both persistence and lag order uncertainty. In particular, Hansen (2010a) adopted a local-to-unity framework to develop combination forecasts that combine forecasts from the restricted (i.e., imposing a unit root) and unrestricted models with the weights obtained by minimizing a one-step Mallows criterion. To address lag order uncertainty, he also proposed a general combination approach that, in addition to the restricted and unrestricted model forecasts, also combines forecasts based on different lag orders. Kejriwal and Yu (2021) provided theoretical justification for the general combination approach and developed improved combination forecasts that employ feasible generalized least squares (FGLS) estimates instead of ordinary least squares (OLS) estimates of the deterministic trend component.

Our paper can be viewed as extending Hansen's (2010a) approach in two practically relevant directions. First, in addition to one-step ahead forecasts, we also analyze the statistical properties of multistep combination forecasts given that uncertainty regarding the presence of a stochastic trend is especially relevant over longer horizons. Second, in contrast to Mallows weighting as advocated by Hansen (2010a), our combination weights are obtained via the APE criterion that directly targets the AFR instead of the AMSE. Our Monte Carlo and empirical comparisons of the performance of combination forecasts based on different weighting schemes clearly illustrate the importance of directly targeting the AFR. Thus, an important implication of our study is that the preferred choice of weighting scheme when combining forecasts can critically depend on whether the variables involved are stationary or not.

The recent machine learning literature has proposed a variety of forecasting methods that exploit information in a large number of potential predictors (see, e.g., Masini et al. 2023, for a survey). In contrast, our study is univariate in that it only utilizes past information

about the variable of interest to develop forecasts. A natural question one may ask, then, is what is the value added by our univariate forecasting approach when more sophisticated machine learning approaches are available? We offer three possible responses. First, our approach is simple to use in practice, since it only requires running OLS regressions. Second, it is transparent in that its statistical properties can be studied analytically, which can be useful for understanding the merits and limitations of the approach. Third, when evaluating the performance of machine learning methods, our preferred forecasting approach can provide a much more competitive univariate benchmark for comparison than a simple autoregressive model with a prespecified/estimated lag order which has routinely been used as the benchmark (see, e.g., [Kim and Swanson 2018](#); [Medeiros et al. 2021](#)).

The rest of the paper is organized as follows. Section 2 provides a review of the related literature. Section 3 presents the model and the related estimators. Section 4 analyzes the AMSE and AFR as alternative measures of forecast accuracy. Section 5 discusses the choice of combination weights based on the APE criterion. Section 6 extends the analysis to allow for lag order uncertainty in the construction of the forecasts. Monte Carlo evidence, including comparisons with various existing methods, is provided in Section 7. Section 8 details an empirical application to forecasting US macroeconomic time series and Section 9 concludes. Appendices A–C contain, respectively, the proofs, details of forecasting methods considered, and additional simulation results. All computations were carried out in the 2022b version of [MATLAB \(2022\)](#).

2. Literature Review

A common practice in the economic forecasting literature is to apply a stationarity-inducing transformation (e.g., differencing or detrending) to the time series of interest and then attempt to forecast the transformed series. Consequently, most of the forecasting procedures in current use have been developed under the assumption of data stationarity. The traditional approach of [Box and Jenkins \(1970\)](#) transforms the data through differencing which amounts to modeling the low-frequency peak in the spectrum as a zero-frequency phenomenon and proceeds to forecast the transformed series using standard stationary autoregressive moving average (ARMA) models. More recently, [Stock and Watson \(2005, 2006\)](#) constructed a extensive database of 132 monthly macroeconomic time series over the period 1959–2003 and applied a variety of transformations to render them stationary before using a handful of common factors extracted from the data set using principal components as predictors (the so-called diffusion-index methodology). Similarly, [McCracken and Ng \(2016\)](#) assembled a publicly available database of 134 monthly time series referred to as FRED-MD and updated on a timely basis by the Federal Reserve Bank of St Louis. They also suggested a set of data transformations which is used to construct factor-based diffusion indexes for forecasting as well as to analyze business cycle turning points.

While convenient in practice, the approach of forecasting the transformed stationary series tends to ignore the information in the levels of the variables. In particular, it does not properly account for the uncertainty arising from the nature of the underlying trends, which can lead to poor forecasts if the trends are misspecified. [Clements and Hendry \(2001\)](#) documented, both analytically and numerically, the detrimental consequences of trend misspecification on the resulting forecasts in the presence of parameter estimation uncertainty. Specifically, they found that when the sample size increases at a faster rate than the forecast horizon, misspecifying a difference stationary process as trend stationary or vice versa yields forecast error variances of a higher order of magnitude relative to the correctly specified model. Consequently, the objective of our study is to construct forecasts of the time series in levels and explicitly model the uncertainty regarding the presence of a stochastic trend instead of transforming the time series to stationarity based on a trend specification that is determined a priori and is possibly misspecified.

Our study is closely related to the existing literature on methods for forecasting nonstationary time series. [Diebold and Kilian \(2000\)](#) showed that a unit root pretesting strategy can improve forecast accuracy relative to restricted or unrestricted estimation. [Ng](#)

and Vogelsang (2002) found that the use of FGLS estimates of the trend component can yield superior forecasts relative to their OLS counterparts. Turner (2004) recommended the use of forecasting thresholds whereby the restricted (unit root) forecast is preferred on one side of these thresholds while the unrestricted (OLS) forecast is preferred on the other. His proposal was based on median unbiased estimation of the local-to-unity parameter to determine the thresholds and was shown to dominate a unit root pretesting strategy. Ing et al. (2012) studied the impact of nonstationarity, model complexity and model misspecification on the AFR in infinite order autoregressions.

A promising approach to addressing both stochastic trend uncertainty and lag order uncertainty is forecast combination. Introduced in the seminal work of Bates and Granger (1969), the idea underlying forecast combination is to exploit the bias–variance trade-off by combining forecasts from restricted (possibly subject to bias) specifications and unrestricted (possibly subject to overfitting) specifications using an appropriate choice of combination weights. A voluminous amount of literature has subsequently developed, which has analyzed the efficacy of several alternative weighting schemes for constructing the combination forecasts (see, e.g., Wang et al. 2022, for a recent survey). Hansen (2010a) proposed one-step ahead combination forecasts within an autoregressive modeling framework that accounts for both aforementioned sources of uncertainty, where the combination weights are obtained by minimizing a Mallows criterion. The Mallows criterion is designed to provide an approximately unbiased estimator of the in-sample AMSE. Hansen’s analysis showed that the unit root pretesting strategy could be subject to high forecast risk for a range of persistence levels, while his combination forecast performed favorably compared to a number of methods popular in applied work and dominated the unrestricted forecast uniformly in terms of finite sample forecast risk. Kejriwal and Yu (2021) proposed a refinement of Hansen’s (2010a) approach, which entails estimating the deterministic trend component by FGLS instead of OLS. Tu and Yi 2017 analyzed one-step forecasting based on the Mallows averaging estimator in a cointegrated vector autoregressive model and found that it dominated the commonly used approach of pretesting for cointegration.

In a stationary setup, combination forecasts based on Mallows/cross-validation (CV) weighting typically target the AMSE, relying on its approximate equivalence with the AFR (e.g., Hansen 2008, 2010b; Liao and Tsay 2020). Such equivalence, however, breaks down in a nonstationary setup. Hansen (2010a) showed, within a local-to-unity framework, that the AMSEs of unrestricted as well as restricted (imposing a unit root) one-step ahead forecasts are different from the corresponding expressions for their AFR in autoregressive models (see Section 4 for further discussion on the issue of equivalence or lack thereof).

To address the lack of equivalence between AMSE and AFR, we develop combination forecasts based on minimizing the APE criterion that directly targets the AFR instead of the AMSE. Previous work in the context of model selection has shown the APE criterion to remain valid whether the process is stationary or has a unit root. Specifically, Ing (2004) showed that a normalized version of the APE is almost certain to the AFR in the stationary case, while a similar result was obtained by Ing et al. (2009) in the unit root case. Focusing on the first-order autoregressive case and one-step ahead forecasts, Yu et al. (2012) extended the validity of the APE to a unit root model with a deterministic time trend. Our study extends the use of the APE criterion to construct combination forecasts in a nonstationary environment.

In summary, there is a plethora of approaches available in the literature for forecasting nonstationary time series, including model selection, pretesting, and forecast combination. Combination forecasts have often been shown to incur lower forecast risk in practice than forecasts based on model selection or pretesting. However, the existing literature has typically employed weighting schemes such as Mallows/CV weighting that have been formally justified only in a stationary framework. Our study contributes to this literature by demonstrating that when the variable of interest is potentially nonstationary, it may be desirable to construct the combination weights using an alternative approach (the APE criterion). Our findings are particularly relevant for macroeconomic applications, given that

several macroeconomic time series have been documented to exhibit a degree of persistence that is difficult to distinguish from a unit root process.

3. Model and Estimation

We consider a univariate time series y_t generated as follows:

$$\begin{aligned}
 y_t &= m_t + u_t \\
 m_t &= \beta_0 + \beta_1 t + \dots + \beta_p t^p \\
 u_t &= \alpha u_{t-1} + \alpha_1 \Delta u_{t-1} + \dots + \alpha_k \Delta u_{t-k} + e_t \\
 \alpha &= 1 + \frac{ac}{T}, \quad a = 1 - \alpha_1 - \dots - \alpha_k, \quad c \leq 0
 \end{aligned} \tag{1}$$

where $p \in \{0, 1\}$ is the order of the trend component and the stochastic component u_t follows a finite order autoregressive process of order $(k + 1)$ process driven by the innovations e_t . The uncertainty about the stochastic trend is captured by the persistence parameter α that is modeled as local-to-unity with $c = 0$ corresponding to the unit root case and $c < 0$ to the stationary case. The initial observations are set at $u_0, u_{-1}, \dots, u_{-k} = O_p(1)$.² This section treats the true lag order k as known. Lag order uncertainty is addressed in Section 6. Our analysis is based on the following assumptions:

Assumption 1. *The sequence $\{e_t\}$ is a martingale difference sequence with $E(e_t | \mathcal{F}_{t-1}) = 0$ and $E(e_t^2 | \mathcal{F}_{t-1}) = \sigma^2$, where $0 < \sigma^2 < \infty$, and \mathcal{F}_t is the σ -field generated by $\{e_s; s \leq t\}$. Moreover, there exist small positive numbers ϕ_1 and ϕ_2 and a large positive number M_1 , such that for $0 \leq s - s' \leq \phi_2$,*

$$\sup_{1 \leq m \leq t < \infty, \|\mathbf{v}_m\|=1} |F_{t,m,\mathbf{v}_m}(s) - F_{t,m,\mathbf{v}_m}(s')| \leq M_1(s - s')^{\phi_1},$$

where $\mathbf{v}_m = (v_1, \dots, v_m)' \in \mathbb{R}^m$, $\|\mathbf{v}_m\| = \sum_{j=1}^m v_j^2$ and $F_{t,m,\mathbf{v}_m}(\cdot)$ denotes the distribution of $\sum_{l=1}^m v_l e_{t+1-l}$.

Assumption 2. *All roots of $A(L) = 1 - \sum_{i=1}^k \alpha_i L^i$ lie outside the unit circle.*

The data generating process in (1) and Assumptions 1 and 2 are adopted from Hansen (2010a) with an additional restriction on the distribution of $\{e_t\}$, which ensures that the sample second moments of the regressors are bounded in expectation (see Ing et al. 2009). The difference between our modeling framework and Ing et al. (2009) is that they impose an exact unit root ($c = 0$), while we allow $c \leq 0$. For $h \geq 1$, let the optimal (infeasible) mean squared error minimizing h -step ahead forecast of y_t be denoted as μ_{t+h} . It is the conditional mean of y_{t+h} given \mathcal{F}_t , which is obtained from the following recursion (Hamilton 1994, pp. 80–82):

$$\begin{aligned}
 \mu_{t+h} &= z'_{t+h} \beta + \alpha(\mu_{t+h-1} - z'_{t+h-1} \beta) + \alpha_1(\Delta \mu_{t+h-1} - \Delta z'_{t+h-1} \beta) \\
 &+ \dots + \alpha_k(\Delta \mu_{t+h-k} - \Delta z'_{t+h-k} \beta)
 \end{aligned} \tag{2}$$

with $\mu_{t+j} = y_{t+j}$ if $j \leq 0$; $\beta = \beta_0$ and $z_t = 1$ if $p = 0$; $\beta = (\beta_0, \beta_1)'$ and $z_t = (1, t)'$ if $p = 1$. We can further rewrite (2) as

$$\mu_{t+h} = z'_{t+h} \beta^* + \alpha \mu_{t+h-1} + \sum_{j=1}^k \alpha_j \Delta \mu_{t+h-j} \tag{3}$$

where $\beta^* = (1 - \alpha)\beta_0$ if $z_t = 1$ and $\beta^* = (\beta_0^*, \beta_1^*)'$ with $\beta_0^* = (1 - \alpha)\beta_0 + (\alpha - \sum_{j=1}^k \alpha_j)\beta_1$, $\beta_1^* = (1 - \alpha)\beta_1$ if $z_t = (1, t)$.

We consider three alternative estimators of μ_{t+h} . The first is the unrestricted estimator $\hat{\mu}_{t+h}$ obtained as

$$\hat{\mu}_{t+h} = z'_{t+h}\hat{\beta}^* + \hat{\alpha}\hat{\mu}_{t+h-1} + \sum_{j=1}^k \hat{\alpha}_j \Delta \hat{\mu}_{t+h-j} \quad (4)$$

with $\hat{\mu}_{t+j} = y_{t+j}$ if $j \leq 0$ where $(\hat{\beta}^*, \hat{\alpha}, \hat{\alpha}_j)$ are the OLS estimates from the regression

$$y_s = z'_s \beta^* + \alpha y_{s-1} + \sum_{j=1}^k \alpha_j \Delta y_{s-j} + e_s, \quad s = k+2, \dots, T$$

Instead of using (4), one may consider a two-step strategy for estimating μ_{t+h} that entails regressing y_t on z_t and obtaining the estimate $\hat{\beta}$ of β and the residuals $\hat{u}_t = y_t - z'_t \hat{\beta}$ in a first step and then estimating an autoregression of order $k+1$ in \hat{u}_t to obtain the estimates of $(\alpha, \alpha_1, \dots, \alpha_k)$. The forecasts are obtained from (4). However, as shown in Ng and Vogelsang (2002), the one-step estimate $\hat{\mu}_{t+h}$ is preferable to the two-step estimate with persistent data.

The second estimator is the restricted estimator $\tilde{\mu}_{t+h}$ that imposes the unit root restriction $\alpha = 1$ and is obtained as

$$\tilde{\mu}_{t+h} = \Delta z'_{t+h} \tilde{\beta}^* + \tilde{\mu}_{t+h-1} + \sum_{j=1}^k \tilde{\alpha}_j \Delta \tilde{\mu}_{t+h-j}$$

with $\tilde{\mu}_{t+j} = y_{t+j}$ if $j \leq 0$ where $(\tilde{\beta}^*, \tilde{\alpha}, \tilde{\alpha}_j)$ are the OLS estimates from the regression

$$\Delta y_s = \Delta z'_s \beta^* + \sum_{j=1}^k \alpha_j \Delta y_{s-j} + e_s, \quad s = k+2, \dots, T$$

Finally, the third estimator is based on taking a weighted average of the unrestricted and restricted forecasts. Letting $w \in [0, 1]$ be the weight assigned to the unrestricted estimator, the averaging estimator is given by

$$\hat{\mu}_{t+h}(w) = w\hat{\mu}_{t+h} + (1-w)\tilde{\mu}_{t+h}$$

The relative accuracy of the three foregoing estimators can be evaluated using the asymptotic forecast risk (AFR), which is the limit of the h -step ahead expected squared forecast error:

$$\begin{aligned} f_0(c, p, k, h) &= \lim_{T \rightarrow \infty} \frac{T}{\sigma^2} E(\tilde{\mu}_{T+h} - \mu_{T+h})^2 \\ f_1(c, p, k, h) &= \lim_{T \rightarrow \infty} \frac{T}{\sigma^2} E(\hat{\mu}_{T+h} - \mu_{T+h})^2 \\ f_w(c, p, k, h) &= \lim_{T \rightarrow \infty} \frac{T}{\sigma^2} E(\hat{\mu}_{T+h}(w) - \mu_{T+h})^2 \end{aligned}$$

In order to derive analytical expressions for the AFR, we introduce the following notation. Let $W(\cdot)$ denote a standard Brownian motion on $[0, 1]$ and define the Ornstein–Uhlenbeck process

$$dW_c(r) = cW_c(r) + dW(r)$$

For $p \in \{0, 1\}$, let $X_c(r) = (r^p, W_c(r))'$ and define the stochastic processes

$$W_c^*(r, p) = \begin{cases} W_c(r) & \text{if } p = 0 \\ W_c(r) - \int_0^1 W_c(s)ds & \text{if } p = 1 \end{cases}$$

$$X_c^*(r, p) = \begin{cases} X_c(r) & \text{if } p = 0 \\ X_c(r) - \int_0^1 X_c(s)ds & \text{if } p = 1 \end{cases}$$

and the functionals

$$T_{0c} = -cW_c^*(1, p) + I(p = 1)W(1)$$

$$T_{1c} = X_c^*(1, p)' \left(\int_0^1 X_c^*(r, p)X_c^*(r, p)' \right)^{-1} \int_0^1 X_c^*(r, p)dW(r) + I(p = 1)W(1)$$

Next, note that from (1), we can write

$$y_{t+h} = E_t(y_{t+h}) + \eta_{t,h}$$

where $\eta_{t,h} = \sum_{j=0}^{h-1} b_j e_{t+h-j}$, $E_t(\cdot)$ denotes conditional expectation with respect to information at time t and the coefficients b_j ($j = 0, \dots, h - 1$) are obtained by equating coefficients of L^j on both sides of the equation

$$b(L)d(L) = 1$$

where $b(L) = \sum_{j=0}^{h-1} b_j L^j$ and $d(L) = 1 - \alpha L - (1 - L) \sum_{j=1}^k \alpha_j L^j$. When $\alpha = 1$, $b_j = \sum_{i=0}^j v_i$, $v_0 = 1$ and $v_j; j \geq 1$, satisfies $1 + \sum_{j=1}^{\infty} v_j L^j = 1/A(L)$ (see Ing et al. 2009).

Denoting $\alpha(k) = (\alpha_1, \dots, \alpha_k)'$, we define the following quantities:

$$S_M(k) = \begin{pmatrix} \alpha(k-1) & I_{k-1} \\ \alpha_k & \mathbf{0}'_{k-1} \end{pmatrix}, S_M^0(k) = I_k$$

$$M_h(k) = \sum_{j=0}^{h-1} b_j S_M^{h-1-j}(k), \Gamma(k) = \lim_{j \rightarrow \infty} E(\mathbf{s}_j(k)\mathbf{s}_j(k)'), \mathbf{s}_j(k) = (\Delta y_j, \dots, \Delta y_{j-k+1})'$$

$$g_h(k) = \begin{cases} 0 & \text{if } k = 0 \\ tr(\Gamma(k)M_h(k)\Gamma^{-1}(k)M_h'(k)) & \text{if } k \geq 1 \end{cases}$$

With the above notation in place, we obtain the following result, which provides an analytical representation for the AFR of the unrestricted and restricted forecasts:

Theorem 1. Under Assumptions 1 and 2 and $\sup_t E(|e_t|^{\theta_h}) < \infty$, where $\theta_h = \max\{8, 2(h + 2)\} + \psi$ for some $\psi > 0$,

$$(a) f_1(c, p, k, h) = f_1(c, p, h) + g_h(k), f_1(c, p, h) = \left(\sum_{j=0}^{h-1} b_j \right)^2 E(T_{1c}^2).$$

$$(b) f_0(c, p, k, h) = f_0(c, p, h) + g_h(k), f_0(c, p, h) = \left(\sum_{j=0}^{h-1} b_j \right)^2 E(T_{0c}^2).$$

Theorem 1 shows that the AFR of both the restricted and unrestricted forecasts can be decomposed into two components: the first component $f_j(c, p, h)$, $j = 0, 1$, depends on both the underlying stochastic/deterministic trends as well as the short-run dynamics through the coefficients $\{b_j\}$; the second component $g_h(k)$ is common to the restricted and unrestricted estimators and depends on the parameters governing the short-run dynamics of the time series. The result generalizes Theorem 2 of Hansen (2010a) for one-step forecasts to multistep forecasts. Interestingly, when $h = 1$, the AFR can be expressed as the sum of a purely nonstationary component representing the stochastic/deterministic trends (since $b_0 = 1$) and a stationary short-run component which is simply the number of

first-differenced lags, i.e., $g_1(k) = k$. However, as Theorem 1 shows, when $h > 1$, such a stationary-nonstationary decomposition no longer holds, since both components now depend on the short-run coefficients $\{\alpha_j\}$. Theorem 1 also generalizes Theorem 2.2 of Ing et al. (2009), which derived an expression for AFR assuming an exact unit root ($c = 0$) and no deterministic component.

The next result, which follows as a direct consequence of Theorem 1, shows that the optimal combination weight is independent of the forecast horizon and the moving average coefficients $\{b_j\}$ but depends on the nuisance parameter c :

Corollary 1. *The AFR of the combination forecast is given by*

$$f_w(c, p, k, h) = \left(\sum_{j=0}^{h-1} b_j \right)^2 \left\{ w^2 E(T_{1c}^2) + (1-w)^2 E(T_{0c}^2) + 2w(1-w) E(T_{1c}T_{0c}) \right\} + g_h(k)$$

with optimal (i.e., AFR minimizing) weight

$$w^* = \frac{E(T_{0c}^2) - E(T_{0c}T_{1c})}{E(T_{0c}^2) + E(T_{1c}^2) - 2E(T_{0c}T_{1c})}$$

4. Asymptotic Mean Squared Error and Asymptotic Forecast Risk

An alternative measure of forecast accuracy is the in-sample asymptotic mean squared error (AMSE) defined as

$$m_u(c, p, k, h) = \lim_{T \rightarrow \infty} \frac{1}{\sigma^2} \sum_{t=1}^{T-h} E(\hat{\mu}_{t+h} - \mu_{t+h})^2$$

for the unrestricted estimator with similar expressions in place for the restricted and averaging estimators. Hansen (2008) established the approximate equivalence between this measure and the AFR under the assumption of strict stationarity. Accordingly, existing forecast combination approaches developed in the stationary framework are based on targeting the AMSE by appealing to its equivalence with the AFR. Hansen (2008) proposed estimating the weights by minimizing a Mallows (2000) criterion which yields an asymptotically unbiased estimate of the AMSE. Similarly, Hansen (2010b) demonstrated that a leave- h -out cross validation criterion delivers an asymptotically unbiased estimate of the AMSE.

This equivalence result, however, breaks down in a nonstationary setup. For instance, when the process has a unit root with no drift and the regression does not include a deterministic component, it follows from the results in Hansen (2010a) that the AMSE of the one-step ahead forecast coincides with the expected value of the squared limiting Dickey–Fuller t -statistic. This expectation has been shown to be about 1.141 by Gonzalo and Pitarakis (1998) and Meng (2005) using analytical and numerical integration techniques, respectively. In contrast, Ing (2001) theoretically established that the AFR of the one-step ahead forecast for the same data generating process and regression is two. More recently, Hansen (2010a) demonstrated the lack of equivalence within a local-to-unity framework, showing that the AMSE of unrestricted as well as restricted (imposing a unit root) one-step ahead forecasts are different from the corresponding expressions for their AFR in autoregressive models with a general lag order and a deterministically trending component. Notwithstanding this result, he suggested using a Mallows criterion to estimate the combination weights and evaluated the adequacy of the resulting combination forecast in finite samples via simulations. A similar approach was taken by Kejriwal and Yu (2021), who also employed Mallows weighting but estimated the deterministic component by FGLS in order to improve upon the accuracy of OLS-based forecasts.

To illustrate the failure of equivalence, Figure 1 plots the AMSE and the AFR of the unrestricted estimator for the case $p = 0$ and $k = 0$.³ The figure clearly illustrates that while

the two measures of forecast accuracy follow a similar path for c sufficiently far from zero, they tend to diverge as the process becomes more persistent. This pattern remains robust across different forecast horizons and suggests that a forecast combination approach that directly targets AFR instead of AMSE can potentially generate more accurate forecasts of highly persistent time series when forecast risk is used as a metric for forecast evaluation.

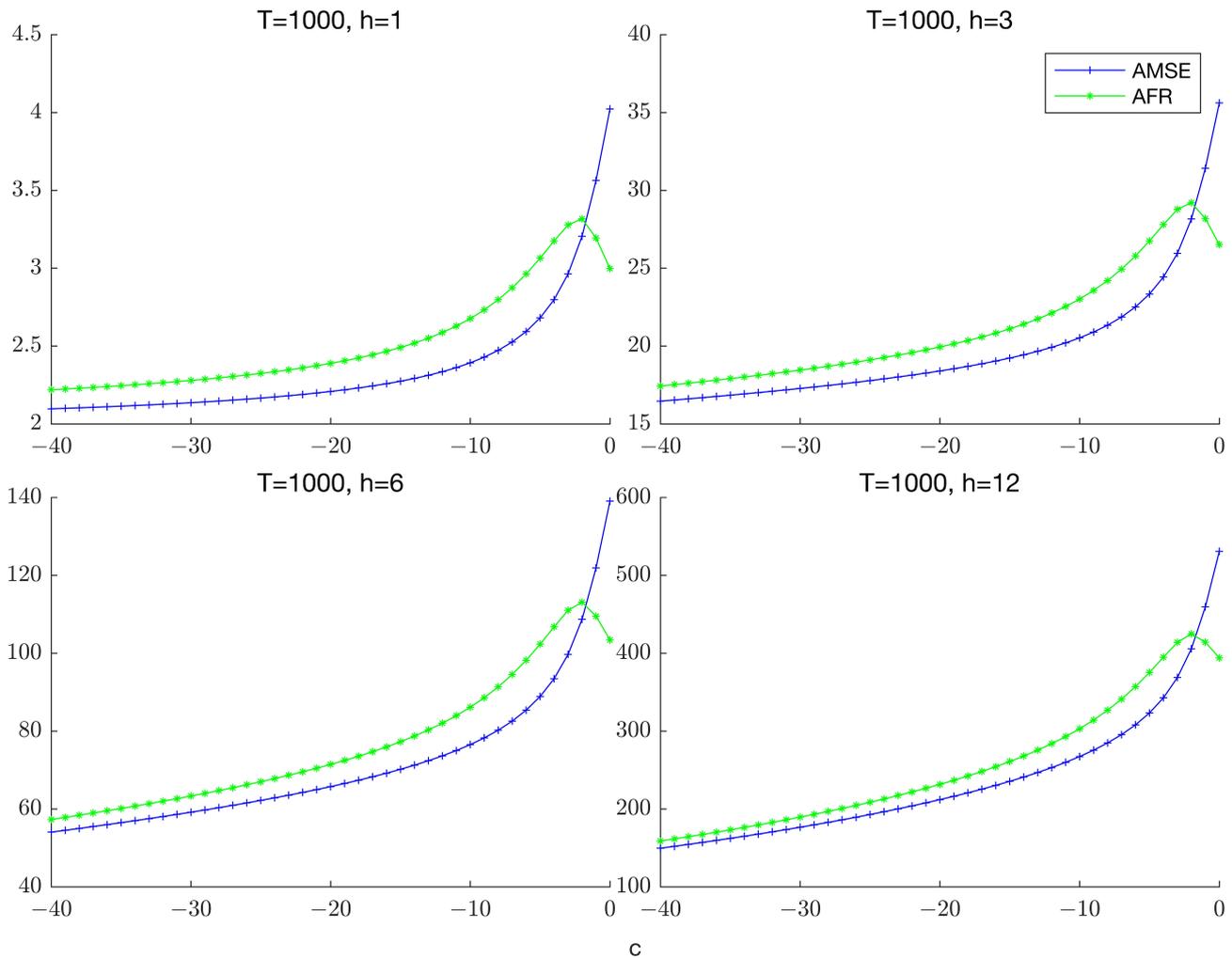


Figure 1. In-sample AMSE versus asymptotic forecast risk ($p = 0, k = 0$).

5. Choice of Combination Weights

The optimal combination forecast $\hat{\mu}_{t+h}(w^*)$ is infeasible in practice, since the weight w^* depends on the unknown local-to-unity parameter c that is not consistently estimable. Given the lack of equivalence between AMSE and AFR for nonstationary time series as discussed in the previous section, we pursue an alternative approach to estimating the combination weights that directly targets the AFR, which is a more direct and practical measure of forecast accuracy than AMSE. In particular, the estimated weight \hat{w} is obtained by minimizing the so-called accumulated prediction errors (APE) criterion defined as

$$APE(w) = \sum_{i=m_h}^{T-h} \{y_{i+h} - \hat{\mu}_{i+h}(w)\}^2 = \sum_{i=m_h}^{T-h} \{w(y_{i+h} - \hat{\mu}_{i+h}) + (1-w)(y_{i+h} - \tilde{\mu}_{i+h})\}^2 \quad (5)$$

with respect to w , where $w \in [0, 1]$, $\hat{\mu}_{i+h}(w)$ is the h -step ahead combination forecast based only on data up to period i , and m_h denotes the smallest positive number such that the forecasts $\hat{\mu}_{i+h}$ and $\tilde{\mu}_{i+h}$ are well defined for all $i \geq m_h$. The solution is given by

$$\hat{w} = \frac{\sum_{i=m_h}^{T-h} (y_{i+h} - \tilde{\mu}_{i+h})^2 - \sum_{i=m_h}^{T-h} (y_{i+h} - \hat{\mu}_{i+h})(y_{i+h} - \tilde{\mu}_{i+h})}{\sum_{i=m_h}^{T-h} (y_{i+h} - \tilde{\mu}_{i+h})^2 + \sum_{i=m_h}^{T-h} (y_{i+h} - \hat{\mu}_{i+h})^2 - 2 \sum_{i=m_h}^{T-h} (y_{i+h} - \hat{\mu}_{i+h})(y_{i+h} - \tilde{\mu}_{i+h})}$$

The APE criterion with $h = 1$ was first introduced by [Rissanen \(1986\)](#) in the context of model selection. [Wei \(1987\)](#) derived the asymptotic properties of APE in general regression models and specialized his results to stationary and nonstationary autoregressive processes with $h = 1$. [Ing \(2004\)](#) demonstrated the strong consistency of the APE-based lag order estimator in stationary autoregressive models for $h \geq 1$. In particular, he showed that a normalized version of the APE is almost certain to converge to the AFR in large samples. [Ing et al. \(2009\)](#) extended the analysis to autoregressive processes with a unit root. The results in [Wei \(1987\)](#), [Ing \(2004\)](#) and [Ing et al. \(2009\)](#) all relied on the law of iterated logarithm which ensures that, in large samples, APE is almost certain to be equivalent to $\log T$ times the AFR. It is, however, important to note that while this convergence result holds pointwise for $|\alpha| \leq 1$, it does not hold uniformly over α . In particular, it does not hold in the local-to-unity setup considered in this paper for $c < 0$.⁴ Nevertheless, the following result shows that the APE criterion remains asymptotically valid in the current framework at the two limits of c which represent the unit root and fixed stationary cases:

Theorem 2. For a given k , let $APE_0 = \sum_{i=m_h}^{T-h} \{y_{i+h} - \tilde{\mu}_{i+h}\}^2$, $APE_1 = \sum_{i=m_h}^{T-h} \{y_{i+h} - \hat{\mu}_{i+h}\}^2$. Under Assumptions 1 and 2 and $\sup_t E(|e_t|^r) < \infty$, for some $r > 2$,

(a) For $c = O(T)$, $\lim_{T \rightarrow \infty} (\sigma^2 \log T)^{-1} (APE_1 - \sum_{i=m_h}^{T-h} \eta_{i,h}^2) = \lim_{c \rightarrow -\infty} f_1(c, p, k, h)$.

(b) $\lim_{c \rightarrow 0} \lim_{T \rightarrow \infty} (\sigma^2 \log T)^{-1} (APE_0 - \sum_{i=m_h}^{T-h} \eta_{i,h}^2) = \lim_{c \rightarrow 0} f_0(c, p, k, h)$.

Remark 1. In a similar vein, [Hansen \(2010a\)](#) developed feasible combination weights by evaluating the Mallows criterion at the two limits of c , given that the criterion depends on c and is therefore infeasible in practice. Thus, while his analysis demonstrated that the infeasible Mallows criterion is an asymptotically unbiased estimate of the AMSE for any c , the feasible version of the criterion remains valid only in the two limit cases. When estimation is performed using FGLS instead of OLS, [Kejriwal and Yu \(2021\)](#) showed that the infeasible Mallows criterion also depends on the parameter a in (1) which governs the short-run dynamics. Evaluating the criterion at the two limits, however, eliminates the dependence on both nuisance parameters.

Figure 2 plots the AFR of the optimal (infeasible) and APE-based combination forecasts for $p = 1$ and $k = 0$.⁵ For comparison, the unrestricted and restricted forecasts are also presented. As expected, the forecast risk of the restricted estimator increases with $|c|$, while the risk function of the unrestricted estimator is relatively flat as a function of c . Regardless of the forecast horizon, the feasible combination forecast maintains a risk profile close to that of the optimal forecast. In particular, the risk of the APE-weighted forecast is uniformly lower than that of the unrestricted estimator across values of c , as well as lower than that of the restricted estimator unless c is very close to zero. These results suggest that the loss in forecast accuracy due to the unknown degree of persistence is relatively small when constructing the combination weights based on the APE criterion. In Sections 7 and 8, we conduct extensive comparisons of the APE-based combination forecasts with both the Mallows and cross-validation based combination forecasts.

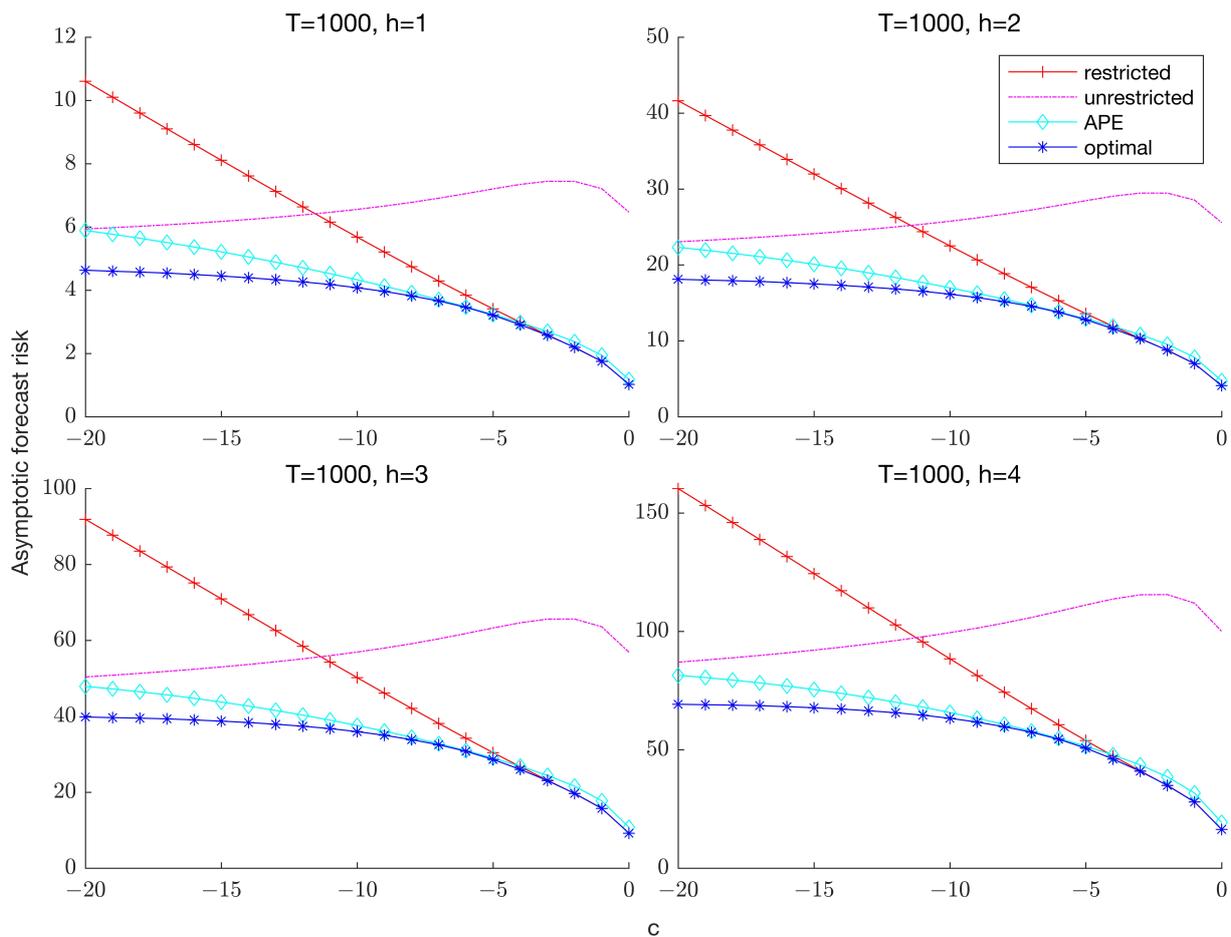


Figure 2. Asymptotic forecast risk of infeasible (optimal) and feasible (APE-based) combination forecasts ($p = 1, k = 0$).

6. Lag Order Uncertainty

This section extends the preceding analysis to the case where the lag order k is unknown. In order to accommodate lag order uncertainty, the set of models on which the combination forecast is based needs to be expanded to include models with different lag orders. Such a forecast can potentially trade off the misspecification bias inherent from the omission of relevant lags against the problem of overfitting induced by the inclusion of unnecessary lags. Kejriwal and Yu (2021) showed that the essence of this trade-off can be captured analytically by adopting a local asymptotic framework in which the coefficients of the short-run dynamics lie in a $O(T^{-1/2})$ -neighborhood of zero in addition to the $O(T^{-1})$ parameterization for the persistence parameter as specified in (1). Specifically, we make the following assumption as in Kejriwal and Yu (2021):

Assumption 3. We assume that $\alpha_i = \frac{\delta_i}{T^{1/2}}$, $i = 1, \dots, k$, where $\delta = (\delta_1, \dots, \delta_k)'$ is fixed and independent of T .

Assumption 3 ensures that the squared misspecification bias from omitting relevant lags is of the same order as the sampling variance introduced by estimating additional lags. Modeling $\{\alpha_i\}$ as fixed would make the bias due to misspecification diverge with the sample size and thus leave no scope for exploiting the trade-off between inclusion and exclusion of lags when constructing the combination forecasts.

We include sub-models with $l \in \{0, 1, \dots, K\}$, $K \geq k$, with the corresponding restricted and unrestricted forecasts given by $\tilde{\mu}_t(l)$ and $\hat{\mu}_t(l)$, respectively. Let $I(l < k) = 1$ if $l < k$, and zero otherwise. Define $\zeta_h(\delta, l, k) = M_h(k)(\mathbf{0}'_l, \delta_{l+1}, \dots, \delta_k)'$, where $\mathbf{0}_l$ is an $(l \times$

1) vector of zeros. Further, let $r_h(\delta, l, k) = \zeta_h(\delta, l, k)' \zeta_h(\delta, l, k)$. The following result derives the AFR of the forecasts in the presence of lag order uncertainty:

Theorem 3. Under Assumptions 1–3 and $\sup_t E(|e_t|^{\theta_h}) < \infty$, where $\theta_h = \max\{8, 2(h + 2)\} + \psi$ for some $\psi > 0$,

$$(a) \lim_{T \rightarrow \infty} \frac{T}{\sigma^2} E(\hat{\mu}_{T+h}(l) - \mu_{T+h})^2 = h^2 E(T_{1c}^2) + g_h(l) + r_h(\delta, l, k).$$

$$(b) \lim_{T \rightarrow \infty} \frac{T}{\sigma^2} E(\check{\mu}_{T+h}(l) - \mu_{T+h})^2 = h^2 E(T_{0c}^2) + g_h(l) + r_h(\delta, l, k).$$

Theorem 3 shows that large sample forecast accuracy now depends on an additional misspecification component $[r_h(\delta, l, k)]$ emanating from the omission of relevant lags. The larger the magnitudes of the coefficients corresponding to the omitted lags, the larger the contribution of this component to the forecast risk. Moreover, under Assumption 3, $g_h(l) = \text{tr}(M_h(l)M_h'(l))$ varies with h for $h < l$ but is constant for all $h \geq l$. Similarly, $r_h(\delta, l, k)$ varies with h for $h < k$ but is constant thereafter. Thus, the forecast horizon only makes a limited contribution to the two short-run components of the asymptotic forecast risk. Another notable feature of Theorem 3 is that, in contrast to the case where the lag order is assumed known (Theorem 1), the contribution of the trend component is now proportional to the square of the forecast horizon. This difference is due to the fact that the coefficients $b_j \rightarrow 1$ for all j since $\alpha_i \rightarrow 0$ for $i = 1, \dots, k$ by virtue of Assumption 3.

We consider two types of combination forecasts. The first is a “partial averaging” forecast that only addresses lag order uncertainty by averaging over the $K + 1$ unrestricted forecasts:

$$\hat{\mu}_{t+h}(\hat{W}) = \sum_{l=0}^K \hat{w}_l \hat{\mu}_t(l) \tag{6}$$

The weights $\hat{W} = (\hat{w}_0, \hat{w}_1, \dots, \hat{w}_K)'$ are obtained by minimizing the APE criterion

$$APE_P(W) = \sum_{i=m_h}^{T-h} \left\{ \sum_{l=0}^K [w_l (y_{i+h} - \hat{\mu}_{i+h}(l))] \right\}^2 \tag{7}$$

where $w_l \geq 0$ ($l = 0, \dots, K$), $\sum_{l=0}^K w_l = 1$. We refer to (6) as the APE-based partial averaging (APA) forecast.

The second forecast is a “general averaging” forecast that accounts for both persistence and lag order uncertainty and thus combines the forecasts from all $2(K + 1)$ sub-models:

$$\check{\mu}_{t+h}(\check{W}) = \sum_{l=0}^K (\check{w}_{1l} \hat{\mu}_t(l) + \check{w}_{0l} \check{\mu}_t(l)) \tag{8}$$

The weights $\check{W} = (\check{w}_{01}, \check{w}_{02}, \dots, \check{w}_{0K}, \check{w}_{11}, \check{w}_{12}, \dots, \check{w}_{1K})'$ are obtained by minimizing a generalized APE criterion of the form

$$APE_G(W) = \sum_{i=m_h}^{T-h} \left\{ \sum_{l=0}^K [w_{1l} (y_{i+h} - \hat{\mu}_{i+h}(l)) + w_{0l} (y_{i+h} - \check{\mu}_{i+h}(l))] \right\}^2 \tag{9}$$

where $w_{1l} \geq 0, w_{0l} \geq 0$ ($l = 0, \dots, K$), $\sum_{l=0}^K (w_{0l} + w_{1l}) = 1$. We refer to (8) as the APE-based general averaging (AGA) forecast. Comparing the APA and AGA forecasts will serve to isolate the effects of the two sources of uncertainty on forecast accuracy.

The following result establishes the limiting behavior of the APE criterion in the presence of lag-order uncertainty:

Theorem 4. Let $APE_0(l) = \sum_{i=m_h}^{T-h} \{y_{i+h} - \check{\mu}_{i+h}(l)\}^2$, $APE_1(l) = \sum_{i=m_h}^{T-h} \{y_{i+h} - \hat{\mu}_{i+h}(l)\}^2$. Under Assumptions 1–3 and $\sup_t E(|e_t|^r) < \infty$, for some $r > 2$,

$$\begin{aligned}
 (a) & \text{ For } c = O(T), \lim_{T \rightarrow \infty} (\sigma^2 \log T)^{-1} \left(APE_1(l) - \sigma^2 r_h(\delta, l, k) - \sum_{i=m_h}^{T-h} \eta_{i,h}^2 \right) = h^2 \lim_{c \rightarrow -\infty} E(T_{1c}^2) + g_h(l). \\
 (b) & \lim_{c \rightarrow 0} \lim_{T \rightarrow \infty} (\sigma^2 \log T)^{-1} \left(APE_0(l) - \sigma^2 r_h(\delta, l, k) - \sum_{i=m_h}^{T-h} \eta_{i,h}^2 \right) = h^2 \lim_{c \rightarrow 0} E(T_{0c}^2) + g_h(l).
 \end{aligned}$$

Theorem 4 shows that while APE captures the components of AFR that are attributable to persistence uncertainty and estimation of the short-run dynamics, it does not account for lag-order uncertainty in the limit. As shown in the proof of Theorem 4, this is because the former two components grow at a logarithmic rate while the bias component due to lag order misspecification is bounded [i.e., $O(1)$]. Nonetheless, given that the logarithmic function is slowly varying, it can be expected that in small samples, the APE is still effective at capturing the bias that occurs due to misspecification of the number of lags. Indeed, as shown subsequently in the simulations, the APE criterion offers considerable improvements over its competitors, even under lag-order misspecification.

7. Monte Carlo Simulations

This section reports the results of a set of Monte Carlo experiments designed to (1) evaluate the finite sample performance of the proposed approach relative to extant approaches; (2) quantify the importance of accounting for each source of uncertainty in terms of its effect on finite sample forecast risk. Section 7.1 lays out the experimental design. Section 7.2 details the different forecasting procedures included in the analysis. Sections 7.3 and 7.4 present the results. Results are obtained for $p \in \{0, 1\}$. For brevity, we report the results only for $p = 1$. The results for $p = 0$ are qualitatively similar, although the improvements offered by the proposed approach are more pronounced for $p = 1$ than $p = 0$. The full set of results is available upon request.

7.1. Experimental Design

We adopt a design similar to that in Hansen (2010a) and Kejriwal and Yu (2021) to facilitate direct comparisons. The data generating process (DGP) is based on (1) and specified as follows: (a) the innovations $\epsilon_t \stackrel{i.i.d.}{\sim} N(0, 1)$; (b) the trend parameters are set at $\beta_0 = \beta_1 = 0$; (c) the true lag order $k \in \{0, 6, 12\}$ with $\alpha_j = -(-\theta)^j$ for $j = 1, \dots, k$ and $\theta = 0.6$. The maximum number of first-differenced lags included is set at $K = 12$. The sample size is set at $T \in \{100, 200\}$. The local-to-unity parameter c varies from -20 to 0 , implying α ranging from 0.8 to 1 for $T = 100$ and α ranging from 0.9 to 1 for $T = 200$. At each c value, the finite-sample forecast risk $TE[(\hat{\mu}_{T+h} - \mu_{T+h})^2]$ is computed for all estimators considered, where $h \in \{1, 3, 6, 12\}$. All experiments are based on 10,000 Monte Carlo replications.

We report two sets of results. The first assumes k is known, thereby allowing us to demonstrate the effect of persistence uncertainty on forecast accuracy while abstracting from lag order uncertainty. The second allows k to be unknown and facilitates the comparison between forecasts that address both forms of uncertainty with those that only account for lag order uncertainty.

7.2. Forecasting Methods

The benchmark forecast in both the known and unknown lag cases is calculated from a standard autoregressive model of order $K + 1$ estimated by OLS:

$$y_t = \beta_0^* + \beta_1^* t + \alpha y_{t-1} + \sum_{j=1}^K \alpha_j \Delta y_{t-j} + \epsilon_t \tag{10}$$

When the number of lags is assumed to be known (Section 7.3), we compare a set of six forecasting methods: (1) Mallows selection (Mal-Sel); (2) Cross-validation selection (CVh-

Sel); (3) APE selection (APE-Sel); (4) Mallows averaging (Mal-Ave); (5) Cross-validation averaging (CVh-Ave); (6) APE averaging (APE-Ave). With an unknown number of lags, the following six methods are compared⁶: (1) Mallows partial averaging (MPA); (2) Cross-validation partial averaging (CPA); (3) APE partial averaging (APA); (4) Mallows general averaging (MGA); (5) Cross-validation general averaging (CGA); (6) APE general averaging (AGA). For brevity, a detailed description of these methods is not presented here but is included in Appendix B.

Both the APE selection and combination forecasts require a choice of m_h . To our knowledge, no data-dependent methods for choosing m_h are available in the existing literature. We therefore examined the viability of alternative choices via simulations. Specifically, for each persistence level (value of c), we computed the minimum forecast risk over all values of m_h in the range [15, 70] with a step-size of 5 (assuming a known number of lags k). While no single value was found to be uniformly dominant across persistence levels/horizons, $m_h = 20$ turned out to be a reasonable choice overall.⁷ To justify this choice, Figure A1 in Appendix C plots the difference between the optimal forecast risk and the risk of the APE selection forecasts for $m_h = 20$ expressed as a percentage of the forecast risk for $m_h = 20$. The corresponding results for the APE combination forecasts are presented in Figure A2. It is evident that using $m_h = 20$ entails only a marginal increase in forecast risk (at most 5%) for the combination forecasts over the optimal forecast risk across different persistence levels and horizons. In contrast, the optimal choice of m_h for the selection forecasts is somewhat more unstable and appears to depend more heavily on the forecast horizon and the level of persistence. This robustness in behavior provides additional motivation for employing a combination approach to forecasting in practice.

7.3. Forecast Risk with Known Lag Order

Figures 3a–5b plot the risk of the six methods relative to the benchmark. First, we consider the case $k = 0$. Several features of the results are noteworthy.

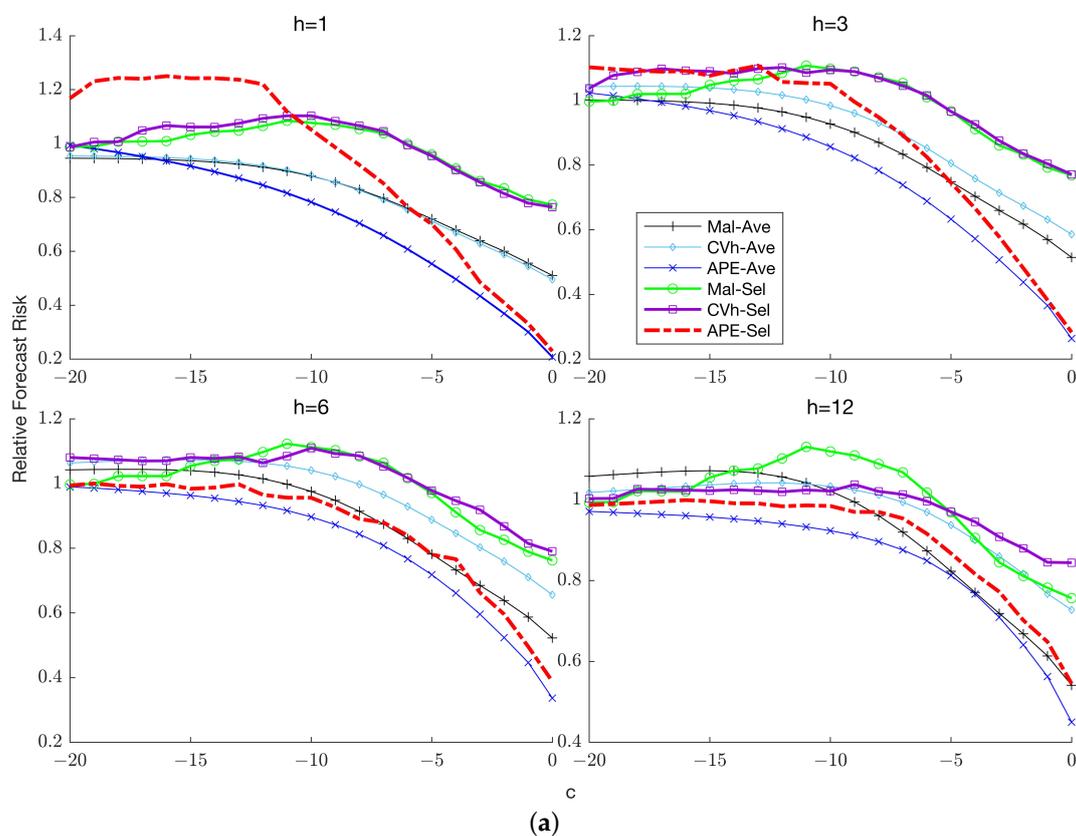


Figure 3. Cont.

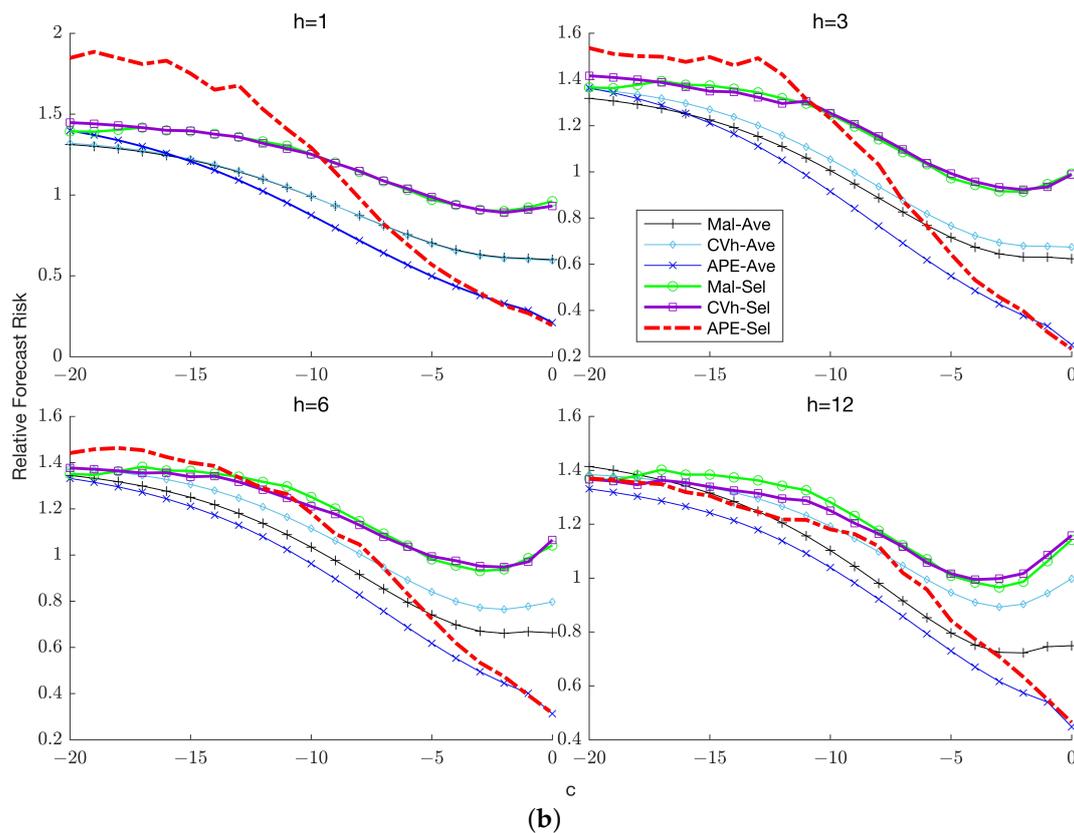
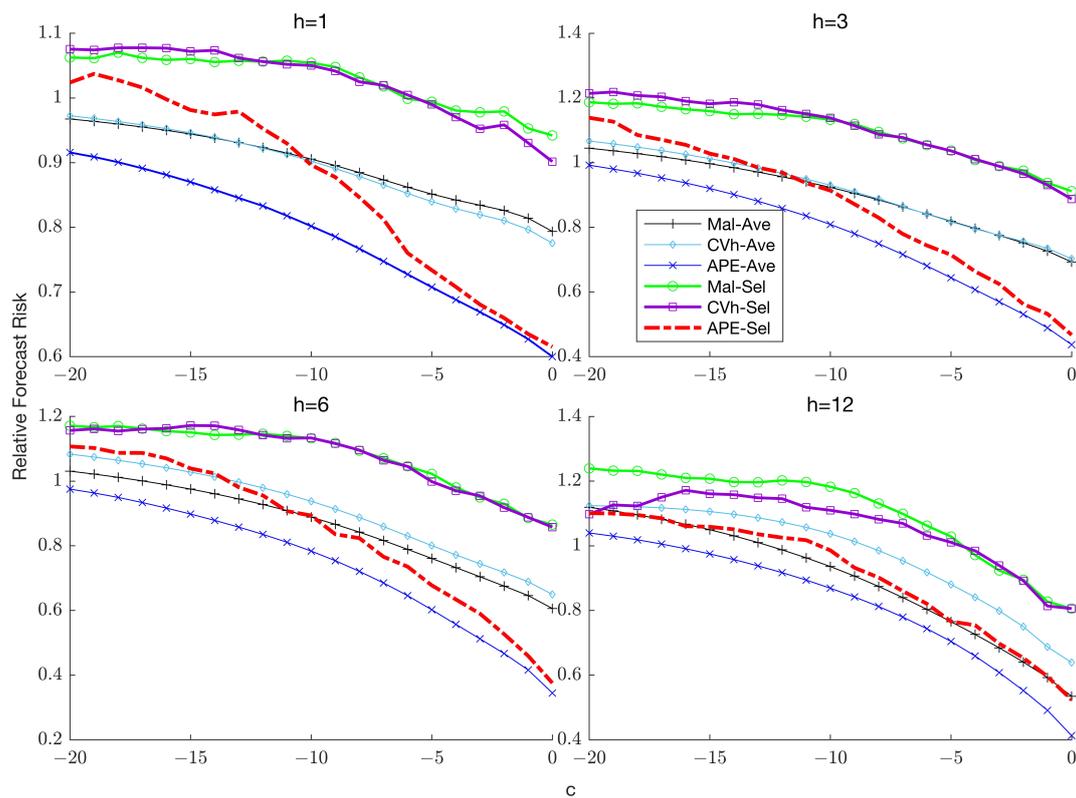


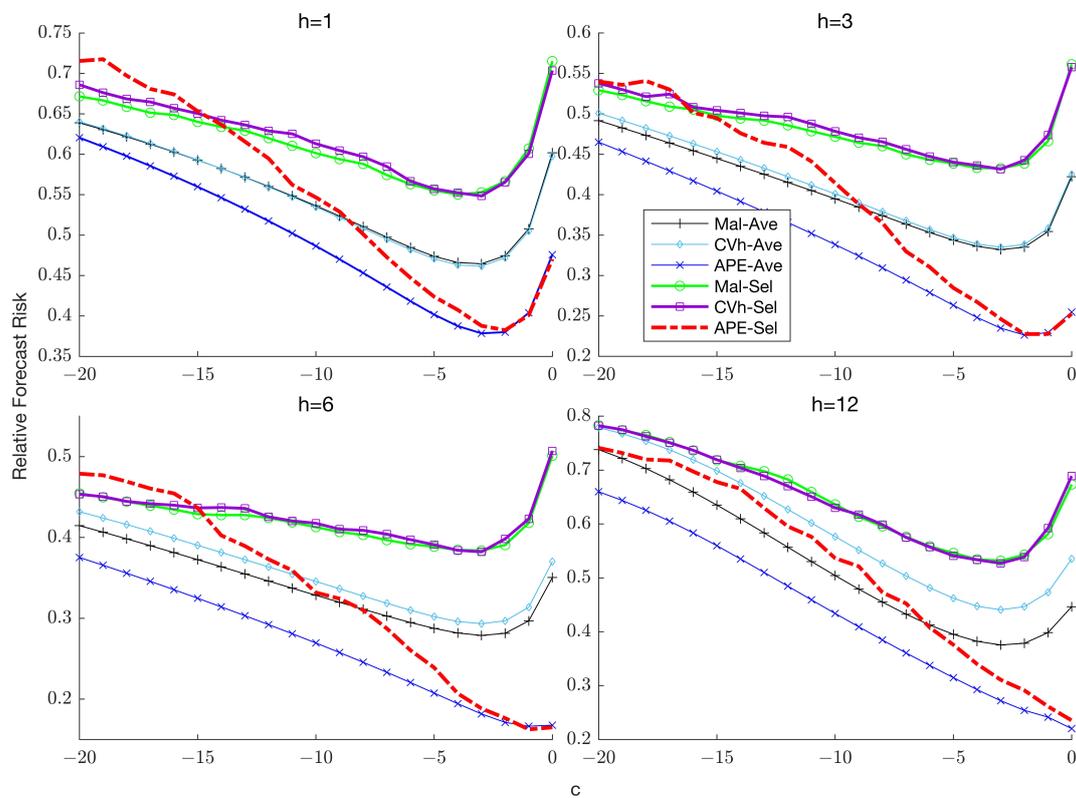
Figure 3. (a). Forecast risk with known lag order ($k = 0, T = 100$). (b). Forecast risk with known lag order ($k = 0, T = 200$).

First, the selection forecasts typically exhibit higher risk than the corresponding combination forecasts across sample sizes and horizons. Second, when $T = 100$, the APE combination forecast is clearly the dominant method, performing discernibly better than forecasts based on either of the two competing weighting schemes. When $T = 200$, its dominance continues except when $|c|$ is sufficiently large (the exact magnitude being horizon-dependent), in which case the benchmark delivers the most accurate forecasts and averaging over the restricted model becomes less attractive. Third, the relative performance of the Mallows and cross-validation weighting schemes depends on the horizon: at $h = 1$, the two schemes yield virtually indistinguishable forecasts; when $h \in \{3, 6\}$, Mallows weighting yields uniformly lower risk over the parameter space; at $h = 12$, Mallows weighting is preferred when persistence is high (c close to zero) while cross-validation weighting dominates for lower levels of persistence.

In the presence of higher order serial correlation ($k > 0$), the superior performance of the APE combination forecast becomes even more evident: it now dominates all competing forecasts regardless of horizon and sample size. In particular, APE weighting outperforms the benchmark at all persistence levels, even at $T = 200$, unlike the $k = 0$ case. The intuition for this difference in relative performance between the cases with and without higher-order serial correlation is that in the former case, averaging is comparatively more beneficial, since imposing the unit root restriction can potentially reduce the estimation uncertainty associated with the coefficients of the lagged differences. This reduction in sampling uncertainty in turn engenders a reduction in the overall risk of the combination forecast relative to the unrestricted benchmark forecast. Another notable difference from the $k = 0$ case is that while Mallows and cross-validation weighting are comparable for $h \in \{1, 3\}$, the former now dominates for $h \in \{6, 12\}$ uniformly over the parameter space.

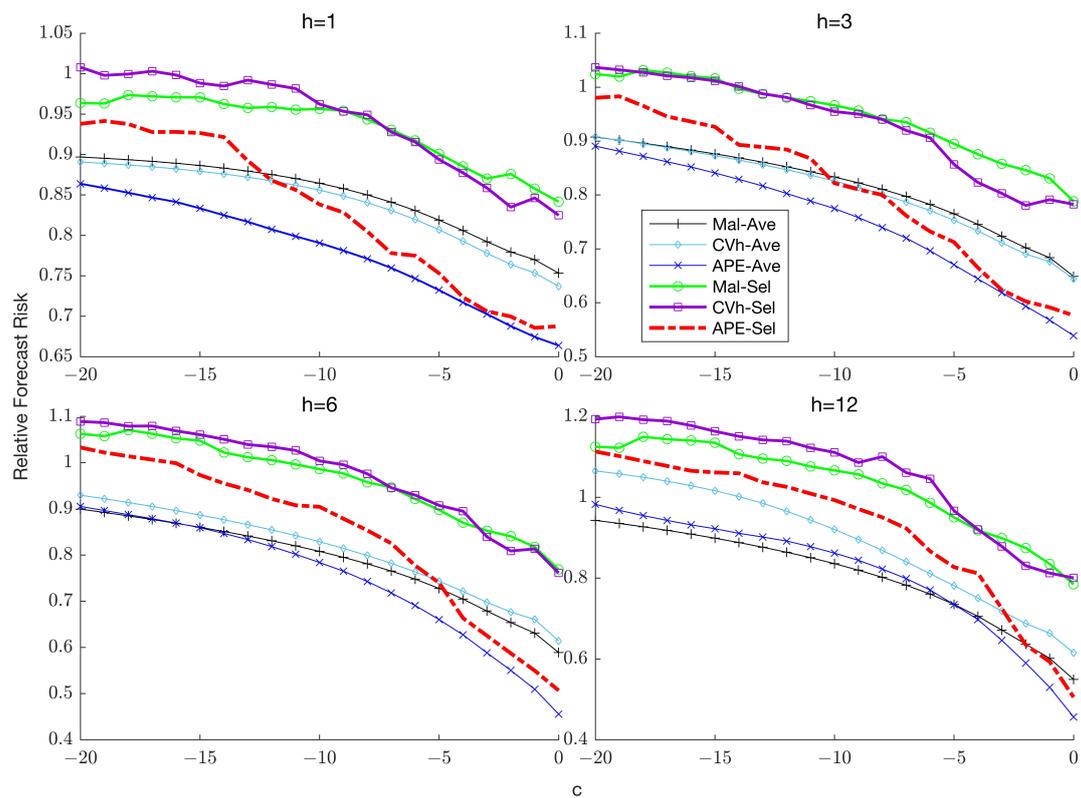


(a)

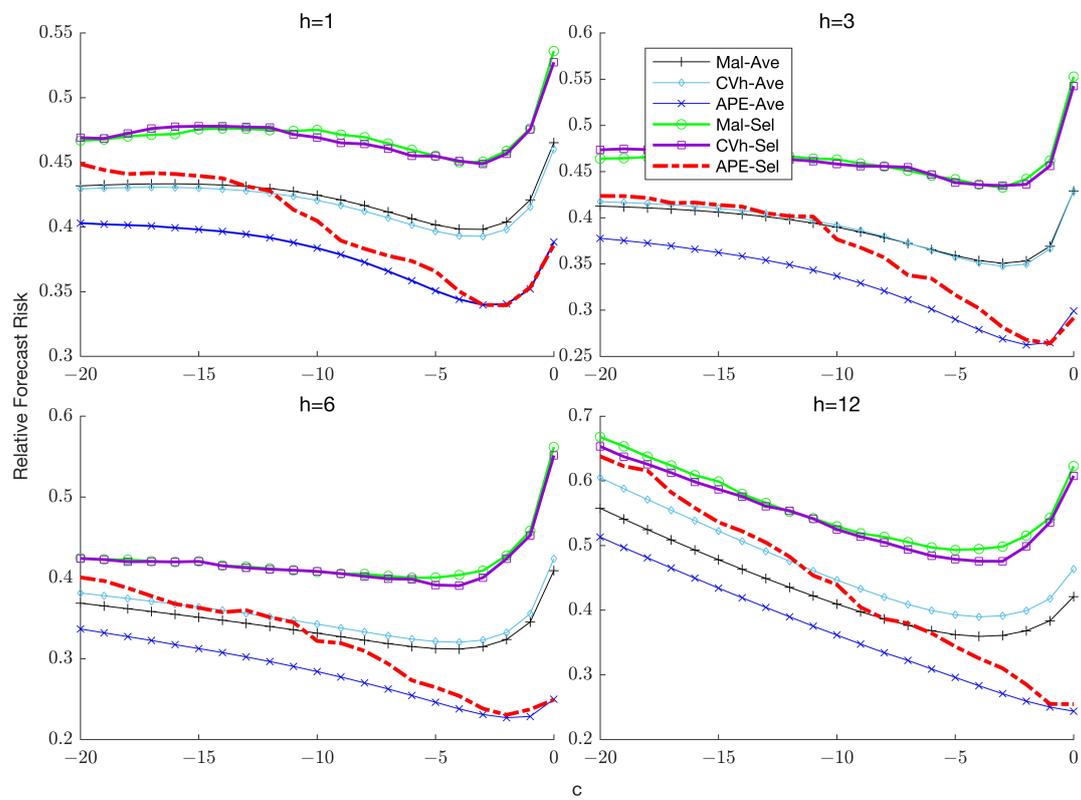


(b)

Figure 4. (a). Forecast risk with known lag order ($k = 6, T = 100$). (b). Forecast risk with known lag order ($k = 6, T = 200$).



(a)



(b)

Figure 5. (a). Forecast risk with known lag order ($k = 12, T = 100$). (b). Forecast risk with known lag order ($k = 12, T = 200$).

7.4. Forecast Risk with Unknown Lag Order

Figures 6a–8b plot the relative risk of the six combination forecasts which comprise the three partial forecasts that only account for lag-order uncertainty and the three general forecasts that account for both lag-order and stochastic trend uncertainty. A clear implication of these results is that general averaging methods typically exhibit considerably lower forecast risk than partial averaging methods unless the process has relatively low persistence, in which case averaging over the unit root model increases the forecast risk incurred by the general averaging methods. The improvements offered by general averaging hold across both horizons and the number of lags (k) in the true DGP and become more prominent as the sample size increases.

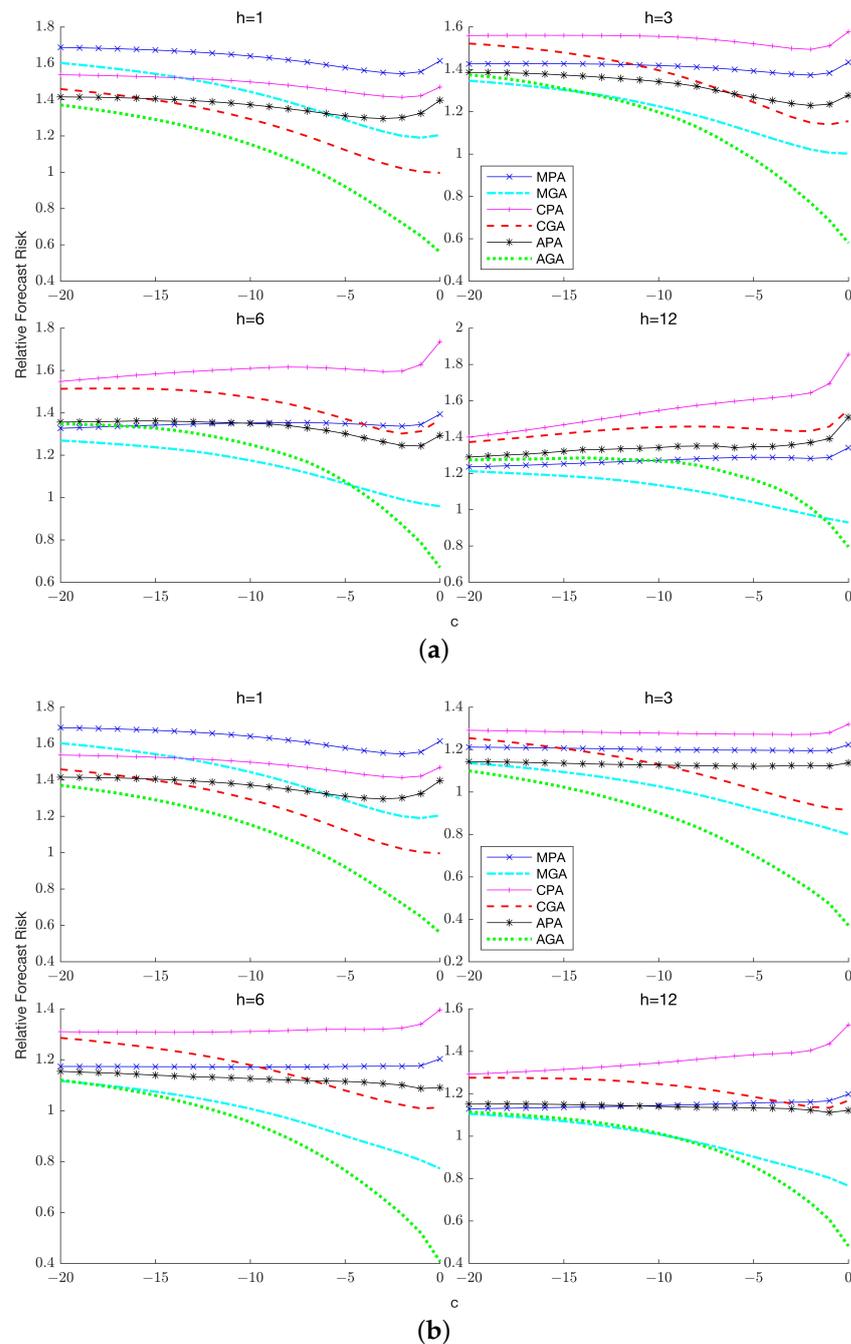


Figure 6. (a). Forecast risk with unknown lag order ($k = 0, T = 100$). (b). Forecast risk with unknown lag order ($k = 0, T = 200$).

Among the three weighting schemes, APE-based weights are the preferred choice except when $h \in \{6, 12\}$ and $T = 100$, where Mallows weighting turns out to be the dominant approach if persistence is relatively low. A potential explanation for this result is that with long horizons and a small sample size, the APE criterion is based on a relatively smaller number of prediction errors, which increases the sampling variability associated with the resulting weights, thereby increasing the risk of the combination forecast. As in the known lag-order case, the choice between Mallows and cross-validation weighting is horizon-dependent: when $h = 1$, cross-validation weighting is preferred while when $h > 1$, Mallows weighting is preferred, with the magnitude of reduction in forecast risk increasing as h increases.

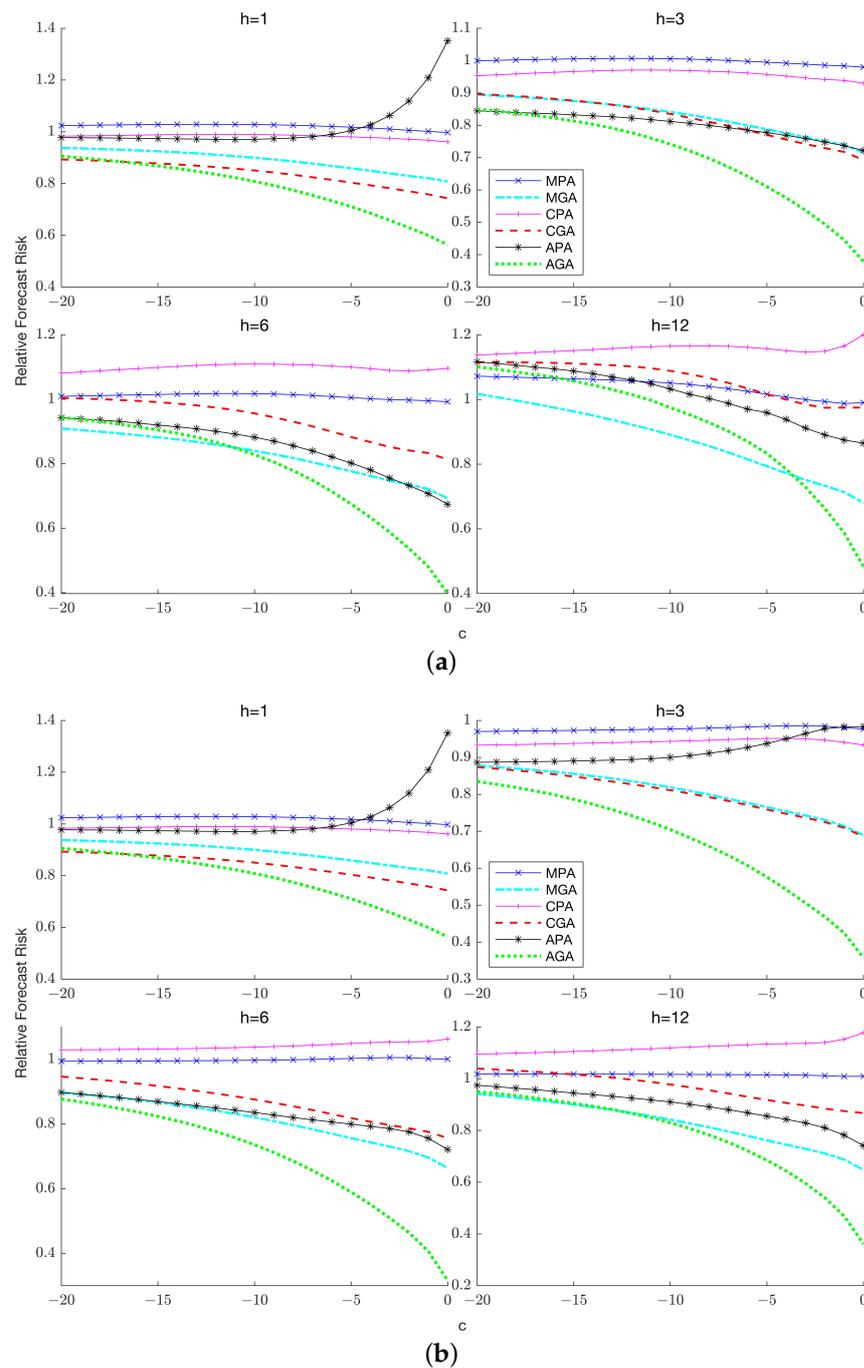


Figure 7. (a). Forecast risk with unknown lag order ($k = 6, T = 100$). (b). Forecast risk with unknown lag order ($k = 6, T = 200$).

In summary, the results from the simulation experiments make a strong case for employing APE weights when constructing the combination forecasts and clearly highlight the benefits of targeting forecast risk rather than in-sample mean squared error. The comparison of general and partial combination forecasts also underscore the importance of concomitantly controlling for both stochastic trend uncertainty and lag-order uncertainty in generating accurate forecasts.

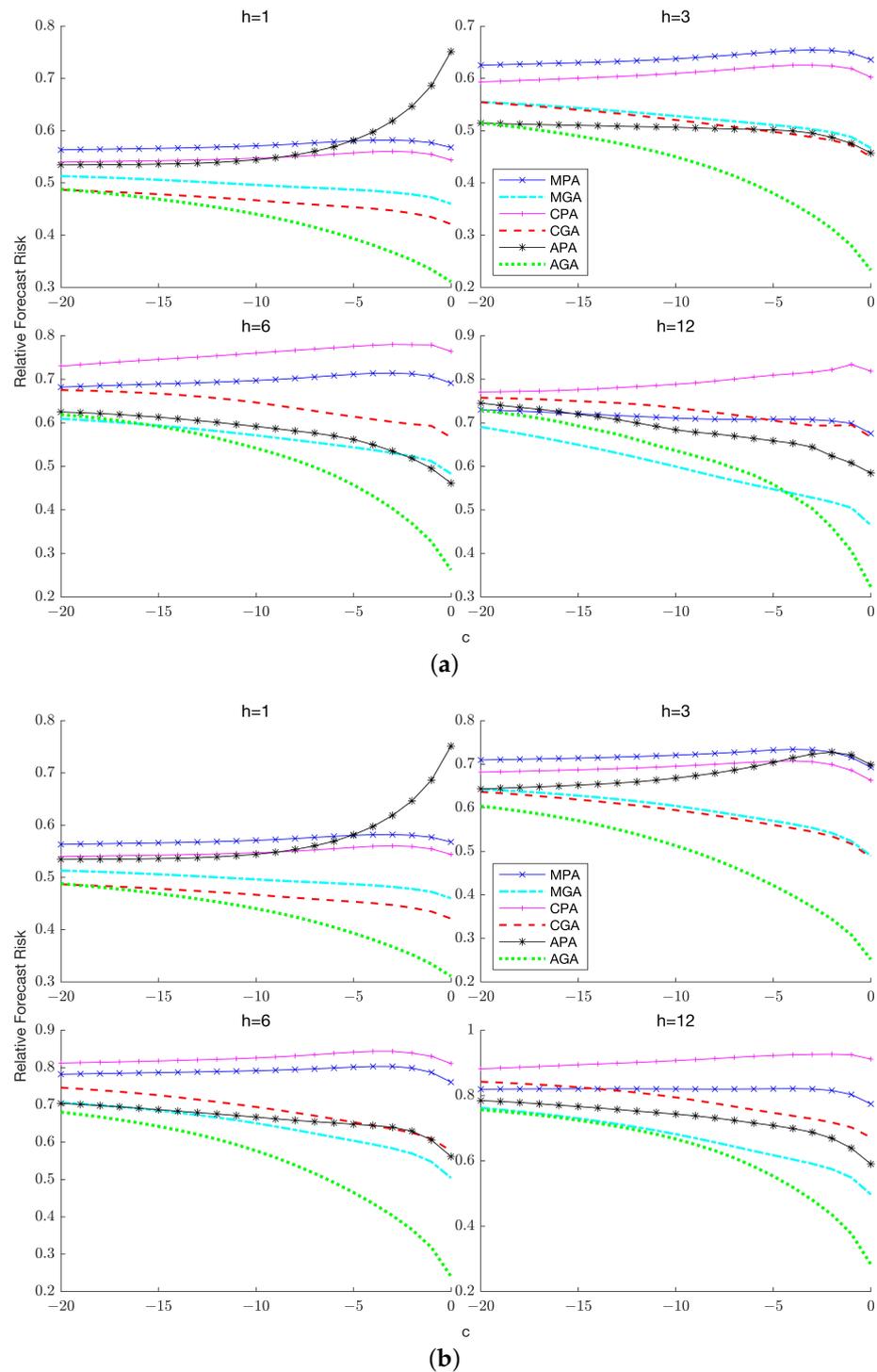


Figure 8. (a). Forecast risk with unknown lag order ($k = 12, T = 200$). (b). Forecast risk with unknown lag order ($k = 12, T = 100$).

8. Empirical Application

This section conducts a pseudo out-of-sample forecast comparison of the different multistep forecast combination methods using a set of US macroeconomic time series. Our objectives are to empirically assess (1) the efficacy of different averaging/selection methods relative to a standard autoregressive benchmark; (2) the importance of averaging over both the persistence level and the lag order; and (3) the relative performance of alternative weight choices for constructing the combination forecasts.

Our analysis employs the FRED-MD data set compiled by [McCracken and Ng \(2016\)](#), which contains 123 monthly macroeconomic variables over the period January 1960–December 2018.⁸ [McCracken and Ng \(2016\)](#) suggested a set of seven transformation codes designed to render each series stationary: (1) no transformation; (2) Δy_t ; (3) $\Delta^2 y_t$; (4) $\log(y_t)$; (5) $\Delta \log(y_t)$; (6) $\Delta^2 \log(y_t)$; (7) $\Delta(y_t/y_{t-1} - 1)$. To ensure that the series fit our framework, which allows for highly persistent time series with/without deterministic trends, we adopt the following transformation codes as modified by [Kejriwal and Yu \(2021\)](#): (1') no transformation; (2') y_t ; (3') Δy_t ; (4') $\log(y_t)$; (5') $\log(y_t)$; (6') $\Delta \log(y_t)$; (7') $y_t/y_{t-1} - 1$. For series that correspond to codes (1') and (4'), we construct the forecasts from a model with no deterministic trend ($p = 0$), while for the remaining codes, we use forecasts from a model that include a linear deterministic trend ($p = 1$). We also report results for eight core series as in [Stock and Watson \(2002\)](#), comprising four real and four nominal variables. As in the simulation experiments, four alternative forecast horizons are considered: $h \in \{1, 3, 6, 12\}$. We use a rolling window scheme with an initial estimation period of January 1960–December 1969 so that the forecast evaluation period is January 1970–December 2018 (588 observations). The size of the estimation window changes depending on the forecast horizon h . For example, when $h = 1$, the initial training sample contains 120 observations from January 1960–December 1969 while for $h = 3$, it contains only 118 observations from January 1960–October 1969. This ensures that the forecast origin is January 1970 for all forecast horizons considered. We compare ten different methods in terms of the mean squared forecast error (MSFE) computed as the average of the squared forecast errors: (1) MPA: Mallows partial averaging over the number of lags only in the unrestricted model; (2) MGA: Mallows general averaging over both the unit root restriction and the number of lags; (3) CPA: leave- h -out cross-validation (CV- h) averaging over the number of lags only in the unrestricted model; (4) CGA: leave- h -out cross-validation averaging over both the unit root restriction and the number of lags; (5) APA: accumulated prediction error averaging over the number of lags only in the unrestricted model; (6) AGA: accumulated prediction error averaging over both the unit root restriction and the number of lags; (7) MS: Mallows selection from all models (unrestricted and restricted) that vary with the number of lags; (8) CVhS: leave- h -out cross-validation selection from all models (unrestricted and restricted) that vary with the number of lags; (9) APES: accumulated prediction error selection from all models (unrestricted and restricted) that vary with the number of lags; (10) AR: unrestricted autoregressive model (benchmark). The maximum number of allowable first differenced lags in each method is set at $K = 12$. The benchmark forecast is computed from unrestricted OLS estimation of an autoregressive model of the form (10) that uses 12 first-differenced lags of the dependent variable and includes/excludes a deterministic trend depending on the transformation code the series corresponds to, as discussed above.

Table 1a ($h = 1, 3$) and Table 1b ($h = 6, 12$) report the percentages of wins and losses based on the MSFE for the 123 series. Specifically, they show the percentage of 123 series for which a method listed in a row outperforms a method listed in a column, and all other methods (last column). A summary of the results in Table 1(a and b) is given below:

1. The averaging methods uniformly dominate their selection counterparts at all forecast horizons. For instance, Mallows/cross-validation averaging outperform the corresponding selection procedures in more than 90% of the series at each horizon. The performance of AGA relative to APES is relatively more dependent on the horizon, with improvements observed in 77% (65%) of the series for $h = 1$ ($h = 12$), respectively.

2. Given a particular weighting scheme, averaging over both the unit root restriction and number of lags (general averaging) outperforms averaging over only the number of lags (partial averaging) at all horizons. For instance, when $h = 1$, MGA (CGA, AGA) dominate MPA (CPA, APA) in 95% (81%, 79%) of the series, respectively, based on pairwise comparisons. A similar pattern is observed for multi-step forecasts.
3. Across all horizons, AGA emerges as the leading procedure due to its ability to deliver forecasts with the lowest MSFE among all methods for the maximum number of series (last column of Table 1(a and b)). This approach also dominates each of the competing approaches in terms of pairwise comparisons. The APES approach ranks second among all methods so that forecasting based on the accumulated prediction errors criterion (either AGA or APES) outperforms the other approaches for more than 50% of the series over each horizon (the specific percentages are 68.3% for $h = 1, 3$; 57.7% for $h = 6$; 55.3% for $h = 12$).

Next, we examine the performance of the forecasting methods for different types of series based on their groupwise classification by [McCracken and Ng \(2016\)](#) in an attempt to uncover the extent to which the best methods vary by the type of series analyzed. In particular, [McCracken and Ng \(2016\)](#) classified the series into eight distinct groups: (1) output and income; (2) labor market; (3) housing; (4) consumption, orders and inventories; (5) money and credits; (6) interest and exchange rates; (7) prices; (8) stock market. For each of these groups, Table 2 reports the method(s) with the lowest MSFE for the most series compared to all other competing methods. We also report the number of horizons in which (a) averaging outperforms selection and vice-versa; (b) averaging over both the unit root restriction and number of lags (general averaging—GA) methods is superior to averaging over only the number of lags (partial averaging—PA) and vice-versa; (c) each of the three weighting schemes dominates the other two. The results are consistent with those in Table 1(a and b) and clearly demonstrate (1) the dominance of averaging over selection (with the exception of Group 3); (2) the benefits of accounting for both stochastic trend uncertainty and lag order uncertainty (GA) relative to only the latter (PA) for five out of the eight groups; (3) the superiority of APE weighting over the two competing weighting schemes (the exception is Group 5, where cross-validation weighting is the dominant approach).

Finally, we present a comparison of the different methods with respect to their ability to forecast the eight core series analyzed in [Stock and Watson \(2002\)](#). Table 3 reports the MSFE of the eight methods relative to the benchmark model (10) for four real variables (industrial production, real personal income less transfers, real manufacturing and trade sales, number of employees on nonagricultural payrolls) while Table 4 reports the corresponding results for four nominal variables (the consumer price index, the personal consumption expenditure implicit price deflator, the consumer price index less food and energy, and the producer price index for finished goods). To assess whether the difference between the proposed methods and the benchmark model is statistically significant, we use a two-tailed Diebold–Mariano test statistic ([Diebold and Mariano 1995](#)). A number less than one indicates better forecast performance than the benchmark and vice versa. The method with smallest relative MSFE for a given series is highlighted in bold.

Consider first the results for real variables (Table 3). The performance of the best method is statistically significant (at the 10% level) relative to the benchmark in twelve out of the sixteen cases. Consistent with the results in Tables 1(a and b) and 2, general averaging typically dominates partial averaging, the exceptions being nonagricultural employment for $h \leq 6$, industrial production at $h = 12$, and real manufacturing and trade sales for $h = 6, 12$, where APES is the dominant procedure. The AGA approach turns out to have the highest relative forecast accuracy in 50% of all cases, with the improvements offered over rival approaches being particularly notable at $h = 12$. While cross-validation weighting does not yield the best forecasting procedure in any of the cases, Mallows weighting is the preferred approach in only two cases, although the improvements are statistically insignificant. Turning to the nominal variables (Table 4), the best method significantly outperforms the benchmark in ten cases. Again, general averaging is usually preferred to

partial averaging, the exception being the case $h = 12$, where APA outperforms all other methods for three of the four variables. As with the real variables, the AGA forecast is the most accurate in 50% of all cases, though the improvements are now comparable across horizons. Finally, cross-validation weighting partly redeems itself by providing the best forecast in four cases, while Mallows weighting is the preferred method in only one case.

Table 1. Percentage wins/losses of different forecasting methods for (a) $h = 1$ and $h = 3$ and (b) $h = 6$ and $h = 12$.

Methods	MPA	MGA	CPA	CGA	APA	AGA	MS	CVhS	APES	AR	All
(a) $h = 1$ & $h = 3$											
MPA	0.0	4.9	22.0	3.3	43.9	7.3	96.7	58.5	25.2	100.0	0.0
MGA	95.1	0.0	61.8	13.8	61.8	13.0	99.2	71.5	40.7	100.0	4.9
CPA	78.0	38.2	0.0	18.7	66.7	16.3	95.1	81.3	36.6	100.0	5.7
CGA	96.7	86.2	81.3	0.0	78.0	23.6	98.4	95.9	53.7	100.0	8.9
APA	56.1	38.2	33.3	22.0	0.0	21.1	71.5	54.5	35.0	78.9	11.4
AGA	92.7	87.0	83.7	76.4	78.9	0.0	97.6	93.5	77.2	100.0	54.5
MS	3.3	0.8	4.9	1.6	28.5	2.4	0.0	18.7	9.8	90.2	0.0
CVhS	41.5	28.5	18.7	4.1	45.5	6.5	81.3	0.0	22.0	95.1	0.8
APES	74.8	59.3	63.4	46.3	65.0	22.8	90.2	78.0	0.0	95.9	13.8
AR	0.0	0.0	0.0	0.0	21.1	0.0	9.8	4.9	4.1	0.0	0.0
MPA	0.0	6.5	43.1	8.1	50.4	10.6	87.0	58.5	23.6	97.6	0.8
MGA	93.5	0.0	65.9	43.1	67.5	17.1	97.6	73.2	36.6	99.2	8.9
CPA	56.9	34.1	0.0	15.4	59.3	16.3	81.3	68.3	30.1	96.7	4.1
CGA	91.9	56.9	84.6	0.0	84.6	23.6	95.1	93.5	43.1	99.2	10.6
APA	49.6	32.5	40.7	15.4	0.0	14.6	62.6	48.0	29.3	71.5	6.5
AGA	89.4	82.9	83.7	76.4	85.4	0.0	95.9	93.5	69.9	100.0	41.5
MS	13.0	2.4	18.7	4.9	37.4	4.1	0.0	34.1	10.6	80.5	0.0
CVhS	41.5	26.8	31.7	6.5	52.0	6.5	65.9	0.0	14.6	89.4	0.8
APES	76.4	63.4	69.9	56.9	70.7	30.1	89.4	85.4	0.0	95.1	26.8
AR	2.4	0.8	3.3	0.8	28.5	0.0	19.5	10.6	4.9	0.0	0.0
(b) $h = 6$ & $h = 12$											
MPA	0.0	7.3	61.8	12.2	57.7	21.1	82.9	57.7	32.5	95.9	3.3
MGA	92.7	0.0	83.7	52.0	83.7	34.1	98.4	83.7	46.3	98.4	13.0
CPA	38.2	16.3	0.0	13.8	60.2	21.1	65.0	60.2	30.1	87.0	8.1
CGA	87.8	48.0	86.2	0.0	86.2	35.8	93.5	96.7	44.7	98.4	13.8
APA	42.3	16.3	39.8	13.8	0.0	21.1	56.1	47.2	30.9	65.9	3.3
AGA	78.9	65.9	78.9	64.2	78.9	0.0	92.7	82.9	69.9	99.2	34.1
MS	17.1	1.6	35.0	6.5	43.9	7.3	0.0	36.6	17.1	78.0	0.0
CVhS	42.3	16.3	39.8	3.3	52.8	17.1	63.4	0.0	19.5	78.0	0.8
APES	67.5	53.7	69.9	55.3	69.1	30.1	82.9	80.5	0.0	96.7	23.6
AR	4.1	1.6	13.0	1.6	34.1	0.8	22.0	22.0	3.3	0.0	0.0
MPA	0.0	10.6	60.2	16.3	74.8	21.1	85.4	46.3	23.6	91.9	1.6
MGA	89.4	0.0	75.6	48.0	84.6	27.6	97.6	70.7	30.9	97.6	10.6
CPA	39.8	24.4	0.0	13.8	66.7	25.2	58.5	47.2	26.8	81.3	8.9
CGA	83.7	52.0	86.2	0.0	81.3	36.6	91.9	91.9	40.7	98.4	10.6
APA	25.2	15.4	33.3	18.7	0.0	22.8	47.2	35.8	28.5	65.9	8.1
AGA	78.9	72.4	74.8	63.4	77.2	0.0	87.8	77.2	65.0	95.9	31.7
MS	14.6	2.4	41.5	8.1	52.8	12.2	0.0	33.3	15.4	78.9	0.8
CVhS	53.7	29.3	52.8	8.1	64.2	22.8	66.7	0.0	26.8	82.9	4.1
APES	76.4	69.1	73.2	59.3	71.5	35.0	84.6	73.2	0.0	91.9	23.6
AR	8.1	2.4	18.7	1.6	34.1	4.1	21.1	17.1	8.1	0.0	0.0

Note: Percentage of the 123 series for which a method listed in a row outperforms a method in a column, and all other methods (last column). AR refers to the benchmark autoregressive model that uses 12 lags of the first differences (see Section 8 of the main text for details). The best method overall is highlighted in bold.

Table 2. Best forecasting methods by group.

		Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
h	1	AGA	AGA	AGA	AGA	APA	AGA	AGA	AGA
	3	AGA	AGA	APES	AGA	CGA	APES	APA	AGA
	6	AGA	APES	APES	AGA	CPA	AGA	MGA	AGA
	12	APES	AGA	AGA APES	AGA	CGA	AGA	AGA APA	AGA
GA > PA	3	3	1	4	1	3	2	4	
PA > GA	0	0	0	0	2	0	1	0	
AVE > SEL	3	3	1	4	4	3	4	4	
SEL > AVE	1	1	2	0	0	1	0	0	
M > (CVh,APE)	0	0	0	0	0	0	0	0	
CVh > (APE,M)	0	0	0	0	3	0	0	0	
APE > (M,CVh)	4	4	4	4	1	4	3	4	

Note: The groups are defined as in [McCracken and Ng \(2016\)](#): (1) output and income; (2) labor market; (3) housing; (4) consumption, orders, and inventories; (5) money and credits; (6) interest and exchange rates; (7) prices; (8) stock market. The last 7 rows counts excluded pairwise ties.

Table 3. Relative MSFE of core real macroeconomic time series.

Method	Ind. Prod.	Per. Income	M&T Sales	Nonag. Emp.	Ind. Prod.	Per. Income	M&T Sales	Nonag. Emp.
h			1				6	
MPA	0.971 **	0.964	0.983	0.956 ***	0.978	0.991	0.985	0.952 *
MGA	0.965 ***	0.959 *	0.970 **	0.948 ***	0.945	0.949	0.934	0.918 ***
CPA	0.967 **	0.919 *	0.983	0.961 ***	0.993	0.999	1.009	0.961 *
CGA	0.957 ***	0.908 *	0.968 **	0.949 ***	0.952	0.968	0.953	0.936 **
APA	1.040	0.930	0.995	1.262 ***	1.122	1.033	1.043	1.360 ***
AGA	0.947 ***	0.889 **	0.956 **	0.923 ***	0.974	0.891 *	0.916 *	0.925 *
MS	0.984	0.992	0.992	0.946 ***	1.009	1.021	0.983	0.939 **
CVhS	0.985	0.946	0.986	0.945 ***	1.007	1.003	1.037	0.953 **
APES	0.972	0.905 *	0.973	0.922 ***	0.981	0.918	0.915	0.911 *
h			3				12	
MPA	0.976	0.979	0.986	0.955 **	0.970	0.972	0.995	0.968
MGA	0.962	0.961	0.957	0.937 ***	0.906 ***	0.905 **	0.886 **	0.895 ***
CPA	0.987	0.978	0.997	0.963 **	1.027	1.007	1.023	1.005
CGA	0.972	0.957 *	0.965	0.950 ***	0.852 *	0.957	0.848	0.912 *
APA	1.121 *	0.997	1.031	1.419 ***	1.173	1.010	1.023	1.305 ***
AGA	0.970	0.922 **	0.955	0.921 **	0.837 **	0.775 **	0.751 **	0.841 **
MS	0.990	1.006	0.983	0.942 ***	0.988	1.001	0.985	0.962
CVhS	0.999	1.004	1.017	0.957 ***	0.914	1.020	0.926	0.966
APES	0.974	0.941	0.969	0.902 ***	0.814 **	0.777 **	0.737 *	0.851 **

Note: * denotes 10%, ** denotes 5%, and *** denotes 1% significance level for a two-sided [Diebold and Mariano \(1995\)](#) test. The benchmark is an unrestricted OLS estimation method with 12 lags (see Section 8 for details). The best method in each case is highlighted in bold.

It is useful to briefly discuss the recent, related literature to place our empirical findings in perspective. [Cheng and Hansen \(2015\)](#) conducted a comparison of several shrinkage-type forecasting approaches using 143 quarterly US macroeconomic time series (transformed to stationarity) from 1960 to 2008. Their methods included factor-augmented forecast combination based on Mallows/cross-validation/equal weights, Bayesian model averaging, empirical Bayes, pretesting and bagging. They found that while the methods were comparable at the one-quarter horizon, cross-validation weighting clearly emerged as the preferred approach at the four-quarter horizon. [Tu and Yi \(2017\)](#) found that, when forecasting US inflation one-quarter ahead, Mallows-based combination forecasts that

combine forecasts from unrestricted and restricted (imposing no error correction) vector autoregressions under the assumption of cointegration dominated both unrestricted and restricted forecasts. Using the same data set as ours, [Kejriwal and Yu \(2021\)](#) compared partial and general combination forecasts using Mallows weights with forecasts based on pretesting and Mallows selection. Consistent with our results, they found that a general combination strategy that averages over both the unit root restriction and different lag orders delivered the best forecasts overall.

Table 4. Relative MSFE of core nominal macroeconomic time series.

Method	CPI	Con. Deflator	CPI e. Food	PPI	CPI	Con. Deflator	CPI e. Food	PPI
h		1				6		
MPA	0.966 *	0.963 **	0.965 *	0.952 **	0.985	0.975	0.970	0.957
MGA	0.957 **	0.957 **	0.961 **	0.944 **	0.964	0.954	0.964	0.950 *
CPA	0.962	0.955 **	0.957 *	0.938 **	0.973	0.984	0.961	0.953
CGA	0.955 *	0.952 **	0.957 *	0.937 **	0.952	0.952	0.943 *	0.942 *
APA	0.958	0.953 **	0.960	0.930 **	0.971	0.966	0.959	0.944 *
AGA	0.948 *	0.949 **	0.949 *	0.941 **	0.949	0.947	0.945	0.955
MS	1.001	0.994	0.999	0.999	1.038	1.009	1.011	0.983
CVhS	0.980	0.979	0.967	0.946 *	0.977	0.992	0.984	0.969
APES	0.989	0.982	0.968	0.939 *	0.967	0.985	0.964	0.968
h		3				12		
MPA	0.969	0.966	0.963	0.946 **	0.972	0.984	0.955 *	0.954 *
MGA	0.957	0.956	0.959	0.938 **	0.948 *	0.961	0.945	0.949 *
CPA	0.957	0.959	0.952	0.933 *	0.955	0.967	0.939 **	0.954
CGA	0.942	0.946	0.945	0.926 **	0.947 *	0.949	0.928 **	0.952 *
APA	0.956	0.951	0.962	0.927 **	0.937 **	0.935	0.920 **	0.958
AGA	0.938	0.936	0.945	0.936 **	0.944	0.939	0.938	0.972
MS	1.029	1.017	0.989	0.991	1.006	1.035	0.998	0.992
CVhS	0.983	0.987	0.981	0.938 *	0.960	0.977	0.946 *	0.956
APES	0.970	0.957	0.974	0.942 *	0.958	0.957	0.968	0.972

Note: Here, * denotes 10% and ** denotes 5% significance level for a two-sided [Diebold and Mariano \(1995\)](#) test. The benchmark is an unrestricted OLS estimation method with 12 lags (see Section 8 for details). The best method in each case is highlighted in bold.

In summary, our empirical results were found to be consistent with the simulation results in that (1) addressing both persistence uncertainty and lag-order uncertainty are crucial for generating accurate forecasts; (2) a weighting scheme that directly targets forecast risk instead of in-sample mean squared error yields an efficacious forecast combination approach at all horizons.

9. Conclusions

This paper has developed new multistep forecast combination methods for a time series driven by stochastic and/or deterministic trends. In contrast to existing methods based on Mallows/cross-validation weighting, our proposed combination forecasts were based on constructing weights obtained from an accumulated prediction errors criterion that directly targets the asymptotic forecast risk instead of the in-sample AMSE. Our analysis found strong evidence in favor of a version of the proposed approach that simultaneously addresses stochastic trend and lag order uncertainty. A practical implication of our results is that the degree of persistence in a time series can play an important role in the choice of combination weights. Our preferred approach can potentially serve as a useful univariate benchmark when evaluating the effectiveness of methods designed to exploit information in large data sets.

We conclude with a discussion of four possible directions for future research. First, the APE-based combination forecasts can potentially be used in conjunction with FGLS estimation of the deterministic component, given that the latter has been shown to yield

improved forecasts over OLS estimation (Kejriwal and Yu 2021). Second, it may be useful to explore the possibility of allowing for a nonlinear deterministic component through, say, the inclusion of polynomial trends or a few low-frequency trigonometric components (Gallant 1981). To the extent that the specific nonlinear modeling structure captures the observed nonlinearities, such an approach may contribute to a further improvement in forecasting performance. Third, it would be useful to develop prediction intervals around our combination forecasts in order to quantify the associated sampling uncertainty. Fourth, and perhaps most challenging, while our numerical and empirical analyses clearly document the desirability of the proposed approach based on APE weighting relative to Mallows/CV weighting, an analytical comparison may shed further light on the relative merits of the different methods. To our knowledge, such results are primarily available in the context of the standard stationary framework with Mallows/cross-validation weighting (e.g., Hansen 2007; Zhang et al. 2013; Liao and Tsay 2020). Extending these results to the present nonstationary framework would be a potentially fruitful endeavor.

Author Contributions: Conceptualization, M.K., L.N. and X.Y.; methodology, M.K., L.N. and X.Y.; Software, L.N. and X.Y.; validation, M.K., L.N. and X.Y.; formal analysis, M.K., L.N. and X.Y.; investigation, M.K., L.N. and X.Y.; resources, M.K., L.N. and X.Y.; data curation, L.N. and X.Y.; writing—original draft preparation, M.K., L.N. and X.Y.; writing—review and editing, M.K., L.N. and X.Y.; visualization, L.N. and X.Y.; supervision, M.K.; project administration, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: Yu is supported by Shanghai Sailing Program No. 23YF1402000 and National Natural Science Foundation of China (72303040, 72192845).

Data Availability Statement: The data set is publicly available for download at <https://research.stlouisfed.org/econ/mccracken/fred-databases/> (accessed on 1 February 2023). Simulation data are available upon request.

Acknowledgments: The authors wish to thank the three anonymous referees for their constructive feedback which helped improve the paper. They also gratefully acknowledge many useful discussions from participants in the Purdue University seminar and the Midwest Econometrics conference.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

Let $W(\cdot)$ denote a standard Brownian motion on $[0, 1]$ and define the Ornstein-Uhlenbeck process: $dW_c(r) = cW_c(r) + dW(r)$. For $p \in \{0, 1\}$, let $X_c(r) = (r^p, W_c(r))'$ and define the detrended processes

$$W_c^*(r, p) = \begin{cases} W_c(r) & \text{if } p = 0 \\ W_c(r) - \int_0^1 W_c(s) ds & \text{if } p = 1 \end{cases}$$

$$X_c^*(r, p) = \begin{cases} X_c(r) & \text{if } p = 0 \\ X_c(r) - \int_0^1 X_c(s) ds & \text{if } p = 1 \end{cases}$$

and the functionals

$$T_{0c} = -cW_c^*(1, p) + I(p = 1)W(1)$$

$$T_{1c} = X_c^*(1, p)' \left(\int_0^1 X_c^*(r, p) X_c^*(r, p)' \right)^{-1} \int_0^1 X_c^*(r, p) dW(r) + I(p = 1)W(1).$$

Let $\beta = (\beta_0, \beta_1)'$, $z_t = (1, t)'$. Without loss of generality, we assume that $\beta_0 = \beta_1 = 0$ in the true data generating process. For a matrix A , $\|A\|^2 = \sup_{\|v\|=1} v' A' A v$ with $\|v\|$ denoting the Euclidean norm for vector v . Unless otherwise defined, for any variable x , we use x^* to denote its demeaned version. For a random quantity δ , we write $\delta = \delta_0 + o_p(\delta_0)$ as $\delta = \delta_0 + s.o.$, where $s.o.$ represents a term of smaller order in probability. For brevity, all

proofs are provided only for the case $p = 1$. The proofs for $p = 0$ are simpler and follow analogous arguments.

We start by noting that if u_t is generated by (1), it has the $AR(k + 1)$ representation $u_t = \sum_{i=1}^{k+1} a_i u_{t-i} + e_t$, where $a_1 = \alpha + \alpha_1$, $a_i = \alpha_i - \alpha_{i-1}$ ($i = 2, \dots, k$), $a_{k+1} = -\alpha_k$. The companion VAR(1) form of the model is expressed as

$$Y_t = B(k)'Y_{t-1} + v_t$$

where

$$\begin{aligned} Y_t &= (1, t + 1, y_t, \dots, y_{t-k})', \quad v_t = (0, 0, e_t, 0, \dots, 0)' \\ & \quad (k+3) \times 1 \qquad \qquad \qquad (k+3) \times 1 \\ B(k) &= \begin{pmatrix} B_1 & B_2 \\ \mathbf{0}_{(k+1) \times 2} & F(k) \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix} \\ & \quad (k+3) \times (k+3) \qquad \quad (2 \times 2) \qquad \quad 2 \times (k+1) \\ F(k) &= \left(a(k) \middle| \begin{matrix} I_k \\ \mathbf{0}'_k \end{matrix} \right), \quad F^0(k) = I_{k+1}, \quad a(k) = (a_1, \dots, a_{k+1})' \\ & \quad (k+1) \times (k+1) \end{aligned}$$

With “hat” and “tilde” denoting the unrestricted and restricted OLS estimates, respectively, the unrestricted and restricted forecasts can then be expressed as (see, e.g., Ing 2003):

$$\hat{\mu}_{T+h} = \mathbf{y}_T(k+1)' \hat{B}^{h-1}(k) \hat{\gamma}(k) \tag{A1}$$

$$\tilde{\mu}_{T+h} = \mathbf{y}_T(k+1)' \tilde{B}^{h-1}(k) \tilde{\gamma}(k) \tag{A2}$$

where $\mathbf{y}_T(k+1) = (1, T + 1, y_T, \dots, y_{T-k})'$, $\hat{\gamma}(k) = (\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{a}_1, \dots, \hat{a}_{k+1})'$ and

$$\begin{aligned} \hat{B}(k) &= \begin{pmatrix} B_1 & \hat{B}_2 \\ \mathbf{0}_{(k+1) \times 2} & \hat{F}(k) \end{pmatrix}, \\ & \quad (k+3) \times (k+3) \\ B_1 &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \hat{B}_2 = \begin{pmatrix} \hat{\beta}_0^* & 0 & \dots & 0 \\ \hat{\beta}_1^* & 0 & \dots & 0 \end{pmatrix} \\ & \quad (2 \times 2) \qquad \quad 2 \times (k+1) \\ \hat{F}(k) &= \left(\hat{a}(k) \middle| \begin{matrix} I_k \\ \mathbf{0}'_k \end{matrix} \right), \quad \hat{F}^0 = I_{k+1}, \quad \hat{a}(k) = (\hat{a}_1, \dots, \hat{a}_{k+1})' \\ & \quad (k+1) \times (k+1) \\ \tilde{B}(k) &= \begin{pmatrix} B_1 & \tilde{B}_2 \\ \mathbf{0}_{(k+1) \times 2} & \tilde{F}(k) \end{pmatrix}, \quad \tilde{B}_2 = \begin{pmatrix} \tilde{\beta}_0^* & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix} \\ & \quad (k+3) \times (k+3) \qquad \quad 2 \times (k+1) \end{aligned} \tag{A3}$$

The matrix $\tilde{F}(k)$ is constructed in the same way as $\hat{F}(k)$ with $\hat{a}(k)$ replaced by $\tilde{a}(k)$, where $\tilde{a}(k) = (\tilde{a}_1, \dots, \tilde{a}_{k+1})' = (1 + \tilde{\alpha}_1, \tilde{\alpha}_2 - \tilde{\alpha}_1, \dots, \tilde{\alpha}_k - \tilde{\alpha}_{k-1}, -\tilde{\alpha}_k)'$ with $\tilde{\gamma}(k) = (\tilde{\beta}_0^*, 0, \tilde{a}_1, \dots, \tilde{a}_{k+1})'$. Next, we state a set of lemmas that will be useful in developing the proofs of the main results. Lemmas A.1–A.4, A.7–A.9 below parallel Lemmas A.1–A.4, B.1–B.3 in Ing et al. (2009) who assumed an exact unit root ($c = 0$). Since the sample moments have the same order whether $c = 0$ or $c < 0$, the proofs of the following lemmas also follow directly those in Ing et al. (2009) and are hence omitted.

Lemma A1. Suppose $\{y_t\}$ satisfies (1) and Assumptions 1 and 2. Then for any $q > 0$,

$$E \|\hat{R}_T^{-1}(k)\|^q = O(1)$$

where

$$\hat{R}_T(k) = T^{-1} D_T(k) \sum_{j=k+1}^{T-1} \mathbf{y}_j(k+1) \mathbf{y}_j(k+1)' D_T(k)'$$

with

$$D_T(k) = \underset{(k+3) \times (k+3)}{\text{diag}}(1, T^{-1}, \bar{D}_T(k)),$$

$$\bar{D}_T(k) = \underset{(k+1) \times (k+1)}{\begin{pmatrix} \frac{1}{\sqrt{T}} & \frac{-\alpha_1}{\sqrt{T}} & \cdots & \cdots & \frac{-\alpha_k}{\sqrt{T}} \\ 1 & -1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \end{pmatrix}}$$

Lemma A2. Suppose $\{y_t\}$ satisfies (1) and Assumptions 1 and 2 and for some $q_1 \geq 2$, $\sup_{-\infty \leq t \leq \infty} E|e_t|^{2q_1} < \infty$. Then for any $0 < q < q_1$,

$$E\|\hat{R}_T^{-1}(k) - \hat{R}_T^{*-1}(k)\|^q = O(T^{-q/2})$$

where

$$\hat{R}_T^*(k) = \underset{(k+3) \times (k+3)}{\text{diag}}(\underset{3 \times 3}{\hat{R}_c^*(k)}, \underset{k \times k}{\hat{\Gamma}_T(k)})$$

$$\hat{R}_c^*(k) = \begin{pmatrix} T^{-1}(T-1-k) & T^{-1} \sum_{j=k+1}^{T-1} X'_t \\ T^{-1} \sum_{j=k+1}^{T-1} X_t & T^{-2} \sum_{j=k+1}^{T-1} X_t X'_t \end{pmatrix}$$

$$X_t = [T^{-1}(t+1), T^{-1/2}N_t]', N_j = A(L)y_j$$

$$\hat{\Gamma}_T(k) = T^{-1} \sum_{j=k+1}^{T-1} s_j(k)s_j(k)', s_j(k) = (\Delta y_j, \dots, \Delta y_{j-k+1})'$$

Lemma A3. Suppose $\{y_t\}$ satisfies (1) and Assumptions (1) and (2) with $\sup_{-\infty \leq t \leq \infty} E|e_t|^q < \infty$ for some $q \geq 2$. Then,

$$E\|T^{-1/2}D_T(k) \sum_{j=k+1}^{T-1} y_j(k+1)e_{j+1}\|^q = O(1)$$

Lemma A4. Suppose $\{y_t\}$ satisfies (1) and Assumptions 1 and 2 with $\sup_{-\infty \leq t \leq \infty} E|e_t|^r < \infty$ for some $r > 4$. Then,

$$\lim_{T \rightarrow \infty} E(F_{T,k}) = 0$$

where

$$F_{T,k} = s_T(k)' M_h(k) \hat{\Gamma}_T^{-1}(k) \left\{ \sum_{j=k+1}^{T-1} s_j(k) e_{j+1} \right\} X'_T \left(\sum_{j=k+1}^{T-1} X_j X'_j \right)^{-1} \left\{ \sum_{j=k+1}^{T-1} X_j e_{j+1} \right\}$$

Lemma A5. Let $X = \underset{T \times (p+1)}{[X_1, X_2]}$, $X_1 = (1, \dots, 1)'$, and assume $X'X$ is invertible. Define $M_1 = \underset{T \times T}{I} - X_1(X_1'X_1)^{-1}X_1'$, $X_2^* = M_1X_2$. For any $T \times 1$ vector e and any $p \times 1$ vector x_2 , we have $x'(X'X)^{-1}X'e = x_1(X_1'X_1)^{-1}X_1'e + x_2^*(X_2^*X_2^*)^{-1}X_2^*e$, where $x = (x_1, x_2)'$, $x_1 = 1$, $x_2^* = x_2 - (X_1'X_1)^{-1}X_2'X_1$.

Lemma A6. Under Assumptions 1 and 2, $\frac{\sqrt{T}\hat{\beta}_0^*}{\sigma} \xrightarrow{d} W_c(1)$.

Lemma A7. Under Assumptions 1 and 2 and $\sup_{-\infty \leq t \leq \infty} E|e_t|^q < \infty$ for some $q > 2$,

- (i) For some $\kappa_1 > 0$, $\|\hat{\Gamma}(k) - \Gamma(k)\| = o(T^{-\kappa_1})$ a.s.;
- (ii) For some $\kappa_2 > 0$, $\|\hat{R}_T - R_T^*\| = o(T^{-\kappa_2})$ a.s.;
- (iii) $\|\hat{R}_T^{-1}\| = O(\log \log T)$ a.s..

Lemma A8. Under Assumptions 1 and 2 and $\sup_{-\infty \leq t \leq \infty} E|e_t|^q < \infty$ for some $q > 2$, $\sum_{i=m_h}^{T-h} F_{i,k} = o(T)$ a.s., where

$$F_{i,k} = s_i(k)' M_h(k) \hat{\Gamma}_i^{-1}(k) \left\{ \sum_{j=k+1}^{i-1} s_j(k) e_{j+1} \right\} X_i' \left(\sum_{j=k+1}^{i-1} X_j X_j' \right)^{-1} \left\{ \sum_{j=k+1}^{i-1} X_j e_{j+1} \right\}$$

Lemma A9. Let $\{x_T\}$ be a sequence of real numbers.

- (i) If $x_T \geq 0$, $T^{-1} \sum_{j=1}^T x_j = O(1)$, and for some $\xi > 1$, $\liminf_{T \rightarrow \infty} v_T / T^\xi > 0$, then, $\sum_{j=1}^T x_j / v_j = O(1)$;
- (ii) If $T^{-1} \sum_{j=1}^T x_j = o(1)$, then, $\sum_{j=1}^T x_j / j = o(\log T)$.

Proof of Lemma A5. Note, by block matrix inversion,

$$\begin{aligned} (X'X)^{-1} &= \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (X_1'X_1)^{-1} + (X_1'X_1)^{-1}X_1'X_2(X_2'M_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1} & -(X_1'X_1)^{-1}X_1'X_2(X_2'M_1X_2)^{-1} \\ -(X_2'M_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1} & (X_2'M_1X_2)^{-1} \end{pmatrix} \end{aligned}$$

then

$$(X'X)^{-1}X'e = \begin{pmatrix} (X_1'X_1)^{-1}X_1'[I - X_2(X_2'M_1X_2)^{-1}X_2'M_1]e \\ (X_2'M_1X_2)^{-1}X_2'M_1e \end{pmatrix}$$

Recall $x = (x_1, x_2')' = (x_1, x_2^{*'})' + (0, X_1'X_2(X_1'X_1)^{-1})'$, we have,

$$\begin{aligned} &x'(X'X)^{-1}X'e \\ &= \underbrace{[(x_1, x_2^{*'})](X'X)^{-1}X'e}_{\text{Term 1}} + \underbrace{[(0, X_1'X_2(X_1'X_1)^{-1})](X'X)^{-1}X'e}_{\text{Term 2}} \\ &= \underbrace{x_1(X_1'X_1)^{-1}X_1'[I - X_2(X_2'M_1X_2)^{-1}X_2'M_1]e}_{\text{Term 1}} + \underbrace{x_2^{*'}(X_2'M_1X_2)^{-1}X_2'M_1e}_{\text{Term 2}} \\ &\quad + \underbrace{X_1'X_2(X_1'X_1)^{-1}(X_2'M_1X_2)^{-1}X_2'M_1e}_{\text{Term 2}} \\ &= x_1(X_1'X_1)^{-1}X_1'e + x_2^{*'}(X_2'M_1X_2)^{-1}X_2'M_1e \\ &\quad - \underbrace{x_1(X_1'X_1)^{-1}X_1'X_2(X_2'M_1X_2)^{-1}X_2'M_1e + X_1'X_2(X_1'X_1)^{-1}(X_2'M_1X_2)^{-1}X_2'M_1e}_{=0, \text{ since } x_1 = 1, (X_1'X_1)^{-1} = 1/T, \text{ which is a constant}} \\ &= x_1(X_1'X_1)^{-1}X_1'e + x_2^{*'}(X_2^{*'}X_2^*)^{-1}X_2^{*'}e \end{aligned}$$

□

Proof of Lemma A6. The true DGP can be expressed as

$$\Delta y_t = \beta_0^* + \sum_{j=1}^k \alpha_j \Delta y_{t-j} + e_t^*$$

where $\beta_0^* = 0$ and $e_t^* = \frac{ac}{T}u_{t-1} + e_t$. Let $\dot{Z}_t = (\dot{Z}_1, \dot{Z}_2, t)$, $\dot{Z}_1 = 1$, $\dot{Z}_{2,t} = (\Delta y_{t-1}, \dots, \Delta y_{t-k})'$, $\iota_1 = (1, 0, \dots, 0)'$, $\iota_{[2:k+1]} = (0, 1, \dots, 1)'$. Now

$$\begin{aligned} \frac{\sqrt{T}\tilde{\beta}_0^*}{\sigma} &= \frac{\sqrt{T}}{\sigma} \iota_1' \left(\sum_{t=k+1}^T \dot{Z}_t \dot{Z}_t' \right)^{-1} \sum_{t=k+1}^T \dot{Z}_t' \left(\frac{ac}{T}u_{t-1} + e_t \right) \tag{A4} \\ &= \frac{\sqrt{T}}{\sigma} \left(\sum_{t=k+1}^T \dot{Z}_1^2 \right)^{-1} \sum_{t=k+1}^T \dot{Z}_1 \left(\frac{ac}{T}u_{t-1} + e_t \right) + o_p(1) \\ &= \frac{ca}{\sigma T^{3/2}} \sum_{t=k+1}^T u_{t-1} + \frac{1}{\sigma\sqrt{T}} \sum_{t=k+1}^T e_t + o_p(1) \xrightarrow{d} c \int_0^1 W_c + W(1) = W_c(1) \end{aligned}$$

□

Proof of Theorem 1. (a) Defining $\gamma(k) = (\beta_0^*, \beta_1^*, a_1, \dots, a_{k+1})'$, $\hat{L}_h(k) = \sum_{j=0}^{h-1} b_j \hat{B}^{h-1-j}(k)$ and $L_h = \sum_{j=0}^{h-1} b_j B^{h-1-j}(k)$, we can write

$$\begin{aligned} &\frac{T}{\sigma^2} E(\hat{\mu}_{T+h} - \mu_{T+h})^2 \\ &= \frac{T}{\sigma^2} E[\mathbf{y}_T(k+1)' \hat{L}_h(k) (\hat{\gamma}(k) - \gamma(k))]^2 \\ &= \frac{T}{\sigma^2} \left[E[\mathbf{y}_T(k+1)' L_h(k) (\hat{\gamma}(k) - \gamma(k))]^2 \right. \\ &\quad \left. + E[\mathbf{y}_T(k+1)' \{ \hat{L}_h(k) - L_h(k) \} (\hat{\gamma}(k) - \gamma(k))]^2 + o(1) \right] \\ &= \frac{1}{\sigma^2} E \left[\mathbf{y}_T(k+1)' L_h(k) D_T'(k) (\hat{R}_T^{*-1}(k)) \frac{D_T(k)}{\sqrt{T}} \sum_{j=k+1}^{T-1} \mathbf{y}_j(k+1) e_{j+1} \right]^2 \\ &\quad + \frac{1}{\sigma^2} E \left[\mathbf{y}_T(k+1)' L_h(k) D_T'(k) (\hat{R}_T^{-1}(k) - \hat{R}_T^{*-1}(k)) \frac{D_T(k)}{\sqrt{T}} \sum_{j=k+1}^{T-1} \mathbf{y}_j(k+1) e_{j+1} \right]^2 \\ &\quad + \frac{T}{\sigma^2} E \left[\mathbf{y}_T(k+1)' \{ \hat{L}_h(k) - L_h(k) \} (\hat{\gamma}(k) - \gamma(k)) \right]^2 + o(1) \\ &= (I) + (II) + (III) \tag{A5} \end{aligned}$$

The (II) and (III) terms in (A5) are each $o(1)$ by Lemmas A1–A3 and Holder’s inequality [see, e.g., the proof of Theorem 2.2 in Ing et al. 2009].

The term (I) can be written as:

$$\begin{aligned} &\frac{1}{\sigma^2} E[\mathbf{y}_T(k+1)' L_h(k) D_T'(k) \hat{R}_T^{*-1}(k) \frac{D_T(k)}{\sqrt{T}} \sum_{j=k}^{T-1} \mathbf{y}_j(k+1) e_{j+1}]^2 \\ &= \frac{1}{\sigma^2} E[\mathbf{y}_T(k+1)' D_T'(k) \bar{L}_h(k) \hat{R}_T^{*-1}(k) \frac{D_T(k)}{\sqrt{T}} \sum_{j=k}^{T-1} \mathbf{y}_j(k+1) e_{j+1}]^2 \tag{A6} \end{aligned}$$

where $\bar{L}_h(k) = \sum_{j=0}^{h-1} b_j \text{diag}(G_T^{h-1-j}, \bar{F}(k)^{h-1-j})$ with $G_T = \begin{pmatrix} 1 & T^{-1} \\ 0 & 1 \end{pmatrix}$, $\bar{F}(k) = \text{diag}(1, S_M(k))$

and $S_M(k) = \begin{pmatrix} \alpha(k-1) & I_{k-1} \\ \alpha_k & \mathbf{0}'_{k-1} \end{pmatrix}$, $S_M^0(k) = I_k$.

Note that $\mathbf{y}_T(k+1)' D_T' = (1, T^{-1}(T+1), T^{-1/2}N_T, s_T(k))$. Further, since G_T is upper triangular, (A6) converges to

$$\begin{aligned} & \frac{1}{\sigma^2} \left(\sum_{j=0}^{h-1} b_j \right)^2 \lim_{T \rightarrow \infty} E \left\{ T^{-1/2} \sum_{j=k+1}^{T-1} e_{j+1} + X_T^{*'} \left(\sum_{j=k+1}^{T-1} X_j^* X_j^{*'} \right)^{-1} \sum_{j=k+1}^{T-1} X_j^* e_{j+1} \right\}^2 \\ & + \lim_{T \rightarrow \infty} \frac{1}{\sigma^2} E \left\{ s_T'(k) M_h(k) \hat{\Gamma}_T^{-1}(k) T^{-1/2} \sum_{j=k+1}^{T-1} s_j(k) e_{j+1} \right\}^2 + \frac{2}{\sigma^2} \left(\sum_{j=0}^{h-1} b_j \right) \lim_{T \rightarrow \infty} E(F_{T,k}) \\ & = B.1 + B.2 + B.3 \end{aligned} \tag{A7}$$

where B.1 utilizes Lemma A.5. Since B.2 = $g_h(k)$ by Theorem 1 of Ing (2003) and B.3 = 0 by Lemma A4, (A7) simplifies to:

$$\begin{aligned} B.1 + B.2 &= \left(\sum_{j=0}^{h-1} b_j \right)^2 \lim_{T \rightarrow \infty} E \left\{ W(1) + X_c^*(1)' \left(\int_0^1 X_c^* X_c^{*'} \right)^{-1} \int_0^1 X_c^* dW \right\}^2 + g_h(k) \\ &= \left(\sum_{j=0}^{h-1} b_j \right)^2 E[T_{1c}^2] + g_h(k) \end{aligned} \tag{A8}$$

The required result then follows from (A5), (A7) and (A8).

(b) Defining $\tilde{L}_h(k) = \sum_{j=0}^{h-1} b_j \tilde{B}^{h-1-j}(k)$, with similar arguments as in (a), we can write:

$$\begin{aligned} \frac{T}{\sigma^2} E(\tilde{\mu}_{T+h} - \mu_{T+h})^2 &= \frac{T}{\sigma^2} E \left[\mathbf{y}_T(k+1)' \tilde{L}_h(k) (\tilde{\gamma}(k) - \gamma(k)) \right]^2 \\ &= \frac{T}{\sigma^2} E \left[\mathbf{y}_T(k+1)' L_h(k) (\tilde{\gamma}(k) - \gamma(k)) \right]^2 + o(1) \end{aligned} \tag{A9}$$

Note that

$$L_h(k) = \sum_{j=0}^{h-1} b_j B^{h-1-j}(k) = \sum_{j=0}^{h-1} b_j \begin{pmatrix} B_1 & 0 \\ 0 & F(k) \end{pmatrix}^{h-1-j} = \sum_{j=0}^{h-1} b_j \begin{pmatrix} B_1^{h-1-j} & 0 \\ 0 & F^{h-1-j}(k) \end{pmatrix}$$

Since B_1 is upper triangular with $B_1(1, 1) = 1$,

$$\begin{aligned} (A9) &= \frac{T}{\sigma^2} E \left[\mathbf{y}_T(k+1)' \begin{pmatrix} \sum_{j=0}^{h-1} b_j \begin{bmatrix} \tilde{\beta}_0^* \\ 0 \end{bmatrix} \\ \sum_{j=0}^{h-1} b_j F^{h-1-j}(k) [\tilde{a}(k) - a(k)] \end{pmatrix} \right]^2 + o(1) \\ &= \frac{T}{\sigma^2} E \left[\left(\sum_{j=0}^{h-1} b_j \right) \tilde{\beta}_0^* + (y_T, \dots, y_{T-k}) \sum_{j=0}^{h-1} b_j F^{h-1-j}(k) [\tilde{a}(k) - a(k)] \right]^2 + o(1) \end{aligned} \tag{A10}$$

Now, consider the term

$$\begin{aligned} & \frac{\sqrt{T}}{\sigma} (y_T, \dots, y_{T-k}) \sum_{j=0}^{h-1} b_j F^{h-1-j}(k) [\tilde{a}(k) - a(k)] = \frac{\sqrt{T}}{\sigma} (y_T, \dots, y_{T-k}) L_h^{(F)}(k) [\tilde{a}(k) - a(k)] \\ &= \frac{\sqrt{T}}{\sigma} (y_T, \dots, y_{T-k}) L_h^{(F)}(k) \left[\hat{a}(k) - a(k) + H_k D_T'(k) \hat{R}_T^{-1}(k) D_T(k) R_k'(k) (R_k D_T'(k) \hat{R}_T^{-1}(k) D_T(k) R_k')^{-1} (r - R_k \hat{\gamma}(k)) \right] \end{aligned} \tag{A11}$$

where

$$L_h^{(F)}(k) = \sum_{j=0}^{h-1} b_j F(k)^{h-1-j}, \quad H_k = \begin{bmatrix} \mathbf{0}_{(k+1) \times 2} & I_{(k+1)} \end{bmatrix}, \quad R_k = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 1 \end{bmatrix}, \quad r = (0, 1)'$$

Next, define $\tilde{L}_h^{(F)}(k) = \text{diag}(\sum_{j=0}^{h-1} b_j, M_h(k))$, $\hat{\theta}(k) = (T(1 - \hat{\alpha})/a, 0, \dots, 0)'$, we have

$$\begin{aligned}
 (A11) &= \frac{1}{\sigma} (y_T, \dots, y_{T-k}) \left[\sqrt{T} L_h^{(F)}(k) \{ \hat{a}(k) - a(k) \} + L_h^{(F)}(k) \bar{D}'_T(k) \hat{\theta}(k) \right] \\
 &= \frac{1}{\sigma} (y_T, \dots, y_{T-k}) \left[\sqrt{T} L_h^{(F)}(k) \{ \hat{a}(k) - a(k) \} + \bar{D}'_T(k) L_h^{(F)}(k) \hat{\theta}(k) \right] \\
 &= \frac{1}{\sigma} (y_T, \dots, y_{T-k}) \sqrt{T} L_h^{(F)}(k) (\hat{a}(k) - a(k)) + \frac{1}{\sigma} (N_T / \sqrt{T}, s_T(k)) \text{diag} \left(\sum_{j=0}^{h-1} b_j, M_h(k) \right) \hat{\theta}(k) \\
 &= \frac{1}{\sigma} (y_T, \dots, y_{T-k}) \sqrt{T} L_h^{(F)}(k) (\hat{a}(k) - a(k)) + \frac{N_T}{\sigma \sqrt{T}} \sum_{j=0}^{h-1} b_j [-c - T(\hat{\alpha} - \alpha) / a] \\
 &= \frac{N_T}{\sigma \sqrt{T}} \sum_{j=0}^{h-1} b_j \{ T(\hat{\alpha} - \alpha) / a \} + \frac{1}{\sigma} s_T(k)' M_h(k) \hat{\Gamma}_T^{-1}(k) T^{-1/2} \sum_{j=k}^{T-1} s_j(k) e_{j+1} \\
 &\quad + \frac{N_T}{\sigma \sqrt{T}} \sum_{j=0}^{h-1} b_j [-c - T(\hat{\alpha} - \alpha) / a] \\
 &= -c \frac{N_T}{\sigma \sqrt{T}} \sum_{j=0}^{h-1} b_j + \frac{1}{\sigma} s_T(k)' M_h(k) \hat{\Gamma}_T^{-1}(k) T^{-1/2} \sum_{j=k}^{T-1} s_j(k) e_{j+1} \tag{A12}
 \end{aligned}$$

Then, combining (A10) with (A12) and using Lemma A6, we finally get

$$\begin{aligned}
 \lim_{T \rightarrow \infty} \frac{T}{\sigma^2} E(\tilde{\mu}_{T+h} - \mu_{T+h})^2 &= E \left[\sum_{j=0}^{h-1} b_j (W_c(1) - cW_c(1)) \right]^2 \\
 &\quad + \frac{1}{\sigma^2} \lim_{T \rightarrow \infty} E \left\{ s_T^*(k) M_h(k) \hat{\Gamma}_T^{-1}(k) T^{-1/2} \sum_{j=k}^{T-1} s_j(k) e_{j+1} \right\}^2 \\
 &= \left(\sum_{j=0}^{h-1} b_j \right)^2 E [T_{0c}^2] + g_h(k)
 \end{aligned}$$

which uses the fact that $W_c(1) - cW_c(1) = W(1) - cW_c^*(1)$, thereby proving the result. \square

Proof of Theorem 2. Henceforth, estimated parameters and quantities with subscript i denotes the estimates using observations from 1 to i . We prove (a) first. It follows from Chow (1965) and Ing (2004) that

$$APE_1 - \sum_{i=m_h}^{T-h} \eta_{i,h}^2 = \sum_{i=m_h}^{T-h} \left[y_i'(k+1) \hat{L}_{i,h}(k) (\hat{\gamma}_i(k) - \gamma(k)) \right]^2 (1 + o(1)) + O(1) \text{ a.s.}$$

Using similar algebra as in Theorem 1, we have:

$$\begin{aligned}
 & \sum_{i=m_h}^{T-h} \left[\mathbf{y}'_i(k+1) \hat{L}_{i,h}(k) (\hat{\gamma}_i(k) - \gamma(k)) \right]^2 \\
 &= \sum_{i=m_h}^{T-h} \left[\left[\mathbf{y}'_i(k+1)' L_h(k) (\hat{\gamma}_i(k) - \gamma(k)) \right]^2 \right. \\
 & \quad \left. + \left[\mathbf{y}'_i(k+1)' \{ \hat{L}_{i,h}(k) - L_h(k) \} (\hat{\gamma}_i(k) - \gamma(k)) \right]^2 \right] + s.o. \\
 &= \sum_{i=m_h}^{T-h} \frac{1}{i} \left[\mathbf{y}'_i(k+1)' L_h(k) D'_i(k) (\hat{R}_i^{*-1}(k)) \frac{D_i(k)}{\sqrt{i}} \sum_{j=k+1}^{i-1} \mathbf{y}_j(k+1) e_{j+1} \right]^2 \\
 & \quad + \sum_{i=m_h}^{T-h} \frac{1}{i} \left[\mathbf{y}'_i(k+1)' L_h(k) D'_i(k) (\hat{R}_i^{-1}(k) - \hat{R}_i^{*-1}(k)) \frac{D_i(k)}{\sqrt{i}} \sum_{j=k+1}^{i-1} \mathbf{y}_j(k+1) e_{j+1} \right]^2 \\
 & \quad + \sum_{i=m_h}^{T-h} \left[\mathbf{y}'_i(k+1)' \{ \hat{L}_{i,h}(k) - L_h(k) \} (\hat{\gamma}_i(k) - \gamma(k)) \right]^2 + s.o. \\
 &= (IV) + (V) + (VI) \tag{A13}
 \end{aligned}$$

The (V) and (VI) terms in (A13) are each $O(1)$ following similar arguments in Ing et al. (2009) which build on Lemmas A7–A9.

Analogous to (A6) and (A7) in the proof of Theorem 1, (IV) can be rewritten as:

$$\begin{aligned}
 (IV) &= \left(\sum_{j=0}^{h-1} b_j \right)^2 \sum_{i=m_h}^{T-h} \left\{ Z'_i \left(\sum_{j=k+1}^{i-1} Z_j Z'_j \right)^{-1} \sum_{j=k+1}^{i-1} Z_j e_{j+1} \right\}^2 \\
 & \quad + \sum_{i=m_h}^{T-h} \left\{ s'_i(k) M_h(k) \hat{\Gamma}_i^{-1}(k) \frac{1}{i} \sum_{j=k+1}^{i-1} s_j(k) e_{j+1} \right\}^2 + 2 \left(\sum_{j=0}^{h-1} b_j \right) \sum_{i=m_h}^{T-h} \frac{1}{i} F_{i,k} \\
 &= C.1 + C.2 + C.3
 \end{aligned}$$

where $Z_j = (1, t + 1, N_j)'$. In analogy with Theorem 3.1 of Ing (2004),

$$C.2 = g_h(k) \sigma^2 \log T + o_p(\log T) \tag{A14}$$

By Lemmas A8 and A9, $C.3 = o_p(\log T)$. Now we focus on C.1. By Theorem 4 of Wei (1987), we have

$$C.1 = \left(\sum_{j=0}^{h-1} b_j \right)^2 \sigma^2 \log \det \left(\sum_{j=k+1}^{T-1} Z_j Z'_j \right) + o_p(\log T)$$

Defining the 3×3 matrix $Y_T = \text{diag}(T, T^3, T^2/|c|)$ and using Lemma A of Phillips (2014) in conjunction with the fact that $|c|T^{-2} = O(T^{-1})$, we can calculate

$$\begin{aligned}
 \log \det \left(\sum_{j=k+1}^{T-1} Z_j Z'_j \right) &= \log \det \left(Y_T^{1/2} Y_T^{-1/2} \sum_{j=k+1}^{T-1} Z_j Z'_j Y_T^{-1/2} Y_T^{1/2} \right) \\
 &= \log \det(Y_T) + O_p(1) = \log(T^5) + O_p(1) \\
 &= 5 \log(T) + O_p(1) \tag{A15}
 \end{aligned}$$

which leads to $C.1 = 5\sigma^2 \left(\sum_{j=0}^{h-1} b_j \right)^2 \log(T) + o_p(\log T)$. Thus,

$$\lim_{T \rightarrow \infty} \frac{1}{\sigma^2 \log T} (APE_1 - \sum_{i=m_h}^{T-h} \eta_{i,h}^2) = 5 \left(\sum_{j=0}^{h-1} b_j \right)^2 + g_h(k) \tag{A16}$$

where the right hand side of (A16) is the limit of $f_1(c, p, k, h) = f_1(c, p, h) + g_h(k)$ as $c \rightarrow -\infty$. We next prove (b). Following similar steps as in the proof of (a) and the proof of Theorem 1 for the restricted case, we can derive

$$APE_0 - \sum_{i=m_h}^{T-h} \eta_{i,h}^2 = \left(\sum_{j=0}^{h-1} b_j \right)^2 \sum_{i=m_h}^{T-h} \left(\tilde{\beta}_{0,i}^* - c \frac{N_i}{i} \right)^2 + \sum_{i=m_h}^{T-h} \left\{ \mathbf{s}'_i(k) M_h(k) \hat{\Gamma}_i^{-1}(k) \frac{1}{i} \sum_{j=k+1}^{i-1} \mathbf{s}_j(k) e_{j+1} \right\}^2 + o_p(\log T) = D.1 + D.2$$

In view of (A4), taking the limit $c \rightarrow 0$, we have

$$\begin{aligned} \sum_{i=m_h}^{T-h} \left(\tilde{\beta}_{0,i}^* - c \frac{N_i}{i} \right)^2 &= \sum_{i=m_h}^{T-h} \left[\mathbf{1}'_i \left(\sum_{t=k+1}^i \dot{Z}_t \dot{Z}'_t \right)^{-1} \sum_{t=k+1}^i \dot{Z}'_t e_t \right]^2 \\ &= \sum_{i=m_h}^{T-h} \left[\left(\sum_{t=k+1}^i \dot{Z}'_1 \right)^{-1} \sum_{t=k+1}^i \dot{Z}_1 e_t \right]^2 + s.o. \\ &= \log \det \left(\sum_{j=k+1}^{T-1} \dot{Z}'_1 \right) + o_p(\log T) = \sigma^2 \log T + o_p(\log T) \end{aligned}$$

Further, using the same argument as in (A14), we have $D.2 = g_h(k) \sigma^2 \log T + o_p(\log T)$. Thus,

$$\lim_{c \rightarrow 0} \lim_{T \rightarrow \infty} \frac{1}{\sigma^2 \log T} (APE_0 - \sum_{i=m_h}^{T-h} \eta_{i,h}^2) = \left(\sum_{j=0}^{h-1} b_j \right)^2 + g_h(k) \tag{A17}$$

where the right hand side of (A17) is the limit of $f_0(c, p, k, h) = f_0(c, p, h) + g_h(k)$ as $c \rightarrow 0$ since $\lim_{c \rightarrow 0} E(T_{0c}^2) = E[W(1)^2] = 1$. \square

For the proofs of Theorems 3 and 4, we focus on the misspecified case where $l < k$. For the case $l \geq k$, the proofs follow directly from the arguments in Theorems 1 and 2 above and those in Ing et al. (2009).

Proof of Theorem 3. First, note that under Assumption 3, Lemmas A1–A4 continue to hold with k replaced by l ($l < k$), and e_t replaced by $\varepsilon_t = e_t + \omega_t$, where $\omega_t = \sum_{j=l+1}^k \alpha_j \Delta y_{t-j}$. Define $L_h^*(l) = \sum_{j=0}^{h-1} b_j [B^*(l)]^{h-1-j}$, where $B^*(l)$ is defined similarly to $B(l)$ except that $F(l)$ is replaced by $F^*(l) = \left(a^*(l) \middle| \begin{matrix} I_l \\ \mathbf{0}'_l \end{matrix} \right)$, $F^{*0}(l) = I_{l+1}$, $a^*(l) = (a_1, \dots, a_l, a_{l+1}^*)'$ and $a_{l+1}^* = -\alpha_l$. Also, let $\gamma^*(l) = (\beta_0^*, \beta_1^*, a_1, \dots, a_l, a_{l+1}^*)'$ and $\alpha^*(l, k) = (\mathbf{0}'_l, \alpha_{l+1}, \dots, \alpha_k)'$. Finally, note that under Assumption 3, $b_j \rightarrow 1$ for all j .

(a) We can write

$$\begin{aligned} \frac{T}{\sigma^2} E(\hat{\mu}_{T+h}(l) - \mu_{T+h})^2 &= \frac{T}{\sigma^2} E[\mathbf{y}_T(l+1)' \hat{L}_h(l) (\hat{\gamma}(l) - \gamma^*(l)) + \mathbf{s}_T(k)' M_h(k) \alpha^*(l, k)]^2 \\ &= E[\hat{f}_{T,h}^2(l, k)] + E[r_{T,h}^2(l, k)] + o(1) \\ &= (I) + (II) + o(1) \end{aligned} \tag{A18}$$

where

$$\begin{aligned} \hat{f}_{T,h}(l, k) &= (\sqrt{T}/\sigma) \mathbf{y}_T(l+1)' \hat{L}_h(l) (\hat{\gamma}(l) - \gamma^*(l)) \\ r_{T,h}(l, k) &= (\sqrt{T}/\sigma) \mathbf{s}_T(k)' M_h(k) \alpha^*(l, k) \end{aligned}$$

We now derive the limit of the terms (I) and (II) in (A18). First, consider the term (I) in (A18). Noting that the effective errors are now $\{\varepsilon_t\}$ instead of $\{e_t\}$, we can write, similar to (A5),

$$\begin{aligned}
 E[\hat{f}_{T,h}^2(l,k)] &= \frac{1}{\sigma^2} E \left[\mathbf{y}_T(l+1)' L_h^*(l) D_T'(l) (\hat{R}_T^{*-1}(l)) \frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) \varepsilon_{j+1} \right]^2 \\
 &+ \frac{1}{\sigma^2} E \left[\mathbf{y}_T(l+1)' L_h^*(l) D_T'(l) (\hat{R}_T^{-1}(l) - \hat{R}_T^{*-1}(l)) \frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) \varepsilon_{j+1} \right]^2 \\
 &+ \frac{T}{\sigma^2} E [\mathbf{y}_T(l+1)' \{ \hat{L}_h(l) - L_h^*(l) \} (\hat{\gamma}(l) - \gamma^*(l))]^2 + o(1) \\
 &= (T1) + (T2) + (T3) + o(1)
 \end{aligned} \tag{A19}$$

The terms (T2) and (T3) are each $o(1)$ by the fact that Lemmas A1–A3 hold when $\{e_t\}$ is replaced by $\{\varepsilon_t\}$. Now consider the following term appearing in term (T1):

$$\begin{aligned}
 &\frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) \varepsilon_{j+1} \\
 &= \frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) e_{j+1} + \frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) \omega_{j+1} \\
 &= \frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) e_{j+1} + \frac{1}{T} \sum_{i=l+1}^k \delta_i \sum_{j=l+1}^T D_T(l) \mathbf{y}_j(l+1) \Delta y_{j-i+1} \\
 &= \frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) e_{j+1} + \frac{1}{\sqrt{T}} \sum_{i=l+1}^k \delta_i \frac{1}{\sqrt{T}} \sum_{j=l+1}^T D_T(l) \mathbf{y}_j(l+1) e_{j-i+1} + o_p(1) \\
 &= \frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) e_{j+1} + \frac{1}{\sqrt{T}} \left(\sum_{i=l+1}^k \delta_i \right) O_p(1) + o_p(1) \\
 &= \frac{D_T(l)}{\sqrt{T}} \sum_{j=l+1}^{T-1} \mathbf{y}_j(l+1) e_{j+1} + o_p(1)
 \end{aligned} \tag{A20}$$

where the third equality follows from Assumption 3. Then, substituting (A20) in (A19) and following the same arguments used in deriving (A8), we get

$$\lim_{T \rightarrow \infty} E[\hat{f}_{T,h}^2(l,k)] = \lim_{T \rightarrow \infty} (T1) = h^2 E(T_{1c}^2) + g_h(l) \tag{A21}$$

Now, consider the term (II) in (A18). Defining $\xi_h(\delta, l, k) = M_h(k) (\mathbf{0}'_l, \delta_{l+1}, \dots, \delta_k)'$, we can write

$$\begin{aligned}
 E[r_{T,h}^2(l,k)] &= \frac{1}{\sigma^2} \left[\sqrt{T} \alpha^*(l,k)' M_h(k)' E \{ \mathbf{s}_T(k) \mathbf{s}_T(k)' \} M_h(k) \sqrt{T} \alpha^*(l,k) \right] \\
 &= \frac{1}{\sigma^2} \left[\sqrt{T} \alpha^*(l,k)' M_h(k)' \sigma^2 I_k M_h(k) \sqrt{T} \alpha^*(l,k) \right] + o(1) \\
 &\rightarrow \xi_h(\delta, l, k)' \xi_h(\delta, l, k)
 \end{aligned} \tag{A22}$$

where the second equality in (A22) follows from the facts that under Assumption 3, $\sqrt{T} \alpha^*(l,k) = (\mathbf{0}'_l, \delta_{l+1}, \dots, \delta_k)'$ and $E \{ \mathbf{s}_T(k) \mathbf{s}_T(k)' \} \rightarrow \sigma^2 I_k$. Finally, substituting (A21) and (A22) in (A18), the result follows.

(b) We can write

$$\begin{aligned}
 \frac{T}{\sigma^2} E(\tilde{\mu}_{T+h}(l) - \mu_{T+h})^2 &= \frac{T}{\sigma^2} E [\mathbf{y}_T(l+1)' \tilde{L}_h(l) (\tilde{\gamma}(l) - \gamma^*(l)) + \mathbf{s}_T(k)' M_h(k) \alpha^*(l,k)]^2 \\
 &= E[\hat{f}_{T,h}^2(l,k)] + E[r_{T,h}^2(l,k)] + o(1) \\
 &= (I) + (II) + o(1)
 \end{aligned} \tag{A23}$$

where $\tilde{f}_{T,h}(l, k) = (\sqrt{T}/\sigma)\mathbf{y}_T(l+1)'\tilde{L}_h(l)(\tilde{\gamma}(l) - \gamma^*(l))$ and $r_{T,h}(l, k)$ is as defined in (A18). The limit of term (II) is derived in (A22). To obtain the limit of (I), we follow the same steps as in the proof of Theorem 1(b) with k replaced by l and use (A20) to get

$$E[\tilde{f}_{T,h}^2(l, k)] \rightarrow h^2E(T_{0c}^2) + g_h(l) \tag{A24}$$

The result then follows by substituting (A24) and (A22) in (A23). \square

Proof of Theorem 4. (a) For each $i \in \{m_h, T - h\}$, we can write

$$y_{i+h} - \hat{\mu}_{i+h}(l) = \eta_{i,h} - \underbrace{\mathbf{y}'_i(l+1)\hat{L}_{i,h}(l)(\hat{\gamma}_i^*(l) - \gamma^*(l))}_{\hat{f}_{i,h}(l,k)} + \underbrace{\mathbf{s}'_i(k)M_h(k)\boldsymbol{\alpha}^*(l,k)}_{r_{i,h}(l,k)}$$

We then have

$$\begin{aligned} APE_1(l) &= \sum_{i=m_h}^{T-h} \left[\eta_{i,h} - \hat{f}_{i,h}(l, k) + r_{i,h}(l, k) \right]^2 \\ &= \sum_{i=m_h}^{T-h} \eta_{i,h}^2 + \sum_{i=m_h}^{T-h} \hat{f}_{i,h}^2(l, k)[1 + o_p(1)] + \sum_{i=m_h}^{T-h} r_{i,h}^2(l, k) + 2 \sum_{i=m_h}^{T-h} \eta_{i,h}r_{i,h}(l, k) \\ &\quad - 2 \sum_{i=m_h}^{T-h} \hat{f}_{i,h}(l, k)r_{i,h}(l, k) + O_p(1) \end{aligned} \tag{A25}$$

Note that $\sum_{i=m_h}^{T-h} \eta_{i,h}r_{i,h}(l, k) = \left[T^{-1/2} \sum_{i=m_h}^{T-h} \eta_{i,h}\mathbf{s}'_i(k) \right] \left[M_h(k)T^{1/2}\boldsymbol{\alpha}^*(l, k) \right] = O_p(1) \cdot O(1) = O_p(1)$. We will now consider $\sum_{i=m_h}^{T-h} \hat{f}_{i,h}^2(l, k)$, $\sum_{i=m_h}^{T-h} r_{i,h}^2(l, k)$, and $\sum_{i=m_h}^{T-h} \hat{f}_{i,h}(l, k)r_{i,h}(l, k)$ in turn.

$$\begin{aligned} \sum_{i=m_h}^{T-h} \hat{f}_{i,h}^2(l, k) &= \sum_{i=m_h}^{T-h} \frac{1}{i} \left[\mathbf{y}'_i(l+1)L_h(l)D'_i(l)\hat{R}_i^{*-1}(l) \frac{D_i(l)}{\sqrt{i}} \sum_{j=1}^{i-1} \mathbf{y}_j(l+1)\varepsilon_{j+1} \right]^2 + O_p(1) \\ &= \sum_{i=m_h}^{T-h} \frac{1}{i} \left[\mathbf{y}'_i(l+1)L_h(l)D'_i(l)\hat{R}_i^{*-1}(l) \frac{D_i(l)}{\sqrt{i}} \sum_{j=1}^{i-1} \mathbf{y}_j(l+1)e_{j+1} \right]^2 + O_p(1) \\ &= 5\sigma^2h^2 \log T + g_h(l)\sigma^2 \log T + o_p(\log T). \end{aligned} \tag{A26}$$

where the first equality in (A26) follows in analogy with (A13), the second follows in analogy with (A20), and the third follows from the same arguments used to derive (A14) and (A15) as well as the fact that $b_j \rightarrow 1$.

Next, consider the term

$$\begin{aligned} \sum_{i=m_h}^{T-h} r_{i,h}^2(l, k) &= \sum_{i=m_h}^{T-h} \left[\boldsymbol{\alpha}^{*l}(l, k)M'_h(k)\mathbf{s}_i(k)\mathbf{s}'_i(k)M_h(k)\boldsymbol{\alpha}^*(l, k) \right] \\ &= \zeta_h(\delta, l, k)' \left[T^{-1} \sum_{i=m_h}^{T-h} \mathbf{s}_i(k)\mathbf{s}'_i(k) \right] \zeta_h(\delta, l, k) \\ &= \zeta_h(\delta, l, k)'(\sigma^2I_k)\zeta_h(\delta, l, k) + o_p(1) \\ &\rightarrow \sigma^2\zeta_h(\delta, l, k)'\zeta_h(\delta, l, k) \end{aligned} \tag{A27}$$

where the third equality in (A27) is a consequence of Assumption 3.

Finally, consider the cross-product term

$$\begin{aligned} \sum_{i=m_h}^{T-h} \hat{f}_{i,h}(l,k)r_{i,h}(l,k) &\leq \left[\sum_{i=m_h}^{T-h} \hat{f}_{i,h}^2(l,k) \right]^{1/2} \left[\sum_{i=m_h}^{T-h} r_{i,h}^2(l,k) \right]^{1/2} \\ &= O_p\left(\sqrt{\log T}\right)O_p(1) = o_p(\log T) \end{aligned} \tag{A28}$$

Substituting (A26)–(A28) in (A25), the result follows.

(b) For each $i \in \{m_h, T - h\}$, we can write

$$y_{i+h} - \tilde{\mu}_{i+h}(l) = \eta_{i,h} - \underbrace{y'_i(l+1)\tilde{L}_{i,h}(l)(\hat{\gamma}_i^*(l) - \gamma^*(l))}_{\tilde{f}_{i,h}(l,k)} + \underbrace{s'_i(k)M_h(k)\alpha^*(l,k)}_{r_{i,h}(l,k)}$$

We then have

$$\begin{aligned} APE_0(l) &= \sum_{i=m_h}^{T-h} [\eta_{i,h} - \tilde{f}_{i,h}(l,k) + r_{i,h}(l,k)]^2 \\ &= \sum_{i=m_h}^{T-h} \eta_{i,h}^2 + \sum_{i=m_h}^{T-h} \tilde{f}_{i,h}^2(l,k)[1 + o_p(1)] + \sum_{i=m_h}^{T-h} r_{i,h}^2(l,k) - 2 \sum_{i=m_h}^{T-h} \tilde{f}_{i,h}(l,k)r_{i,h}(l,k) + O_p(1) \end{aligned} \tag{A29}$$

By similar arguments as in (a), we have

$$\begin{aligned} \sum_{i=m_h}^{T-h} \tilde{f}_{i,h}^2(l,k) &= \sigma^2 h^2 \log T + g_h(l)\sigma^2 \log T + o_p(\log T) \\ \sum_{i=m_h}^{T-h} \tilde{f}_{i,h}(l,k)r_{i,h}(l,k) &= o_p(\log T) \end{aligned} \tag{A30}$$

Substituting (A27) and (A30) in (A29), the result follows. □

Appendix B. Description of Methods

This Appendix provides a detailed description of the forecasting methods compared in the Monte Carlo analysis presented in Section 7 and the empirical analysis presented in Section 8.

Unrestricted Autoregressive Model (Benchmark). The benchmark forecast is calculated from a standard autoregressive model of order $K + 1$ estimated by OLS:

$$y_t = \beta_0^* + \beta_1^* t + \alpha y_{t-1} + \sum_{j=1}^K \alpha_j \Delta y_{t-j} + \epsilon_t,$$

Mallows Selection. Hansen (2010a) demonstrated the validity of the Mallows criterion for selecting between the restricted and unrestricted models when $h = 1$. When the number of lags k is known, the criteria for the restricted and unrestricted models are, respectively, given by

$$\begin{aligned} M_0 &= T\tilde{\sigma}^2 + 2\tilde{\sigma}^2(p+k) \\ M_1 &= T\tilde{\sigma}^2 + 2\tilde{\sigma}^2(2+p+k) \end{aligned}$$

where $\tilde{\sigma}^2 = T^{-1} \sum_{t=1}^T (y_t - \tilde{\mu}_t)^2$ and $\sigma^2 = T^{-1} \sum_{t=1}^T (y_t - \hat{\mu}_t)^2$. The Mallows selection estimator picks the restricted model if $M_0 < M_1$ and the unrestricted model otherwise. This is equivalent to picking the unrestricted model when $F_T = T\left(\frac{\tilde{\sigma}^2 - \sigma^2}{\sigma^2}\right) \geq 4$. The Mallows selection forecast can then be expressed as $\hat{\mu}_{t+h,M} = \hat{\mu}_{t+h}\mathbf{1}(F_T \geq 4) + \tilde{\mu}_{t+h}\mathbf{1}(F_T < 4)$. When

the number of lags is unknown, the relevant Mallows criteria are obtained as (see [Kejriwal and Yu 2021](#)):

$$M_0(l) = T\hat{\sigma}_l^2 + 2\hat{\sigma}_K^2(p + l)$$

$$M_1(l) = T\hat{\sigma}_l^2 + 2\hat{\sigma}_K^2(2 + p + l)$$

for $l = 0, 1, \dots, K$, where $\hat{\sigma}_j^2 = T^{-1} \sum_{t=1}^T (y_t - \hat{\mu}_t(j))^2$, $j = l, K$ and $\hat{\sigma}_l^2 = T^{-1} \sum_{t=1}^T (y_t - \tilde{\mu}_t(l))^2$. Then, defining $\tilde{l} = \arg \min_{l \in S} \{M_0(l)\}$, $\hat{l} = \arg \min_{l \in S} \{M_1(l)\}$, where $S = \{0, 1, \dots, K\}$, the Mallows selection forecast is obtained as

$$\check{\mu}_{t+h,M} = \begin{cases} \hat{\mu}_{t+h}(\hat{l}), & \text{if } \min_{l \in S} \{M_1(l)\} \leq \min_{l \in S} \{M_0(l)\} \\ \tilde{\mu}_{t+h}(\tilde{l}), & \text{if } \min_{l \in S} \{M_1(l)\} > \min_{l \in S} \{M_0(l)\} \end{cases}$$

Mallows Averaging. As an alternative to Mallows selection, [Hansen \(2010a\)](#) developed the Mallows combination forecast that entails taking a weighted average of the unrestricted and restricted forecasts where the weights are chosen by minimizing a Mallows criterion. When the number of lags is known, the criterion is

$$M_w = \sum_{t=1}^T (y_t - \hat{\mu}_t(w))^2 + 2\hat{\sigma}^2(2w + p + k) \tag{A31}$$

with $\hat{\mu}_t(w) = w\hat{\mu}_t + (1 - w)\tilde{\mu}_t$ and $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T (y_t - \hat{\mu}_t)^2$. The Mallows selected weight \hat{w} is derived from minimizing (A31) over $w \in [0, 1]$. The solution is

$$\hat{w} = \begin{cases} 1 - 2/F_T & \text{if } F_T > 2 \\ 0 & \text{otherwise} \end{cases}$$

The Mallows averaging estimator is then defined as

$$\hat{\mu}_{t+h,M}(\hat{w}) = \hat{w}\hat{\mu}_{t+h} + (1 - \hat{w})\tilde{\mu}_{t+h} = \begin{cases} \hat{\mu}_{t+h} & \text{if } F_T \leq 2 \\ (1 - \frac{2}{F_T})\hat{\mu}_{t+h} + \frac{2}{F_T}\tilde{\mu}_{t+h} & \text{otherwise} \end{cases} \tag{A32}$$

When the number of lags is unknown, [Hansen \(2010a\)](#) considered two alternative Mallows combination forecasts. The first is the so-called partial averaging forecast that averages only over unrestricted forecasts that vary according to the number of first-differenced lags included. With a maximum of K lags, this forecast is given by

$$\hat{\mu}_{t+h,M}(\hat{W}) = \sum_{l=0}^K \hat{w}_l \hat{\mu}_{t+h}(l) \tag{A33}$$

where $\hat{W} = (\hat{w}_0, \hat{w}_1, \dots, \hat{w}_K)'$ minimizes the criterion (with $\hat{\mu}_t(W) = \sum_{l=0}^K w_l \hat{\mu}_t(l)$),

$$M_P(W) = \sum_{t=1}^T (y_t - \hat{\mu}_t(W))^2 + 2\hat{\sigma}_K^2 \left(\sum_{l=0}^K [w_l(2 + l + p)] \right)$$

subject to the restrictions $w_j \geq 0$ ($j = 0, 1, \dots, K$), $\sum_{j=0}^K w_j = 1$. The second combination forecast is the so-called general averaging forecast that averages over the forecasts from all $2(K + 1)$ models that include the $(K + 1)$ restricted models. This forecast is given by

$$\check{\mu}_{t+h,M}(\check{W}) = \sum_{l=0}^K (\check{w}_{1l} \hat{\mu}_{t+h}(l) + \check{w}_{0l} \tilde{\mu}_{t+h}(l)) \tag{A34}$$

with $\check{W} = (\check{w}_{00}, \check{w}_{01}, \dots, \check{w}_{0K}, \check{w}_{10}, \check{w}_{11}, \check{w}_{12}, \dots, \check{w}_{1K})'$ minimizing the criterion (with $\check{\mu}_t(W) = \sum_{l=0}^K (w_{0l}\check{\mu}_t(l) + w_{1l}\hat{\mu}_t(l))$,

$$M_G(W) = \sum_{t=1}^T (y_t - \check{\mu}_t(W))^2 + 2\hat{\sigma}_K^2 \left(\sum_{l=0}^K [w_{0l}l + w_{1l}(2+l)] + p \right)$$

where the weights are non-negative and sum to one: $w_{1l} \geq 0, w_{0l} \geq 0, \sum_{l=0}^K (w_{0l} + w_{1l}) = 1$. In what follows, we will refer to (A33) and (A34) as the MPA (Mallows Partial Averaging) and MGA (Mallows General Averaging) forecasts, respectively.

Leave- h -out Cross Validation Selection. Hansen (2010b) provided theoretical justification for constructing h -step ahead forecasts using leave- h -out cross validation under the assumption that the data are strictly stationary. For model selection with a known number of lags, let CV_0 and CV_1 denote the cross-validation criteria for the restricted and unrestricted models, respectively. These criteria are computed as

$$CV_0 = \sum_{t=k+1}^{T-h} (y_{t+h} - \check{\mu}_{t+h}^{(t)})^2 \tag{A35}$$

$$CV_1 = \sum_{t=k+1}^{T-h} (y_{t+h} - \hat{\mu}_{t+h}^{(t)})^2 \tag{A36}$$

where $\check{\mu}_{t+h}^{(t)}$ and $\hat{\mu}_{t+h}^{(t)}$ are the restricted and unrestricted leave- h -out forecasts, respectively. Specifically, $\check{\mu}_{t+h}^{(t)}$ is obtained using parameter estimates from the restricted model after leaving out the observations $\{t+1, \dots, t+h\}$ ⁹:

$$\Delta y_j = \beta_0^* + \sum_{s=1}^k \alpha_k \Delta y_{j-s} + \epsilon_j, \quad j \neq t+1, \dots, t+h$$

Similarly, $\hat{\mu}_{t+h}^{(t)}$ is obtained from estimating the unrestricted model after leaving out the observations $\{t+1, \dots, t+h\}$:

$$y_j = \beta_0^* + \beta_1^* j + \alpha y_{j-1} + \sum_{s=1}^k \alpha_k \Delta y_{j-s} + \epsilon_j, \quad j \neq t+1, \dots, t+h$$

Then the cross-validation based forecast is $\hat{\mu}_{t+h,CV} = \hat{\mu}_{t+h} \mathbf{1}(CV_1 \leq CV_0) + \check{\mu}_{t+h} \mathbf{1}(CV_1 > CV_0)$. When the number of lags is unknown, the cross-validation criterion is computed for each of the $2(K+1)$ possible models and the selected forecast, denoted $\check{\mu}_{t+h,CV}$, is the one that corresponds to the model with the minimum value of this criterion.

Leave- h -out Cross Validation Averaging. When the number of lags is known, the cross validation weights $(\hat{w}, 1 - \hat{w})$ are obtained by minimizing the criterion

$$CV_w = \sum_{t=k+1}^{T-h} \left\{ w(y_{t+h} - \hat{\mu}_{t+h}^{(t)}) + (1-w)(y_{t+h} - \check{\mu}_{t+h}^{(t)}) \right\}^2$$

and the resulting forecast is $\hat{\mu}_{t+h,CV}(\hat{w}) = \hat{w}\hat{\mu}_{t+h} + (1-\hat{w})\check{\mu}_{t+h}$. When the number of lags is unknown, the partial combination forecast that only combines the unrestricted forecasts with different lags is obtained as

$$\hat{\mu}_{t+h,CV}(\hat{W}) = \sum_{l=0}^K \hat{w}_l \hat{\mu}_{t+h}(l) \tag{A37}$$

where $\hat{W} = (\hat{w}_0, \hat{w}_1, \dots, \hat{w}_K)'$ minimizes the criterion

$$CV_P(W) = \sum_{t=k+1}^{T-h} \left\{ \sum_{l=0}^K w_l (y_{t+h} - \hat{\mu}_{t+h}^{(l)}) \right\}^2 \tag{A38}$$

subject to the restrictions $w_j \geq 0$ ($j = 0, 1, \dots, K$), $\sum_{j=0}^K w_j = 1$, and $\hat{\mu}_{t+h}^{(l)}$ is the unrestricted leave- h -out forecast assuming l first-differenced lags. As with weight selection using the Mallows criterion, we also construct a general combination forecast that combines forecasts from the $K + 1$ unrestricted models as well as the $K + 1$ restricted models. This forecast is given by

$$\check{\mu}_{t+h,CV}(\check{W}) = \sum_{l=0}^K (\check{w}_{1l} \hat{\mu}_{t+h}(l) + \check{w}_{0l} \tilde{\mu}_{t+h}(l)) \tag{A39}$$

with $\check{W} = (\check{w}_{01}, \check{w}_{02}, \dots, \check{w}_{0K}, \check{w}_{11}, \check{w}_{12}, \dots, \check{w}_{1K})'$ minimizing the criterion

$$CV_G(W) = \sum_{t=k+1}^{T-h} \left\{ \sum_{l=0}^K \left[w_{1l} (y_{t+h} - \hat{\mu}_{t+h}^{(l)}) + w_{0l} (y_{t+h} - \tilde{\mu}_{t+h}^{(l)}) \right] \right\}^2$$

where $w_{1l} \geq 0, w_{0l} \geq 0, \sum_{l=0}^K (w_{0l} + w_{1l}) = 1$, $\hat{\mu}_{t+h}^{(l)}$ is as defined in (A38) and $\tilde{\mu}_{t+h}^{(l)}$ is the restricted leave- h -out forecast assuming l first-differenced lags. In what follows, we will refer to (A37) and (A39) as the CPA (Cross-Validation Partial Averaging) and CGA (Cross-Validation General Averaging) forecasts, respectively.

APE Selection. With a known number of lags, this forecast is computed from the model that corresponds to the lower APE between the restricted and unrestricted models:

$$\hat{\mu}_{t+h,S} = \tilde{\mu}_{t+h} I(APE_0 \leq APE_1) + \hat{\mu}_{t+h} I(APE_0 > APE_1)$$

$$APE_0 = \sum_{i=m_h}^{T-h} \{y_{i+h} - \tilde{\mu}_{i+h}\}^2, \quad APE_1 = \sum_{i=m_h}^{T-h} \{y_{i+h} - \hat{\mu}_{i+h}\}^2$$

In the unknown lags case, the forecast is computed from the model that minimizes the APE criterion among all $2(K + 1)$ possible models, comprising the $K + 1$ restricted and $K + 1$ unrestricted models.

Appendix C. Additional Simulation Results

This Appendix shows simulation evidence to justify the choice of $m_h = 20$. Figures A1 and A2 plot the difference between the optimal forecast risk (a grid from 20 to 70 in increments of 5) and the risk of the APE selection and APE averaging forecasts respectively for $m_h = 20$ expressed as a percentage of the forecast risk for $m_h = 20$.

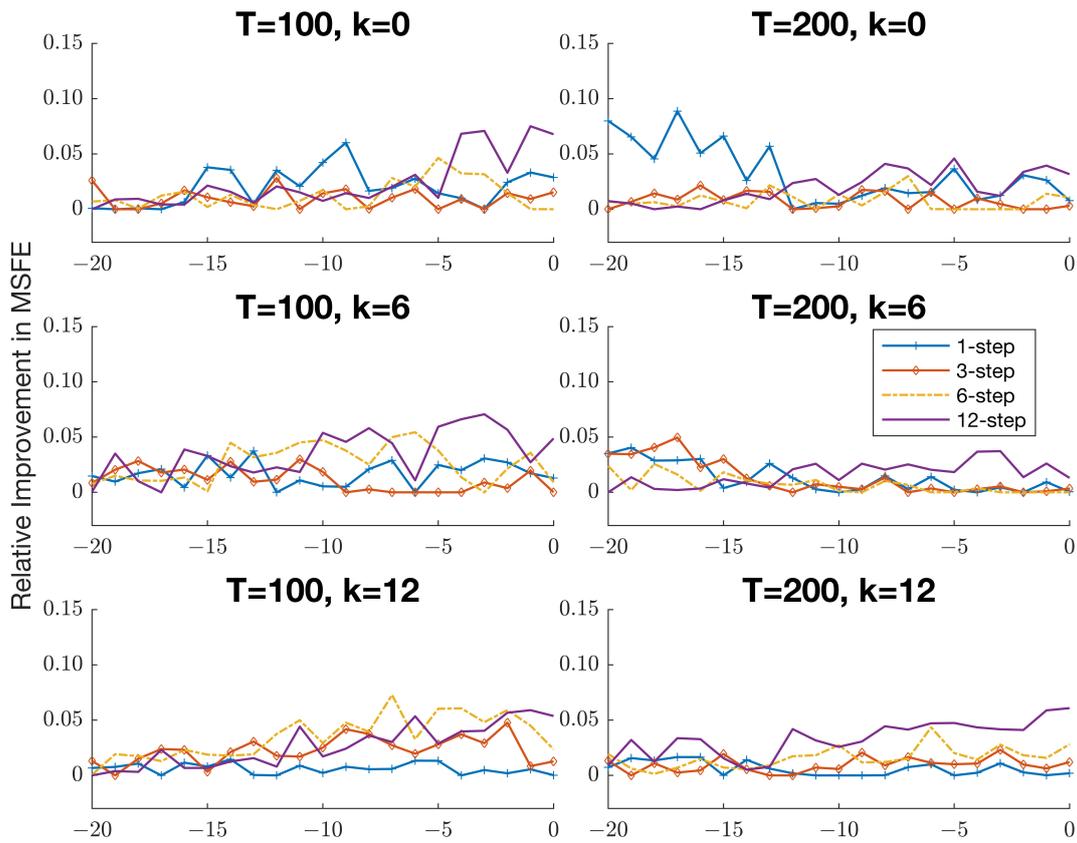


Figure A1. Forecast risk with optimal m_h and $m_h = 20$, APE selection ($p = 1$).

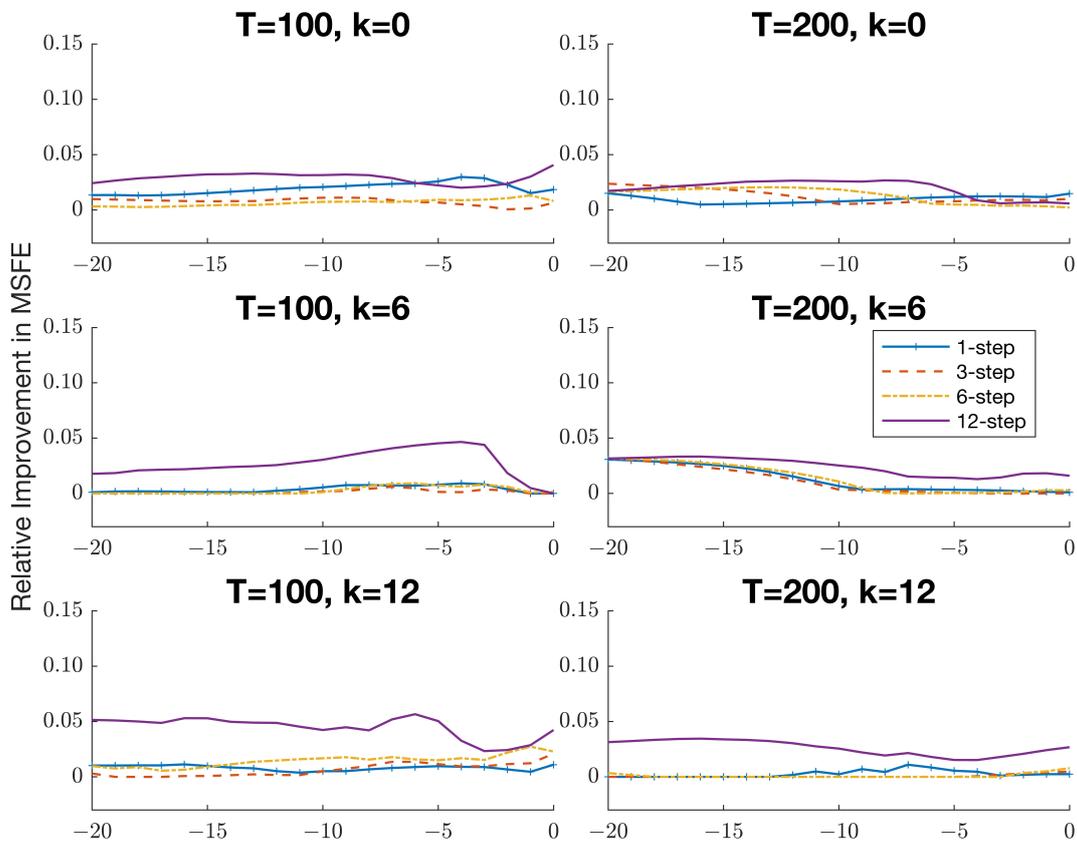


Figure A2. Forecast risk with optimal m_h and $m_h = 20$, APE averaging ($p = 1$).

Notes

- 1 Analytically, the importance of the trend component over long horizons can be seen by noting that the trend/drift coefficient is multiplied by the forecast horizon when constructing forecasts so that any specification/estimation error is magnified linearly as the forecast horizon increases (Sampson 1991).
- 2 The conclusion for the subsequent analysis will not be affected as long as the initial observations are $o_p(T^{1/2})$.
- 3 The figure was obtained by simulating the AMSE and AFR assuming i.i.d. normal errors with $T = 1000$. 5000 replications were used.
- 4 To illustrate the lack of uniformity, consider the case $p = 1$ with $k = 0$. Using the same arguments as in the proof of Theorem 2 of Yu et al. (2012), it follows that, for any finite $c \leq 0$, $\sum_{i=m_h}^{T-h} \{y_{i+h} - \hat{\mu}_{i+h}\}^2 = E(T_{10}^2) \log T + o_p(\log T)$, where $E(T_{10}^2) = 6$. The lack of uniformity follows since $E(T_{1c}^2) \neq E(T_{10}^2)$ for any $c < 0$.
- 5 This figure was obtained using the same method as in Figure 1.
- 6 We do not report the results for the selection forecasts, since their performance relative to the combination forecasts is qualitatively similar to the known lag-order case. The results are nevertheless available upon request.
- 7 This choice was also adopted by Ing and Yang (2014) in their Monte Carlo analysis of forecasting using autoregressive models with positive-valued errors.
- 8 The data set is publicly available for download at <https://research.stlouisfed.org/econ/mccracken/fred-databases/> (accessed on 1 February 2023).
- 9 Hansen (2010b) instead left out the $2h - 1$ observations $\{t - h + 1, \dots, t, t + 1, \dots, t + h - 1\}$. The difference emanates from the fact that he constructs direct forecasts while our forecasts are constructed iteratively which exploit the autoregressive structure and hence necessitate leaving out only the h observations $\{t + 1, \dots, t + h\}$.

References

- Bates, John M., and Clive W. J. Granger. 1969. The combination of forecasts. *Journal of the Operational Research Society* 20: 451–68. [CrossRef]
- Box, George E. P., and Gwilym M. Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Cheng, Xu, and Bruce E. Hansen. 2015. Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics* 186: 280–93. [CrossRef]
- Chow, Yuan Shih. 1965. Local convergence of martingales and the law of large numbers. *The Annals of Mathematical Statistics* 36: 552–58. [CrossRef]
- Clements, Michael P., and David F. Hendry. 2001. Forecasting with difference-stationary and trend-stationary models. *The Econometrics Journal* 4: 1–19.
- Diebold, Francis X., and Lutz Kilian. 2000. Unit-root tests are useful for selecting forecasting models. *Journal of Business & Economic Statistics* 18: 265–73.
- Diebold, Francis X., and Roberto S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253–63.
- Elliott, Graham. 2006. *Unit Root Pre-Testing and Forecasting*. Technical Report, Working Paper. San Diego: UCSD.
- Gallant, A. Ronald. 1981. On the bias in flexible functional forms and an essentially unbiased form: The fourier flexible form. *Journal of Econometrics* 15: 211–45. [CrossRef]
- Gonzalo, Jesus, and Jean-Yves Pitarakis. 1998. On the exact moments of asymptotic distributions in an unstable ar(1) with dependent errors. *International Economic Review* 39: 71–88. [CrossRef]
- Granger, Clive W. J. 1966. The typical spectral shape of an economic variable. *Econometrica* 34: 150–61. [CrossRef]
- Hamilton, James Douglas. 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Hansen, Bruce E. 2007. Least squares model averaging. *Econometrica* 75: 1175–89. [CrossRef]
- Hansen, Bruce E. 2008. Least-squares forecast averaging. *Journal of Econometrics* 146: 342–50. [CrossRef]
- Hansen, Bruce E. 2010a. Averaging estimators for autoregressions with a near unit root. *Journal of Econometrics* 158: 142–55. [CrossRef]
- Hansen, Bruce E. 2010b. Multi-step forecast model selection. Paper presented at the 20th Annual Meetings of the Midwest Econometrics Group, St. Louis, MO, USA, October 1–2.
- Ing, Ching-Kang. 2001. A note on mean-squared prediction errors of the least squares predictors in random walk models. *Journal of Time Series Analysis* 22: 711–24. [CrossRef]
- Ing, Ching-Kang. 2003. Multistep prediction in autoregressive processes. *Econometric Theory* 19: 254–79. [CrossRef]
- Ing, Ching-Kang. 2004. Selecting optimal multistep predictors for autoregressive processes of unknown order. *The Annals of Statistics* 32: 693–722. [CrossRef]
- Ing, Ching-Kang, Jin-Lung Lin, and Shu-Hui Yu. 2009. Toward optimal multistep forecasts in non-stationary autoregressions. *Bernoulli* 15: 402–37. [CrossRef]
- Ing, Ching-Kang, Chor-yiu Sin, and Shu-Hui Yu. 2012. Model selection for integrated autoregressive processes of infinite order. *Journal of Multivariate Analysis* 106: 57–71. [CrossRef]
- Ing, Ching-Kang, and Chiao-Yi Yang. 2014. Predictor selection for positive autoregressive processes. *Journal of the American Statistical Association* 109: 243–53. [CrossRef]

- Kejriwal, Mohitosh, and Xuewen Yu. 2021. Generalized forecast averaging in autoregressions with a near unit root. *The Econometrics Journal* 24: 83–102. [CrossRef]
- Kim, Hyun Hak, and Norman R. Swanson. 2018. Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting* 34: 339–54. [CrossRef]
- Liao, Jen-Che, and Wen-Jen Tsay. 2020. Optimal multistep var forecast averaging. *Econometric Theory* 36: 1099–126. [CrossRef]
- Mallows, Colin L. 2000. Some comments on cp. *Technometrics* 42: 87–94.
- Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes. 2023. Machine learning advances for time series forecasting. *Journal of Economic Surveys* 37: 76–111. [CrossRef]
- McCracken, Michael W., and Serena Ng. 2016. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34: 574–89.
- Medeiros, Marcelo C., Gabriel F. R. Vasconcelos, Álvaro Veiga, and Eduardo Zilberman. 2021. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics* 39: 98–119.
- Meng, Xiao-Li. 2005. From unit root to stein's estimator to fisher'sk statistics: If you have a moment, i can tell you more. *Statistical Science* 20: 141–62. [CrossRef]
- Ng, Serena, and Timothy Vogelsang. 2002. Forecasting autoregressive time series in the presence of deterministic components. *The Econometrics Journal* 5: 196–224. [CrossRef]
- Phillips, Peter C. B. 2014. On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82: 1177–95. [CrossRef]
- Rissanen, Jorma. 1986. Order estimation by accumulated prediction errors. *Journal of Applied Probability* 23: 55–61. [CrossRef]
- Sampson, Michael. 1991. The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models. *Journal of Applied Econometrics* 6: 67–76. [CrossRef]
- Stock, James H., and Mark W. Watson. 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20: 147–62.
- Stock, James H., and Mark W. Watson. 2005. *An Empirical Comparison of Methods for Forecasting Using Many Predictors*. Manuscript. Princeton: Princeton University, p. 46.
- Stock, James H., and Mark W. Watson. 2006. Forecasting with many predictors. *Handbook of Economic Forecasting* 1: 515–54.
- The MathWorks Inc. 2022. MATLAB version: 9.13.0 (R2022b) Natick, Massachusetts: The MathWorks Inc. Available online: <https://www.mathworks.com> (accessed on 1 February 2023)
- Tu, Yundong, and Yanping Yi. 2017. Forecasting cointegrated nonstationary time series with time-varying variance. *Journal of Econometrics* 196: 83–98. [CrossRef]
- Turner, John L. 2004. Local to unity, long-horizon forecasting thresholds for model selection in the ar (1). *Journal of Forecasting* 23: 513–39. [CrossRef]
- Wang, Xiaoqian, Rob J. Hyndman, Feng Li, and Yanfei Kang. 2022. Forecast combinations: An over 50-year review. *International Journal of Forecasting* 39: 1518–47. [CrossRef]
- Wei, Ching-Zong. 1987. Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *The Annals of Statistics* 15: 1667–82. [CrossRef]
- Yu, Shu-Hui, Chien-Chih Lin, and Hung-Wen Cheng. 2012. A note on mean squared prediction error under the unit root model with deterministic trend. *Journal of Time Series Analysis* 33: 276–86. [CrossRef]
- Zhang, Xinyu, Alan T. K. Wan, and Guohua Zou. 2013. Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics* 174: 82–94. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.