



Munich Personal RePEc Archive

# **Direct and iterated multistep AR methods for difference stationary processes**

Proietti, Tommaso

SEFEMEQ University of Rome "Tor Vergata"

1 October 2008

Online at <https://mpa.ub.uni-muenchen.de/10859/>  
MPRA Paper No. 10859, posted 01 Oct 2008 08:41 UTC

# Direct and iterated multistep AR methods for difference stationary processes

Tommaso Proietti\*  
University of Rome "Tor Vergata"

1st October 2008

## Abstract

The paper focuses on the comparison of the direct and iterated AR predictors when  $X_t$  is a difference stationary process. In particular, it provides some useful results for comparing the efficiency of the two predictors and for extracting the trend from macroeconomic time series using the two methods. The main results are based on an encompassing representation for the two predictors which enables to derive their properties quite easily under a maintained model. The paper provides an analytic expression for the mean square forecast error of the two predictors and derives useful recursive formulae for computing the direct and iterated coefficients. From the empirical standpoint, we propose estimators of the AR coefficients based on the tapered Yule-Walker estimates; we also provide a test of equal forecast accuracy which is very simple to implement and whose critical values can be obtained with the bootstrap method. Since multistep prediction is tightly bound up with the estimation of the long run component in a time series, we turn to the role of the direct method for trend estimation and derive the corresponding multistep Beveridge-Nelson decomposition.

*Keywords:* Multistep estimation. Tapered Yule-Walker estimates. Beveridge-Nelson decomposition. Forecast combination.

*JEL codes:* C22, C51, C53, E32.

---

\*Address for Correspondence: Dipartimento SEFEMEQ, Università di Roma "Tor Vergata", Via Columbia 2, 00133 Rome, Italy. *E-mail:* tommaso.proietti@uniroma2.it. This paper was prepared for the 5-th Eurostat Colloquium on Modern Tools for Business Cycle Analysis, Luxembourg 29 Sept. - 1st Oct. 2008.

# 1 Introduction

Let  $X_t, t = 1, 2, \dots, n$ , be an integrated stochastic process, so that  $\Delta X_t = X_t - X_{t-1}$  is a stationary zero mean process. We assume throughout that we are interested in predicting  $h$ -steps ahead the levels (rather than the differences) of the series. Two important linear predictors, the *direct* (labelled by  $D$  henceforth) and *iterated* predictors (labelled by  $I$ ), are obtained by fitting the following finite order autoregressive (AR) model, usually by least squares:

$$X_{t+h} = X_t + \sum_{j=1}^p \phi_{jh}^{(i)} \Delta X_{t-j+1} + \epsilon_{t+h|t}^{(i)}, \quad i = D, I, \quad (1)$$

where  $\epsilon_{t+h|t}^{(i)}$  denotes the  $h$ -steps ahead prediction error. The two predictors use the same information set, represented by the vector  $\Delta \mathbf{X}'_t = [\Delta X_t, \Delta X_{t-1}, \dots, \Delta X_{t-p+1}]$ , but differ in the definition of the coefficients  $\phi_{jh}^{(i)}$ .

The *direct* predictor of the levels  $X_{t+h}$  arises from the direct projection of  $\Delta_h X_{t+h} = X_{t+h} - X_t$  on  $\Delta \mathbf{X}'_t$ ; it can be expressed as  $X_{t+h|t}^{(D)} = X_t + \Delta_h X_{t+h|t}^{(D)}$ , where  $\Delta_h X_{t+h|t}^{(D)} = \sum_{j=1}^p \phi_{jh}^{(D)} \Delta X_{t-j+1}$ , and the coefficients minimize the  $h$ -step ahead mean square forecast error,  $\text{MSFE}_D(h, p) = E[(X_{t+h} - X_{t+h|t}^{(D)})^2]$ . Notice that this is different from the direct predictor of the *changes*  $\Delta X_{t+h}$ , which arises from projecting  $\Delta X_{t+h}$  onto  $\Delta \mathbf{X}'_t$ .

The *indirect* (or *plug-in*) predictor is obtained from the AR( $p$ ) model by iterating via the chain rule the one-step-ahead predictor, so as to obtain forecasts of all the intermediate future changes  $\Delta X_{t+k}$ , for  $k = 1, \dots, h$ , which are combined to yield:  $X_{t+h|t}^{(I)} = X_t + \sum_{k=1}^h \Delta X_{t+k|t}^{(I)}$ , where  $\Delta X_{t+k|t}^{(I)} = \sum_{j=1}^p \phi_{j1}^{(I)} \Delta X_{t+k-j|t}^{(I)}$  (with  $\Delta X_{t+k-j|t}^{(I)} = \Delta X_{t+k-j}$ , if  $j \geq k$ ), and the coefficients  $\phi_{j1}^{(I)}, j = 1, \dots, p$ , minimize  $\text{MSFE}_D(1, p) = E[(X_{t+1} - X_{t+1|t}^{(I)})^2] = E[(\Delta X_{t+1} - \Delta X_{t+1|t}^{(I)})^2]$ . Obviously,  $\phi_{j1}^{(I)} = \phi_{j1}^{(D)}$ . From the application of the chain rule we can express the indirect predictor as  $X_{t+h|t}^{(I)} = X_t + \sum_{j=1}^p \phi_{jh}^{(I)} \Delta X_{t+j-1}$ , where  $\phi_{jh}^{(I)}$  are the iterated AR multistep coefficients (which will be defined more properly in a later section).

The efficiency of the two methods is judged by comparing  $\text{MSFE}_D(h, p)$  with  $\text{MSFE}_I(h, p) = E[(X_{t+h} - X_{t+h|t}^{(I)})^2]$ ; if we are given a finite realisation of  $X_t$ , the comparison will be based on their sample counterparts. There is a vast and well established literature comparing the performance of the two predictors for the purpose of forecasting more than one step ahead, not exclusively in the AR case. We refer to Bhansali (2002) and Chevillon (2007) for comprehensive surveys of the literature. Actually, the seminal paper by Cox (1961) concerned multistep estimation of a first order integrated moving average model, yielding exponential smoothing forecasts. Other essential references are Findley (1983), Weiss (1991), Tiao and Xu (1993), Tiao and Tsay (1994), Clements and Hendry (1996), and Ing (2003, 2004). In a recent paper, Marcellino, Stock and Watson (2008) carry out an extensive real time multistep forecasting exercise comparing the performance of the direct and the iterated predictors for a set of U.S. macroeconomic time series. Their main finding is that, despite the theoretical superiority of the direct forecasts, the iterated predictor stands out as a winner.

In this paper we focus on the comparison of the direct and iterated AR predictors when  $X_t$  is a difference stationary process. In particular, we aim at comparing the efficiency of the direct approach for out of sample forecasting at different horizons and we discuss its role for trend extraction from macroeconomic time series. For this purpose we derive an encompassing representation for the two predictors (see section 2), according to which they result from the application of stable AR filters to the stationary changes of the series. This sets

up a common ground for the comparison of their theoretical properties, which are easily obtained under a maintained model. The paper provides an analytic expression for the mean square forecast error of the two predictors and derives useful recursive formulae for the direct and iterated coefficients.

Section 3 illustrates these results when the true model is ARIMA(1, 1, 1); an important finding is that the comparative efficiency gains of the direct predictor over the iterated one are larger when the AR model is grossly misspecified, in which case the predictive performance of the direct AR predictor is poor anyway, in comparison with the minimum MSFE predictor. It would be preferable in these occurrences to move away from the AR representation and to look for an alternative specification, but large improvements can be obtained by combining the direct predictor with a multistep exponential smoothing predictor.

In section 4 we discuss several empirical issues. For consistency with the theoretical framework, we propose estimators of the coefficients  $\phi_{jh}^{(i)}$  based on the solution of tapered Yule-Walker systems. We also provide a test of equal forecast accuracy which is very simple to implement and whose critical values can be obtained with the bootstrap method.

Since multistep prediction is tightly bound up with the estimation of the long run component in a time series, we turn to the role of the direct method for trend estimation (section 5) and derive the corresponding long-run Beveridge-Nelson decomposition (Beveridge and Nelson, 1981). Section 6 provides two empirical illustration concerning the U.S. gross domestic product and monthly inflation; in the second case, the direct method produces a significant increase in predictive accuracy at horizons greater than a year and yields estimates of smoother trends. In section 7 we summarize the contribution of the paper and draw our conclusions.

## 2 A convenient representation

In this section we establish a simple and fundamental result which derives the two competitor predictors, direct and iterated, as arising from the application of a stable AR filter to the stationary changes of the series. Let us denote the  $h$ -step ahead prediction error associated to the  $i$ -th predictor,  $X_{t+h|t}^{(i)}$ ,  $i = D, I$ , by  $\epsilon_{t+h|t}^{(i)} = X_{t+h} - X_{t+h|t}^{(i)}$ . Since both predictors take the form  $X_{t+h|t}^{(i)} = X_t + \sum_{j=1}^p \phi_{jh}^{(i)} \Delta X_{t+j-1}$ , the prediction error is rewritten as

$$\epsilon_{t+h|t}^{(i)} = X_{t+h} - X_t - \sum_{j=1}^p \phi_{jh}^{(i)} \Delta X_{t+j-1}.$$

The linear combination of past and lagged values of the process on the right hand side can be expressed in terms of the first differences  $\Delta = 1 - L$ , where  $L$  is the lag operator,  $L^j X_t = X_{t-j}$ :

$$\epsilon_{t+h|t}^{(i)} = [S_{h-1}(L) + L^{h-1} \phi_h^{(i)}(L)] \Delta X_{t+h}. \quad (2)$$

Here we have denoted  $S_m(L) = 1 + L + L^2 + \dots + L^{m-1}$ ,  $\Delta_h = 1 - L^h = \Delta S_h(L)$ , and  $\phi_h^{(i)}(L) = 1 - \phi_{1h}^{(i)} L - \dots - \phi_{ph}^{(i)} L^p$ .

The corresponding MSFE is obtained as the variance of the filtered first differences of the process. Writing the multistep prediction filter as  $\nu_i(L) = S_{h-1}(L) + L^{h-1} \phi_h^{(i)}(L)$ , it is immediate to show that

$$\text{MSFE}_i(h, p) = \gamma(0) \sum_j \nu_{ij}^2 + 2 \sum_k \gamma(k) \sum_j \nu_{ij} \nu_{i,j+k}, \quad i = D, I, \quad (3)$$

where  $\gamma(k) = E(\Delta X_t \Delta X_{t-k})$  is the autocovariance function of  $\Delta X_t$  and  $\nu_{ij}$  is the coefficient of the polynomial  $\nu_i(L)$  associated with the  $j$ -th power of the lag operator.

Expression (3) is useful since it allows to express the MSFE of the direct and indirect predictors as a function of true underlying process, via its autocovariance function. It is the AR counterpart of the result obtained for the exponential smoothing predictor by Tiao and Xu (1993, formula 2.3). In the frequency domain, the equivalent expression is

$$\text{MSFE}_i(h, p) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\nu_i(e^{-i\omega})|^2 g(\omega) d\omega,$$

with  $|\nu_i(e^{-i\omega})|^2 = \nu_i(e^{-i\omega})\nu_i(e^{i\omega})$ , the squared gain of the filter  $\nu_i(L)$ , and  $g(\omega)$  represents the spectral generating function of  $\Delta X_{t+h}$ .

In the light of (3), the differences in the two predictors lie in the AR coefficients  $\phi_{jh}^{(i)}$ . For the direct predictor,  $i = D$ , the coefficients  $\phi_h^{(D)} = [\phi_{1h}^{(D)}, \dots, \phi_{ph}^{(D)}]'$  are obtained by minimizing  $\text{MSFE}(h, p)^{(D)}$  with respect to  $\phi_h^D$ . The optimization problem leads to the following linear system of equations:

$$\mathbf{\Gamma} \phi_h^{(D)} = \gamma_h, \quad (4)$$

with

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \ddots & \gamma(p-2) \\ \vdots & \ddots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{bmatrix}, \quad \gamma_h = \begin{bmatrix} \gamma(1) + \cdots + \gamma(h) \\ \gamma(2) + \cdots + \gamma(h+1) \\ \vdots \\ \gamma(p) + \cdots + \gamma(h+p-1) \end{bmatrix}.$$

Notice that, from

$$\gamma_h = \gamma_{h-1} + \gamma^{(h)}, \quad \gamma^{(h)} = \begin{bmatrix} \gamma(h) \\ \gamma(h+1) \\ \vdots \\ \gamma(h+p-1) \end{bmatrix}, \quad h = 2, \dots, \gamma_1 = \gamma^{(1)},$$

it follows

$$\phi_h^{(D)} = \phi_{h-1}^{(D)} + \phi^{(h)}, \quad \phi^{(h)} = \mathbf{\Gamma}^{-1} \gamma^{(h)}. \quad (5)$$

Bondon (2001) and Brockwell and Dahlhaus (2004) provide generalized Levinson–Durbin recursions for computing the coefficients  $\phi^{(h)}$ , which operate both on the order  $p$  and the forecast lead  $h$ .

The iterated method obtains the coefficients  $\phi_{jh}^{(I)}$ ,  $j = 1, \dots, p$  in (2) recursively from the one-step-ahead coefficients, which are obtained from the linear system  $\phi_1^{(I)} = \phi_1^{(D)} = \mathbf{\Gamma}^{-1} \gamma_1$ :

$$\phi_h^{(I)'} = \mathbf{e}'_1 (\mathbf{I} - \mathbf{T}^h) (\mathbf{I} - \mathbf{T})^{-1} \mathbf{T} = \mathbf{e}'_1 \sum_{j=1}^h \mathbf{T}^j$$

where

$$\mathbf{T} = \begin{bmatrix} \phi_1^{(I)} & \phi_2^{(I)} & \cdots & \phi_{p-1}^{(I)} & \phi_p^{(I)} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

The iterated AR coefficients satisfy the following first order recursion:

$$\phi_h^{(I)} = \phi_{h-1}^{(I)} + \mathbf{T}' \mathbf{e}_1, \quad (6)$$

with starting value  $\phi_1^{(I)} = \mathbf{T}' \mathbf{e}_1 = \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_1$ .

It is important to remark that the indirect predictor can also be obtained by replacing, in the expression for the direct predictor, the autocovariances  $\gamma(p+k)$ ,  $k \geq 1$  with the values implied by the AR( $p$ ) model:

$$\tilde{\gamma}(p+k) = \sum_{j=1}^p \phi_{1j} \tilde{\gamma}(p+k-j)$$

where  $\tilde{\gamma}(p+k-j) = \gamma(p+k-j)$  for  $k \leq j$ .

In matrix notation, setting

$$\mathbf{T}^* = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \phi_p^{(I)} & \phi_{p-1}^{(I)} & \phi_{p-2}^{(I)} & \cdots & \phi_1^{(I)} \end{bmatrix},$$

$$\tilde{\boldsymbol{\gamma}}^{(h)} = \mathbf{T}^* \tilde{\boldsymbol{\gamma}}^{(h-1)},$$

An obvious (the coefficient of the AR direct predictor are chosen so as to minimize the MSFE at horizon  $h$ ) but important result is that, if  $\mathbf{\Gamma}$  is positive definite,  $\text{MSFE}_I(h, p) \geq \text{MSFE}_D(h, p)$ . This fact can be proven using e.g. the results in Ing (2003), who proves a more general theorem, referring to the case when  $X_t$  is stationary, and taking into account the estimation uncertainty.

### 3 Comparison for an ARIMA(1,1,1) process

In this section we illustrate the use of expression (3) for characterising the comparative forecasting performances of the direct and iterated predictors. We assume that  $X_t$  is generated by the ARIMA(1,1,1) process  $\Delta X_t = \phi \Delta X_{t-1} + \xi_t + \theta \xi_{t-1}$ , with  $|\phi| < 1$  and  $|\theta| \leq 1$ , so that  $\gamma(k)$  in (3) is the autocovariance function of the stationary ARMA(1,1) process for  $\Delta X_t$ . The true generating process is simple, but at the same time sufficiently rich to illustrate a few important facts.

Figure 1 refers to the case  $h = 4$  and  $p = 2$  and displays in the first panel the efficiency ratio  $\text{ER}_{ID}(4, 2) = 100 \times \text{MSFE}_I(4, 2) / \text{MSFE}_D(4, 2)$ , as a function of the values of the AR and MA parameters  $\phi$  and  $\theta$ . Obviously, the ratio cannot be smaller than 100. An important evidence is that the superiority of the direct predictor is not overwhelming, as the scale of the vertical axis suggests, e.g. around 5% when  $\phi = 0.95$  and  $\theta = -0.65$ . The greatest efficiency gains arise when  $\theta$  is close to -1 and  $\phi$  is close to 1, and no cancelation of roots occurs.

The second figure (top right) serves to assess how good are the direct forecasts as compared to the true model forecasts, by displaying the efficiency ratio  $\text{ER}_{DT}(4, 2) = 100 \times \text{MSFE}_D(4, 2) / \text{MSFE}(4)$ , where the denominator is the true MSFE of the ARIMA(1,1,1) optimal forecasts,  $\text{MSFE}(h) = \text{E}\{[X_{t+h} - \text{E}(X_{t+h} | \mathcal{F}_t)]^2\}$ , where  $\mathcal{F}_t$  is the information set at time  $t$ , which is the minimum value that can be attained by any predictor. The interesting fact is that for the parameters combinations of interest ( $\phi$  and  $-\theta$  are

close to 1) the performance of the direct predictor is poor anyway, as the efficiency loss with respect to the minimum MSFE predictor can reach up to 40%.

It is worth the while to compare the predictive accuracy of the direct AR predictor with an important competitor simple predictor that has been proposed by Cox (1963), Tiao and Xu (1993) and Haywood and Tunnicliffe-Wilson (1997), namely the multistep *exponential smoothing* (ES) predictor,

$$X_{t+h|t}^{(ES)} = \sum_{j=0}^{\infty} w_j X_{t-j}, \quad w_j = (1 - \lambda_h) \lambda_h^j,$$

where the weights sum to one and depend on a single smoothing constant,  $\lambda_h$ , taking values between 0 and 1, which is chosen so as to minimise MSFE at forecast horizon  $h$ . The prediction error can be expressed in terms of the stationary changes of  $X_t$  as follows:

$$\begin{aligned} \epsilon_{t+h|t}^{(ES)} &= X_{t+h} - X_{t+h|t}^{(ES)} \\ &= X_t + \sum_{k=1}^h \Delta X_{t+k} - \sum_{j=0}^{\infty} w_j X_{t-j} \\ &= [S_{h-1} + L^{h-1} \phi^{(ES)}(L)] \Delta X_{t+h} \end{aligned} \quad (7)$$

The lag polynomial  $\phi^{(ES)}(L)$  is of infinite order and its coefficients satisfy the first order difference equation  $\phi_j^{(ES)} = \phi_{j-1}^{(ES)} + w_j$ , with starting value  $\phi_1^{(ES)} = w_0 - 1$ . Representation (7) follows directly from the fact that  $\sum_{j=0}^{\infty} w_j = 1$ .

The ES predictor uses all the available observations, but since it depends on a single parameter, it has less flexibility with respect to the direct predictor, which changes also with the lag order  $p$ . The plot of the MSFE ratio  $100 \times \text{MSFE}_D(4, 2) / \text{MSFE}_{ES}(4)$  shows (see the bottom left panel of figure 1), the ES predictor outperforms the direct AR one when  $\theta$  is close to -1 and greater than  $-\phi$ . The reverse holds for  $\phi > -\theta$ . This finding opens the way to combining the forecasts. The MSFE of the combined predictor,

$$X_{t+h|t}^C = \tau X_{t+h|t}^D + (1 - \tau) X_{t+h|t}^{ES},$$

is compared to the minimum MSFE of the optimal forecasts in the right bottom panel of figure 1. The weight  $\tau \in (0, 1)$  is the first element of the vector  $(\mathbf{i}' \Sigma_h^{-1} \mathbf{i})^{-1} \Sigma_h^{-1} \mathbf{i}$ , where  $\Sigma_h$  is the variance covariance matrix of the vector  $[\epsilon_{t+h|t}^{(D)}, \epsilon_{t+h|t}^{(ES)}]'$ . The combined predictor outperforms uniformly the direct AR predictor as it emerges from the comparison of the left panels of figure 1.

For higher values of  $h$  the predictive gains are more substantial; for instance, for  $h = 12$  and  $p = 2$ , the direct forecast are 20% more accurate than the iterated ones, when  $\phi = 0.95$  and  $\theta = -0.65$ . This is visible from figure 2, whose top left panel shows the values  $\text{ER}_{ID}(12, 2)$ , corresponding to different values of  $(\phi, \theta)$ . The right panel illustrates that once again that for values of  $\theta$  close to -1 and  $\phi$  close to 1 the performance of the direct predictor improves considerably over the iterated one. Finally, by increasing the order of the AR approximation, for  $h$  fixed, the gap between the two predictors narrows (see the bottom left panel) and the direct predictor outperforms the iterated one  $\theta$  is close to -1, which is also the case when the direct predictor displays the poorer performance compared to the true predictor (see the bottom right panel).

The conclusions that we may draw from this simple example are the following.

- The comparative gains of the direct over the iterated predictor may not be very large, especially for small  $h$  and large  $p$ .
- Choosing a large  $p$  exposes the analysis to the dangers of overfitting. See Granger and Jeon (2006) for the consequences on the estimated AR polynomials.

- Very large predictive accuracy gains are obtainable when the AR model is grossly misspecified, in which case the predictive performance of the direct AR predictor is poor in comparison with the minimum MSFE predictor. It would be preferable in these occurrences to move away from the AR representation and look for an alternative specification, or the combination with alternative forecasts.
- The previous observations suggests that one may use the difference in the two predictors as evidence for model misspecification and use the direct forecast only in the absence of a better representation of the series.
- The commonest source of misspecification is due to the presence of an MA component close to the non-invertibility region. The combination of the direct forecasts with exponential smoothing forecasts yields a predictor which is almost as efficient as the optimal predictor.

## 4 Estimation issues and a bootstrap test of predictive ability

Given a realization of the stochastic process  $X_t$ , denoted  $x_t, t = 1, \dots, n$ , there are several alternative estimators of the direct and indirect coefficients,  $\phi_h^{(i)}, i = I, D$ . The most common estimation method is ordinary least squares (LS), by which the vector  $\hat{\phi}_h^{(D)}$  minimizes  $\sum_t (\Delta_h x_{t+h} - \hat{\phi}_h^{(D)' \Delta \mathbf{x}_t})^2$ , where  $\Delta \mathbf{x}_t = [\Delta x_t, \Delta x_{t-1}, \dots, \Delta x_{t-p+1}]'$ . The properties of the corresponding predictor have been discussed by Ing (2004) in the stationary case; Marcellino, Stock and Watson (2006) provide an empirical comparison of the direct and plug in least squares predictors in terms of their capability of forecasting a large set of macroeconomic time series, both stationary and non stationary.

The problems with the least square estimates are twofold. First, the AR estimated parameters may be nonstationary. Secondly, for given horizon and AR order the empirical MSEF of the iterated predictor can be smaller than that of the direct predictor. On the contrary, the Yule-Walker estimates, which are obtained by replacing the theoretical autocovariances in (5) by their sample counterparts  $\hat{\gamma}(k) = n^{-1} \sum_{t=1}^{n-k} \Delta x_t \Delta x_{t+k}$ , are guaranteed to correspond to a stationary AR process and they enforce the condition  $\widehat{\text{MSFE}}_I(h, p) \geq \widehat{\text{MSFE}}_D(h, p)$ .

On the other hand, it is well known that the Yule-Walker estimators suffer from larger bias than the least squares estimates for short time series and when the root of the AR polynomial is close to one (Parzen, p. 351, Tjostheim and Paulsen, 1983, Kang, 1987, Shaman and Stine, 1988). These drawbacks are alleviated by *tapering*. A taper is a data window taking the form of a sequence of positive weights  $h_t, t = 1, \dots, n$  that leaves unaltered the series in the middle of the sample and downweights the observations at the extremes. In other words, tapering amounts to smoothing the observed sample transition from zero to the observed values when estimating convolutions of data sequences such as the autocovariances and the periodogram.

### 4.1 Tapered Yule-Walker estimates

The tapered Yule-Walker estimates of the AR coefficients are obtained by replacing the theoretical autocovariances with those computed on the sequence  $h_t \Delta x_t$ , by the estimator:

$$\hat{\gamma}(k) = \frac{n}{(\sum_{t=1}^n h_t^2)^2} \sum_{t=1}^{n-k} h_t \Delta x_t h_{t+k} \Delta x_{t+k}.$$



In our applications we consider the Tukey-Hanning data taper (see e.g. Bloomfield, 1985, p. 84, and Dahlhaus, 1988), such that, defining  $u = (t - 0.5)/n$ ,

$$h_t = \begin{cases} 0.5 [1 - \cos(2\pi u/\varrho)], & u \leq 0.5\varrho, \\ 1, & 0.5\varrho \leq u \leq 1 - 0.5\varrho, \\ 0.5 [1 - \cos(2\pi(1 - u)/\varrho)], & u \geq 1 - 0.5\varrho, \end{cases}$$

The  $\varrho$  parameter, regulating the fraction of the initial and final stretch of data that are tapered, is set equal to 0.1 (see Hurvich, 1988, for a method to estimate the optimal degree of tapering). Notice that the standard biased estimator of the autocovariance arise when the boxcar taper, with  $h_t = 1, 1 \leq t \leq n$  and 0 otherwise, is adopted.

The tapered Yule-Walker estimates have better small sample properties with respect to the non-tapered counterparts. In particular they can reduce substantially the bias affecting the Yule Walker estimates of the AR parameters, see e.g. Dahlhaus (1988). Zhou and Roy (2006) document the reduction of the bias and the corresponding improvement in forecast accuracy in the vector AR case. Tapering was originally proposed as a device for removing leakage in spectrum estimation using the periodogram (see Percival and Walden, 1983). Velasco and Robinson (2000) discuss its merits for the estimation of the long-memory parameter by pseudo maximum likelihood in the frequency domain. The reduction of the bias is achieved at the expenses of an increase in the variance of the estimates. An interesting strategy to avoid it is to use multitapered estimates (see Walden, 2000).

For solving the system  $\hat{\Gamma}\hat{\phi}_h^{(D)} = \hat{\gamma}_h$ , we use the functions for Toeplitz systems built in the package Ox 4.00 by Doornik (2006), which make use of the Levinson-Durbin algorithm.

## 4.2 Order Selection

The choice of the AR order  $p$  is done by information criteria. The selection of  $p$  for stationary time series has been considered by Shibata (1980) and Bhansali (1996), who advocate the use of the Akaike Information Criterion, where the estimated one-step innovation variance is replaced by the estimated  $h$ -step prediction error variance. Hurvich and Tsai (1997) introduced a multistep generalization of the corrected AIC, given by

$$\text{AIC}_C(h, p) = n[\log \widehat{\text{MSFE}}_D(h, p) + 1] + 2(p + 1)\frac{n}{n - p - 2}. \quad (8)$$

## 4.3 A Bootstrap Test of Predictive Efficiency

We can take advantage of the properties of the tapered Yule-Walker estimates to build up a test of the significance of the improved predictive performance of the direct predictor. In fact, the statistic representing the difference between the mean square forecast error  $\widehat{\text{MSFE}}_I(h, p) - \widehat{\text{MSFE}}_D(h, p)$  is always nonnegative and can be written as a linear combination of the first  $p$  autocovariances. However, the weights of the combination depend on the estimated coefficients  $\nu_{ij}$ , which in turn depend on the autocovariance function of  $\Delta X_t$ .

To judge the significance of the reduction of the MSFE arising from using the direct predictor at horizon  $h$  we propose the following  $F$ -type test statistic, defined in terms of the Granger and Newbold (1986, p. 310) measure of forecastability at horizon  $h$ :

$$F(h, p) = \frac{(R_D^2 - R_I^2)/p}{(1 - R_D^2)/(n - p)} \quad (9)$$

where

$$R_i^2(h, p) = 1 - \frac{\widehat{\text{MSFE}}_i(h, p)}{\hat{\gamma}(0)}, i = I, D.$$

is the forecastability index. The statistic (9) is the standard test for the  $p$  restrictions  $\phi_h^{(D)} = \phi_h^{(I)}$ , but it has not the usual  $F$  distribution in finite samples.

In the light of (3),

$$R_i^2(h, p) = 1 - \left[ \sum_j \hat{\nu}_{ij}^2 + 2 \sum_k \hat{\rho}(k) \sum_j \hat{\nu}_{ij} \hat{\nu}_{i, j+k} \right], i = D, I,$$

with  $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$  and  $\hat{\nu}_i(L) = S_{h-1}(L) + L^{h-1} \hat{\phi}_h^{(i)}(L)$ . It follows from the positive-definiteness of the tapered autocovariance sequence that  $R_i^2(h, p) \geq 0$  and  $R_D^2(h, p) \geq R_I^2(h, p)$ , so that  $F(h, p) \geq 0$ . The null of equal forecast accuracy will thus be rejected for "large" values of the test statistic.

The  $p$ -values of the finite sample distribution of the statistic (9) are obtained by the bootstrap method, using the *sieve bootstrap* to obtain replicates of the observed time series (see Bühlmann, 1997, 2002, and the references therein). The test procedure takes the following steps.

1. For a given pair  $(h, p)$  compute the direct and iterated predictors and the statistic  $\hat{F}(h, p)$  in (9).
2. Determine the AR order  $p^*$  of the one-step-ahead model ( $h = 1$ ) by selecting the value in  $(1, [n/10])$  that minimizes the Hurvich and Tsai (1989) corrected AIC given above in (8).
3. Estimate the AR coefficients model by the Yule-Walker method, solving  $\hat{\Gamma} \hat{\phi} = \hat{\gamma}_1$ , where  $\hat{\Gamma}, \hat{\gamma}_1$  contain either the standard or the tapered sample autocovariances.
4. Generate  $B$  bootstrap replicates of the series by sampling with replacement the centered innovations  $e_t - \bar{e}$ ,  $e_t = \Delta x_t - \sum_{j=1}^{p^*} \hat{\phi}_j \Delta x_{t-j}$ ,  $t = p^* + 1, \dots, n$ ,  $\bar{e} = (n - p^*)^{-1} \sum e_t$ , and computing recursively for  $t = p^* + 1, \dots, n$ ,  $x_t^{(b)} = x_{t-1}^{(r)} + \sum_{j=1}^{p^*} \hat{\phi}_j \Delta x_{t-j}^{(b)} + e_t^b$ , using the starting values  $x_{p^*}, \Delta x_j, j = 2, 3, \dots, p^*$ , where  $e_t^b, b = 1, \dots, B$ , is a draw from the empirical distribution of  $e_t - \bar{e}$ .
5. For each bootstrap replication compute the statistic  $F^{(b)}(h, p)$ . The distribution function of  $F^{(b)}(h, p)$ ,  $b = 1, \dots, B$ , is used to approximate the unknown distribution of the F-test statistic (9). Bootstrap  $p$ -values are obtained as the proportion of the bootstrap statistics  $F^{(b)}(h, p)$ , that are more extreme than the actual statistic  $\hat{F}(h, p)$  computed at the first step.

## 5 Long range forecasting and trend estimation: the multistep Beveridge-Nelson decomposition

Using the identity

$$X_{t+h} = X_t + \sum_{j=1}^h \Delta X_{t+j},$$

the  $h$ -step ahead predictor based on the information set available at time  $t$ , denoted  $\mathcal{F}_t$ , is obtained by adding to the current  $X_t$  all forecastable future changes up to time  $t + h$ , i.e.:

$$\tilde{X}_{t+h|t} = X_t + \sum_{j=1}^h \widetilde{\Delta X}_{t+j|t}, \quad (10)$$

where  $\widetilde{\Delta X}_{t+j|t} = E(\Delta X_{t+j} | \mathcal{F}_t)$ .

If  $h$  is allowed to go infinity in (10) and we assume that the drift is zero, then  $\tilde{X}_{t+h|t}$  tends to the Beveridge-Nelson (1981, BN henceforth) trend, or permanent, component, and  $\lim_{h \rightarrow \infty} \sum_{j=1}^h \widetilde{\Delta X}_{t+j|t}$  is minus the BN cycle (transitory component). In the case when the drift is nonzero,  $E(\Delta X_t) = \beta \neq 0$ , the BN trend is redefined as  $\tilde{X}_{t+h|t} - \beta h$ , which equals the current value of the series plus "all forecastable future changes beyond the mean rate of drift".

The direct and iterated AR methods provide two different approximations to  $\lim_{h \rightarrow \infty} \sum_{j=1}^h \widetilde{\Delta X}_{t+j|t}$ . As a matter of facts, the integration of all forecastable future changes up to time  $h$ ,  $\sum_{j=1}^h \widetilde{\Delta X}_{t+j|t}$ , is approximated by  $\phi_h^{(i)'} \Delta \mathbf{X}_t$ ,  $i = I, D$ , and thus the BN trend arising from both methods is

$$m_{it} = \lim_{h \rightarrow \infty} X_{t+h|t} = X_t + \lim_{h \rightarrow \infty} \phi_h^{(i)'} \Delta \mathbf{X}_t, \quad i = I, D. \quad (11)$$

Letting  $\phi_\infty^{(i)} = \lim_{h \rightarrow \infty} \phi_h^{(i)}$ , (an approximation to) the BN trend can be expressed as the following one sided moving average of the series:

$$m_{it} = (1 + \phi_{\infty,1}^{(i)})X_t + (\phi_{\infty,2}^{(i)} - \phi_{\infty,1}^{(i)})X_{t-1} + \cdots + (\phi_{\infty,p}^{(i)} - \phi_{\infty,p-1}^{(i)})X_{t-p+1} - \phi_{\infty,p}^{(i)}X_{t-p}. \quad (12)$$

The filter weights, which add up to one, can be obtained directly by letting  $h \rightarrow \infty$  in the expressions (5) and (6), respectively.

## 5.1 The BN trend for AR(1) predictors

In the AR(1) case ( $p = 1$ ), letting  $h \rightarrow \infty$  in (5) gives

$$\phi_\infty^{(D)} = \frac{g(0) - \gamma(0)}{2\gamma(0)} = \frac{1}{2}(\mathcal{P} - 1), \quad \mathcal{P} = \frac{g(0)}{\gamma(0)}.$$

The parameter  $\mathcal{P}$  is often referred to in the literature as the persistence parameter (being equal to the normalized spectral generating function at the zero frequency, or equivalently, to the ratio of the long run variance to the variance of  $\Delta X_t$ ).

In the iterated case  $\mathbf{T}$  is a scalar matrix; taking the limit of (6), and denoting  $\phi_1 = \phi_1^{(I)}$ ,

$$\phi_\infty^{(I)} = \frac{\phi_1}{1 - \phi_1} = \frac{\gamma(1)}{\gamma(0) - \gamma(1)},$$

since  $\phi_1 = \gamma(1)/\gamma(0)$ . The BN trend is  $m_{it} = (1 + \phi_\infty^{(i)})X_t - \phi_{\infty,p}^{(i)}X_{t-1} = X_t + \phi_\infty^{(i)}\Delta X_t$ .

If we assume that  $X_t$  is the IMA(1,1) process  $\Delta X_t = (1 + \theta L)\epsilon_t$ , then we have, respectively,

$$m_{It} = \frac{1}{1 - \rho(1)}X_t - \frac{\rho(1)}{1 - \rho(1)}X_{t-1}, m_{Dt} = (1 + \rho(1))X_t - \rho(1)X_{t-1}, \rho(1) = \frac{\theta}{1 + \theta^2}.$$

When  $\theta = 0$  ( $X_t$  is a pure random walk), the two expressions are equivalent. When  $\theta$  is equal to -1,  $X_t$  is white noise and the BN trends are, respectively,  $m_{It} = \frac{2}{3}X_t + \frac{1}{3}X_{t-1}$ ,  $m_{Dt} = \frac{1}{2}(X_t + X_{t-1})$ . Notice that  $m_{it}, i = I, D$ , can be regarded as estimators of the mean of the process and that the second is more efficient.

In general, the true BN trend is

$$m_t = \frac{1 + \theta}{1 + \theta L} X_t = X_t + \frac{\theta}{1 + \theta L} \Delta X_t,$$

so that the mean square error ratio equal to

$$\text{Eff}(\theta) = \frac{\text{Var}(m_{It} - m_t)}{\text{Var}(m_{Dt} - m_t)} = \frac{(\phi_\infty^{(I)} - \theta)^2 + (\theta \phi_\infty^{(I)})^2}{(\phi_\infty^{(D)} - \theta)^2 + (\theta \phi_\infty^{(D)})^2}.$$

The ratio is always greater than 1 for  $|\theta| \leq 1$ , except for  $\theta = 0$ , in which case it is exactly 1. For  $\theta = 1$  the direct approximation is twice as efficient; the maximum of the ratio is when  $\theta = 0.5$ , for which  $\text{Eff}(0.5) = 2.78$ . Finally,  $\text{Eff}(-1) = 1.11$ .

## 5.2 The iterated case

In the iterated case the coefficients  $\phi_\infty^{(I)}$  can be expressed in terms of the one-step AR polynomial coefficients,  $\phi_1^{(I)}$ . Intuitively, this is so since in (5) all the autocovariances beyond lag  $p$  are made dependent upon the first  $p$  autocovariances. Hence, we can derive an explicit limit for the iterated coefficients:

$$\phi_\infty^{(I)'} = \mathbf{e}_1' (\mathbf{I} - \mathbf{T})^{-1} \mathbf{T},$$

where the matrix  $\mathbf{T}$  was given in section 2, and depends solely on  $\phi_1^{(I)}$ .

*Theorem:* The BN trend implied by the indirect method can be expressed

$$m_{It} = \frac{\phi_1^{(I)}(L)}{\phi_1^{(I)}(1)} X_t. \quad (13)$$

*Proof:* The proof is direct. Writing for simplicity of notation  $\phi_1^{(I)} = \boldsymbol{\phi} = [\phi_1, \dots, \phi_p]'$ ,  $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ ,  $\phi(1) = 1 - \boldsymbol{\phi}' \mathbf{i}$ ,  $\mathbf{i} = [1, 1, \dots, 1]'$ , and defining  $\mathbf{C}$  as the matrix with unit elements on the main diagonal, -1 on the first subdiagonal and zero elsewhere, so that  $\mathbf{C}^{-1}$  is a lower triangular matrix with all elements equal to one (sometimes referred to as the random walk generating matrix),

$$\mathbf{I} - \mathbf{T} = \mathbf{C} - \mathbf{e}_1 \boldsymbol{\phi}', \quad (\mathbf{I} - \mathbf{T})^{-1} = \mathbf{C}^{-1} + \frac{1}{\phi(1)} \mathbf{i} \boldsymbol{\phi}' \mathbf{C}^{-1};$$

using  $(\mathbf{I} - \mathbf{T})^{-1} \mathbf{T} = (\mathbf{I} - \mathbf{T})^{-1} - \mathbf{I}$  and replacing into (12), yields the nice representation (13).

## 5.3 The BN trend at horizon $h$

The estimator  $m_{Dt}$  is clearly unfeasible, unless we know the true model that generated  $X_t$ . In fact, the example presented in section 5.1 postulated that the true model is IMA(1,1) and considered the long run forecast function implied by the AR(1) predictors for  $\Delta X_t$ .

Hence, the analytic form of  $m_{Dt}$  is only useful for theoretical discussion. Nevertheless, we can construct an approximation at horizon  $h$ , with  $h$  sufficiently large,  $m_{it}^{(h)} = X_t + \phi_h^{(D)'} \Delta \mathbf{X}_t$ , or, equivalently,

$$m_{Dt}^{(h)} = (1 + \phi_{h,1}^{(D)})X_t + (\phi_{h,2}^{(D)} - \phi_{h,1}^{(D)})X_{t-1} + \cdots + (\phi_{h,p}^{(D)} - \phi_{h,p-1}^{(D)})X_{t-p+1} - \phi_{h,p}^{(D)}X_{t-p}.$$

Obviously,  $m_{Dt}^{(h)} = m_{Dt}$  if  $X_t$  is an IMA(1, $q$ ) process with  $q \leq p$ .

Another possibility is to construct an estimate of the forecastable future changes of the series by deriving the one step ahead predictor implied by the  $h$ -step ahead coefficients, and applying the chain rule for forecasting any step ahead in the future. Hence, having obtained the  $h$ -step ahead AR prediction coefficients  $\phi_h^{(D)}$ , we can obtain the corresponding one-step ahead coefficients as those coefficients that, when propagated  $h$ -steps ahead by the chain rule, would produce exactly  $\phi_h^{(D)}$ . Denoting by  $\phi_h^*$  the vector of implied one-step coefficients, the above argument leads to the solution of the following nonlinear system of equations:

$$\phi_h^{(D)'} = \mathbf{e}_1'(\mathbf{I} - \mathbf{T}_h^h)(\mathbf{I} - \mathbf{T}_h)^{-1}\mathbf{T}_h$$

where  $\phi_h^{(D)}$  is known and

$$\mathbf{T}_h = \begin{bmatrix} \cdots & \phi_h^{*'} & \cdots \\ \mathbf{I}_{p-1} & \vdots & \mathbf{0} \end{bmatrix}$$

We can equivalently obtain  $\phi_h^*$  as the vector, containing the coefficients of the projection of  $\Delta X_{t+1}$  onto  $\Delta \mathbf{X}_t$ , that minimize the  $h$  step ahead prediction error variance. Hence, the model is the same as for the iterated method, i.e. a standard AR( $p$ ) autoregressive model, but the coefficients are obtained by minimizing the  $h$ -step ahead, rather than the one-step ahead, prediction error variance (this is sometimes referred to as *multistep estimation* of a standard AR model).

From  $\phi_h^*$  we construct the corresponding AR lag polynomial  $\phi_h^*(L)$ , and we obtain the following approximation, indexed by the forecast horizon  $h$ , of the BN trend:

$$m_{Dt}^{(h)*} = \frac{\phi_h^*(L)}{\phi_h^*(1)} X_t. \quad (14)$$

Obviously,  $\lim_{h \rightarrow \infty} m_{Dt}^{(h)*} = m_{Dt}$ .

## 5.4 The BN smoother

As shown in Proietti and Harvey (2000), when the true model is AR( $p$ ), under suitable conditions, there exists a two sided Beveridge-Nelson smoother, given by the following two-sided symmetric weighted average of the series:

$$\mu_{It} = \frac{\phi_1^{(I)}(L)\phi_1^{(I)}(L^{-1})}{[\phi_1^{(I)}(1)]^2} X_t = \frac{\phi_1^{(I)}(L^{-1})}{\phi_1^{(I)}(1)} m_{It}. \quad (15)$$

A sufficient condition for the interpretation of the BN smoother as the Wiener-Kolmogorov trend extraction filter for the decomposition into orthogonal components with uncorrelated disturbances, using the identifying assumption that the trend is a random walk and the cycle is stationary, is that the persistence parameter,  $[\phi_1(1)^{(I)}]^{-1}$ , is less than one.

For the BN smoother at forecast horizon  $h$  we have two possibilities. The first is to apply the BN smoother above using the implied AR( $p$ ) lag polynomial obtained by multistep estimation:

$$\mu_{Dt}^{(h)*} = \frac{\phi_h^*(L)\phi_h^*(L^{-1})}{[\phi_h^*(1)]^2} X_t = \frac{\phi_h^*(L^{-1})}{\phi_h^*(1)} m_{Dt}^{(h)*}. \quad (16)$$

For  $h \rightarrow \infty$  this estimator coincides with the final BN smoother estimator:

$$\mu_{Dt} = (1 + \phi_{\infty,1}^{(D)})m_{Dt} + (\phi_{\infty,2}^{(D)} - \phi_{\infty,1}^{(D)})m_{D,t+1} + \cdots + (\phi_{\infty,p}^{(D)} - \phi_{\infty,p-1}^{(D)})m_{D,t+p-1} - \phi_{\infty,p}^{(D)}m_{D,t+p}.$$

An alternative approximate BN smoother is obtained by replacing in the expression for the final BN smoother the quantities arising from  $h$ -step ahead estimation:

$$\mu_{Dt}^{(h)} = (1 + \phi_{h,1}^{(D)})m_{Dt}^{(h)} + (\phi_{h,2}^{(D)} - \phi_{h,1}^{(D)})m_{D,t+1}^{(h)} + \cdots + (\phi_{h,p}^{(D)} - \phi_{h,p-1}^{(D)})m_{D,t+p-1}^{(h)} - \phi_{h,p}^{(D)}m_{D,t+p}^{(h)}.$$

## 6 Illustrations

This section presents two illustrations dealing with two relevant macroeconomic indicators of the U.S. economy: quarterly real gross domestic product (logarithms, sample period: 1947.q1–2008.q2) and monthly inflation, obtained as  $x_t = \ln \text{CPI}_t - \ln \text{CPI}_{t-1}$ , where CPI is the consumer price index for all urban consumers released by the U.S. Bureau of Labor Statistics (seasonally adjusted, January 1960 - April 2008).

### 6.1 Gross domestic product

The top panel of figure 3 displays the percent gain in forecast accuracy arising from the direct method:

$$G(h, p) = 100 \times \left( 1 - \frac{\widehat{\text{MSFE}}_D(h, p)}{\widehat{\text{MSFE}}_I(h, p)} \right)$$

for  $h = 1, \dots, 40$  and  $p = 1, \dots, 15$ . The maximum gain, around 9%, is obtained for  $h = 32$  and  $p = 10$ . The second panel shows for each forecast lead time the order selected by Hurvich and Tsai's corrected AIC criterion (8);  $p$  is around 3 for small lead times and increases up to 12 for horizons around 7-8 years (28-32 quarters). For convenience, the plot also displays the efficiency of the direct predictor  $G(h, p)$  for the selected  $p$ .

The next question is whether the empirical accuracy gains are statistically significant. The bottom panel answers this by plotting against  $h$  the bootstrap p-values (using  $B = 9999$  replicates) of the test statistic  $F(h, p)$  in (9), where  $p$  is equal to the value selected by  $AIC_C$ . The steady line is drawn at the value 0.05. The plot reveals that none of these gains are significant at the 5% level, although the value for  $h = 32, p = 10$ , gets very close to it (the estimated p-value is 0.056).

Hence, it appears that the direct predictor does not outperform the iterated one, unless one is willing to adopt a very long horizon (8 years) and a significance level greater than 0.056. Suppose that we are willing to do that; the BN permanent and cyclical component that would originate from estimating the AR coefficients assuming a long run horizon are depicted in figure 4 and compared to the standard BN components, obtained from fitting an AR( $p$ ) model to the series by minimizing the one-step ahead prediction error variance. In particular, the top panels display the estimated trend  $\hat{m}_{It}$ , as given in (12) or (13), with the coefficients replaced by the tapered Yule-Walker estimates, and the deviations  $x_t - \hat{m}_{It}$ ; the one-step estimated BN

trend closely follows the observed GDP and the resulting BN cycle has small amplitude. The bottom panels display  $\hat{m}_{Dt}^{h*}$ , the trend obtained assuming the eight years forecast horizon,  $h = 32$ , and the corresponding cycle  $x_t - \hat{m}_{Dt}^{h*}$ . See section 5.3 for details.

The estimated trend  $\hat{m}_{Dt}^{h*}$  (which is the sample counterpart of (14)) is much smoother than its one-step trend; as a consequence, the estimated cycle has larger amplitude and displays the alternation of phases and the persistence that is characteristic of the U.S. business cycle. This is so since the estimation criterion enhances the long run properties of the estimated model, in that the filter (2) assigns more weight to the low frequencies and discard the high frequency fluctuations for estimating the AR coefficients.

## 6.2 Monthly inflation

The U.S. monthly inflation series is often modelled by an IMA(1,1) model, as in Stock and Watson (2007) and the references therein, with a negative MA coefficient. Hence, we expect that the AR representation is misspecified. Figure 5 presents the percent reduction in the prediction error variance obtained from using the direct predictor en lieu of the iterated predictor. The efficiency gains increase with the forecast horizon and with the order of the AR polynomial reaching a maximum at  $h = 41, p = 26$  (where it reaches around 20%).

The order  $p$  minimizing the corrected AIC is typically very large (see the middle panel of figure 5): for the one-step ahead predictor ( $h = 1$ ) it is already equal to  $p = 12$  and jumps to around 30 for  $h \geq 12$ . The finding that long autoregressions are required is consistent with the presence of a MA component close to the non invertibility region.

The reduction of the MSFE produced by the direct predictor is highly significant at all horizons greater than  $h = 16$ , as it is visible from the bottom panel of figure 5.

We thus turn to the implications of adopting the multistep direct AR model for the estimation of the underlying level of inflation. The top right panel of figure 6 displays the standard BN trend arising from the AR(12) model for  $\Delta x_t$  fitted by minimizing the one-step ahead prediction error variance. The second panel on the right depicts the BN smoothed trend computed according to the two-sided symmetric filter in (15).

These plots should be compared with the multistep BN and smoothed BN trends, estimated respectively using the sample counterpart of  $m_{Dt}^{(h)*}$ , see equation (14), and  $\mu_{Dt}^{(h)*}$ , given in (16). These estimates are characterized by a higher degree of smoothness, which is motivated by the fact that the estimates of the AR polynomial optimize the predictive performance at an horizon,  $h = 41$ , between 3 and 4 years. Consequently, the multistep BN trend, the estimation of the value the series would take if it were on its long-run path does not carry over to a great extent the high frequency fluctuations present in the series. The comparison of the real time and the smoothed estimates further reveals that the former suffer from a phase shift, due to the one sided nature of the signal extraction filter, which is not present in the smoothed estimates.

## 7 Conclusive remarks

We think that the paper can contribute to the already substantive literature on multistep estimation, and on the comparison of direct and iterated AR predictors, in the following ways.

- By providing an encompassing representation for the direct and iterated predictors that enables the derivation of the analytic mean square forecast error and recursive formulae for the AR coefficients.

- By proposing inferences (parameter estimates, bootstrap tests of equal predictive accuracy) based on the tapered autocovariance function. The estimation methodology has several advantages over ordinary least squares.
- By establishing the connection between multistep forecasting and trend estimation. In particular we introduced the multistep Beveridge-Nelson decomposition for the direct method and extended it to the two-sided BN decomposition.

There are several issues that we would like to address in our future research. As far as the estimation methodology is concerned, we can improve the sampling properties of the Yule-Walker estimates by multitapering, see Walden (2000); moreover, the class of Burg estimators (see Hurvich and Tsai, 1997, and Brockwell, Dahlhaus and Trinidade, 2005) deserves further investigation. While the extension to second order difference stationary, or  $I(2)$ , processes seems only labor intensive, the multivariate extensions are more promising. We find particularly attractive the notions of multistep impulse response functions and of multistep trends in a multiple time series, as devices that can provide a different characterization of the dynamic behavior and interrelations between time series.



## References

- Bhansali, R.J. (1996). Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics*, 48, 577-602.
- Bhansali, R.J. (2002). Multi-step forecasting, in Clements, M.P. and D. F. Hendry (eds.), *A Companion to Economic Forecasting*, Blackwell Publishers, Oxford, 206–221.
- Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, 3, 123–148.
- Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science*, 17, 52–72.
- Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. New York: Wiley.
- Bondon, P. (2001). Recursive relations for multistep prediction of a stationary time series. *Journal of Time Series Analysis*, 22, 309–410.
- Brockwell, P.J. and Dahlhaus, P. (2003). Generalized Levinson-Durbin and Burg algorithms. *Journal of Econometrics*, 118, 129-149.
- Brockwell, P.J., Dahlhaus, P. and Trinidad, A. (2005). Modified Burg algorithms for multivariate subset autoregression. *Statistica Sinica*, 15, 197-213.
- Clements, M.P. and Hendry, D.F. (1996). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*, 58, 657–684.
- Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21, 746-785.
- Cox, D.R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society, Series B*, 23, 414–422.
- Dahlhaus, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Annals of Statistics*, 16, 808-841.
- Doornik, J.A. (2006), *Ox. An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.
- Findley, D.F. (1983). On the use of multiple models for multi-period forecasting. *Proceedings of Business and Economic Statistics, American Statistical Association*, 528-531.
- Granger, C. W. J. and Jeon, Y. (2006). Dynamics of model overfitting measured in terms of autoregressive roots. *Journal of Time Series Analysis*, 27, 347-365.
- Granger, C. W. J. and Newbold, P. (1986). *Forecasting Economic Time Series*. Academic Press.
- Harvey, A.C. (1989), *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge, UK.
- Haywood, G. and Tunnicliffe Wilson, G. (1997), Fitting Time Series Models by Minimizing Multistep-ahead Errors: a Frequency Domain Approach, *Journal of the Royal Statistical Society, Series B*, 59, 237–254.

- Hurvich, C.M. (1988). A Mean Squared Error Criterion for Time Series Data Windows. *Biometrika*, 75, 485–490.
- Hurvich, C.M. and Tsai C.-L. (1997). Selection of a multistep linear predictor for short time series. *Statistica Sinica*, 7, 395–406.
- Ing, C.-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory*, 19, 254–279.
- Ing, C.-K. (2004). Selecting optimal multistep predictors for autoregressive process of unknown order. *Annals of Statistics*, 32, 693–722.
- Kang, H. (1987). The Tapering Estimation of the First-Order Autoregressive Parameters. *Biometrika*, 74, 643–645.
- Marcellino, M., Stock, J.H. and Watson, M. (2006). A comparison of direct and iterated multistep AR methods for forecasting microeconomic time series. *Journal of Econometrics*, 135, 499–526.
- Percival D., Walden A. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.
- Proietti, T., and Harvey, A.C. (2000). A Beveridge Nelson Smoother. *Economics Letters*, 67, 139–146.
- Shaman, P., and Stine, R.A. (1988). The Bias of Autoregressive Coefficient Estimators. *Journal of the American Statistical Association*, 83, 842–848.
- Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics*, 8, 147–164.
- Stock, J.H. and Watson, M. (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, Supplement to Vol. 39, No. 1, 13–33.
- Tiao, G. C., and Xu, D. (1993). Robustness of Maximum Likelihood Estimates for Multi-Step Predictions: The Exponential Smoothing Case. *Biometrika*, 80, 623–641.
- Tiao, G. C., and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series, *Journal of Forecasting*, 13, 109–131.
- Tjostheim, D. Paulsen J. (1983). Bias of some Commonly-Used Time Series Estimates, *Biometrika*, 70, 389–399.
- Velasco, C. and Robinson, P.M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, 95, 1229–1243.
- Walden, A.T. (2000). A unified view of multitaper multivariate spectral estimation. *Biometrika*, 87, 767–787.
- Weiss, A.A. (1991). Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics*, 48, 135–149.
- Zhou, Y.Y. and Roy A. (2006). Effect of tapering on accuracy of forecasts made with stable estimators of vector autoregressive processes. *International Journal of Forecasting*, 22, 169–180.

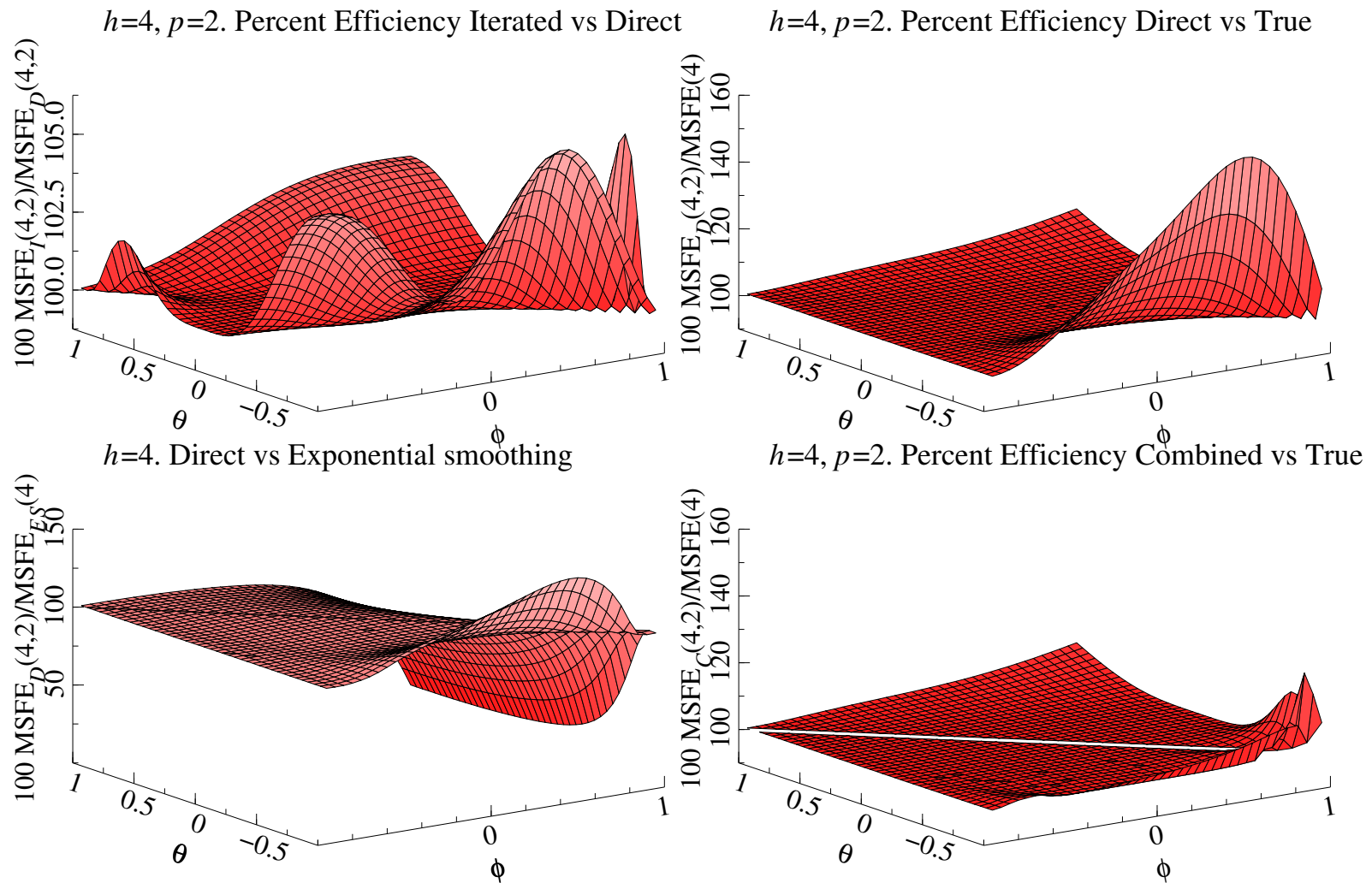


Figure 1: ARIMA(1,1,1) process. Comparison of the efficiency of the iterated and the direct AR predictors for horizon  $h = 4$  and order  $p = 2$ .

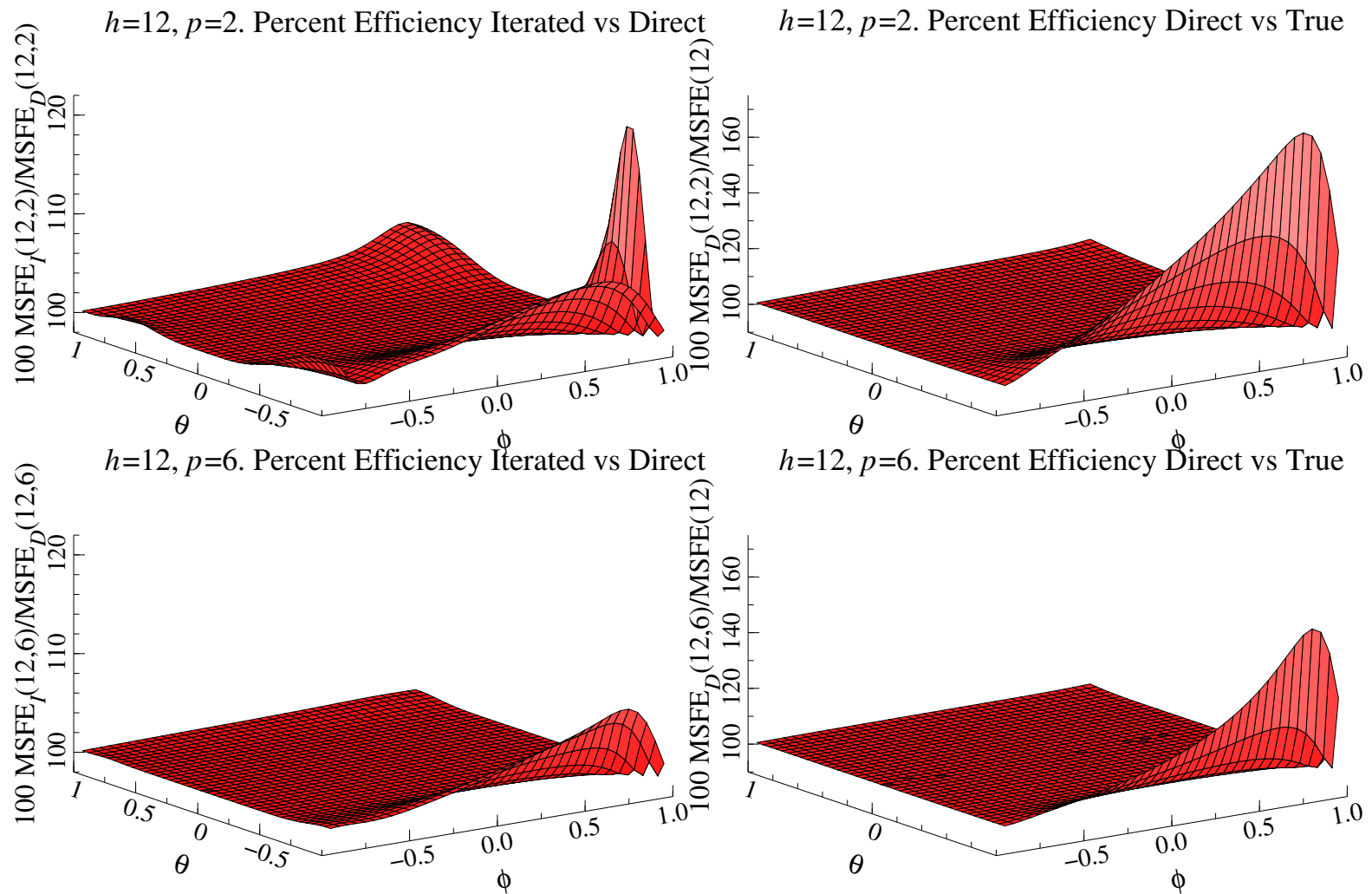


Figure 2: ARIMA(1,1,1) process. Comparison of the efficiency of the iterated and the direct AR predictors for horizon  $h = 12$  and order  $p = 2$  (top panels) and  $p = 6$  (bottom panels).

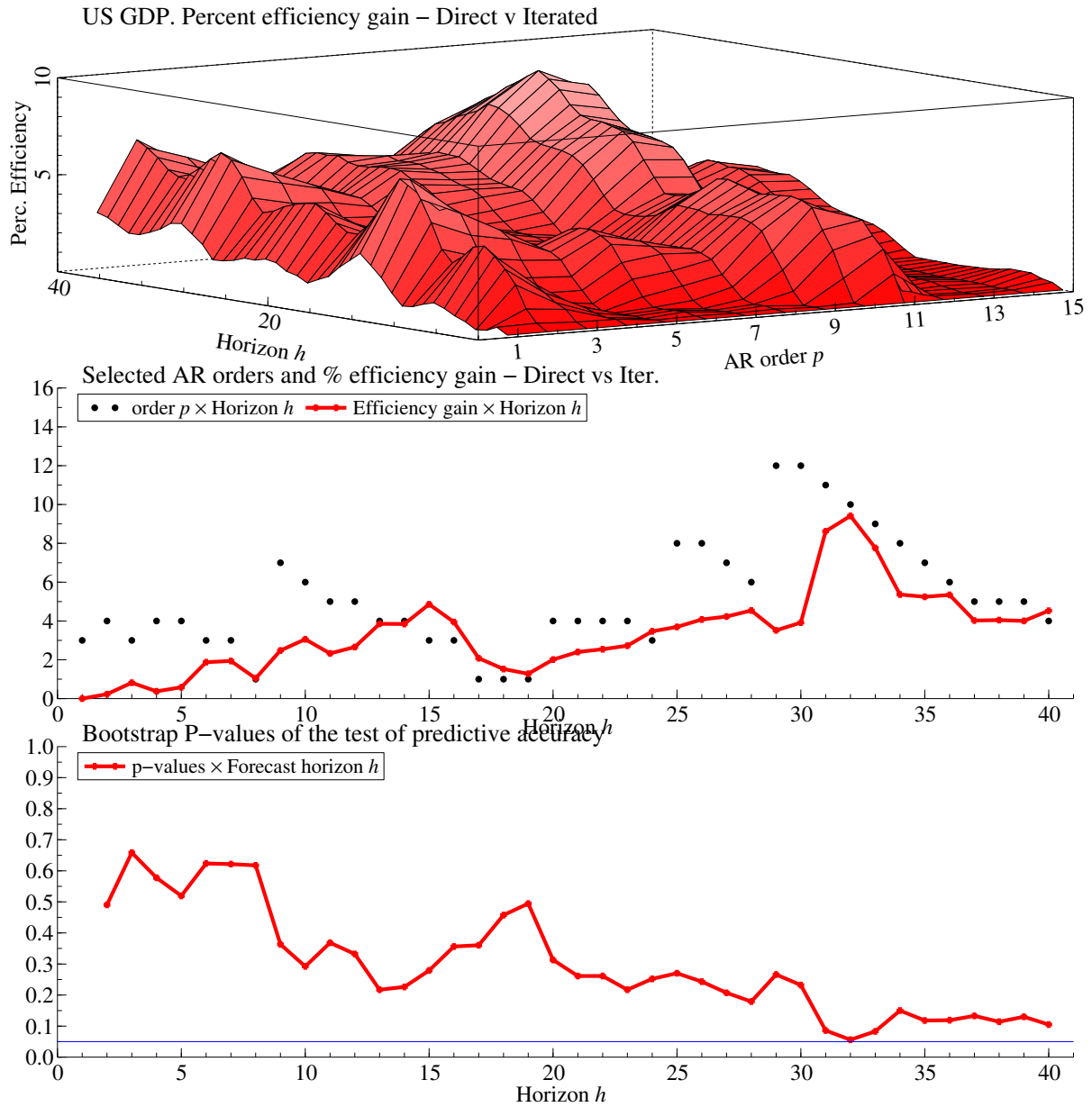


Figure 3: U.S. real gross domestic product. Percentage reduction in MSFE:  $100 \left( 1 - \widehat{MSFE}_D(h, p) / \widehat{MSFE}_I(h, p) \right)$  (top panel); selected AR orders and percent efficiency gain versus forecast horizon  $h$  (middle panel); bootstrap p-values of the predictive accuracy test statistic.

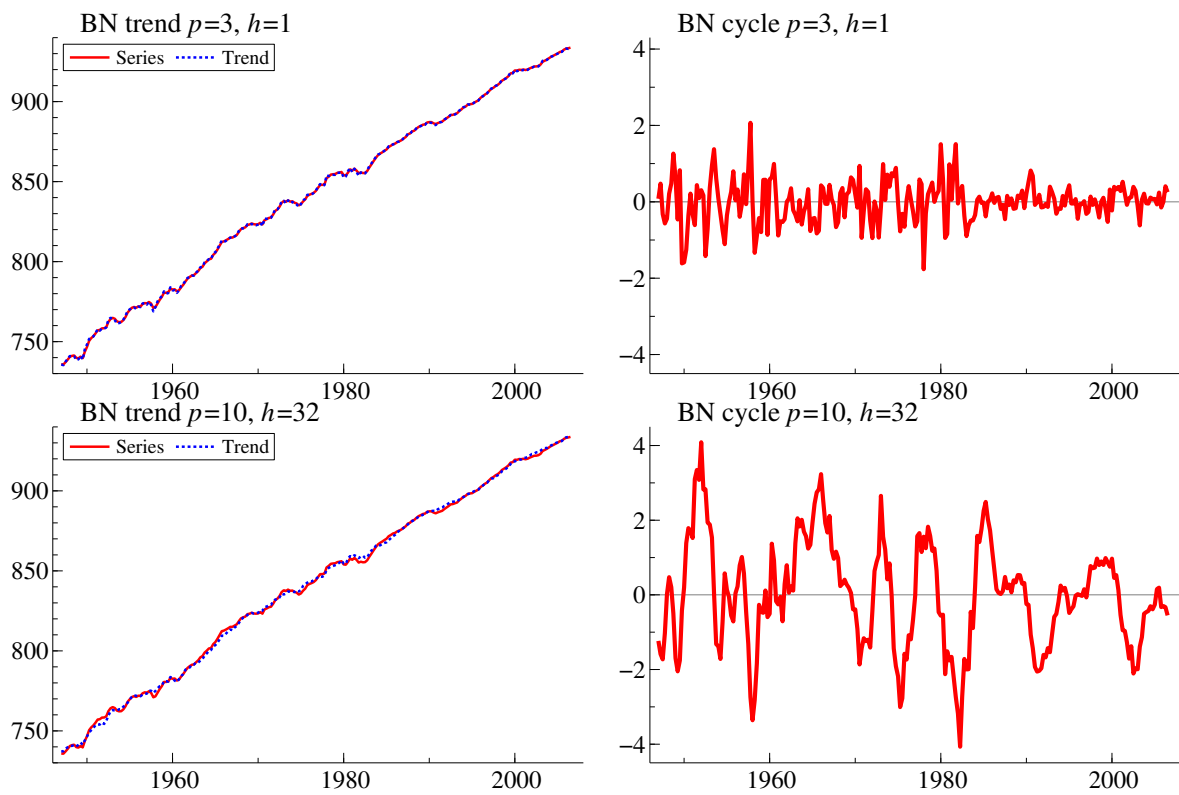


Figure 4: U.S. real gross domestic product. Beveridge-Nelson trends and cycles: standard (iterated) decomposition for  $h = 1$  and multistep direct decomposition at horizon  $h = 32$ . The AR order has been selected according to the modified AIC criterion.

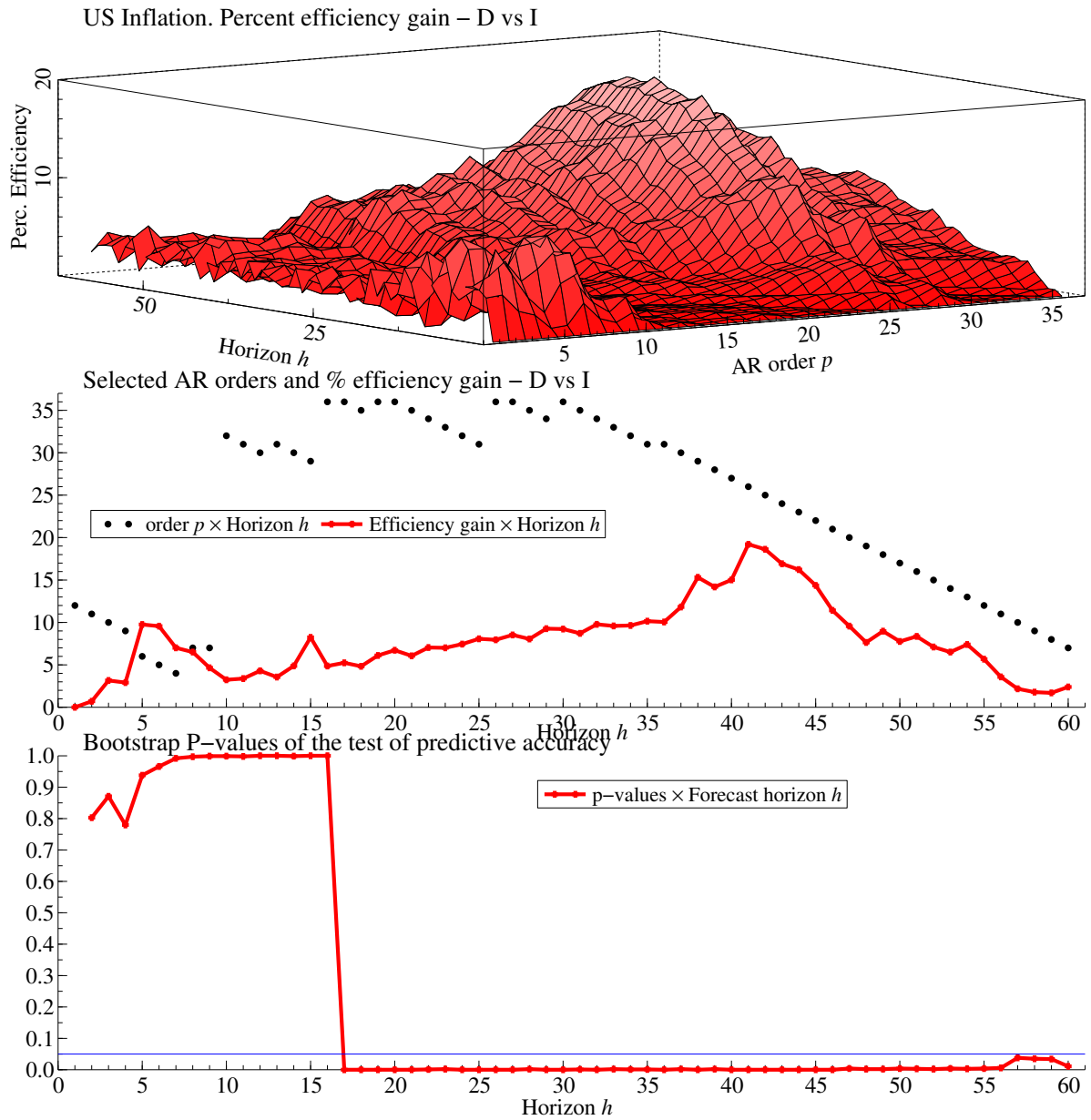


Figure 5: U.S. monthly inflation. Percentage reduction in MSFE:  $100 \left( 1 - \widehat{MSFE}_D(h, p) / \widehat{MSFE}_I(h, p) \right)$  (top panel); selected AR orders and percent efficiency gain versus forecast horizon  $h$  (middle panel); bootstrap p-values of the predictive accuracy test statistic.

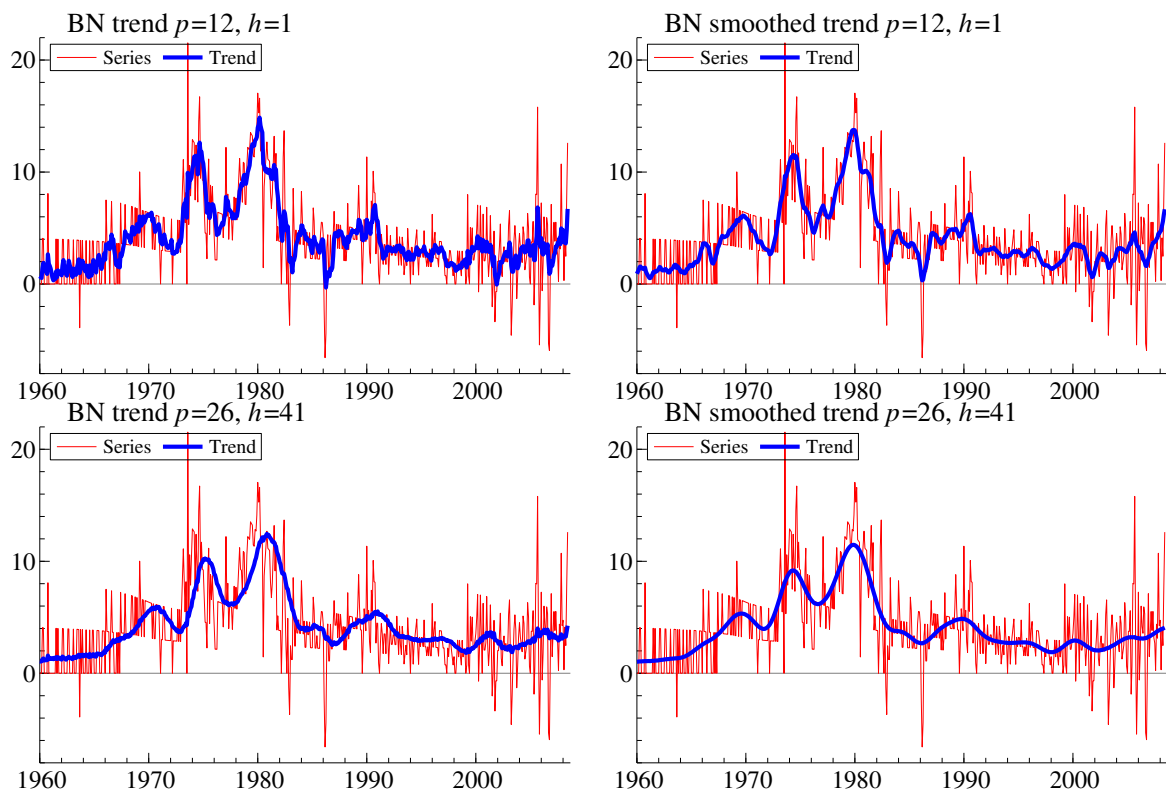


Figure 6: U.S. monthly CPI inflation. Beveridge-Nelson trends and smoothed (two-sided) trends: standard (iterated) decomposition for  $h = 1$  and multistep direct decomposition at horizon  $h = 41$ . The AR order has been selected according to the modified AIC criterion.