Direct and iterated multistep AR methods for difference stationary processes

Proietti, Tommaso

SEFEMEQ University of Rome "Tor Vergata"

1 October 2008

Direct and iterated multistep AR methods for difference stationary processes

Abstract

The paper focuses on the comparison of the direct and iterated AR predictors for difference stationary processes. In particular, it provides new methods for comparing the efficiency of the two predictors and for extracting the trend from macroeconomic time series using the two methods. The methods 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 are obtained with the bootstrap method.

Keywords: Multistep estimation. Tapered Yule-Walker estimates. Forecast combination.

JEL codes: C22, C51, C53.

1 Introduction

The direct and iterated autoregressive predictors play an important role in macroeconomic forecasting. This paper is concerned with the case when they are used to forecast the future levels of a difference stationary process, i.e. a process which is stationary in first differences. Typical occurrences are the level of the inflation rate and the level of gross domestic product.

In general, let us denote by $X_t, t = 1, 2, \ldots, n$, an integrated stochastic process, so that $\Delta X_t = X_t - X_{t-1}$ is a stationary process. Without generality loss we assume that the mean is zero. Our interest lies in predicting h-steps ahead the levels (rather than the differences) of the series. The *direct* (labelled by D henceforth) and *iterated* predictors (labelled by I), arise from the following linear projection:

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

where $\epsilon_{t+h|t}^{(i)}$ denotes the h-steps ahead prediction error.

It should be noticed that the two predictors use the same information set, represented by the vector $\boldsymbol{\Delta X}_t' = [\Delta X_t, \ \Delta X_{t-1}, \ldots, \Delta X_{t-p+1}]$, but differ in the definition of the coefficients $\phi_{jh}^{(i)}$. In particular, the *direct* predictor of X_{t+h} arises from the projection of $\Delta_h X_{t+h} = X_{t+h} - X_t$ on $\boldsymbol{\Delta X}_t$; 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* ΔX_{t+h}, which arises from projecting ΔX_{t+h} onto $\boldsymbol{\Delta X}_t'$.

The *indirect* (or *plug-in*, *iterated*) 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 ΔX_{t+k}, for $k = 1, \ldots, h$, 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 $\phi_{j1}^{(I)}, j = 1, \ldots, p$, 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 Chevillon (2007) for an up to date and comprehensive survey 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 emerges as the 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.

Section 5 illustrate the proposed methods using representative sample of U.S. macroeconomic time series. In section 6 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, $L^j X_t = X_{t-j}$:

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

3

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

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

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

where $\gamma(k) = \text{E}(\Delta X_t \Delta X_{t-k})$ is the autocovariance function of ΔX_t and ν_{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

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

with $|\nu_i(e^{-\iota\omega})|^2 = \nu_i(e^{-\iota\omega})\nu_i(e^{\iota\omega})$, the squared gain of the filter $\nu_i(L)$, and $g(\omega)$ represents the spectral generating function of ΔX_{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 $\phi_h^{(D)} = [\phi_{1h}^{(D)}, \ldots, \phi_{ph}^{(D)}]'$ are obtained by minimizing $\text{MSFE}(h,p)^{(D)}$ with respect to ϕ_h^D. The optimization problem leads to the following linear system of equations:

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

with

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

Notice that, from

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

it follows

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

4

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

The coefficients of the iterated predictor $\phi_{jh}^{(I)}, j = 1, \ldots, p$, in (2) are obtained recursively from the one-step-ahead coefficients. The latter are computed from the linear system $\phi_1^{(I)} = \phi_1^{(D)} = \Gamma^{-1}\gamma_1$:

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

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:

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

with starting value $\phi_1^{(I)} = \mathbf{T}'\mathbf{e}_1 = \Gamma^{-1}\gamma_1$.

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

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

we have

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

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 Γ 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 ΔX_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 $\mathrm{ER}_{ID}(4, 2) = 100 \times \mathrm{MSFE}_I(4, 2)/\mathrm{MSFE}_D(4, 2)$, as a function of the values of the AR and MA parameters ϕ and θ. 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 θ is close to -1 and ϕ 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 $\mathrm{ER}_{DT}(4, 2) = 100 \times \mathrm{MSFE}_D(4, 2)/\mathrm{MSFE}(4)$, where the denominator is the true MSFE of the ARIMA(1,1,1) optimal forecasts, $\mathrm{MSFE}(h) = \mathrm{E}\{[X_{t+h} - \mathrm{E}(X_{t+h}|\mathcal{F}_t)]^2\}$, 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 (ϕ 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,

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

where the weights sum to one and depend on a single smoothing constant, λ_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:

$$
\begin{aligned}
\epsilon_{t+h|t}^{(ES)} &= X_{t+h} - X_{t+h|t}^{(ES)} \\
&= X_t + \sum_{k=1}^{h} \Delta X_{t+k} - \sum_{j=0}^{\infty} \mathrm{w}_j X_{t-j} \\
&= \left[S_{h-1} + L^{h-1}\phi^{(ES)}(L)\right] \Delta X_{t+h}.
\end{aligned}
\tag{7}
$$

The lag polynomial $\phi^{(ES)}(L)$ is of infinite order and its coefficients satisfy the first order difference equation $\phi_j^{(ES)} = \phi_{j-1}^{(ES)} + \mathrm{w}_j$, with starting value $\phi_1^{(ES)} = \mathrm{w}_0 - 1$. Representation (7) follows directly from the fact that $\sum_{j=0}^{\infty} \mathrm{w}_j = 1$.

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 $100 \times \mathrm{MSFE}_D(4, 2)/\mathrm{MSFE}_{ES}(4)$ shows (see the bottom left panel of figure 1), the ES predictor outperforms the direct AR one when θ is close to -1 and greater that

$-\phi$. The reverse holds for $\phi > -\theta$. This finding opens the way to combining the forecasts. The MSFE of the combined predictor,

$$X^C_{t+h|t} = \tau X^D_{t+h|t} + (1 - \tau) X^{ES}_{t+h|t},$$

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 of the vector $(\mathbf{i}'\Sigma_h^{-1}\mathbf{i})^{-1}\Sigma_h^{-1}\mathbf{i}$, where Σ_h is the variance covariance matrix of the vector $[\epsilon^{(D)}_{t+h|t}, \epsilon^{(ES)}_{t+h|t}]'$. 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 $\mathrm{ER}_{ID}(12, 2)$, corresponding to different values of (ϕ, θ). The right panel illustrates that once again that for values of θ close to -1 and ϕ 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 θ 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 $x_t, t = 1, \ldots, n$, 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 $\sum_t (\Delta_h x_{t+h} -$

$\hat{\boldsymbol{\phi}}^{(D)'}\boldsymbol{\Delta x}_t)^2$, where $\boldsymbol{\Delta x}_t = [\Delta x_t, \Delta x_{t-1}, \ldots, \Delta x_{t-p+1}]'$. 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 $\widehat{\mathrm{MSFE}}_I(h, p) \geq \widehat{\mathrm{MSFE}}_D(h, p)$.

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 (Priestley, 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 $h_t, t = 1, \ldots, n$ 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:

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

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

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

The ϱ 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 autocovariances 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 improved 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).

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

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

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 ν_{ij}, which in turn depend on the autocovariance function of ΔX_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:

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

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),

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

9

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 form 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 $e_t - \bar{e}$, $e_t = \Delta x_t - \sum_{j=1}^{p^*} \hat{\phi}_j \Delta x_{t-j}, t = p^* + 1, \ldots, n, \bar{e} = (n - p^*)^{-1} \sum e_t$, and computing recursively for $t = p^* + 1, \ldots, n, x_t^{(b)} = x_{t-1}^{(r)} + \sum_{j=1}^{p^*} \hat{\phi}_j \Delta x_{t-j}^{(b)} + e_t^b$, using the starting values $x_{p^*}, \Delta x_j, j = 2, 3, \ldots, p^*$, where $e_t^b, b = 1, \ldots, B$, 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 $F^{(b)}(h,p)$, $b = 1, \ldots, B$, 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   Illustrations

This section illustrates the techniques proposed in the previous sections with reference to a small but representative subset of U.S. macroeconomic time series, available in the FRED®(Federal Reserve Economic Data) database. The series are listed in table 1. They are seasonally adjusted and analyzed in logarithms. All are considered difference stationary except for the price and earnings series, AHETPI, CPIAUCSL, GDPCTPI, GPDICTPI, PCECTPI, which are considered as integrated of order two. We assume that for these series we are interested in predicting their growth rate (e.g. in the case of CPIAUCSL x_t is the monthly inflation rate).

Tables 2 and 3 display, for different forecast horizons, the AR orders p^* that minimize the corrected AIC given in equation (8), along with the p-value of the bootstrap test of equal predictive accuracy (see section 4.3). The maximum p is 12 for monthly data, and 8 for quarterly data. All the computations have been carried out in Ox 4.00 by Doornik (2006). For solving the system $\hat{\Gamma}\hat{\phi}_h^{(D)} = \hat{\gamma}_h$, we use the functions for Toeplitz systems built in the package, which make use of the Levinson-Durbin algorithm.

Series ID	Title	Frequency	Sample
AHETPI	Average Hourly Earnings: Total Private Industries	M	1964:1-2008:12
AWHMAN	Average Weekly Hours: Manufacturing	M	1960:1-2008:12
CE16OV	Civilian Employment	M	1960:1-2008:12
CPIAUCSL	Consumer Price Index For All Urban Consumers: All Items	M	1960:1-2008:12
DSPIC96	Real Disposable Personal Income	M	1960:1-2008:12
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	M	1960:1-2008:12
INDPRO	Industrial Production Index	M	1960:1-2008:12
RSXFS	Retail Sales: Total (Excluding Food Services)	M	1992:1-2008:12
UNRATE	Civilian Unemployment Rate	M	1960:1-2008:12
DPIC96	Real Disposable Personal Income	Q	1947:1-2008:4
FPIC96	Real Private Fixed Investment, 3 Decimal	Q	1947:1-2008:4
GDPC96	Real Gross Domestic Product	Q	1947:1-2008:4
GDPCTPI	Gross Domestic Product: Chain-type Price Index	Q	1947:1-2008:4
GPDICTPI	Gross Private Domestic Investment: Chain-type Price Index	Q	1947:1-2008:4
PCECC96	Real Personal Consumption Expenditures	Q	1947:1-2008:4
PCECTPI	Personal Consumption Expenditures: Chain-type Price Index	Q	1947:1-2008:4

Table 1: Lists the time series used in the empirical analysis. Source: FRED®(Federal Reserve Economic Data) database.

The result confirm the findings of Marcellino, Stock and Watson (2008): in particular, there are no significant gains in predictive accuracy arising from the direct methods when time series dealing with the level of economic time series in real terms, such as GDP (GDPC96), industrial production (INDPRO), employment (CE16OV), the unemployment rate (UNRATE), hours worked (AWHMAN), real private fixed investment (FPIC96), real disposable income (DPIC96). An exception is provided by HOUST, for which the iterated forecasts are outperformed by the direct ones for short and long horizons.

On the contrary, for the inflation rate series, Δ AHETPI, Δ CPIAUCSL, Δ GDPCTPI Δ GPDICTPI, Δ PCECTPI, the direct method is more successful. Also, very large values of p are selected for the iterated predictor. This evidence is not surprising if we think that the U.S. monthly inflation series are often modelled by an IMA(1,1) model, as in Stock and Watson (2007) and the references therein, with a negative MA coefficient. Under these circumstances, we expect that the AR representation is misspecified; thus the order p minimizing the corrected AIC is typically very large and long autoregressions are required is consistent with the presence of a MA component close to the non invertibility region.

Figure 3 displays the percent gain in forecast accuracy arising from the direct method:

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

for 4 times series. In the first two cases the gains are small and not significantly different from zero. For HOUST and ΔGDPCTPI the gains are significant.

11

h	1	2	6	12	24	36	48
			Δ AHETPI				
p^*	11	10	11	12	8	10	7
p-value	-	0.92	0.99	0.95	0.79	0.76	0.87
			AWHMAN				
p^*	2	2	1	12	12	12	12
p-value	-	0.95	0.55	0.13	0.06	0.09	0.08
			CE16OV				
p^*	4	7	7	4	3	1	1
p-value	-	0.90	0.45	0.21	0.74	0.64	0.99
			Δ CPIAUCSL				
p^*	12	11	5	8	8	12	12
p-value		0.68	0.99	0.00	0.02	0.00	0.00
			DSPIC96				
p^*	3	2	1	1	1	1	1
p-value	-	0.57	0.74	0.90	0.79	0.97	0.98
			HOUST				
p^*	12	11	8	1	1	12	12
p-value	-	0.00	0.13	0.61	0.45	0.08	0.02
			INDPRO				
p^*	5	3	3	3	1	1	1
p-value	-	0.34	0.64	0.53	0.62	0.83	0.90
			RSXFS				
p^*	12	12	6	6	1	1	1
p-value	-	1.00	0.00	0.06	0.55	0.90	0.32
			UNRATE				
p^*	12	5	12	4	12	12	12
p-value	-	0.23	0.43	0.52	0.20	0.11	0.12

Table 2: U.S. monthly time series: comparison of direct and iterated predictors. AR orders selected by AIC and bootstrap p-values of the predictive accuracy test statistic.

h	1	2	3	4	8	12	16
				DPIC96			
p^*	2.00	4.00	1.00	1.00	1.00	1.00	1.00
p-value	-	0.23	0.19	0.67	0.76	0.82	0.93
				FPIC96			
p^*	1.00	3.00	3.00	5.00	8.00	8.00	8.00
p-value	-	0.64	0.77	0.10	0.11	0.07	0.09
				GDPC96			
p^*	3.00	4.00	3.00	3.00	1.00	5.00	3.00
p-value	-	0.47	0.62	0.71	0.60	0.28	0.35
				Δ GDPCTPI			
p^*	2.00	8.00	7.00	8.00	8.00	8.00	8.00
p-value	-	0.97	0.01	0.01	0.01	0.01	0.02
				Δ GPDICTPI			
p^*	7.00	8.00	7.00	8.00	8.00	8.00	8.00
p-value	-	0.97	0.01	0.04	0.05	0.03	0.07
				PCECC96			
p^*	4.00	4.00	4.00	4.00	1.00	1.00	1.00
p-value	-	0.43	0.26	0.45	0.82	0.92	0.61
				Δ PCECTPI			
p^*	2.00	8.00	8.00	8.00	8.00	8.00	8.00
p-value	-	0.98	1.00	0.00	0.00	0.00	0.00

Table 3: U.S. quarterly time series: comparison of direct and iterated predictors. AR orders selected by AIC and bootstrap p-values of the predictive accuracy test statistic.

6 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.

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. The extensions to a multivariate system of time series is also interesting.

References

Bhansali, R.J. (1996). Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics*, 48, 577-602.

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 Trinidade, 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.

Priestley, M.B. (1981), *Spectral Analysis and Time Series*. London, Academic Press.

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.
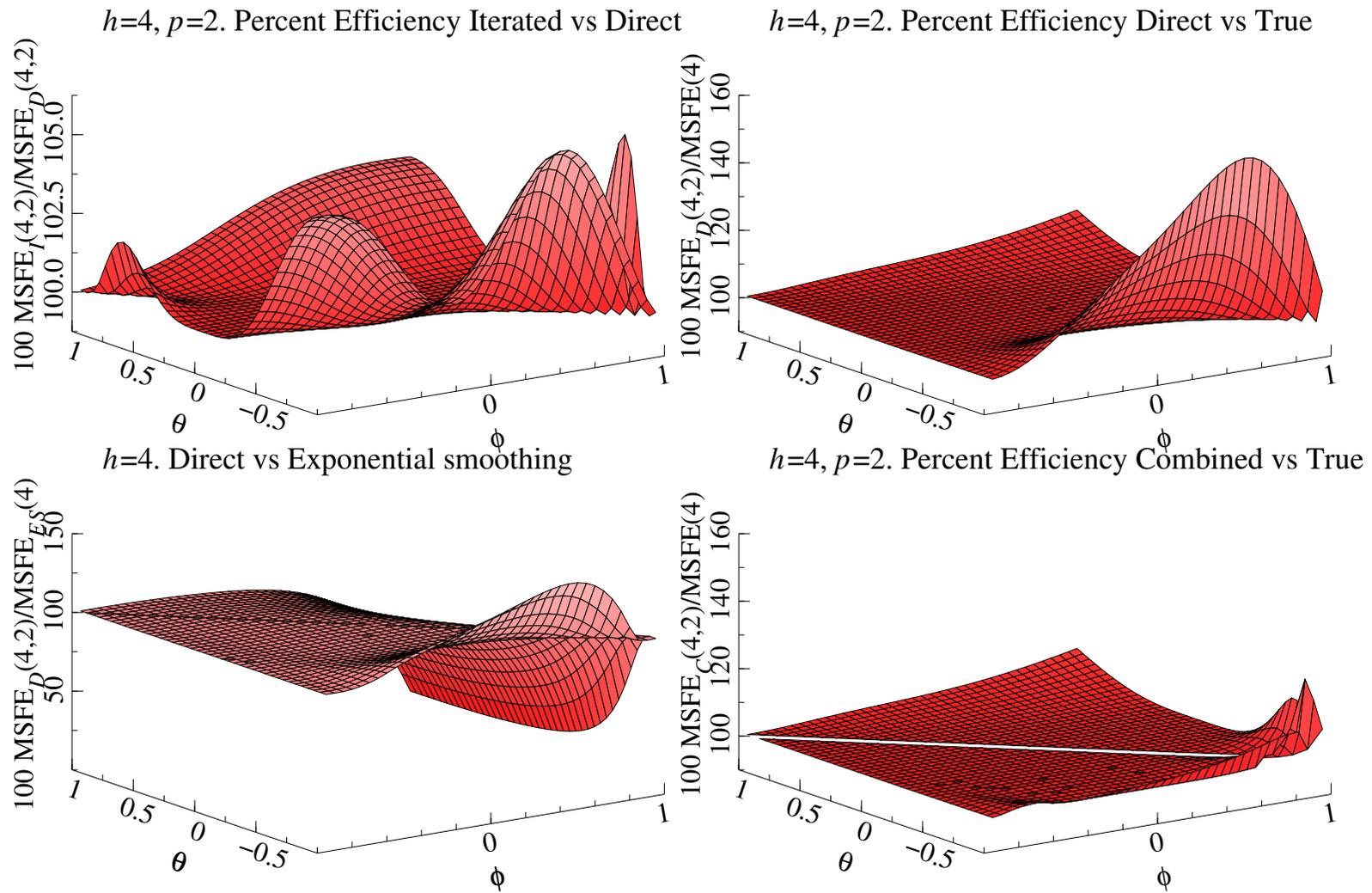
Figure 1: ARIMA(1,1,1) process. Comparison of the efficiency of the iterated and the direct AR predictors for horizon $h = 4$ and order $p = 2$.
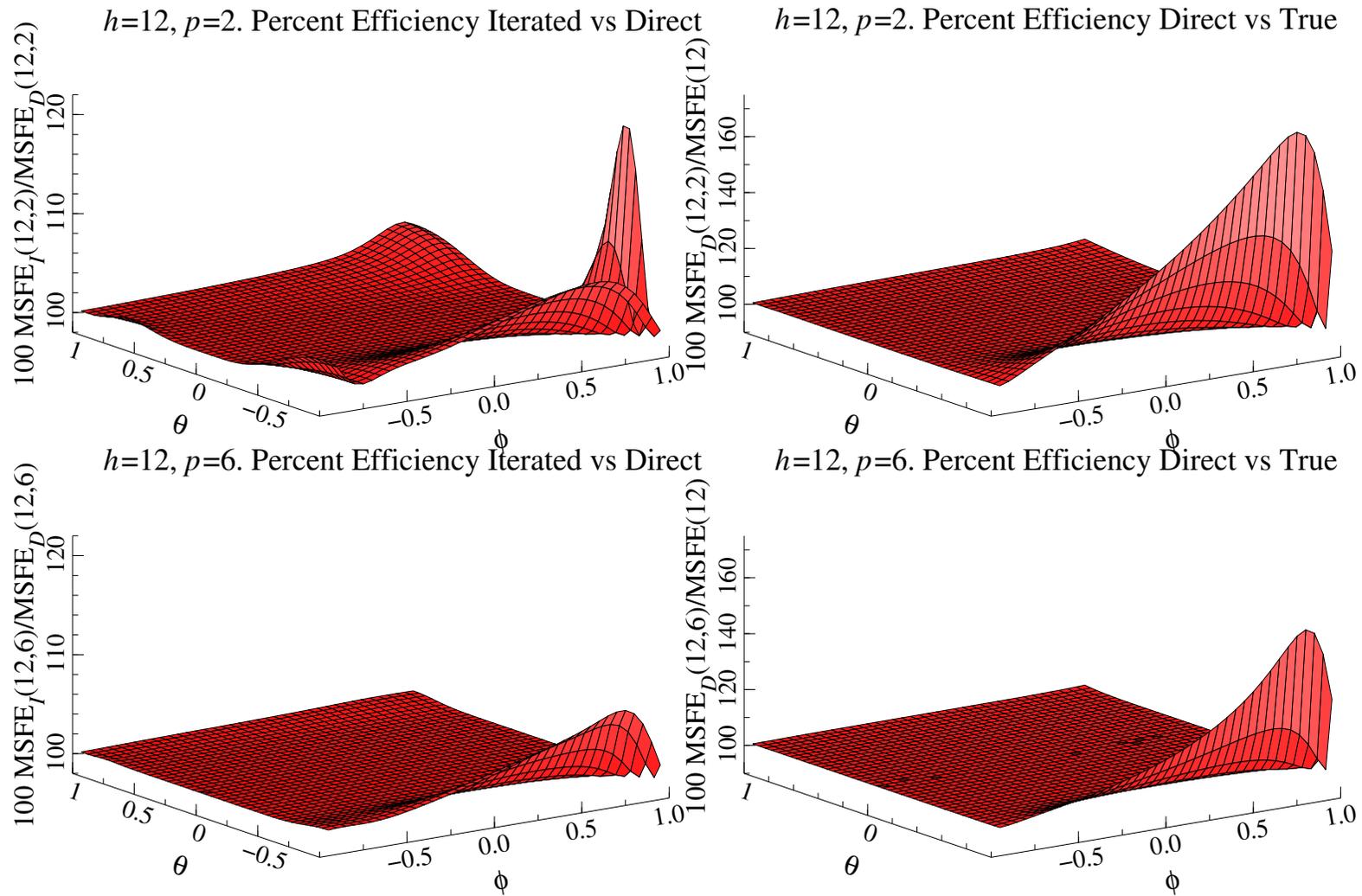
Figure 2: ARIMA(1,1,1) process. Comparison of the efficiency of the iterated and the direct AR predictors for horizon $h = 12$ and order $p = 2$ (top panels) and $p = 6$ (bottom panels).
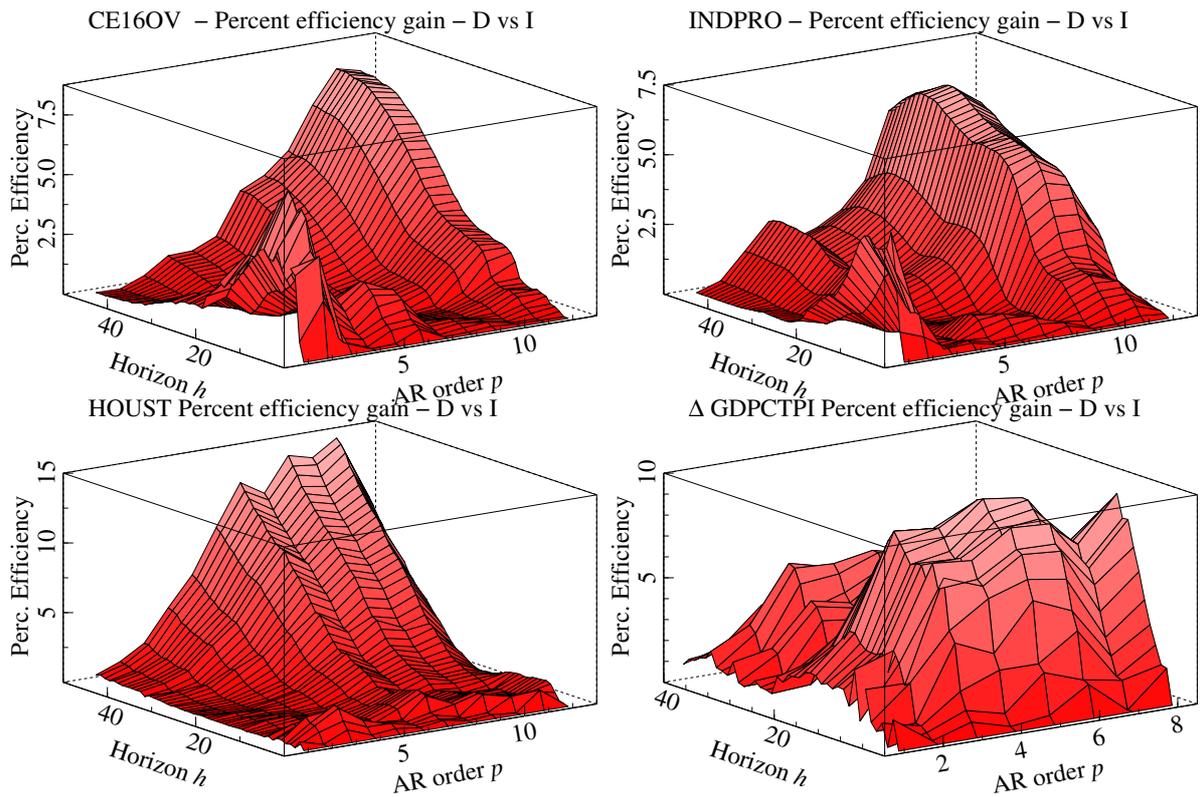
Figure 3: Comparison of direct and iterated AR predictors. Percentage reduction in MSFE: $100\left(1 - \widehat{\mathrm{MSFE}_D}(h,p)/\widehat{\mathrm{MSFE}_I}(h,p)\right)$. Series CE16OV (top left), series INDPRO (top right), series HOUST (bottom left), series Δ GDPCTPI (bottom right).