

Proietti, Tommaso

Working Paper

The Multistep Beveridge-Nelson Decomposition

EERI Research Paper Series, No. 24/2009

Provided in Cooperation with:

Economics and Econometrics Research Institute (EERI), Brussels

Suggested Citation: Proietti, Tommaso (2009) : The Multistep Beveridge-Nelson Decomposition, EERI Research Paper Series, No. 24/2009, Economics and Econometrics Research Institute (EERI), Brussels

This Version is available at:

<https://hdl.handle.net/10419/142556>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

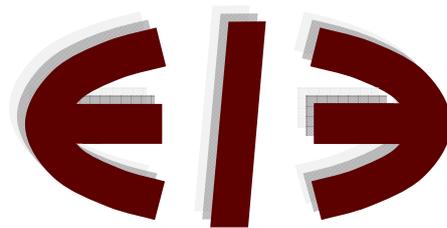
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Multistep Beveridge-Nelson Decomposition

Tommaso Proietti

EERI Research Paper Series No 24/2009

ISSN: 2031-4892



EERI
Economics and Econometrics Research Institute
Avenue de Beaulieu
1160 Brussels
Belgium

Tel: +322 299 3523
Fax: +322 299 3523
www.eeri.eu

The Multistep Beveridge-Nelson Decomposition

Tommaso Proietti*
University of Rome “Tor Vergata”

Abstract

The Beveridge-Nelson decomposition defines the trend component in terms of the eventual forecast function, as the value the series would take if it were on its long-run path. The paper introduces the multistep Beveridge-Nelson decomposition, which arises when the forecast function is obtained by the direct autoregressive approach, which optimizes the predictive ability of the AR model at forecast horizons greater than one. We compare our proposal with the standard Beveridge-Nelson decomposition, for which the forecast function is obtained by iterating the one-step-ahead predictions via the chain rule. We illustrate that the multistep Beveridge-Nelson trend is more efficient than the standard one in the presence of model misspecification and we subsequently assess the predictive validity of the extracted transitory component with respect to future growth.

Keywords: Trend and Cycle. Forecasting. Filtering.
JEL codes: C22, C52, E32.

*Address for Correspondence: Dipartimento SEFEMEQ, Università di Roma “Tor Vergata”, Via Columbia 2, 00133 Rome, Italy. *E-mail:* tommaso.proietti@uniroma2.it. A preliminary version of this paper was presented at the 5-th Eurostat Colloquium on Modern Tools for Business Cycle Analysis, Luxembourg 29 Sept. - 1st Oct. 2008, and at the Econometrics Seminar in Leicester, 15 Feb. 2009.

1 Introduction

The Beveridge-Nelson decomposition (BN, henceforth, Beveridge and Nelson, 1981), defines the trend component at time t as the value of the eventual forecast function at that time, or, equivalently, as the value that the series would take if it were on its long run path.

In the case of a difference stationary process with no drift, denoted X_t , such that the changes $\Delta X_t = X_t - X_{t-1}$ are stationary, letting \mathcal{F}_t denote the information set available at time t , the long run prediction equals the current level plus all forecastable future changes:

$$\begin{aligned}\lim_{t \rightarrow \infty} E[X_{t+h} | \mathcal{F}_t] &= \lim_{t \rightarrow \infty} E[X_t + \sum_{j=1}^h \Delta X_{t+j} | \mathcal{F}_t] \\ &= X_t + \sum_{j=1}^{\infty} E[\Delta X_{t+j} | \mathcal{F}_t].\end{aligned}$$

The resulting trend component is a random walk process, whereas the transitory component is a stationary process.

The role of the decomposition for characterizing the nature of macroeconomic fluctuations, and its relation to other unobserved components models, are discussed in Watson (1986), Morley, Nelson and Zivot (2003), Proietti (2006), Oh, Zivot and Creal (2008), and Morley (2009), among others. A recent issue of the Journal of Econometrics (JoE, Volume 146, Issue 2, October 2008), celebrating the 25th anniversary of the publication of the Beveridge and Nelson paper, featured some interesting extensions of this result in various directions.

One important finding, also remarked by the paper by Nelson (2008) opening the JoE issue, is that most of the variation in macroeconomic time series can be ascribed to permanent shocks, which are largely unpredictable, whereas the transitory component has very little amplitude. In particular, Nelson tests the predictive validity of the BN cycle, as well as other model-based cycle measures, with respect to future growth. It turns out that for U.S. real Gross Domestic Product (GDP), all the cyclical measures have little value for predicting economic growth in real time, and neither outperforms the BN cycle, which is significantly and negatively correlated with one-quarter-ahead output growth.

The univariate BN trend is typically obtained from a parametric model that is estimated by minimizing the one-step-ahead prediction error variance. The forecast of the future changes are obtained by iterating the one-step-ahead predictor, using the chain rule for multistep forecasting. When the model for ΔX_t is ARMA exact computational algorithms based on the state space representation of the ARMA model are available; see Proietti (1995) and Morley (2002). Under model and parameter uncertainty, the iterated predictor can be seen at best as an approximation to the conditional expectation of the future change, $E[\Delta X_{t+j} | \mathcal{F}_t]$, which is the optimal predictor for a symmetric square loss function. A better approximation could be provided by the so-called direct predictor, which optimizes the forecasting ability at a longer run horizon. Hence, reduced form models are short-run forecasting tools, and they may not be appropriate for forecasting the long-run.

The objective of this paper is to propose the multistep BN decomposition, which is the BN decomposition that arises when reference model is linear autoregressive and the out of sample predictions are obtained by the so-called direct method. As matter of fact, the notion of the BN decomposition is tightly bound up with long-run forecasting and a forecasting rule that optimizes

the predictability at longer horizons may be more suitable for the task of extracting the trend from a time series. A difference with the standard BN decomposition obtained from an autoregressive model for ΔX_t would emerge in the case of model misspecification, when the direct method is known to provide a better approximation to the true expectation $E[\Delta X_{t+j}|\mathcal{F}_t]$.

The paper sets off by reviewing the theory of forecasting using the direct and iterated predictors for difference stationary processes (section 2). The multistep BN decomposition is introduced and illustrated in section 3. Section 4 deals with the possibility of defining a two-sided symmetric filter for estimating the trend. In section 5 we present the estimation methodology. Section 6 provides two empirical illustrations 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 estimates smoother trends. Following Cogley (2002) and Nelson (2008) we then validate the BN cycles obtained by the indirect and direct AR methods by the effectiveness by which they predict future growth. In section 7 we summarize the contribution of the paper and draw our conclusions.

2 Direct and iterated forecasting

Let us assume that $\Delta X_t = X_t - X_{t-1}$ is a stationary zero mean process. Two important linear predictors of the levels or the series, the *direct* (labelled by D henceforth) and *iterated* predictors (labelled by I), are obtained by the following projection (see Marcellino, Stock and Watson, 2008):

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

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* ΔX_{t+h} , which arises from projecting ΔX_{t+h} onto $\Delta \mathbf{X}'_t$.

The *indirect* (or *iterated*) predictor is obtained from the AR(p) model by iterating the one-step-ahead predictor via the chain rule, so as to obtain forecasts of all the intermediate future changes ΔX_{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 $\Delta X_{t+k|t}^{(I)} = \sum_{j=1}^p \phi_{j1}^{(I)} \Delta X_{t+k-j|t}$ (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).

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 a comprehensive survey of the literature. 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.

The differences between 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 $\text{MSFE}_D(h, p)$ with respect to ϕ_h^D . The optimization problem leads to the following linear system of equations:

$$\mathbf{\Gamma} \phi_h^{(D)} = \gamma_h, \quad (2)$$

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

$$\phi_h^{(D)} = \phi_{h-1}^{(D)} + \phi^{(h)}, \quad \phi^{(h)} = \mathbf{\Gamma}^{-1} \gamma^{(h)}. \quad (3)$$

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 (1) recursively from the one-step-ahead coefficients, which are in turn 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

$$\mathbf{T} = \begin{bmatrix} \phi_1^{(I)} & \phi_2^{(I)} & \cdots & \phi_{p-1}^{(I)} & \phi_p^{(I)} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \quad (4)$$

The iterated AR coefficients satisfy the following first order recursion:

$$\phi_h^{(I)} = \phi_{h-1}^{(I)} + \mathbf{T}^{h'} \mathbf{e}_1, \quad (5)$$

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

An obvious 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 establishes a more general theorem, referring to the case when X_t is stationary, and taking into account the estimation uncertainty.

3 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.:

$$\tilde{X}_{t+h|t} = X_t + \sum_{j=1}^h \widetilde{\Delta X}_{t+j|t}, \quad (6)$$

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

If h is allowed to go infinity in (6) and we assume that the drift is zero, then $\tilde{X}_{t+h|t}$ tends to the BN 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, $\text{E}(\Delta X_t) = \beta \neq 0$, 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 fact, 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

$$m_{it} = \lim_{h \rightarrow \infty} X_{t+h|t} = X_t + \lim_{h \rightarrow \infty} \phi_h^{(i)'} \Delta \mathbf{X}_t, \quad i = I, D. \quad (7)$$

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:

$$m_{it} = (1 + \phi_{\infty,1}^{(i)})X_t + (\phi_{\infty,2}^{(i)} - \phi_{\infty,1}^{(i)})X_{t-1} + \cdots + (\phi_{\infty,p}^{(i)} - \phi_{\infty,p-1}^{(i)})X_{t-p+1} - \phi_{\infty,p}^{(i)}X_{t-p}. \quad (8)$$

The filter weights, which add up to one, can be obtained directly by letting $h \rightarrow \infty$ in the expressions (3) and (5), respectively.

3.1 The BN trend for AR(1) predictors

A simple example can be used to illustrate that in the case of model misspecification, the multistep BN trend is a more efficient estimator of the true underlying trend. Let us consider the AR(1) case ($p = 1$). Letting $h \rightarrow \infty$ in (3) 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 ΔX_t).

In the iterated case \mathbf{T} is a scalar matrix; taking the limit of (5), 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 $m_{it} = (1 + \phi_\infty^{(i)})X_t - \phi_\infty^{(i)}X_{t-1} = X_t + \phi_\infty^{(i)}\Delta X_t$.

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 θ 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

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

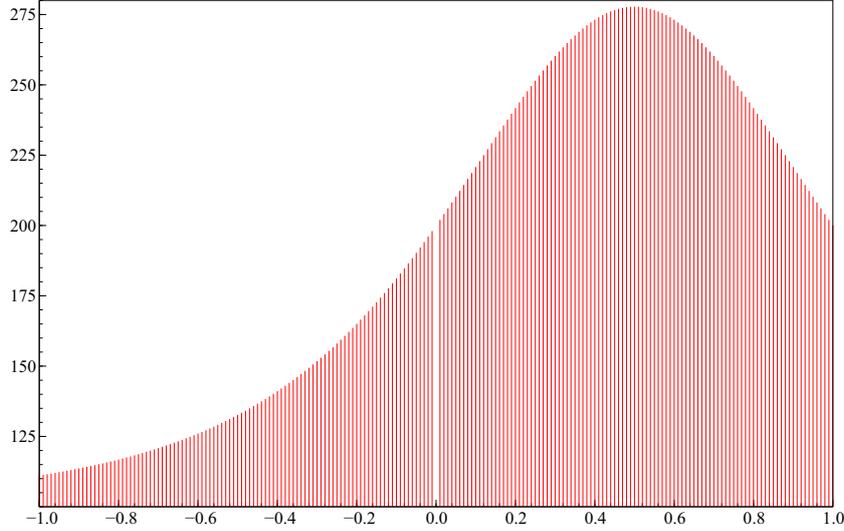


Figure 1: Efficiency of the multistep BN trend estimator versus the indirect AR(1) BN trend estimator: plot of $100 \times \text{Eff}(\theta)$ against the value of $\theta \in [-1, 1]$.

so that the mean square error ratio is 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 measures the precision of the direct method relative to that of the iterated one.

Figure 1 displays $100 \times \text{Eff}(\theta)$ against the value of $\theta \in [-1, 1]$. 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, $\text{Eff}(-1) = 1.11$.

3.2 A closed form expression for 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 (3) 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 (4), and depends solely on the elements of $\phi_1^{(I)}$.

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

$$m_{It} = \frac{\phi_1^{(I)}(L)}{\phi_1^{(I)}(1)} X_t. \quad (9)$$

Proof: The proof is direct. Writing for simplicity of notation $\phi_1^{(I)} = \phi = [\phi_1, \dots, \phi_p]'$, $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$, $\phi(1) = 1 - \phi' \mathbf{i}$, $\mathbf{i} = [1, 1, \dots, 1]'$, 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),

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

using $(\mathbf{I} - \mathbf{T})^{-1} \mathbf{T} = (\mathbf{I} - \mathbf{T})^{-1} - \mathbf{I}$ and replacing into (8), yields the nice representation (9).

3.3 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 3.1 postulated that the true model is IMA(1,1) and considered the long run forecast function implied by the AR(1) predictors for ΔX_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,

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

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 ϕ_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

$$\mathbf{T}_h = \begin{bmatrix} \dots & \phi_h^* & \dots \\ \dots & \dots & \dots \\ \mathbf{I}_{p-1} & \vdots & \mathbf{0} \end{bmatrix}$$

We can equivalently obtain ϕ_h^* as the vector, containing the coefficients of the projection of ΔX_{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 ϕ_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:

$$m_{Dt}^{(h)*} = \frac{\phi_h^*(L)}{\phi_h^*(1)} X_t. \quad (10)$$

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

4 The Beveridge-Nelson 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:

$$\mu_{It} = \frac{\phi_1^{(I)}(L)\phi_1^{(I)}(L^{-1})}{[\phi_1^{(I)}(1)]^2} X_t = \frac{\phi_1^{(I)}(L^{-1})}{\phi_1^{(I)}(1)} m_{It}. \quad (11)$$

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

$$\mu_{Dt}^{(h)*} = \frac{\phi_h^*(L)\phi_h^*(L^{-1})}{[\phi_h^*(1)]^2} X_t = \frac{\phi_h^*(L^{-1})}{\phi_h^*(1)} m_{Dt}^{(h)*}. \quad (12)$$

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

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

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:

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

5 Estimation issues

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 $\sum_t (\Delta_h x_{t+h} -$

$\hat{\phi}^{(D)'} \Delta \mathbf{x}_t)^2$, 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 iterated 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 (3) 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 (Priestley, 1981, p. 351, Tjøstheim 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.

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 $u = (t - 0.5)/n$,

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

The ϱ parameter, regulating the fraction of the initial and final stretch of data that are tapered, is set equal to 0.1. 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). 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.

The choice of the AR order p is according to the Hurvich and Tsai (1997) multistep generalization of the corrected AIC, given by

$$\text{AIC}_C(h, p) = n[\log \widehat{\text{MSFE}}_D(h, p) + 1] + 2(p + 1) \frac{n}{n - p - 2}. \quad (13)$$

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 :

$$F(h, p) = \frac{(R_D^2 - R_I^2)/p}{(1 - R_D^2)/(n - p)} \quad (14)$$

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 (14) 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. The p -values of the finite sample distribution of the statistic (14) 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).

6 Illustrations

This section presents two illustrations dealing with two relevant macroeconomic indicators of the U.S. economy: quarterly real gross domestic product (GDP, $100 \times$ logarithms, sample period: 1947.q1–2008.q4) and monthly inflation, obtained as $x_t = 100(\ln \text{CPI}_t - \ln \text{CPI}_{t-1})$, where CPI is the consumer price index for all urban consumers released by the U.S. Bureau of Labor Statistics (seasonally adjusted, January 1960 - December 2008). The series were downloaded from the FRED®(Federal Reserve Economic Data) database.

In both cases we first discuss whether the direct predictor provides a significant improvement in the predictive accuracy at a given horizon, we compare the standard one-step-ahead and the multistep BN decompositions, and finally we apply the predictive validity test proposed by Cogley (2002) and Nelson (2008), which aims at evaluating whether the BN transitory components contain information that is useful for predicting the future growth of GDP and inflation.

6.1 U.S. Gross Domestic Product

The top panel of figure 2 shows, for each forecast lead time h on the horizontal axis, the order selected by Hurvich and Tsai corrected AIC criterion given in (13); p is around 3 for small lead times and increases up to 12 for horizons around 7-8 years (28-32 quarters). The plot also displays the efficiency of the direct predictor for the selected p , measured by

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

i.e. percent gain in forecast accuracy arising from the direct method.

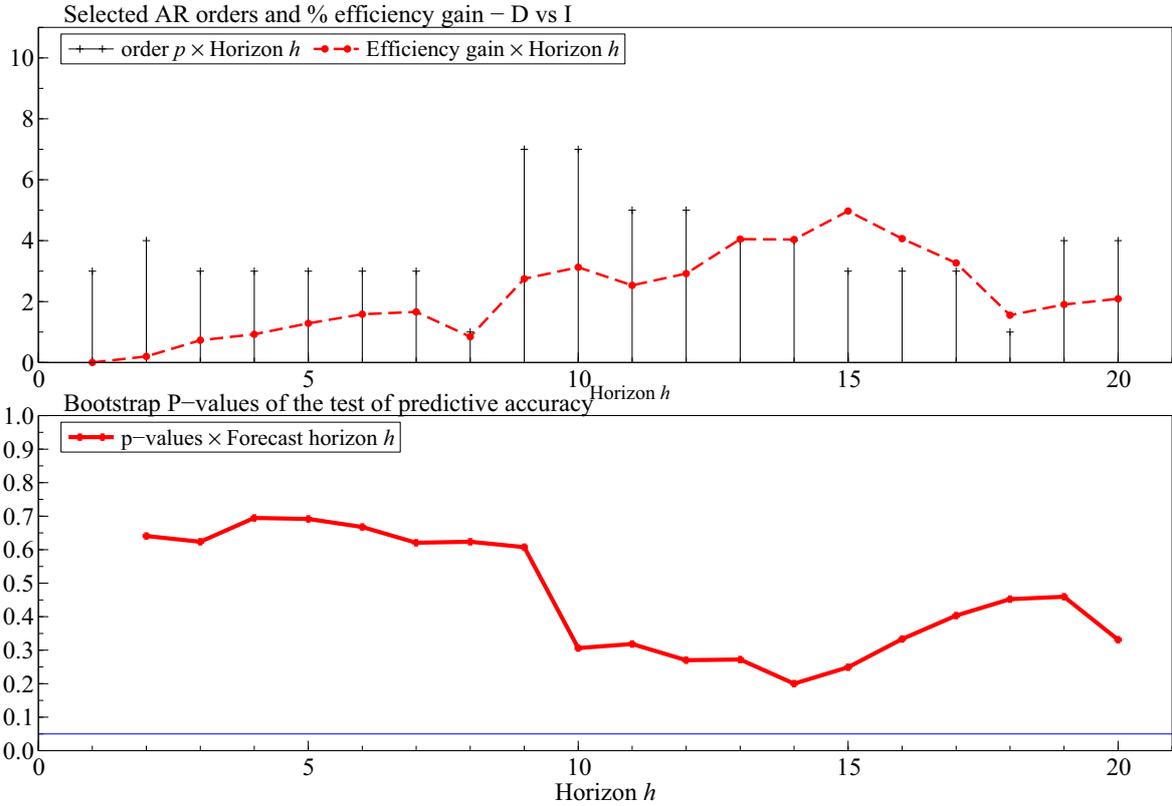


Figure 2: U.S. real gross domestic product. Selected AR orders and percent efficiency gain versus forecast horizon h (top panel); bootstrap p-values of the predictive accuracy test statistic (bottom panel).

The gains do not appear to be substantial; the maximum value, around 5%, is obtained for $h = 15$ and $p = 3$. 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 (14), 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 is significant at the 5% level.

We would like to remark at this point that our proposed test of equal forecasting accuracy serves as a preliminary screening for situations of potential interest, where the direct predictor can lead to a significant improvement in predictive accuracy, possibly since the $AR(p)$ representation is misspecified. From an applied standpoint, if one carries out a rolling forecasting exercise, using least squares for parameter estimation, it may well turn out that the direct predictor is actually less accurate. Hence, our results are not in contrast with those reported in Marcellino, Stock and Watson (2008); rather, they confirm that for U.S. GDP the direct predictor does not outperform the iterated one.

Be that as it may, it is nevertheless instructive to construct the multistep BN decomposition.

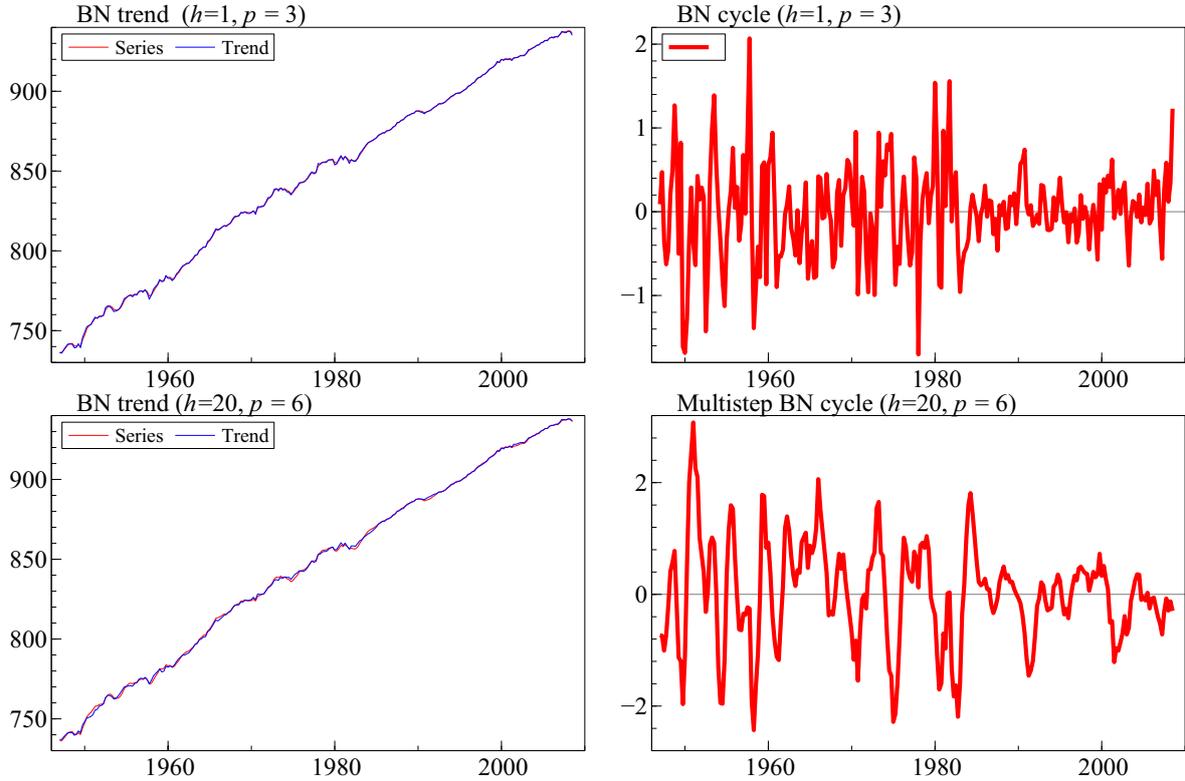


Figure 3: U.S. real gross domestic product. Beveridge-Nelson trends and cycles: standard (iterated) decomposition for $h = 1$ and multistep direct decomposition at horizon $h = 20$.

Figure 3 compares to the standard BN components, obtained from fitting an AR(3) model and the multistep ones at horizon $h = 20$ (5 years) and $p = 6$. In particular, the top panels display the estimated trend $\hat{m}_{I,t}$, as given in (8) or (9), with the coefficients replaced by the tapered Yule-Walker estimates, and the deviations $x_t - \hat{m}_{I,t}$; the one-step estimated BN trend closely follows the observed GDP and the resulting BN cycle has small amplitude. The bottom panels display $\hat{m}_{D,t}^{h*}$, the trend obtained assuming the eight years forecast horizon, $h = 32$, and the corresponding cycle $x_t - \hat{m}_{D,t}^{h*}$. See section 3.3 for details.

The estimated trend $\hat{m}_{D,t}^{h*}$ (which is the sample counterpart of (10)) is 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, say, of the Hodrick and Prescott (1999) estimate of the U.S. GDP business cycle.

To validate the predictive content of the BN transitory component we run the OLS regression of Δx_{t+1} on $x_t - \hat{m}_{i,t}$, $i = D, I$. For simplicity we report only the results for two representative values of p and h . The following table presents the correlation coefficient between the two series, the estimated regression coefficient and the associated t -value, and finally the coefficient of determination, R^2 of the regression.

	$h = 1, p = 3$	$h = 1, p = 7$
Correlation	-0.27	-0.19
Coefficient	-0.63	-0.47
t -value	-5.81	-3.89
R^2	0.07	0.04

	$h = 20, p = 3$	$h = 20, p = 7$
Correlation	-0.03	0.09
Coefficient	-0.05	0.07
t -value	-0.60	1.91
R^2	0.00	0.01

The results confirm Nelson’s overall conclusion that only a small fraction of future GDP growth is predictable using the BN transitory component. This fraction is significant only when $h = 1$, which confirms that the multistep decomposition plays no differential role in explaining the GDP fluctuations. It should be noticed that, as it is also evident from figure 3, the sign of the correlation is reversed.

6.2 U.S. Monthly Inflation

The U.S. monthly inflation series is often modeled 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 for this series.

In fact, the order p minimizing the corrected AIC is typically very large, as it can be seen from the top panel of figure 4): for the one-step ahead predictor ($h = 1$) it is already equal to $p = 14$ and jumps to around 32 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 $h = 18$, as it is visible from the bottom panel of figure 4. For horizons around 4 years the gain in predictive accuracy can reach up to 20%. It is also noticeable that the order p selected by AIC is negatively correlated with h .

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 5 displays the standard BN trend arising from the AR(14) model for Δx_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 (11). 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 (10), and $\mu_{Dt}^{(h)*}$, given in (12). 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, $h = 48$, i.e. 4 years of monthly observations. 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.

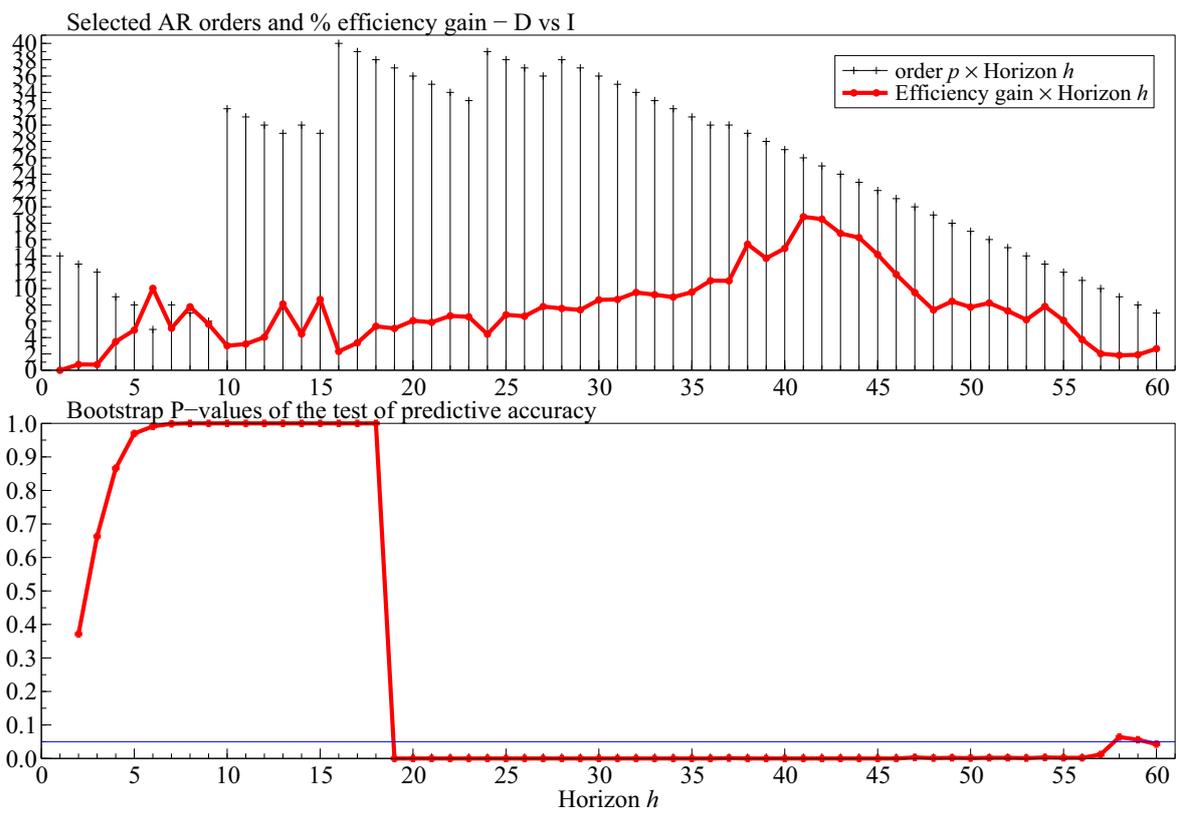


Figure 4: U.S. monthly inflation. Selected AR orders and percent efficiency gain versus forecast horizon h (middle panel); bootstrap p-values of the predictive accuracy test statistic.

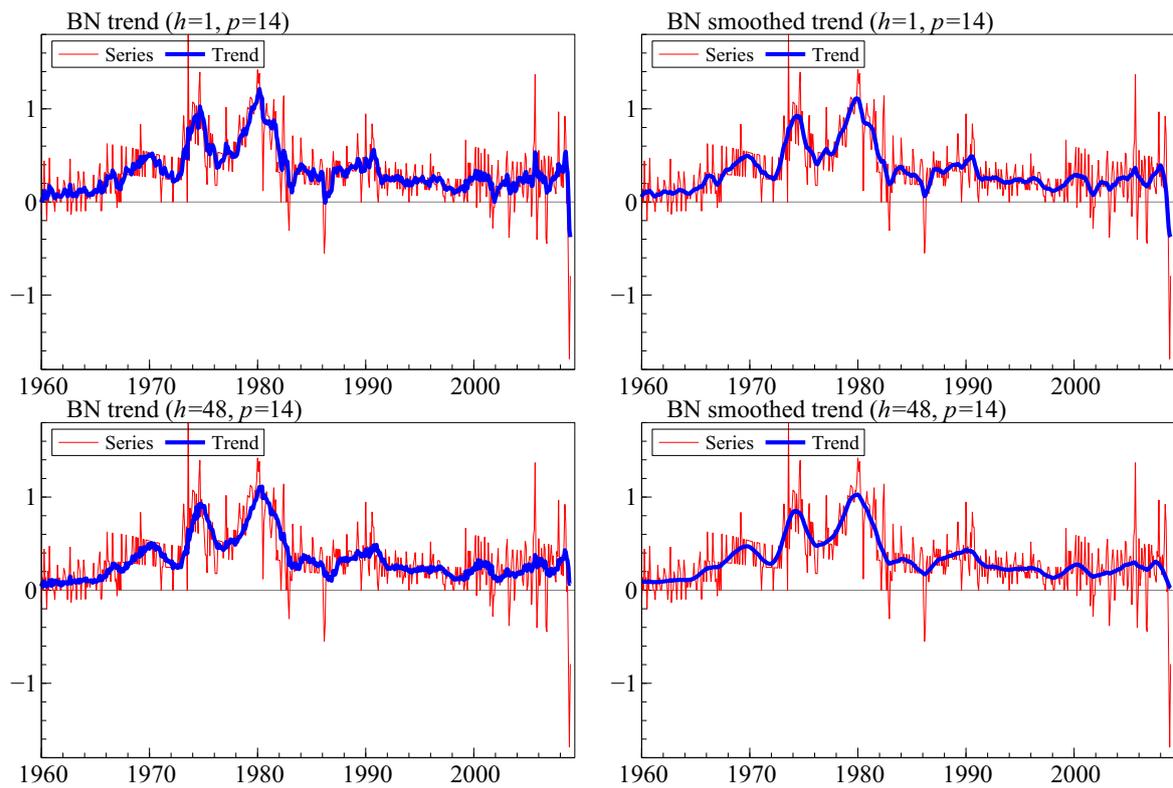


Figure 5: U.S. monthly CPI inflation. Beveridge-Nelson trends and smoothed (two-sided) trends: standard (iterated) decomposition for $h = 1$ and multistep direct decomposition at horizon $h = 48$.

Turning to the predictive content of the transitory component with respect to the next period change in inflation, the regression of Δx_{t+1} on the the BN transitory component at time t , $x_t - \hat{m}_{it}$, $i = I, D$, produced the following results, which refer to a few representative cases:

	$h = 1, p = 2$	$h = 1, p = 10$	$h = 1, p = 20$
Correlation	-0.39	-0.50	-0.50
Coefficient	-1.01	-0.69	-0.71
t -value	-10.15	-13.95	-13.83
R^2	0.15	0.26	0.25

	$h = 48, p = 2$	$h = 48, p = 10$	$h = 48, p = 20$
Correlation	-0.39	-0.51	-0.49
Coefficient	-0.63	-0.75	-0.61
t -value	-10.31	-14.38	-13.48
R^2	0.15	0.27	0.24

	$h = 60, p = 2$	$h = 60, p = 10$	$h = 60, p = 20$
Correlation	-0.39	-0.51	-0.50
Coefficient	-0.66	-0.66	-0.63
t -value	-10.34	-14.11	-13.72
R^2	0.15	0.26	0.25

The main evidence is that the transitory component has a large predictive power for the next change in monthly inflation. The best performance is the case $h = 48, p = 10$, for which $R^2 = 0.27$. Increasing the AR order from a low value ($p = 2$) to a moderately high value $p = 10$ yields a transitory component with higher predictive content. Increasing the forecast horizon helps improving the performance, but not as dramatically.

It should be noticed that if the true model was IMA(1,1), that is $\Delta X_t = \epsilon_t + \theta\epsilon_{t-1}$, $|\theta| < 1$, $\epsilon_t \sim \text{NID}(0, \sigma^2)$, the BN transitory component would be $\psi_t = -\theta\epsilon_t$, so that $\Delta X_{t+1} = -\psi_t + \epsilon_{t+1}$, which implies $E[\Delta X_{t+1}\psi_t] = -\theta^2\sigma^2$, the theoretical regression coefficient is -1, and $\text{Corr}(\Delta X_{t+1}, \psi_t) = \frac{\theta}{\sqrt{1+\theta^2}}$.

7 Conclusive remarks

The paper has introduced the multistep Beveridge-Nelson decomposition, which is the BN decomposition that arises when the long-run predictions are generated by the direct autoregressive predictor. We have also discussed how to construct a two sided BN decomposition. This result plays a role when the AR model is misspecified, so that minimizing the multistep prediction mean square error is likely to yield more accurate long-run predictions. The components can be validated according to their capability of predicting future growth, as proposed by Cogley (2002) and Nelson (2008).

The application to the U.S. GDP provide strong support to the conclusion by Nelson (2008), that much of the variation is due to permanent shocks, which are largely unpredictable: the BN

transitory component has negligible amplitude and has very low predictive content for the next period GDP growth rate. The multistep decomposition has even less predictive power and there is no statistical support for it.

However, the reverse is true when we consider the U.S. monthly inflation rate, for which about one fourth of the variation of the future changes of the inflation rate is explained by the BN cycle, and the multistep BN components yield more accurate estimates of the true underlying BN components.

References

- 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.
- Cogley, T. (2002), A simple adaptive measure of core inflation. *Journal of Money Credit and Banking*, 34, 94-113.
- 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.
- Hodrick R.J., and Prescott, E.C. (1997). Postwar U.S. Business Cycles: an Empirical Investigation. *Journal of Money, Credit and Banking*, 29, 1-16.
- 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.
- Morley, J., (2002), A state-space approach to calculating the Beveridge-Nelson decomposition, *Economics Letters* 75, 123-127.
- Morley, J., (2009), The Two Interpretations of the Beveridge-Nelson Decomposition. *Working paper*, Department of Economics, Washington University.
- Morley, J.C., Nelson, C.R. and Zivot E. (2003), Why are the Beveridge-Nelson and unobserved components decompositions of GDP so different? *The Review of Economics and Statistics*, 85, 235-243.
- Nelson, C.R. (2008), The BeveridgeNelson decomposition in retrospect and prospect. *Journal of Econometrics*, 146, 202–206.
- Oh, K.H., E. Zivot, and Creal, D. (2008) The Relationship between the Beveridge-Nelson decomposition and Unobserved Components Models with Correlated Shocks. *Journal of Econometrics*, 146, 207–219.

- 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.
- Proietti, T. (1995). The Beveridge-Nelson decomposition: Properties and extensions, *Statistical Methods and Applications*, 4, 101-124.
- Proietti, T. (2006). Trend-cycle decompositions with correlated components, *Econometric Reviews*, 25, 61-84.
- 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.
- Watson, M.W. (1986). Univariate detrending methods with stochastic trends. *Journal of Monetary Economics*, 18, 49-75.
- Weiss, A.A. (1991). Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics*, 48, 135-149.