An Alternative to Transfer Function Forecasting Based on Subspace Methods

Victor Gomez, Felix Aparicio-Perez, Angel Sanchez-Avila

An Alternative to Transfer Function Forecasting Based on Subspace Methods

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2008-0260.R1
Manuscript Type:	Review Article
Date Submitted by the Author:	23-Aug-2010
Complete List of Authors:	Gomez, Victor; Ministerio de Economia y Hacienda Aparicio-Perez, Felix; Instituto Nacional de Estadística Sanchez-Avila, Angel; Ministerio de Economia y Hacienda
Keywords:	ARMAX; forecasting; Kalman Filter;, Subspace methods; state space model; , transfer function;
Abstract:	In the time series literature recent interest has focused on the so--called subspace methods. These techniques use canonical correlations and linear regressions to estimate the system matrices of an ARMAX model expressed in state space form. In this article, we use subspace methods to forecast two series with the help of some exogenous variables related to them. We compare the results with those obtained using traditional transfer function models and find that the forecasts obtained with both methods are similar. This result is very encouraging because, in contrast to transfer function models, subspace methods can be considered as almost automatic.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

AN ALTERNATIVE TO TRANSFER FUNCTION FORECASTING BASED ON SUBSPACE METHODS

Víctor Gómez, Ministerio de Economía y Hacienda, Spain

Félix Aparicio-Pérez, Instituto Nacional de Estadística, Spain

Ángel Sánchez-Ávila, Ministerio de Economía y Hacienda, Spain

Corresponding author: Víctor Gómez

Address:

Ministerio de Economía y Hacienda, Dirección Gral. de Presupuestos
Subdirección Gral. de Análisis y P. E.
Alberto Alcocer 2, 1-P, D-34, 28046, Madrid, Spain
vgomez@sgpg.meh.es

Short title: FORECASTING BASED ON SUBSPACE METHODS

Key Words: ARMAX; forecasting; Kalman Filter; subspace methods; state space model; transfer function.

ABSTRACT

In the time series literature recent interest has focused on the so–called subspace methods. These techniques use canonical correlations and linear regressions to estimate the system matrices of an ARMAX model expressed in state space form. In this article, we use subspace methods to forecast two series with the help of some exogenous variables related to them. We compare the results with those obtained using traditional transfer function models and find that the forecasts obtained with both methods are similar. This result is very encouraging because, in contrast to transfer function models, subspace methods can be considered as almost automatic.

1. INTRODUCTION

This article presents two empirical studies of forecasting using subspace methods. In the first study, we forecast the Spanish Consumer Price Index (SCPI), base year 2001, whereas in the second study we forecast the frequency deviations considered by Jenkins and Watts (1968) and studied also by Reinsel (1997). In both cases we use traditional transfer function models as the benchmark. For forecasting the SCPI we make use of three exogenous inputs that are believed to have some information about SCPI. These are the Spanish Producer Price Index for Consumer Goods (base year 2000), Spanish Import Price Index for Consumer Goods (base year 2000), and the Spot Prices of Crude Oil in pesetas of the Brent barrel. This last series is considered as deterministic and its forecasts are the corresponding Futures Market Prices. In the second study, it is assumed that the two series of in–phase and out–of–phase current deviations are inputs to the system and the frequency deviations are an output. All of the variables in the second study are expressed as deviations from nominal values.

The SCPI series and the three aforementioned inputs can be seen in Figures 1 and 2. The data cover the period 1993:1–2006:3, except for the Spanish Import Price Index for Consumer Goods that extends from 1993:1 to 2006:2.

The data in the second example consists of 100 observations on the three variables and they have been coded by multiplying by 10.

The aim of the article is to gain some insight into the potential of subspace methods for forecasting the two output variables, SCPI and frequency, using some exogenous inputs. We compare the forecasts obtained with the subspace method with those given by a traditional transfer function model used as the benchmark. The forecasts are generated with a rolling estimator. In the first example, the rolling estimator begins with January 2004 as the initial origin of forecast and ends with December 2004 as the final origin. In the second example, it begins with observation number 74 as the initial origin of forecast and ends with observation number 85 as the final origin.

The results show that the forecasts obtained with subspace methods are similar to those

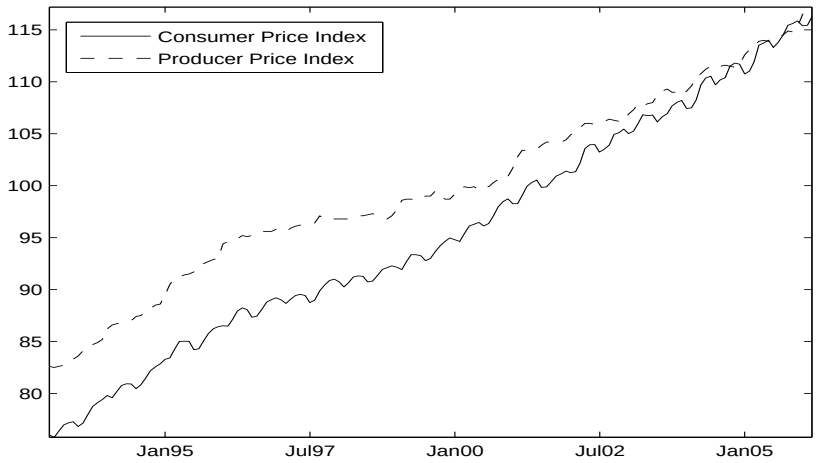


Figure 1: Spanish Consumer Price Index (SCPI) and Spanish Producer Price Index for Consumer Goods, 1993:1–2006:3.

of the benchmark. Given that the identification of a transfer function model is not easy, that it requires considerable expertise in time series analysis, and that subspace methods are kind of black box models, we believe that the results are very encouraging.

Subspace methods are identification methods for state space models with exogenous inputs that bypass the traditional difficulties associated with specifying the canonical form and maximum likelihood estimation of such models. It seems that the most promising of subspace methods and the one that is more easily understood from a statistical point of view is the one based on canonical correlation analysis (CCA), proposed as CVA (canonical variate analysis) by Larimore (1983). The CCA method is based on a singular value decomposition (SVD), that is very reliable numerically, to identify the system order and some regressions to estimate the system matrices. The method is suboptimal but gives estimators with good quality in general.

In this article, we apply the CCA subspace method to identify and estimate the state space models with exogenous inputs that will be used to forecast the two outputs, SCPI and frequency.

Subspace methods have been widely used in engineering for some time and only recently

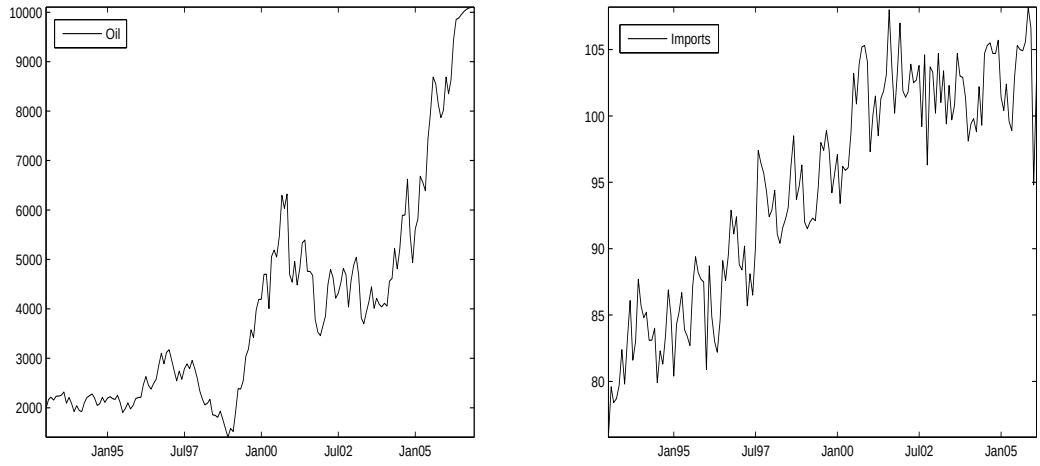


Figure 2: Spot Prices of Crude Oil in pesetas of the Brent barrel, 1993:1–2006:3, and Spanish Import Price Index for Consumer Goods, 1993:1–2006:2.

have econometricians started to pay attention to them. One potential advantage of subspace methods over the popular vector autoregression (VAR) approximations to vector autoregressive moving average (VARMA) models used in econometrics is that the state space models identified by them have minimal order and take the moving average part into consideration. In addition, subspace methods are not computationally expensive because they are based on regressions and the computation of an SVD.

A traditional method for forecasting time series, and economic time series in particular, is to use transfer function models. These models were popularized by Box and Jenkins (1976) in their seminal work. However, the method proposed by these authors to identify transfer function models with one input, based on the so–called prewhitening of the input and output, has a series of drawbacks. For example, a model for the input series is necessary and it is not easy to generalize the method for multiple–input models.

For these reasons, Liu and Hanssens (1982) and Tsay (1985) have proposed alternative methods for transfer function identification that seem to work well in practice and can be used with several inputs.

The methodology used in this article to identify the transfer function model makes use

of some of the procedures proposed by Liu and Hanssens (1982) and Tsay (1985).

The outline of the article is as follows. Section 2 briefly reviews subspace methods and transfer function models. Section 3 compares the out–of–sample performance of one– to fifteen–months–ahead forecasts constructed from the state space models given by the subspace method and the transfer function models. Section 4 summarizes the conclusions.

2. BRIEF REVIEW OF SUBSPACE METHODS AND TRANSFER FUNCTION MODELS

In this section we briefly discuss subspace methods, see Ljung (1999), and transfer function models, see Box and Jenkins (1976).

2.1 SUBSPACE METHODS

Let us assume that the output process $\{y_t : t = 1, 2, \ldots, N\}$ has dimension s and follows the minimal state space model

$$x_{t+1} = Ax_t + Bu_t + Ke_t, \tag{1}$$

$$y_t = Cx_t + Du_t + e_t, \qquad t = 1, 2, \ldots, N \tag{2}$$

where $\{e_t\}$ is the sequence of s–dimensional innovations, which for simplicity are assumed to be i.i.d. Gaussian random variables with zero mean and covariance matrix $\Omega > 0$, $\{u_t\}$ is the m–dimensional input process, which is assumed to be strongly exogenous, and $\{x_t\}$ is the sequence of n–dimensional unobserved states. The system matrices, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{s\times n}$, $D \in \mathbb{R}^{s\times m}$ and $K \in \mathbb{R}^{n\times s}$ are to be estimated.

Choosing two integers, f and p, that stand for "future" and "past", we can define the following vectors

$$Y_{t,f}^{+} = \begin{bmatrix} y_t \\ y_{t+1} \\ \vdots \\ y_{t+f-1} \end{bmatrix} \in \mathbb{R}^{fs}, \; Y_{t,p}^{-} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} \in \mathbb{R}^{ps} \quad \text{and} \quad Z_{t,p}^{-} = \begin{bmatrix} U_{t,p}^{-} \\ Y_{t,p}^{-} \end{bmatrix} \in \mathbb{R}^{p(m+s)},$$

where $U_{t,p}^{-}$ is defined analogously to $Y_{t,p}^{-}$ using the input vector u_t instead of y_t. In a similar way, we define $U_{t,f}^{+}$ and $E_{t,f}^{+}$ using the input vector u_t and the innovations e_t instead of y_t.

By repeated application of equations (1) and (2), we get

$$y_{t+i} = CA^i x_t + \sum_{j=0}^{i} L_j u_{t+i-j} + \sum_{j=0}^{i} K_j e_{t+i-j},$$

where $\{L_j\}$ and $\{K_j\}$ are the impulse response sequences, $L_j \in \mathbb{R}^{s\times m}$, $K_j \in \mathbb{R}^{s\times s}$, given by $L_0 = D$, $K_0 = I$, $L_j = CA^{j-1}B$, $K_j = CA^{j-1}K$, $j > 0$. From this, it is obtained that

$$Y_{t,f}^{+} = O_f x_t + U_f U_{t,f}^{+} + E_f E_{t,f}^{+}, \tag{3}$$

where $O_f = [C', A'C', \ldots, (A^{f-1})'C']'$ denotes the extended observability matrix,

$$U_f = \begin{bmatrix} D & & & \\ CB & D & & \\ \vdots & \vdots & \ddots & D \\ CA^{f-2}B & \cdots & \cdots & CB \quad D \end{bmatrix}, \quad \text{and} \quad E_f = \begin{bmatrix} I & & & \\ CK & I & & \\ \vdots & \vdots & \ddots & I \\ CA^{f-2}K & \cdots & \cdots & CK \quad I \end{bmatrix}.$$

Note that U_f and E_f are the Toeplitz matrices of the impulse responses $\{L_j\}$ and $\{K_j\}$.

From (2) we get $e_t = y_t - Cx_t - Du_t$. Substituting this expression in (1) and iterating, it is obtained that

$$x_t = K_p Z_{t,p}^{-} + (A - KC)^p x_{t-p}, \tag{4}$$

where $K_p = [K_{u,p}, K_{y,p}]$, $K_{u,p} = [B_K, A_K B_K, ..., A_K^{p-1} B_K]$, $K_{y,p} = [K, A_K K, ..., A_K^{p-1} K]$, $A_K = A - KC$ and $B_K = B - KD$.

Using (4), we can rewrite equation (3) as

$$Y_{t,f}^{+} = O_f K_p Z_{t,p}^{-} + U_f U_{t,f}^{+} + N_t, \tag{5}$$

where $N_t = O_f (A - KC)^p x_{t-p} + E_f E_{t,f}^{+}$. Defining $\beta_z = O_f K_p$ and $\beta_u = U_f$ in equation (5), we can write

$$Y_{t,f}^{+} = \beta_z Z_{t,p}^{-} + \beta_u U_{t,f}^{+} + N_t, \quad t = p+1, \ldots, T-f. \tag{6}$$

If the process y_t satisfies the minimum phase assumption, which is equivalent to the assumption that the process admits an infinite autoregressive representation, the eigenvalues of the matrix $A - KC$ have modulus less than one. Thus, under this last assumption, the central equation (6) has the following features:

- The vector N_t is asymptotically uncorrelated with the remaining terms on the right hand side of equation (6) as $p \to \infty$ because $(A - KC)^p \to 0$.
- The matrix β_z has rank n, the system order.
- $N_t \to E_f E_{t,f}^+$ as $p \to \infty$.
- $E_f E_{t,f}^+$ is an MA(f) process.

The CCA subspace algorithm is based on the previous observations. There are several versions of this algorithm. One of them is the following:

1. Since the rank of β_z in the regression equation (6) is equal to the system order n, use the results of Anderson (1951) based on partial canonical correlations to test H_0: rank$(\beta_z) = n$ for suitable values of n. After the rank of β_z has been determined, estimate $\beta_z = O_f K_p$, where O_f and K_p both have rank n. From this and equations (3) to (5), an estimate of the state is obtained as $\hat{x}_t = K_p Z_{t,p}^-$. Also, the matrices C and A can be estimated from the observability matrix $O_f = [C', A'C', \ldots, (A^{f-1})'C']'$. The estimate of C is simply the first s rows of O_f and the estimate of A is obtained by solving the following overdetermined system in the least squares sense,

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-2} \end{bmatrix} A = \begin{bmatrix} CA \\ CA^2 \\ \vdots \\ CA^{f-1} \end{bmatrix}.$$

   It may be possible at this stage that the estimated A matrix is not stable (not all of its eigenvalues have modulus less than one). In this case, we should transform it into a stable matrix.

2. After having estimated the regression equation (6), the residuals, $\hat{N}_t$, are an estimator of $E_f E_{t,f}^+$. If $M = M^{1/2}(M^{1/2})'$ is the Cholesky decomposition of the covariance matrix

of $\hat{N}_t$, where $M^{1/2}$ is a lower triangular matrix, then $M^{1/2}$ is an estimator of

$$E_f \operatorname{diag}(\Omega^{1/2}, \ldots, \Omega^{1/2}) = \begin{bmatrix} \Omega^{1/2} & 0 & \cdots & 0 \\ CK\Omega^{1/2} & \Omega^{1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}K\Omega^{1/2} & CA^{f-3}K\Omega^{1/2} & \cdots & \Omega^{1/2} \end{bmatrix}$$

and the first block column of the previous matrix is used, together with the estimated observability matrix, to estimate K. If the estimated $A - KC$ matrix is not stable, we would solve the following discrete algebraic Riccati equation

$$P = APA' - (APC' + S)(CPC' + R)^{-1}(APC' + S)' + Q$$

to get the stabilizing solution $P \geq 0$ and the corresponding Kalman gain $K = (APC' + S)(CPC' + R)^{-1}$ making $A - KC$ stable, where

$$\widehat{W} = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix}$$

is an estimate of $W = \operatorname{Var}(e_t'K', e_t')'$ in (1) and (2).

3. The matrices B and D and the initial state vector, x_1, are estimated by regression using the estimated matrices A, C and, possibly, K. If K is not used, the regression equation is

$$y_t = \left[C(zI - A)_t^{-1}B + D\right] u_t + CA^{t-1}x_1 + v_t,$$

where z is the forward operator, $zy_t = y_{t+1}$, $(zI - A)_t^{-1} = \sum_{j=1}^{t-1} z^{-j}A^{j-1}$, and $v_t = \left[C(zI - A)_t^{-1}K + I\right] e_t$. If, on the contrary, K is used, then the regression equation is

$$\left[I - C(zI - A_K)_t^{-1}K\right] y_t = \left[C(zI - A_K)_t^{-1}B_K + D\right] u_t + CA_K^{t-1}x_1 + e_t, \qquad (7)$$

where $(zI - A_K)_t^{-1} = \sum_{j=1}^{t-1} z^{-j}A_K^{j-1}$ and, as before, $A_K = A - KC$ and $B_K = B - KD$. Note that in this last equation the residuals are white.

4. The matrix $\Omega = \text{Var}(e_t)$ can be estimated in several ways. For example, one can use the regression equation (7) to estimate the residuals e_t. Another possibility is to use the observation equation (2) with the state x_t replaced with the estimated state $\hat{x}_t$ to estimate the residuals.

It is to be noted that the test statistic for H_0: $\text{rank}(\beta_z) = n$ in (6) based on partial canonical correlations (see Anderson, 1951) used in step 1 of the previous algorithm has an asymptotic chi–squared distribution under the assumption that the errors are white noise. This is not the case in equation (6), however, because the errors are asymptotically a moving average process. For this reason, Tsay (1989) has proposed a modification of this statistic.

An alternative to use partial canonical correlations to estimate the system order is to use an information criterion like AIC or BIC. To this end, for each order, we would replace x_t and C in equation (2) with the estimated state, $\hat{x}_t$, and the estimated C matrix and we would estimate by regression the D matrix, the residuals, e_t, and their covariance matrix.

As mentioned earlier, the procedure in step 1 of the previous algorithm is based on the computation of an SVD. More specifically, let $\Sigma_{ff|u}$, $\Sigma_{pp|u}$ and $\Sigma_{fp|u}$ the covariance matrices of the residuals of the regression of $Y_{t,f}^+$ and $Z_{t,p}^-$ onto $U_{t,f}^+$. Then, computing the SVD

$$\Sigma_{ff|u}^{-1/2}\Sigma_{fp|u}(\Sigma_{pp|u}^{-1/2})' = USV',$$

one estimates the partial canonical correlations between $Y_{t,f}^+$ and $Z_{t,p}^-$, given $U_{t,f}^+$, as the elements in the diagonal of S. After having tested for the number of canonical correlations that are nonzero, we have the approximation

$$\Sigma_{ff|u}^{-1/2}\Sigma_{fp|u}(\Sigma_{pp|u}^{-1/2})' \simeq \widehat{U}\widehat{S}\widehat{V}',$$

where $\widehat{S}$ has rank n, the estimated system order. The matrices O_f and K_p are estimated as

$$\widehat{O_f} = \Sigma_{ff|u}^{1/2}\widehat{U}\widehat{S}^{1/2} \quad \text{and} \quad \widehat{K_p} = \widehat{S}^{1/2}\widehat{V}'\Sigma_{pp|u}^{-1/2}.$$

As regards the orders f and p in the previous algorithm, there is no general consensus. One possibility is to fit an ARX approximation to the data and make p equal to the number of

lags in the fitted autoregression. The number of future lags, f, should be at least equal to the maximum system order, n, considered. Some authors simply put $f = p$. Another possibility is to make f grow with the sample size according to a formula of the type $f = \log^a(N)$ with $1.5 \leq a \leq 2$.

The function **n4sid** of the Toolbox SystemIdent of MATLAB implements several versions of subspace methods, including CCA. The CCA algorithm used by **n4sid** is similar to the one we have described in this section.

2.2 TRANSFER FUNCTION MODELS

Assuming an output variable, y_t, and m input variables, $x_{1t}, \ldots, x_{mt}$, a transfer function model can be written as

$$y_t = C + \frac{\omega_1(B)}{\delta_1(B)}x_{1t} + \frac{\omega_2(B)}{\delta_2(B)}x_{2t} + \cdots + \frac{\omega_m(B)}{\delta_m(B)}x_{mt} + \frac{\theta(B)}{\phi(B)}a_t,$$

where B is the backshift operator, $By_t = y_{t-1}$,

$$\begin{aligned}
\omega_i(B) &= (\omega_{i0} + \omega_{i1}B + \omega_{i2}B^2 + \cdots + \omega_{ih_i}B^{h_i})B^{b_i} \\
\delta_i(B) &= 1 + \delta_{i1}B + \cdots + \delta_{ir_i}B^{r_i} \\
\phi(B) &= 1 + \phi_1 B + \cdots + \phi_p B^p \\
\theta(B) &= 1 + \theta_1 B + \cdots + \theta_q B^q,
\end{aligned}$$

$\{a_t\}$ is white noise, usually assumed to be i.i.d. and Gaussian with zero mean. In addition, $\{a_t\}$ and the $\{x_{it}\}$ are assumed to be mutually and serially uncorrelated. The polynomials $\phi(z)$ and $\theta(z)$ can have multiplicative form in case seasonality is present.

As mentioned earlier, Box and Jenkins (1976) proposed a prewhitening method to identify transfer function models that has several drawbacks. For this reason, we will consider in this article the identification method proposed by Liu and Hanssens (1982), known as linear transfer function (LTF), and also the procedure proposed by Tsay (1985).

The LTF method is based on the following ideas. To simplify the notation, suppose only one input in the transfer function equation and denote by $\nu(z) = \omega(z)/\delta(z)$ its filter. Then,

we can consider the approximation

$$\nu(z) = \nu_0 + \nu_1 z + \nu_2 z^2 + \cdots,$$

and we can try to estimate the weights $\{\nu_j\}$ first. The whole procedure is as follows:

1. Estimate the weights $\{\nu_j\}$ assuming some model for N_t in the transfer function equation, $y_t = \nu(z)x_t + N_t$.

2. Identify a model for $\{N_t\}$.

3. Identify the polynomials $\omega(z)$ and $\delta(z)$ for the best approximation $\omega(z)/\delta(z) \simeq \nu(z)$.

In practice, a finite approximation for the filter $\nu(z)$ is used, so that a model of the form

$$y_t = C + (\nu_0 + \nu_1 B + \nu_2 B^2 + \cdots + \nu_k B^k)x_t + N_t$$

is considered. After steps 1 and 2, a generalization of the corner method (see [10]) is used to identify the polynomials $\omega(z)$ and $\delta(z)$ such that $\omega(z)/\delta(z) \simeq \nu(z)$.

Tsay (1985) proposed as a first step in transfer function identification to fit an autoregressive vector model to the random vector formed with the output and all of the inputs. In this way, a test of unidirectional causality can be implemented, the number of lags in the approximation for the input filters can be determined, and the weights of the approximation can be estimated.

Based on the identification and estimation of this autoregressive model, Tsay (1985) proposed a method to identify the output model and filters for the inputs. These last filters are also identified using the corner method.

3. COMPARATIVE PERFORMANCE OF ROLLING FORECASTS

Comparisons of models is usually based on their likelihoods. For Gaussian linear time series models one usually considers only the MSE of the residuals. This in turn is equivalent to assessing the models according to their within–sample one–step–ahead forecasts. Because our primary focus in this article is forecasting, an alternate and perhaps more relevant

criterion in judging model performance is the average absolute percent error of out–of–sample multistep–ahead forecasts.

In this section, we consider out–of–sample forecasts generated from a rolling forecast method. First, the model parameters are estimated using all observations through a given forecasting origin. Next, the forecasts are generated for this origin. This procedure is then repeated for all forecasting origins in the period of interest. In the first example, we consider all forecasting origins beginning with January 2004 and ending with December 2004. The initial 132 months, that corresponds to the dates 1993:1–2003:12, allow for fairly stable estimates for the time series models. In the second example, the forecasting origins start with observation number 74 and end with observation number 85. The results obtained with the first 73 observations are not very different from those obtained by Reinsel (1997) using the whole sample of 100 observations.

3.1 EXAMPLE 1

In the first example, the output series, y_t, is the log of the SCPI series and the input series are the logs of the inputs mentioned in Section .

To fit a transfer function model for the sample 1993:1–2003:12, we first fit a vector autoregressive model to the random vector formed with the output and the three inputs. Using the sequential likelihood ratio test statistic starting with a maximum number of lags equal to 19 the criterion selects 13 as the optimum number of lags. We note that all of the series considered are nonstationary but that the criterion is also valid in this case. We have obtained the maximum number of lags using the formula $\log(N)^{1.88}$, where N is the sample size. We have chosen the exponent 1.88 because the series is a monthly seasonal series and the number of lags in a VAR approximation can be big.

Looking at the t–values of the estimated parameters we see that the relation between the input series and the output series is indeed unidirectional. In addition, we find that the third input, the Spanish Import Price for Consumer Goods, does not seem to be related with the other series, so we decide to drop it from all of the models that we will fit. Thus, in the sequel only two inputs will be considered, the Spanish Producer Price Index for Consumer

Goods, x_{1t}, and the Spot Prices of Crude Oil of the Brent barrel, x_{2t}.

Using 13 lags as approximation for the filter inputs, we use the LTF method to identify a transfer function model. This model is

$$\nabla\nabla_{12}y_t = \frac{\omega_{10}+\omega_{11}B}{1+\delta_1 B+\delta_2 B^2}\nabla\nabla_{12}x_{1t} + (\omega_{20}+\omega_{21}B)\nabla\nabla_{12}x_{2t} + \frac{1}{1+\Phi B^{12}}a_t,$$

where $\nabla = 1-B$ and $\nabla_{12} = 1-B^{12}$ are the usual differencing operators. We perform the usual tests for diagnostic checking and confirm that the fit is good. The model is thus accepted. We also check whether the model changes when we add one observation at a time until we get the sample 1993:1–2004:12. We conclude that the model does not change.

After applying the operator $\nabla\nabla_{12}$, the output and both input series are stationary. Thus, we apply the CCA subspace method using the differenced series. We adjust the forecast of the differenced series accordingly to obtain the forecasts in levels.

The software for model identification that we use is the function **n4sid** of the SystemIdent Toolbox of MATLAB, with the options "CVA" (CCA in our terminology) and "simulation". This last option guarantees that the estimated A matrix is stable. The procedure selects a system order, i.e. dimension of the state vector, equal to 12 for the sample 1993:1–2003:12 among all orders between 1 and 19. We keep this order fixed in all of the samples in the forecasting exercise because the model does not change substantially. Because different versions of the software can give different results, we use the following rule to select values for f and p. First, set $f = \log(N)^{1.88}$, where N is the number of observations in the sample. Then, obtain p by fitting a VARX. Using this rule, we get the following values. For sample sizes between 133 and 136 (120 and 123 for the differenced series), $f = 18$ and $p = 11$, and for the other sample sizes, $f = 19$ and $p = 12$. To obtain the forecasts with the CCA method, we use a set of functions programmed in MATLAB by the first two authors that can forecast any state space model by applying the Kalman filter. The state space model that we use encompasses the model (1) and (2) and the model for the input x_{1t}. The other input series, x_{2t}, is considered as deterministic and the forecasts (the Future Prices) are simply added to the observed series at each forecast origin.

The variables entered to the function **n4sid** should in principle be centered. However,

Table 1. Comparison of Forecasts for both Methods

	Average absolute percent error of forecasts							
	$y_{t+1\|t}$	$y_{t+2\|t}$	$y_{t+3\|t}$	$y_{t+4\|t}$	$y_{t+5\|t}$	$y_{t+6\|t}$	$y_{t+7\|t}$	$y_{t+8\|t}$
CCA[1)]	0.0293	0.0427	0.0603	0.0881	0.0922	0.0768	0.0707	0.0646
TF[1)]	0.0267	0.0369	0.0568	0.0723	0.0814	0.0831	0.0901	0.0853
CCA[2)]	0.3486	0.5316	0.5713	0.4603	0.3788	0.6805	1.0469	1.3050
TF[2)]	0.3473	0.5201	0.5412	0.5506	0.4778	0.6174	0.9551	1.2125

	$y_{t+9\|t}$	$y_{t+10\|t}$	$y_{t+11\|t}$	$y_{t+12\|t}$	$y_{t+13\|t}$	$y_{t+14\|t}$	$y_{t+15\|t}$
CCA[1)]	0.0644	0.0645	0.0661	0.0675	0.0756	0.0903	0.0981
TF[1)]	0.0883	0.0777	0.0709	0.0881	0.1137	0.1365	0.1588
CCA[2)]	1.7040	2.1477	2.6050	3.0265	3.2568	3.3404	3.4249
TF[2)]	1.6106	2.0417	2.5265	2.9526	3.1303	3.1752	3.2659

1) and 2) refer to Examples 1 and 2, respectively.

we enter the differenced variables without centering because we found after some testing that the means are not significant. If we used centered variables, even when the means are insignificantly small, the forecasts could be severely biased due to the multiplicative effect of the differencing operator when passing from the predicted values in differences to the predicted levels.

The model for the input series is an airline model with one intervention,

$$x_{1t} = \frac{\omega}{1-\delta B} I_t^{37} + \frac{(1+\theta B)(1+\theta B^{12})}{\nabla \nabla_{12}} a_t,$$

where $0 < \delta < 1$ and $I_t^{37} = 1$ if $t = 37$ and $I_t^{37} = 0$ otherwise.

Tables 1–3 summarize the forecasting performance of the two models. Table 1 summarizes the average absolute percent error, defined as

$$100 \times \frac{1}{12} \sum_{t=N+h}^{N+11+h} \left| \frac{y_t - y_{t|t-h}}{y_t} \right|,$$

where N is the number of observations in the sample whose last point is the first forecasting origin, h is the number of periods ahead for the forecast, and $y_{t|t-h}$ is the forecast, of one– to

Table 2. Comparison of Forecasts for both Methods

	Means of forecast errors							
	$y_{t+1\|t}$	$y_{t+2\|t}$	$y_{t+3\|t}$	$y_{T+4\|t}$	$y_{t+5\|t}$	$y_{t+6\|t}$	$y_{t+7\|t}$	$y_{t+8\|t}$
CCA[1)]	-0.0003	-0.0003	0.0000	0.0004	0.0001	- 0.0001	- 0.0003	-0.0003
TF[1)]	-0.0003	-0.0002	0.0001	0.0006	0.0007	0.0009	0.0013	0.0016
CCA[2)]	0.0298	0.0489	0.0924	0.0727	-0.0636	-0.2182	-0.3076	-0.5776
TF[2)]	0.0261	0.0256	0.0529	0.0612	-0.0598	-0.1930	-0.3432	-0.5037

	$y_{t+9\|t}$	$y_{t+10\|t}$	$y_{t+11\|t}$	$y_{t+12\|t}$	$y_{t+13\|t}$	$y_{t+14\|t}$	$y_{t+15\|t}$
CCA[1)]	0.0003	0.0008	0.0013	0.0023	0.0032	0.0040	0.0047
TF[1)]	0.0023	0.0029	0.0033	0.0042	0.0054	0.0065	0.0075
CCA[2)]	-0.6637	-0.6225	-0.4747	-0.2794	0.0211	0.3237	0.5301
TF[2)]	-0.5877	-0.5673	-0.4586	-0.3077	-0.0462	0.2332	0.4388

1) and 2) refer to Examples 1 and 2, respectively.

fifteen–month–ahead forecasts for the two models considered. Table 2 summarizes the mean of the forecast errors, $y_t - y_{t|t-h}$, for the two models. It is seen that the transfer function forecasts better for one– to five–month–ahead and that the subspace model is better for the other periods ahead. Also, there is a systematic bias in the forecasts obtained with both models, although of similar magnitude. This is due to a great increase of oil prices during the period in which the forecasts are computed and that both models are unable to capture.

Table 3 summarizes the average absolute errors (AAE) of one– to fifteen–month–ahead forecasts with the two models. All the AAEs of the forecasts are expressed in relative terms. That is, relative to the corresponding AAEs of the transfer function model which serves as benchmark.

3.2 EXAMPLE 2

The second example involves data from a 50 megawatt turbo–alternator given in Jenkins and Watts (1968) and studied also by Reinsel (1997). Theory indicates that the in–phase and out–of–phase current deviations, x_{1t} and x_{2t}, can represent inputs to the system, and

the corresponding frequency deviations are an output, y_t.

Reinsel (1997) performs a preliminary VAR analysis with the three series. The results of this analysis confirm that x_{1t} and x_{2t} are indeed exogenous and that the inclusion of lag zero terms of $\{x_{1t}\}$ and $\{x_{2t}\}$ will not lead to any improvement in the equation for y_t. Reinsel further identifies the following transfer function model

$$\begin{aligned}&(1 - 0.801B + 0.276B^4)y_t \\ &= \hat{C} + (-0.155 + 0.487B - 0.820B^2 + 0.352B^3)x_{1,t-1} + 0.035B^2x_{2,t-1} + a_t.\end{aligned}$$

We confirmed that the fit of this model is good when we use a sample consisting of the first 73 observations and that the model does not change much when we increase the sample with one observation at a time until we get a sample of 85 observations. We thus accept this transfer function model for the forecasting exercise.

To apply subspace methods to this example, we again use the function **n4sid** of the SystemIdent Toolbox of MATLAB, with the options "CVA" and "simulation". Given that the transfer function model contains a constant, we increase the inputs with an input variable consisting of ones to account for the mean. The procedure selects a system order equal to 2 for the sample consisting of the first 73 observations among all orders between 1 and 6. The maximum order, m, is selected according to the formula $m = \log(N)^{1.3}$, where N is the sample size. Note that the exponent, 1.3, is now smaller than that in the first example, 1.88. This is due to the fact that the series considered have no seasonality and, therefore, the number of lags in a VAR approximation should be smaller. We keep the order equal to two fixed in all of the samples for the forecasting exercise. To select values for f and p, we use the following rule. First, set $f = \log(N)^{1.3}$, where N is the number of observations in the sample. Then, obtain p by fitting a VARX. Using this rule, we get $f = 6$ and $p = 3$ for all sample sizes.

Reinsel (1997) states on p. 297 that each of the two input variables, x_{1t} and x_{2t}, can be adequately modeled by a univariate ARMA(4,1) model and that a bivariate model shows only mild improvement over these univariate models. Thus, we use univariate ARMA(4,1) models to obtain the forecasts of the input variables. Since we are not interested in the

Table 3. Comparison of Forecasts for both Methods

	Relative AAE of forecasts							
	$y_{t+1\|t}$	$y_{t+2\|t}$	$y_{t+3\|t}$	$y_{t+4\|t}$	$y_{t+5\|t}$	$y_{t+6\|t}$	$y_{t+7\|t}$	$y_{t+8\|t}$
CCA[1)]	**1.0984**	**1.1579**	**1.0610**	**1.2181**	**1.1333**	**0.9245**	**0.7842**	**0.7573**
CCA[2)]	**1.0038**	**1.0220**	**1.0556**	**0.8359**	**0.7927**	**1.1022**	**1.0960**	**1.0763**

	$y_{t+9\|t}$	$y_{t+10\|t}$	$y_{t+11\|t}$	$y_{t+12\|t}$	$y_{t+13\|t}$	$y_{t+14\|t}$	$y_{t+15\|t}$
CCA[1)]	**0.7292**	**0.8295**	**0.9321**	**0.7656**	**0.6643**	**0.6616**	**0.6179**
CCA[2)]	**1.0580**	**1.0519**	**1.0311**	**1.0250**	**1.0404**	**1.0520**	**1.0487**

1) **and** 2) **refer to Examples 1 and 2, respectively.**

computation of the mean squared errors of the forecasts, these forecasts are simply added to the observed input variables at each forecast origin and the input variables are thus treated as deterministic when we use the Kalman filter to obtain the forecasts with the CCA method.

Tables 1–3 summarize the forecasting performance of the two models. Table 1 shows the average absolute percent error and table 2 reports the mean of the forecast errors for the two models. It is seen that the transfer function forecasts are similar to those of the CCA method. Unlike in Example 1, it seems that there is no bias in the forecasts in this case.

In Table 3 we can see again that the forecasts obtained with both models are similar.

4. CONCLUSIONS

In this article we have compared the forecasting performance of CCA subspace methods and transfer function models when these methods are applied to two output series, SCPI and frequency deviations, using some exogenous inputs. The main finding is that the forecasts obtained with both the CCA method and the traditional transfer function models are similar. This result is very encouraging because the CCA method can be considered as an automatic method in which the user has to make very few choices. This contrasts with the transfer

function methodology where considerable expertise in time series analysis is needed and the whole procedure is not automatic.

BIBLIOGRAPHY

Anderson, T. W., (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, **22**, 327–351.

Beguin, J. M., Gourieroux, C. and Monfort, A. (1980). Identification of a mixed autoregressive–moving average process: the corner method. In: Anderson, O. D. ed., *Time Series*. Amsterdam: North–Holland.

Box, G.E.P., and Jenkins, G.M. (1976), *Time Series Analysis, Forecasting and Control*. San Francisco: Holden–Day.

Jenkins, G. M., and Watts, D. G. (1968), *Spectral Analysis and its Applications*. San Francisco: Holden Day.

Larimore, W. E. (1983). System identification, reduced order filters and modeling via canonical variate analysis. In: Rao, H.S. Dorato, P. ed., *Proceedings of the 1983 American Control Conference 2*. Piscataway, NJ, pp. 445–451.

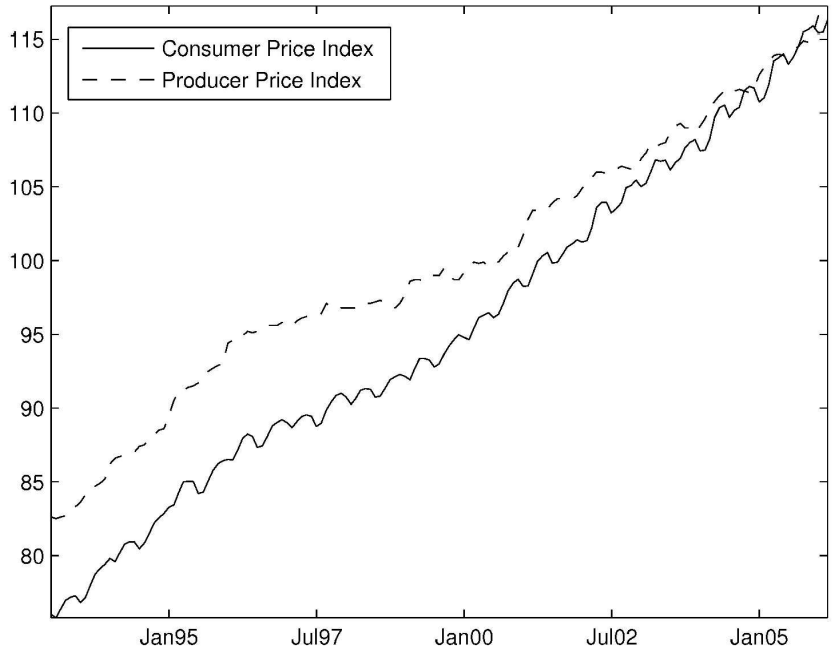
Liu, L. M., and Hanssens, D. M. (1982). Identification of multiple–input transfer function models. *Communications in Statistics, Theory and Methods*, **11**, 297–314.

Ljung, L. (1999), *System Identification: Theory for the User*. New Jersey: Prentice Hall.

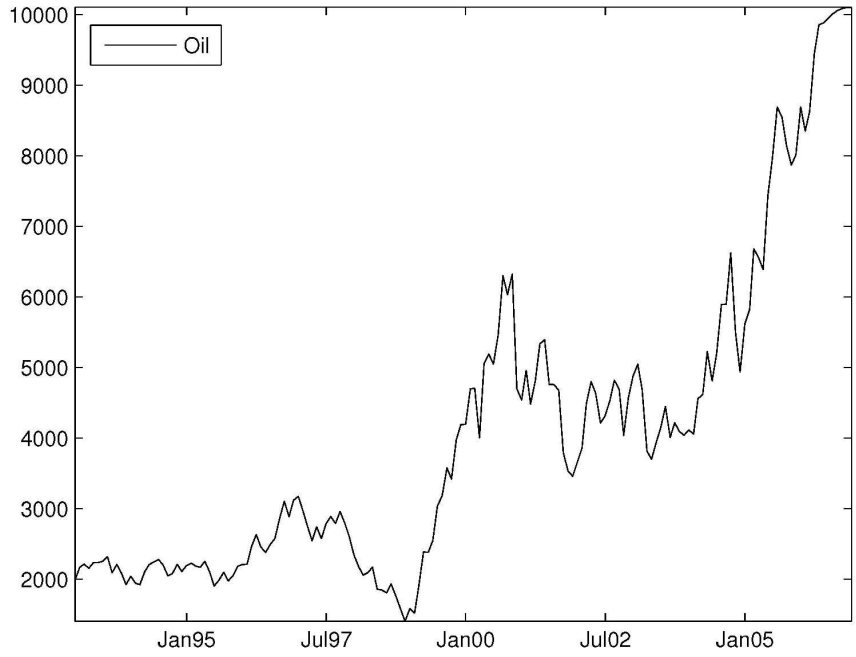
Reinsel, G. C. (1997), *Elements of Multivariate Time Series Analysis*. New York: Springer Verlag.

Tsay, R. S. (1985). Model identification in dynamic regression (distributed lag) models. *Journal of Business and Economic Statistics*, **3**, 228–237.

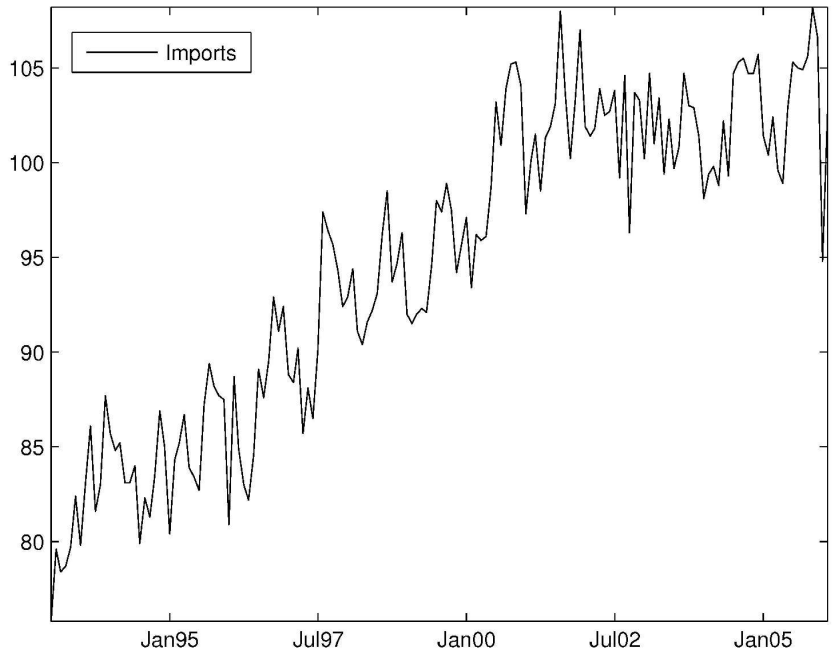
Tsay, R. S. (1989). Parsimonious parameterization of vector autoregressive moving average models. *Journal of Business and Economic Statistics*, **7**, 327–341.



148x111mm (600 x 600 DPI)



148x111mm (600 x 600 DPI)



148x111mm (600 x 600 DPI)