

Justification of Wold's Theorem and the Unbiasedness of a Stable Vector Autoregressive Time Series Model Forecasts

Iberedem A. Iwok¹

¹Department of Mathematics/Statistics, University of Port-Harcourt, P. M. B. 5323, Port-Harcourt, Rivers State, Nigeria

Correspondence: Iberedem A. Iwok, Department of Mathematics/Statistics, University of Port-Harcourt, P. M. B. 5323, Port-Harcourt, Rivers State, Nigeria. E-mail: ibywok@yahoo.com

Received: November 15, 2016 Accepted: December 30, 2016 Online Published: February 13, 2017

doi:10.5539/ijsp.v6n2p1

URL: <https://doi.org/10.5539/ijsp.v6n2p1>

Abstract

In this work, the multivariate analogue to the univariate Wold's theorem for a purely non-deterministic stable vector time series process was presented and justified using the method of undetermined coefficients. By this method, a finite vector autoregressive process of order p [VAR(p)] was represented as an infinite vector moving average (VEMA) process which was found to be the same as the Wold's representation. Thus, obtaining the properties of a VAR(p) process is equivalent to obtaining the properties of an infinite VEMA process. The proof of the unbiasedness of forecasts followed immediately based on the fact that a stable VAR process can be represented as an infinite VEMA process.

Keywords: VAR model, VEMA model, stable VAR process, univariate wold's theorem, multivariate wold's theorem and white noise process

1. Introduction

World's theorem is widely embraced in theoretical time series frame work because of its seemingly special representation and simplicity. The simple representation of the Wold's theorem is very remarkable and is a special case of the moving average process (Kenneth, 2014). The most advantageous part of the Wold's theorem is that it allows a dynamic evolution of a process to be approximated by a linear model and because of this feature; many researches consider the Wold's theorem as an existence theorem. Wold's theorem is sometimes referred to as Wold's decomposition theorem and is mostly applied in time series analysis. According to Borghers and Wessa (Borghers & Wessa, 2015), the fundamental justification for time series analysis is due to Wold's decomposition theorem.

Kallianpur (Kallianpur, 1979) presented a method for constructing the Wold's decomposition for multivariate stationary stochastic processes. The method was based on orthogonal decompositions obtained by forming orthogonal projections onto its component processes. However, it was discovered that the method does not give a complete solution to the Wold decomposition problem.

Papoulis (Papoulis, 1985) examined the concepts of predictability and band-limitedness in the Wold's decomposition theorem. He considered a real discrete time wide sense stationary process with autocorrelation and power spectrum. An extreme case of a bilinear process whose spectrum consists of lines only was also considered. It was shown that the values of a bilinear process were linearly dependents. This result was used to proof the wold's decomposition theorem in the context of mean-square estimation.

Jaydeb (Jaydeb, 2015) obtained a complete description of the class of n -tuples of doubly commuting isometrics. In particular, he presented a several variables analogue of the Wold's decomposition for isometrics on Hilbert spaces. The main result obtained was the generalization of the Slocinski's Wold-type decomposition of a pair of doubly commuting isometrics.

De Nicolao and Ferrari-Trecate (De Nicolao & Ferrari-Trecate, 2015) considered the Wold's decomposition of discrete-time cyclostationary (CS) process into regular and predictable component. The main result showed that predictable CS processes are linear combination of sine waves. The frequency of the sinusoids was found to be associated with the location on the unit circle of the zeros of the periodic error filter. From a spectral view point, it was discovered that an m th order predictable CS process exhibited a sequence of spectral lines. In practical prediction, De Nicolao and Ferrari-Trecate (De Nicolao & Ferrari-Trecate, 2015) found that the detection of spectral lines can then be used to separate the regular and predictable parts of a CS process, in analogy with deseasoning techniques used in the analysis of stationary time series.

Ansley (Ansley, et al, 1976) used Hilbert space methods to develop a rigorous proof that the sum of two uncorrelated moving average processes of order q_1 and q_2 is an MA process of order $q_1 \leq \max(q_1, q_2)$. The methods established the existence of suitable random shocks for the summed process, they illuminate relationships between the coefficients of such processes and their random shocks, and they provided means for proving that the random shocks of the summed processes are normal when the shocks of the underlying processes are normal. The Wold's decomposition was examined in terms of multiple representations of an MA process.

Caines and Gerencser (Caines & Gerencser, 1991) showed that the transform $Z(z), z \in \mathbb{C}$, of the coefficient sequence of the Wold's decomposition of any full rank wide sense stationary purely non-deterministic stochastic process satisfies the condition that $Z(z) \in H^2(D)$ and $Z^{-1}(z) \in H(D)$. It was further shown that all spectral factors satisfying the two conditions are equal up to right multiplication by orthogonal matrices, and that among these, the normalized ($Z(0) = 1$) spectral factors are equal to the transform of the Wold's decomposition. An elementary proof of Youla's theorem was then given together with a simple proof that the rows of a Cholesky factor of a banded block Toeplitz matrix converge to the coefficients of a stable matrix polynomial.

Olofsson (Olofsson, 2004) presented a Wold's decomposition of a two-isometric operator on a general Hilbert space. A pure two isometry was shown to be unitarily equivalent to a shift operator on a Dirichlet space corresponding to a positive operator measure on the unit circle. The result contained a previous result by Richter (Richter, 1998) as well as the result of von Neumann- Wold's decomposition of an isometry.

Katsoulis and Kribs (Katsoulis & Kribs, 2005) applied the Wold's decomposition to the study of row contractions associated with directed graphs. The work extended several fundamental theorems from the case of single vertex graphs to the general case of countable directed graphs with no sinks. The Szego-type factorization theorem for Countz-Krieger-Teoplitz families was proved which led to information on the structure of the unit ball in a free semi-group algebra. This showed that the joint similarity implied joint unitary equivalence for such families. For each group, Katsoulis and Kribs (Katsoulis & Kribs, 2005) proved a generalization of von Neumann's inequality which applied to row contractions of operators on Hilbert space which are related to the graph in conventional way. The results yielded a functional calculus determined by quiver algebras and free semi-group algebras.

As noted in the review, many researchers have proved and applied the Wold's theorem in different frame works. In this work, however, the intention is to proof the unbiasedness of forecasts of a stable VAR process based on the fact that a finite $\text{VAR}(p)$ is shown to have the same representation as the Wold's theorem.

2. Methodology

In this work, the underlined letters are used to represent vectors and matrices.

2.1 Stationarity

A time series is said to be stationary if the statistical property e.g. the mean and variance are constant through time. For a multivariate process, stationarity of a time series \underline{y}_t is achieved if $E[\underline{y}_t] = \underline{\mu}$; a vector of constants.

$$\underline{y}_t = \underline{Y}_t - \underline{\mu} \quad (1)$$

2.2 Backward shift Operator

The Backward shift Operator B is defined by

$$B^k \underline{y}_t = \underline{y}_{t-k} \quad (2)$$

2.3 White Noise Process

A white noise process $\underline{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{kt})'$ is a continuous random vector satisfying $E(\underline{\varepsilon}_t) = \underline{0}$, $E(\underline{\varepsilon}_t \underline{\varepsilon}_s') = \Sigma_\varepsilon$ and $E(\varepsilon_t \varepsilon_s') = 0$ for $s \neq t$.

where Σ_ε is covariance matrix which is assume to be non singular.

2.4 Vector Autoregressive (VAR) Model

Several multivariate time series model are in existence. However, the most commonly used model is the Vector Autoregressive (VAR) Model. VAR model is an independent reduced form dynamic model which involves constructing an equation that makes each endogenous variable a function of its own past values and past values of all other endogenous variables. A stable p -lagged vector autoregressive [VAR(p)] model has the form:

$$\underline{y}_t = \underline{\phi}_1 \underline{y}_{t-1} + \underline{\phi}_2 \underline{y}_{t-2} + \dots + \underline{\phi}_p \underline{y}_{t-p} + \underline{\varepsilon}_t \quad ; t = 0, \pm 1, \pm 2, \dots \quad (3)$$

Where $\underline{y}_t = (y_{1t}, \dots, y_{kt})'$ is a $(k \times 1)$ vector of time series variable, ϕ_j are fixed $(k \times k)$ coefficient matrices, $\underline{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{kt})'$ is a $(k \times 1)$ vector white noise process or innovation process.

The model (3) can be written explicitly in matrix form as:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{kt} \end{pmatrix} = \begin{pmatrix} \phi_{11}^1 & \phi_{12}^1 & \cdot & \cdot & \phi_{1k}^1 \\ \phi_{21}^1 & \phi_{22}^1 & \cdot & \cdot & \phi_{2k}^1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{k1}^1 & \phi_{k2}^1 & \cdot & \cdot & \phi_{kk}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ \cdot \\ \cdot \\ y_{kt-1} \end{pmatrix} + \begin{pmatrix} \phi_{11}^2 & \phi_{12}^2 & \cdot & \cdot & \phi_{1k}^2 \\ \phi_{21}^2 & \phi_{22}^2 & \cdot & \cdot & \phi_{2k}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{k1}^2 & \phi_{k2}^2 & \cdot & \cdot & \phi_{kk}^2 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \\ \cdot \\ \cdot \\ y_{kt-2} \end{pmatrix} \\ + \dots + \begin{pmatrix} \phi_{11}^p & \phi_{12}^p & \cdot & \cdot & \phi_{1k}^p \\ \phi_{21}^p & \phi_{22}^p & \cdot & \cdot & \phi_{2k}^p \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{k1}^p & \phi_{k2}^p & \cdot & \cdot & \phi_{kk}^p \end{pmatrix} \begin{pmatrix} y_{1t-p} \\ y_{2t-p} \\ \cdot \\ \cdot \\ y_{kt-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \cdot \\ \cdot \\ \varepsilon_{kt} \end{pmatrix} \quad (4)$$

2.5 Stable VAR(p) Processes

The process (3) is stable if its reverse characteristic polynomial of the VAR(p) has no roots in and on the complex unit circle. Formally y_t is stable if

$$\det(I_k - \phi_1 z - \dots - \phi_p z^p) \neq 0 \text{ for } |z| \leq 1. \quad (5)$$

A stable VAR(p) process y_t , $t = 0, \pm 1, \pm 2, \dots$, is also said to be stationary.

2.6 The Wold's Decomposition Theorem for a Univariate Time Series

The Wold's decomposition theorem for a univariate time series states that any zero-mean discrete stationary process $\{Y_t\}$ can be expressed as the sum of two uncorrelated components (processes):

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + v_t \quad (6)$$

Where $y_t = Y_t - \mu$, $\psi_0 = 1$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, $\{\varepsilon_t\} \sim N(0, \sigma_\varepsilon^2)$, $Cov(\varepsilon_t, \varepsilon_s) = 0$, $\forall s, t \in \mathbb{Z}$ and $\{v_t\}$ is deterministic. The sequence $\{\varepsilon_t\}$, $\{\psi_j\}$ and $\{v_t\}$ are unique. Thus, we can express

$$v_t = y_t - \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

In most cases, $v_t = 0$, $\forall t$ and y_t becomes purely non-deterministic. The Wold decomposition consists of two parts: the purely deterministic and purely non-deterministic. Since the process is zero-mean stationary,

$$y_t = Y_t - \mu_t ;$$

so that (6) can equivalently be presented as:

$$\begin{aligned} Y_t - \mu_t &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}; \\ \psi_0 &= 1, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty \end{aligned} \quad (7)$$

Also, since ε_t is a pure random process,

$$E[\varepsilon_t] = 0 \quad (8)$$

and

$$E[\varepsilon_t \varepsilon_s] = \begin{cases} \sigma^2 & , \quad t = s \\ 0 & , \quad \text{otherwise} \end{cases} \quad (9)$$

From (7), we obtain the mean,

$$\begin{aligned} E[Y_t - \mu_t] &= E\left[\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right] \\ &= \sum_{j=0}^{\infty} \psi_j E[\varepsilon_{t-j}] \\ &= 0 \\ \Rightarrow E[Y_t] &= \mu_t. \end{aligned}$$

For the variance, we have

$$\begin{aligned} V[X_t] &= E[(Y_t - \mu_t)^2] = E\left[\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right)^2\right] \\ &= E[(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots)^2] \end{aligned}$$

But

$$\begin{aligned} E[\varepsilon_t \varepsilon_{t-j}] &= 0, \forall j \neq 0 \\ \Rightarrow V[Y_t] &= E[\varepsilon_t^2] + \psi_1^2 E[\varepsilon_{t-1}^2] + \psi_2^2 E[\varepsilon_{t-2}^2] + \dots \end{aligned}$$

From (9), we have

$$V[Y_t] = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2$$

Thus, the variance is finite and not time dependent.

For the autocovariances, we have

$$\begin{aligned} Cov(Y_t, Y_{t+k}) &= \gamma_k = E[(Y_t - \mu_t)(Y_{t+k} - \mu_{t+k})] \\ &= E\left[\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right)\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t+k-j}\right)\right] \\ &= E[(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots)(\varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \psi_2 \varepsilon_{t+k-2} + \dots)] \\ &= \sigma_\varepsilon^2 (\psi_k + \psi_1 \psi_{k+1} + \psi_2 \psi_{k+2} + \dots) \\ &= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{k+j} = \gamma(k) < \infty. \end{aligned}$$

Thus, the autocovariances are only functions of the lag k and the process is covariance stationary.

The autocorrelation function at lag k is given as

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{k+j}}{\sum_{j=0}^{\infty} \psi_j^2}; \quad k = 1, 2, 3, \dots$$

According to the Wold decomposition theorem, any discrete stationary model (especially ARMA) can be represented on the basis of this decomposition.

2.7 Wold's Theorem for the Multivariate Model

The multivariate analogue to the univariate Wold's theorem is that if $\{Y_t\}$ is a purely non-deterministic stationary process with mean vector $\underline{\mu}$; then $\underline{Y}_t - \underline{\mu}$ can be represented as a linear combination of weighted lagged vector white noise processes. That is,

$$\begin{aligned}\underline{Y}_t - \underline{\mu} &= \underline{\varepsilon}_t + \underline{\psi}_1 \underline{\varepsilon}_{t-1} + \underline{\psi}_2 \underline{\varepsilon}_{t-2} + \cdots \\ &= \sum_{j=0}^{\infty} \underline{\psi}_j \underline{\varepsilon}_{t-j}; \text{ where } \underline{\psi}_0 = \underline{I}\end{aligned}\quad (10)$$

Let $\underline{y}_t = \underline{Y}_t - \underline{\mu}$; then

$$\underline{y}_t = \sum_{j=0}^{\infty} \underline{\psi}_j \underline{\varepsilon}_{t-j} = \underline{\psi}(B) \underline{\varepsilon}_t \quad (11)$$

where $\underline{\psi}(B) = \sum_{j=0}^{\infty} \underline{\psi}_j B^j$ is a $(k \times k)$ matrix such that $B^j \underline{\varepsilon}_t = \underline{\varepsilon}_{t-j}$;

\underline{y}_t is a $(k \times 1)$ random vector;

$\underline{\psi}_j$'s are fixed $(k \times k)$ coefficient matrices;

$\underline{\varepsilon}_{t-j}$ is a $(k \times 1)$ vector of white noise process at lag j ;

$\underline{\varepsilon}_t$ is a k - dimensional white noise or innovation process such that

$E[\underline{\varepsilon}_t] = \underline{0}$, $E[\underline{\varepsilon}_t, \underline{\varepsilon}_t'] = \underline{\Sigma}_\varepsilon$ and $E[\underline{\varepsilon}_t, \underline{\varepsilon}_s'] = \underline{0}$ for $s \neq t$. The covariance matrix $\underline{\Sigma}_\varepsilon$ is assumed to be non-singular.

3. Justification of the Wold's Expression (11)

To justify the Wold's theorem, there is need to define some stationary processes that will lead to the results.

3.1 Vector Moving Average Model of Order q

A vector time series process $\{\underline{y}_t\}$ is said to be a vector moving average model of order q denoted as $VEMA(q)$ if it can be represented as

$$\underline{y}_t = \underline{\varepsilon}_t + \underline{\psi}_1 \underline{\varepsilon}_{t-1} + \cdots + \underline{\psi}_q \underline{\varepsilon}_{t-q} \quad (12)$$

where $\underline{\varepsilon}_t$ is a vector of white noise processes. It should be noted here that the expressions (10) and (12) seems to be the same; except that the $VEMA(q)$ model is finite while the Wold's model is infinite.

Equation (12) can be compressed as:

$$\begin{aligned}\underline{y}_t &= \sum_{j=0}^q \underline{\psi}_j \underline{\varepsilon}_{t-j}; \underline{\psi}_0 = \underline{1} \\ \Rightarrow \underline{y}_t &= \left(1 + \underline{\psi}_1 B + \underline{\psi}_2 B^2 + \cdots + \underline{\psi}_q B^q\right) \underline{\varepsilon}_t \\ &\Rightarrow \underline{y}_t = \underline{\psi}(B) \underline{\varepsilon}_t\end{aligned}\quad (13)$$

where,

$$\underline{\psi}(B) = \left(1 + \underline{\psi}_1 B + \underline{\psi}_2 B^2 + \cdots + \underline{\psi}_q B^q\right) = \sum_{j=1}^q \underline{\psi}_j B^j; \underline{\psi}_0 = \underline{1} \quad (14)$$

and $B^k \underline{\varepsilon}_t = \underline{\varepsilon}_{t-k}$.

3.2 Vector Autoregressive Model of Order p

We earlier noted that a stationary $VAR(p)$ can be represented as

$$\begin{aligned}\underline{y}_t &= \underline{\phi}_1 \underline{y}_{t-1} + \underline{\phi}_2 \underline{y}_{t-2} + \cdots + \underline{\phi}_p \underline{y}_{t-p} + \underline{\varepsilon}_t \\ &= \sum_{j=0}^p \underline{\phi}_j \underline{y}_{t-j}; \underline{\phi}_0 = \underline{1} \\ \Rightarrow \underline{y}_t - \underline{\phi}_1 \underline{y}_{t-1} - \underline{\phi}_2 \underline{y}_{t-2} - \cdots - \underline{\phi}_p \underline{y}_{t-p} &= \underline{\varepsilon}_t \\ \Rightarrow \left(\underline{1} - \underline{\phi}_1 B - \underline{\phi}_2 B^2 - \cdots - \underline{\phi}_p B^p\right) \underline{y}_t &= \underline{\varepsilon}_t\end{aligned}\quad (15)$$

$$\Rightarrow \underline{\phi}(B) \underline{y}_t = \underline{\varepsilon}_t \quad (16)$$

$$\Rightarrow \underline{y}_t = \underline{\phi}^{-1}(B) \underline{\varepsilon}_t \quad (17)$$

where,

$$\underline{\phi}(B) = (1 - \underline{\phi}_1 B - \underline{\phi}_2 B^2 - \dots - \underline{\phi}_p B^p) \text{ and } B^k \underline{y}_t = \underline{y}_{t-k}$$

Expression (17) simply shows that a $VAR(p)$ process can be represented as an infinite $VEMA(q)$ process.

The Wold's theorem emphasizes that any stationary process can be represented as a linear combination of weighted lagged vector white noise processes.

Now, we have a stationary $VAR(p)$ process represented as (15) which result in equation (16).

From equation (16), we can write

$$\underline{\varepsilon}_t = \underline{\phi}(B) \underline{y}_t \quad (18)$$

Substituting equation (13) in (18) gives

$$\underline{\varepsilon}_t = \underline{\phi}(B) \underline{y}_t = \underline{\phi}(B) \underline{\psi}(B) \underline{\varepsilon}_t$$

Dividing by $\underline{\varepsilon}_t$

$$\Rightarrow \underline{\phi}(B) \underline{\psi}(B) = 1$$

But

$$\underline{\psi}(B) = \sum_{j=1}^q \underline{\psi}_j B^j \text{ [from (14)]}$$

$$\Rightarrow \underline{\phi}(B) \sum_{j=1}^q \underline{\psi}_j B^j = 1$$

On expansion,

$$\Rightarrow (1 - \underline{\phi}_1 B - \underline{\phi}_2 B^2 - \dots - \underline{\phi}_p B^p) \sum_{j=1}^q \underline{\psi}_j B^j = 1 \quad (19)$$

Thus given a $VAR(p)$ process, the $\underline{\psi}_j$ s ($j = 0, 1, 2, \dots$) can easily be obtained by equating the coefficients of like powers of B s on both sides of (19). Since the $\underline{\phi}_j$ s are the coefficient matrix of the given $VAR(p)$ multivariate model, the $\underline{\psi}_j$ s can easily be obtained by the method of undetermined coefficients. These $\underline{\psi}_j$ s are the coefficients of an infinite VEMA process represented as:

$$\underline{y}_t = \underline{\varepsilon}_t + \underline{\psi}_1 \underline{\varepsilon}_{t-1} + \underline{\psi}_2 \underline{\varepsilon}_{t-2} + \dots \quad (20)$$

This simply means that a $VAR(p)$ process can be converted to an infinite VEMA process which is nothing else but the Wold's representation (10) or (11).

It further implies that obtaining the properties of a $VAR(p)$ process is equivalent to obtaining the properties of an infinite VEMA process or the Wold's representation (10).

The proof of unbiasedness of forecast shall be based on this fact that a stationary VAR process can be represented as an infinite VEMA process.

4. Proof of the Unbiasedness of Forecasts

Let the forecast origin be t and the lead time be l . The major concern is to forecast a vector \underline{y}_{t+l} ($l \geq 1$) when we are currently at time t .

From (20), the future value can be expressed as \underline{y}_{t+l}

$$\underline{y}_{t+l} = \underline{\varepsilon}_{t+l} + \underline{\psi}_1 \underline{\varepsilon}_{t+l-1} + \underline{\psi}_2 \underline{\varepsilon}_{t+l-2} + \dots \quad (21)$$

Since $\underline{\varepsilon}_t$ is a white noise vector process and $E[\underline{\varepsilon}_t] = 0$; then

$$E[\underline{\varepsilon}_{t+h} \mid \underline{y}_t, \underline{y}_{t-1}, \dots] = 0 \quad ; \quad \forall h > 0.$$

Thus, the best predictor of \underline{y}_{t+l} is

$$\begin{aligned}\hat{y}_t(l) &= E \left[\varepsilon_{t+h} \mid y_t, y_{t-1}, \dots \right] = \underline{\psi}_l^* \underline{\varepsilon}_t + \underline{\psi}_{l+1}^* \underline{\varepsilon}_{t-1} + \underline{\psi}_{l+2}^* \underline{\varepsilon}_{t-2} + \dots \\ \Rightarrow \hat{y}_t(l) &= \sum_{j=l}^{\infty} \underline{\psi}_j^* \underline{\varepsilon}_{t+l-j}\end{aligned}\quad (22)$$

Since y_{t+l} and $\hat{y}_t(l)$ are linear combinations of the $\underline{\varepsilon}_t$'s; the forecast error $\underline{\varepsilon}_t(l)$ for lead time l will also be a linear combination of the $\underline{\varepsilon}_t$'s. That is,

$$\underline{\varepsilon}_t(l) = \underline{\varepsilon}_{t+l} + \underline{\psi}_1 \underline{\varepsilon}_{t+l-1} + \underline{\psi}_2 \underline{\varepsilon}_{t+l-2} + \dots + \underline{\psi}_{l-1} \underline{\varepsilon}_{t+1} \quad (23)$$

From (21) and (22), the mean square error of forecast is

$$\begin{aligned}E \left[\left(y_{t+l} - \hat{y}_t(l) \right) \left(y_{t+l} - \hat{y}_t(l) \right)' \right] &= \underline{\Sigma}_\varepsilon + \underline{\psi}_1 \underline{\Sigma}_\varepsilon \underline{\psi}_1' + \underline{\psi}_2 \underline{\Sigma}_\varepsilon \underline{\psi}_2' + \dots + \underline{\psi}_{l-1} \underline{\Sigma}_\varepsilon \underline{\psi}_{l-1}' \\ &\quad + \sum_{j=0}^{\infty} \left[\left(\underline{\psi}_{l+j} - \underline{\psi}_{l+j}^* \right) \underline{\Sigma}_\varepsilon \left(\underline{\psi}_{l+j} - \underline{\psi}_{l+j}^* \right)' \right] \\ &= \sum_{j=0}^{l-1} \underline{\psi}_j \underline{\Sigma}_\varepsilon \underline{\psi}_j' + \sum_{j=0}^{\infty} \left[\left(\underline{\psi}_{l+j} - \underline{\psi}_{l+j}^* \right) \underline{\Sigma}_\varepsilon \left(\underline{\psi}_{l+j} - \underline{\psi}_{l+j}^* \right)' \right]\end{aligned}\quad (24)$$

which is minimised by setting $\underline{\psi}_{l+j} - \underline{\psi}_{l+j}^* = 0$. Since the variance of the error forecast has attained minimum; then for lead time l ,

$$\begin{aligned}y_{t+l} &= \underline{\varepsilon}_{t+l} + \underline{\psi}_1 \underline{\varepsilon}_{t+l-1} + \underline{\psi}_2 \underline{\varepsilon}_{t+l-2} + \dots + \underline{\psi}_{l-1} \underline{\varepsilon}_{t+1} + \underline{\psi}_l^* \underline{\varepsilon}_t + \underline{\psi}_{l+1}^* \underline{\varepsilon}_{t-1} + \underline{\psi}_{l+2}^* \underline{\varepsilon}_{t-2} + \dots \\ \Rightarrow y_{t+l} &= \underline{\varepsilon}_t(l) + \hat{y}_t(l)\end{aligned}$$

Since $\left[\underline{\varepsilon}_t(l) \right] = 0$; the forecast is unbiased.

5. Discussion and Conclusion

Since it has been shown in this work that a finite VAR(p) can have an equivalent Wold's representation; it means obtaining the properties of the Wold's process is equivalent to obtaining the properties of the VAR(p) process that gives rise to the Wold's process. To be more scientific, it can be said here that the underlying linear structure in the time domain that generate the VAR process in general is the Wold's process. This idea is specific for linear processes only. However, the presence of non linear structure in most time varying quantities cannot be completely ruled out.

References

- Ansley, C. F., Spivey, W. A., & Wroblewski, W. J. (1976). On the structure of moving average processes. *Journal of Econometrics*, 6(8), 121-134.
- Borghers, E. & Wessa, P. (2015). *Statistics – econometric – Forestry*. Office for research Development and education.
- Caines, P.E. & Gerencser, L. (1991). A simple proof for a spectral factorization theorem. *IMA Journal of Mathematical control and Inf. Oxford University Press*, 8(1), 39-44. <https://doi.org/10.1093/imamci/8.1.39>
- De Nicolao, G. & Ferrari-Trecate, G. (2015). *On the Wold's Discrete-time cyclostationary processes*. Dipartimento di Informatica e Sistemistica, Universita di Pavia, Via Ferrata 1, 27100 Paria, Italy.
- Jaydeb, S. (2015). Wold's decomposition for doubly commuting isometrics. *Indian Statistical institute, Statistics and Mathematics unit.*, 8(12), 21-34.
- Kallianpur, P. (1979). On the construction of Wold's Decomposition for multivariate stationary process. *Journal of Multivariate Analysis*, 9(4). [http://dx.doi.org/10.1016/0047-259X\(79\)90057-5](http://dx.doi.org/10.1016/0047-259X(79)90057-5)
- Katsoulis, E. & Kribs, D. W. (2005). Application of the Wold's decomposition to the study of Row Contractions associated with direct graphs. *Transactions of the American Mathematical Society*, 357(9), 3739-3755. <https://doi.org/10.1090/S0002-9947-05-03692-5>
- Kenneth, T. S. (2014). Wold's decomposition and Time Series. *Journal of Scientific Research.*, 8(5), 48-56.
- Lutkepohl H.(2005). New introduction to multiple Time Series Analysis. *Springer Berlin Heidebelg*, New York. ISBN 3-540-40172-5. SPIN 10932797. <https://doi.org/10.1007/978-3-540-27752-1>
- Olofsson, A. (2004). *Acta Sci. Math. Szeged*, 70 (10), 715-726.

Papoulis, A. (1985). Predictable processes and Wold's decomposition-A review. *IEEE transactions on Acoustics, speech, and signal processing.*, 33(4). <https://doi.org/10.1109/TASSP.1985.1164637>

Ritcher, S. (1998). Invariant subspaces of Dirichlet shift. *J. Reine Angew. Math.*, 386 (14), 205-220.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).