Journal of Economics and Technology Research
ISSN 2690-3695 (Print) ISSN 2690-3709 (Online)
Vol. 3, No. 2, 2022
www.scholink.org/ojs/index.php/jetr

Original Paper

A Novel Approach to Forecasting High Dimensional S&P500 Portfolio Using VARX Model with Information Complexity

Jana Salim[1*] & Hamparsum Bozdogan[2]

[1] School of Economics, Shandong University, Jinan, China

[2] Department of Business Analytics, University of Tennessee, Knoxville, USA

[*] Jana Salim, School of Economics, Shandong University, Jinan, China

Received: April 14, 2022 Accepted: April 28, 2022 Online Published: May 17, 2022

doi:10.22158/jetr.v3n2p1 URL: http://dx.doi.org/10.22158/jetr.v3n2p1

Abstract

This study considers vector autoregressive models that allow for endogenous and exogeneous regressors VARX using multivariate OLS regression. For the model selection, we follow bozdogan's entropic or information-theoretic measure of complexity ICOMP criterion of the estimated inverse Fisher information matrix IFIM in choosing the best VARX lag parameter and we established that ICOMP outperform the conventional information criteria. As an empirical illustration, we reduced the dimension of the S&P500 multivariate time series using Sparse Principal Component Analysis (SPCA) and chose the best subset of 37 stocks belonging to six sectors. We then performed a portfolio of stocks based on the highest SPC loading weight matrix, plus the S&P500 index. Furthermore, we applied the proposed VARX model to predict the price movements in the constructed portfolio, where the S&P500 index was treated as an exogeneous regressor of the VARX model. It has been deduced too that the buy-sell decision making in response to VARX (4,0) for a stock outperforms investing and holding the stock over the out-of-sample period.

Keywords

ICOMP, Sparse Principal Component Analysis (SPCA), S&P500, VARX

1. Introduction

Vector Autoregressive (VAR) model since its introduction by (Sims, 1980) almost 40 years ago has become a popular model for econometricians, economists, financial modelers to analyze multivariate time-series data. VAR model provides a convenient flexible model for model order determination, forecasting, and structural inference in many applications. VAR model is attractive because it can capture and estimate the covariance matrix of the innovations and the coefficient estimates which achieves a more accurate prediction as compared to the traditional univariate AR(1) models that has been used conventionally in several data (Salim, 2017).

Once VAR is incorporated with unmodeled or exogeneous variables, then VARX models are produced improving the Minimum Square Error (MSE) forecast. For example, oil price has been widely used as an exogeneous macroeconomic variable, and where the lag order may differ between modeled and unmodeled series. Thus, model selection should fit all subset models up to the predetermined maximal lag order for both the exogeneous and the endogenous series (Nicholson et al., 2017).

In the literature of VARX modeling, there does not appear to be convincing study to establish the performance of the ICOMP criteria. Few progress has been made in the literature concerning model selection in VARX models, for example (Hoover, 1999), (Hendry, 1999), and (Hlawka, 2014) have analyzed different reduced rank vector autoregression VAR model selection and model reduction methods by Monte Carlo methods.

In this paper, our aspiration is to fill the gap in the literature by first studying the efficacy of the information-theoretic model selection criteria in multivariate Gaussian vector autoregressive VARX models under a certain innovations covariance structure to choose the optimal true number of lags in the VARX model. We carry out one Monte Carlo Simulation to study the performance of Akaike's information criterion to study the performance of Akaike's Information Criterion (AIC) (Schwarz, 1978), and Bozdogan's entropic or information-theoretic measure of the complexity ICOMP criterion of the estimated Inverse-Fisher Information Matrix (IFIM) (1990, 2000, 2004) under a unique scenario with varying sample sizes.

As is well known in the VARX model, each time series is interacted linearly with both its own lag variable values and those of every other series included (Nicholson et al., 2017), then large number of parameters needs to be estimated. Hence high dimensional time series require dimension reduction techniques as a preprocessing approach to estimate the parameters of the VARX model to avoid overparameterization.

In the literature, L1-type space estimation procedures have been used to resolve the dimensionality issue. However, they are not well suited to solve lag selection problem and time series dependence. The resulting models are not easily interpretable. In fact, Nicholson et al., 2017 shows that the prediction performance of the L1-type estimators substantially deteriorates for VARs with large lag orders as they select high lag order coefficients. To simultaneously address the dimensionality and lag selection issues, in a series of papers, Bearse and Bozdogan (1998), Bearse and Bozdogan (2003), and Howe and Bozdogan (2010) introduced a novel Genetic Algorithm (GA) procedure along with Information Complexity (ICOMP) criterion to choose the optimal lag length and to reduce the dimension of the high-dimensional time series by subset selection of the best variables.

In this paper, our novel approach is motivated and in the spirit of their approach. Instead of using the GA, for dimension reduction and selecting the best subset of features or variables, we have used the

Sparse Principle Component Analysis (SPCA), introduced by Zou et al. (2004) and hybridize it with the VARX model. SPCA works by writing the usual Principal Component Analysis (PCA) as a regression-type optimization problem and applies Least Absolute Shrinkage and Selection Operator (LASSO) of Tibshirani (1996), a penalization technique based on the L1-norm. We used SPCA as our pre-processing first to reduce the dimension and select the features, and not for lag length selection.

Next, we aim to forecast the directional movement of the constructed portfolio of the six S&P500 stock prices after employing a VARX model; for investors according to (Reboredo et al., 2012), the directional predictability ability of a model has practical implications for market timing and asset allocation management. We provide a real example to select the best subset of stocks for investment purposes and to predict the movement of a high dimensional portfolio which is constructed from 500 constituents of stock prices related to eleven sectors. This is a high-frequency data measured in every minute for a couple of months totaling T=41266 minutes (time series observations). To reduce the dimension of the S&P500 stocks and construct the optimal portfolio weights, we propose to use Sparse Principal Component Analysis (SPCA). The chosen best subset of stocks are then classified belonging into six market sectors. Then a portfolio of stocks is formed from this list based on the highest SPCs loading weight matrix, plus the S&P500 index to carry out a multivariate time series, called VARX model. To set the stage, the remainder of this paper is organized as follows. In Section 2, we provide the brief background of the VARX modeling and estimation of the parameters of the Gaussian VARX models. Section 3 introduces the several model selection criteria and their analytical form in VARX(p,s) model lag order determination. In Section 4, we provide one Monte Carlo simulation result and investigate the performance of the model selection under a specific innovation covariance matrix as we vary the number of observations for a 3-dimensional endogenous and 1-dimensional exogenous VARX (2,0) model (denoted by $3 \times 1 - D$ VARX (2,0). Then, we present the Sparse Principle Component Analysis (SPCA) to extract the best subset of features (stocks) which is an extension of the usual Principle Components Analysis (PCA) used as a dimension-reduction technique and later we provide our numerical results on a high-dimensional data set composed of 500 constituents of S$P500 stock prices. We then examine how a portfolio is constructed by the SPCA, and showing how ICOMP can substantially improve the out-of-sample forecasting performance using the VARX model as the data generating process. Section 5 concludes the paper.

2. Method

2.1 Vector Autoregressive Model with Exogeneous Variable

A multivariate time series model that contains lagged endogenous and exogenous variables is called the VARX model which is an extended form of the VAR model. A VARX model can be affected by the presence of other observable variables that are determined outside the system of interest. Such variables are called exogeneous (independent) or unmodelled variables (Lutkepohl, 2005, p. 387).

Exogenous variables can be stochastic or non-stochastic. The model can also be affected by the lags of exogenous variables. In short hand notation, VARX model is denoted by VARX(p,s), where p is the number of lags of the endogenous variables, and s is the number of lags of the exogenous variables.

A d-dimensional VARX(p, s) model can be written as:

$$\mathbf{y}t = \mathbf{c} + \boldsymbol{\Sigma}\Phi i\mathbf{y}t-i + \boldsymbol{\Sigma}\Theta i\mathbf{x}t-i + \varepsilon t, \quad (1)$$

where $\mathbf{y}_t = (\mathbf{y}_{1t},\mathbf{y}_{2t},...,\mathbf{y}_{dt})^0$ is a $(d\times1)$ vector of time series variables with a sample size t=1,..,T observations for each of the d-variables, the Φ_i are fixed $(d \times d)$ coefficient matrices, $c = (c_1,c_2,...,c_d)^0$ is a fixed $(d \times 1)$ vector of intercept terms, where $\mathbf{x}_t = (\mathbf{x}_{1t},...,\mathbf{x}_{rt})^0$ is a r-dimensional time series vector, Θ is a $(d \times r)$ coefficient matrix and $\varepsilon_t = (\varepsilon_{1t},...,\varepsilon_{dt})$ is a $(d \times 1)$ unobservable zero mean white noise vector process (serially uncorrelated or independent) with time invariant positive definite covariance matrix Σ. i.e., is, $\varepsilon_t \sim N(0,\Sigma_\varepsilon)$, where $E(\varepsilon_t) = 0$**.** For example, a VARX(1,0) model with 3-dimensional endogenous with 2-dimensional exogenous variables denoted by $3 \times 2 - D$ VARX(1,0) is $\mathbf{y}t = \mathbf{c} + \Phi 1\mathbf{y}t-1 + \Theta 0\mathbf{x}t + \varepsilon t, \quad (2)$

where $\mathbf{y}t = (\mathbf{y}1t,\mathbf{y}2t,\mathbf{y}3t)0$ and $\mathbf{x}t = (\mathbf{x}1t,\mathbf{x}2t)0$ Or, equivalently this can be written as

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \\ \theta_{31} & \theta_{32} \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} \quad (3)$$

We note that if $\mathbf{x}_t$ contains just a constant and $s = 0$, then VARX model reduces to VAR model, assuming ε_t is white noise. Although a VARX model may be a simple model, but, it is highly overparameterized. For a d- dimensional VARX(p,s) model, there are $d(d + 1)/2$ free parameters in Σ, in total, there are $m = d(dp + 1 + r(s + 1)) + d(d + 1)/2$ parameters to estimate. In the absence of further restrictions on the parameter space, we refer to eq.(1) as a saturated VARX(p,s) model.

2.2 Parameter Estimation of VARX Model

This VARX model has the general multivariate regression-like linear structure:

$$\underset{(T\times d)}{\mathbf{Y}} = \underset{(T\times q)(q\times d)}{\mathbf{ZB}} + \underset{(T\times d)}{\mathbf{E}} \quad (4)$$

where $q \equiv dp + 1 + r(q + 1)$, $\mathbf{Y} = (\mathbf{y}'_1, \mathbf{y}'_2, \ldots, \mathbf{y}'_T)'$, $\mathbf{Z} = (\mathbf{z}'_1, \mathbf{z}'_2, \ldots, \mathbf{z}'_T)'$

with the q-vector $\mathbf{z}t = (1, \mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_{t-p}, \mathbf{x}'_{t-1}, \mathbf{x}'_{t-2}, \ldots, \mathbf{x}'_{t-s})$, and $\mathbf{E} = (\varepsilon'_1, \varepsilon'_2, \ldots, \varepsilon'_T)'$. The coefficient matrix is given by $\mathbf{B} = (\mathbf{c}', \boldsymbol{\Phi}'_1, \boldsymbol{\Phi}'_2, \ldots, \boldsymbol{\Phi}'_p, \boldsymbol{\Theta}'_1, \boldsymbol{\Theta}'_2, \ldots, \boldsymbol{\Theta}'_s)'$.

A VARX model can be expressed as the saturated VARX in eq. (4) subject to linear constraints on the coefficient matrix **B**. We adopt as notation for VAR model, and rewrite the saturated VARX in its vectorized form as,

$$\mathrm{vec}(\mathbf{Y}) = \mathrm{vec}(\mathbf{ZB} + \mathbf{E}) = \mathrm{vec}(\mathbf{ZB}) + \mathrm{vec}(\mathbf{E})$$
$$= (\mathbf{I}_d \otimes \mathbf{Z})vec(\mathbf{B}) + vec(\mathbf{E}). \quad (5)$$

Then VARX(p,s) can be rewritten again in vectorized form as:

$$\underset{(dT\times 1)}{\mathbf{y}} = \underset{(dT\times dq)}{(\mathbf{I}_d \otimes \mathbf{Z})}\underset{(dq\times 1)}{\beta} + \underset{(dT\times 1)}{\varepsilon} \quad (6)$$

$= Z_{sup}\beta + \varepsilon, \quad \varepsilon \sim N(0, \Omega)$

where $\mathbf{y} = vec(\mathbf{Y})$, $\mathbf{Z}_{sup} = (\mathbf{I}_d \otimes \mathbf{Z})$, $\beta = vec(\mathbf{B})$, $\varepsilon = vec(\mathbf{E})$, $\mathbf{\Omega} = \mathbf{\Sigma} \otimes \mathbf{I}_T$, and $\otimes$ denotes the Kronecker product.

2.2.1 The loglihood function and maximum likelihood estimation

Assuming that the distribution of the model is Gaussian, to estimate the parameters of the VARX model, we have written the loglihood function given by:

$$logL(\beta, \Sigma) = -\frac{dT}{2}log(2\pi) - \frac{T}{2}log \mid \Sigma_\varepsilon \mid -\frac{1}{2}tr\left[(\mathbf{Y} - \mathbf{ZB})'\Sigma_\varepsilon^{-1}(\mathbf{Y} - \mathbf{ZB})\right]$$
$$= -\frac{dT}{2}log(2\pi) - \frac{T}{2}log \mid \Sigma_\varepsilon \mid -\frac{1}{2}\left[(\mathbf{y} - (I_d \otimes \mathbf{Z})\beta)'(\Sigma_\varepsilon^{-1} \otimes I_T)(\mathbf{y} - (I_d \otimes \mathbf{Z})\beta)\right]$$
$$= -\frac{dT}{2}log(2\pi) - \frac{T}{2}log \mid \Sigma_\varepsilon \mid -\frac{dT}{2} \quad (7)$$

After some work, the maximum likelihood (ML) estimator *of β becomes*

$$\hat{\beta} = vec(\hat{B}) = (Z'\hat{\Omega}_\varepsilon^{-1}Z)^{-1}\hat{\Omega}_\varepsilon^{-1}y, \quad (8)$$

where $\hat{\Omega}_\varepsilon = I_d \otimes \hat{\Sigma}_\varepsilon$. Hence

$$\hat{\beta} = \left[\hat{\Sigma}_\varepsilon^{-1} \otimes (Z'Z)^{-1}\right](\hat{\Sigma}_\varepsilon^{-1} \otimes Z')y$$
$$= \left[I_d \otimes (Z'Z)^{-1}Z'\right]y. \quad (9)$$

Now reshaping $\hat{\beta}$, in matrix form ,

$$\hat{B} = (Z0Z)-1)Z0Y \quad (10)$$

Further since

$$\hat{\varepsilon} = vec(\hat{E}) = y - Z\hat{\beta} = y - \hat{y} \quad (11)$$

and reshaping $\hat{\varepsilon}$ into $\hat{E}$, then the estimated covariance matrix of the innovations E is given by ,

$$\hat{\Sigma}_\varepsilon = \frac{1}{T}\hat{E}'\hat{E}. \quad (12)$$

We note that $\hat{\Sigma}_\varepsilon$ is a biased estimate of Σ_ε. Adjusting for the degrees of freedom an unbiased estimate of the covariance matrix is

$$\tilde{\Sigma}_\varepsilon = \frac{1}{T - dp - 1}\hat{E}'\hat{E}. \qquad (13)$$

Therefore the estimated covariance matrix of the coefficient matrix B is given by

$$\hat{\mathbf{Cov}}(\hat{\mathbf{B}}) = \hat{\Sigma}_\varepsilon \otimes (Z^0 Z)^{-1}. \qquad (14)$$

It has been shown (Hamilton, 1994) that the multivariate OLS estimators and the maximum likelihood ML-estimtors are identical since they acquire both similar asymptotic distribution under the assumption of Gaussianity.

In the next section, for us to opertionalize the Information Complexity (ICOMP) criterion and its other forms, we need to obtain the estimated Fisher Information Matrix (FIM) and its inverse (IFIM) for the d-dimensional VARX(p,s) model. Following Bearse and Bozdogan (1998) and Bearse and Bozdogan (2002)[8], the estimated inverse fisher information matrix (IFIM) is given by

$$\hat{\mathcal{F}}^{-1} = \begin{bmatrix} \hat{\Sigma} \otimes (Z'Z)^{-1} & 0 \\ 0' & \frac{2}{T} D_d^+ (\hat{\Sigma} \otimes \hat{\Sigma}) D_d^{+\prime} \end{bmatrix},$$

where $D_d^+ = (D_d' D_d)^{-1} D_d$ is the Moore-Penrose inverse of the duplication matrix, D_d;D_d is a unique $d^2 \times \frac{1}{2}d(d + 1)$ duplication matrix, such that $vec(\hat{\Sigma}) = D_d vech(\hat{\Sigma})$. In other words, $D_d^+ vec(\hat{\Sigma}) = vech(\hat{\Sigma})$.

For more on duplication matrices, we refer the readers for Magnus (1988), Magnus and Neudecker (1988), and Lutkepohl (2005, p. 387).

3. Information Criteria and Complexity under Specified VARX Models

Model Selection criteria are used to fit models under a specified parametric probability distribution by minimizing the overall risk that results from modeling and estimating the model under the maximum likelihood. For VARX models, model specification is to select optimally the lag order of the endogenous and exogenous variables. In other words, for a d-dimensional VARX(p,s), what is the optimal order (*p,s*) ?

In order to be able to answer this question, in this section, we introduce and save the information theoretic model selection criteria to choose the best VARX(p,s) to fit the time series data. The use of the conventional *F* and *t* tests are no longer valid when the significance of the OLS coefficient estimates is distorted.

AIC is inconsistent (in fact, it overestimates the true order with positive probability) while *ICOMP* (*IFIM*) is consistent and when $n > 1$, strongly consistent (i.e., they will choose the correct model almost surely). Intuitively, AIC is inconsistent because the penalty function used does not simultaneously goes to infinity as T ∞ and to zero when scaled by T. Although, consistent methods may have poor small sample properties. All of them are based on the maximal value of the likelihood function with an additional penalizing factor related to the number of estimated parameters. The suggested criteria differ regarding the strength of the penalty associated with the increase in model parameters as a result of adding more lags. The idea is to calculate the test criterion for different values of p and then choose the value of p that corresponds to the smallest value. When using these criteria for the choice of lag length, it is important to remember that they are valid under the assumption of a correctly specified model.

As we infered before, penalizing complexity by adding the term of twice the number of estimated parameters is not sufficient to overcome over-parametrization and unnecessary complexity within the chosen model. So, in contrast to AIC, and all the above mentioned different variants of AIC, Bozdogan developed a new Information complexity criterion based on the generalization of the information or entropic covariance complexity index of (van Emden,1971).

3.1 Conventional Model Selection Criteria, AIC, SBC, CAIC

Akaike's (1973) AIC reinforces the tradeoff between the maximized log likelihood (the lack of fit) which is an asymptotic estimate of the KullbackLiebler distance (KL) and the number of parameters estimated in the model (the penalty term measuring the complexity). Akaike's (1973) abounded the principle of parsimony by extending the entropic derivation of the method of maximum likelihood. The entropic derivation or interpretation is based on the maximization of the entropy or on the minimzation of the KullbackLeibler (KL) distance or information quantity.

The lack or bad of fit and the lack-of parsimony are penalized by negative twice the maximum loglikelihood and by twice the number of estimated parameters consecutively, and that is given by the following equation,

Akaike's information criterion (AIC): Let $\{L_m, m = 1,2,...,M\}$ be a set of competing models indexed by $m = 1,2,...,M$. Then,

$$AIC = -2logL(\hat{\theta}_m) + 2m. \quad (15)$$

where $logL(\hat{\theta})$ is the maximized log-likelihood function and $\hat{\theta}_m$is the maximum likelihood estimate of the parameter vector θ_s in the model $m = 1,2,...,$M

For multivariate Gaussian errors in VARX(p,s) model, with m being the number of the fitted unrestricted VARX estimates in the model, we have,

$$-2logL(\hat{\theta}_m) = \frac{dT}{2}log(2\pi) + Tlog \mid \hat{\Sigma}_\epsilon \mid + dT. \quad (16)$$

Hence,

$$AIC = dTlog(2\pi) + Tlog\left|\hat{\Sigma}_\epsilon\right| + dT + 2m.$$

$$m = d(dp + 1 + r(s+1)) + \frac{d(d+1)}{2}$$

(17)

Where (18)

Other information criteria in the literature are introduced but they are similar to AIC by using the lack of fit penalty term whereas they differ in penalizing the model complexity term through adding the second term. Schwarz (1978) proposed the criterion SBC given by,

$SBC = -2\log L(\hat{\theta}_m) + m\log(T)$, (19)

$$SBC = dTlog(2\pi) + Tlog\left|\hat{\Sigma}_\epsilon\right| + dT + mlog(T), \quad (20)$$

where $log(T)$ is the natural logarithm of the sample size T.

Choosing the minimal value of AIC that corresponds to the best model does not lead to an assympotic consistent estimate of the model order (Bozdogan, 2000), since the penalty term remains constant even if the first term of AIC increases as the sample size increase. So, Bozdogan (1987) extended AIC through different approaches to make AIC consistent . Consistent AIC or (CAIC) is defined as

$$CAIC = -2logL(\hat{\theta}_m) + m[log(T) + 1] \quad (21)$$

$$CAIC = dTlog(2\pi) + Tlog\left|\hat{\Sigma}_\epsilon\right| + dT + mlog(T+1), \quad (22)$$

3.2 Information Complexity Criteria

To determine the lag length p of the VARX model, the usual standard statistical procedures such as the likelihood ratio type procedure or sequential likelihood ratio procedure cannot be used to analyze high-dimensional time series data. This is due to the fact that overall level of significance is not known a priori as we fit $p = 1,2,...,p_{\max}$ VARX model. Moreover, it is difficult to decide on appropriate level of siginifcance to carry out the likelihood ratio and type test procedures.

Therefore, we introduce in this subsection the Bozdogan information complexity ICOMP criteria.

Consider a multivariate linear or non-linear structural model defined by

StatisticalModel = Signal + Noise

Then ICOMP [25] is structured to estimate the loss function as follows, Loss = Lackoffit + Lackofparsimony + ProfusionOfComplexity (23)

This loss estimate is achieved by the additive property of information theory and the entropic developments in (Rissanen, 1976) in his final estimation criterion for estimation and model identification problems, as well as Akaike's and its analytic extensions in (Bozdogan,1987).

ICOMP (IFIM) can explicitly regulate for the number of estimated parameters, the sample size, and adjust the risks of redundant unnecessary overparametrized models since model complexity depends intrinsically on other factors other than the dimension of variables, it takes into account for example parameter redundancy, parameter consistency, random error structure of the model, and linearity and non-linearity of the parameter estimates of the model.

Complexity is defined as the degree of interdependence among the components of the model, so the objective of Bozdogan was to incorporate a more judicious penalty term that prevents this complexity, therefore instead of counting and penalizing the number of parameters in a certain model, Bozdogan added an adaptively adjusted entropic complexity of the estimated inverse-fisher information matrix IFIM (known as Cramer-Rao lower bound matrix).

Profusion of complexity attributes into the interdependencies or the correlations among the parameter estimates and among the random error terms of the underlying model. So, this IFIM is a trade-off between the accuracy of the estimated model parameters and the interdependency of the innovations adhering a simpler model with minimum covariance matrix of the parameter estimates and minimum covariance matrix of the innovations too. A simple Model is always preferred to a more complex one as it was mentioned before. ICOMP is a more refined criteria that aim to balance the overfitting and underfitting risks withstanding in that given model.

Instead of penalizing the free parameters directly, ICOMP penalizes the covariance complexity of the model. Its complexity term of any random vector is a measure of the interaction or the dependency between its components. Hence ICOMP is given in (Bozdogan,1990)[6] as

$$ICOMP = -2logL(\hat{\theta}_m) + 2C(\hat{\Sigma}); \quad (24)$$

$$ICOMP = -2logL(\hat{\theta}_m) + 2C(\hat{\Sigma}); \quad (25)$$

where $logL(\hat{\theta}_m)$ is the natural logarithm of the maximized likelihood function, $\hat{\theta}_m$ is the maximum likelihood estimate of the parameter vector θ. $C(.)$ represents a real-valued complexity measure and $\hat{cov}(\hat{\theta}_m) = \hat{\Sigma}$ represents the estimated covariance matrix of the parameter vector of the model. Two forms of $C(.)$ are defined in Bozdogan's paper (Bozdogan,1990)[6].

ICOMP (IFIM) is defined as:

$$ICOMP(\hat{F}^{-1}) = -2logL(\hat{\theta}_m) + 2C_1(\hat{F}^{-1}). \quad (26)$$

where $C_1(\hat{F}^{-1})$ is the entropic complexity defined by :

$$\frac{rank(\hat{\mathcal{F}}^{-1})}{2}log(\frac{tr(\hat{\mathcal{F}}^{-1})}{rank(\hat{\mathcal{F}}^{-1})}) - \frac{1}{2}log \mid \hat{\mathcal{F}}^{-1} \mid \quad (27)$$

For the d-dimensional VARX (p,s) model, ICOMP(IFIM) is given by

$$\begin{aligned} ICOMP(\hat{\mathcal{F}}^{-1}) &= dTlog(2\pi) + Tlog \mid \hat{\Sigma} \mid + dT \\ &+ d(d+q)log\left(\frac{1}{d(d+q)}[tr(\hat{\Sigma})tr[(Z'Z)^{-1}] + \frac{1}{2T}(tr(\hat{\Sigma}^2) + tr(\hat{\Sigma})^2 + 2\sum_{j=1}^{d}(\sigma_{jj}^2)^2)]\right. \\ &- (d+q)log \mid \hat{\Sigma} \mid + dlog \mid Z'Z \mid + \frac{d(d+1)}{2}log(T) - dlog(2), \end{aligned}$$

!

(28)

where $(\sigma_{jj}^2)^2$ indicates the square of the jth diagonal element of $\hat{\Sigma}$, and where $q = dp+rs$ is the dimension of the matrix of lagged covariates Z . To compute $ICOMP(\hat{F}^{-1})_{PEU}$, one must simply add m to the above equation.

$ICOMP(\hat{F}^{-1})_{PEU} = ICOMP(\hat{F}^{-1}) + m$

$$= -2logL(\hat{\theta}_m) + 2C_1(\hat{F}^{-1}) + m, \quad (29)$$

and

$$ICOMP(\hat{\mathcal{F}}^{-1})_{PEU_{Miss}} = -2logL(\hat{\theta_m}) + 2\frac{m(T-p)}{m-T-p-2} + 2C_1(\hat{\mathcal{F}}^{-1}) + m. \quad (30)$$

CICOMP is defined by

$$\begin{aligned} CICOMP &= CAIC + 2C_1(\hat{\mathcal{F}}^{-1}) \\ &= dTlog(2\pi) + Tlog\left|\hat{\Sigma}\right| + dT + mlog(T+1) + 2C_1(\hat{\mathcal{F}}^{-1}). \quad (31) \end{aligned}$$

4. Numerical Examples

4.1 Monte Carlo Simulation Study

In this section, through a large monte Carlo Simulation, for space considerations, under one covariance pattern structure of the we show the practical utility and the empirical performance of model selection criteria. In what follows, we show the design of the protocol of the Monte Carlo simulation and our results of the percent relative frequency of the choice of the true lag order p of the VARX model where the lags coefficients of the regressors are ignored or in other words, we are simulating VARX(p,0) to study the performance of each criteria in terms of their hit ratios and overfitting and underfitting percentages. Our computations are carried out using MATLAB modules that we have developed for this study as our computational platform.

4.2 Simulation: 3×1-D VARX(2,0) Model

In this simulation example, we consider a stationary 3× 1-dimensional VARX(2,0) model with a known homogeneous and correlated innovation covariance. The simulation protocol is given by

$$\mathbf{y}t = \mathbf{c} + \Phi 1\mathbf{y}t-1 + \Phi 2\mathbf{y}t-2 + \Theta 0\mathbf{x}t + \varepsilon t \quad (32)$$

The multivariate time series innovation is generated from a multivariate Gaussian distribution with zero mean vector and structured covariance matrix Σ. That is, $\varepsilon_t \sim N(0,\Sigma)$, $t = 1,2,...,T$. We repeat our simulation 1000 times. To our knowledge such a study has not been done before in the literature to study the efficacy of the model selection criteria of VARX process.

where $\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} \sim N\left(\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{bmatrix}\right)$.

Based on the results in tables 1 and 2, the best performing information criteria are: *SBC*, *CAIC*, *ICOMP*(*IFIM*), $ICOMP_{PEU}$,$ICOMP_{PEUMiss}$, and *CICOMP at* sample size T = 600. Their underfitting or overfitting reach 0.0%. As we note there is a slight overfitting in *AIC* (4.1%). However, this overfitting diminish as we increase the number of time series observations T.

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} 1 \\ 0.5 \\ -0.5 \end{pmatrix} + \begin{pmatrix} 0.3 & -0.1 & 0.05 \\ 0.1 & 0.2 & 0.1 \\ -0.1 & 0.2 & 0.4 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \begin{pmatrix} 0.1 & 0.05 & 0.001 \\ 0.001 & 0.1 & 0.01 \\ -0.01 & -0.01 & 0.2 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \\ y_{3t-2} \end{pmatrix} + + \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x_{1t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

And it is not so surprising since AIC has the tendency of overfitting. What this means is that just counting and penalizing the number of parameters may not be sufficient to capture the correlated stucture in the innovations. The average of underfitting and overfitting of *ICOMP* reach 0.0% and 0.0% respectively. As a result, a common consensus reveals that the *ICOMP* is the most efficient Information criteria capable of capturing or identifying the correct true parsimonious lag structure for larger dimensional VARX models and of being consistent with the increase in the sample size. Therefore, we can declare that *SBC*, *CAIC*, *ICOMP*(*IFIM*), $ICOMP_{PEU}$, $ICOMP_{PEUMiss}$, and *CICOMP* are consistent criteria of VARX order estimation. For more consistency of model selection criteria, we refer the readers to Bozdogan (1987).

4.3 Real Data Example: Model Selection and Forecasting of S&P500 stock portfolio

We consider an application based on predicting the movement of the U.S. S&P's 500 stock prices minute by minute. We use updated data from the STATWORX team (https://medium.com/mlreview/a-simple-deep-learning-\model-for-stock-price-prediction-using-tensorflow-30505541d877) members who scraped minute by minute the data from the Google Finance API spanning 2017: 04-03: 09.01. The data consisted of index as well as the stock prices of the S& P 500 constituents. Thus our goal is to select the best VARX lag parameter using *ICOMP* model selection criteria and thus predict the direction of the flow of holding on or selling the portfolio constructed of the S&P 500 stocks based on the 500 constituents prices one minute ago and compare the out-of-sample forecast.

4.3.1. Dimension Reduction via SPCA

VARX model estimation on a high dimensional data set will definitely lead to an increase in the size of the model which consecutively makes OLS estimation impossible. Thus variable selection is one of the techniques used to reduce the dimension of the model. For example, to produce an efficient portfolio constructed from large number of stocks, financial or risk managers have to estimate the covariance matrix of the portfolio assets and which would be impossible unless dimension reduction techniques are adopted.

Referring to (Zou et al.,2006),Principle Component Analysis has been widely used since PCs explain the maximum variance among the variables enhancing the minimal information loss and all the PCs are uncorrelated where each principle component is a linear combination of all the variables presented in the model.

Thus (Zou et al.,2006) introduced the Sparse principle Components (SPCA) where they imposed an elastic net (lasso) $L1$ Norm penalty on the regression optimization and will eventually emit a new interpretable loading factor matrix B with possible zero loadings and a new variance structure explained by the SPCs. While the Elastic-net optimization allows including all the highly correlated variables, the Lasso (Tibshirani, 1996) chooses one variable from a highly correlated set of variables since those highly correlated variables produce same information on the response.

The steps of the procedure of **SPCA algorithm** can be summarized as follows (Zou et al., 2006),

1) PCA is computed either using the covariance decomposition or thethrough the Singular-Value-Decomposition (SVD) of the data represented as a centred matrix X.

X is an $n \times p$ matrix where n is the number of observations and p is the dimension or number of variables. If X is centred, or in other words, all the columns of X have mean zero and if the SVD of X is given by $X = UDV^T$ where U is an $(n \times n)$ unitary matrix and D is an $(n \times p)$ diagonal matrix of singular values D_{ii}, then $PC = UD$. The loadings of the principal components is the column vectors of V and the sample variance of the each PCA is $\frac{D_{ii}^2}{n}$ (Zou et al., 2006).

2) Let A start at $V[,1:k]$, the loadings of the first k ordinary principle components.

3) Given a fixed $A = [\alpha_1,...\alpha_k]$, solve the following elastic net problem for $j = 1,..k$ $\beta_j = \text{argmin}(\alpha_i - \beta)^T X^T X(\alpha_i - \beta) + \lambda k\beta k^2 + \lambda_{1,j} k\beta^k{}_1$ (33)

4) For a fixed $B = [\beta_1,...\beta_k]$, compute the SVD of $X^T XB = UDV^T$, then update $A = UV^T$.

5) Repeat Steps 2-3 until convergence or in other words untill we get thedesired number of non-zero loadings.

6) Normalization: $\hat{V}_j = \frac{\beta_j}{\|\beta_j\|}, j = 1, ..k$.

The best tuning parameter $\lambda_{1,j}$ is chosen by the above algorithm that approximates the sparse of each PCA promoting the best trade-off between variance and sparsity.

The entire regularization path of the parameters (λ in eq. (28)) varies from zero active variables (a high value of λ)to the point where $\lambda_{1,j}$ criterion is met.

We chose λ equal to infinity and $\lambda_{1,j}$ equal to [100,150,200] as the number of non-zero coefficients chosen for the corresponding three SPCs in the data application part below.

First, The data is centred by subtracting the mean of each column of the corresponding 500 S&P500 stocks and then Sparse Principle Component Analysis (SPCA) is applied to reduce the dimension of the 500 stocks into 3 sparse components SPCs. Then the sample data was partitioned at 90:10 that is 90% for training (used for VARX model building) and 10% for validation or testing out-of sample.

These components are nothing but the linear combination of the existing 500 variables (companies stocks) arranged in decreasing order of their variances. In other words, the first SPC explains the

maximum amount of the variance in the data. SPCA converts the inputs which is the highly correlated data and outputs SPC's that explain the variance in the data to make the data uncorrelated with each other. Table 1 shows the percentage explained variances by the first three SPC's which is 56.11%.
We note that after the third SPC, the percent explained variances decreases monotonically, showing that the first three SPCs are sufficient to explain the original data without much loss of information.

3. Result

Results of the VARX Simulation:

Table 1. Percent Relative Frequency of Choosing True 3 × 1 − *D* VARX(2,0) Model, T=200

$T = 200$		Lag Length				Percent	
Criteria\V AR Order	1	**2∗**	3	4	5	*Overfitting*	*Underfitting*
AIC	0.0	**94.3**	4.7	0.8	0.2	5.7	0.0
SBC	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0
CAIC	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0
ICOMP(*IFIM*)	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0
$ICOMP_{PEU}$	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0
$ICOMP_{PEUMiss}$	0.3	**99.7**	0.0	0.0	0.3	0.0	0.3
CICOMP	9.4	**90.6**	0.0	0.0	0.0	0.0	9.4

Table 2. Percent Relative Frequency of Choosing True 3 × 1 − *D* VARX(2,0) Model, T=600

$T = 600$		Lag Length				Percent	
Criteria\V AR Order	1	**2∗**	3	4	5	*Overfitting*	*Underfitting*
AIC	0.0	**95.9**	3.6	0.4	0.1	4.1	0.0
SBC	0	**100.0**	0.0	0.0	0.0	0.0	0.0
CAIC	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0
ICOMP(*IFIM*)	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0
$ICOMP_{PEU}$	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0
$ICOMP_{PEUMiss}$	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0
CICOMP	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0

Table 3. Explained Variances of SPCs

	Percentage Explained Variances
SPC1	31.36
SPC2	14.74

SPC3	10.01

SPCA outperforms the conventional Principal Component Analysis (PCA), since the loading matrix gives the most significant stocks among the whole S&P500 stocks. In Table 2 we show the classification of the most 37 significant S&P500 stocks corresponding to six different sectors which have nonzero rows of the loading factor matrix B.

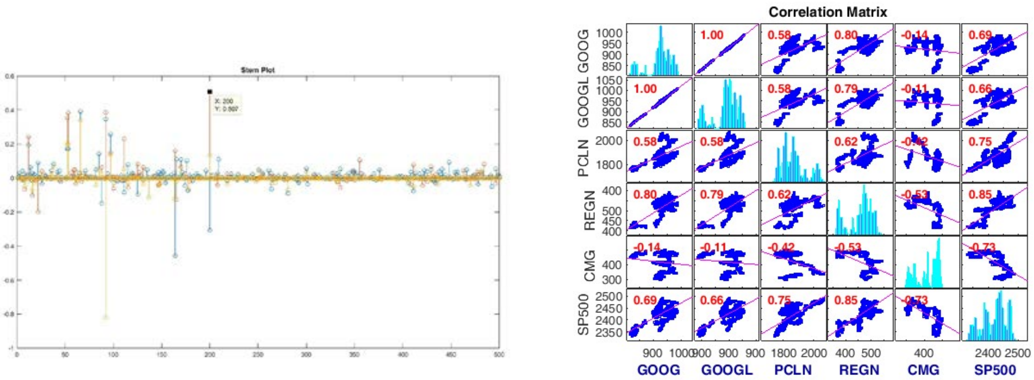
Based on the SPCA, it can be concluded that the Health Care Sector, followed by the Consumer Discretionary, Industrial and Technology sectors compose the highly significant number of stocks among the 37 chosen. We

Table 4. Classification of the Relevant Stocks Extracted from the three SPC's

Discretionary	Health	Industrials	Staples	Financials	Energy	Technology	Materials
ORLY	ALXN	MMM	COST	AMG	BHI/BHGE	GOOG	
BKNG	AMGN	COL		AMP	CXO	GOOGL	
DISCA	BIIB	GWW			PXD	INTU	PX
ULTA	CELG	BA				LRCX	SHW
AZO	REGN						
	VRTX					ADS	
BDX	EW						
CMG	HUM						
FL	LH						
	TMO						
	UHS						
	WAT						

can interpret the loadings after estimating the loading matrix B through SPCA. The stem plot of the loading matrix B shows graphically in Figure 1 (a), the non-zero loadings related to the highly significant contributing stocks in the SPC's. Thus, it can deduced that several portfolios can be created to mimic the factor loadings.

We note that matrix B laodings range is in the range [|0.0006|,|0.8221|]. But for space considerations, we have chosen one portfolio which constitutes of the highest loading negative or positive weights in matrix *B* that falls in the interval [|0.1|,|0.8|], as follows,



(a) Stem plot of the loading matrix B of the SPCA. (b) Correlation plot of the portfolio constructed.

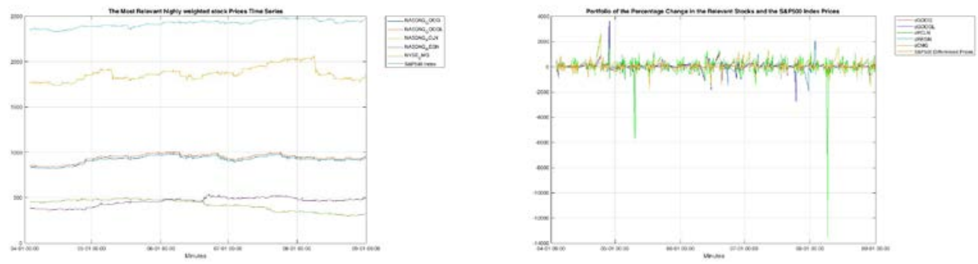
Figure 1. Stem Plot of the Selected 5 Stocks and Correlation Plot

0.1736	0.3536	0.2013
0.1719	0.3831	0.2026
B=0.3453	0.3876	−0.8221
0.2578	0.1487	0.1368
	0.5070	0.1336
−0.3075		

This set of portfolio is formed from the following stocks, [GOOG, GOOGL, BKNG, REGN, CMG], plus the S&P500 index itself. Figure 1 (b) shows the correlation plot. It can be inferred that this portfolio sample is noisy. All the recorded T = 41,266 successive trading minutes (i.e., observations) of the above constructed portfolio up to the end of August 31, 2017 are displayed in Figure 2 (a).

a. Testing the Stationarity

Some statistical diagnostics are investigated to assess the validity of the stationarity of the 5-dimensional VARX model for the chosen portfolio using autocorrelation (ACF) and partial autocorrelations functions (PACF). ACF is shown in Figure 3. As discussed in (Granger & Newbold, 1976) differencing the time series of the portfolio is one kind of transformation that can be used to address non-stationarity and remove the spurious regression.



(a) The Portfolio plot **before** percent differ-(b) The portfolio plot **after** percent differencing. encing.

Figure 2. The Portfolio Plot before and after Percentage

Differencing of the multivariate time series. effects caused by cointegration, thus reliable forecasts are obtained. Let

and consider

$$Y_t = 100(P_t - P_{t-1}) \tag{34}$$

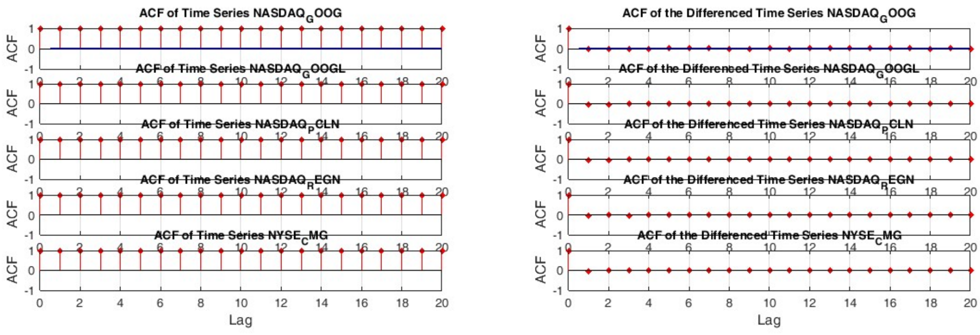
$$Y_t = \begin{matrix} Yt,1 \\ Yt,2 \\ Y_{t,3} \\ Yt,4 \\ Yt,5 \end{matrix} \quad ,t=1,...4,1266$$

Looking at Figure 2(b), we see that the transformed 5-dimensional multivariate time series appear to be stationary. However observations appear serially correlated. In other words, the lagged regressors are not predetermined. Thus an exogeneous or an instrument variable should be used.

b. Lag Order Selection of VARX for S&P500 Stocks

Next, we estimate a multivariate time series *V ARX* model that contains lagged endogenous and exogenous variables. The response series is the percentage change in Stock prices minute by minute, while the exogenous time series is the percentage change in the S&P500 index.

The best choice among the competing models is determined by some of the well-known information criteria mentioned above, especially producing ICOMP (IFIM) as an efficient consistent criteria.



(a) ACF **before** % Differencing. (b) ACF **after** % Differencing.

Figure 3. Autocorrelation ACF before and after Percent Differencing

Looking at Table 5, *ICOMP N*(*IFIM*) is minimized at VARX(4,0) model showing that VARX(4,0) is the best parsimonious model. The estimated covariance matrix of the innovations is given below.

2840.7435	2656.8020	1683.0826	416.5625	69.3029
2656.8020 Σ = ˆ	3126.9337	1793.1231	437.5604	55.8588
1683.0826	1793.1231	17127.7013	584.2448	162.8382
	437.5604	584.2448	2589.7513	27.8892
416.5625	55.8588	162.8382	27.8892	1378.0130
69.3029				

Whereas, *CAIC* and *SBC* select lag 6. *AIC* overfits the model and chooses lag 8 as the best VARX lag parameter which is not so surprising result since *AIC* imposes a weaker penalty for higher lag orders and consequently overfits the model.

d. Out-of-Sample Forecasting

Using the best VARX lag parameter, we construct the MSE out-of- sample forecasts for the corresponding 5-dimesnsional time series and detect its minute by minute movement direction.

The out-of-sample calculations are used in predicting the conditional mean time series. We estimated the approximate 95% forecast intervals for the best VARX(4,0) fitting model and calculated the mean squared errors (MSE) forecasts.

Table 5. Model selection results from fitting VARX(4,0) model

	AIC	*CAIC*	$ICOMP_{IFIM}$	*SBC*	*ICOMPPEUMiss*
Lag1	1986264.817	1986598.101	1987342.062	1986563.101	1987447.132
Lag2	1983359.652	1983930.994	1985035.297	1983870.994	1985215.498
Lag3	1982124.905	1982934.304	1984390.584	1982849.304	1984645.983
Lag4	1981285.299	1982332.753	**1984134.959**	1982222.753	**1984465.625**
Lag5	1980818.996	1982104.505	1984247.490	1981969.505	1984653.490
Lag6	1980547.914	**1982071.476**	1984551.098	**1981911.476**	1985032.501
Lag7	1980344.031	1982105.645	1984918.189	1981920.645	1985475.061
Lag8	**1980270.113**	1982269.777	Inf	1982059.777	Inf

Figure 4 shows the Mean Squared Erros (MSE) forecast predictions of the out-of sample and the 1-trading day prediction.

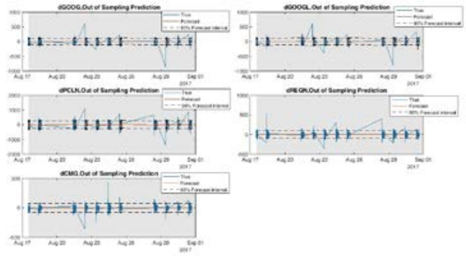
Figure 4(a) shows the 95% forecast interval of the fitted VARX(4,0) model of the percent differenced data set. While Figure 4(b), displays the 1-trading day non-transformed prices prediction in comparison to the prices of the last day in the original non-transformed data set, the gap in the figure refers to the non-trading hours at night when the stock market is closed.

The accuracy of the forecasts of the fitted VARX(4,0) is calculated by using the root Mean Squared Error (RMSE). MSE and RMSE are given by

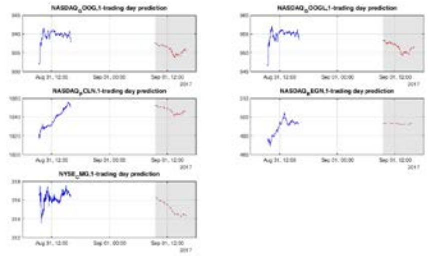
$$MSE = \frac{1}{T}\sum_{i=1}^{T}(Y_i - \hat{Y}_{i+h})^2 \quad (35)$$

$$RMSE = \sqrt{MSE} \quad (36)$$

Table 6 illustrates the basic performance test by comparing the root mean square error (RMSE) of the out-of-sample forecasts to the RMSE of a simple, baseline forecast that holds the last in-sample value of the response constant. If the model forecast does not significantly improve on the baseline forecast, then it is reasonable to suspect that the model has not captured the relevant variables in the VARX model (Diebold, 2007). However in our case, Table 6 shows that RMSE baseline forecast RMSE Base is larger than the RMSE of the model RMSE Pred, therefore the fitted model has captured the best



(a) Out-Of-Sampling Forecast of the % Differenced 5-dimensional Times Series.



(b) The 1-Trading day Prediction of the Pricesof the 5-dimensional Time Series.

Figure 4: Out-of-sampling forecast and the 1-trading day prediction contributing variables of the corresponding 5-dimensional fitted VARX(4,0). It can be noted too that the RMSEPred and RMSEBase of VARX(4,0) are smaller than those of VARX(6,0). This can assert that ICOMP performs well in choosing the correct best lag parameter.

Table 6. The Prediction Accuracy of the Portfolio with Stationary Return Time Series

Multivariate Gaussian Innovations	**RMSEPred**	**RMSEBase**
Multivariate OLS Estimation of VARX(4,0), selected by ICOMP	**1.3009257**	**3.8486868**
Multivariate OLS Estimation of VARX(6,0), selected by SBC	**1.300932**	**3.8487068**

d. Directional Movement of the Portfolio

To determine the directional movement of the portfolio constructed from the five stocks, we have to construct the one-step-out-of-sample forecasts of the prices minute by minute. According to Akbilgic and Bozdogan (2014), each forecast is translated into a directional sign +1 and −1. The +1 sign refers to an increase in the price value by the next minute, while −1 indicates a drop in value. Consequently, increase and decrease in value can be interpreted too as buy-sell signal decision.

Table 11 shows how to manage an investment according to the out-of-sample forecasts for the original time series prices for **GOOGL** stock only for the first 10 minute-period by simply using our best fitted VARX(4,0) model. It can be discovered from Table 7 that the Buy-Sell decision making in response to VARX(4,0) for **GOOGL** outperforms investing and holding the position over the out-of-sample period. Out of all the 4,126 periods of minutes' forecasts (out-of-sample periods), 2,071 periods recognized correctly the direction of the portfolio price movement.

Table 7. Direction Prediction of the Portfolio Stock Movement

Out-of-sample forecast periods	GOOGL	Forecast	Buy-Sell decision by forecast
Period 1	+	−	Sell
Period 2	−	+	Buy
Period 3	+	+	Keep
Period 4	−	+	Keep
Period 5	+	−	Sell
Period 6	−	+	Buy
Period 7	−	−	Sell
Period 8	+	+	Buy
Period 9	−	−	Sell
Period 10	−	−	Keep

4. Discussion

The paper presents a hybrid novel approach of VARX-SPCA with emphasis of Bozdogan's Information Complexity Criterion efficacy under OLS estimation and one real application data set, having VARX as the vector autoregressive model that allows for endogenous and exogenous variables and SPCA as the sparse principal component analysis employed to reduce the dimension of the S&P500 multivariate time series. First, it has been inferred that Bozdogan ICOMP outperforms all the other information criteria in selecting the simulated true $3 \times 1 - D$ VARX(2,0) since ICOMP penalizes the covariance complexity of the model instead of penalizing the free parameters directly. Second, the application data set indicates that the Buy-Sell decision making in response to VARX(4,0) for **GOOGL** outperforms investing and holding the position over the out-of-sample period. However, Further studies aim to highlight the effect of Feasible Generalized Least Squares Estimation (FGLS) on VARX simulated model selection under misspecification where the Gaussianity assumption of the innovations is relaxed and other distributions are considered.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267-281). Budapest: Academiai Kiado.

Akbilgic, O., Bozdogan, H., & Balaban, M. E. (2014). A novel Hybrid RBF Neural Networks model as a forecaster, *Statistics and Computing*, *3*, 365-375.

Bearse, P. M., & Bozdogan, H. (1998). Subset selection in vector autoregressive models using the genetic algorithm with informational complexity as the fitness function. *Systems Analysis Modelling Simulation*, *31*, 61-91. Gordon and Breach Science Publishers, Inc.

Bozdogan, H. (1987). ICOMP: A new model-selection criterion, 1. *Conference of the International Federation of Classification Societies*, inproceedings (pp. 599-608).

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345-370.

Bozdogan, H. (1990). On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics-Theory and Methods*, *19*, 221-278.

Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of mathematical psychology*, *44*, 62-91.

Bozdogan, H., & Beasrse, P. (2002). Multivariate regressions, Genetic Algorithms, and Information Complexity: A Three Way Hybrid. *Measurement and Multivariate Analysis* (pp. 269-278).

Bozdogan, H., & Beasrse, P. (2003). Information complexity criteria for detecting influential observations in dynamic multivariate linear models using the genetic algorithm. *Journal of Statistical Planning and Inference*, *114*, 31-44.

Bru¨ggemann, R. (2004). Model reduction methods for vector autoregressive processes. *Springer Science & Business Media*, *536*.

Fildes, R. (2008). In F. X. Diebold (Ed.), Elements of Forecasting (4th ed.). *International Journal of Forecasting*, *24*, 552-553. Retrieved from https://EconPapers.repec.org/RePEc:eee:intfor:v:24:y:2008:i:3:p:552553

Granger, C. W. J., & Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 189-203.

Hamilton, J. D. (1994). Time series analysis, *2*. Princeton University Press: Princeton.

Hendry, D. F., & Krolzig, H.-M. (1999). In K. D. Hoover, & S. J. Perez (Eds.), Improving on "Data mining reconsidered". *The econometrics journal*, *2*, 202-219.

Hendry, H lawka, M., & Zagdan´ski, A. (2014). Model selection criteria for reduced rank multivariate time series: a simulation study. *Journal of Statistical Computation and Simulation*, *84*, 1851-1883.

Hoover, K. D., & Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *The econometrics journal*, *2*, 167-191.

Howe, A., & Bozdogan, H. (2010). Predictive subset VAR modeling using the genetic algorithm and information complexity. *European Journal of Pure and Applied Mathematics*, *3*, 382-405.

Lu¨tkepohl, H. (2005). *New introduction to multiple time series analysis*, *Springer Science & Business Media*.

Magnus, Jan R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics, Wiley series in probability and mathematical statistics*.

Magnus, J. (1998). *Linear Structures*. New York: Oxford University Press.

Nicholson, W. B, Matteson, D. S., & Bien, J. (2017). VARX-L: Structured regularization for large vector autoregressions with exogenous variables. *International Journal of Forecasting*, *33*, 627-651.

Pamuk¸cu, E., Bozdogan, H., & C¸alık, S. (2015). A novel hybrid dimension reduction technique for undersized high dimensional gene expression data sets using information complexity criterion for cancer classification. *Computational and mathematical methods in medicine*.

Reboredo, J. C., Mat´ıas, J. M., & Garcia-Rubio, R. (2012). Nonlinearity in forecasting of high-frequency stock returns. *Computational Economics*, *40*, 245-264.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*, 465-471.

Salim, J., Ren, Y., & Fang, X. (2017). Real Estate Bubbles in China, Causes and Future Expectations/Solutions. *International Journal of Trade, Economics and Finance*, *8*, 179-183.

Schwarz, G., & others. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*, 461-464. Institute of Mathematical Statistics.

Sims, C. (1980). Macroeconomics and reality. *Econometrica*, *48*, 1-48.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

Van Emden, M. H. (1971). An Analysis of Complexity. *Mathematical Centre Tracts*, *35*, Amsterdam, Google Scholar.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, *15*, 265-286.