International Journal of Data and Network Science 8 (2024) 523–536

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Data forecasting performance evaluation of threshold spatial vector autoregressive with exogenous variables

Gama Putra Danu Sohibien[a,b], Setiawan[a*] and Dedy Dwi Prastyo[a]

[a]*Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia*
[b]*Department of Applied Statistics, Politeknik Statistika STIS, Jakarta 10710, Indonesia*

C H R O N I C L E

Article history:
Received: July 4, 2023
Received in revised format: July 27, 2023
Accepted: August 31, 2023
Available online: September 1, 2023

Keywords:
Spatial
Threshold
SpVAR
Nonlinear
Vector Autoregressive
Simulation

A B S T R A C T

One time series model developed to predict economic data is Spatial Vector Autoregressive with Exogenous Variables (SpVARX). This model can accommodate simultaneously the interrelationships between variables, the impact of exogenous variables, and Inter-regional linkages. However, this model has not adjusted the nonlinearity relationships between variables. The relationship between economic variables is usually not linear. Therefore, we introduce the Threshold Spatial Vector Autoregressive with exogenous variables (TSpVARX). This paper assesses the forecasting performance of TSpVARX and compares it with SpVARX models. We conducted a simulation study by generating 100 times the simulation data with twelve scenarios. We found that the forecasting performance of the TSpVARX model is better than SpVARX when there is a nonlinear relationship between variables. In addition, we find that the forecasting performance of TSpVARX models will improve as the sample size increases.

1. Introduction

Several models have been created for forecasting, especially economic data. The most popular univariate model is the Autoregressive Moving Average (ARIMA). This model can be used to forecast by making the lag of the endogenous variable a predetermined variable (Kong et al., 2022). However, this model cannot capture the impact of exogenous variables, so it triggers the ARIMA model's development with exogenous variables, consisting of transfer functions, ARIMA with intervention variables, and ARIMA with Outliers. Economic variables are often related to each other (Sohibien, 2018). Both ARIMA and ARIMA with exogenous variables can still not capture the interrelationships between endogenous variables. It became the forerunner of the development of Vector Autoregressive (VAR) and Vector Autoregressive with exogenous variables (VARX).

VAR and VARX models can be used to predict by making all lag endogenous variable into predetermined variables so that the interrelationships between endogenous variables can be accommodated (Rajab et al., 2022; Wang et al., 2021; Andreas et al., 2022). Meanwhile, VARX can accommodate the influence of exogenous variables, which VAR does not have (Zhu, 2021; Sohibien et al., 2022). However, in addition to the interrelationships between economic variables, the influence of endogenous variables in one area on other areas is also very possible. Spatial Vector Autoregressive (SpVAR) is a model that can overcome the shortcomings of ARIMA, ARIMAX, VAR, and VARX models that cannot capture inter-regional linkages. SpVAR can accommodate connections between regions for multiple endogenous variables (Ramajo et al., 2017; Andrés-Rosales et al., 2021). SpVAR evolved into the SpVAR with calendar variations to capture exogenous variables in seasonal patterns with varying periods.

* Corresponding author.
E-mail address: setiawan@statistika.its.ac.id (Setiawan)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2023.9.004

Another aspect that also needs to be considered in the time series model is the possibility of nonlinear relationships between variables. The relationship between economic variables is usually not linear. If the nonlinear relationship is not captured in the model, it can lead to incorrect estimation of the model coefficient (Balke & Fomby, 1997; Stigler, 2010). Several models have been developed to accommodate nonlinear relationships and interrelationships between variables, namely Threshold Vector Autoregressive (TVAR). We make this model by creating multiple VAR models that will be divided into different regimes (Yuhan & Sohibien, 2018; Jiang et al., 2021). Each regime will produce different coefficients of the VAR model. The TVAR model then evolved into a TVAR model with an exogenous metric variable (E-TVAR) to capture the influence of the exogenous metric variable (Tsagkanos et al., 2018).

Based on the previous explanation about the models that have been developed, we can see that several things are essential to include in time series modeling, especially economic data, namely interrelationships between variables, interrelationships between areas, the influence of exogenous variables, and nonlinear relationships between variables. However, there is currently no model that can accommodate all four things simultaneously. Therefore, in this study, we want to introduce the Threshold Spatial Vector Autoregressive model with Exogenous Variables (TSpVARX). This model can include these four things simultaneously to improve time series data forecasting. Based on literacy studies, we have not found any research that evaluates the forecasting performance of the TSpVARX by using simulation studies in many scenarios. This study aims to show the TSpVARX model's forecasting performance and compare it with SpVARX through simulation studies. It is imperative to get more convincing results related to the forecasting performance of the TSpVARX compared to SpVARX.

2. Methodology

2.1. SpVARX

The SpVARX model with spatial lag 1, autoregressive lag length *p*, and exogenous variable lag length *q* can be written as SpVARX (1,*p*,*q*). The form of SpVARX (1,*p*,*q*) for *N* areas, *K* endogenous variables, and one exogenous variable is as follows (Ramajo et al., 2017; Andrés-Rosales et al., 2021):

$$\begin{aligned} S_{1,t}^{1} &= \tau^{1}_{10} + \delta_{11}^{1,(1)} R_{1,t-1}^{1} + \cdots + \delta_{11}^{1,(q)} R_{1,t-q}^{1} + \cdots + \delta_{1m}^{1,(i)} R_{m,t-1}^{1} + \cdots + \delta_{1M}^{1,(1)} R_{M,t-1}^{1} + \cdots + \delta_{1M}^{1,(q)} R_{M.t-q}^{1} + \eta_{11}^{1,(1,0)} S^{1}_{1,t-1} \\ &+ \eta_{11}^{1,(1,1)} \left(w_{11(1,2)} S_{1,t-1}^{2} + \cdots + w_{11(1,u)} S_{1,t-1}^{u} + \cdots + w_{11(1,N)} S_{1,t-1}^{N} \right) + \cdots + \eta_{1r}^{1,(j,0)} S^{1}_{r,t-j} \\ &+ \eta_{1r}^{1,(j,1)} \left(w_{1r(1,2)} S_{r,t-j}^{2} + \cdots + w_{1r(1,u)} S_{r,t-j}^{u} + \cdots + w_{1r(1,N)} S_{r,t-j}^{N} \right) + \cdots + \eta_{1K}^{1,(p,0)} S_{K,t-p}^{1} \\ &+ \eta_{1K}^{1,(p,1)} \left(w_{1K(1,2)} S_{K,t-p}^{2} + \cdots + w_{1K(1,u)} S_{K,t-p}^{u} + \cdots + w_{1K(1,N)} S_{K,t-p}^{N} \right) + \varepsilon_{1t}^{1}, \end{aligned} \tag{1a}$$

$$\vdots$$

$$\begin{aligned} S_{kt}^{n} &= \tau_{k0}^{n} + \delta_{k1}^{n,(1)} R_{1,t-1}^{n} + \cdots + \delta_{k1}^{n,(q)} R_{1,t-q}^{n} + \cdots + \delta_{km}^{n,(i)} R_{m,t-i}^{n} + \cdots + \delta_{kM}^{n,(1)} R_{M,t-1}^{n} + \cdots + \delta_{kM}^{n,(q)} R_{M,t-q}^{n} + \eta_{k1}^{n,(1,0)} S_{1,t-1}^{n} \\ &+ \eta_{k1}^{n,(1,1)} \left(w_{k1(n,1)} S_{1,t-1}^{1} + \cdots + w_{k1(n,n-1)} S_{1,t-1}^{n-1} + w_{k1(n,n+1)} S_{1,t-1}^{n+1} + \cdots + w_{k1(n,N)} S_{1,t-1}^{N} \right) + \cdots + \eta_{kr}^{n,(j,0)} S_{r,t-j}^{n} \\ &+ \eta_{kr}^{n,(j,1)} \left(w_{kr(n,1)} S_{1,t-j}^{1} + \cdots + w_{kr(n,n-1)} S_{r,t-1}^{n-1} + w_{kr(n,n+1)} S_{r,t-1}^{n+1} + \cdots + w_{kr(n,N)} S_{r,t-j}^{N} \right) + \cdots + \eta_{kK}^{n,(p,0)} S_{K,t-p}^{n} \\ &+ \eta_{kK}^{n,(p,1)} (w_{kK(n,1)} S_{K,t-p}^{n} + \cdots + w_{kK(n,n-1)} S_{K,t-1}^{n-1} + w_{kK(n,n+1)} S_{K,t-1}^{n+1} + \cdots + w_{1K(n,N)} S_{K,t-p}^{N}) + \varepsilon_{kt}^{n}, \end{aligned} \tag{1b}$$

$$\vdots$$

$$\begin{aligned} S_{Kt}^{N} &= \tau_{K0}^{N} + \delta_{K1}^{N,(1)} R_{1t-1}^{N} + \cdots + \delta_{K1}^{N,(q)} R_{1,t-q}^{N} + \cdots + \delta_{Km}^{N,(i)} R_{m,t-i}^{N} + \cdots + \delta_{KM}^{N,(1)} R_{M,t-1}^{N} + \cdots + \delta_{KM}^{N,(q)} R_{M,t-q}^{N} + \eta_{K1}^{N,(1,0)} S_{1,t-1}^{N} \\ &+ \eta_{K1}^{N,(1,1)} \left(w_{K1(N,1)} S_{1,t-1}^{1} + \cdots + w_{K1(N,u)} S_{1,t-1}^{u} + \cdots + w_{K1(N,N-1)} S_{1,t-1}^{N-1} \right) + \cdots + \eta_{Kr}^{N,(j,0)} S_{r,t-j}^{N} \\ &+ \eta_{Kr}^{N,(j,1)} \left(w_{Kr(N,1)} S_{r,t-j}^{1} + \cdots + w_{Kr(N,u)} S_{r,t-j}^{u} + \cdots + w_{Kr(N,N-1)} S_{r,t-j}^{N-1} \right) + \cdots + \eta_{KK}^{N,(p,0)} S_{K,t-p}^{N} \\ &+ \eta_{KK}^{N,(p,1)} (w_{KK(N,1)} S_{K,t-p}^{1} + \cdots + w_{KK(N,u)} S_{K,t-p}^{u} + \cdots + w_{KK(N,N-1)} S_{K,t-p}^{N-1}) + \varepsilon_{Kt}^{N}, \end{aligned} \tag{1c}$$

where *N* is the amount of the area; *K* is the amount of endogenous variables; $R_{M,t-q}^{n}$ is the *M*-th metric exogenous variable at the *n*-th area and the *q*-th lag; S_{kt}^{n} is the *k*-th endogenous variable at the *n*-th area; $S_{r,t-j}^{u}$ is the *r*-th endogenous variable at the *u*-th area and the *j*-th lag; τ_{k0}^{n} is the interception of the SpVARX(1,*p*,*q*) model for equations at the *k*-th endogenous variable and the *n*-th area; $\delta_{kM}^{n,(q)}$ is the model coefficient for the exogenous metric variable of $R_{M,t-q}^{n}$; $\eta_{kr}^{n,(j,0)}$ is the model coefficient for the endogenous lag variable $S_{r,t-j}^{n}$; $\eta_{kr}^{n,(j,1)}$ is the space-time coefficient for $w_{kr(n,1)} S_{1,t-j}^{1} + \cdots + w_{kr(n,n-1)} S_{r,t-1}^{n-1} + w_{kr(n,n+1)} S_{r,t-1}^{n+1} + \cdots + w_{kr(n,N)} S_{r,t-j}^{N}$; $w_{kr(n,N)}$ is the spatial weight between S_{kt}^{n} and $S_{r,t-j}^{N}$, with *j*= 1, 2, …, *p* ; ε_{kt}^{n} is the error of the SpVARX model for the equation of S_{kt}^{n}. The estimation of the SpVARX model coefficient can be done using the Maximum Likelihood Estimation method, which can be seen at Tsay (2014), Tsay (2010), and Bickel & Doksum (2013).

2.2. TSpVARX

The TSpVARX model is formed by dividing data into several parts according to the selected value (υ) of threshold variable $\left(S_{k,t-d}^{u}\right)$, where d is the delay value. Each rule will have a SpVARX model with different model coefficients. The coefficient estimation of SpVARX in each regime can be resulted using the Maximum Likelihood Estimation (MLE) method. A TSpVARX model with spatial lag 1, autoregressive lag length p, exogenous variable lag length q, and delay value d can be written as TSpVARX (1,p,d,q). The equation of TSpVARX (1,p,d,q) with two regimes for K endogenous variables and N areas is as follows:

- The first regime model of TSpVARX when $S_{k,t-d}^{u} \le \upsilon$

$$\begin{aligned} S_{1,t}^{1} &= \left(\tau_{10}^{1}\right)^{(1)} + \left(\delta_{11}^{1,(1)}\right)^{(1)} R_{1,t-1}^{1} + \cdots + \left(\delta_{11}^{1,(q)}\right)^{(1)} R_{1,t-q}^{1} + \cdots + \left(\delta_{1m}^{1,(i)}\right)^{(1)} R_{m,t-1}^{1} + \cdots + \left(\delta_{1M}^{1,(1)}\right)^{(1)} R_{M,t-1}^{1} + \cdots + \left(\delta_{1M}^{1,(q)}\right)^{(1)} R_{M,t-q}^{1} + \left(\eta_{11}^{1,(1,0)}\right)^{(1)} S_{1,t-1}^{1} \\ &+ \left(\eta_{11}^{1,(1,1)}\right)^{(1)} \left(w_{11(1,2)} S_{1,t-1}^{2} + \cdots + w_{11(1,u)} S_{1,t-1}^{u} + \cdots + w_{11(1,N)} S_{1,t-1}^{N}\right) + \cdots + \left(\eta_{1r}^{1,(j,0)}\right)^{(1)} S_{r,t-j}^{1} + \left(\eta_{1r}^{1,(j,1)}\right)^{(1)} \left(w_{1r(1,2)} S_{r,t-j}^{2} + \cdots + w_{1r(1,u)} S_{r,t-j}^{u} + \cdots + w_{1r(1,N)} S_{r,t-j}^{N}\right) \\ &+ \cdots + \left(\eta_{1K}^{1,(p,0)}\right)^{(1)} S_{K,t-p}^{1} + \left(\eta_{1K}^{1,(p,1)}\right)^{(1)} \left(w_{1K(1,2)} S_{K,t-p}^{2} + \cdots + w_{1K(1,u)} S_{K,t-p}^{u} + \cdots + w_{1K(1,N)} S_{K,t-p}^{N}\right) + \varepsilon_{1t}^{1}, \end{aligned} \tag{2a}$$

$$\vdots$$

$$\begin{aligned} S_{kt}^{n} &= \left(\tau_{k0}^{n}\right)^{(1)} + \left(\delta_{k1}^{n,(1)}\right)^{(1)} R_{1,t-1}^{n} + \cdots + \left(\delta_{k1}^{n,(q)}\right)^{(1)} R_{1,t-q}^{n} + \cdots + \left(\delta_{km}^{n,(i)}\right)^{(1)} R_{m,t-g}^{n} + \cdots + \left(\delta_{kM}^{n,(1)}\right)^{(1)} R_{M,t-1}^{n} + \cdots + \left(\delta_{kM}^{n,(q)}\right)^{(1)} R_{M,t-q}^{n} + \left(\eta_{k1}^{n,(1,0)}\right)^{(1)} S_{1,t-1}^{n} \\ &+ \left(\eta_{k1}^{n,(1,1)}\right)^{(1)} \left(w_{k1(n,1)} S_{1,t-1}^{1} + \cdots + w_{k1(n,n-1)} S_{1,t-1}^{n-1} + w_{k1(n,n+1)} S_{1,t-1}^{n+1} + \cdots + w_{k1(n,N)} S_{1,t-1}^{N}\right) + \cdots + \left(\eta_{kr}^{n,(j,0)}\right)^{(1)} S_{r,t-j}^{n} + \left(\eta_{kr}^{n,(j,1)}\right)^{(1)} \left(w_{kr(n,1)} S_{1,t-j}^{1} + \cdots + w_{kr(n,n-1)} S_{r,t-1}^{n-1} + w_{kr(n,n+1)} S_{r,t-1}^{n+1} + \cdots + w_{kr(n,N)} S_{r,t-j}^{N}\right) \\ &+ \cdots + \left(\eta_{kK}^{n,(p,0)}\right)^{(1)} S_{K,t-p}^{n} + \left(\eta_{kK}^{n,(p,1)}\right)^{(1)} (w_{kK(n,1)} S_{K,t-p}^{n} + \cdots + w_{kK(n,n-1)} S_{K,t-1}^{n-1} + w_{kK(n,n+1)} S_{K,t-1}^{n+1} + \cdots + w_{1K(n,N)} S_{K,t-p}^{N}) + \varepsilon_{kt}^{n}, \end{aligned} \tag{2b}$$

$$\vdots$$

$$\begin{aligned} S_{Kt}^{N} &= \left(\tau_{K0}^{N}\right)^{(1)} + \left(\delta_{K1}^{N,(1)}\right)^{(1)} R_{1t-1}^{N} + \cdots + \left(\delta_{K1}^{N,(q)}\right)^{(1)} R_{1,t-q}^{N} + \cdots + \left(\delta_{Km}^{N,(i)}\right)^{(1)} R_{m,t-g}^{N} + \cdots + \left(\delta_{KM}^{N,(1)}\right)^{(1)} R_{M,t-1}^{N} + \cdots + \left(\delta_{KM}^{N,(q)}\right)^{(1)} R_{M,t-q}^{N} + \left(\eta_{K1}^{N,(1,0)}\right)^{(1)} S_{1,t-1}^{N} \\ &+ \left(\eta_{K1}^{N,(1,1)}\right)^{(1)} \left(w_{K1(N,1)} S_{1,t-1}^{1} + \cdots + w_{K1(N,u)} S_{1,t-1}^{u} + \cdots + w_{K1(N,N-1)} S_{1,t-1}^{N-1}\right) + \cdots + \left(\eta_{Kr}^{N,(j,0)}\right)^{(1)} S_{r,t-j}^{N} + \left(\eta_{Kr}^{N,(j,1)}\right)^{(1)} \left(w_{Kr(N,1)} S_{r,t-j}^{1} + \cdots + w_{Kr(N,u)} S_{r,t-j}^{u} + \cdots + w_{Kr(N,N-1)} S_{r,t-j}^{N-1}\right) \\ &+ \cdots + \left(\eta_{KK}^{N,(p,0)}\right)^{(1)} S_{K,t-p}^{N} + \left(\eta_{KK}^{N,(p,1)}\right)^{(1)} (w_{KK(N,1)} S_{K,t-p}^{1} + \cdots + w_{KK(N,u)} S_{K,t-p}^{u} + \cdots + w_{KK(N,N-1)} S_{K,t-p}^{N-1}) + \varepsilon_{Kt}^{N}, \end{aligned} \tag{2c}$$

- The second regime model of TSpVARX when $S_{k,t-d}^{u} > \upsilon$

$$\begin{aligned} S_{1,t}^{1} &= \left(\tau_{10}^{1}\right)^{(2)} + \left(\delta_{11}^{1,(1)}\right)^{(2)} R_{1,t-1}^{1} + \cdots + \left(\delta_{11}^{1,(q)}\right)^{(2)} R_{1,t-q}^{1} + \cdots + \left(\delta_{1m}^{1,(i)}\right)^{(2)} R_{m,t-1}^{1} + \cdots + \left(\delta_{1M}^{1,(1)}\right)^{(2)} R_{M,t-1}^{1} + \cdots + \left(\delta_{1M}^{1,(q)}\right)^{(2)} R_{M,t-q}^{1} + \left(\eta_{11}^{1,(1,0)}\right)^{(2)} S_{1,t-1}^{1} \\ &+ \left(\eta_{11}^{1,(1,1)}\right)^{(2)} \left(w_{11(1,2)} S_{1,t-1}^{2} + \cdots + w_{11(1,u)} S_{1,t-1}^{u} + \cdots + w_{11(1,N)} S_{1,t-1}^{N}\right) + \cdots + \left(\eta_{1r}^{1,(j,0)}\right)^{(2)} S_{r,t-j}^{1} + \left(\eta_{1r}^{1,(j,1)}\right)^{(2)} \left(w_{1r(1,2)} S_{r,t-j}^{2} + \cdots + w_{1r(1,u)} S_{r,t-j}^{u} + \cdots + w_{1r(1,N)} S_{r,t-j}^{N}\right) \\ &+ \cdots + \left(\eta_{1K}^{1,(p,0)}\right)^{(2)} S_{K,t-p}^{1} + \left(\eta_{1K}^{1,(p,1)}\right)^{(2)} \left(w_{1K(1,2)} S_{K,t-p}^{2} + \cdots + w_{1K(1,u)} S_{K,t-p}^{u} + \cdots + w_{1K(1,N)} S_{K,t-p}^{N}\right) + \varepsilon_{1t}^{1}, \end{aligned} \tag{2d}$$

$$\vdots$$

$$\begin{aligned} S_{kt}^{n} &= \left(\tau_{k0}^{n}\right)^{(2)} + \left(\delta_{k1}^{n,(1)}\right)^{(2)} R_{1,t-1}^{n} + \cdots + \left(\delta_{k1}^{n,(q)}\right)^{(2)} R_{1,t-q}^{n} + \cdots + \left(\delta_{km}^{n,(i)}\right)^{(2)} R_{m,t-g}^{n} + \cdots + \left(\delta_{kM}^{n,(1)}\right)^{(2)} R_{M,t-1}^{n} + \cdots + \left(\delta_{kM}^{n,(q)}\right)^{(2)} R_{M,t-q}^{n} + \left(\eta_{k1}^{n,(1,0)}\right)^{(2)} S_{1,t-1}^{n} \\ &+ \left(\eta_{k1}^{n,(1,1)}\right)^{(2)} \left(w_{k1(n,1)} S_{1,t-1}^{1} + \cdots + w_{k1(n,n-1)} S_{1,t-1}^{n-1} + w_{k1(n,n+1)} S_{1,t-1}^{n+1} + \cdots + w_{k1(n,N)} S_{1,t-1}^{N}\right) + \cdots + \left(\eta_{kr}^{n,(j,0)}\right)^{(2)} S_{r,t-j}^{n} + \left(\eta_{kr}^{n,(j,1)}\right)^{(2)} \left(w_{kr(n,1)} S_{1,t-j}^{1} + \cdots + w_{kr(n,n-1)} S_{r,t-1}^{n-1} + w_{kr(n,n+1)} S_{r,t-1}^{n+1} + \cdots + w_{kr(n,N)} S_{r,t-j}^{N}\right) \\ &+ \cdots + \left(\eta_{kK}^{n,(p,0)}\right)^{(2)} S_{K,t-p}^{n} + \left(\eta_{kK}^{n,(p,1)}\right)^{(2)} (w_{kK(n,1)} S_{K,t-p}^{n} + \cdots + w_{kK(n,n-1)} S_{K,t-1}^{n-1} + w_{kK(n,n+1)} S_{K,t-1}^{n+1} + \cdots + w_{1K(n,N)} S_{K,t-p}^{N}) + \varepsilon_{kt}^{n}, \end{aligned} \tag{2e}$$

$$\vdots$$

$$\begin{aligned} S_{Kt}^{N} &= \left(\tau_{K0}^{N}\right)^{(2)} + \left(\delta_{K1}^{N,(1)}\right)^{(2)} R_{1t-1}^{N} + \cdots + \left(\delta_{K1}^{N,(q)}\right)^{(2)} R_{1,t-q}^{N} + \cdots + \left(\delta_{Km}^{N,(i)}\right)^{(2)} R_{m,t-g}^{N} + \cdots + \left(\delta_{KM}^{N,(1)}\right)^{(2)} R_{M,t-1}^{N} + \cdots + \left(\delta_{KM}^{N,(q)}\right)^{(2)} R_{M,t-q}^{N} + \left(\eta_{K1}^{N,(1,0)}\right)^{(2)} S_{1,t-1}^{N} \\ &+ \left(\eta_{K1}^{N,(1,1)}\right)^{(2)} \left(w_{K1(N,1)} S_{1,t-1}^{1} + \cdots + w_{K1(N,u)} S_{1,t-1}^{u} + \cdots + w_{K1(N,N-1)} S_{1,t-1}^{N-1}\right) + \cdots + \left(\eta_{Kr}^{N,(j,0)}\right)^{(2)} S_{r,t-j}^{N} + \left(\eta_{Kr}^{N,(j,1)}\right)^{(2)} \left(w_{Kr(N,1)} S_{r,t-j}^{1} + \cdots + w_{Kr(N,u)} S_{r,t-j}^{u} + \cdots + w_{Kr(N,N-1)} S_{r,t-j}^{N-1}\right) \\ &+ \cdots + \left(\eta_{KK}^{N,(p,0)}\right)^{(2)} S_{K,t-p}^{N} + \left(\eta_{KK}^{N,(p,1)}\right)^{(2)} (w_{KK(N,1)} S_{K,t-p}^{1} + \cdots + w_{KK(N,u)} S_{K,t-p}^{u} + \cdots + w_{KK(N,N-1)} S_{K,t-p}^{N-1}) + \varepsilon_{Kt}^{N}, \end{aligned} \tag{2f}$$

where:

i is the notation to indicate regime (i=1 and 2),

$\left(\tau_{k0}^{n}\right)^{(i)}$ is an interception of the equation with endogenous variable S_{kt}^{n} at regime i,

$\left(\delta_{kM}^{n,(q)}\right)^{(i)}$ is the coefficient of the metric exogenous variable $R_{m,t-q}^{n}$ of the equation with endogenous variable S_{kt}^{n} at regime i,

$\left(\eta_{kr}^{n,(j,0)}\right)^{(i)}$ is the coefficient of the endogenous lag variable $S_{r,t-j}^{n}$ of the equation with endogenous variable S_{kt}^{n} at regime i,

$\left(\eta_{kr}^{n,(j,1)}\right)^{(i)}$ is the space-time coefficient for $w_{kr(n,1)} S_{1,t-j}^{1} + \cdots + w_{kr(n,n-1)} S_{r,t-1}^{n-1} + w_{kr(n,n+1)} S_{r,t-1}^{n+1} + \cdots + w_{kr(n,N)} S_{r,t-j}^{N}$ of the equation with endogenous variable S_{kt}^{n} at regime i.

There are some stages to determine the estimator of threshold $(\hat{\upsilon})$ and delay values $(\hat{d})$.

1. Determine the lag of the endogenous variable that will be used as a threshold variable $\left(S_{k,t-d}^{u}\right)$.
2. Determine the maximum value d is p.
3. Determine the 10-th percentile of $S_{k,t-d}^{u}$ as the lower limit of the candidate threshold value (υ_{dL}) and the 90-th percentile of $S_{k,t-d}^{u}$ as the upper limit of the candidate threshold value (υ_{dU}) so that $\upsilon_{dL} \le \upsilon \le \upsilon_{dU}$.

4. Specify the data of the 1st regime when $S^{u}_{k,t-d} \le \upsilon$ and the 2nd regime when $S^{u}_{k,t-d} > \upsilon$ for each possibility of υ.
5. Estimate the coefficients of TSpVARX in the 1st and 2nd regimes using the MLE method.
6. Calculate the log-likelihood function values $l\left(\hat{\boldsymbol{\eta}}^{(1)}_{MLE,SpVARX}(d,\upsilon)\,|\,\boldsymbol{\Omega}^{(1)}\right)$ on the 1st and 2nd regimes for all possible data divisions, where $\boldsymbol{\Omega}^{(i)}$ is the matrix of error covariance in the regime *i*. How to determine the log-likelihood function can be seen in Bickel & Doksum (2013).
7. Calculate the total log-likelihood for all possibilities *d* and *v* with the following formula:
$l\left(\hat{\boldsymbol{\eta}}_{MLE,TSpVARX}(d,\upsilon)\,|\,\boldsymbol{\Omega}\right) = l\left(\hat{\boldsymbol{\eta}}^{(1)}_{MLE,SpVARX}(d,\upsilon)\,|\,\boldsymbol{\Omega}^{(1)}\right) + l\left(\hat{\boldsymbol{\eta}}^{(2)}_{MLE,SpVARX}(d,\upsilon)\,|\,\boldsymbol{\Omega}^{(2)}\right)$.
8. The delay estimator $(\hat{d})$ and threshold value $(\hat{\upsilon})$ are obtained from *d* and *v* which can maximize the value of $l\left(\hat{\boldsymbol{\eta}}_{MLE,TSpVARX}(d,\upsilon)\,|\,\boldsymbol{\Omega}\right)$.

2.3 Simulation method

This study evaluates the forecasting performance of TSpVARX and compares it with SpVARX. We simulate the TSpVARX and SpVARX models for three areas, two endogenous at each area and one exogenous variable at each area. Thus, there are six endogenous variables used in this simulation. Our simulation has 12 scenario combinations based on the following:

1. two scenarios of the conditions of model error distribution used to generate data, namely multivariate normal and multivariate T distributed errors,
2. two scenarios of the error correlation between equations used to generate data, namely:
    a. the correlation error between equations of 0.1 (covariance error between endogenous variables 0.01),
    b. the correlation error between equations of 0.9 (covariance error between endogenous variables 0.09),
3. three scenarios of sample size conditions are used, namely 120, 240, and 360 samples.

The simulation is done by generating data and then creating SpVARX and TSpVARX models 100 times for each combination of scenarios. There are several steps performed in our simulation.

1. We specify the spatial weight used is uniform with the following shape:

$$\boldsymbol{W} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

2. We specify the true parameter by using the following TSpVARX equation:

The 1st regime (when $S^{1}_{1,t-1} \le \mathbf{0}$ **)**

$$S^{1}_{1t} = 0.4R_{1,t-1} + 0.25S^{1}_{1,t-1} + 0.3\left(0.5S^{2}_{1,t-1} + 0.5S^{3}_{1,t-1}\right) + 0.2S^{1}_{2,t-1} + 0.5\left(0.5S^{2}_{2,t-1} + 0.5S^{3}_{2,t-1}\right) + e^{1}_{1}, \tag{3a}$$

$$S^{2}_{1} = 0.55R_{1,t-1} + 0.2S^{2}_{1,t-1} + 0.12\left(0.5S^{1}_{1,t-1} + 0.5S^{3}_{1,t-1}\right) + 0.3S^{2}_{2,t-1} + 0.25\left(0.5S^{1}_{2,t-1} + 0.5S^{3}_{2,t-1}\right) + e^{2}_{1}, \tag{3b}$$

$$S^{3}_{1} = 0.2R_{1,t-1} + 0.3S^{3}_{1,t-1} + 0.3\left(0.5S^{1}_{1,t-1} + 0.5S^{2}_{1,t-1}\right) + 0.35S^{3}_{2,t-1} + 0.4\left(0.5S^{1}_{2,t-1} + 0.5S^{2}_{2,t-1}\right) + e^{3}_{1}, \tag{3c}$$

$$S^{1}_{2} = 0.5R_{1,t-1} + 0.15S^{1}_{1,t-1} + 0.05\left(0.5S^{2}_{1,t-1} + 0.5S^{3}_{1,t-1}\right) + 0.125S^{1}_{2,t-1} + 0.2\left(0.5S^{2}_{2,t-1} + 0.5S^{3}_{2,t-1}\right) + e^{1}_{2}, \tag{3d}$$

$$S^{2}_{2} = 0.3R_{1,t-1} + 0.25S^{2}_{1,t-1} + 0.3\left(0.5S^{1}_{1,t-1} + 0.5S^{3}_{1,t-1}\right) + 0.3S^{2}_{2,t-1} + 0.05\left(0.5S^{1}_{2,t-1} + 0.5S^{3}_{2,t-1}\right) + e^{2}_{2}, \tag{3e}$$

$$S^{3}_{2,t} = 0.4R_{1,t-1} + 0.2S^{3}_{1,t-1} + 0.3\left(0.5S^{1}_{1,t-1} + 0.5S^{2}_{1,t-1}\right) + 0.25S^{3}_{2,t-1} + 0.02\left(0.5S^{1}_{2,t-1} + 0.5S^{2}_{2,t-1}\right) + e^{3}_{2}. \tag{3f}$$

The 2nd regime (when $S^{1}_{1,t-1} > \mathbf{0}$ **)**

$$S^{1}_{1t} = 0.2R_{1,t-1} + 0.2S^{1}_{1,t-1} + 0.24\left(0.5S^{2}_{1,t-1} + 0.5S^{3}_{1,t-1}\right) + 0.16S^{1}_{2,t-1} + 0.4\left(0.5S^{2}_{2,t-1} + 0.5S^{3}_{2,t-1}\right) + e^{1}_{1}, \tag{3g}$$

$$S^{2}_{1t} = 0.25R_{1,t-1} + 0.16S^{2}_{1,t-1} + 0.096\left(0.5S^{1}_{1,t-1} + 0.5S^{3}_{1,t-1}\right) + 0.24S^{2}_{2,t-1} + 0.2\left(0.5S^{1}_{2,t-1} + 0.5S^{3}_{2,t-1}\right) + e^{2}_{1}, \tag{3h}$$

$$S^{3}_{1} = 0.1R_{1,t-1} + 0.24S^{3}_{1,t-1} + 0.24\left(0.5S^{1}_{1,t-1} + 0.5S^{2}_{1,t-1}\right) + 0.28S^{3}_{2,t-1} + 0.32\left(0.5S^{1}_{2,t-1} + 0.5S^{2}_{2,t-1}\right) + e^{3}_{1}, \tag{3i}$$

$$S^{1}_{2} = 0.25R_{1,t-1} + 0.12S^{1}_{1,t-1} + 0.04\left(0.5S^{2}_{1,t-1} + 0.5S^{3}_{1,t-1}\right) + 0.17S^{1}_{2,t-1} + 0.016\left(0.5S^{2}_{2,t-1} + 0.5S^{3}_{2,t-1}\right) + e^{1}_{2}, \tag{3j}$$

$$S^{2}_{2} = 0.1R_{1,t-1} + 0.2S^{2}_{1,t-1} + 0.24\left(0.5S^{1}_{1,t-1} + 0.5S^{3}_{1,t-1}\right) + 0.24S^{2}_{2,t-1} + 0.04\left(0.5S^{1}_{2,t-1} + 0.5S^{3}_{2,t-1}\right) + e^{2}_{2}, \tag{3k}$$

$$S^{3}_{2t} = 0.15R_{1,t-1} + 0.16S^{3}_{1,t-1} + 0.24\left(0.5S^{1}_{1,t-1} + 0.5S^{2}_{1,t-1}\right) + 0.2S^{3}_{2,t-1} + 0.016\left(0.5S^{1}_{2,t-1} + 0.5S^{2}_{2,t-1}\right) + e^{3}_{2}. \tag{3l}$$

3. We generate a normally distributed error $(\in_t)$ with an average of zero and a variance of 0.2 $(\in_t \sim N(0, 0.2))$ to form simulation data of exogenous variables. Simulation data of exogenous variables is developed by following an autoregressive order one or AR (1). The model coefficients are specified such that $|\phi| < 1$. The following equation forms the data of an exogenous variable:

$$R_t = 0.5R_{t-1} + \in_t,$$

4. We generate multivariate normally distributed errors with mean vectors $\boldsymbol{\mu}$ and covariance matrices $\Sigma \otimes \boldsymbol{I}_T$ $(\boldsymbol{\varepsilon} \sim N(\boldsymbol{\mu}, (\Sigma \otimes \boldsymbol{I}_T)))$. We also create errors distributed multivariate t with a degree of freedom (v) equal to 2 $(\boldsymbol{\varepsilon} \sim t_2(\boldsymbol{\mu}, (\Sigma \otimes \boldsymbol{I}_T)))$,

   where:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1^1 \\ \varepsilon_1^2 \\ \varepsilon_1^3 \\ \varepsilon_2^1 \\ \varepsilon_2^2 \\ \varepsilon_2^3 \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \text{var}(\varepsilon_1^1) & \text{cov}(\varepsilon_1^1,\varepsilon_1^2) & \text{cov}(\varepsilon_1^1,\varepsilon_1^3) & \text{cov}(\varepsilon_1^1,\varepsilon_2^1) & \text{cov}(\varepsilon_1^1,\varepsilon_2^2) & \text{cov}(\varepsilon_1^1,\varepsilon_2^3) \\ \text{cov}(\varepsilon_1^2,\varepsilon_1^1) & \text{var}(\varepsilon_1^2) & \text{cov}(\varepsilon_1^2,\varepsilon_1^3) & \text{cov}(\varepsilon_1^2,\varepsilon_2^1) & \text{cov}(\varepsilon_1^2,\varepsilon_2^2) & \text{cov}(\varepsilon_1^2,\varepsilon_2^3) \\ \text{cov}(\varepsilon_1^3,\varepsilon_1^1) & \text{cov}(\varepsilon_1^3,\varepsilon_1^2) & \text{var}(\varepsilon_1^3) & \text{cov}(\varepsilon_1^3,\varepsilon_2^1) & \text{cov}(\varepsilon_1^3,\varepsilon_2^2) & \text{cov}(\varepsilon_1^3,\varepsilon_2^3) \\ \text{cov}(\varepsilon_2^1,\varepsilon_1^1) & \text{cov}(\varepsilon_2^1,\varepsilon_1^2) & \text{cov}(\varepsilon_2^1,\varepsilon_1^3) & \text{var}(\varepsilon_2^1) & \text{cov}(\varepsilon_2^1,\varepsilon_2^2) & \text{cov}(\varepsilon_2^1,\varepsilon_2^3) \\ \text{cov}(\varepsilon_2^2,\varepsilon_1^1) & \text{cov}(\varepsilon_2^2,\varepsilon_1^2) & \text{cov}(\varepsilon_2^2,\varepsilon_1^3) & \text{cov}(\varepsilon_2^2,\varepsilon_2^1) & \text{var}(\varepsilon_2^2) & \text{cov}(\varepsilon_2^2,\varepsilon_2^3) \\ \text{cov}(\varepsilon_2^3,\varepsilon_1^1) & \text{cov}(\varepsilon_2^3,\varepsilon_1^2) & \text{cov}(\varepsilon_2^3,\varepsilon_1^3) & \text{cov}(\varepsilon_2^3,\varepsilon_2^1) & \text{cov}(\varepsilon_2^3,\varepsilon_2^2) & \text{var}(\varepsilon_2^3) \end{bmatrix}.$$

   The diagonal value (variance of error) is determined by the same magnitude of 0.1. Off-diagonal components of $\boldsymbol{\Sigma}$ (covariance of error) are specified with the same value. In this simulation, the off-diagonal part of $\boldsymbol{\Sigma}$ is determined such that the resulting correlation error between the equation is 0,1 or 0,9. Because error covariance is obtained by multiplying error correlation with error variance, two scenarios of the magnitude of error covariance will be simulated in this study, namely 0,01 and 0,09.

5. By using $\varepsilon_1^1, \varepsilon_1^2, \varepsilon_1^3, \varepsilon_2^1, \varepsilon_2^2, \varepsilon_2^3$ as the result of the fourth step, we generate simulation data using Eq. (3a) until Eq. (3l).
6. We use ninety percent of the simulation data as training data, while ten percent of the data as testing data.
7. We use MLE to estimate the model coefficients of SpVARX and TSpVARX from training data.
8. We perform forecasting for endogenous variables throughout the testing data period.
9. We calculate the RMSE of SpVARX and TSpVARX models from testing data. The RMSE formula can be seen in Wei (2006).
10. We do the simulation for 100 times replication, and then a hypothesis test is carried out to determine whether TSpVARX could improve forecasting performance. If the difference data between TSpVARX's RMSE and SpVARX's RMSE are normally distributed, then the average hypothesis test for paired observations is used (Illowsky et al., 2013). The null (Ho) and alternative hypotheses (H_1) used are:
    $\text{Ho}: \mu(\text{RMSE TSpVARX - RMSE SpVARX}) \geq 0$ (The average of the difference between TSpVARX's RMSE and SpVARX's RMSE is equal to or greater than zero)
    $H_1: \mu(\text{RMSE TSpVARX - RMSE SpVARX}) < 0$ (The average of the difference between TSpVARX's RMSE and SpVARX's RMSE is less than zero).
    Meanwhile, Gibbons & Chakraborti (2011) stated that if the difference data between TSpVARX's RMSE and SpVARX's RMSE is not normally distributed, we can use the Wilcoxon Signed-Rank (WSR) test for testing the data difference. The hypotheses used are:
    $\text{Ho}: \text{Median}(\text{RMSE TSpVARX - RMSE SpVARX}) \geq 0$ (The median of the difference between TSpVARX's RMSE and SpVARX's RMSE is equal to or greater than zero)
    $H_1: \text{Median}(\text{RMSE TSpVARX - RMSE SpVARX}) < 0$ (The median of the difference between TSpVARX's RMSE and SpVARX's RMSE is less than zero).
    We reject Ho if the probability value obtained is smaller than the significance level used. Ho rejection indicates that TSpVARX's forecasting performance is better than SpVARX. Because six endogenous variables and 12 scenarios are used, there are 72 hypothesis tests for testing 72 data groups of the difference between RMSE TSpVARX and SpVARX.
11. After generating data up to testing the RMSE difference hypothesis (as in steps 1 through 9) for six endogenous variables of 12 scenarios, we calculate the Ho rejection percentage of 72 hypothesis test results.

We provide a flowchart in Figure 1 to make it easier to understand the simulation steps, from generating data to testing hypotheses on RMSE (step 1 until step 9) in our research. The flowchart in Fig. 1 describes only the simulation steps for the scenario: normally distributed error, a covariance error of 0.01, and a sample size of 120.

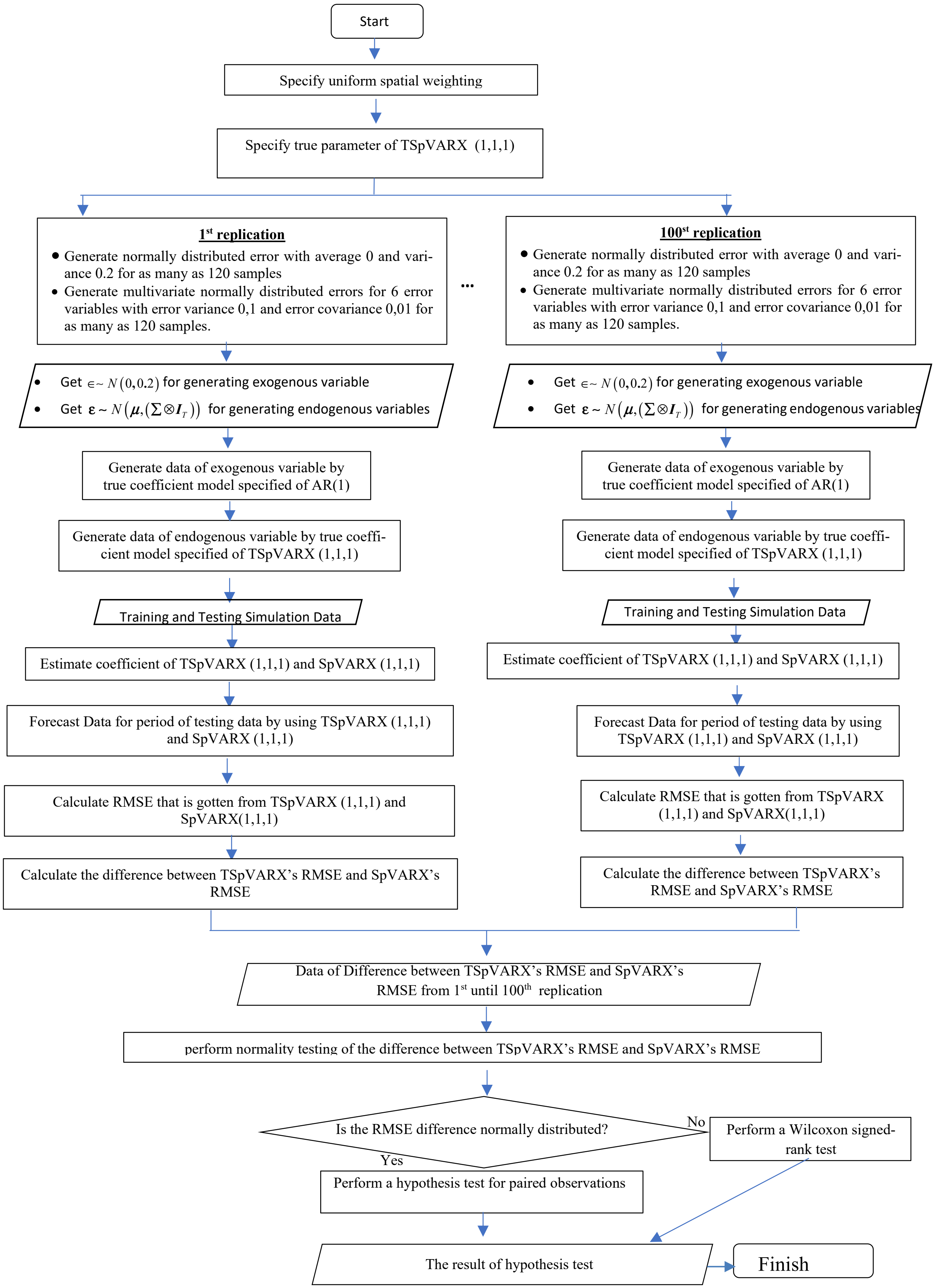

Fig. 1. Flow Chart of the simulation steps from generating data to testing hypotheses on RMSE for the scenario: multivariate normally distributed error, a covariance error of 0.01, and a sample size of 120.

A flowchart regarding the calculation of the percentage of Ho reject decisions from the results of the RMSE difference hypothesis test is as shown in Fig. 2.

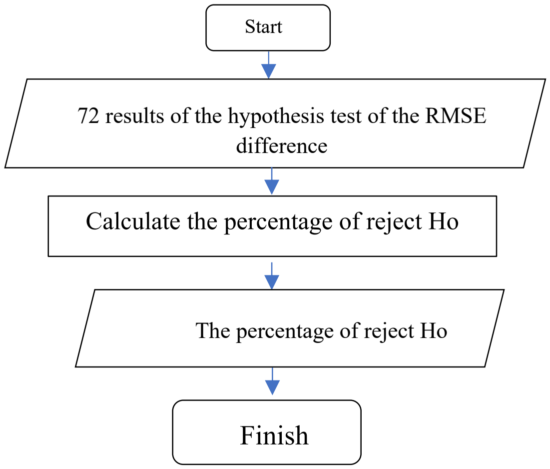

Fig. 2. Flow Chart of the calculation of reject Ho percentage of the hypothesis tests of the RMSE difference

3. Result

Based on subsection 2.3, there are 72 hypothesis tests of the difference between TSpVARX's RMSE and SpVARX's RMSE. This test helps compare TSpVARX and SpVARX forecasting performance. Before we perform a hypothesis test to compare the forecasting performance of TSpVARX and SpVARX, we first test the distribution of the difference between TSpVARX's RMSE and SpVARX's RMSE for the 72 data groups. If the difference is normally distributed, then the hypothesis test used is the average hypothesis test for paired observations. The hypothesis test uses the Wilcoxon signed-rank test if the difference is not normally distributed.

We use the Shapiro-Wilk method for testing the normality data of all 72 data groups. The result can be seen in Table 1. In normality hypothesis testing, normal data condition becomes the Ho, and the alternative hypothesis (H_1) is that data does not follow a normal condition or distribution. If the probability value obtained is more than the significance level, the data has normal distribution. The data does not follow normal distribution if the probability value is less than the significance level. In Table 1, we can see that the results of 72 normality hypothesis tests get a probability value of < 0.010. If we use the five percent significance level, then the normality test rejects the null hypothesis for all 72 normality hypothesis tests. It means the difference between TSpVARX's RMSE and SpVARX's RMSE for all 72 groups is not normally distributed.

Table 1
The Normality Test of The Difference between TSpVARX's RMSE and SpVARX's RMSE for All 72 Data Groups (Combination of 12 Scenarios and 6 Endogenous Variables)

No	Sample Size	Error Distribution	Error correlation between equations	Endogenous Variable	*The Probability Value of Shapiro Wilk*	Conclusion
1	120	Multivariate Normal	0,1	S_{1t}^{1}	<0,010	not normal
2	120	Multivariate Normal	0,1	S_{1t}^{2}	<0,010	not normal
3	120	Multivariate Normal	0,1	S_{1t}^{3}	<0,010	not normal
4	120	Multivariate Normal	0,1	S_{2t}^{1}	<0,010	not normal
5	120	Multivariate Normal	0,1	S_{2t}^{2}	<0,010	not normal
6	120	Multivariate Normal	0,1	S_{2t}^{3}	<0,010	not normal
7	120	Multivariate Normal	0,9	S_{1t}^{1}	<0,010	not normal
8	120	Multivariate Normal	0,9	S_{1t}^{2}	<0,010	not normal
9	120	Multivariate Normal	0,9	S_{1t}^{3}	<0,010	not normal
10	120	Multivariate Normal	0,9	S_{2t}^{1}	<0,010	not normal
11	120	Multivariate Normal	0,9	S_{2t}^{2}	<0,010	not normal
12	120	Multivariate Normal	0,9	S_{2t}^{3}	<0,010	not normal
13	120	Multivariate T	0,1	S_{1t}^{1}	<0,010	not normal

14	120	Multivariate T	0,1	S_{1t}^{2}	<0,010	not normal
15	120	Multivariate T	0,1	S_{1t}^{3}	<0,010	not normal
16	120	Multivariate T	0,1	S_{2t}^{1}	<0,010	not normal
17	120	Multivariate T	0,1	S_{2t}^{2}	<0,010	not normal
18	120	Multivariate T	0,1	S_{2t}^{3}	<0,010	not normal
19	120	Multivariate T	0,9	S_{1t}^{1}	<0,010	not normal
20	120	Multivariate T	0,9	S_{1t}^{2}	<0,010	not normal
21	120	Multivariate T	0,9	S_{1t}^{3}	<0,010	not normal
22	120	Multivariate T	0,9	S_{2t}^{1}	<0,010	not normal
23	120	Multivariate T	0,9	S_{2t}^{2}	<0,010	not normal
24	120	Multivariate T	0,9	S_{2t}^{3}	<0,010	not normal
25	240	Multivariate Normal	0,1	S_{1t}^{1}	<0,010	not normal
26	240	Multivariate Normal	0,1	S_{1t}^{2}	<0,010	not normal
27	240	Multivariate Normal	0,1	S_{1t}^{3}	<0,010	not normal
28	240	Multivariate Normal	0,1	S_{2t}^{1}	<0,010	not normal
29	240	Multivariate Normal	0,1	S_{2t}^{2}	<0,010	not normal
30	240	Multivariate Normal	0,1	S_{2t}^{3}	<0,010	not normal
31	240	Multivariate Normal	0,9	S_{1t}^{1}	<0,010	not normal
32	240	Multivariate Normal	0,9	S_{1t}^{2}	<0,010	not normal
33	240	Multivariate Normal	0,9	S_{1t}^{3}	<0,010	not normal
34	240	Multivariate Normal	0,9	S_{2t}^{1}	<0,010	not normal
35	240	Multivariate Normal	0,9	S_{2t}^{2}	<0,010	not normal
36	240	Multivariate Normal	0,9	S_{2t}^{3}	<0,010	not normal
37	240	Multivariate T	0,1	S_{1t}^{1}	<0,010	not normal
38	240	Multivariate T	0,1	S_{1t}^{2}	<0,010	not normal
39	240	Multivariate T	0,1	S_{1t}^{3}	<0,010	not normal
40	240	Multivariate T	0,1	S_{2t}^{1}	<0,010	not normal
41	240	Multivariate T	0,1	S_{2t}^{2}	<0,010	not normal
42	240	Multivariate T	0,1	S_{2t}^{3}	<0,010	not normal
43	240	Multivariate T	0,9	S_{1t}^{1}	<0,010	not normal
44	240	Multivariate T	0,9	S_{1t}^{2}	<0,010	not normal
45	240	Multivariate T	0,9	S_{1t}^{3}	<0,010	not normal
46	240	Multivariate T	0,9	S_{2t}^{1}	<0,010	not normal
47	240	Multivariate T	0,9	S_{2t}^{2}	<0,010	not normal
48	240	Multivariate T	0,9	S_{2t}^{3}	<0,010	not normal
49	360	Multivariate Normal	0,1	S_{1t}^{1}	<0,010	not normal
50	360	Multivariate Normal	0,1	S_{1t}^{2}	<0,010	not normal
51	360	Multivariate Normal	0,1	S_{1t}^{3}	<0,010	not normal
52	360	Multivariate Normal	0,1	S_{2t}^{1}	<0,010	not normal
53	360	Multivariate Normal	0,1	S_{2t}^{2}	<0,010	not normal
54	360	Multivariate Normal	0,1	S_{2t}^{3}	<0,010	not normal
55	360	Multivariate Normal	0,9	S_{1t}^{1}	<0,010	not normal
56	360	Multivariate Normal	0,9	S_{1t}^{2}	<0,010	not normal
57	360	Multivariate Normal	0,9	S_{1t}^{3}	<0,010	not normal
58	360	Multivariate Normal	0,9	S_{2t}^{1}	<0,010	not normal
59	360	Multivariate Normal	0,9	S_{2t}^{2}	<0,010	not normal

60	360	Multivariate Normal	0,9	S_{2t}^{3}	<0,010	not normal
61	360	Multivariate T	0,1	S_{1t}^{1}	<0,010	not normal
62	360	Multivariate T	0,1	S_{1t}^{2}	<0,010	not normal
63	360	Multivariate T	0,1	S_{1t}^{3}	<0,010	not normal
64	360	Multivariate T	0,1	S_{2t}^{1}	<0,010	not normal
65	360	Multivariate T	0,1	S_{2t}^{2}	<0,010	not normal
66	360	Multivariate T	0,1	S_{2t}^{3}	<0,010	not normal
67	360	Multivariate T	0,9	S_{1t}^{1}	<0,010	not normal
68	360	Multivariate T	0,9	S_{1t}^{2}	<0,010	not normal
69	360	Multivariate T	0,9	S_{1t}^{3}	<0,010	not normal
70	360	Multivariate T	0,9	S_{2t}^{1}	<0,010	not normal
71	360	Multivariate T	0,9	S_{2t}^{2}	<0,010	not normal
72	360	Multivariate T	0,9	S_{2t}^{3}	<0,010	not normal

Because the difference between TSpVARX's RMSE and SpVARX's RMSE of all 72 data groups is not normally distributed, the evaluation TSpVARX forecasting performance compared to SpVARX is carried out using the WSR test. The hypothesis of the Wilcoxon signed-rank test in this study are:

$$\text{Ho} : \text{Median}(\text{RMSE TSpVARX - RMSE SpVARX}) \geq 0$$

$$\text{H}_1 : \text{Median}(\text{RMSE TSpVARX - RMSE SpVARX}) < 0$$

The rejection of Ho is occurred when the probability value of the Wilcoxon signed-rank test is less than the significant level (α) used. If we reject Ho, TSpVARX has better forecasting performance than SpVARX at a certain α.

The results of the WSR test for all 72 data groups (a combination of 12 scenarios and six endogenous variables) can be seen in Table 2. Meanwhile, concise results containing the percentage of Ho reject and Ho reject failure results of the WSR test are presented in Figure 3 and Figure 4. The greater the percentage of Ho rejection of the test, the better TSVARX forecasting performance compared to SpVARX.

Table 2
Wilcoxon Signed-Rank (WSR) Test of All 72 Data Groups (Combination of 12 Scenarios and 6 Endogenous Variables)

No	Sample Size	Error Distribution	Error correlation between equations	Endogenous variables	*Probability Value*	Reject Decision (Yes/ No)
1	120	Multivariate Normal	0,1	S_{1t}^{1}	0,102	No
2	120	Multivariate Normal	0,1	S_{1t}^{2}	0,008**	Yes
3	120	Multivariate Normal	0,1	S_{1t}^{3}	0,036**	Yes
4	120	Multivariate Normal	0,1	S_{2t}^{1}	0,182	No
5	120	Multivariate Normal	0,1	S_{2t}^{2}	0,005**	Yes
6	120	Multivariate Normal	0,1	S_{2t}^{3}	0,045**	Yes
7	120	Multivariate Normal	0,9	S_{1t}^{1}	0,180	No
8	120	Multivariate Normal	0,9	S_{1t}^{2}	0,062*	Yes
9	120	Multivariate Normal	0,9	S_{1t}^{3}	0,038**	Yes
10	120	Multivariate Normal	0,9	S_{2t}^{1}	0,317	No
11	120	Multivariate Normal	0,9	S_{2t}^{2}	0,031*	Yes
12	120	Multivariate Normal	0,9	S_{2t}^{3}	0,052*	Yes
13	120	Multivariate T	0,1	S_{1t}^{1}	0,028**	Yes
14	120	Multivariate T	0,1	S_{1t}^{2}	0,073*	Yes
15	120	Multivariate T	0,1	S_{1t}^{3}	0,117	No
16	120	Multivariate T	0,1	S_{2t}^{1}	0,021**	Yes
17	120	Multivariate T	0,1	S_{2t}^{2}	0,165	No
18	120	Multivariate T	0,1	S_{2t}^{3}	0,165	No

19	120	Multivariate T	0,9	S_{1t}^{1}	0,608	No
20	120	Multivariate T	0,9	S_{1t}^{2}	0,048**	Yes
21	120	Multivariate T	0,9	S_{1t}^{3}	0,096*	Yes
22	120	Multivariate T	0,9	S_{2t}^{1}	0,553	No
23	120	Multivariate T	0,9	S_{2t}^{2}	0,098*	Yes
24	120	Multivariate T	0,9	S_{2t}^{3}	0,075*	Yes
25	240	Multivariate Normal	0,1	S_{1t}^{1}	0,002*	Yes
26	240	Multivariate Normal	0,1	S_{1t}^{2}	0,006*	Yes
27	240	Multivariate Normal	0,1	S_{1t}^{3}	0,001*	Yes
28	240	Multivariate Normal	0,1	S_{2t}^{1}	0,003*	Yes
29	240	Multivariate Normal	0,1	S_{2t}^{2}	0,002*	Yes
30	240	Multivariate Normal	0,1	S_{2t}^{3}	0,000*	Yes
31	240	Multivariate Normal	0,9	S_{1t}^{1}	0,217	No
32	240	Multivariate Normal	0,9	S_{1t}^{2}	0,067*	Yes
33	240	Multivariate Normal	0,9	S_{1t}^{3}	0,116	No
34	240	Multivariate Normal	0,9	S_{2t}^{1}	0,258	No
35	240	Multivariate Normal	0,9	S_{2t}^{2}	0,024**	Yes
36	240	Multivariate Normal	0,9	S_{2t}^{3}	0,306	No
37	240	Multivariate T	0,1	S_{1t}^{1}	0,099*	Yes
38	240	Multivariate T	0,1	S_{1t}^{2}	0,001**	Yes
39	240	Multivariate T	0,1	S_{1t}^{3}	0,002**	Yes
40	240	Multivariate T	0,1	S_{2t}^{1}	0,098*	Yes
41	240	Multivariate T	0,1	S_{2t}^{2}	0,001**	Yes
42	240	Multivariate T	0,1	S_{2t}^{3}	0,000**	Yes
43	240	Multivariate T	0,9	S_{1t}^{1}	0,034**	Yes
44	240	Multivariate T	0,9	S_{1t}^{2}	0,034**	Yes
45	240	Multivariate T	0,9	S_{1t}^{3}	0,049**	Yes
46	240	Multivariate T	0,9	S_{2t}^{1}	0,002**	Yes
47	240	Multivariate T	0,9	S_{2t}^{2}	0,027**	Yes
48	240	Multivariate T	0,9	S_{2t}^{3}	0,066**	Yes
49	360	Multivariate Normal	0,1	S_{1t}^{1}	0,005**	Yes
50	360	Multivariate Normal	0,1	S_{1t}^{2}	0,037**	Yes
51	360	Multivariate Normal	0,1	S_{1t}^{3}	0,000**	Yes
52	360	Multivariate Normal	0,1	S_{2t}^{1}	0,000**	Yes
53	360	Multivariate Normal	0,1	S_{2t}^{2}	0,009**	Yes
54	360	Multivariate Normal	0,1	S_{2t}^{3}	0,000**	Yes
55	360	Multivariate Normal	0,9	S_{1t}^{1}	0,049**	Yes
56	360	Multivariate Normal	0,9	S_{1t}^{2}	0,012**	Yes
57	360	Multivariate Normal	0,9	S_{1t}^{3}	0,029**	Yes
58	360	Multivariate Normal	0,9	S_{2t}^{1}	0,067*	Yes
59	360	Multivariate Normal	0,9	S_{2t}^{2}	0,04**	Yes
60	360	Multivariate Normal	0,9	S_{2t}^{3}	0,032**	Yes
61	360	Multivariate T	0,1	S_{1t}^{1}	0,000**	Yes
62	360	Multivariate T	0,1	S_{1t}^{2}	0,001**	Yes
63	360	Multivariate T	0,1	S_{1t}^{3}	0,000**	Yes
64	360	Multivariate T	0,1	S_{2t}^{1}	0,000**	Yes

65	360	Multivariate T	0,1	S_{2t}^{2}	0,000**	Yes
66	360	Multivariate T	0,1	S_{2t}^{3}	0,001**	Yes
67	360	Multivariate T	0,9	S_{1t}^{1}	0,000**	Yes
68	360	Multivariate T	0,9	S_{1t}^{2}	0,001**	Yes
69	360	Multivariate T	0,9	S_{1t}^{3}	0,000**	Yes
70	360	Multivariate T	0,9	S_{2t}^{1}	0,000**	Yes
71	360	Multivariate T	0,9	S_{2t}^{2}	0,000**	Yes
72	360	Multivariate T	0,9	S_{2t}^{3}	0,001**	Yes

Note: ** Reject Ho with α= 5%

* Reject Ho with α= 10%

Fig. 3 shows that at a significance level (α) of 5 percent, 66.67 percent (48 out of 72 Wilcoxon signed-rank test results) reject Ho. Meanwhile, based on Figure 4, if the α used is 10 percent, 81.94 percent of the hypothesis test (59 out of 72 Wilcoxon signed-rank test results) result Ho rejection. It shows that most forecasting results using TSpVARX are better than SpVARX when there is a nonlinear relationship between endogenous variables.

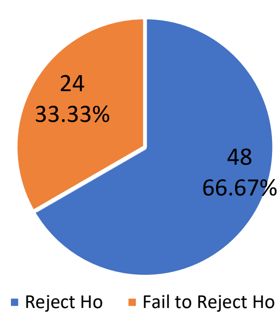

Fig. 3. Percentage of Reject Ho (Blue) and Fail to Reject Ho (Orange) on Wilcoxon Signed-Rank Test with α 5 Percent

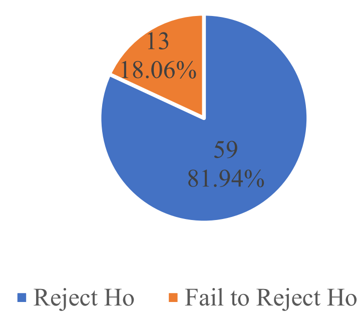

Fig. 4. Percentage of Reject Ho (Blue) and Fail to Reject Ho (Orange) on Wilcoxon Signed-Rank Test with α 10 Percent

The results of evaluating TSpVARX forecasting performance compared to SpVARX can also be seen according to the sample size, error distribution, and magnitude of error correlation between equations determined for generating simulation data. The discussion will be based on the Wilcoxon signed-rank test results using α five percent. Fig. 5 shows that the more samples used, the more Ho rejection results on the Wilcoxon signed-rank test. Only when the sample size is 120 can we see that the number of Ho reject failures exceeds Ho rejection. It indicates that in addition to the nonlinear relationship between endogenous variables, TSpVARX forecasting performance will be better than SpVARX as the sample size increases.

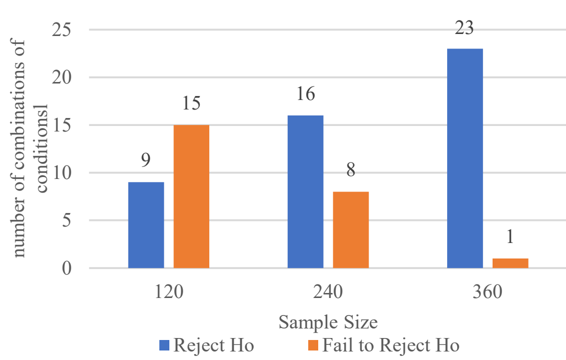

Fig. 5. The Number of Reject Ho (Blue) and Fail to Reject Ho (Orange) of Wilcoxon Signed-Rank Test Based on Sample Size

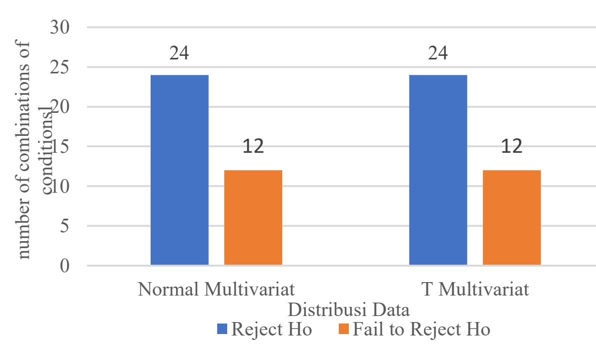

Fig. 6. The Number of Reject Ho (Blue) and Fail to Reject Ho (Orange) Results on the Wilcoxon Sign-Rank Test by The Data Distribution

TSpVARX forecasting performance compared to SpVARX can also be seen according to the data distribution. Figure 6 shows that the number of Ho rejections is more fantastic than the failed ones when multivariate normal and multivariate t-distributed.

It means that the forecasting performance of TSpVARX is better than SpVARX under any data distribution conditions. In addition, there is no difference in the number of reject Ho of the Wilcoxon signed-rank test when the data are multivariate normally distributed and multivariate t distributed. Evaluation of TSpVARX forecasting performance compared to SpVARX according to the error correlation between equations can be seen in Fig. 7 and Fig. 8. Fig. 7 shows that when the error correlation between equations is 0.1 or 0.9, the number of Ho rejections is greater than the failed ones. It means that the forecasting performance of TSpVARX is better than SpVARX under any conditions of error correlation between model equations. We can also see that the number of Ho rejections is more at the time of error correlation 0.1. Besides that, Fig. 8 shows that at each sample size used, the Ho rejection results are obtained more when the error correlation between equations is 0.1 compared to 0.9. It indicates that in addition to the nonlinear relationship between endogenous variables and the larger sample size, the TSpVARX model's forecasting performance will improve when the error correlation between equations gets smaller.

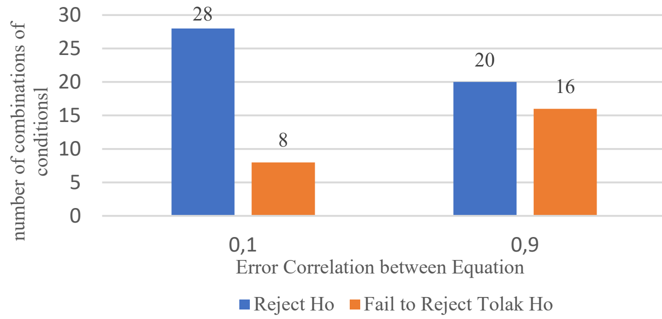

Fig. 7. The Number of Reject Ho (Blue) and Fail to Reject Ho (Orange) Results of the Wilcoxon Signed-Rank Test Based on the Error Correlation between Equation

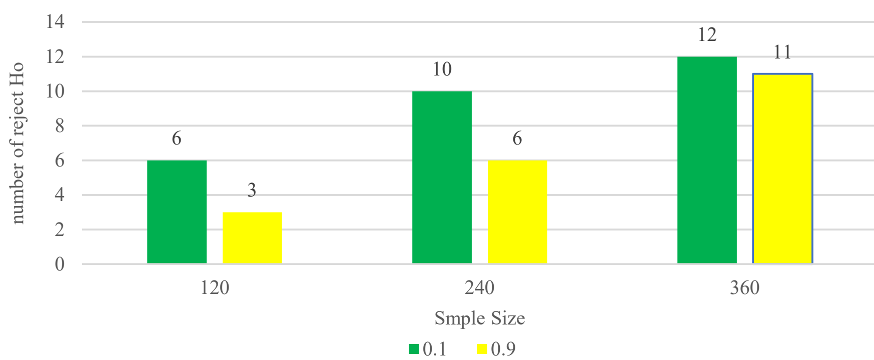

Fig. 8. The Number of Reject Ho Results on the Wilcoxon Signed-Rank Test Based on The Sample Size and Error Correlation between Equations (green for 0,1 and yellow for 0,9)

4. Conclusion

TSpVARX model can be applied to actual data, especially those derived from economic variables that have a nonlinear relationship. Based on our simulation study, we can conclude that when the nonlinearity between endogenous variables exists, the forecasting performance of the TSpVARX model is better than SpVARX. The increasing sample size can also impact the growing forecasting performance of the TSpVARX. TSpVARX forecasting performance is better than SpVARX in low or high error correlation conditions between model equations. In addition, our simulation proves that we will get better forecasting using TSpVARX with the small error correlation between equations. TSpVARX's forecasting performance is better than SpVARX's in multivariate normal or multivariate T data distribution. There is no difference in forecasting performance of TSpVARX between multivariate normally distributed and multivariate T distributed data.

Author Contributions: "Conceptualization, G.P.D.S. and S.S.; methodology, G.P.D.S and S.S.; software, G.P.D.S, and D.D.P.; validation, S.S., G.P.D.S and D.D.P.; formal analysis, G.P.D.S., S.S., and D.D.P.; investigation, G.P.D.S, and S.S.; resources, G.P.D.S.; data curation, G.P.D.S.; writing—original draft preparation, G.P.D.S, S.S., and D.D.P; writing—review and editing, G.P.D.S., and D.D.P.; visualization, G.P.D.S, and S.S.; supervision, S.S. and D.D.P.; All authors have read and agreed to the published version of the manuscript".

Conflicts of Interest

The authors declare no conflict of interest.

References

Andreas, C., Faricha, A., Ulyah, S. M., Susanti, R., Mardhiana, H., Nanda, M. A., & Firman Adi, R. (2022). Comparison study using ARIMAX and VARX in cash flow forecasting. *AIP Conference Proceedings*, *2641*(December). https://doi.org/10.1063/5.0118519

Andrés-Rosales, R., Quintana-Romero, L., de Jesús-Almonte, L., & del Río-Rama, M. de la C. (2021). Spatial spillovers of economic growth and public spending in Mexico: Evidence from a SpVAR model, 1999–2019. *Economic Analysis and Policy*, *71*, 660–673. https://doi.org/10.1016/j.eap.2021.07.004.

Balke, N. S., & Fomby, T. (1997). Threshold Cointegration. *International Economic Review*, *38*, 627–645. http://www.jstor.org.libezp.utar.edu.my/stable/pdfplus/2527201.pdf.

Bickel, P. J., & Doksum, K. A. (2013). *Mathematical statistics,* Vol. 1, Second Edition, CRC Press, New York.

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric Statistical Inference*. CRC Press, New York.

Illowsky, B., College, D. A., & Dean, S. (2013). Introductory Statistics. In *Journal of Chemical Information and Modeling*, Vol. 53, Issue 9, OpenStax.

Jiang, Y., Wang, G. J., Ma, C., & Yang, X. (2021). Do credit conditions matter for the impact of oil price shocks on stock returns? Evidence from a structural threshold VAR model. *International Review of Economics and Finance*, *72*(November 2018), 1–15. https://doi.org/10.1016/j.iref.2020.10.019

Kong, L., Li, G., Rafique, W., Shen, S., He, Q., Khosravi, M. R., Wang, R., & Qi, L. (2022). Time-Aware Missing Healthcare Data Prediction Based on ARIMA Model. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *PP*(8), 1–10. https://doi.org/10.1109/TCBB.2022.3205064

Rajab, K., Kamalov, F., & Cherukuri, A. K. (2022). Forecasting COVID-19: Vector Autoregression-Based Model. *Arabian Journal for Science and Engineering*, *47*(6), 6851–6860. https://doi.org/10.1007/s13369-021-06526-2

Ramajo, J., Márquez, M. A., & Hewings, G. J. D. (2017). Spatiotemporal Analysis of Regional Systems: A Multiregional Spatial Vector Autoregressive Model for Spain. *International Regional Science Review*, *40*(1), 75–96. https://doi.org/10.1177/0160017615571586

Sohibien, G. P. D. (2018). Analysis of the effect of fuel price policy on Jakarta inflation by using multi-input intervention model. *AIP Conference Proceedings*, *2014*(September). https://doi.org/10.1063/1.5054529

Sohibien, G. P. D., Laome, L., Choiruddin, A., & Kuswanto, H. (2022). COVID-19 Pandemic's Impact on Return on Asset and Financing of Islamic Commercial Banks: Evidence from Indonesia. *Sustainability*, *14*(3), 1–13. https://doi.org/10.3390/su14031128

Stigler, M. (2010). *Threshold cointegration: overview and implementation in R*. http://r.meteo.uni.wroc.pl/web/packages/tsDyn/vignettes/ThCointOverview.pdf

Tsagkanos, A., Evgenidis, A., & Vartholomatou, K. (2018). Financial and monetary stability across Euro-zone and BRICS: An exogenous threshold VAR approach. *Research in International Business and Finance*, *44*, 386–393. https://doi.org/10.1016/j.ribaf.2017.07.108

Tsay و R. S (2014). *Multivariate Time Series Analysis with R and Financial Applications*, Wiley, New Jersey.

Tsay, R. S. (2010). *Analysis of financial time series*, Third Edition, Wiley, New Jersey.

Wang, D., Zheng, Y., Lian, H., & Li, G. (2021). High-Dimensional Vector Autoregressive Time Series Modeling via Tensor Decomposition. *Journal of the American Statistical Association*, Vol. 117, No. 539, 1338–1356. https://doi.org/10.1080/01621459.2020.1855183

Wei, W. W. S. (2006). *Time Series Analysis Univariate and Multivariate Methods*, Second Edition, Pearson, New York.

Yuhan, R. J., & Sohibien, G. P. D. (2018). Relationship between inflation, exchange rate, and money supply in Indonesia using threshold vector autoregressive (TVAR). *AIP Conference Proceedings*, *2014*. https://doi.org/10.1063/1.5054532.

Zhu, H. (2021). Real-time prognostics of engineered systems under time-varying external conditions based on the COX PHM and VARX hybrid approach. *Sensors*, *21*(5), 1–23. https://doi.org/10.3390/s21051712.