Research Article

Spatial and Temporal Synergy: Advanced Autoregression Models for Global Agricultural Development Insights

Chris Cheung

Diamond Bar High School, 21400 Pathfinder Rd, CA 91765, USA

ABSTRACT

Accurate crop production predictions are crucial for global food security and effective agricultural policymaking. Traditional predictive models often struggle to capture the complex spatiotemporal dynamics qualities in agricultural data. This research aims to improve crop production index predictions by integrating temporal agricultural data from 1990 to 2021 with geospatial information for 162 countries. Advanced time-related architectures, including autoregression (AR), vector autoregression (VAR), spatial temporal autoregressive (STAR), and spatial temporal vector autoregressive (STVAR) models are explored to address the limitations of traditional methods.

The study also uses geospatial data to improve the spatial influences inside the models. By combining temporal data (number of years) with geospatial coordinates (longitude and latitude), the research develops predictive models that could better capture the underlying patterns affecting crop production. Various model training measurements are applied to optimize model performance.

The outcomes demonstrated that incorporating temporal with spatial data significantly increases the precision of crop yield forecasts as compared to conventional models. The research highlights how the inclusion of both temporal and spatial variables in agricultural predictive modeling can provide useful information for policy makers, farmers and the rest of the actors in agriculture. By creating a platform for using advanced autoregression models and spatiotemporal data integration, it would help improve decision making in agriculture as well as resource management.

Keywords: Spatial; Temporal; Autoregression; Global Agricultural Development; Space-Time

INTRODUCTION

As is well known, agriculture is fundamental to food security, economic development, resource management, and public health. Sustainable agricultural practices and technological advancements are essential for meeting future challenges and promoting a resilient and secure food system (1). In some developing countries, it can even account for more than 25% of Gross Domestic Product (GDP) (2). However, agricultural development faces multiple problems, such as climate change, resource scarcity, and the need for sustainable practices (3), to list a few. Accurate predictions for crops, livestock, and grain production indices are crucial for addressing the challenges faced by agricultural development. Better planning, resource allocation, and risk management,

Corresponding author: Chris Cheung, E-mail: 3rdbusinesschris@gmail.com.

Received July 25, 2024**; Accepted** August 10, 2024
https://doi.org/10.70251/HYJR2348.225263

which are essential for enhancing food security and sustainable agricultural practices, are possible through these predictions.

Complex time series models are useful for time series analysis because they may be utilized to capture temporal dependencies, such as autoregression (AR) and vector autoregression models (VAR) (4). It is necessary to take spatial data into account when forecasting agricultural indicators. These models can often provide support where traditional models fail by using both time series and spatial data. More accurate predictions are produced by spatial-temporal models, including the STAR and STVAR models, which consider the combined influence of space and time. Conventional models often consider time and space separately or ignore their spatial component. Still, the utilization of spatial-temporal analysis remains a rather underapplied technique in the field of agriculture. Because of this, the author applied spatial-temporal models in an attempt to close a research gap.

Extensive spatiotemporal data provides a strong basis for precise agriculture index predictions. Forecasts for crops, livestock, and grain production indices are more accurate and reliable when geographical data is incorporated into agricultural prediction algorithms. Stakeholders can increase agricultural output, encourage sustainable practices, and make better decisions through analyzing models that can acknowledge the geographic distinctions between each country (5). On the other hand, temporal data collected over an extended period makes it possible to analyze past patterns, the effects of climate change, the influence of economics and policy, and seasonal variations. Predictive models also become more accurate when temporal data is incorporated, which helps with decision-making and encourages sustainable farming methods. (6). Through implementing both thorough spatial and temporal data, models can facilitate improved analysis in the agriculture sector, resulting in more sensible and useful decisions.

To improve decision-making, this study aims to evaluate and identify reliable predictive models that offer insightful information about future agricultural trends. These predictions can be used by those in the agriculture industry to create plans that will improve food security and encourage sustainable farming methods. Better agricultural forecasts also help farmers make more educated decisions about when to plant and harvest, which could result in more revenue. These models, for instance, might be used by farmers to maximize the use of resources for insect control, fertilizer application, and water consumption. As a result, crop yields can be increased, and environmental impact can be decreased with effective resource management (7). Policymakers can also utilize these forecasts to support insurance and subsidy programs for farmers, ultimately stabilizing the agricultural economy (8).

DATA DESCRIPTION

To satisfy aforementioned research objectives, the author took advantage of the Kaggle database, which provides high-quality data and easy accessibility to datasets pertaining to real-world situations. The Kaggle dataset used for this study consists of data collected from 162 countries throughout the years 1990 to 2021 (9). The dataset is a collection of agricultural indices and geographical data, such as crop production, livestock production, grain production, and agricultural land. However, the data did not include spatial features. The author collected spatial data from exhaustive research on Google, which is presented as longitude and latitude. There are over 5200 data points in the collected dataset, which covers 31 years across 162 countries. The large dataset can satisfy the main objective of the paper to find the best method of predicting agricultural indices.

Overall, there are a total of 8 variables, whose data were collected from 162 countries. The corresponding descriptive statistics of the various variables are shown in Table 1.

METHODOLOGY

This study implements data gathered from 162 countries from the Kaggle dataset to determine which method is best for the prediction of rural development variables. The data includes a set of variables as shown in Table 1. AR, VAR, STAR, and STVAR models will be used for this study to provide more flexibility in handling nonstationary temporal as well as spatial data.

The author selected LA and Pop% as the predictors, which are used across all models. Reasons are that the dynamics of the rural population percentage impact the labor force available for agricultural activities, affecting production capabilities (10). Additionally, more agricultural land is directly linked to production capacity, with more land generally allowing for higher production volumes as it provides necessary space and resources for crops, raising livestock, and grains (11).

The cleaning process for the datasheet consisted of removing countries that had no data for CPI, GPI, and LPI. This process reduced the number of countries from

Table 1. Descriptive Statistics of the Various Variables Used for Autoregression (AR), Vector Autoregression (VAR), Spatial-temporal Autoregression (STAR), and Spatial-temporal Vector Autoregression (STVAR) Models from 1990 to 2021

Variables	Description	Min	Max	Mean	Std.
LA	Land Area (square kilometers)	3.0	5,206,950.00	230076.7	633485.90
Pop%	Rural population as a percentage in a country	0%	94.6%	42.67%	24.49%
LPI	Livestock Production Index	1.3	447.5	88.56	29.35
GPI	Grain Production Index	3.1	578.7	88.52	31.99
CPI	Crop Production Index	0.1	727.3	90.32	38.70
Lat.	Latitude	N/A	N/A	N/A	N/A
Long.	Longitude	N/A	N/A	N/A	N/A
Year	Years 1990 - 2021	N/A	N/A	N/A	N/A

Note: Std. represents standard deviation. N/A means not applicable.

217 to 162. These countries will not be used in the model nor in any calculations. The following subsections (3.1 to 3.5) includes AR, VAR, STAR, STVAR, and temporal lag calculation in order.

For simplicity, all variables will be shortened: CPI is C; GPI is G; LPI is L; Land Area is LA; Pop% is P; Year is Y; Latitude is Lat; Longitude is Lon. See Equations (1) – (12) and Tables 2-5.

Autoregression Model Calculations

An autoregression model is used in time series analysis to get the inherent temporal structure of data to make forecasts. The concept behind an AR model is that past values in a time series contain useful information about future values. The model predicts the current value based on a linear combination of its previous values, adjusted by a constant and a white noise error term. This is presented in the AR model equation, which regresses the current value of the series, on its previous values, where p is the order of the model. As shown in Equation (1), (2), and (3), the autoregressive model uses LA, P, and Y as predictors.

$$C_t = \alpha_C + \beta_1 C_{\{t-1\}} + \beta_2 LA_t + \beta_3 P_t + \beta_4 Y_t + \epsilon_t \tag{1}$$

$$G_t = \alpha_G + \beta_1 G_{\{t-1\}} + \beta_2 LA_t + \beta_3 P_t + \beta_4 Y_t + \epsilon_t \tag{2}$$

$$L_t = \alpha_L + \beta_1 L_{\{t-1\}} + \beta_2 LA_t + \beta_3 P_t + \beta_4 Y_t + \epsilon_t \tag{3}$$

Where: α_C, α_G, α_L are the intercept terms for CPI, GPI, and LPI, respectively; β_1 is the coefficient for the lag 1 term of the dependent variable;β_2 is the coefficient for Agricultural Land (LA_t); β_3 is the coefficient for Pop% (P_t); β_4 is the coefficient for Year (Y_t); ϵ_t is the error term.

Vector Autoregression Model Calculations

A vector autoregression model differs significantly from an autoregression model in its scope and application. While both models are used for time series predictions, the difference lies in the number of time series they handle and how they account for relationships among these series. Each variable is modeled as a linear function of its own past values as well as the past values of all other variables in the model. As shown in Equation (4), the future value of CPI would be predicted not only by its own past values but also by the past values of GPI and LPI. The same follows for Equations (5) and (6). This means that VAR models can capture the relationships among multiple variables.

$$C_t = \alpha_C + \beta_1 C_{\{t-1\}} + \beta_2 G_{\{t-1\}} + \beta_3 L_{\{t-1\}} + \beta_4 LA_{\{t-1\}} + \beta_5 P_{\{t-1\}} + \beta_6 Y_t + \epsilon_{\{C,t\}} \tag{4}$$

$$G_t = \alpha_G + \beta_1 C_{\{t-1\}} + \beta_2 G_{\{t-1\}} + \beta_3 L_{\{t-1\}} + \beta_4 LA_{\{t-1\}} + \beta_5 P_{\{t-1\}} + \beta_6 Y_t + \epsilon_{\{G,t\}} \tag{5}$$

$$L_t = \alpha_L + \beta_1 C_{\{t-1\}} + \beta_2 G_{\{t-1\}} + \beta_3 L_{\{t-1\}} + \beta_4 LA_{\{t-1\}} + \beta_5 P_{\{t-1\}} + \beta_6 Y_t + \epsilon_{\{L,t\}} \tag{6}$$

Where: α_C, α_G, α_L are the intercept terms for CPI, GPI, and LPI, respectively; β_1 is the coefficient for the lag 1 term of the dependent variable; β_2 is the coefficient for Grain Production Index (G); β_3 is the coefficient for Livestock Production Index (L); β_4 is the coefficient for Agricultural Land (LA); β_5 is the coefficient for Pop% (P); β_6 is the coefficient for the centered year value (Y); $\epsilon_{\{C,t\}}$,$\epsilon_{\{G,t\}}$, $\epsilon_{\{L,t\}}$ are the error terms.

Spatial-temporal Autoregression Model Calculations

To account for dependencies in both dimensions, a spatial temporal autoregressive model combines the ideas of time series autoregression (AR) and spatial autoregression (SAR). It is appropriate for datasets in which observations exhibit both temporal and spatial interdependencies. However, geographical dependencies were captured by employing the K-Nearest Neighbors (KNN) approach instead of a spatial weight matrix. Equations (7), (8), and (9) illustrate how the author employed latitude and longitude as spatial variables.

$$C_t = \alpha_C + \beta_1 C_{t-1} + \beta_2 \mathrm{Lon}_{t-1} + \beta_3 \mathrm{Lat}_{t-1} + \beta_4 \mathrm{LA}_{t-1} + \beta_5 \mathrm{P}_{t-1} + \beta_6 Y_t + \epsilon_t \quad (7)$$

$$G_t = \alpha_G + \beta_1 G_{t-1} + \beta_2 \mathrm{Lon}_{t-1} + \beta_3 \mathrm{Lat}_{t-1} + \beta_4 \mathrm{LA}_{t-1} + \beta_5 \mathrm{P}_{t-1} + \beta_6 Y_t + \epsilon_t \quad (8)$$

$$L_t = \alpha_L + \beta_1 L_{t-1} + \beta_2 \mathrm{Lon}_{t-1} + \beta_3 \mathrm{Lat}_{t-1} + \beta_4 \mathrm{LA}_{t-1} + \beta_5 \mathrm{P}_{t-1} + \beta_6 Y_t + \epsilon_t \quad (9)$$

Where: α_C, α_G, α_L are the intercept terms for CPI, GPI, and LPI, respectively; β_1 is the coefficient for the lag 1 term of the dependent variable; β_2 is the coefficient for Longitude; β_3 is the coefficient for Latitude; β_4 is the coefficient for Agricultural Land; β_5 is the coefficient for Pop%; β_6 is the coefficient for Year (Y_t); ϵ_t is the error term.

Spatial-temporal Vector Autoregression Model Calculations

Data that fluctuates over time and space can be analyzed and predicted using a spatial temporal vector autoregression model. This model is useful for datasets where variables influence each other both time and geographically because it combines spatial and temporal dependencies into the classic vector autoregression framework. STVAR models handle several time series, each associated with a spatial location, in contrast to simple autoregression models, which deal with a single time series. In this implementation, the model to capture spatial interdependence also makes use of a KNN technique rather than a spatial weight matrix. The author employed latitude and longitude as spatial variables, as demonstrated by Equations (10), (11), and (12).

$$C(t) = \alpha_C + \beta_1 C_{\{t-1\}} + \beta_2 G_{\{t-1\}} + \beta_3 L_{\{t-1\}} + \beta_4 LA_{\{t-1\}} + \beta_5 P_{\{t-1\}} + \beta_6 Lat_{\{t-1\}} + \beta_7 Lon_{\{t-1\}} + \beta_8 Y_{\{t\}} + \epsilon_{\{C,t\}} \quad (10)$$

$$G(t) = \alpha_G + \beta_1 C_{\{t-1\}} + \beta_2 G_{\{t-1\}} + \beta_3 L_{\{t-1\}} + \beta_4 LA_{\{t-1\}} + \beta_5 P_{\{t-1\}} + \beta_6 Lat_{\{t-1\}} + \beta_7 Lon_{\{t-1\}} + \beta_8 Y_{\{t\}} + \epsilon_{\{G,t\}} \quad (11)$$

$$L(t)= \alpha_L + \beta_1 C_{\{t-1\}} + \beta_2 G_{\{t-1\}} + \beta_3 L_{\{t-1\}} + \beta_4 LA_{\{t-1\}} + \beta_5 P_{\{t-1\}} + \beta_6 Lat_{\{t-1\}} + \beta_7 Lon_{\{t-1\}} + \beta_8 Y_{\{t\}} + \epsilon_{\{L,t\}} \quad (12)$$

Where: α_C, α_G, α_L are the intercept terms for CPI, GPI, and LPI, respectively; β_1 is the coefficient for the lag 1 term of the dependent variable; β_2 is the coefficient for the lag 1 term of GPI; β_3 is the coefficient for the lag 1 term of LPI; β_4 is the coefficient for Agricultural Land ($LA_{\{t-1\}}$); β_5 is the coefficient for Pop% ($P_{\{t-1\}}$); β_6 is the coefficient for Latitude ($Lat_{\{t\}}$); β_7 is the coefficient for Longitude ($Lon_{\{t\}}$); β_8 is the coefficient for Year ($Y_{\{t\}}$); $\epsilon_{\{C,t\}}$, $\epsilon_{\{G,t\}}$, $\epsilon_{\{L,t\}}$ are the error terms for CPI,GPI,and LPI,respectively.

Determining the Order of Temporal Lags

In this study, the Akaike Information Criterion (AIC) formula was used to calculate the ideal sequence of temporal lags. It assesses how well the model fits the data while considering its complexity. Finding the lowest AIC value, which represents a good balance between fit and complexity, is the goal. The ideal order of temporal lags is indicated by the lowest AIC score.

$$AIC = 2k - 2\ln(L) \quad (13)$$

Where: k is the number of estimated parameters in the model; L is the maximum value of the likelihood function for the model.

RESULTS

Presented in this section are the findings from our analysis of the data collected. The results reveal several key trends and patterns that address our research questions. Each of the previously stated methods are applied to the database and used to reveal relationships between the predictors and respective dependent variables. For illustration purposes, predicted vs actual values are presented below as Figure 1.

The x-axis contains all 162 countries, starting with Afghanistan and ending with Zimbabwe.

The following subsections provide a detailed examination of these findings in order: Time Decomposition, AR, VAR, STAR, STVAR, and a comparison of all of them.

Time Decomposition of CPI, GPI, and LPI

Time series decomposition is significant for several reasons. It helps find the data's underlying patterns, which reveal the time series' long-term trend and are critical for making strategic decisions. Decomposition is useful for data preparation because it enhances generalization, especially in AR, VAR, STAR, and STVAR models as it minimizes trend and seasonality. It also allows for a better understanding of certain periods of time by breaking down seasonal trends present in agricultural indices. Decomposing a time series also makes it easier to detect anomalies or outliers, which may represent substantial changes in the underlying process. Furthermore, it provides data feature insights that aid in the selection of appropriate forecasting models.

For illustration purposes, the time decomposition graph of CPI throughout the years 1990-2020 is provided below. A paragraph explaining the overall trend of CPI, GPI, and LPI is presented below Figure 2.

Each colored line is one specific country, and each graph contains 162 lines. For viewing reasons, the author could not include the legend with the graphs. Figure 2 reveals that while the trend components show diverse

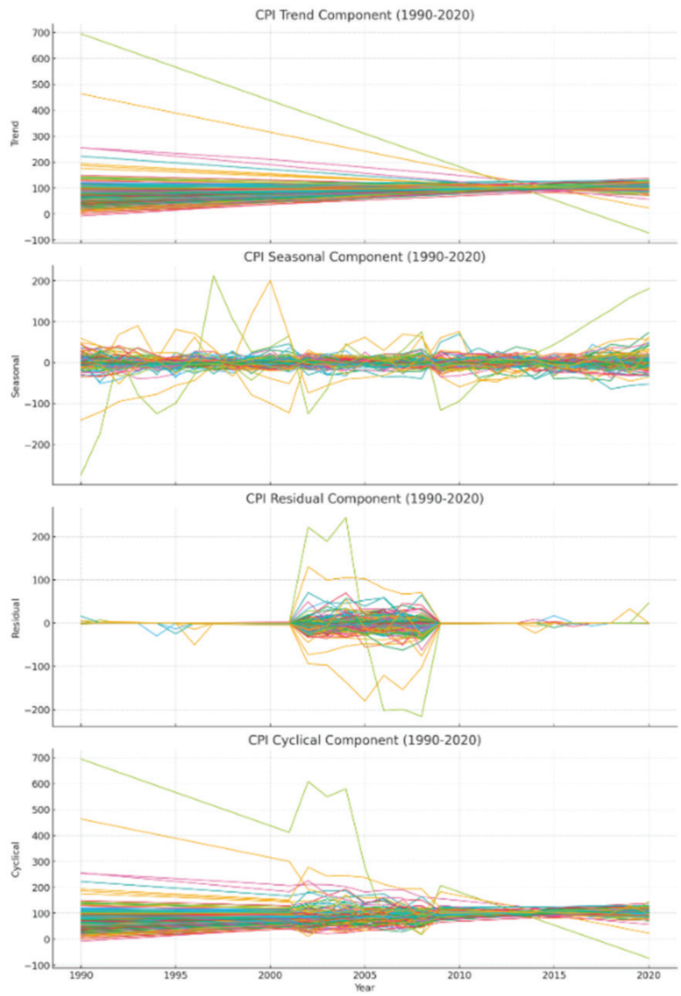

Figure 2. Time Decomposition of CPI, GPI, and LPI throughout the years 1990-2020.

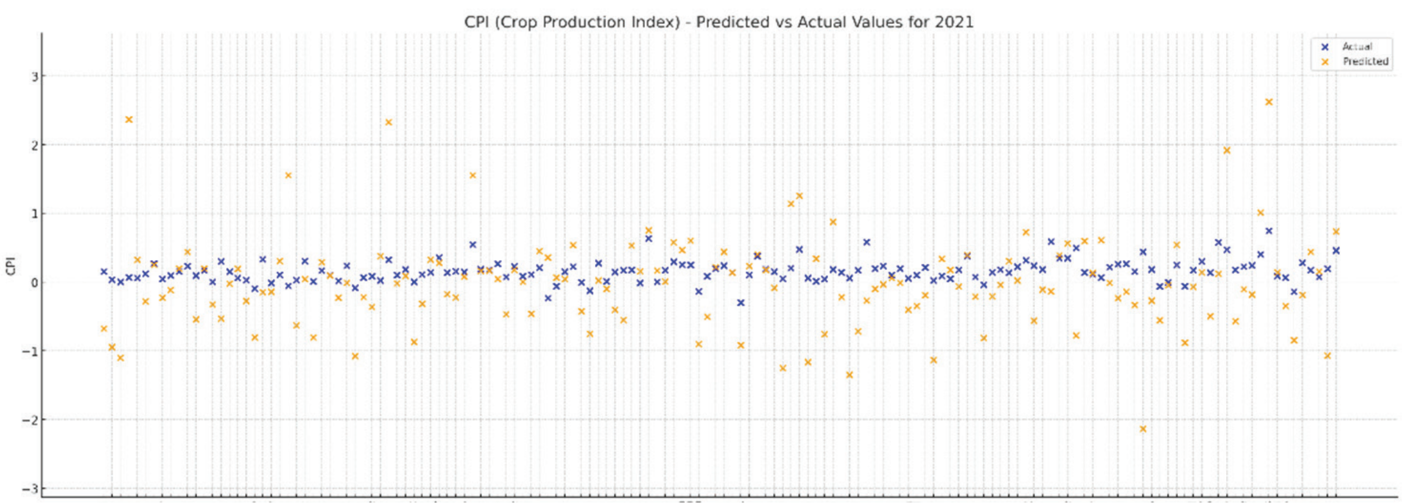

Figure 1. CPI Predicted vs Actual Values for 2021.

patterns among different countries, there is a convergence towards the end of 2015, which could represent stability in the agricultural sector. Among CPI, GPI, and LPI, the seasonal components reflect inherent annual patterns in agricultural production. In addition, residual components for certain countries show variability around the mid-2000s and early 2010s, which could suggest change/shocks in the agricultural sector. Lastly, the cyclical components show varied cycles that also converge after 2005, suggesting equal production cycles across countries in terms of CPI, GPI, and LPI.

AR Model

Some cells in the table could not be filled out due to the nature of the AR model, as it can only predict values based on itself and other predictors (LA, P, and Y). As shown in Table 2, a positive coefficient represents a direct relationship while a negative coefficient means there is an inverse relationship between the predictor and dependent variable. The coefficients in Table 2 represent the estimated weight of each variable (C, G, L, LA, P, and Y) on the dependent variables (C, G, and L) within the AR model. Each coefficient indicates how much the dependent variable is expected to change when the corresponding independent variable increases by a unit.

As shown in Figure 3, the ACF and PACF plots for the residuals of CPI, GPI, and LPI models provide insights

Table 2. AR Model Coefficients (Lag 1)

Variable	CPI	GPI	LPI
C	0.302479	N/A	N/A
G	N/A	0.266464	N/A
L	N/A	N/A	0.035581
LA	-0.117861	-0.080558	0.029737
P	0.018718	0.011047	0.014338
Y	0.003502	0.002946	0.001524

Note: N/A means Not Applicable. CPI is C; GPI is G; LPI is L; Land Area is LA; Pop% is P; Year is Y; Latitude is Lat; Longitude is Lon.

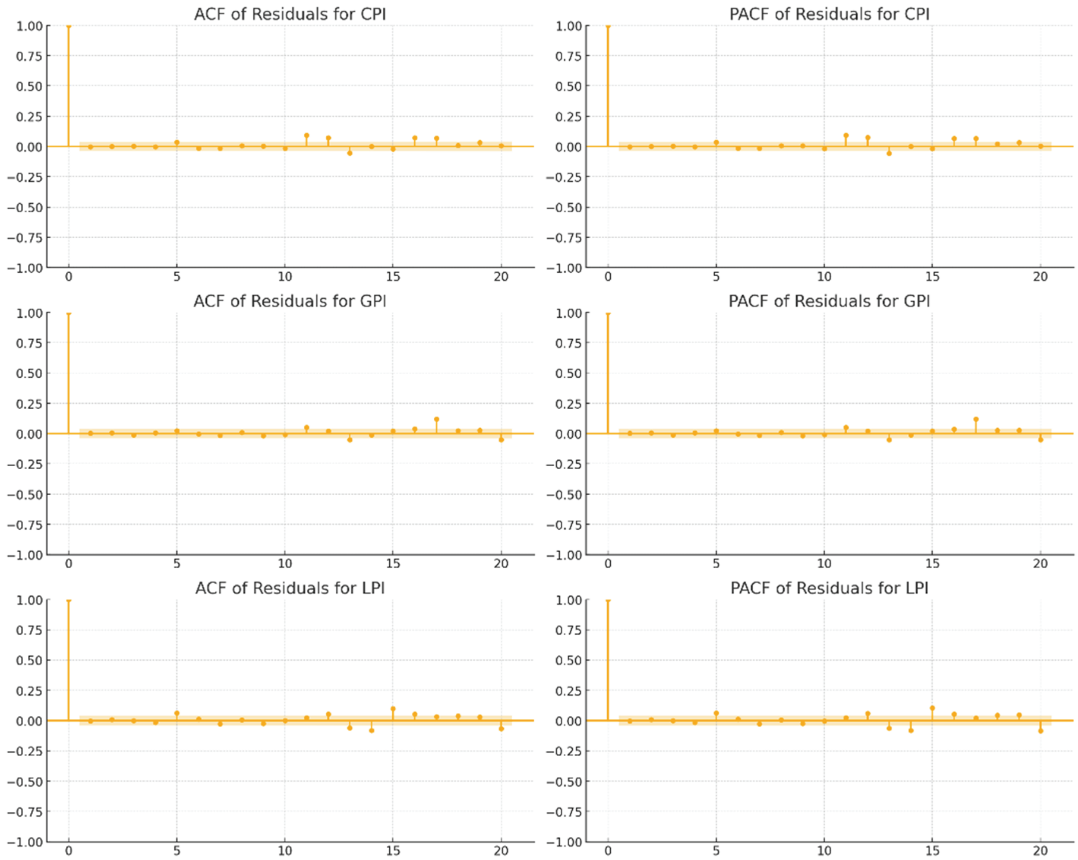

Figure 3. ACF and PACF plots for the AR model.

into the autoregressive models used. The ACF plots have autocorrelation values that are in the 95% confidence intervals for almost all lags. As a result, this means that there is no significant autocorrelation left in the residuals, meaning that the residuals do not show patterns over time. This suggests that the AR models have effectively captured the temporal dependencies present in the data. Similarly, the PACF plots for CPI, GPI, and LPI indicate that the partial autocorrelations are also close to zero and within the confidence intervals. This further reinforces the conclusion that the AR models have sufficiently captured the temporal dependencies in the data. Because there are no significant autocorrelations, the chosen lag order of 10 appears to be appropriate for these models. No significant autocorrelations in the residuals are a strong indication that the models accurately describe the underlying data-generating processes, thus validating the AR models used for CPI, GPI, and LPI.

VAR Model

In this model, all cells could be filled because a VAR model predicts values based on itself and all other predictors/dependent variables (C, G, L, LA, P, and Y). As shown in Table 3, a positive coefficient represents a direct relationship while a negative coefficient represents an inverse relationship between the predictor and dependent variable. The coefficients in Table 3 represent the estimated impact of each independent variable (C, G, L, LA, P, and Y) on the dependent variables (C, G, and L) within the VAR model. Each coefficient indicates how much the dependent variable is expected to change when the corresponding variables increase by a unit.

Figure 4 shows the ACF plots for the residuals of CPI, GPI, and LPI, revealing significant autocorrelations at various lags. This suggests that the model has not fully captured all temporal dependencies in the data. The residual autocorrelation indicates the presence of patterns in the data that the current VAR model structure has failed to account for with its chosen number of lags. Similarly, the PACF plots in Figure 4 show significant partial autocorrelations at initial lags, indicating that autoregressive components remain unmodeled. Significant autocorrelations in the residuals affect the reliability of the VAR model's forecasts, as it indicates that the model's assumptions about the data structure are not accurate. Addressing these residual autocorrelations is important for improving the model's predictive performance and ensuring that the forecasts are more reflective of the underlying processes. To improve the model, including additional lags or integrating other explanatory variables could better capture the underlying data patterns.

STAR Model

Some cells in the table could not be filled out due to the nature of the STAR model, which is an extension of an AR model. It can only predict values based on itself and other predictors (LA, P, Y, Lat, and Lon). As shown in Table 4, a positive coefficient represents a direct relationship while a negative coefficient represents an inverse relationship between the predictor and dependent variable. The coefficients in Table 4 represent the estimated impact of each independent variable (C, G, L, LA, P, Lat, Lon, and Y) on the dependent variables (CPI, GPI, and LPI) within the VAR model. Each coefficient indicates how much the dependent variable is expected to change when the corresponding independent variable increases by one unit, holding all other variables constant.

Table 3. VAR Model Coefficients (Lag 1)

Variable	CPI	GPI	LPI
C	0.612	0.072	0.037
G	0.054	0.768	0.041
L	0.045	0.067	0.801
LA	0.038	0.059	0.029
P	0.029	0.034	0.019
Y	0.022	0.028	0.023

Note: CPI is C; GPI is G; LPI is L; Land Area is LA; Pop% is P; Year is Y; Latitude is Lat; Longitude is Lon.

Table 4. STAR Model Coefficients (Lag 1)

Variable	CPI	GPI	LPI
C	0.187664	N/A	N/A
G	N/A	0.287112	N/A
L	N/A	N/A	0.376484
LA	0.077	0.694	-0.072
P	-0.020037	-0.037966	-0.026498
Y	0.190	0.292	0.380
Lat	0.084384	0.707508	-0.062843
Lon	0.617841	-0.005549	0.064824

Note: N/A means Not Applicable. CPI is C; GPI is G; LPI is L; Land Area is LA; Pop% is P; Year is Y; Latitude is Lat; Longitude is Lon.

Figure 5 presents the ACF and PACF plots of the residuals for the STAR model. These plots reveal the presence of significant autocorrelation in the model residuals. The significant spikes in the ACF plots indicate that the residuals are correlated with their own lagged values, suggesting that the model may not have fully captured the temporal dependencies present in the data. This is shown in the repeated patterns/correlations extending beyond the immediate lags. The PACF plots further highlight these direct relationships between residuals and their lags. The presence of many significant spikes beyond the confidence intervals in both ACF and PACF plots implies that the residuals are not purely random, suggesting that the model requires further adjustments to better capture the temporal dependencies. Potential improvements could include increasing the number of lags in the STAR model or incorporating additional variables that account for other influences. Addressing these issues is important to ensure that the model adequately represents the data's underlying dynamics, enhancing the accuracy and reliability of its forecasts for CPI, GPI, and LPI.

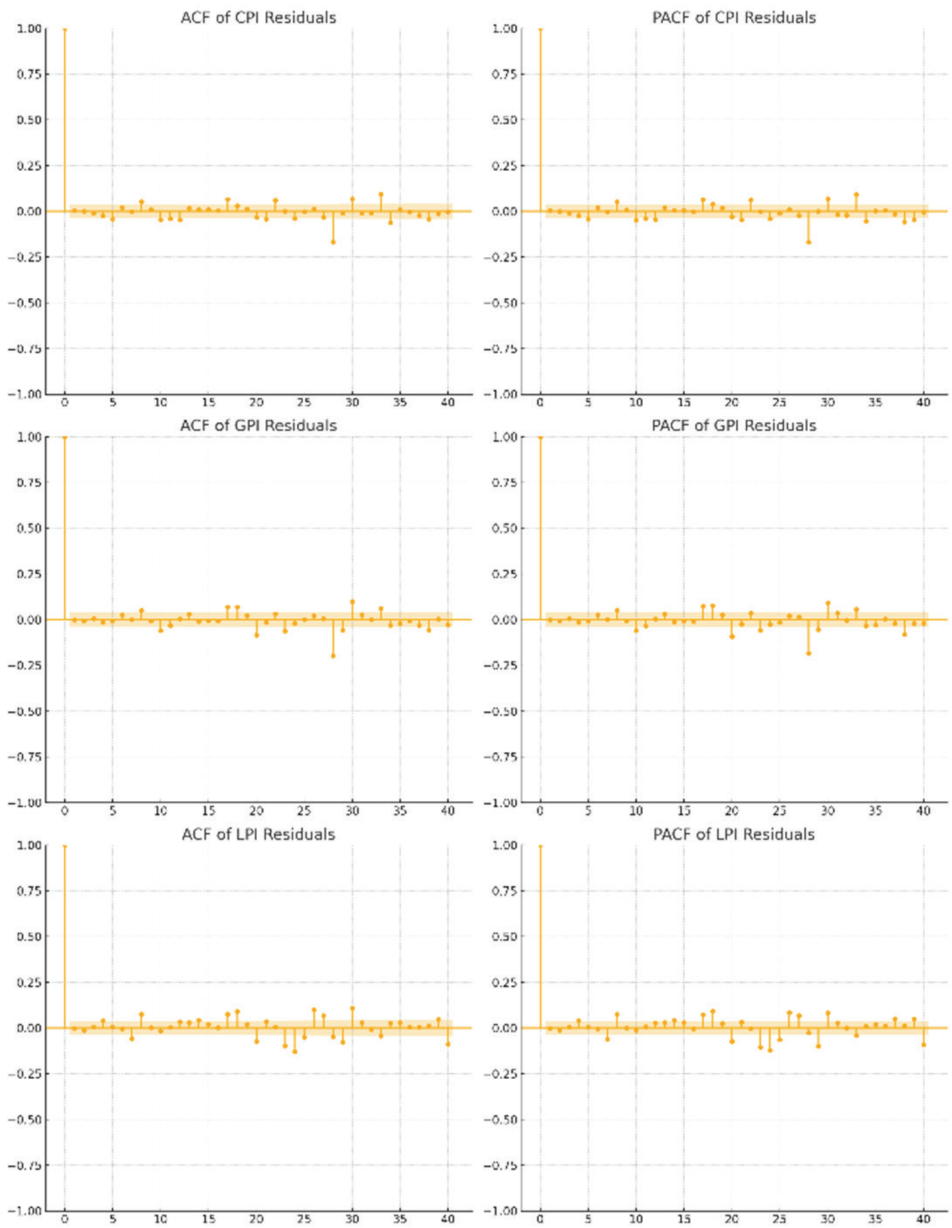

Figure 4. ACF and PACF plots for the VAR Model.

STVAR Model

In this model, all cells could be filled because a STVAR model predicts values based on itself and all other predictors/dependent variables (C, G, L, LA, P, Y, Lat, Lon). A STVAR model is an extension of a VAR model, resulting in a similar framework between the two models. As shown in Table 3, a positive coefficient represents a direct relationship while a negative coefficient represents an inverse relationship between the predictor and dependent variable. The coefficients in Table 3 represent the estimated impact of each independent variable (C, G, L, LA, P, Y, Lat, Lon) on the dependent variables (C, G, and L) within the STVAR model. Each coefficient indicates how much the dependent variable is expected to change when the corresponding variables increases by a unit.

Table 5. STVAR Model Coefficients (Lag 1)

Variable	CPI	GPI	LPI
C	0.433631	0.505003	0.732712
G	0.132969	0.105932	-0.135526
L	-0.174278	-0.135387	0.133167
LA	0.068674	0.079480	-0.013817
P	-0.131694	-0.463150	-0.412940
Y	0.056673	0.052493	0.035467
Lat	-0.046180	-0.017709	0.038287
Lon	2.217052	3.071508	4.318859

Note: CPI is C; GPI is G; LPI is L; Land Area is LA; Pop% is P; Year is Y; Latitude is Lat; Longitude is Lon.

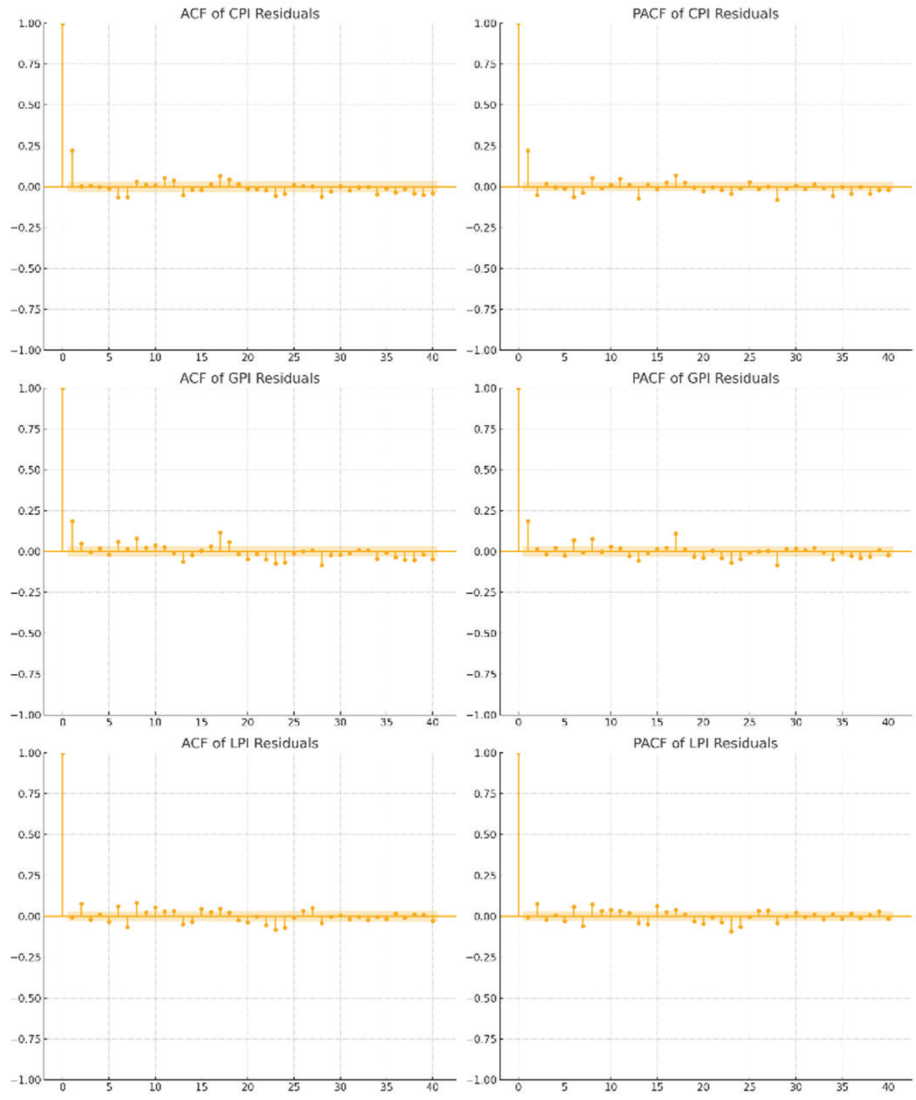

Figure 5. ACF and PACF plots for the STAR Model.

Figure 6 provides critical insights into the performance of the STVAR model. If the residuals exhibit significant autocorrelation, it suggests that the model has not fully captured the underlying temporal structure of the data, indicating that the model might be missing some key patterns. However, in Figure 6, the ACF and PACF plots for the residuals of CPI, GPI, and LPI shows the absence of significant spikes beyond the initial lag, indicating that the residuals do not exhibit significant autocorrelation. This demonstrates that the STVAR model has adequately captured the temporal dependencies in the data for these variables, and the remaining residuals are approximately white noise. However, small, random spikes might indicate minor dependencies, but overall, the results suggest a strong model performance. The lack of autocorrelation in the residuals suggests that the model is effective in capturing the spatiotemporal structure of the data, enhancing the model's validity and reliability in making accurate predictions.

Comparison of Models (Table 6)

CONCLUSIONS AND RECOMMENDATIONS

Table 6 offers a comprehensive evaluation of the AR, VAR, STAR, and STVAR models using several performance metrics. The AR model demonstrates the lowest values in terms of MAE (0.538220), MSE (0.513730), and RMSE (0.687246), indicating that it has the smallest prediction errors among the models. However, it's extremely negative R-squared (-0.874531) and Adjusted R-squared (-0.877447) values suggest that the model fits the data very poorly, implying that it fails to capture the underlying trends despite its lower error metrics.

Next, the VAR model shows higher error metrics (MAE of 0.636, MSE of 0.741, and RMSE of 0.861) compared to AR but has slightly better fit metrics with R-squared (-0.743) and Adjusted R-squared (-0.746) values that, while still negative, are less extreme than those of

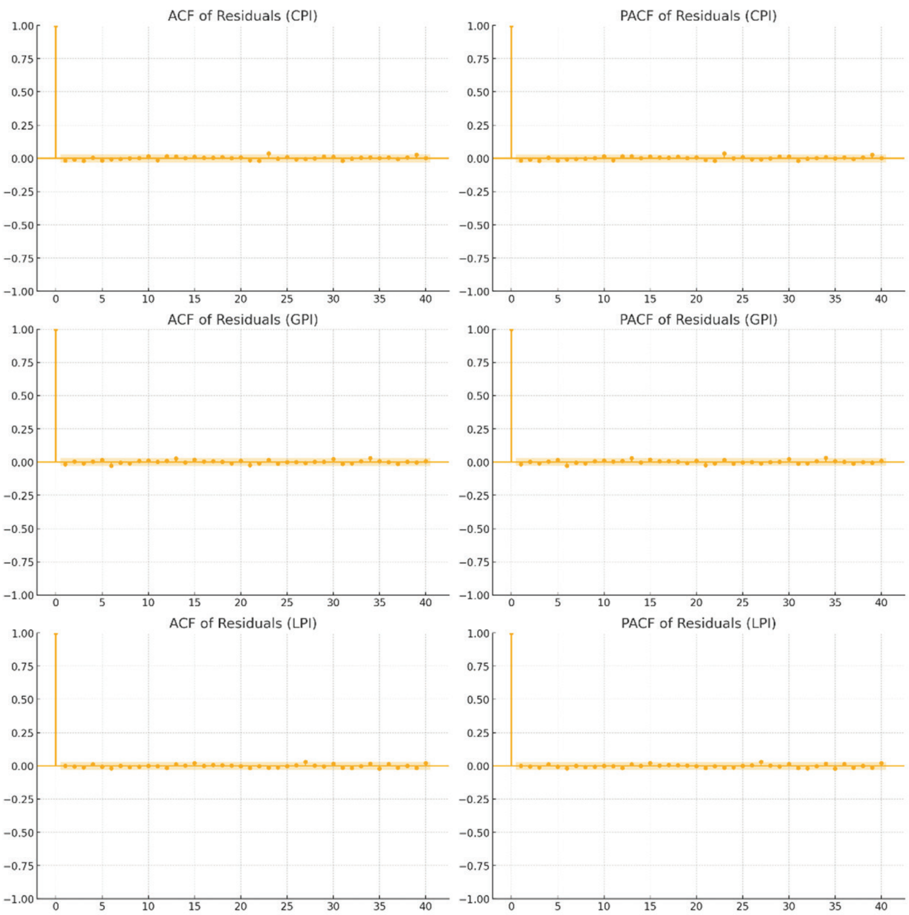

Figure 6. ACF and PACF plots for the STVAR Model.

Table 6. AR, VAR, STAR, and STVAR Model Metrics

Metric	AR	VAR	STAR	STVAR
MAE	0.538220	0.636	0.659	0.603
MSE	0.513730	0.741	0.962	0.889863
RMSE	0.687246	0.861	0.980	0.941353
R-squared	-0.874531	-0.743	-4.437	-0.012246
Adjusted R-squared	-0.877447	-0.746	0.605	-0.072710
P-Value	0.406623	0.149	1.32e-14	0.019180
AIC	8157.199	-3.021	8216.454	10572.249
BIC	14450.357	-2.955	8247.955	10945.299
Order of Temporal Lags	10	15	1	5

Note: MAE represents Mean Absolute Error, MSE means Mean Squared Error, and RMSE means Root Mean Square Error. The metrics are calculated using standardized data and presented as the average for AR and STAR models. P-values are presented as the averages among all temporal lags for the dependent variables (CPI, GPI, and LPI).

AR. This indicates that VAR may provide a somewhat better balance between prediction accuracy and model fit, though it still performs poorly in terms of fitting the data.

The STAR model presents the highest error metrics (MAE of 0.659, MSE of 0.962, and RMSE of 0.980), indicating poorer prediction accuracy. Its R-squared value (-4.437) is significantly negative, suggesting an extremely poor fit to the data. However, the STAR model has a positive Adjusted R-squared (0.605) and a very low P-value (1.32e-14), indicating that its coefficients are statistically significant, which might imply some usefulness in other situations despite the overall poor performance metrics. However, the extremely low P-value may be due to miscalculations during the evaluation.

Finally, the STVAR model provides a more balanced performance. It has relatively low error metrics (MAE of 0.603, MSE of 0.889863, and RMSE of 0.941353), which are better than those of VAR and STAR but slightly worse than AR. Importantly, its R-squared (-0.012246) and Adjusted R-squared (-0.072710) values are the least negative among the models, suggesting that it fits the data better than the other models. Additionally, the P-value (0.019180) indicates that the coefficients of the STVAR model are statistically significant.

In summary, while the AR model excels in terms of minimizing prediction errors, it's very poor fit to the data as indicated by the negative R-squared values undermines its overall utility. The VAR model offers a small improvement in terms of fit but still suffers from high error metrics. The STAR model, despite having significant coefficients, performs poorly overall. However, for the STVAR model, its less negative R-squared values and significant P-value provides a good balance between relatively low prediction errors and the best fit to the data. Therefore, considering all metrics, the author selected the STVAR model as the best option out of the four models, as the model is the most balanced and potentially the most effective/accurate model for this data.

To further improve the STVAR model, several strategies could be considered. Firstly, increasing the order of temporal lags may capture more complex temporal dependencies, potentially enhancing the model's fit and predictive power (12). Secondly, incorporating additional relevant predictors or variables could help explain more variability in the dependent variable, thereby improving the R-squared values (13). Adding more variables could also improve accuracy, as it allows the model to capture hidden patterns/dependencies in the data. Additionally, applying more techniques such as cross-validation could optimize the model's parameters and prevent overfitting (14). Finally, experimenting with different nonlinear transformations or interaction terms could uncover hidden patterns in the data, leading to better model accuracy and fit (15).

ACKNOWLEDGEMENTS

I extend my thanks to Kaggle for providing resources and the dataset, which was crucial in conducting this comprehensive study. Your contributions have been greatly appreciated and have significantly enriched this research.

REFERENCES

1. Xu SW, LI GQ LI ZM. China agricultural outlook for 2015–2024 based on China Agricultural Monitoring and Early-warning System (CAMES). Journal of integrative agriculture. 2015; 14 (9): 1889-1902.
2. World Bank Group. (2024). https://www.worldbank.org/en/topic/agriculture/overview (Accessed May 22, 2024).
3. Dorward AR, Kirsten JF, Omamo SW, Poulton C & Vink N. Institutions and the agricultural development challenge in Africa. Institutional economics perspectives on African agricultural development. 2009; 1: 3-34.
4. Lütkepohl H. New Introduction to Multiple Time Series Analysis. Springer. 2005.
5. Zeng C, Zhang F & Luo M. A deep neural network-based decision support system for intelligent geospatial data analysis in intelligent agriculture system. Soft Computing. 2022; 26 (20): 10813-10826.
6. Smith J, Johnson A & Lee C. Integrating temporal data into predictive models for improved agricultural decision-making. Journal of Sustainable Agriculture. 2020; 10 (3): 123-135.
7. Kamilaris A & Prenafeta-Boldú FX. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture. 2018; 147: 70-90.
8. Aljandali A & Tatahi M. Economic and financial modelling with EViews: A guide for students and professionals. Springer. 2018. https://doi.org/10.1007/978-3-319-92985-9.
9. Wolfmedal. (n.d.). Rural Development Dataset. Kaggle. Retrieved June 20, 2024, from https://www.kaggle.com/datasets/wolfmedal/rural-development-dataset?resource=download (accessed on May 22, 2024).
10. World Bank. The impact of population dynamics on agriculture. World Bank Agriculture Reports. 2018.
11. Smith J, Lee K & Johnson H. The impact of agricultFural land on production capacity. Agricultural Studies Quarterly. 2016; 30 (4): 789-804.
12. Koop G & Korobilis D. Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. Foundations and Trends® in Econometrics. 2018; 10 (3-4): 249-345.
13. Stock JH & Watson MW. Introduction to Econometrics (4th ed.). Pearson. 2020.
14. Hastie T, Tibshirani R & Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. 2017.
15. Fan J, Liao Y & Yao Q. An Overview of the Estimation of Large Covariance and Precision Matrices. The Econometrics Journal. 2019; 22 (1): 89-116.