

## *Retraction*

# **Retracted: Application of Bayesian Vector Autoregressive Model in Regional Economic Forecast**

### **Complexity**

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] J. Ma, Y. Shang, and H. Zhang, "Application of Bayesian Vector Autoregressive Model in Regional Economic Forecast," *Complexity*, vol. 2021, Article ID 9985072, 10 pages, 2021.

## Research Article

# Application of Bayesian Vector Autoregressive Model in Regional Economic Forecast

Jinghao Ma,<sup>1</sup> Yujie Shang,<sup>1</sup> and Hongyan Zhang<sup>1,2</sup> 

<sup>1</sup>*Pai Chai University, Daejeon 35345, Republic of Korea*

<sup>2</sup>*School of Business Administration, Shandong Women's University, Jinan City, Shandong Province 250300, China*

Correspondence should be addressed to Hongyan Zhang; 30094@sdwu.edu.cn

Received 22 March 2021; Revised 20 April 2021; Accepted 23 April 2021; Published 6 May 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Jinghao Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Bayesian vector autoregressive (BVAR) model introduces the statistical properties of variables as the prior distribution of the parameters into the traditional vector autoregressive (VAR) model, which can overcome the problem of too little freedom. The BVAR model established in this paper can overcome the problem of short time series data by using prior statistical information. In theory, it should have a good effect in China's regional economic forecasting. Most regional forecasting model literature lacks out-of-sample forecasting error evaluation research in the real sense, but our early forecasts of major economic indicators provide an excellent opportunity for this paper to evaluate the actual forecast errors of the BVAR model in detail. The analysis in this paper shows that the prediction error of the BVAR model is very small and the prediction ability is very satisfactory. At the same time, this article also analyzes and points out the direction of efforts to further improve the prediction accuracy of the BVAR model.

## 1. Introduction

Economic forecasting is the application of forecasting technology in the economic field, and it is a method of forecasting fiscal revenue [1]. Methods are divided into two categories: one is qualitative prediction methods and the other is quantitative prediction methods, such as regression analysis methods. In actual economic forecast analysis, qualitative forecasts are often combined with fixed forecasts, based on qualitative analysis and quantitative analysis as a means, organically combining the nature and quantity of economic activities and complementing each other. And this article will introduce how to use regression analysis to forecast fiscal revenue. In regression analysis and prediction, the prediction object (dependent variable) is firstly analyzed qualitatively to determine one or more factors that affect its change, and then an appropriate regression prediction model is established for prediction through multiple sets of observation values of the prediction object and influencing factors [2]. This method uses the causal relationship between the dependent variable and the independent variable, so it is also called causal regression analysis. In the economic field,

an economic variable is often affected by multiple factors. Therefore, it is necessary to establish a multiple regression model for prediction. However, the conclusion that the regression equation is significant in the multiple regression model is not satisfied because the regression equation is the most significant and it does not mean that the influence of the independent variable on the dependent variable is important. The main task is to remove those secondary and dispensable variables from these variables to establish a simpler regression equation, so as to better understand predictions, and stepwise regression analysis is a good way to solve this problem [3]. The specific method of stepwise regression is to introduce variables individually, but each time a variable is introduced, the variables of the selected person must be tested one by one. When the originally introduced variable becomes no longer significant due to the introduction of subsequent variables, it must be removed. Introduce a variable or remove a variable from the regression equation. For stepwise regression, an  $F$  test must be performed at each step to ensure that only significant variables are included in the regression equation before each new variable is introduced. This process is repeated until no

significant independent variables are selected into the regression equation, and all significant independent variables are eliminated from the regression equation. Both introduction and removal are based on a given level of significance. On the basis of economic development, maintaining stable and sustained growth in fiscal revenue has always been one of the main goals pursued by the government [4].

The traditional quantitative forecast of regional economy is based on econometric models, and simultaneous equations are established based on macroeconomic theory to describe the internal connections between economic variables. The advantage of this method is that the model is based on economic theory and the parameters have clear economic meanings. At present, most of the regional economic forecasting models constructed in the literature adopt this method. However, the econometric model has a large number of equations, with 10–30 equations at every turn, and the error of parameter estimation will mutually affect the final forecasting effect. Econometric models did not perform well in the stagflation period at the end of the 1970s, and modern time series analysis based on the ARIMA model gradually emerged. Sargent and Sim proposed the VAR model and extended it to multivariate forecasting. Feldkircher et al. [5] pointed out that in terms of regional economic forecasting, the VAR model is theoretically better than the econometric model, because the econometric model requires data on all variables, but there is almost no data on interregional trade and investment, while the VAR model does not require all. From the perspective of practical application, the VAR model has achieved good results in foreign regional economic forecasting. However, as the number of variables increases, the parameters in the VAR model increase rapidly so that the degrees of freedom are consumed too quickly. Considering that the economic structure has undergone tremendous changes before and after China's reform and opening up, most of the regional economic data began in 1978, and the data is relatively small. The VAR model cannot guarantee the accuracy of parameter estimation and prediction. The Bayesian vector autoregressive model [6] uses the statistical properties of variables as the prior information of the VAR model parameters to overcome the overparameterization defect of the VAR model to a certain extent. From a theoretical point of view, the BVAR model has certain advantages in China's regional economic forecasting. The initial application of the BVAR model in foreign regional economic forecasts has achieved some results. In domestic research, Rickman [7] research shows that the prediction error of regional BVAR model is significantly smaller than other models. It should be pointed out that most of the regional prediction model literature lacks the "real" sense of out-of-sample prediction error evaluation research. In these literatures, most of the known time series data are divided into two parts. The data in the previous period is called "in-sample" data, the data in the latter period is called "out-of-sample" data, and the in-sample data is used to build the model and get the forecast value of the latter period, compared with the actual value of the latter period to get the so-called "out-of-sample" forecast accuracy. Obviously, this is not a prediction in the true sense,

but an approximation of the actual prediction behavior, so it is also called "pseudo-out-of-sample" prediction. The lack of true out-of-sample prediction error evaluation makes it impossible for us to accurately measure the pros and cons of a prediction model.

Based on this, this article will evaluate the actual forecasting ability of the BVAR model for the regional economy and analyze the main factors affecting the forecast accuracy of the BVAR model and the direction of efforts to improve the forecast accuracy of the BVAR model. A BVAR model was established to illustrate its application in regional economic forecasting and was compared with other forecasting models including ARIMA to evaluate the forecasting effect of the BVAR model. The comparison of forecast errors confirms the feasibility of the BVAR model in regional economic forecasting. The content of this article is divided into the following four parts. First, it briefly introduces the BVAR model; then, it explains the data sources of modeling, the selection of variables, and the evaluation methods of forecasting effects; third, it makes a comparison of the forecasting effects of various models and the BVAR model on economic growth in 2020 predicted value; and finally the conclusion of this article is given.

## 2. Related Work

Sims pioneered the vector autoregressive model (VAR) to explore the impact of temporary economic policies on economic development and thus won the 2011 Nobel Prize in Economics [8]. Because different macroeconomic data is different in statistical difficulty, statistical cost, etc., different macroeconomic data will also have different statistical frequency. For example, China's GDP is counted quarterly, while it is generally counted monthly. Traditional VAR models often use low-frequency data for macroeconomic research, mainly annual or quarterly data. For monthly data, they often use the method of directly adding quarterly data and then use the same frequency VAR model to perform macroeconomic indicators [9–13]. This simple summation of monthly data leads to the loss of high-frequency data information. In order to make full use of high-frequency monthly data and information to make the forecast and analysis of national macroeconomic indicators more accurate, this article does not simply make high-frequency monthly data. Instead of using all the low-frequency and high-frequency data completely, the low-frequency data is regarded as the high-frequency data with missing intervals, and the state space model is established. The low-frequency quarterly data and the high-frequency monthly data that can be actually observed in each period are used as observed variables, the potential monthly data of each variable is used as the state variable, the mixed frequency VAR model is used as the state equation of the state space model, and the corresponding relationship between the state variable and the observed variable is used as the measurement equation of the state space model.

Many scholars have found in their research that although the VAR model is concise and clear and is widely used in the mid- and short-term macroeconomic forecasts, the excessive

parameters of the VAR model have become one of its main shortcomings [14–16]. For an ordinary VAR model, assuming that the model has a  $p$ -order lag and has  $m$  endogenous variables, the number of parameters that need to be estimated is  $m(m + p + 1)$ , even if there are only 5 endogenous variables. The model also has 125 parameters in total. For macroeconomic data, even with 40 years of quarterly data, there are a total of 160 quarterly data samples, and each parameter contains only 1.25 samples on average. Too many model parameters and too few samples will lead to overfitting and failure of the model. For macroeconomic research, since the number of macroeconomic data samples is certain, the only way to solve this problem is to reduce the model parameters proceeds.

It can be seen from the above formula that there are three conventional methods to reduce the parameters of the VAR model: one is to reduce the number  $m$  of endogenous variables in the model, the other is to reduce the lag order  $p$ , and the third is to directly set some parameters to zero. The above approach can improve the fit level of the training set, but there is no way to improve the accuracy of the estimation of the out-of-sample test set, and the model may lose its economic theoretical foundation due to model changes. In order to overcome the problems caused by too many parameters of the VAR model, the paper adopts the Bayesian estimation method based on Minnesota prior to obtain the posterior distribution of each parameter of the VAR model and uses the method based on Kalman filter and Kalman smoothing to obtain the potential [17]. Then according to the posterior distribution of the VAR model parameters and the posterior distribution of the potential state variables, the Gibbs sampling method is used to obtain the estimation of the VAR model and state variables, so as to predict and analyze the national macroeconomic data.

Many scholars have discovered in their research that although the VAR model is concise and clear and is widely used in macroeconomic short- and medium-term forecasts, the excessive parameters of the VAR model have become one of its main shortcomings. As pointed out above, there are three conventional methods to reduce the VAR model parameters: one is to reduce the number  $m$  of endogenous variables in the model, the other is to reduce the lag order, and the third is to directly set some parameters to zero. Although the above approach can improve the fitting level of the training set, there is no way to improve the accuracy of the estimation of the out-of-sample test set, and the model may lose its economic theoretical foundation due to model changes. In order to solve the problem of too many parameters in the VAR model, Litterman introduced the Bayesian method into the estimation of the parameters of the VAR model, assuming that the parameters of the VAR model obey the Minnesota prior distribution and use the BVAR model to predict the macroeconomic indicators of Minnesota. The forecasting model has achieved excellent forecasting results. This pioneering research has provided the basis for the widespread application of Bayesian methods in the future. For example, Dua [18] developed a number of time series models for predicting Irish inflation. The research results show that the application of Bayesian vector

autoregressive method makes the inflation prediction more accurate. Qiu et al. [19] constructed a Bayesian vector autoregressive model optimized by the posterior information criterion (PIC), in which hyperparameters are determined in the same way as the lag length and trend order. The paper uses historical data and Monte Carlo simulation, which evaluates the performance of the selected model through one-step prediction in advance. The research results show that, compared with the ordinary VAR model, the BVAR model has superior performance in forecasting. Laouan [20] applied the BVAR method to the analysis of the factors that cause inflation in our country. The study found that currency oversupply was the main reason for the recent round of inflation, and the influence of external shocks did not dominate. Korobilis [21] pointed out that due to the volatile economic environment, economic forecasting faces a modeling problem with a small amount of data. Bayesian methods have obvious advantages in modeling small sample data. Vosseler and Weber [22] used the BVAR model based on the independent Minnesota-Wishart conjugate prior distribution to predict the yield of interbank treasury bonds through Gibbs sampling.

### 3. Bayesian Vector Autoregressive Model

*3.1. Bayesian Method.* Bayesian reasoning was put forward by the British scholar Thomas Bayes in the 18th century. It describes the relationship between the conditional probabilities of two random events. Suppose there are random event  $x$  and random event  $y$ ,  $P(x)$  represents the probability of event  $x$ , also called the prior probability of event  $x$ .  $P(y)$  represents the probability of event  $B$  occurring, also called the prior probability of event  $y$ .  $P(x|y)$  represents the probability of event  $x$  occurring under the condition that event  $y$  occurs, also called the posterior probability of event  $x$ . Similarly,  $P(y|x)$  represents the probability of event  $y$  occurring under the condition that event  $x$  occurs. It is also called the posterior probability of event  $y$ . We have the following relationship:

$$P\left(\frac{x}{y}\right) = \frac{P(y)P(y|x)}{P(x)}. \quad (1)$$

The above formula is our famous Bayes theorem. Sometimes we also call  $P(y|x)/P(y)$  the standard likelihood, so Bayes' theorem can also be expressed as

$$P_p = L_l \times p_p. \quad (2)$$

The above is a brief introduction to Bayes' theorem.  $P_p$  is the posterior probability,  $L_l$  is the likelihood function, and  $p_p$  is the prior probability. Let us return to the subject of our paper. In their research, the majority of scholars have found that although the VAR model is concise and clear and is widely used in macroeconomic mid- and short-term forecasts, the VAR model having too many parameters becomes one of its main disadvantages. In order to overcome the problems caused by too many parameters of the VAR model, the paper adopts the Bayesian estimation method based on Minnesota prior to obtain the posterior distribution of each

parameter of the VAR model and uses the method based on Kalman filter and Kalman smoothing to obtain the potential. Then according to the posterior distribution of the VAR model parameters and the posterior distribution of the potential state variables, the Gibbs sampling method is used to obtain the estimation of the VAR model and state variables, so as to predict and analyze the national macro-economic data.

**3.2. Vector Autoregressive Model.** Different from the classical estimation method, the basic idea of the Bayesian estimation method is treat the parameters of the model to be estimated as random variables and obey a certain distribution, and then give the prior distribution of the parameters to be estimated based on experience and put it combined with the sample information. Bayes' theorem can be used to calculate the posterior distribution of the parameter to be estimated, thereby obtaining the estimated value with the estimated parameter. In this paper, the parameters to be estimated are obtained by Gibbs Sampler. Pfarrhofer and Piribauer [23] applied Bayesian Gibbs Sampling (BGS) to the estimation of the parameters of the state space model. The results show that the estimation results of the Gibbs sampling on the state space parameter values and state vectors differ from the true values. It is not big and is obviously better than the estimation result based on Kalman filter method. The results show that the BVAR model has poor short-term effects, but medium- and long-term prediction effects significantly better than other models. The vector autoregressive model construction process is shown in Figure 1.

**3.3. Basic Principles of BVAR Model.** The BVAR model is developed on the basis of the ordinary VAR model. Consider the unconstrained VAR model:

$$y_t = \lambda X_t + \sum_{i=1}^k \gamma_i y_{t-i} + \varepsilon_t. \quad (3)$$

Among them,  $t$  represents the time and  $k$  represents the lag order.  $y_t$  is the value of the  $K$ -dimensional random vector  $y$  at time  $t$  and is the endogenous variable in the model.  $X_t$  is a block diagonal matrix of  $K \times Kn$  order,  $x_t = x_t \otimes I_k$ . Among them  $x_t$  is an  $n$ -dimensional vector, which is the deterministic part of every equation, such as constant terms or exogenous variables.

$\lambda$  is the  $K$ -dimensional coefficient vector  $X_t$ .  $\gamma_i$  is the coefficient of the  $K$ -dimensional vector  $y_{t-i}$  and the square matrix of the  $K$ -dimensional parameter.

$\varepsilon_t$  is the  $K$ -dimensional random error term. Assume that  $\text{Exp}(\varepsilon_t)$  is related in the same period, but not serially related:

$$\text{Exp}(\varepsilon_t \varepsilon_t') = \begin{cases} 0, & t = i, \\ \Sigma, & t \neq i. \end{cases} \quad (4)$$

It can be seen from the VAR model formula (4) that there are a total of parameters to be estimated.  $\Sigma$  is the variance of  $K$ -dimensional random error term. Even if there are fewer variables to be predicted, considerable

observations are needed. Therefore, unless there is enough data, in general, VAR models are faced with the problem of fewer degrees of freedom and therefore low prediction accuracy. By imposing certain constraints on the parameters, such as reducing the lag order or removing some variables in individual equations, the problem of too little freedom can be effectively alleviated. But from the Bayesian point of view, this means that the predictor believes that the probability that the coefficient of the removed lag term is 0 is 100%. Unfortunately, it is impossible to know whether this constraint holds. The BVAR model creatively combines the predictor's prior information on the island with the above unconstrained VAR model to solve the above problems. Because this method was first proposed at the University of Minnesota and the Minnesota branch of the Federal Reserve, it is also called the Minnesota Prior.

**3.4. Minnesota Prior.** In Bayesian theory, when the posterior probability distribution and the prior probability distribution belong to the same type of distribution, we call it conjugate prior distribution. When we assume that the parameters are conjugate prior distributions, we can obtain the parameters. The advantage of assuming a conjugate prior distribution is that it can greatly reduce the amount of our calculations. A large number of studies have shown that this assumption is reliable in many cases. The Minnesota prior model is one of the conjugate prior models. When the posterior distribution and the prior distribution do not belong to the same type of distribution, we cannot obtain the analytical value of the parameter. We need to use the MCMC method to simulate the best value of the parameter. This requires a great amount of calculation, but this model is more flexible; for example, we can allow the parameters of the VAR model to change over time. This article uses the Minnesota prior distribution. It is mainly used to solve the problem of too many parameters in the VAR model under the conjugate prior distribution and improve the prediction accuracy of the model. For a  $k$ -order lag, a VAR model with  $n$  endogenous variables and no constant term, the general form is

$$y_i(t) = \sum_{i=1}^n \sum_{j=1}^k \alpha_{ijt} y_{i,t-j} + \beta_{it}. \quad (5)$$

If the random parameters  $\alpha_{ijt}$  obey the normal distribution, then the number of hyperparameters that needs to be determined at this time is at least  $2m^2$ , of which the prior mean is  $m^2$ , and the prior variance is  $m^2$ . The Minnesota prior distribution reduces the number of hyperparameters to 5. One or three, the method used in this article is based on the research of Negro and Schorfheide, using 5 hyperparameters. The basic assumptions of Minnesota prior model include the following aspects:

- (1) Normality: the random disturbance vector obeys the multivariate normal distribution.
- (2) Independence: the covariance matrix  $\Sigma$  and the model coefficient are independent of each other.

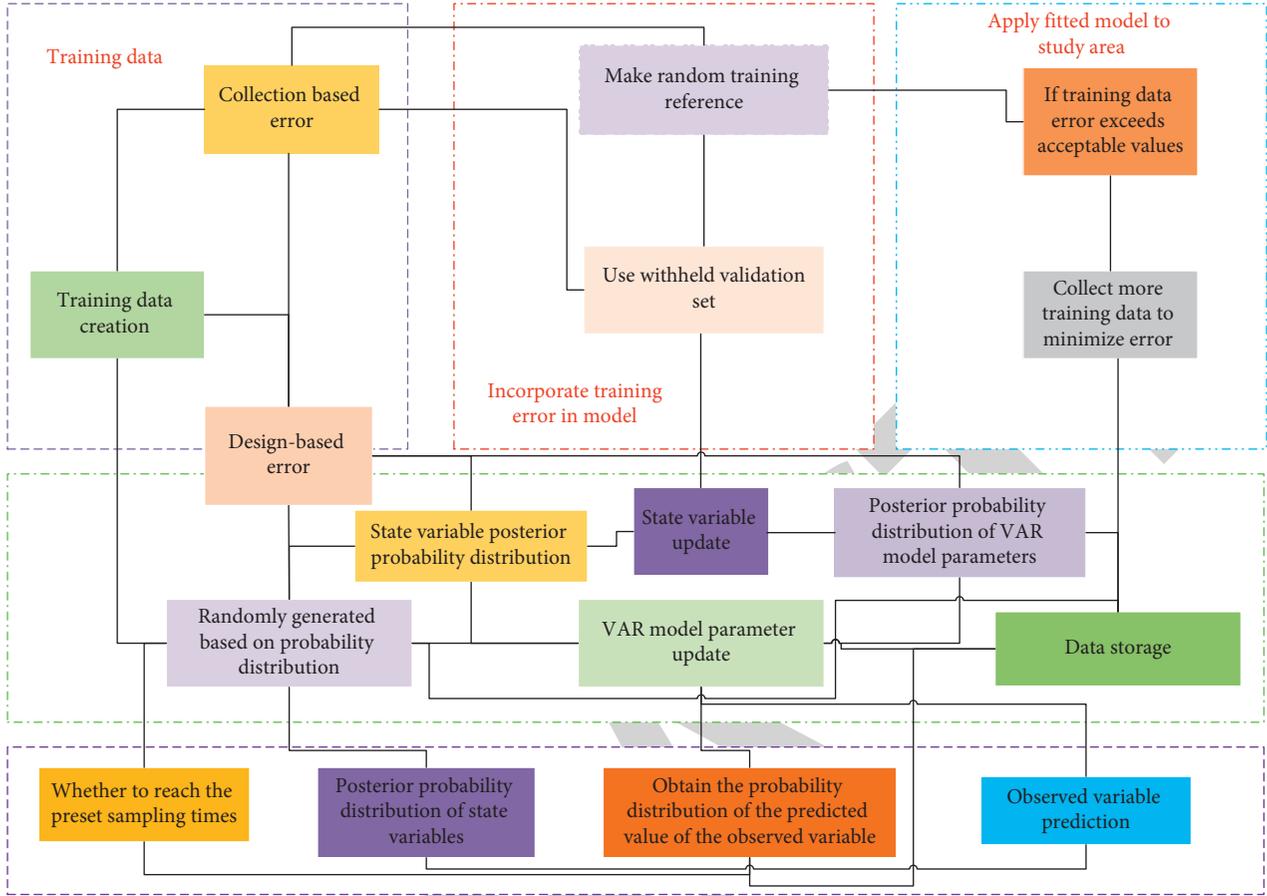


FIGURE 1: Flow chart of measurement model construction.

- (3) The prior distribution of the covariance matrix  $\Sigma$  is taken as the diffusion distribution.
- (4) Model coefficients are mutually independent and obey normal distribution.
- (5) The prior variance is determined by 5 hyperparameters:  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ , and  $\alpha_5$ , which respectively represent overall tightness, lagging delay function attenuation factor, innovation covariance matrix dispersion  $i$ -th equation variable  $x_i$ , and the aggregation degree of the innovation variance matrix. The lag term coefficient sum is the credibility of a fixed value, relative tightness.

The estimation of BVAR model can use the mixed estimation method proposed by Yang and Zhang [24]. Estimating the BVAR model requires the predictor to determine the value of the above hyperparameters. Because the main purpose of the BVAR model is prediction, so, unlike other models, the value standard of hyperparameters is to obtain the optimal prediction effect, rather than relying on various model settings and tests. The determination of hyperparameters is actually a process similar to raster search, searching for the value that can obtain the best prediction effect within the range of hyperparameters. For this reason, the total sample  $T$  obtained is usually divided into two periods  $T$  and  $T - T_0$ . The data of period  $T$  is used to estimate

the BVAR model and forecast, and the data of  $T - T_0$  is used to calculate and compare the forecast error and determine the final hyperparameter value.

#### 4. Regional Economic Bayes Vector Autoregressive Model Operation

**4.1. Data Description.** When we set out to construct a Bayesian vector autoregressive model of the regional economy, we use the annual import and export price index calculated by the World Bank to perform a comparable conversion of the quarterly import and export sequence and finally convert it into RMB units according to the official exchange rate. In the latter stage of the model, we extend the data length to the second quarter of 2021. Then, we use the model to forecast the interval from the first quarter of 2020 to the first quarter of 2021 and compare the forecast results with the actual situation, so that we could analyze the model's forecast for external samples.

Given the posterior distribution and initial value, the  $i$ -th cycle of Gibbs sampling is as follows. The specific process can be understood in conjunction with the flow chart, as shown in Figure 2 of the previous measurement model.

- (1) Update state variables: According to the VAR model parameters, use Kalman filtering and Kalman

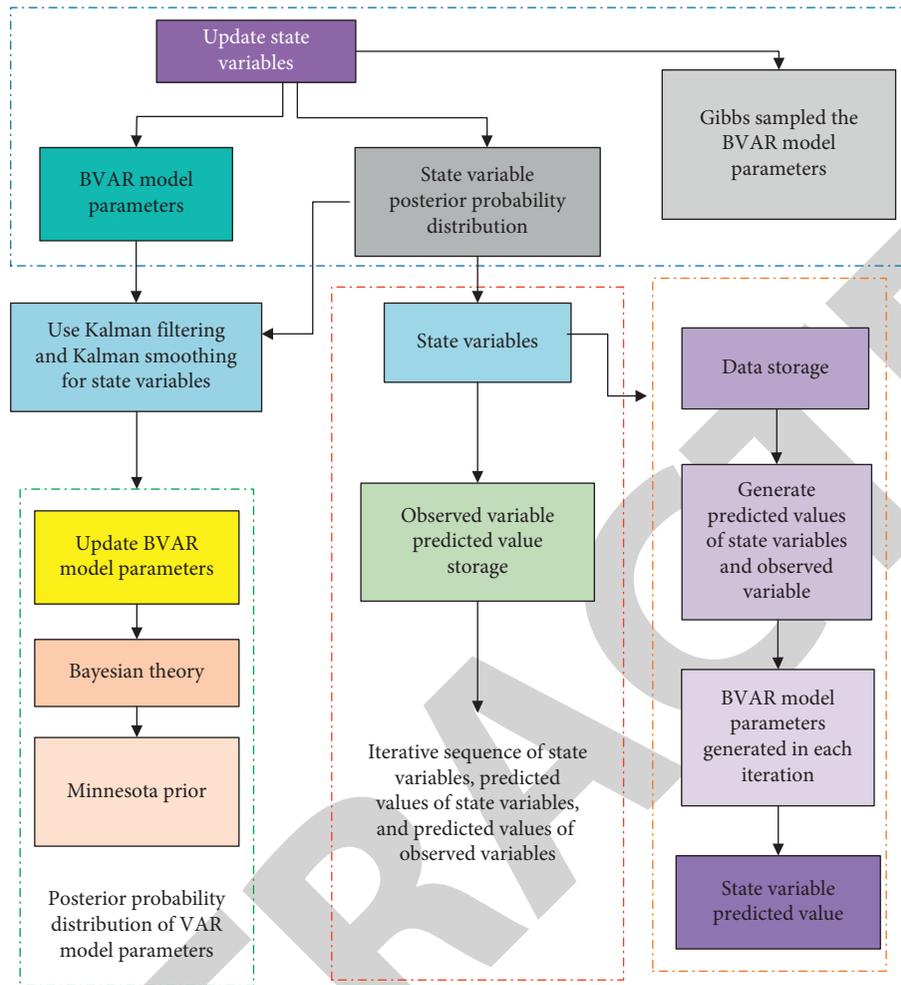


FIGURE 2: Gibbs sampling flow chart.

smoothing on state variable to obtain the posterior probability distribution of state variables.

- (2) Update VAR model parameters: According to Bayesian theory and Minnesota priors, obtain the posterior probability distribution of VAR model parameters.
- (3) Data storage: Generate predicted values of state variables and observed variables, and store the VAR model parameters, state variables, predicted values of state variables, and predicted values of observed variables generated in each iteration.
- (4) Enter the next cycle: So far, we get the iterative sequence of VAR model parameters, state variables, predicted values of state variables, and predicted values of observed variables through Gibbs sampling.

**4.2. Sample Model Run Results.** Figure 3 shows the various statistical measurement results of the prediction accuracy calculated by the above recursive method from the first quarter of 2010 to the fourth quarter of 2019. The forecast statistics results are arranged in the order of one to five (quarterly) forecasting plans for different variables.

**4.3. Comparison of Prediction Results of Various Regional Economic Prediction Models.** In order to illustrate the predictive effect of the BVAR model, Figure 4 compares the Theil U statistics of the intrasample predictions of the three models of BAVR, VAR, and ARIMA from 2014 to 2019. The VAR model selects the lag order of the endogenous variable as the 2nd order according to the Akaike Information Criterion (AIC) and Schwartz Information Criterion (BIC). The lag order of the exogenous variable is the same as that of the BVAR model, which is the same time and 1 time lag. The ARIMA model is based on the autocorrelation graph and partial correlation graph of regional GDP, the coefficient value, residual test, and AIC value are determined as the ARIMA process. From the Theil U statistics in Figure 4, it can be seen that the BVAR model has the best predictive effect, and the prediction error is the smallest from 1 step ahead to 5 steps ahead. Moreover, similar to the prediction results of Litterman, the longer the prediction period of the BVAR model, the better the prediction effect relative to the VAR and ARIMA models. The prediction effects of VAR model and ARIMA model are not very satisfactory. Similar to the BVAR model, the longer the prediction period of the VAR model, the higher the prediction accuracy relative to

	Average error	absolute average error	standard error	Theil U
1	0.012	0.05	0.053	0.201
2	0.012	0.022	0.028	0.112
3	0.012	0.053	0.061	0.197
4	0.014	0.034	0.043	0.364
5	0.015	0.062	0.084	0.241

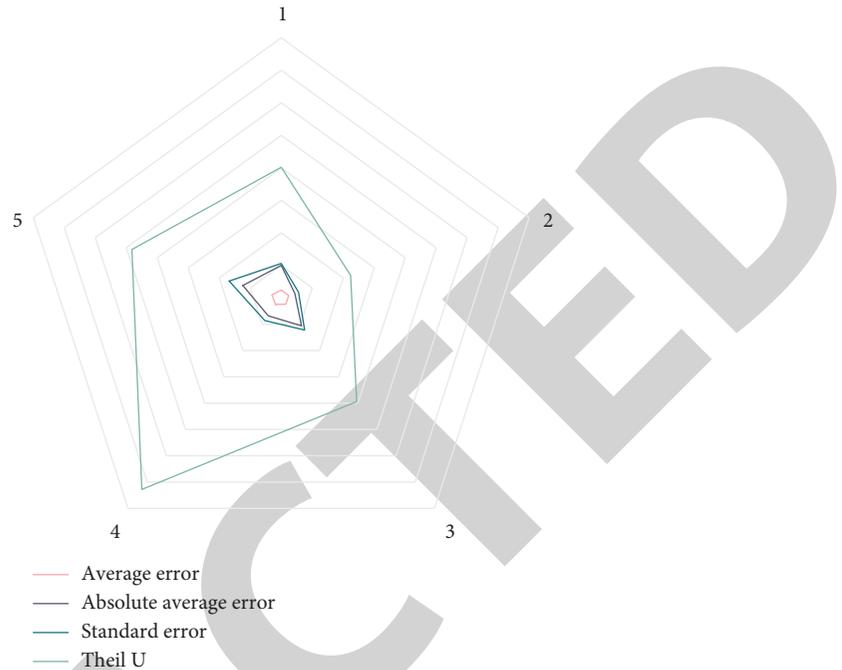


FIGURE 3: Sample prediction statistics results.

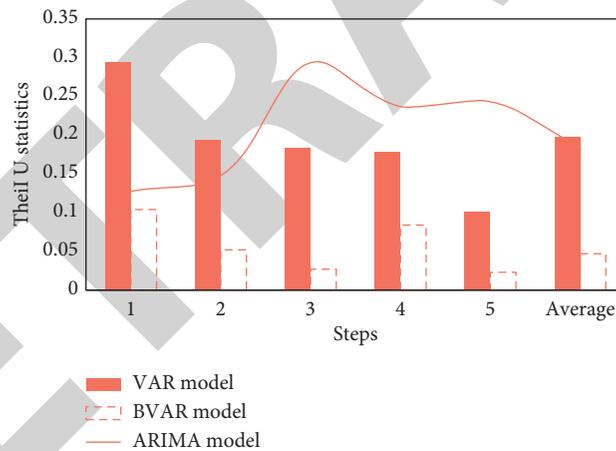


FIGURE 4: Theil's U statistics of in-sample predicted values of each model from 2014 to 2019.

the ARIMA model. This feature of BVAR and VAR models may be due to the consideration of the relationship between economic variables, while ARIMA is a univariate model. However, the ARIMA model is far better than the VAR model by one step ahead of prediction, which may be caused by the VAR model's too little freedom.

Figure 5 further compares the out-of-sample prediction effects of several models. The difference with in-sample prediction is that the value of exogenous variables of the same period in out-of-sample prediction uses its predicted value to simulate the real prediction process. The two exogenous variables, national GDP and central government transfer payment (local fiscal revenue and expenditure balance), are predicted by a univariate ARIMA model.

According to their autocorrelation graph and partial correlation graph, the coefficient value, residual test, and AIC value are determined as ARIMA processes, respectively. The Provincial Academy of Social Sciences and the Information Office of the National Development and Reform Commission regularly release the forecast results of the next year's macroeconomic indicators every year. Their forecast models mainly rely on the multiplier model of investment data. In order to facilitate the comparison with their forecast results, Figure 5 only compares the advanced forecast errors from 2015 to 2019 and converts the horizontal forecast values into growth rates. It can be seen from Figure 5 that the BVAR model has the smallest prediction error among all models. Except for the larger prediction error in 2015, the prediction

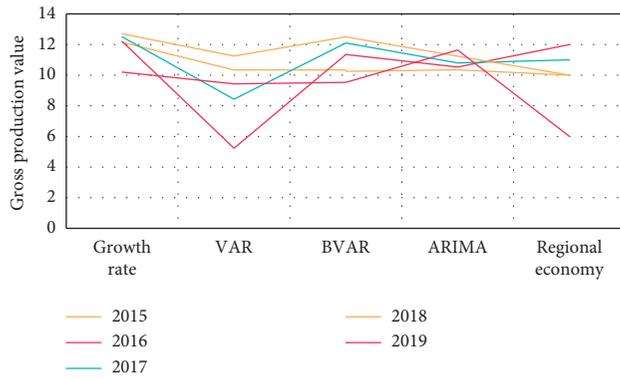


FIGURE 5: The gross production value of each model from 2015 to 2019.

in other years is more accurate. The prediction effect of the ARIMA model is second only to the BAVR model among all models and is better than the predictions of the regional academy of social sciences and the National Development and Reform Commission.

The VAR model has the largest prediction error, which may be caused by too small freedom. Moreover, the BVAR model is more accurate in predicting the turning point of economic growth in 2019. Comparing with the Theil U statistic of the in-sample forecast value of one step ahead in Figure 5, it can be found that the out-of-sample forecast error of the BVAR model is smaller than the in-sample forecast error, which may imply the expectation of the national economic growth situation and the transfer payment of the central government will affect the actual regional economic growth process. It can be seen from the comparison of the prediction effects of various models that the BVAR regional economic prediction model is better than other prediction models.

**4.4. Analysis of Actual Errors in Forecasting Regional Economic Development.** We first evaluate the forecast errors of the BVAR model for major economic indicators. This region is chosen as an example because the regional information center will release the forecast values of the main economic indicators for the next year every year, which is convenient for us to make detailed out-of-sample forecast errors. The BVAR model gives the forecast value of the growth rate (at comparable prices), while the regional information center gives the forecast data of the nominal value (absolute amount). In order to be able to directly compare, we calculated the 2020 value of the two forecasting methods: the predicted value of the absolute amount of change and the predicted value of the growth rate (at comparable prices). The forecast errors of the two methods are shown in Figure 6.

The forecast error of the absolute constant price is measured by the average absolute percentage error (APE) and the Theil U statistic, and the forecast error of the growth rate is measured by the difference of the average growth rate. It should be pointed out that BVAR gives a forecast of 3 years ahead, while the regional information center forecasts a year

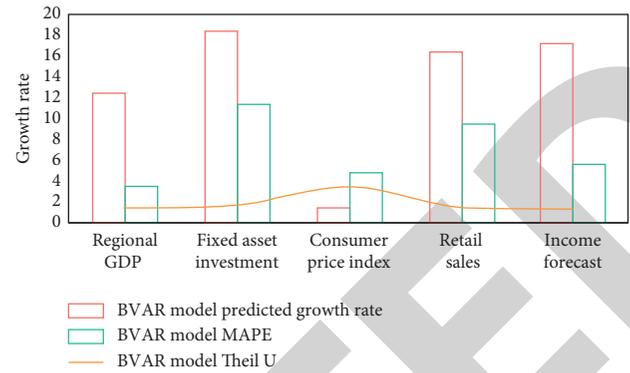


FIGURE 6: Predicted values and actual values of BVAR model and regional information center in 2020.

ahead. In theory, the average forecast error of the former is greater than the latter.

It can be seen from Figure 7 that the BVAR model predicts that the average growth rate of regional GDP from 2018 to 2019 is 11.35%, which is 1.14 percentage points lower than the actual value, which is slightly better than the prediction of the regional information center and MAPE. The forecast errors measured by  $U$  statistics also indicate that the forecast error of the BVAR model is smaller in the forecast of regional gross product. However, the BVAR model has large errors in forecasting regional fixed asset investment, consumer price index, and retail sales of consumer goods. After observing the data in detail, we found that it may be caused by the four-trillion economic stimulus plan introduced in 2019. Starting from 2019, the regional fixed asset investment has suddenly increased sharply and its average annual growth rate from 2017 to 2019 only 10%. Because it is different from the regional information center, the BVAR model is forecasted three years in advance, and the model fails to adjust the forecast value of fixed asset investment in time according to the latest data, which causes the BVAR model to have a large forecast error. The changes in the actual value of the regional CPI are very close to the changes in the national CPI. We believe that the reason for the large error of the BVAR model to predict regional CPI is that the linkage relationship between regional CPI and national CPI was not considered at the beginning of the model setting, which is a congenital deficiency in the model setting. As for the BVAR model overestimating the retail sales of consumer goods in the region, we believe that the retail sales of consumer goods depend to a large extent on the prices of consumer goods because the BVAR model underestimates the consumer price index and therefore will overestimate the retail sales of consumer goods. The BVAR model and the regional information center have their own advantages in predicting the general budget revenue of regional local finance. The BVAR model has a lower MAPE value, while the regional information center Theil's  $U$  index is lower, and the average growth rate prediction error is smaller. It can be seen that, except for the regional GDP and local fiscal revenue, the BVAR model has large forecast errors in other indicators, which is obviously not as accurate as the regional information center forecast. However, if it is

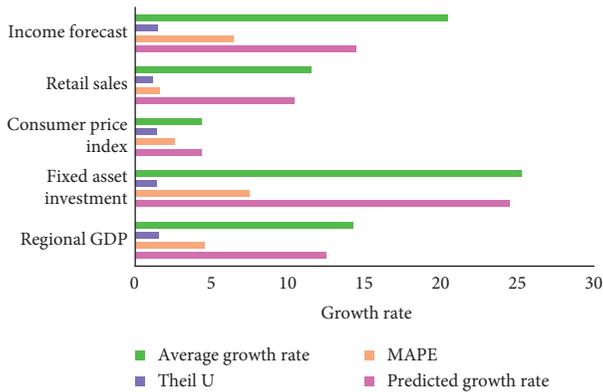


FIGURE 7: Actual values of BVAR model and regional information center in 2020.

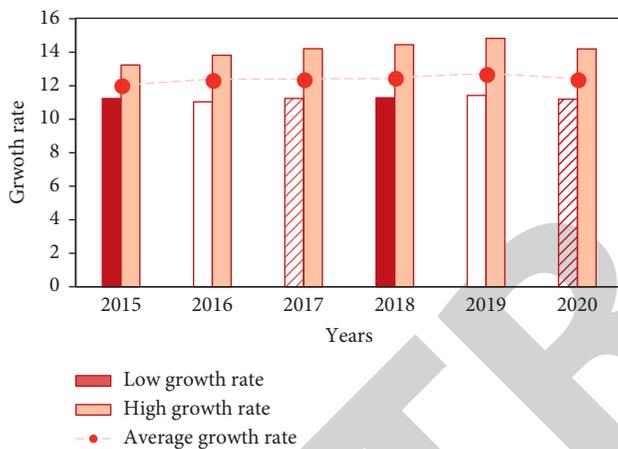


FIGURE 8: Regional GDP growth rate in each quarter in 2020 BVAR model forecast.

considered that the BVAR model is a three-year forecast rather than a one-year forecast and has a clear advantage in the forecast of regional GDP, the forecast error of the BVAR model for regional economic indicators is within an acceptable range.

**4.5. Forecast Results of BVAR Model.** In Figure 8, the predicted value of the regional economic growth rate from 2010 to 2015 is given based on the BVAR model. According to the values of the two exogenous variables, the following three situations are considered:

- (1) The predicted value of national GDP is predicted by the univariate ARIMA process (average growth rate is 9.31%); government transfer payments have grown at an average growth rate of 22.53% in the past five years.
- (2) National GDP has grown at a relatively low rate of 8.21%, and central government transfer payments have grown at an average growth rate of 22.52% in the past five years.

- (3) National GDP has increased by 10.42%, and the central government transfer payment increased at a relatively high rate of 26%. These three situations can be regarded as moderate, sluggish, and high external economic environments, respectively. It can be seen from Figure 8 that the central transfer payment plays an important role in regional economic growth. Even if the national economic growth situation is not very optimistic, as long as the central transfer payment maintains a high growth rate, the regional economy can also achieve faster growth. Figure 8 shows the BVAR model forecast value of growth rate of regional GDP from 2015 to 2020.

## 5. Conclusion

Conventional regional econometric forecasting models and modern time series analysis and forecasting models such as ARIMA and VAR require more observations, and it is difficult to obtain the expected predictive effect in China's regional economic forecasting. The BVAR model developed on the basis of the VAR model not only has the advantages of the VAR model, but also overcomes the problem of excessive consumption of the VAR model's freedom. This paper takes the region as an example and establishes a BVAR model to illustrate the application of this model in regional economic forecasting. The comparison of the prediction errors of various models shows that, regardless of whether the BVAR model is in-sample or out-of-sample, the prediction errors are significantly smaller than the actual prediction models of VAR, ARIMA models, and regional provincial academies of social sciences and development and reform commissions. The BVAR model is also more accurate in predicting the turning point of the regional economic growth in 2020. This shows that the application of the BVAR model to regional economic forecasting can achieve better forecasting results. The modeling process of regional econometric models is complicated and the workload is heavy. Even the simpler quantitative forecasting models of the regional provincial academies of social sciences and the National Development and Reform Commission need to collect a lot of investment data, and some data are not public, and the forecasting process relies more on the empirical judgment of experts. The modeling process of the BVAR model is relatively simple, and there are clear standards for the determination of parameters and hyperparameters. Therefore, it has broad application prospects in China's regional economic forecasting.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] A. G. Assaf, G. Li, H. Song, and M. G. Tsionas, "Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model," *Journal of Travel Research*, vol. 58, no. 3, pp. 383–397, 2019.
- [2] R. Gupta, C. K. M. Lau, V. Plakandaras, and W.-K. Wong, "The role of housing sentiment in forecasting U.S. home sales growth: evidence from a Bayesian compressed vector autoregressive model," *Economic Research-Ekonomska Istrazivanja*, vol. 32, no. 1, pp. 2554–2567, 2019.
- [3] J. P. LeSage and D. Hendrikz, "Large Bayesian vector autoregressive forecasting for regions: a comparison of methods based on alternative disturbance structures," *The Annals of Regional Science*, vol. 62, no. 3, pp. 563–599, 2019.
- [4] P. G. Coulombe and M. Göbel, "Arctic amplification of anthropogenic forcing: a vector autoregressive analysis," *Journal of Climate*, vol. 2, no. 1, pp. 1–52, 2021.
- [5] M. Feldkircher, F. Huber, and M. Pfarrhofer, "Factor Augmented vector autoregressions, panel VARs, and global VARs," *Macroeconomic Forecasting in the Era of Big Data*, vol. 2, no. 1, pp. 65–93, 2020.
- [6] A. R. Putri, M. Usman, and E. Virginia, "Application of vector autoregressive with exogenous variable: case study of closing stock price of PT INDF. Tbk and PT ICBP. Tbk," *Journal of Physics: Conference Series. IOP Publishing*, vol. 1751, no. 1, pp. 12–21, 2021.
- [7] D. S. Rickman, "Regional science research and the practice of regional economic forecasting: less is not more," *Regional Research Frontiers*, vol. 1, no. 1, pp. 135–149, 2017.
- [8] A. Anagnostou and P. Gajewski, "Heterogeneous impact of monetary policy on regional economic activity: empirical evidence for Poland," *Emerging Markets Finance and Trade*, vol. 55, no. 8, pp. 1893–1906, 2019.
- [9] M. Feldkircher and N. Hauzenberger, "How useful are time-varying parameter models for forecasting economic growth in CESEE?" *Focus on European Economic Integration Q*, vol. 1, pp. 29–48, 2019.
- [10] F. Huber and M. Feldkircher, "Adaptive shrinkage in Bayesian vector autoregressive models," *Journal of Business & Economic Statistics*, vol. 37, no. 1, pp. 27–39, 2019.
- [11] M. Chen, S. Lu, and Q. Liu, "Uniform regularity for a Keller-segel-navier-stokes system," *Applied Mathematics Letters*, vol. 107, p. 106476, 2020.
- [12] S. A. Brave, R. A. Butters, and A. Justiniano, "Forecasting economic activity with mixed frequency BVARs," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1692–1707, 2019.
- [13] J. Chai, L.-M. Xing, X.-Y. Zhou, Z. G. Zhang, and J.-X. Li, "Forecasting the WTI crude oil price by a hybrid-refined method," *Energy Economics*, vol. 71, no. 1, pp. 114–127, 2018.
- [14] A. Zollanvari and E. R. Dougherty, "Optimal bayesian classification with vector autoregressive data dependency," *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3073–3086, 2019.
- [15] P. I. Rahayu, A. Famalika, and P. R. Sihombing, "Penerapan model vector autoregressive (var (2)) pada data inflasi di provinsi jawa timur dan bali," *Bayesian Analysis*, vol. 1, no. 1, pp. 55–66, 2021.
- [16] M. Balcilar, R. Gupta, and C. Jooste, "Long memory, economic policy uncertainty and forecasting US inflation: a Bayesian VARFIMA approach," *Applied Economics*, vol. 49, no. 11, pp. 1047–1054, 2017.
- [17] A. Madhou, T. Sewak, I. Moosa, and V. Ramiah, "GDP nowcasting: application and constraints in a small open developing economy," *Applied Economics*, vol. 49, no. 38, pp. 3880–3890, 2017.
- [18] P. Dua, "Macroeconomic modelling and bayesian methods," *Journal of Quantitative Economics*, vol. 15, no. 2, pp. 209–226, 2017.
- [19] J. Qiu, H. Wang, and L. Hu, "Spatial transmission network construction of influenza-like illness using dynamic Bayesian network and vector-autoregressive moving average model," *BMC Infectious Diseases*, vol. 21, no. 1, pp. 1–9, 2021.
- [20] A. M. Laouan, "Developed countries agricultural subsidies and cotton production in Burkina Faso: an analysis using the vector autoregressive model (VAR)," *Modern Economy*, vol. 12, no. 01, pp. 46–68, 2021.
- [21] D. Korobilis, "Quantile regression forecasts of inflation under model uncertainty," *International Journal of Forecasting*, vol. 33, no. 1, pp. 11–20, 2017.
- [22] A. Vosseler and E. Weber, "Forecasting seasonal time series data: a Bayesian model averaging approach," *Computational Statistics*, vol. 33, no. 4, pp. 1733–1765, 2018.
- [23] M. Pfarrhofer and P. Piribauer, "Flexible shrinkage in high-dimensional Bayesian spatial autoregressive models," *Spatial Statistics*, vol. 29, no. 9, pp. 109–128, 2019.
- [24] Y. Yang and H. Zhang, "Spatial-temporal forecasting of tourism demand," *Annals of Tourism Research*, vol. 75, no. 5, pp. 106–119, 2019.