3 A NOVEL SPATIAL MIXED FREQUENCY FORECASTING MODEL WITH APPLICATION TO CHINESE REGIONAL GDP

Xianning WANG[1*]
Jingrong DONG[2]
Zhi XIAO[3]
Guanjie HE[4]

Abstract

Direct use of economic indicators for different frequencies is important to improve the regional forecast performance, and is the quantitative basis for improving the awareness on regional economic cycle changes, growth drivers and regional differences Considering the spatial mixed frequency data using a high frequency variable to predict a low frequency one in regional prediction problems, this paper proposes a novel spatial mixed frequency forecasting model. Firstly, it analyzes the commonly used spatial forecasting models and the most classical MIDAS. Secondly, it adopts the soft spatial weights to describe the spatial correlation of economic variables to amend the polynomial weighting method of MIDAS. Thirdly, it analyzes the main characteristics and puts forward the prediction error or precision index to test the validity of the model. Finally, it applies the new method to forecast the real GDP growth rate in 30 provinces and autonomous regions in China and compares the different weightings, which show a great feasibility. Futher, it discusses and gets some help findings about the characteristics of parameters such as weighs polynomials, prediction weighs, and lag period in different regions. Finally, it implements the Diebold Mariano tests for RMSEs between different model settings and obtains meaningful conclusions.

Keyword: Spatial mixed frequency; Forecastingl; MIDAS; Chinese regional GDP
JEL Classification: E17, C31, C53

[1] *School of Economics and Management, Chongqing Normal University, Chongqing, China. xianningwang@sina.com. Big Data Marketing Research and Applications Center of Chongqing Normal University Chongqing Normal University, Chongqing, China.*
* *For their contributions, the authors thank to Pro.Bo Zhong, Dr.Xiaodong Feng and Wenjing Li, and the support of National Natural Science Foundation of China (71671019), National Social Science Fund Key Project of China, 14AJL015 and Chongqing Basic Research and Frontier Exploration Project in 2019 (Chongqing Natural Science Foundation: cstc2019jcyj-msxm1001): Research on the Spatial Mixed frequency Forecasting Model and Method for Open Network Data.*
2 School of Economics and Management, Chongqing Normal University, Chongqing, China.
3 School of Economics and Business Administration, Chongqing University, Chongqing, China.
4 Chongqing Rural Commercial Bank-Bishan Branch, Chongqing, China.

I. Introduction

Considering the short-term economic fluctuations and long-term trends contained in the indicators simultaneously will help to improve the regional forecasts' performance, and it forms the quantitative basis for improving the awareness on regional economic cycle changes, growth drivers and regional differences. Direct use of economic indicators for different frequencies is the focus of research and hotspots. However, the existing prediction methods are limited, due to their inability to solve the two problems of spatial correlation and mixing data in the regional economy. Spatial forecasting has greatly improved the forecasting accuracy in space-related issues fully considers the spatial correlation between certain economic variables in a certain geographical scope, which. However, that high-frequency data and low-frequency data emerge at the same time as single problem is very common and expected to be fully and timely analysed. Data foundation of the spatial forecasting models gradually expands from the same frequency data (the consistent statistical frequency in one model) to the mixed frequency data (the inconsistent statistical frequency in one model). MIDAS does not perform artificially homogenous processing of mixed-frequency data. It uses the original data directly to make full use of the information of the original data to build a mixed-frequency prediction model. But MIDAS does not take into account the spatial correlation hypothesis of spatial economics, by which regional relations between regional economic phenomena cannot be well understood. To tackle this, the current paper tries to put forward a novel spatial mixed-frequency forecasting model to help analysing quantitatively the regional economic cycle changes, growth drivers and regional differences.

Literature review shows that there are many researches related to the forecasting of regional economic growth. Both academic and business communities are trying to grasp future trends for the first time within the rapidly changing market environment. Due to the disclosure system, GDP and its related macroeconomic indicators are mostly end-of-quarter or end-of-month data, appearing earlier in the current year (usually in the corresponding next quarter or monthly). After amendments and announcements are completed, the annual GDP is announced in the following year. However, it is so crucial for consumers and investors to acquire information about the economic situation in the future that they try to use available short-term information to predict the long-term tendency. The regional correlation of GDP and its related macroeconomic indicators emphasized by spatial econometrics makes the above problems even more challenging. Therefore, how to use relatively high-frequency monthly and quarterly data to predict relatively low-frequency annual data, taking into account the spatial effects caused by geography, history, society, economy, population, and other spatial factors, is an urgent issue to be solved in the regional GDP forecasting.

Literature review also shows that there are many researches related to spatial forecasting and mixed-frequency prediction. However, the existing literature that involves spatial mixed-frequency forecasting is rare. In contrast to spatial data with same frequency, spatial mixed-frequency data contains both complete information that reflects the spatial relationship of economic variables and samples with different statistical frequencies. Throughout mixed-frequency data, relatively high frequency data can draw momentary changes, fluctuations and shocks and relatively low frequency data are inclined to reflect the long-term steady trend. Inconsistent data sampling frequency makes it impossible to directly use the prediction model constructed based on the same frequency data. Furthermore, commonly used co-frequency processing (Interpolation or Merge) or abandoning one part will result in a decline in the samples' information quality, which in turn will cause prediction errors. How to solve the mixed data and its spatial correlation, using mixing data to construct an effective spatial forecasting model to predict the future trend is increasingly attracting the attention of the academic community and the industry. This article aims to explore this area.

Our objective is to build a spatial mixed-frequency data prediction model, with the advantage that spatial weights are used to introduce the spatial correlation in spatial forecasting and MIDAS can better handle the mixed frequency data, further applying the new model to forecast the regional GDP in China. More concretely, first, we summarize the modelling idea of MIDAS about how its polynomial function (Beta or Almon) solves the problem of mixed-frequency data, and analyze the basic setting principles of the spatial forecasting model about how to introduce the spatial weights matrices. Second, given frequency inconsistency between the dependent variable and single explanatory variable, and the significantly spatial correlation of explanatory variables, we introduce the soft spatial weights matrix to modify the polynomials in MIDAS, which means the coefficient in the new model is given together by mixed frequency data distributed lag weights and soft spatial weights. The model is built so that high-frequency explanatory series forecast a low-frequency variable. Then, we deeper analyze the main features of the new model, and the proposed errors or accuracy measurements of effectiveness of the model. Third, we apply the new model to the GDP forecasts of 30 provinces and cities in China. Through the forecasting conclusions analysis by RMSE of quarterly GDP, we further confirmed the feasibility of the model.

Empirical study shows a good performance and a great feasibility overall. Some helping findings about the characteristics of parameters such as weighs polynomials, prediction weighs, and lag period in different regions are additionally obtained. In detail, the forecasting effectiveness analysis shows more detailed information about how the model performs, different K and H, with or without autoregression in different regions: (1) The spatial mixed frequency forecasting model with auto-regression is less accurate than that without auto-regression terms; (2) When predicting with autoregression, although the prediction errors are relatively close, the minimum error below K=12 and H=3 shows better than K＝12，H=1 and K=12, H=2; (3) Regional models in each province and city are well-fitted (the greater the maximum likelihood, the better the model fits). They fit better overall with 3 different weight functions; and forecast weight analysis shows that the 3rd-order Almon polynomial function in the Eastern region, the Central region, the Western region, and the Northeast region is relatively stable, and the trend of change is more consistent, with “0” being the center of fluctuation. The main contributions can be summarized as below.

(a) Spatial mixed-frequency forecasting broadens the range of statistical data on which economic forecasting depends, and in particular improves the information efficiency of directly using the mixed-frequency data in one model without prior treatments. The economic significance of prediction is based on historical data and market operating mechanisms to achieve scientific forecasting of the economic effects and development trends that economic or social activities may produce. Economists (such as Parsons, Klein, Leontief, etc.) have been working on short-term, medium-term, and long-term economic cycle forecasts to better understand, circumvent, and mitigate the dangers of economic or financial crises. In addition to mathematical and normative theoretical studies of internal and external mechanisms of the economic system, the statistics (data) and forecasting model building have become extremely important drivers.

(b) Spatial mixed-frequency forecasting integrates the spatial correlation of the theoretical cornerstone of spatial econometrics into the prediction model, taking full account of the regional economic relevance to regional economic development. Regional economics studies the coordination of regional economic development and regional relations from the perspective of economics, considering the optimal allocation and combination of production resources in a certain space (region) to obtain maximum output. The theories such as balanced development developed on the basis of the Harrod-Domar neoclassical economic growth model, the unbalanced development proposed by Hirschman, the gradient transfer proposed by Vernon, and the growth poles first proposed by Francois Peru, center-periphery theory and urban circle economic theory, are aware of the interaction of economic activities or decisions in different regions and try to explain the differences in regional economic growth. Spatial mixed-frequency forecasting is one of the quantitative analysis methods that can be studied above, which takes into account the interaction (such as population mobility, capital cross-regional speculation, technology sharing, policy externalities and so on) and forecasting needs of economic development. The forecasting results derived from spatial mixed-frequency forecasting have great reference value for the effective allocation of regional economic factors.

(c) Spatial mixed-frequency forecasting meets the needs of traditional regional economic forecasting, and in the big data era it also caters to the requirements of economic decision makers for various types of data (especially spatial mixed-frequency data, considering the coexistence of high frequency and low frequency). In the era of big data, data awareness continues to strengthen, and the driving force you need behind economic forecasting is increasingly relying on how to mine previous information value out of various styles (for example, text data, audio data, video data, etc., structured or unstructured data, etc.). Regardless of the types of data, the frequency problem is a fundamental issue. The inspiration of spatial mixed-frequency forecasting is that while attaching importance to the economic mechanism, it attempts to rely on rich statistical evidence to test the theory and gain time and convenience for the accurate and scientific decision-making of economic subjects.

The rest of the paper is organized as follows: Section 2 “Literature Review” introduces the recent research status and achievements of spatial forecasting and mixed-frequency data forecasting comprehensively. Section 3 “Model Construction” shows the main modelling steps of spatial mixed-frequency forecasting model, including the preliminary used methods and character analysis of the new model. Section 4 shows how the new model is used to forecast the regional GDP in China, while empirical findings and results comparisons are also presented here. Conclusion is given in Section 5.

2. Literature Review

2.1 Spatial Forecasting

Many scholars considered earlier the influence of spatial factors on forecasting and the representative achievements below are mainly based on the same frequency of spatial panel data. Baltagi et al. (2012) used Monte-Carlo experiments to compare various panel data prediction methods with related spatial error. Baltagi et al. (2014) further researched the parameter estimation method of dynamic spatial-panel data model and the prediction effect of optimal linear unbiased predictor (BLUP). Baltagi and Liu (2016) studied in detail the BLUP of the spatial-panel data model with the serial correlation under random and fixed effects proposed by Lee and Yu (2012).

Literature review shows that spatial forecasting is an important aspect in the field of spatial econometrics. That is to say, spatial correlation should be taken into account when forecasting. Research in this area is rare but with huge theoretical and practical significance. At the same time, it is not difficult to find that the related research on the spatial prediction of the spatial panel data model mostly uses the same frequency. In the current big data environment, more changes and improvements occurred in data collection and analysis. Decision makers need to effectively use data of different statistical frequencies to timely employ the valuable information and grasp future trends in advance. Prediction methods based on mixed-frequency data are currently an important research direction to improve the predictive power.

2.2 Mixed-frequency Data Forecasting

MIDAS can minimize the loss of sample information and interfere with subjective factors in data pre-processing with its special setting of the forecasting model. When making the economic forecasting, the collection of real-time data often involves the problem that the high-frequency and low-frequency data appear in one predictive model at the same time, that is, mixed-frequency data. Generally, the commonly used method is adding or interpolating to unify the mixed-frequency data to the same-frequency. But it will lead to the loss of part of the sample information. Moreover, the changing relationship between variables with different statistical time is ignored. To avoid this, scholars began to devote to directly use mixed-frequency data. Ghysels et al. (2004) proposed the MIDAS based on the distributed lag model, using the parameter-controlled lag weight polynomial function to weight high-frequency lag data to construct the model, and then use numerical optimization and nonlinear methods to estimate the optimal parameters (Foroni and Marcellino, 2013; Ghysels et al., 2005). Later scholars further improved the model to heighten the applicability. Ghysels et al. (2005) proposed an asymmetric MIDAS to consider the effect of high frequency data on low frequency data; Marcellino and Schumacher (2010) proposed a factorial mixed-frequency data sampling model (FACTOR-MIDAS) to process monthly data on the German economy. Clements and Galvao (2008) proposed a MIDAS-AR model that introduced autoregressive factors to predict quarterly industrial output values of the U.S.; Chen and Ghysels (2011) applied semi-parametric or non-parametric MIDAS models to the modelling of mixed-frequency in the financial field; Anderson et al. (2015) studied the parameter identification problem of the high-frequency multi-factor autoregression with mixed-frequency data.

Literature review reveals that research about mixed-frequency data forecasting are relatively mature and fruitful on model setting and estimation methods. In the application of MIDAS, regional factors are continuously mentioned by scholars. Rondeau (2012) has referred to the regional associations of fluctuation between different countries in Latin America. Kuzin et al. (2013) began to pay attention to the correlations and differences between different countries. Further, in regional economic development and cycle forecast MIDAS has obtained good results (Clements and Galvao, 2008; Kuzin et al., 2013; Ghysels and Ozkan, 2015).

3. Model Construction

3.1 Preliminary Method Preparation

Elhorst et al. (2013) presented the generalized dynamic spatiotemporal panel data model, combining many spatial econometric models under the same framework. Baltagi et al. (2014) presented a first-order dynamic autoregressive spatial lag panel data model, and the error term has spatial correlation. Its specific form is shown in the following equation,

$$y_{it} = \gamma y_{it-1} + \rho_1 \sum_{j=1}^{N} w_{ij}\, y_{jt} + x_{it}\beta + \varepsilon_{it} \qquad (1)$$

$$\varepsilon_{it} = \rho_2 \sum_{j=1}^{N} m_{ij}\, \varepsilon_{jt} + u_{it} \qquad (2)$$

$$u_{it} = \mu_i + v_{it} \tag{3}$$

where: y_{it} denotes the value of the explained variable at time t in region i, x_{it} is a 1×K vector representing the explanatory variable of region i at time t, γ and β are 1×1 and K×1 parameters vector to be estimated, respectively. ρ_1 represents the spatial lag coefficient. μ_i is the variable used to control all specific space and in invariable time; v_{it} is the residual term. Both are independent and identically distributed, obey the normal distribution (mean value is zero and the variance is σ_μ^2andσ_v^2 , respectively).

It should be pointed out that the above w_{ij} and m_{ij} represent two corresponding elements to the j-th column of the i-th row of spatial weight matrices, W_N and M_N (which may be the same), which embody the spatial correlation of the spatial panel data model. This advantage is also a special feature that distinguishes it from ordinary ones. The simplified spatial weight matrix integrates spatial correlations between the specific regional and other related regional economic variables into a new explanatory variable. This term is called the spatial lag term, and its coefficient reflects the size of the spatial interaction. Similarly, this paper refers to the above measurement to improve the spatial panel prediction model over the ordinary panel data prediction model. γ and ρ_1 satisfy the stability assumption (when$\rho_1 > 0$, $\gamma < 1 - \rho_1 r_{max}$, , when$\rho_1 < 0$, $\gamma < 1 - \rho_1 r_{min}$, and $|\gamma| < 1$, r_{min} andr_{max} are the minimum and maximum eigenvalues of the spatial weight matrix.

When $\gamma = \rho_2 = 0$, the optimal linear unbiased predictor for the i-th region (individual) in the future $T + \varrho$ is:

$$\hat{y}_{i,T+\varrho} = \sum_{k=1}^{K} \hat{\beta}_k \sum_{j=1}^{N} h_{ij} x_{k,j,T+\varrho} + \frac{T\sigma_\mu^2}{\sigma_1^2} \sum_{j=1}^{N} h_{ij} \hat{\varepsilon}_j \tag{4}$$

where: h_{ij} denotes the corresponding elements of the i-th row j-th column of $G_N^{-1} = (I_N - \rho_1 W_N)^{-1}$, $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$, and $\hat{\varepsilon}_j = (1/T)\sum_{t=1}^{T} \hat{\varepsilon}_{jt}$. Actually, the variance and spatial lag parameter ρ_1 is unknown, and need to be replaced by the ML estimator. When $\gamma \neq 0$ (the dynamic model), $\rho_1 \neq 0$ (spatial lag of explanatory variables), and $\rho_2 \neq 0$ (the spatial autocorrelation of error), the lagging endogenous variables are related to individual effects, it is more difficult to get the optimal BLUP. The development from general prediction models to spatial prediction is mainly based on introducing the spatial weight matrix to characterize the spatial correlation between economic variables in different regions and quantify the relationship further. In the spatial mixed-frequency forecasting the modeling process also follows the above ideas, but the difference is to integrate the characteristics of mixed-frequency data by introducing the spatial weight matrix.

The basic MIDAS proposed earlier is a univariate model. Its h_q （$h_q = h_m/m$） step forward prediction setting is

$$y_{t_q+mh_q} = y_{t_m+h_m} = \beta_0 + \beta_1 b(L_m;\theta) x_{t_m+w}^{(m)} + \varepsilon_{t_m+h_m} \tag{5}$$

where: $b(L^{1/m};\theta) = \sum_{k=0}^{K} c(k;\theta) L_m^k$, $L_m^k x_{t_m}^{(m)} = x_{t_m-x}^{(m)} \cdot x_{t_m+w}^{(m)}$ represents the jump sampling of high-frequency variable x_{t_m}; $t_q = 1, \cdots T_q$ represents the basic time unit (for example, quarter); m indicates the number of times that the higher sampling frequency occurs in the basic unit of time, which means that the high frequency data is a multiple of the low frequency data (for example, the multiple of the annual to the quarter is 4, m=4.; the quarter to the month is 3, m=3. the month to the day is 30, m=30, etc.); When the explanatory variable is a quarterly indicator, the explanatory variable is a monthly indicator, then m=3. w is the number of monthly index values that are predicted, available and earlier than the time of the explanatory variable (Foroni and Marcellino, 2013). Lower frequency variables can be represented by higher frequency variables in the following manner: $y_{t_q} = y_{t_m}$, $\forall t_m = mt_q$, where t_m is the time index for higher frequency variables.

$b(L^{1/m};\theta)$ is a weight function of the parameter vector θ. It is the most important part of MIDAS. Choosing appropriate function form can deal with the increase in parameters and the choice of the lag order k in the model. The important feature is that the number of parameters that need to be estimated is small, although the use of data increases the lag order of the explanatory variables, which is a great advantage in estimating continuously fluctuating data. $c(k;\theta)$ is taken as the lag coefficient, and its simplified parameter setting form is an important factor for accurate prediction using the MIDAS. The most commonly used parameter setting method is the Almon polynomial function. However, when it comes to geographical macroeconomic predictions, experience of MIDAS, but also need to overcome and solve emerging issues (spatial correlation).

3.2 Spatial Mixed Frequency Forecasting Model

To solve spatial effects of explanatory or explained variable in spatial mixed-frequency prediction, we combine the soft spatial weight matrix W_{NSS}^{cd} (Wang, X., & Xiao, Z., 2016) with the existing MIDAS. The introduction of W_{NSS}^{cd} not only can be used from a purely geospatial perspective, but also can be more realistically and comprehensively characterized from the perspective of economic or social network space (Bhattacharjee and Holly, 2013). w_{ij} is the element of W_{NSS}^{cd}, which is calculated by means of various methods such as proximity, distance relationship, economic and social impact of specific prediction problem $i,j = 1,2,\cdots$, n. n is the known number of spatial units or predetermined regions.

$$(\boldsymbol{F}^*,\boldsymbol{A}^*) = \left\{\begin{array}{c} (e_{u1},\{0,w_{21}^s,w_{31}^s,\cdots,w_{n1}^s\}), \\ (e_{u1},\{w_{12}^s,0,w_{32}^s\cdots,w_{n2}^s\}) \\ ,\cdots, \\ (e_{un},\{w_{1n}^s,w_{2n}^s,w_{3n}^s\cdots,0\}) \end{array}\right\} \quad (6)$$

$$W_{NSS}^{cd} = \begin{bmatrix} 0 & \cdots & w_{1n}^s \\ \vdots & \ddots & \vdots \\ w_{n1}^s & \cdots & 0 \end{bmatrix}. \quad (7)$$

Taking into account the spatial relationship, this study uses the model setting form of the spatial panel prediction model, and combines the basic MIDAS modeling ideas to construct the following spatial mixed frequency forecasting model. The h_q （$h_q = h_m/m$） step forward setting is

$$y_{i,t_m} = \beta_{i,0} + \beta_{i,1} b(L_m;\theta) \sum_{j=1}^{N} w_{i,j}^s\, x_{i,j,t_m+w-h_m}^{(m)} + \varepsilon_{i,t_m} \quad (8)$$

$$i = 1,2,\cdots,N \quad (9)$$

where: $b(L^{1/m};\theta) = \sum_{k=0}^{K} c(k;\theta) L_m^k$, $L_m^k x_{t_m}^{(m)} = x_{t_m-x}^{(m)} \cdot x_{t_m+w}^{(m)}$. $b(L^{1/m};\theta)$ is a weight function of the parameter vector θ. $c(k;\theta)$ is the lag coefficient. It is crucial for equation (8) to solve practical prediction with the simplified and less parameter setting of c(k;θ).The main methods are

Almon polynomial function:

$$c(k;\theta) = \frac{exp(\theta_1 k + \cdots + \theta_Q k^Q)}{\sum_{k=1}^{K} exp(\theta_1 k + \cdots + \theta_Q k^Q)} \quad (10)$$

$c(k;\theta)=\frac{\theta_1 k+\cdots+\theta_Q k^Q}{\sum_{k=1}^{K}\theta_1 k+\cdots+\theta_Q k^Q}$ is a common basic form. It can ensure that the weights are positive, make the equation have good properties of zero approximation error, and eliminate multicollinearity between variables at the same time.

Beta polynomial function with two parameters：

$$c(k;\theta_1,\theta_2)=\frac{f\left(\frac{k}{K},\theta_1;\theta_2\right)}{\sum_{k=1}^{K}f\left(\frac{k}{K},\theta_1;\theta_2\right)} \tag{11}$$

where: $c(x,a,b)=\frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$,$\Gamma(a)=\int_0^\infty e^{-x}x^{a-1}dx$。

The above-presented model is mainly based on MIDAS. The main idea is derived from the distributed lag model. It uses the parameter-controlled lag weight polynomial function to weight high-frequency lag data with heavy weight, and considers spatial correlation by adding the influence of spatial weight to further construct a specific model.

The spatial mixed frequency forecasting model is related to h, and the specific setting depends on the value of h. The optimal parameters in the mixed data model are estimated through data optimization and non-linear methods. Therefore, different prediction intervals require multiple parameter estimations.

More generally, the new model with autoregressive terms is shown below:

$$y_{i,t_m}=\beta_{i,0}+\sum_{h_n=0}^{B}y_{i,t_m-h_n}+\beta_{i,1}b(L_m;\theta)\sum_{j=1}^{N}w_{i,j}^{s}\,x_{i,j,t_m+w-h_m}^{(m)}+\varepsilon_{t_m} \tag{12}$$

where: B is the number of lags for the interpreted variable, y_{i,t_m}. When $B=0$, the model does not include the autoregressive term; when $B\neq 0$, the model contains the autoregressive term, and the determination of B needs to be set according to actual problems. At the same time, the difficulty of recognition of B also limited the use scope of the model.

The important features of the above model are presented below.

The polynomial weighting for the high frequency mixing data in the MIDAS is modified by the soft spatial weight matrix, so that the final weight value of the high frequency data takes into account its spatial correlation. The coefficients of the explanatory variable in the new model are determined by the lagging weights of the mixed-frequency data distribution, the soft spatial weights and the constant coefficients.

In the case of a single high-frequency explanatory variable, the use of multi-phase data increases the lag order, and also uses the simplified polynomial weight function but only requires fewer parameters to obtain the optimal prediction result. This process ensures that high-frequency data is not pre-processed in advance, and the original data is directly reflected in the model, reducing information distortion caused by human factors (loss of information or excessive human intervention).

⑤When the model is applied to a problem, some exogenous parameters may be adjusted according to actual needs, including the selection of a polynomial weight function, the selection of the number of polynomial weight function parameters, the selection of the lag order of high frequency data, the number of periods of inter-period prediction, whether or not to introduce the lagged terms of the explained variable, and so on. The above adjustments, on the one hand, reflect that the constructed model has some flexibility in application to real problems, and on the other hand, it reflects that the model recognition needs further study.

4. Forecasting Analysis of Regional GDP in China

4.1 Variables and Data Description

Not only the government revenue and expenditure but also the corporates' profits and financial status can be influenced by the trend and fluctuation of quarterly GDP. Regional quarterly GDP is more discriminate besides reflecting the economic growth, which from the different geographical perspectives has a more direct guidance value for government, businesses and individual decision-making in each region. Forecasting regional quarterly GDP is an important tool for real-time or short-term forecasting of the total amount of macroeconomic development. It can effectively identify uncertainties in regional economic growth and development trends, provide decision support for macroeconomic policy adjustments, and allows for taking preventive measures.

However, due to the fact that collection and processing of data are both systematic and complicated, the publication of macroeconomic data every year is not announced for the first timing, and annual GDP can only be obtained late in the next year, while quarterly GDP is available in the next quarter. Other data's release dates are with relatively certain lags. For example, the estimated release time for the preliminary verification of GDP announcement of 2015 is in September of 2016. The average annual salary of employed persons in urban and non-state-owned units in urban areas, employees in urban private units and online direct reporting platform employees in survey units is expected to be released in May of 2016. Estimated release date of national value-added data on cultural and related industries and national tourism and related industries of 2015 is October 2016. The estimated publication time of the statistical bulletin on scientific and technical funding in 2015 is October 2016. In addition, with the rapid development of Internet and Internet of Things technologies, domestic regions are closely linked and their respective economic growth is also connected. Considering the spatial relationship among GDP in different regions and then effectively predicting the future trend of regional economic growth is of great significance for regional recognition of local risks and the formulation of development strategies that are consistent with the characteristics of the region (population, resources, environment, and culture).

In response to the above-mentioned problems, the existing regional quarterly GDP forecast studies mostly deal with the same frequency or use the MIDAS that does not consider the spatial relationship (Zheng and Wang, 2013; Liu and Liu, 2011). The spatial mixed-frequency forecasting model proposed in this paper can use the newly published high-frequency data to update real-time and short-term predictions of low-frequency data, so that it can be avoided that the current macroeconomic state and its trend not be judged in a timely and accurate manner due to the time lag in the release of economic data, while better considering the spatial effect of each regional GDP.

Considering the spatial characteristics of regional economy, in this section the spatial correlation of factors (price level) is taken into account, utilizing monthly regional price level to predict regional quarterly real GDP growth rate, model fitting effect (R2, log likelihood function, AIC, and BIC Information Criteria), parameter significance test and prediction accuracy index (root mean square prediction error MSFE, mean square prediction error MSE) to evaluate the comprehensive prediction effect, and then to explore the feasibility. A large number of studies on the growth rate of China's macroeconomic growth show that the fluctuation of China's output growth rate depends not only on the fluctuation of the price level but also on the combined effects of other factors. This part draws on the early results of domestic scholars using the mixed-frequency sampling model to study China's problems, and does not choose a more complex situation - only a single explanatory variables and a single explained variable was used when constructing a spatial mixed-frequency forecasting model for macroeconomic fluctuations in 30 provinces, municipalities and autonomous regions of China.

China has 34 provincial administrative units (including 23 provinces, 5 autonomous regions, 4 municipalities, and 2 special administrative regions), Considering that it is difficult to collect and sort data from Tibet, Hong Kong, Macau, and Taiwan, and the statistics are inconsistent, this section selects 30 of the 34 Chinese provinces, municipalities and autonomous regions as the main research object (excluding the above-mentioned four regions), and selects the sample time interval from 2005 to 2010. The real GDP growth rate is quarterly, from the first quarter of 2005 to the first quarter of 2016; the fluctuation of price level (measured by CPI) is monthly data, from January 2005 to March 2016. Taking into account spatial mixed data between regional GDP variables (annual and quarterly) and spatial variance of explanatory variable price levels (CPI measures), the spatial mixed-frequency forecasting model is used here. The primary data are mainly from the Statistical Yearbook of China 1998-2015 (http://www.stats.gov.cn/tjsj/ndsj/), the statistical compilation of New China's 60 years, the statistical yearbooks of above relevant provinces and cities, and other official websites. To ensure the comparability of data, this section uses the GDP deflator index to eliminate the price fluctuations in the economic indicators during the forecasting process. The data required for the construction of the soft weight matrix is based on the 1:4,000,000 electronic versions of China Maps downloaded from the official website of the National Geographic Information System and Google Earth software. Both in-sample predictions are implemented through Matlab (R2013b 8.0.0.783-Win64) programming, by which we complete statistical analysis and mapping of prediction errors.

4.2 Forecasting Effectiveness Analysis

Forecasting effectiveness analysis aimed to explore the following three aspects in the context of predicting China's regional economic growth rate. First, whether the setting form of autoregressive model helps to improve the accuracy of prediction as compared to a model that does not consider the time dependence of the independent variable. Second, the commonly used weight functions have different trend characteristics of mixed-frequency data. When the new model is applied to a specific problem, how do the specific influences of the weight functions perform on the prediction accuracy? Third, the length of the prediction period is an important factor affecting the prediction (Kuzin et al., 2013), and accuracy of the spatial mixed-frequency prediction. What is the specific change in prediction accuracy during different prediction periods? It is beneficial in order to capture the quantitative difference in regional economic development.

Choosing the weight function $c(k;\theta)$ is an important issue for spatial mixed-frequency forecasting. According to Clements and Galvao's empirical research on US GDP, the forecasting selection of lag-order experience (Clements and Galvao, 2008) is based on the reference AIC and BIC criteria, the previous empirical analysis of lagging orders, and the actual situation in China. The lag order K is selected as 12 (Liu, 2013). In the case of K=12, the predictive effects with and without the autoregressive items of the added explained variables are both selected to analyze the errors caused by the model setting. At the same time, different H values are adjusted according to the prediction needs (1, 2, and 3, respectively) to analyze the prediction effect of different prediction intervals. And to intuitively understand the specific changes of spatial correlation and different weight functions of the explanatory variables, the unconstrained Beta polynomial function, the constrained Beta polynomial function, the exponential Almon polynomial function and the 3rd-order Almon polynomial function are chosen to fit data, respectively (Armesto et al., 2010), where RMSE is the main method for evaluating the estimating parameters and predicting.

The main findings are listed below.

①The spatial mixed-frequency forecasting model with auto-regression is less accurate than that without auto-regression terms. This finding was consistent with the conclusion of Lin and Liu (2010), who find that the Monte Carlo simulation experiment incorporating the autoregressive features of mixed-frequency data could improve the forecasting effect. It is also consistent with the conclusion of Liu and Liu (2011), who pointed out that economic development has strong system inertia, and the mixed-frequency data model with autoregressive terms has significantly improved short-term estimates during the financial crisis. When the spatial mixed-frequency forecasting model used for regional GDP growth forecasting, the prediction errors (with different lag order H) are significantly reduced as the auto-regression items are added to predict the GDP growth rate; this conclusion can be compared by the minimum value of the forecast error of each region in the far right column of Tables 1 to 6. Under different weight functions and different lag order H values, this conclusion is also valid. It proves that considering the influence of the spatial correlation of explanatory variables on the weight of regional GDP mixed-frequency data forecasting, incorporating the autoregressive item at the same time, is consistent with the economic operation regularity and helps improving the prediction accuracy.

②When predicting China's real GDP growth rate modeling with or without autoregression, the optimal weighting functions of the specific forecasting models of the 30 provinces, municipalities and autonomous regions are not different. On the one hand, the new model has considered the regional correlation of inflation, and it is an important aspect of explaining the mutual influence of price levels among regions. On the other hand, it takes into account the spatial relationship and fully considers the heterogeneity in different regions (Yu et al., 2014). According to the specific conditions of provinces and cities, we select the appropriate polynomial function to describe the weight of spatial mixed frequency data, and then reduce the prediction error. The above findings are in line with the regional differences in the geographical environment, population, economy, history and culture. It is helpful to understand the regional heterogeneity of the effectiveness of monetary policy (Yu and Huang, 2015).

Table 1

RMSE of prediction in sample in 30 regions of China with K=12 H=1 and without autoregression (%)

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Beijing	21.237499	21.172914	21.734963	21.230052	21.172914
Tianjin	20.092653	20.262813	20.183034	19.453986	19.453986
Hebei	27.321216	27.136739	27.290288	26.706963	26.706963
Shanxi	26.483404	27.067957	26.513261	25.955635	25.955635
Inner Mongolia	40.627075	41.665039	41.035683	40.585844	40.585844
Liaoning	26.750729	27.766826	29.001111	27.641776	26.750729
Jilin	53.322560	50.795103	53.184781	53.672333	50.795103
Heilongjiang	53.057683	57.189090	54.703327	55.140064	53.057683
Shanghai	14.572734	14.681737	14.716223	16.159271	14.572734

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Jiangsu	23.543050	23.703753	24.154949	24.356039	23.543050
Zhejiang	27.260293	27.039949	27.723229	28.320076	27.039949
Anhui	22.394558	22.261050	22.919065	22.089611	22.089611
Fujian	48.506340	49.784221	50.239526	50.581492	48.506340
Jiangxi	26.081143	24.351141	26.944768	26.303126	24.351141
Shandong	23.512151	21.951094	23.490156	22.878108	21.951094
Henan	14.129320	14.108222	14.079949	14.595358	14.079949
Hubei	29.339557	29.224684	30.401809	29.448311	29.224684
Hunan	30.892879	30.480838	31.869345	31.394021	30.480838
Guangdong	20.557333	21.466671	20.876650	22.001220	20.557333
Guangxi	41.093500	41.599133	40.947320	40.471961	40.471961
Hainan	16.501817	17.618601	16.769492	17.513624	16.501817
Chongqing	23.728368	22.651493	23.614348	23.612025	22.651493
Sichuan	16.519427	16.080777	16.533899	16.746792	16.080777
Guizhou	49.105289	48.674792	48.948105	50.662564	48.674792
Yunnan	35.934940	33.455567	36.138533	34.349705	33.455567
Shaanxi	35.887638	35.993257	36.314635	36.228296	35.887638
Gansu	41.333977	41.447554	41.369486	42.090472	41.333977
Qinghai	36.211606	35.974439	36.244227	35.948824	35.948824
Ningxia	39.417559	39.407683	39.607155	40.044595	39.407683
Xinjiang	57.025441	57.592577	57.372811	59.187834	57.025441
Average	31.414725	31.420190	31.830738	31.845666	30.943852

Table 2

RMSE of prediction in sample in 30 regions of China with K=12 H=2 and without autoregression (%)

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Beijing	22.387426	21.749307	22.394639	21.985737	21.749307
Tianjin	20.166447	20.569120	20.382184	20.087674	20.087674
Hebei	27.235958	26.878749	27.071228	26.911541	26.878749
Shanxi	26.566785	25.916908	26.504587	25.474859	25.474859
Inner Mongolia	41.422298	43.233554	41.503166	40.246913	40.246913
Liaoning	28.500841	27.155654	28.318291	26.320281	26.320281
Jilin	53.684185	53.985834	53.456741	51.544459	51.544459
Heilongjiang	54.363626	56.146808	54.688661	52.790823	52.790823
Shanghai	14.445541	14.164489	14.443346	14.119188	14.119188
Jiangsu	23.807961	23.704973	23.811509	23.326474	23.326474
Zhejiang	27.841980	27.980298	27.876488	27.602139	27.602139
Anhui	22.852235	22.925146	22.838530	23.242274	22.838530
Fujian	49.214284	50.398191	49.413843	47.977923	47.977923

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Jiangxi	26.333084	27.451158	26.537427	25.687410	25.687410
Shandong	23.348329	23.503903	23.277008	23.149064	23.149064
Henan	13.951641	14.211796	13.969009	14.360451	13.951641
Hubei	29.797125	30.713505	29.876640	28.991465	28.991465
Hunan	30.930949	31.121174	30.937953	31.059307	30.930949
Guangdong	20.934578	20.844528	20.863856	20.868927	20.844528
Guangxi	41.746002	42.879323	41.013247	40.961016	40.961016
Hainan	16.367616	16.344732	16.622904	16.400658	16.344732
Chongqing	24.584823	26.056108	25.248475	27.094515	24.584823
Sichuan	16.230766	16.915659	16.402824	17.114591	16.230766
Guizhou	47.759790	50.095877	48.119264	50.001081	47.759790
Yunnan	36.105834	35.605952	36.926418	37.573858	35.605952
Shaanxi	36.204656	35.326766	36.311815	36.149590	35.326766
Gansu	39.788341	39.540187	39.721497	39.577166	39.540187
Qinghai	35.613600	36.644907	35.798486	36.193854	35.613600
Ningxia	39.477022	40.323122	39.481141	39.676450	39.477022
Xinjiang	56.675764	59.059375	56.733341	57.175674	56.675764
Average	31.611316	32.048237	31.684817	31.455512	31.087760

Table 3

RMSE of prediction in sample in 30 regions of China with K=12 H=3 and without autoregression (%)

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Beijing	20.951750	20.974333	21.288048	20.993957	20.951750
Tianjin	20.074547	17.682028	19.770272	20.724602	17.682028
Hebei	24.790026	24.741938	25.124853	25.341579	24.741938
Shanxi	24.927747	24.683628	25.020221	25.992935	24.683628
Inner Mongolia	38.940397	40.690503	39.992346	42.615232	38.940397
Liaoning	26.571242	27.553113	26.407413	29.122515	26.407413
Jilin	52.492903	53.891884	52.054180	52.683670	52.054180
Heilongjiang	55.031193	55.457383	54.782819	55.175671	54.782819
Shanghai	14.093589	14.305025	14.430645	16.130086	14.093589
Jiangsu	22.540265	22.060527	22.927637	24.500005	22.060527
Zhejiang	27.323244	28.261416	27.611545	29.422509	27.323244
Anhui	21.977048	22.279880	22.600740	24.618237	21.977048

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Fujian	47.802004	48.724476	48.806083	53.441148	47.802004
Jiangxi	25.642111	26.760536	26.307961	29.535813	25.642111
Shandong	21.252809	21.239289	21.637516	22.105995	21.239289
Henan	14.108467	13.993075	14.280540	14.289900	13.993075
Hubei	29.122600	29.119596	29.145419	31.039586	29.119596
Hunan	29.663405	32.023574	30.222317	33.723598	29.663405
Guangdong	20.595689	20.520000	20.545625	22.657381	20.520000
Guangxi	41.432222	41.909586	41.476214	43.575959	41.432222
Hainan	16.928335	17.103183	16.887007	17.061257	16.887007
Chongqing	24.463417	24.317244	24.456329	26.985661	24.317244
Sichuan	16.122956	16.378546	16.404849	17.226520	16.122956
Guizhou	47.577710	46.397783	47.597254	53.331603	46.397783
Yunnan	35.570340	36.299707	35.564884	37.734816	35.564884
Shaanxi	35.577541	33.896758	35.413579	38.434557	33.896758
Gansu	41.649336	37.059588	41.176340	40.149491	37.059588
Qinghai	35.008830	35.025604	35.246368	36.268586	35.008830
Ningxia	38.886139	39.381490	39.131244	39.813631	38.886139
Xinjiang	55.866452	53.523119	56.640716	58.165160	53.523119
Average	30.899477	30.875160	31.098366	32.762055	30.425819

Table 4

RMSE of prediction in sample in 30 regions of China with K=12 H=1 and autoregression (%)

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Beijing	7.694750	6.830568	6.359115	6.439115	6.359115
Tianjin	8.179567	7.684003	8.013168	7.602940	7.602940
Hebei	7.392890	7.462474	7.472298	7.381301	7.381301
Shanxi	11.234813	10.912689	11.737130	10.799217	10.799217
Inner Mongolia	12.725142	12.878744	11.808088	12.464901	11.808088
Liaoning	9.573907	9.651736	9.383167	10.176578	9.383167
Jilin	9.360429	8.843984	8.504404	9.024195	8.504404
Heilongjiang	10.283746	9.957243	9.760496	9.719419	9.719419
Shanghai	5.120091	5.226694	4.632218	5.046372	4.632218
Jiangsu	5.528253	5.664509	5.560478	5.658860	5.528253
Zhejiang	5.521753	5.637652	5.483048	5.610067	5.483048

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Anhui	7.399601	7.565069	7.221458	7.515621	7.221458
Fujian	5.683896	5.428706	5.598626	5.455972	5.428706
Jiangxi	7.596300	7.431967	7.444189	7.936560	7.431967
Shandong	5.150457	5.133363	5.255873	4.955206	4.955206
Henan	6.936993	6.417247	6.414829	6.406960	6.406960
Hubei	7.914541	7.815474	7.542492	7.958770	7.542492
Hunan	8.153057	7.752086	7.819927	8.100990	7.752086
Guangdong	4.503692	5.774581	4.333792	5.654066	4.333792
Guangxi	12.235883	10.278977	12.259603	11.395583	10.278977
Hainan	5.347706	5.715207	5.507923	5.953970	5.347706
Chongqing	10.444147	9.540283	10.016660	10.049797	9.540283
Sichuan	7.074839	6.808687	6.970010	7.157314	6.808687
Guizhou	6.298965	5.446079	5.401438	6.913163	5.401438
Yunnan	6.473666	7.345031	6.282402	7.587385	6.282402
Shaanxi	9.225288	9.343468	8.833700	9.397243	8.833700
Gansu	8.563486	8.685756	8.397639	8.756431	8.397639
Qinghai	7.328877	7.314012	7.460869	7.798260	7.314012
Ningxia	11.466279	11.595814	11.280895	11.725163	11.280895
Xinjiang	12.235883	10.278977	12.259603	11.395583	10.278977
Average	8.088297	7.880703	7.833851	8.067900	7.601285

Table 5

RMSE of prediction in sample in 30 regions of China with K=12 H=2 and autoregression (%)

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Beijing	7.831710	7.720176	7.845213	6.605303	6.605303
Tianjin	7.942528	8.121169	7.964332	7.759595	7.759595
Hebei	7.401426	7.431211	7.275001	7.250175	7.250175
Shanxi	4.609807	4.374033	4.425528	4.870534	4.374033
Inner Mongolia	12.264050	12.592238	12.218214	13.132864	12.218214
Liaoning	9.813842	10.840395	9.331189	10.302386	9.331189
Jilin	9.184274	9.360423	8.777076	9.339977	8.777076
Heilongjiang	10.298397	9.550924	9.383201	9.481550	9.383201
Shanghai	4.945677	4.722404	5.108439	5.340966	4.722404
Jiangsu	5.531825	5.449911	5.536504	5.646380	5.449911

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Zhejiang	5.444708	5.417243	5.415354	5.690548	5.415354
Anhui	7.196860	7.539788	7.343352	7.914539	7.196860
Fujian	5.476283	5.471182	5.490997	5.641160	5.471182
Jiangxi	7.110701	7.126706	7.172907	7.843846	7.110701
Shandong	5.098485	5.305575	5.198965	5.236213	5.098485
Henan	6.727528	6.688222	6.705590	6.928370	6.688222
Hubei	7.457856	7.747021	7.469104	7.781708	7.457856
Hunan	7.771651	7.677739	7.955098	8.379262	7.677739
Guangdong	4.609807	4.374033	4.425528	4.870534	4.374033
Guangxi	7.674628	7.703801	7.821550	7.994297	7.674628
Hainan	5.339434	5.932206	5.810313	5.724884	5.339434
Chongqing	9.984939	10.559586	10.118271	11.160746	9.984939
Sichuan	6.713525	6.811258	6.866126	6.856811	6.713525
Guizhou	5.447784	5.352325	5.526502	6.084901	5.352325
Yunnan	6.731111	7.327409	6.207537	7.757610	6.207537
Shaanxi	8.873748	10.122439	9.113830	9.947516	8.873748
Gansu	8.720577	8.132964	8.553630	8.905193	8.132964
Qinghai	7.162161	7.604579	7.447492	7.778587	7.162161
Ningxia	11.165503	11.842413	11.587404	11.781736	11.165503
Xinjiang	10.538791	11.661017	10.538788	11.101279	10.538788
Average	7.502320	7.685346	7.487768	7.836982	7.316903

Table 6

RMSE of prediction in sample in 30 regions of China with K=12 H=3 and autoregression (%)

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Beijing	7.186203	7.297255	7.404276	6.948655	6.948655
Tianjin	7.897008	7.163407	7.994626	8.191396	7.163407
Hebei	6.954879	7.331425	7.452494	7.208597	6.954879
Shanxi	10.671095	11.114958	10.408006	10.348941	10.348941
Inner Mongolia	12.059325	11.315725	12.974716	12.333028	11.315725
Liaoning	9.959806	9.228613	9.863024	10.450744	9.228613
Jilin	9.443794	9.428089	9.016563	9.030724	9.016563
Heilongjiang	10.106195	10.760040	9.908417	9.294345	9.294345
Shanghai	5.153445	4.905340	5.073402	4.840737	4.840737

Region Name	Unconstrained Beta	Constrained Beta	Exponent Almon	3 Order Almon	Minimum
Jiangsu	5.408215	5.324250	5.615415	5.538514	5.324250
Zhejiang	5.459770	5.488168	5.288356	5.095791	5.095791
Anhui	7.227138	7.449028	7.175116	7.396827	7.175116
Fujian	5.750342	5.467447	5.705821	5.588209	5.467447
Jiangxi	7.301851	7.356697	7.390479	7.273148	7.273148
Shandong	5.026713	5.101504	5.223232	5.198276	5.026713
Henan	6.338017	6.861823	6.467371	6.258876	6.258876
Hubei	8.016065	7.382092	7.568766	8.127265	7.382092
Hunan	7.723677	7.281990	7.761818	7.332465	7.281990
Guangdong	4.664753	4.618825	4.550611	4.745877	4.550611
Guangxi	7.752926	7.411763	7.641646	7.716220	7.411763
Hainan	5.654506	6.085889	5.205064	5.359141	5.205064
Chongqing	9.223904	9.567957	9.024309	10.265168	9.024309
Sichuan	7.088281	7.299554	6.911702	7.158181	6.911702
Guizhou	5.863506	5.356546	5.470027	5.266760	5.266760
Yunnan	6.880522	6.558136	6.176768	6.904077	6.176768
Shaanxi	9.033750	9.072926	9.186684	9.467624	9.033750
Gansu	8.424467	8.959466	8.792826	8.967916	8.424467
Qinghai	7.175446	7.156455	7.198687	7.854822	7.156455
Ningxia	11.186155	11.666444	11.242829	11.703625	11.186155
Xinjiang	8.744073	9.842502	10.483754	10.172877	8.744073
Average	7.645861	7.661810	7.672560	7.734628	7.349639

④During the model fitting and parameter estimation of K=12 H=3 auto-regression Beta without constraint for 30 provinces, autonomous regions and autonomous regions of China, using the maximum likelihood as a reference, regional models in each province and city are well-fitted (the greater the maximum likelihood, the better the model fits). They fit better overall under 3 different weight functions. At the same time, the AIC and BIC information criterion are used as the model weights analysis in the MIDAS combined forecasting model. The new model selects different polynomial functions, and the number of parameters to be estimated is different. In general, increasing the number of free parameters can improve the goodness of the fitting, while the AIC encourages the goodness of data fitting while avoiding overfitting as much as possible. The smaller the AIC value, the better the model. BIC is also an index that evaluates the model fitting effect as well as AIC. The smaller the BIC value, the better the model. The values of the AIC and BIC information criteria can be compared to analyze the overall performance of the model between different regions under the same polynomial weight function or between different weight polynomials functions in the same region. To calculate the ratio of AIC to BIC between different methods, the larger the ratio, the better the model on the numerator relative to that on the denominator is. Although many researches do not make stringent requirements for the consistency of model fitting parameters when evaluating the predicting performance, their estimates, significance and fitting effects are important, have certain reference value for understanding and evaluating the structure of prediction models overall.

Figure 1
Prediction weights in the Northeast regions of China under 12 lag phases

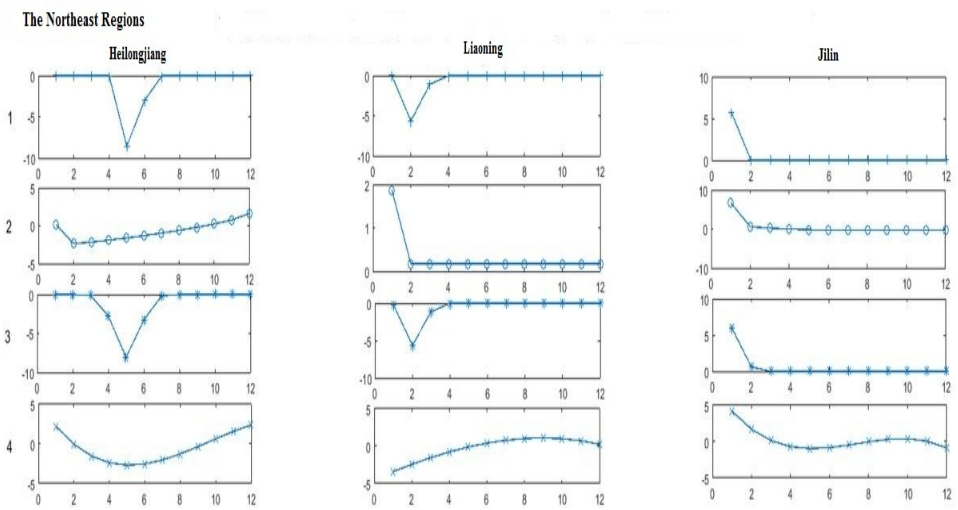


Figure 2
Prediction weights in the Eastern regions of China under 12 lag phases

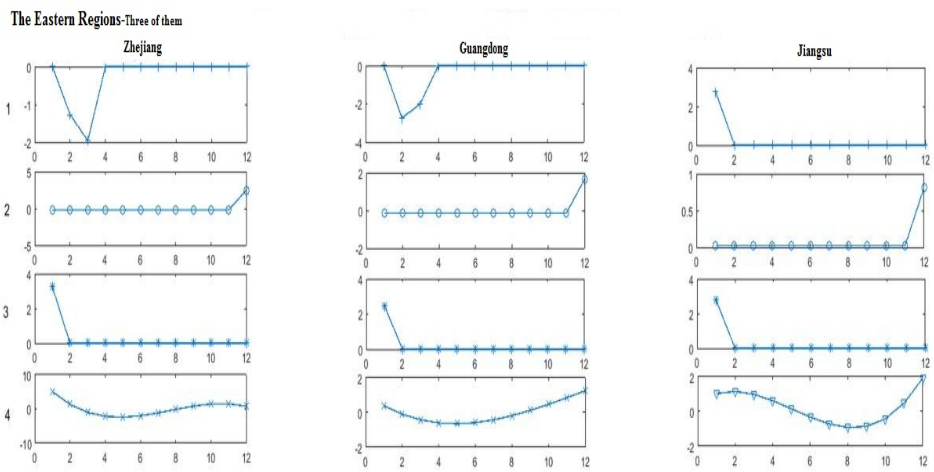


Figure 3
Prediction weights in the Western regions of China under 12 lag phases

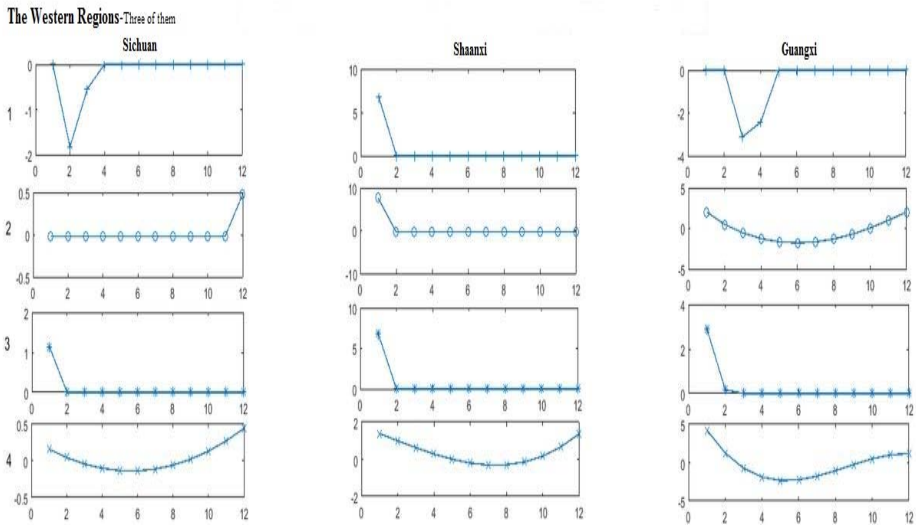
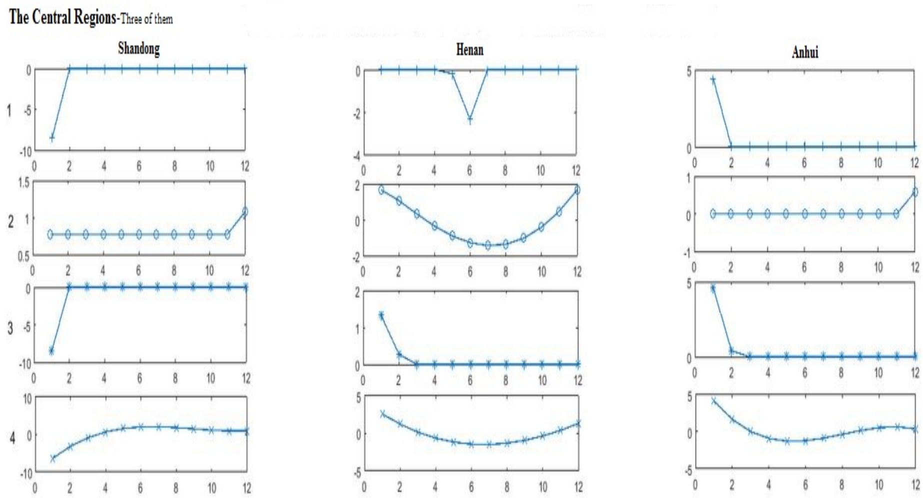


Figure4
Prediction weights in the Central regions of China under 12 different lag phases



In general, good prediction accuracy for each region in the above-presented tables shows that spatial mixed-frequency forecasting can tackle spatial relation and mixed-frequency data by the information efficiency of directly using the mixed-frequency data in one model without prior treatments. The statistics (data) and forecasting model could serve as extremely important drivers for economists who work on short-term, medium-term, and long-term economic cycle forecasts to better understand, circumvent, and mitigate the dangers of economic or financial crises. It also helps to understand the theories such as balanced development developed on the basis of the Harrod-Domar neoclassical economic growth model, unbalanced development proposed by Hirschman, gradient transfer proposed by Vernon, and growth poles first proposed by Francois Peru, center-periphery theory and urban circle economic theory. As one of the quantitative analysis methods, the forecasting results derived from spatial mixed-frequency forecasting have great reference value for the effective allocation of regional economic factors. Spatial mixed frequency forecasting can show confidence to rely on rich statistical evidence for the accurate and scientific decision-making of economic subjects.

4.3 Forecast Weight Analysis

Regional economic development is closely related to spatial locations and corresponding population, natural resources, human history and geographical conditions. Considering the different production methods, resource endowments and industrial structure differences in different regions, this part differentiates 30 provinces and cities in China and analyzes differently their economic trends, and studies the specific influence of corresponding differentiation on the weight function. The conclusions in this section can expand the scope of use of the proposed model and the ability to explain economic problems, which is helpful to understand the regional economic development theories, such as balanced development, unbalanced development, gradient transfer, growth poles, center-periphery theory and urban circle economic theory, from quantitative analysis perspective. For example, which regions are acting as the leader among the local or adjacent areas or neighbors? This is one of the focuses of above-mentioned gradient transfer, growth poles, center-periphery theory, and the predictable differential growth rate tells us the current status of regional imbalance development.

The Chinese government departments divide the national economic region into four regions: the Eastern region, the Central region, the Western region, and the Northeast region. It shows that the provinces and cities in the same region have similar characteristics in terms of social, economic and geographical conditions. To reflect the regional characteristics of quarterly real GDP growth forecast to provide reference for the development of regional government organizations' economic development policies. This section selects three provinces as main objects from each of the above four regional classifications, give a comparative analysis of the weights of different polynomial functions. Findings in this area help to further analyze the performance and applicability of the models constructed for different provinces, cities, and regions under a specific spatial scope.

We calculate the weight values of 12 different lag phases under the setting of the spatial weight of inflation in each province and city, and then build a map (see Figures 1 to 4 for details). The chart shows the specific trend of the change in the weight value of 12 different lag phases in 3 representative provinces and cities. The numbers "1, 2, 3, 4" in the figure indicate the "unconstrained Beta polynomial function, constrained Beta polynomial function, exponential Almon polynomial function, 3rd order Almon polynomial function" under the introduction of spatial weight. Each row represents a consistent weighting function, and each column represents different ways of empowering the same region.

Both horizontal and vertical comparisons provide an intuitive understanding. The abscissa of a single figure represents different lag levels, and the ordinate represents the different weights values. According to the general analysis, the 3rd order Almon polynomial function in the Eastern region, the Central region, the Western region, and the Northeast region is relatively stable, and the trend of change is more consistent, with "0" being the center of fluctuation. The unconstrained Beta polynomial function, the constrained Beta polynomial function, and the exponential Almon polynomial function all show an abnormal change in the weight value. Taking the unconstrained Beta-polynomial function in Northeast China as an example, Heilongjiang and Liaoning experienced a sudden drop in the lag phase of 5 and 2, and then they were stable, the weight value was negative and close to zero, however, Jilin began to stabilize at 2 in the lag phase, and then stabilized, with the weight being a positive number and close to zero. Similarly, the three Eastern provinces (Zhejiang, Guangzhou, Jiangsu), the Western regions (Sichuan, Guangxi, and Shaanxi) and the Central regions (Shandong, Henan, and Anhui) have similar trends, and the location of abnormal locations varies. But overall, the weights conform to the tendency that the absolute value goes smaller as the lag period becomes larger.

4.4 Comparative Analysis with Standard Models

To further verify the proposed method, which can simultaneously utilize and integrate the short-term fluctuation information (high-frequency data) and long-term trend information (low-frequency data), to effectively improve the prediction accuracy, this section chooses the MIDAS model as the standard model to compare the forecasting accuracy to show the relative merits of the proposed method. Firstly, we employ the above relatively best and relatively worst model setting form. The proposed model with the Exponent Almon (K=12 H=1 and without autoregression), the Exponent Almon (K=12 H=2 and autoregression), and the Unconstrained Beta (K=12 H=3 and autoregression) are labeled as PM1, PM2 and PM3, respectively. Secondly, we adopt the basic MIDAS to make the forecasting of 30 regions of China, repectively. That is to say MIDAS is reused 30 times without considering the spatial correlation of variables. Thirdly, we implement the Diebold Mariano tests for RMSEs between every two model settings. We mark the significance test results (P=value) under the levelof 0.1, 0.05 and 0.01 with *, ** and ***, respectively.

Table 8

Comparison Analysis with Standard Models by In-sample Forecasts

DM Value	MIDAS	PM1	PM2	PM3
MIDAS	NG	1.8960*	1.9047**	1.7734*
PM1	1.8960*	NG	1.8767*	1.6984**
PM2	1.9087**	1.8767*	NG	1.7948*
PM3	1.7734*	1.6984**	1.7948*	NG

Table 9

Comparison Analysis with Standard Models by Out-of-sample Forecasts

DM Value	MIDAS	PM1	PM2	PM3
MIDAS	NG	0.2660	1.3445*	0.8764
PM1	0.2660	NG	1.7551*	1.8009*
PM2	1.3445*	1.7551*	NG	1.7565*
PM3	0.8764	1.8009*	1.7565*	NG

Comparison analyses with standard models by in-sample and out-of-sample forecasts have been made. The forecasting step size between each two models is consistent. The main results are listed in Table 8 and 9. As the results show, in-sample forecasts pass the significance test, providing strong evidence that the proposed methods (PM1, PM2 and PM3) dispaly better forecasting results than the standard model. Although the DM value bebween PM1 and MIDAS does not pass the significance test, most comparison between the other two different models have passed the tests during the out-of-sample forecasts analysis. This conclusion proves more quantitatively that the proposed method can improve the prediction accuracy. This benefits from that fact that the new approach reduced the impact of data pre-processing on information intervention, which uses traditional methods to align the statistical frequency between macroeconomic indicators such as quarterly GDP and monthly CPI. It also shows the better impovement of forecasting accrurancy. Meawhile, considering the short-term economic fluctuations and long-term trends contained in the indicators, it will help to improve the forecast performance and it is the quantitative basis for improving the understanding of the regional economic cycle changes, growth drivers and regional differences.

5. Conclusion

This paper proposes a novel spatial mixed-frequency forecasting model and applied the new method to forecast the real GDP growth rate in 30 provinces and autonomous regions in China, which show a good performance and a great feasibility overall. It also provides some helping findings about the characteristics of parameters such as weighs polynomials, prediction weighs, and lag period in different regions. In detail, forecasting effectiveness analysis shows more detailed information about how the model performs, different K and H, with or without autoregression in different regions. Good prediction accuracy for each region shows that spatial mixed-frequency forecasting can tackle spatial relation and mixed-frequency data by the information efficiency of using directly the mixed-frequency data in one model without prior treatments. The statistics (data) and forecasting model could serve as extremely important drivers for economists who work on short-term, medium-term, and long-term economic cycle forecasts to better understand, circumvent, and mitigate the dangers of economic or financial crises. It also helps to understand the theories, such as balanced development developed on the basis of the Harrod-Domar neoclassical economic growth model, unbalanced development proposed by Hirschman, gradient transfer proposed by Vernon, and growth poles first proposed by Francois Peru, center-periphery theory and urban circle economic theory.

The following are unsolved issues deserving further study. (1) There are many factors affecting the forecasting results of practical problems, so there is still big improvement for expansion in the application of multivariate, non-linear set forecasting phenomenon. (2) More good identification and setting of spatial effects of explained and explanatory variables is another direction; for example, choosing of spatial lag, spatial error or dynamic form. (3) What is the optimal lag order and whether needs autoregression requires more systematic and solid mathematical supports besides the empirical results-oriented way?

References

Anderson, B.D.O. *et al.,* 2015. Multivariate AR systems and mixed frequency data: g-identifiability and estimation. *Econometric Theory*, 1(4), pp.793-826.

Andreou, E. Ghysels, E. and Kourtellos, A., 2010. Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2), pp. 246-261.

Armesto, M.T. Engemann, K.M. and Owyang, M.T., 2010. Forecasting with mixed frequencies. *Federal Reserve Bank of St Louis Review*, 92(6), pp.521-536.

Baltagi, B.H. Bresson, G. and Pirotte, A., 2012. Forecasting with spatial panel data. *Computational Statistics and Data Analysis*, 56(11), pp.3381-3397.

Baltagi, B.H. Fingleton, B. and Pirotte, A., 2014. Estimating and forecasting with a dynamic spatial panel data model. *Oxford Bulletin of Economics and Statistics,* 76(1), pp.112-138.

Baltagi, B.H. and Liu, L., 2016. Prediction in a generalized spatial panel data model with serial correlation. *Journal of Forecasting*, 35, pp.573–591

Bhattacharjee, A. and Holly, S., 2013. Understanding Interactions in Social Networks and Committees. *Spatial Economic Analysis*, 8(1), pp.23-53.

Chen, X. L. and Ghysels, E., 2011. News-good or bad-and its impact on volatility predictions over multiple horizons. Review of Financial Studies, 24(1), pp.46-81.

Clements, M. P. and Galvao, A.B., 2008. Macroeconomic forecasting with mixed-frequency data: forecasting output growth in the United States. *Journal of Business and Economic Statistics*, 26(4), pp.546-554.

Elhorst, P. Zandberg, E. and de Haan, J., 2013. The Impact of Interaction Effects among Neighbouring Countries on Financial Liberalization and Reform: A Dynamic Spatial Panel Data Approach. *Spatial Economic Analysis*, 8(3), pp.293-313.

Foroni, C. and Marcellino, M., 2014. A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting*, 30(3), pp.554-568.

Foroni, C. and Marcellino, M., 2013. A survey of econometric methods for mixed-frequency data. *Working Paper 2013/06*, Norges Bank.

Ghysels, E. and Ozkan, N., 2015. Real-time forecasting of the US federal government budget: A simple mixed frequency data regression approach. *International Journal of Forecasting*, 31(4), pp.1009-1020.

Ghysels, E., Santa-Clara, P. and Valkanov, R., 2005. There is a risk-return trade-off after all. Journal of Financial Economics, 76(3), pp.509-548.

Ghysels, E., Sinko, A. and Valkanov, R., 2007. MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1), pp.53-90.

Kuzin, V., Marcellino, M. and Schumacher, C., 2013. Pooling versus model selection for nowcasting GDP with many predictors: empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, 28(3), pp.392-411.

Liu, H., 2013. Research and application of China's macroeconomic mixed data model [D]. Jilin University.

Liu, H. and Liu J., 2011. Real-time forecasting and short-term forecasting of China's macroeconomic aggregates: an empirical study based on mixed data forecasting model. *Economic Research Journal* (in Chinese), (03), pp. 4-17.

Liu, J., Liu H. and Yin, Z., 2010. Application of China's Macroeconomic Mixture Data Model—An Empirical Study Based on MIDAS Model. *Economic Science* (in Chinese), (05), pp. 23-34.

Marcellino, M. and Schumacher, C., 2010. Factor MIDAS for nowcasting and forecasting with ragged-edge data: a model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72(4), pp.518-550.

Rondeau, S., 2012. Sources of fluctuations in emerging markets: DSGE estimation with mixed frequency data. *2012 Meeting Papers 1156*.

Wang, X. and Xiao, Z., 2016. A novel soft spatial weights matrix method based on soft sets. *International Journal of Applied Decision Sciences*, 9(1), pp.19-39.

Yu, H. and Huang, Y., 2015. Regional heterogeneity of monetary policy effects, house price spillover effects, and the cross-region impact of house prices on inflation. *Journal of Financial Research* (in Chinese), (02), pp. 95-113.

Yu, X., Li, Y. and Zhang X., 2014. An Empirical Study on Regional Differences in China's Inflation. Statistical Research (in Chinese), (02), pp.55-60.

Zheng, T. and Wang, X., 2013. Mixture data measurement and real-time analysis of the Chinese economic cycle. *Economic Research Journal* (in Chinese), (06), pp.58-70.