Application Analysis of Macroeconomic Data in Short-term Statistical Forecasting

Siyu Wang

College of Social Science, Michigan State University, 48825, USA

Keywords: Macroeconomic data, Predictive model, Countermeasures

Abstract: This paper introduces the research status of short-term statistical forecasting of macroeconomic data, and introduces several representative analytical models. Among them, the prediction method based on the mixed data model is quite typical. On the basis of introducing and combing related research, the above model was used for data analysis, and the corresponding conclusions were drawn. Finally, on the basis of the analysis, the measures and suggestions for using the macroeconomic data to make short-term economic forecast are put forward, and the problems that should be paid attention to in the research are also pointed out.

1. Application status

From the application of short-term statistical forecasting of macroeconomic data, there are mainly two prediction methods. The first is to predict the operation of the next phase of macroeconomic indicators through a single leading indicator or index. The second is to use the time series itself or related indicators to construct an econometric model to predict the next stage of macroeconomic performance.

1.1 Short-term statistical forecast based on leading indicators.

From the current statistical analysis system of the National Bureau of Statistics, there are three main methods for selecting leading indicators. The first is to use traditional statistical indicators as a macroeconomic leading indicator. The second is to use a multi-index synthetic macroeconomic leading index. The third is to calculate the expected index of the company and consumers' judgment on the future situation through the survey of the economy.

Macroeconomic analysis needs to consider the two levels of supply and demand. At the supply level, it is necessary to grasp the operation of the three industries. At the demand level, it is necessary to examine the operation of consumption, investment and foreign trade. From the supply side, the primary industry mainly investigates agricultural production indicators, and its leading indicators are climate change, international food prices and inventory cycles. The secondary industry mainly investigates industrial added value and output of major industrial products. Its leading indicators include manufacturing purchasing managers' index, new order index, industrial fixed asset investment, industrial electricity consumption, bank credit, and social financing scale. The tertiary industry mainly examines indicators such as added value, consumer spending, and real estate development. Its leading indicators mainly include household income, consumption propensity, new housing starts, land acquisition area, and bank credit.

From the demand side, consumption mainly examines consumer goods zeros and consumer spending. Its leading indicators mainly include employment, household income, and consumer confidence. The investment mainly investigates the total investment, real estate development investment and infrastructure construction investment. Its leading indicators mainly include bank credit growth rate, new construction projects and changes in relevant investment policies. Exports mainly focus on major goods and service exports. The leading indicators mainly include cargo throughput at coastal ports, exchange rate changes, consumer confidence in major trading countries, and export tax rebate policies.

  DOI: 10.25236/ssehr.2018.062

Table 1 Macroeconomic leading indicator classification

Total supply			Total demand			Price
Primary industry	**Secondary industry**	**Tertiary Industry**	**Consumption**	**Investment**	**Export**	**Price level**
International food price	Manufacturing fixed asset investment	Consumption propensity	Employment	Bank credit	Port cargo throughput	Domestic economy
Inventory cycle	Industrial electricity consumption	Land purchase area	Residence income	New construction project	PMI order index	Food price
Climate change	Social financing scale	Bank credit	Consumer confidence	Investment policy	Exchange rate changes	International crude oil

The price level is also an important aspect of monitoring macroeconomic operations. The main indicators are the consumer price level, the industrial producer's ex-factory price index, the purchase price index, and the real estate price index. The leading indicators for observing changes in price levels are mainly domestic economy, money and credit, and investment growth.

The third method of selecting the leading indicator is to calculate the prosperity or expected index of the company and the consumer's expectations for the future through the boom survey. The survey of the economy is based on a sample survey method, which is active in business, production, and entrepreneurs and consumers of consumer activities ask questions about the current economic situation and forecast future economic trends.

In addition, with the development of information technology and the Internet, some research institutions have begun to use the network big data to establish a public opinion index for predictive analysis. The advantage of network big data is that the acquisition cost is low, and it also has advantages in terms of timeliness. Ii can make up for the lack of statistics and survey sampling. Building a network public opinion index based on network big data has the advantages of real-time and intelligent. Through the screening and integration of network data, emotional excavation of the economic market can be achieved, and the expected index can be constructed for high-frequency monitoring and real-time prediction.

1.2 Short-term statistical forecast based on statistical and econometric models.

At present, a large number of government agencies, research think tanks, and academic researcher development wells use econometric models for prediction. On the whole, the commonly used models mainly include multiple regression models, time series models, panel data models and general equilibrium model methods. Using the above methods, government departments, research institutions, and scholars have constructed a large number of predictive models.

With the further application of metrology technology, some scholars have even tried the prediction model containing more high-frequency data in addition to the quarterly and monthly data. Ghvsel et al. proposed a mixed data model (MIDAS), which can effectively use high-frequency information in macroeconomic data to avoid information loss in the process of predicting low-frequency indicators and improve the accuracy of prediction.

2. Macroeconomic data short-term statistical forecasting model

2.1 Mixed data conversion

The combined use of mixed data processing and mixed data results does not actually use the information contained in the high frequency data in the mixed data, but it is the most commonly used method in current model construction.

In the empirical study, when time series data of different frequencies are encountered, for high frequency traffic data, it is generally converted into low frequency data by Temporal Aggregation. For stock, data or exponential data, a sum average is used, or a value in the high frequency data is used to represent the GPD of the quarter as a substitute for low frequency data. In the above method of converting high frequency data to low frequency data, the data processed by the time-averaging method inevitably has a problem of loss of high-frequency data information.

2.2 Distribution lag model

The distributed lag model can directly use the mixing data to construct and estimate the model, and the MIDAS model is the distribution. The evolution of the lag model, Ghysels et al. pointed out that the MIDAS model is technically the mathematical expectation under certain conditions of high frequency explanatory variable distribution lag.

Inspired by the construction of the mixed data sampling model, Chen and Tsay proposed the GADL model, and applied the model to combine the daily data of the currency and capital markets to predict the change of the quarterly commodity price, and obtained a good prediction effect. This model has two significant advantages. First, compared to the standard MIDAS model, the second-order exponential weight polynomial of high-frequency data is imposed with a weighted positive constraint. Second, compared with the mixed data sampling model and the state space model, the GADL model is estimated based on the OLS method, so no nonlinear optimization is needed, and the calculation is simple.

2.3 Mixed data model

The mixed data model is a model that can use the mixed data information without loss. It does not need to preprocess the data, and does not need to be over-parameterized. It can also combine a large number of existing advanced measurement methods. The research of this aspect of MIDAS model mainly focuses on the use of high frequency data as an explanatory variable to improve the prediction of low frequency data. Specifically, it can be divided into the following two aspects. The first is to use the monthly data to predict the quarterly macroeconomic aggregates to predict and update the forecast of the quarterly real output growth rate. The second is to use high-frequency daily data to predict monthly or quarterly macroeconomic indicators. For example, some foreign scholars combined high-frequency daily data to predict real GDP growth rates, inflation rates, short-term Treasury bills, and unemployment rates. The results show that MIDAS is better than a simple stochastic baseline forecast.

2.4 Mixing state space model

The research and application of the mixed-state state space model mainly regards the low-frequency data in the mixed data as unobservable high-frequency data. The research and application of the mixed state space model mainly have the following aspects. First, the method uses the mixed data model to estimate the unobservable state of the economy, and is used for the judgment of the transition point. The accuracy of the data model is slightly higher, but the calculation is more complicated and time consuming.

3. Based on the mixed data model annual prediction method

High frequency data often contains more valuable information. However, some of the more critical macro data in the existing research literature at home and abroad have only quarterly or even annual data. Researchers must not abandon these high-frequency data or use some statistical means to convert high-frequency data into low-frequency data when establishing an economic model, affecting the accuracy of the empirical results. Taking GDP data as an example, the highest frequency of CDP data is quarterly data. Therefore, the method of averaging other monthly data into quarterly data is adopted in the regression, which loses the unique information of high frequency data and reduces the accuracy of the estimation result.

The MIDAS model uses a method of nonlinear estimation to measure the degree of response of an interpreted variable to a high-frequency explanatory variable when the explanatory variable differs from the frequency of the interpreted variable. The specific algorithm is as follows:

3.1 Single explanatory variable MIDAS model predicted by forward h step:

$$y_t = \beta_1 B\left(L^{1/m};\theta\right)x_{t-h}^{m} + \varepsilon_t^{(m)} \tag{1}$$

Among them, $B\left(L^{1/m};\theta\right)x_{t-h}^{m} = \sum_{k=1}^{K} b\left(k;\theta\right)L^{(k-1)}, L^{s/m}x_{t-h}^{(m)} = x_{t-h-s/m}^{(m)}$

Let t be the time unit of the interpreted variable and m the frequency of the explanatory variable. If y is quarterly data and x is monthly data, m = 3. Therefore, $B\left(L^{1/m};\theta\right)$ represents the hysteresis operator of the high frequency data, and $b\left(k;\theta\right)$ is the hysteresis coefficient of the explanatory variable. Common forms are: Exponential Almon polynomial, Step Function polynomial, and so on. According to the analysis of Clements and Calvao, this paper will choose the second-order form of log-alternal lag, which is defined as:

$$b\left(k;\theta\right) = \frac{\exp\left(\theta_1 k + \theta_2 k^2\right)}{\sum_{k=1}^{K}\exp\left(\theta_1 k + \theta_2 k^2\right)} \tag{2}$$

Expand the formula (1) to get:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/m};\theta)x_{t-h}^{(m)} + \varepsilon_t^{(m)} \left(1-\lambda L\right)^{-1} B\left(L^{1/m};\theta\right)x_{t-h}^{m} \tag{3}$$

Where 1 and 2 are estimated parameters. According to the data sample popularity, this paper will choose K=12 and K=24 as the model index. Therefore, equation (3) can be written as:

$$\begin{aligned} y_t &= \beta_0\left(1-\lambda\right)^{-1} + \beta_1\left(1-\lambda L\right)^{-1} \\ B&\left(L^{1/m};\theta\right)x_{t-h}^{m} + \left(1-\lambda L\right)^{-1}\varepsilon_t^{(m)} \end{aligned} \tag{4}$$

It can be seen from equations (1) and (3) that when the explanatory variable in the model is T, the explanatory variable's popularity is only T-h. Therefore, by substituting the explanatory variable of the T into the MIDAS model, the predicted value 1 of the T+h phase can be obtained.

3.2 A single explanatory variable forwards the predicted MIDASL model:

The results of Stock and Waston show that the components of the explanatory variables are added to the economic forecasting model, and the prediction results are better than other prediction models. Rewrite equation (1) to get:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/m};\theta)x_{t-h}^{(m)} + \varepsilon_t^{(m)} \left(1-\lambda L\right)^{-1} B\left(L^{1/m};\theta\right)x_{t-h}^{m} \tag{5}$$

It can be seen from equation (5) that the MIDASL model cannot be established by direct regression. Clements and Calvao proposed an iterative estimation method for the MIDAS}L model, as follows:

First, estimate (1) to get the residual term cloud ε_t, let $\lambda_0 = \left(\sum \varepsilon_{t-h}^2\right)^{-1}\sum \varepsilon_t \varepsilon_{t-h}$. Then take $y_t^* = y_t - y_t - \lambda_0 y_{t-h}$ and $x_t^* = x_t - \lambda_0 x_{t-2h}$ back ((1), estimate the sum of the minimum residual sum of squares by multiple iterations. The parameters obtained at this time are the estimated parameters of MIDASL.

4. Questions and Suggestions

The statistical prediction based on a single statistical indicator or a synthetic leading index has the characteristics of being easy to operate. However, there are also obvious problems. On the one hand, the selection of a single indicator or the leading index system has certain subjectivity, and the difference in the selection of indicators has a relatively large impact on the prediction results. On the other hand, the single statistical indicators and the leading index constitute indicators are relatively limited, it is difficult to contain the complete information of the overall macroeconomics, generally can only predict the short-term trend of economic operation, the prediction of specific values is not high.

The advantages of the online public opinion index are reflected in the large amount of data, easy access and immediacy, but the shortcomings are also prominent. There are “impurities” in the massive data, which requires cumbersome data cleaning. In addition, the construction of the “main word” library of public opinion monitoring is still in its infancy, and it will take time to fully apply it.

Short-term statistical forecasts based on statistical and econometric models are relatively rich in data usage and the information obtained is relatively complete. However, such prediction models also have obvious shortcomings. The model parameters are relatively fixed and cannot adapt to changes in the economic structure. It is easy to produce large errors based on past experience predictions.

Based on the above analysis, the short-term statistical forecast for improving macroeconomics in the future should start from the following two aspects:

The first is to establish and improve a short-term statistical indicator system. First, improve the authenticity, accuracy, completeness and timeliness of statistical data. At present, China's statistical system still needs to be improved, and efforts should be made to solve problems such as insufficient statistical scope and insufficient scientific standards. With the in-depth development of the new statistical work of “new industries, new formats, new business models”, new questions and new requirements have been put forward for the improvement of the statistical indicator system. Second, we should promptly respond to periodic and structural changes in economic operations and adjust the statistical indicator system in a timely and appropriate manner. On the other hand, it is also necessary to combine the actual situation and it is not appropriate to change the existing statistical indicators in the statistical caliber, scope and other systems. At the same time, there is room for further improvement in the release of statistical data. The disconnection of some statistical indicators is one of the main factors for the statistical monitoring and forecasting indicator system is not perfect. Third, in-depth exploration of the application of "big data" in short-term statistical forecasting work, speed up the construction of relevant "main words" library of public opinion monitoring, and establish a reliable and practical public opinion monitoring and forecasting system.

The second is to combine the first indicator monitoring, the prosperity survey, and the measurement model prediction, and at the same time give full play to the advantages of the prosperity survey and measurement forecast. On the basis of fully considering the external impact, the situation is decomposed into the prediction results to enhance the prediction flexibility. At the same time, in the process of statistical or quantitative modeling, not only the use of conventional statistical data, but also the advantages of the survey data should be fully utilized to better analyze the structure of sub-regions and sub-sectors.

References

[1] Bestelmeyer G, Breunsbach M, Hess D. Expected Risk Premium Dynamics Across the Business Cycle and the Stock Market’s Response to Macroeconomic News[J]. Social Science Electronic Publishing, 2011.

[2] Deininger K, Squire L, Basu S. Does Economic Analysis Improve the Quality of Foreign Assistance?[J]. World Bank Economic Review, 1998, 12(3):385-418.

[3]Meier, P., Tilmant, A., Boucher, M., & Anctil, F. (2012). Real-time Reservoir Operation Based on a Combination of Long-term and Short-term Optimization and Hydrological Ensemble Forecasts. AGU Fall Meeting (Vol.52, pp.965-972). AGU Fall Meeting Abstracts.

[4] Duca J V, Vanhoose D D. Goods-market competition and profit sharing: a multisector macro approach[J]. Journal of Economics & Business, 1998, 50(6):525-534.

[5] Qin J, Leonard M, Kuczera G, et al. A high-resolution hierarchical space-time framework for

single storm events and its application for short-term rainfall forecasting.[C]// New Approaches To Hydrological Prediction in Data-Sparse Regions. Proceedings of Symposium Hs.2 at the Joint Convention of the International Association of Hydrological Sciences. 2009:330-340.

[6] Meier P, Tilmant A, Boucher M, et al. Real-time Reservoir Operation Based on a Combination of Long-term and Short-term Optimization and Hydrological Ensemble Forecasts[C]// AGU Fall Meeting. AGU Fall Meeting Abstracts, 2012:965-972.