

Prediction of CPI in China based on MIDAS

Yujing Yin

Beijing Technology and Business university, Beijing, China, 100048

584100240@qq.com

Abstract. Over the past few decades, accurate forecasting of CPI is crucial for government policy making and corporate strategic planning. The purpose of this study is to utilize the mixed data sampling (MIDAS) regression method to forecast the CPI in China and to compare its forecasting performance with the traditional autoregressive moving average (ARIMA) model. We demonstrate that the MIDAS regression model is a promising approach for capturing and forecasting the complex dynamics of macroeconomic indicators. With this approach, policymakers and entrepreneurs can obtain more accurate and comprehensive CPI forecasts and thus make more informed decisions

Keywords: MIDAS regression, ARIMA modeling, CPI forecasting, time series analysis.

1. Introduction

In the context of globalization, the Consumer Price Index (CPI) is regarded as a pivotal indicator for evaluating the economic health of a nation or region. As the largest developing economy in the world, fluctuations in China's CPI not only have profound implications for its domestic economy but also influence the stability and growth of the global economy. Hence, accurate and timely forecasting of China's CPI becomes of paramount importance.

In recent years, with advancements in data acquisition and increased computational capacities, economists and policymakers have gravitated towards more sophisticated and intricate forecasting models. Mixed Data Sampling (MIDAS) regression has emerged as a promising tool, capable of merging data of varying frequencies to enhance predictive accuracy. The strength of MIDAS regression lies in its ability to address heterogeneity and non-linearity, rendering it a powerful and flexible predictive instrument.

Given China's economic backdrop, MIDAS regression holds potential in integrating macroeconomic indicators such as industrial output, employment rates, and monetary policy shifts to predict the CPI. This approach not only bolsters predictive accuracy but also furnishes policymakers with a more detailed foundation for decision-making.

While MIDAS regression has seen successful applications in various research domains, its deployment for predicting China's CPI remains relatively sparse. Considering China's unique economic framework and policy milieu, exploring the applicability and potential efficacy of this method in the Chinese context is crucial.

This study aims to employ MIDAS regression for predicting China's CPI, delving into the advantages and limitations of this methodology. We plan to construct a predictive model utilizing the latest data and techniques and juxtapose it against traditional methodologies to gauge its effectiveness.

We anticipate that MIDAS regression can capture, with enhanced precision, the myriad factors influencing CPI and their intricate interplays. Furthermore, we believe that the model, with its rich data input, will offer a more comprehensive perspective on CPI forecasting.

Through an in-depth analysis and prediction of China's CPI, we aspire to provide policymakers and investors with a pragmatic tool. Additionally, we aim to lay the groundwork for subsequent research, further probing the prospective applications of MIDAS regression in other areas.

To realize these objectives, this paper will first delve into the theoretical underpinnings of MIDAS regression and pertinent literature. Subsequently, we will detail our research data and methods, followed by the presentation of our predictive outcomes. Concluding the paper, we will discuss our findings, their policy implications, and propose directions for future research.

2. Literature review

MIDAS regression, an abbreviation for Mixed Data Sampling Regression, is a sophisticated method adept at processing data spanning various frequencies. Since its inception by Eric Ghysels[1][2] and his team in the early 2000s, MIDAS (Mixed Data Sampling) regression has become a widely used method in macroeconomic and financial forecasting (Ghysels et al., 2004; Ghysels et al., 2007). This approach allows researchers to integrate different frequencies of data in a single model, aiming to improve the accuracy and efficiency of forecasting. One of the pivotal advantages of the MIDAS regression model is its prowess in harmonizing disparate data frequencies. This capability harnesses the richness of high-frequency data to bolster prediction accuracy (Clements and Galvão, 2008)[3]. Moreover, it bestows researchers with the latitude to model interrelationships between variables of different frequencies in a nuanced manner, capturing intricate economic shifts (Zeynalov and Ayaz, 2017)[4]. Clements and Galvão (2008)[3] explore the utility of MIDAS regressions in terms of whether they can improve U.S. output growth. Using a novel methodology that incorporates autoregressive terms, they validate the effectiveness of MIDAS regression in economic forecasting. In addition, Joye Khoo (2021)[5] further demonstrates the usefulness of MIDAS regressions in dealing with corporate finance under geopolitical influences, emphasizing its potential value in modern economic analysis. Breitung and Røling (2015)[6] used nonparametric MIDAS regression to forecast inflation in their study, and they argued that nonparametric estimator can provide a more reliable and flexible approximations to the actual lag distribution than the conventional parametric MIDAS approach based on exponential lag polynomials. In addition, in their study, Clements and Galvão (2008)[3] explore how MIDAS regressions can help improve forecasting on real-time data, showing that the method has a range of applications in macroeconomic forecasting.

The study of China's CPI has been an important topic. Traditional research methods rely on time series analysis and statistical models such as ARIMA and GARCH models (Li et al., 2020)[7][8]. However, in recent years, more and more researchers have begun to try to use more advanced techniques, such as machine learning to forecast Chinese CPI (Wang, 2012)[9]. These studies usually try to improve the accuracy and complexity of the predictions by including more information and data. Although MIDAS regression has been successfully applied in several countries and fields, its application to CPI prediction in China is relatively rare. Considering the special economic environment and data structure of China, exploring the application of MIDAS regression in Chinese CPI forecasting becomes a meaningful research direction. In synthesizing insights from both global and regional studies, it's evident that while MIDAS regression has made a big splash in the field of macroeconomic forecasting, its footprint in China's CPI forecasting landscape is still emerging. Through our forthcoming research endeavors, we aim to navigate this uncharted territory, offering a refreshed perspective to comprehend and prognosticate China's intricate CPI dynamics.

3. Models and variables

3.1 MIDAS Mixed Frequency Modeling

The MIDAS (Mixed Data Sampling) regression model was originally designed to address the issue of effective integration when confronted with data of different frequencies[10]. Unlike traditional models, it allows us to integrate data of different frequencies under a unified modeling framework, which helps to capture more information and improve the accuracy of predictions. Type the formula here.

The core idea of the MIDAS model is to capitalize on the informative nature of high-frequency data while maintaining a low-frequency model structure. Its typical mathematical representation is:

$$Y_t = \beta_0 + \beta_1 + M(X_{t,h}, \theta) + \epsilon_t \quad (1)$$

where Y_t is low-frequency response variables, such as monthly data, $X_{t,h}$ is high-frequency predictor variables, such as daily data, β_0, β_1 is parameters to be estimated, $M()$ is a hybrid data

sampling weight function, whose parameters are controlled by θ , which is responsible for weighting the high-frequency data to generate predictions for low frequencies, and ϵ_t is the error term.

The choice of weighting function is crucial because it determines how to extract information from high frequency data. Common weighting functions include polynomial lag function, exponential lag function, and so on. With these functions, the MIDAS model can capture short-term dynamics in high-frequency data while maintaining the model structure at low frequencies.

The time series model is constructed based on the CPI characteristics, and we can consider the lag effect of the time series. Lag selection is a critical step before using the MIDAS model because different lag choices may affect the forecasting performance of the model.

Lag selection becomes particularly important considering that the MIDAS model can handle data with different frequencies, especially the lag effects of high-frequency data on low-frequency response variables. In order to determine the optimal number of lags, we utilize statistical methods such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)[11].

In the lag selection process, we considered from one lag to multiple lags to determine which number of lags provides the best fit for our MIDAS model.

3.2 Benchmark model

In forecasting CPI, the benchmark model is usually a simple ARIMA model. The forecasting accuracy is compared to the benchmark model.

An ARIMA(p,d,q) model is given by:

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d Y_t = (1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t \quad (2)$$

where L is the lag operator, and ϕ_i , θ_i , ϵ_t are the parameters of the autoregressive part, moving average part and error terms.

3.3 Data Description.

We obtained the daily and monthly frequency data from January 2016 to March 2023. The data were obtained from the National Bureau of Statistics of China (NBS) and CHOICE databases to ensure the accuracy and reliability of the data. The variables are selected as in Table 1.

Table 1

Frequency	Variable Name	Symbol	Description
Monthly	CPI year-on-year growth rate	y	Dependent Variable
Monthly	PPI Year-on-Year Growth Rate	x1	Represents changes in production prices. An indicator of potential inflationary pressures, which can be passed on to consumers.
Monthly	Total Retail Sales of Consumer Goods Cumulative Growth Rate	x2	Reflects consumer spending and is another key indicator of CPI growth.
Daily	Average Wholesale Price of Agricultural Products: 28 Key Vegetables	x3	Captures short-term price fluctuations which can have a direct impact on the overall CPI.
Daily	Average Wholesale Price of Agricultural Products: 7 Key Fruits	x4	Acts as a barometer for short-term food inflation, a significant component of CPI.
Daily	Average Wholesale Produce Prices: 7 Key Meats	x5	Meat prices can significantly impact food inflation, thereby influencing the broader CPI.

Before modeling the mixing model, we first performed a series of data preprocessing steps to ensure the quality and applicability of the data. For the missing values in the dataset, we used linear interpolation to preprocess the data. Also, since our dataset includes both daily and monthly frequency data, we ensured that all the data were properly aligned with the monthly CPI data.

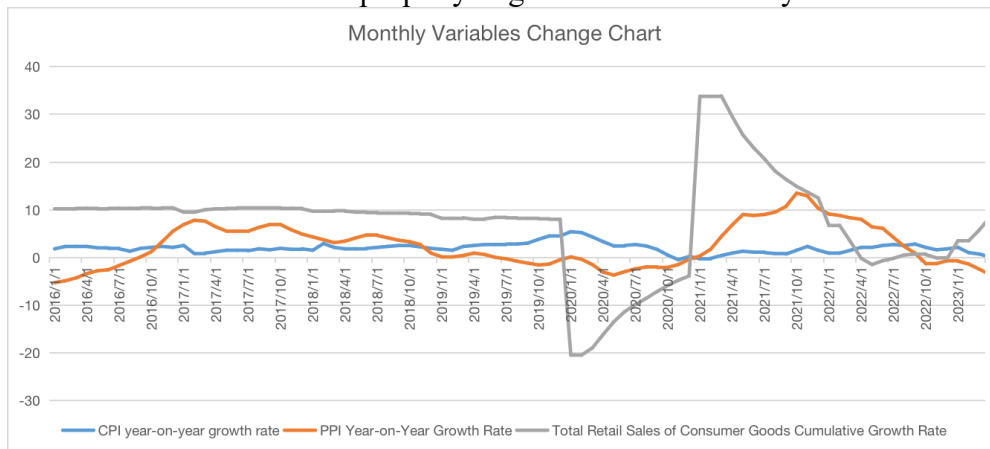


Figure 1. Monthly Frequency Variables Over Time

In Figure 1, we plot the monthly frequency variables, CPI year-on-year growth rate(y), PPI Year-on-Year Growth Rate (x1) and Total Retail Sales of Consumer Goods Cumulative Growth Rate (x2). This visualization provides an overview of their trends and fluctuations over the years, highlighting potential periods of economic stability or volatility.

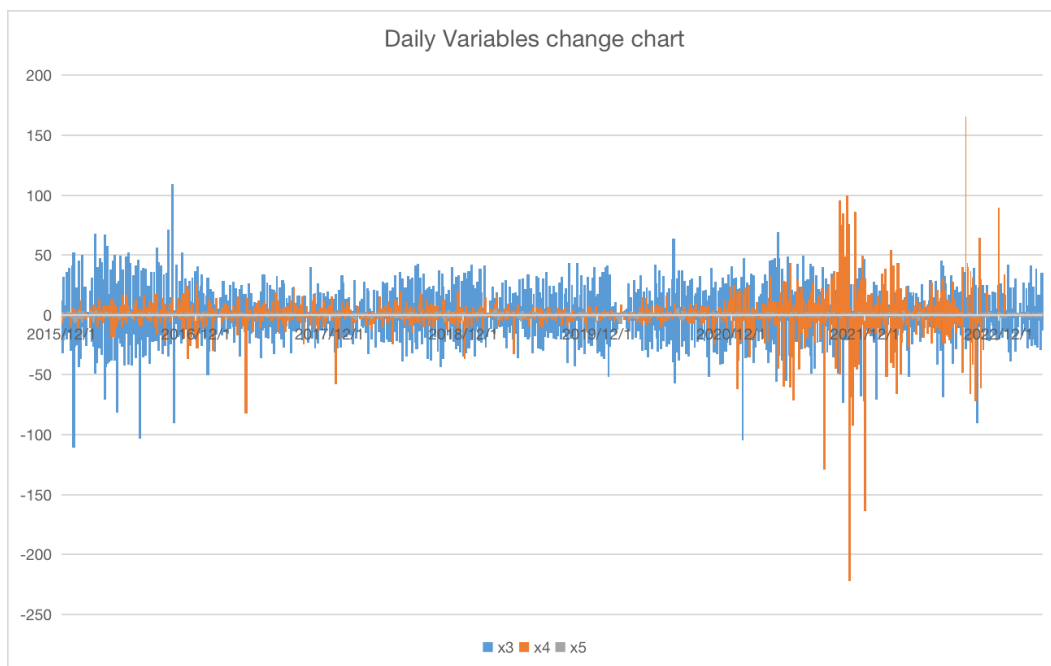


Figure 2. Daily Frequency Variables Over Time

Figure 2 presents a bar chart of daily frequency variables, including the Average Wholesale Price of Agricultural Products for key vegetables (x3), key fruits (x4), and key meats (x5). The graph provides a fine-grained insight into short-term variations, which is critical when using MIDAS regressions to forecast CPI dynamics.

4. Predicted results

We first predicted the year-on-year CPI growth rate using the MIDAS mixing model and ARIMA model. Finally, the predictive accuracy of the models is verified by comparing the RMSE values. The

model with lower RMSE values indicates that the model has better predictive accuracy than other models with higher RMSE values.

The full sample is divided into the in-of-sample forecasting period from 2016M1 to 2022M12 (84 months) and the out-of-sample forecasting period from 2023M1 to 2023M3 (3 months).

The key findings of the MIDAS model and ARIMA model in- and out-of-sample predictions can be seen in figure 3.

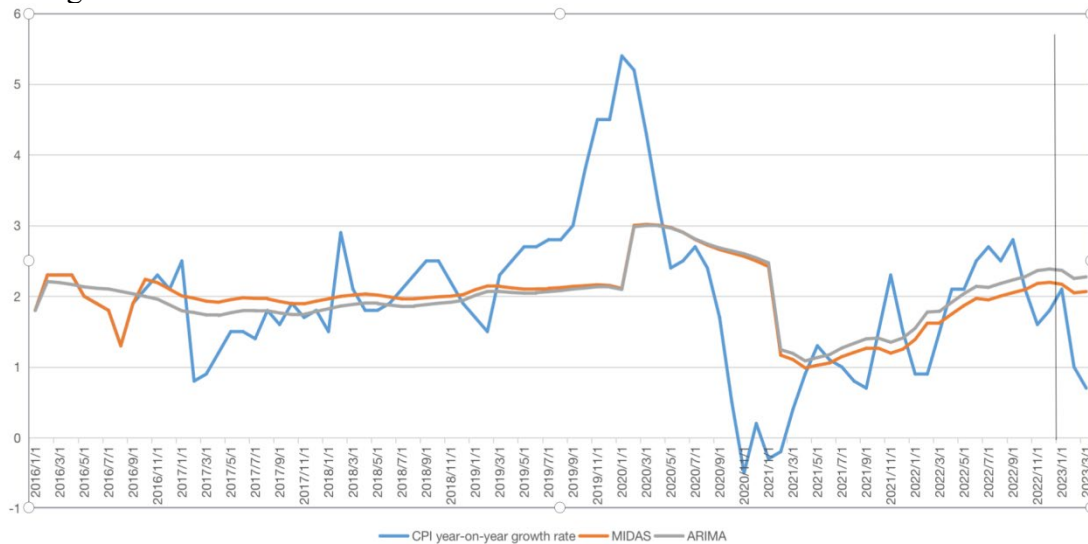


Figure 3. MIDAS and ARIMA in- and out-of-sample predictions Comparison Graph

The comparison graphs clearly show the performance of the two models in both in-sample and out-of-sample predictions. Ideally, both models should be close to the real data in the in-sample prediction, but in the out-of-sample prediction, the MIDAS model may be more accurate due to its ability to handle high frequency data.

In order to compare the two models more accurately, we can compare the RMSE of the two models, and the specific values can be found in Table 2.

Table 2

	MIDAS	ARIMA
RMSE(In-of-sample)	0.954990	0.967935
RMSE(Out-of-sample)	0.995532	1.171349

From the table we can clearly see that the RMSE of the MIDAS model is smaller than the ARIMA model for both in-sample and out-of-sample predictions. Consequently, the MIDAS model performs better in forecasting the year-on-year CPI growth rate compared to the ARMA model. This may be because Midas model can integrate more information sources and capture more complex market dynamics.

5. Conclusion

After in-depth research and analysis, this study successfully employs the MIDAS regression method to forecast China's CPI. By integrating data with different frequencies - daily and monthly - we can construct a more comprehensive and accurate forecasting model.

The results show that the MIDAS regression model is more accurate in predicting China's CPI than the traditional ARIMA model. This suggests that MIDAS regression provides more accurate and reliable results when analyzing a combination of high-frequency and low-frequency data.

In addition, we note that Midas regression model enables us to better capture the dynamic relationship between macroeconomic indicators and CPI, thus providing more valuable insights for decision makers and investors.

However, our study also reveals some limitations, especially in terms of data availability and model complexity. The establishment and estimation of MIDAS regression model requires a large number of high-quality data, which may be a limiting factor in some cases. In addition, although Midas model shows advantages in processing data with different frequencies, it also increases the complexity of the model and requires more time and technical expertise to be applied.

Nonetheless, our research has opened up a promising path, indicating that the MIDAS regression model can become a powerful tool for predicting China's CPI. Future research can further explore how to optimize the performance of models and how to apply this method to predict other economic or financial problems.

References

- [1] Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. "The MIDAS touch: Mixed data sampling regression models." (2004).
- [2] Ghysels, Eric, Arthur Sinko, and Rossen Valkanov. "MIDAS regressions: Further results and new directions." *Econometric reviews* 26.1 (2007): 53-90.
- [3] Clements, Michael P., and Ana Beatriz Galvão. "Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States." *Journal of Business & Economic Statistics* 26.4 (2008): 546-554.
- [4] Zeynalov, Ayaz. "Forecasting tourist arrivals in Prague: Google econometrics." (2017).
- [5] Khoo, Joye. "Does geopolitical uncertainty affect corporate financing? Evidence from MIDAS regression." *Global Finance Journal* 47 (2021): 100519.
- [6] Breitung, Jörg, and Christoph Roling. "Forecasting inflation rates using daily data: A nonparametric MIDAS approach." *Journal of Forecasting* 34.7 (2015): 588-603.
- [7] Tinghui Li, Junhao Zhong & Zimei Huang (2020) Potential Dependence of Financial Cycles between Emerging and Developed Countries: Based on ARIMA-GARCH Copula Model, *Emerging Markets Finance and Trade*, 56:6, 1237-1250.
- [8] Fang T, Lee T H, Su Z. Predicting the long-term stock market volatility: A GARCH-MIDAS model with variable selection[J]. *Journal of Empirical Finance*, 2020, 58: 36-49.
- [9] Wang, Ye, Bo Wang, and Xinyang Zhang. "A new application of the support vector regression on the construction of financial conditions index to CPI prediction." *Procedia Computer Science* 9 (2012): 1263-1272.
- [10] Ghysels, Eric, Virmantas Kvedaras, and Vaidotas Zemlys-Balevičius. "Mixed data sampling (MIDAS) regression models." *Handbook of Statistics*. Vol. 42. Elsevier, 2020. 117-153.
- [11] Acquah, Henry de-Graft. "Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship." (2010).