mathematics

Article

Effectiveness of Principal-Component-Based Mixed-Frequency Error Correction Model in Predicting Gross Domestic Product

Yunxu Wang [1], Chi-Wei Su [1], Yuchen Zhang [2,*], Oana-Ramona Lobonţ [3] and Qin Meng [1]

[1] International Business School, Qingdao Huanghai University, Qingdao 266555, China; wangyx@qdhhc.edu.cn (Y.W.); su.chih@e-uvt.ro (C.-W.S.); magiccubemedia@hhu.edu.cn (Q.M.)
[2] Beijing Public Transport Corporation, Beijing 100161, China
[3] Finance Department, West University of Timisoara, 300223 Timisoara, Romania; oana.lobont@e-uvt.ro
* Correspondence: yc4026@columbia.edu; Tel.: +86-18604889287

Abstract: As an important indicator that can reflect a country's macroeconomic situation and future trend, experts and scholars have long focused on analyses and predictions of gross domestic product (GDP). Combining principal component analysis (PCA), the mixed-frequency data sampling (MIDAS) model and the error correction model (ECM), this investigation constructs the principal-component-based ECM-MIDAS and co-integration MIDAS (CoMIDAS) models, respectively. After that, this investigation uses the monthly consumption, investment and trade data to build a mixed-frequency model to predict quarterly GDP. The empirical results can be summarized as follows: First, the predictive effectiveness of the mixed-frequency model is better than that of the same-frequency model. Second, the three variables have a strong correlation, and applying the principal component idea when modelling the same and mixed frequencies can lead to more favourable predictive effectiveness. Third, adding an error correction term to the principal-component-based mixed-frequency model has a significant coefficient and a higher predictive accuracy. Based on the above, it can be concluded that combining the MIDAS model with error correction and a principal component is effective; thus, this combination may be applied to support real-time and accurate macroeconomic prediction.

Keywords: mixed-frequency data; principal component analysis; error correction model; predictive effectiveness

MSC: 91-08

Citation: Wang, Y.; Su, C.-W.; Zhang, Y.; Lobonţ, O.-R.; Meng, Q. Effectiveness of Principal-Component-Based Mixed-Frequency Error Correction Model in Predicting Gross Domestic Product. *Mathematics* **2023**, *11*, 4144. https://doi.org/10.3390/math11194144

Academic Editor: Manuel Alberto M. Ferreira

Received: 1 September 2023
Revised: 26 September 2023
Accepted: 28 September 2023
Published: 30 September 2023

1. Introduction

This investigation explores the predictive effectiveness of the mixed-frequency data sampling (MIDAS) model with error correction and a principal component. As gross domestic product (GDP) is an important indicator reflecting the macroeconomic operations of countries, the prosperity or recession of the national economy is closely related to this indicator. The time lag of macroeconomic policy means its formulation must be forward-looking, to which end the analysis and prediction of GDP are particularly essential. GDP is released by the National Bureau of Statistics as quarterly data, while the variables used to predict GDP are mainly monthly data, such as consumption, investment, trade, inflation rate and money supply. Due to the difference in data frequency, mixed-frequency data must be processed before establishing the traditional model. The summing method, average approach and alternative technique must be precisely applied to convert the monthly data into quarterly data [1,2] and then build the same-frequency model to predict quarterly GDP. However, these conversion methods may cause a loss of information or a simulated increase, lowering the predictive effectiveness of GDP. Against the background that the current predictive method and accuracy need to be enhanced, the MIDAS model and its expanded forms may be employed to predict quarterly GDP in a more timely and accurate manner, thus supporting real-time and effective macroeconomic policy formulation.

In order to overcome the limitations of the same-frequency models, scholars have proposed mixed-frequency models that can be modelled based on mixed-frequency data without processing the original sequences. Among them, the MIDAS model is widely used, and existing studies have expanded its modelling theory and application field. Ghysels et al. [3,4] developed the MIDAS model based on the modelling theory of the autoregressive distributed lag (ARDL) model, which constructs a mixed-frequency model by giving several weights to high-frequency explanatory variables and puts forward the corresponding weight functions (e.g., almon, beta and step weight functions). Although the existing studies have explored the MIDAS model extensively, the research on the MIDAS model with error correction is insufficient, and few scholars have considered it from the perspective of the principal component. Hence, the contributions of this investigation are primarily threefold: To begin with, we consider the strong correlation among consumption, investment and trade sequences selected in this paper, which was ignored in previous studies. In order to eliminate the estimated bias caused by the collinearity problem, we utilise the principal component analysis (PCA) method to extract the principal component of these three variables, and build the MIDAS model to probe whether the mixed-frequency model based on the principal component is effective. Secondly, the extant research hardly considers the adjustment of the long-term relationship by using the short-term one, in which process some data information is lost. Thus, relying on the error correction is another novelty of this study, in which the predictive effectiveness of the principal-component-based mixed-frequency error correction model is verified by constructing the ECM-MIDAS and CoMIDAS models and taking the extracted principal component as the predictor of GDP. Thirdly, through quantitative analysis, we further improve the modelling theory of the mixed-frequency model and provide new ideas for more accurate GDP prediction, which constitute relevant insights into how related authorities might better predict GDP and formulate macroeconomic policies.

The investigation is organised as follows: Section 2 reviews the extant literature. The materials and methods are introduced in Section 3. Sections 4 and 5 present the results and discussion. Finally, the conclusions are drawn in Section 6.

2. Literature Review

As GDP prediction is essential to the development of a country, existing studies have predicted the GDPs of various countries and regions using different methods, falling into two categories. One is to forecast GDP based on its lag periods, such as with the autoregressive moving average (ARIMA) model [5,6], grey prediction model and its extended form [7,8], BP neural network model [9], etc. But these methods have certain downsides; therefore, some scholars have included GDP-related variables in the prediction, mainly employing same-frequency data [10,11]. In order to avoid information loss, Ghysels et al. [3,4] proposed the MIDAS model, which was initially designed to analyse and predict the stock market's volatility based on mixed-frequency data. Since then, many researchers have used this model to analyse the stock market [12,13]. Although the MIDAS model is effective in analysing the volatility of the stock market, since Engle et al. [14] proposed the generalised autoregressive conditional heteroscedasticity MIDAS (GARCH-MIDAS) model and Colacito et al. [15] developed the dynamic-condition-associated MIDAS (DCC-MIDAS) model, researchers have been more inclined to use these two methods to analyse the stock market [16–22].

Then, Clements and Galvao [23] produced the MIDAS model with autoregression terms (MIDAS-AR), such as GDP, thereby solving the sequences using auto-correlation. They provided evidence that the addition of autoregressive terms to the MIDAS model makes it more effective at predicting quarterly GDP growth in the U.S. based on monthly indicators. Since its launch, scholars have made more use of the MIDAS model to predict the GDPs of various countries. Through in-sample and out-of-sample empirical analyses, Hogrefe [24] proved that a model built based on mixed-frequency data could improve the revision of GDP prediction. According to the modelling theory of the MIDAS model, Andreou et al. [25] offered evidence that the predictive effectiveness for quarterly GDP is improved after the addition of daily financial indicators. Aprigliano et al. [26] predicted the quarterly GDP growth rate of the eurozone based on monthly and daily indicators and suggested that combining the unconstrained MIDAS (U-MIDAS) model with a smaller mean square error and larger weight produces higher predictive accuracy. Fu et al. [27] suggested that the MIDAS model has a smaller root mean squared error (RMSE) than the VAR system in short-term forecasting, which provides more stable real-time predictions and short-term forecasts of quarterly GDP growth rates in China. Mishra et al. [28] found that the values of RMSE were low in their sample and when predicting the out-of-sample one- and four-quarter horizons, while RMSE increased if predicting the ten-quarter horizon. In addition, Chikamatsu et al. [29], Pan et al. [30], Chernis et al. [31], Xu et al. [32], Jiang et al. [33], Pettenuzzo et al. [34], Barsoum and Stankiewicz [35] and Degiannakis [36] also ascertained the effectiveness of MIDAS model in forecasting from different perspectives.

However, in the construction process of the above mixed-frequency models, a growth rate sequence or first-order difference series is adopted to avoid the "spurious regression" problem in the smooth modelling process, which means part of the important information is lost. In order to overcome this problem, Miller [37] contributed the idea of co-integration into the MIDAS model and, accordingly, constructed the co-integration MIDAS (CoMIDAS) model for predicting the variables of real global economic activities. Gotz et al. [38], meanwhile, added an error correction item to the MIDAS model and constructed the error correction MIDAS (ECM-MIDAS) model for forecasting the monthly inflation rate of the U.S. Their results showed that the ECM-MIDAS and CoMIDAS models had improved predictive effectiveness.

3. Materials and Methods

3.1. Data

This investigation was based on the GDP in China (denoted as GDP) from the first quarter of 1998 to the second quarter of 2023, and the total retail sales of consumer goods (denoted as C), completed investment in fixed assets (denoted as I) and total import and export (denoted as T) from January 1998 to June 2023. All the sequences were taken from the National Bureau of Statistics of China, and the descriptive statistics and trends are shown in Table 1 and Figure 1. In Table 1, the mean values of GDP, C, I and T are 115611.0, 15179.32, 22230.41 and 2196.215, indicating that the selected variables were concentrated on these four levels. The skewness of GDP, C, I, and T is positive, highlighting that these sequences had right-skewed distributions. Their kurtosis is less than 3, which indicates the features of low peaks and thin tails. In addition, the Jarque–Bera test result evidences that GDP, C, I and T conformed to the standard normal distributions at the 1% level.

Table 1. The descriptive statistics of GDP, C, I and T.

	GDP	**C**	**I**	**T**
Observations	114	342	342	342
Frequency	quarterly	monthly	monthly	monthly
Mean	115,611.0	15,179.32	22,230.41	2196.215
Median	85,683.55	9985.000	13,419.33	2055.775
Maximum	335,507.9	41,268.90	82,409.00	5861.160
Minimum	12,111.70	1505.300	221.835	156.890
Standard Deviation	94,142.31	12,917.45	21,911.39	1611.037
Skewness	0.676	0.593	0.724	0.291
Kurtosis	2.182	1.854	2.166	1.872
Jarque–Bera	11.863 ***	38.735 ***	39.766 ***	22.968 ***
Probability	0.003	0.000	0.000	0.000

Notes: GDP is quarterly gross domestic product in China, C is monthly total retail sales of consumer goods in China, I is monthly completed investment in fixed assets in China, and T is monthly total import and export in China. *** denotes significance at the 1% level.

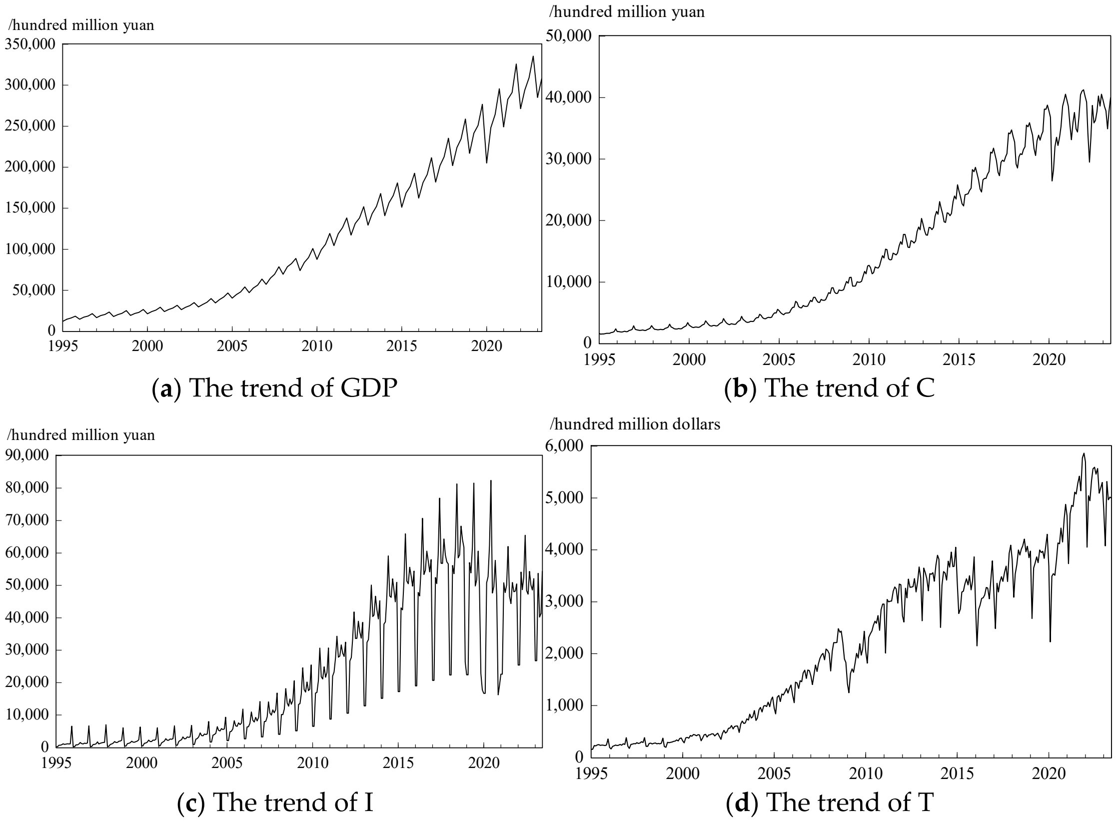

Figure 1. The trends of GDP, C, I and T. Notes: GDP is quarterly gross domestic product in China, C is monthly total retail sales of consumer goods in China, I is monthly completed investment in fixed assets in China, and T is monthly total import and export in China.

It can be observed from Figure 1 that GDP, C, I and T had distinct seasonal effects; thus, we applied the X-12 seasonal adjustment technique [39] to eliminate the seasonal impact on these four sequences, to ensure accuracy. Furthermore, we conducted transformations in GDP, C, I and T in this investigation, which were completed by taking the natural logarithm after seasonal adjustments (denoted as LNGDP, LNC, LNI and LNT) in order to avoid the negative effect of excessive unusual fluctuations. The first-order difference sequences of the above data are denoted as DGDP, DC, DI and DT. In addition, this investigation took the monthly data from February 1998 to June 2021 and the quarterly data from the second quarter of 1998 to the second quarter of 2021 as the in-sample data with which to build models, while the quarterly data from the third quarter of 2021 to the second quarter of 2023 was used as the out-of-sample data for prediction. Then, we used the RMSE to measure the predictive effectiveness of mixed-frequency models [40–42], which can be written with the following formula.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{T}\left(\Delta\mathrm{DGDP}_i - \Delta\mathrm{DG\hat{D}P}_i\right)^2}{T}} \tag{1}$$

where i is the number of periods to predict DGDP, and its value ranges from 1 to T. T is the highest number of periods, which equals 8 in this investigation; for example, i = 8 refers to the prediction of DGDP for the second quarter of 2023. In addition, $\Delta\mathrm{DGDP}_i$ and $\Delta\mathrm{DG\hat{D}P}_i$ are the real and predicted values for period i, respectively.

3.2. Methodology

The MIDAS model produced by Ghysels et al. [3,4] is mainly used to construct the mixed-frequency data in the same model without co-frequency processing. After that, the low-frequency explained variables can be estimated and predicted based on the high-frequency explanatory ones. Since the MIDAS model can then make full use of different-frequency data, it is conducive to enhancing predictive effectiveness.

3.2.1. Basic MIDAS Model

$$Y_t = \beta_0 + \beta_1 B\left(L^{1/m};\theta\right) X_t^{(m)} + \varepsilon_t \tag{2}$$

where $X_t^{(m)}$ is the explanatory variable (e.g., high-frequency data), Y_t is the explained variable (e.g., low-frequency data) and m represents the frequency multiplier between explanatory and explained sequences. This investigation set $X_t^{(m)}$ as the monthly data and Y_t as the quarterly data; thus, the value of m was 3. $B\left(L^{1/m};\theta\right) = \sum_{k=1}^{K} \omega(k;\theta) L^{(k-1)/m}$, where $\omega(k;\theta)$ is the weight function, $K-1$ is the highest lag order calculated at the high frequency, and $L^{(k-1)/m} X_t^{(m)} = X_{t-(k-1)/m}^{(m)}$ is the lag operator. β_0 and β_1 are coefficients, ε_t is the random error term.

Specifically, we set Y_t as the data of the fourth quarter of a certain year. When $K = 6$, there were six high-frequency explanatory variables, which were the monthly data from July to December of the year ($X_{t-5/3}^{(3)}$, $X_{t-4/3}^{(3)}$, $X_{t-1}^{(3)}$, $X_{t-2/3}^{(3)}$, $X_{t-1/3}^{(3)}$, $X_t^{(3)}$). Then, the MIDAS model could be rewritten as the following formula.

$$Y_t = \beta_0 + \beta_1 \left[\omega(1;\theta) X_t^{(3)} + \omega(2;\theta) X_{t-1/3}^{(3)} + \ldots + \omega(6;\theta) X_{t-5/3}^{(3)}\right] + \varepsilon_t \tag{3}$$

3.2.2. MIDAS Model for h-Step Forward Prediction

The introduction of h-step forward prediction into the MIDAS model enabled it to revise the previous prediction according to the latest published high-frequency data, which solved the problem where the predictive effectiveness was weakened due to the time lag of data publication in the same-frequency model [43–45]. This model has the following form:

$$Y_t = \beta_0 + \beta_1 B\left(L^{1/m};\theta\right) X_{t-h/m}^{(m)} + \varepsilon_t \tag{4}$$

When $h = 1$, it is one-step forward prediction. We still set $K = 6$, since the highest lag order of high-frequency data was 6 instead of 5; this was different from Equation (3). The six high-frequency explanatory variables were monthly data from June to November of that year, which can be expressed as:

$$Y_t = \beta_0 + \beta_1 \left[\omega(1;\theta) X_{t-1/3}^{(3)} + \omega(2;\theta) X_{t-2/3}^{(3)} + \ldots + \omega(6;\theta) X_{t-2}^{(3)}\right] + \varepsilon_t \tag{5}$$

3.2.3. MIDAS Model with Autoregression Terms

Generally, economic variables (e.g., GDP) may have strong auto-correlation; therefore, the low-frequency variables with lag order (Y_{t-j}) should be introduced into the MIDAS model to further improve its predictive effectiveness [23]. This is expressed as Equation (6), where p and γ_j are the optimal lag order and coefficient.

$$Y_t = \beta_0 + \sum_{j=1}^{p} \gamma_j Y_{t-j} + \beta_1 B\left(L^{1/m};\theta\right) X_{t-h/m}^{(m)} + \varepsilon_t \tag{6}$$

3.2.4. Weight Functions of the MIDAS Model

- Beta weight function

$$\omega(k;\theta_1,\theta_2) = \frac{f(k/K;\theta_1,\theta_2)}{\sum_{k=1}^{K} f(k/K;\theta_1,\theta_2)} \tag{7}$$

where $f(X;a,b) = \frac{X^{a-1}(1-X)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$, $\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx$.

- BetaNN weight function

$$\omega(k;\theta_1,\theta_2,\theta_3) = \frac{f(k/K;\theta_1,\theta_2)}{\sum_{k=1}^{K} f(k/K;\theta_1,\theta_2)} + \theta_3 \quad (8)$$

θ in Equation (8) cannot be zero, while it could be zero in Equation (7), which was obtained if $\theta_3 = 0$.

- Expalmon weight function

$$\omega(k;\theta_1,\theta_2,\ldots,\theta_P) = \frac{\exp\left(\theta_1 k + \theta_2 k^2 + \ldots + \theta_p k^P\right)}{\sum_{k=1}^{K} \exp\left(\theta_1 k + \theta_2 k^2 + \ldots + \theta_p k^P\right)} \quad (9)$$

This analysis set P = 2, and its general constraints were $\theta_1 \leq 300$, $\theta_2 < 0$.

- Almon weight function

$$\beta_1\omega(k;\theta_1,\theta_2,\ldots,\theta_P) = \theta_0' + \theta_1' k + \theta_2' k^2 + \ldots + \theta_p' k^P \quad (10)$$

This weight function is the most general form, which can be estimated directly by applying the least squares method. Hence, neither $\beta_1\omega(k;\theta)$ nor $\omega(k;\theta)$ on the left side of Equation (10) would significantly impact the result. Then, in the investigation, we chose $\beta_1\omega(k;\theta)$ to reduce the parameters to be estimated, and set P = 3.

- Step weight function

$$\beta_1\omega(k;\theta_1,\theta_2,\ldots,\theta_P) = \theta_1' \mathrm{I}_{k\in[b_0,b_1]} + \sum_{p=2}^{P} \theta_p' \mathrm{I}_{k\in[b_{p-1},b_p]} \quad (11)$$

where $b_0 = 1 < b_1 < \ldots < b_P = K$. $\mathrm{I}_{k\in[b_{p-1},b_p]}$ is an indicative function. If $k \in \left[b_{p-1}, b_p\right]$, $k = 1$; if $k \notin \left[b_{p-1}, b_p\right]$, $k = 0$. In addition, the hysteresis term of the high-frequency variables is segmented by $3n$ (n = 1, 2, 3, . . .).

3.2.5. MIDAS Model with Error Correction

First, we considered the same-frequency ECM. Regardless of whether modelling based on consumption, investment, trade or principal component, the analysis utilised two variables (one of the above and GDP). Therefore, the construction of the ECM was also based on two variables, where Y_t was set as the explained variable and X_t as the explanatory one. The specific process is as follows: Step one is to test the stationarity of the time series. When there is a unit root, according to Granger theorem [46], the ECM must be established to describe their short-term dynamic relationship if there is a co-integration between non-stationary variables. Step two is to construct the same-frequency ECM. We carried out co-integration regression on the non-stationary time series to obtain the residual term, and we took it as the error correction term, denoted as $ecm_{t-1} = Y_{t-1} - \hat{\alpha}_0 - \hat{\alpha}_1 X_{t-1}$. The ECM could be expressed as Equation (12), where $\Delta Y_t = Y_t - Y_{t-1}$, $\Delta X_t = X_t - X_{t-1}$, and β_1 and μ_t are the coefficient and random error term.

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda ecm_{t-1} + \mu_t \quad (12)$$

Then, the coefficient of the error correction term could be obtained using least squares estimation, a method widely used in econometric analysis. In Equation (12), $0 < \lambda < 1$ reflects the speed of error correction, which is negative under normal circumstances. Specifically, when $Y_t > \alpha_0 + \alpha_1 X_{t-1}$, $ecm_{t-1} > 0$ and $-\lambda ecm_{t-1} < 0$, which causes ΔY_t to decrease, and vice versa.

Next, we introduced the ECM-MIDAS and CoMIDAS models. According to Gotz et al. [30], an analysis constructs the ECM-MIDAS model based on the same-frequency co-integration. Their method referred to the data of a certain period being selected from the high-frequency sequence and converted into the low-frequency one, that is, choosing $X_{t-i/m}^{(m)}$ ($i \in [0, m-1]$)

and Y_t for co-integration regression. The residual term after regression is the mixed-frequency error correction term, denoted as $ecm_{t-1} = Y_{t-1} - \hat{\alpha}_0 - \hat{\alpha}_1 X^{(m)}_{t-1-i/m}$. Then, the ECM-MIDAS model can be expressed as the following formula:

$$\Delta Y_t = \beta_0 + \sum_{j=1}^{p-1} \gamma_j \Delta Y_{t-j} + \beta_1 B\left(L^{1/m};\theta\right)\Delta^{1/m} X^{(m)}_{t-h/m} - \lambda ecm_{t-1} + \varepsilon_t \quad (13)$$

where $\Delta^{1/m}$ is the difference in the high-frequency variable. In this case, since the economic data of each month in a quarter would not fluctuate significantly except for exceptional circumstances, the value of i may not cause a huge difference in practical application.

The CoMIDAS model can also be regarded as a mixed-frequency ECM built based on same-frequency co-integration. The difference between the CoMIDAS model and the ECM-MIDAS model lies in how the period of high-frequency and low-frequency variables of the latter is the same (for example, these two variables should be in the same quarter), but in the former, the two sequences' periods may differ. Take h = 1 as an example, the low-frequency variable is in the first quarter, the corresponding high-frequency one is in the second month of the second quarter, and the mixed-frequency error correction term is $ecm_{t-1} = Y_{t-1} - \hat{\alpha}_0 - \hat{\alpha}_1 X^{(m)}_{t-1/m}$.

After testing (Polyzos and Siriopoulos, [47]), we determined that the predictive effectiveness and the parameter significance were better when h = 1 and p = 2 in Equation (13). Therefore, Equation (13) could be simplified as:

$$\Delta Y_t = \beta_0 + \gamma_1 \Delta Y_{t-1} + \beta_1 B\left(L^{1/m};\theta\right)\Delta^{1/m} X^{(m)}_{t-1/m} - \lambda ecm_{t-1} + \varepsilon_t \quad (14)$$

If ecm_{t-1} does not exist, the above formula is the MIDAS model. If $ecm_{t-1} = Y_{t-1} - \hat{\alpha}_0 - \hat{\alpha}_1 X^{(m)}_{t-1-i/m}$, this formula is the ECM-MIDAS model, and only the data in the second month of each quarter are selected for analysis, that is $i = 1$. If $ecm_{t-1} = Y_{t-1} - \hat{\alpha}_0 - \hat{\alpha}_1 X^{(m)}_{t-1/m}$, Equation (14) is the CoMIDAS model; although the data in the second month of each quarter are also used, the quarter may be different from that of the low-frequency variable.

4. Results

4.1. Mixed-Frequency Prediction Based on Consumption, Investment and Trade

4.1.1. Unit Root Test

Our analysis utilised the Augmented Dickey–Fuller (ADF) method [48] to evidence the stationarity in the selected variables, and the results are presented in Table 2. It can be seen that LNGDP, LNC, LNI and LNT had unit roots, and their first-order difference DGDP, DC, DI and DT were stationary sequences.

Table 2. The unit root test results for selected variables.

Variables	ADF Statistics	Test Forms	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)	*p*-Value
LNGDP	−0.8677 ***	(c, t, 1)	−4.083355	−3.470032	−3.161982	0.9539
DGDP	−3.6414 ***	(c, 0, 1)	−3.520307	−2.90067	−2.587691	0.0071
LNC	−2.3322 ***	(c, t, 2)	−3.99828	−3.429398	−3.138192	0.4144
DC	−3.1383 ***	(0, 0, 3)	−2.575099	−1.942218	−1.615776	0.0018
LNI	−1.1899 ***	(c, t, 2)	−3.99828	−3.429398	−3.138192	0.9096
DI	−7.5555 ***	(0, 0, 3)	−2.575099	−1.942218	−1.615776	0.0000
LNT	−0.6592 ***	(c, t, 2)	−3.99828	−3.429398	−3.138192	0.9741
DT	−7.1026 ***	(0, 0, 3)	−2.575099	−1.942218	−1.615776	0.0000

Notes: *** denotes significance at the 1% level.

4.1.2. Results of Mixed-Frequency Prediction

Since there was no unit root in the first-order difference sequences of LNGDP, LNC, LNI and LNT, the MIDAS models could be constructed based on DGDP, DC, DI and DT, respectively. The RMSEs of MIDAS models and corresponding same-frequency models are shown in Table 3.

Table 3. The RMSEs of mixed- and same-frequency models.

Variables	Weight Functions	$K = 6$	$K = 9$	$K = 12$	$K = 15$	$K = 18$	$K = 21$	$K = 24$	$K = 27$	$K = 30$	Average	Same-Frequency
DC	beta	1.0058	1.1229	1.1214	1.1208	1.1205	1.1203	1.1931	1.1483	1.1598	1.1237	1.3383
	betaNN	1.2168	1.1619	0.8979	1.0594	0.9631	0.6876	0.8496	0.8394	0.8616	0.9486	
	expalmon	1.1769	1.1769	1.1769	1.0517	1.1851	1.0169	1.1762	1.1738	1.1667	1.1446	
	almon	1.2288	1.0496	1.0778	1.2767	1.2211	1.0191	0.9817	1.0186	0.9261	1.0888	
	step	1.2449	1.3233	1.33	1.602	1.6516	1.6351	1.6161	1.6744	1.6752	1.5281	
DI	beta	0.7387	0.7464	0.735	0.7351	0.7509	0.7207	0.7632	0.7229	0.8179	0.7479	0.8334
	betaNN	0.7414	0.7237	0.7612	0.7516	0.7409	0.7967	0.7238	0.7396	0.7871	0.7518	
	expalmon	0.7459	0.7543	0.7705	0.7936	0.7804	0.7843	0.7452	0.7385	0.7627	0.7639	
	almon	0.7511	0.7356	0.7566	0.7444	0.7493	0.7477	0.7465	0.7423	0.7582	0.7480	
	step	0.7602	0.7402	0.7554	0.7563	0.7582	0.7564	0.7532	0.7609	0.804	0.7605	
DT	beta	0.865	0.7766	0.7726	0.7707	0.7696	0.7689	0.7528	0.7511	0.7479	0.7750	1.0841
	betaNN	0.8492	0.8117	0.783	0.7641	0.8101	0.8841	0.7976	0.7167	0.7517	0.7965	
	expalmon	0.8606	0.8606	0.8606	0.8606	0.8606	0.8417	0.8516	0.8481	0.8383	0.8536	
	almon	0.8952	0.8972	0.8907	0.8557	0.79	0.7931	0.7196	0.9234	0.8709	0.8484	
	step	0.869	0.8913	0.8967	0.8769	0.8376	0.8225	0.8517	0.8804	0.9116	0.8709	

Notes: The RMSE of the same-frequency model is an average value. The construction of the same-frequency model is to simply average the monthly data of consumption, investment and trade, and then use the quarterly data as explanatory variables to build the ARDL model. After that, the RMSE is obtained by selecting the lag order corresponding to the MIDAS model. In order to simplify the analysis, the specific process of calculating RMSE for the same-frequency model is not listed in this paper, and the following tables are the same.

4.2. Mixed-Frequency Prediction with Principal Component

4.2.1. Principal Component Analysis

Although the mixed-frequency model could improve the predictive effectiveness, there are downsides to constructing a univariate model to predict GDP. Therefore, this analysis utilised the three variables of consumption, investment and trade to predict GDP jointly. Since the direct regression of collinear variables may lead to inaccurate analyses and predictions, in this study, we first conducted the correlation analysis on the original consumption, investment and trade data after seasonal adjustment. The correlation coefficients are shown in Table 4.

Table 4. The correlation coefficients among C, I and T after seasonal adjustment.

	C after Seasonal Adjustment	I after Seasonal Adjustment	T after Seasonal Adjustment
C after seasonal adjustment	1	0.9444	0.9578
I after seasonal adjustment	0.9444	1	0.9598
T after seasonal adjustment	0.9578	0.9598	1

It can be clearly observed from Table 4 that consumption, investment and trade were highly correlated, and the correlation of each pair of variables was greater than 90%. In order to make the analyses and predictions more accurate and avoid problems such as excessive parameters to be estimated in multivariate regression, in this study, we used the PCA technique to extract the principal components of these three variables. The results of the PCA are presented in Table 5.

Table 5. The results of PCA.

	Proportion	Cumulative Proportion
Principal component 1	0.9693	0.9693
Principal component 2	0.0186	0.9879
Principal component 3	0.0121	1

From Table 5, we can observe that the proportion of principal component 1 was 96.93%, meaning that principal component 1 could represent 96.93% of the information on consumption, investment and trade. Thus, the linear combination of principal component 1 could be selected to replace the consumption, investment and trade data. In principal component 1, the linear combination coefficients of these three variables were 0.5762, 0.5766 and 0.5793, respectively. In the remainder of the paper, the weighted and summed linear combination sequence is denoted as PC, and its logarithmic and log-difference sequences are expressed as LNPC and DPC.

4.2.2. Results of Mixed-Frequency Prediction

The ADF test proved that LNPC was a non-stationary sequence, while its first-order difference was stationary. Therefore, when constructing the mixed-frequency model based on DPC and DGDP, the former from February 1998 to June 2021 and the latter from the second quarter of 1998 to the second quarter of 2021 were also selected as data to include in the sample of the MIDAS model to predict the next eight quarters. The results of RMSE with different lags and weight functions and the comparisons are shown in Table 6. Compared with consumption, investment and trade, the principal-component-based mixed-frequency model had better predictive accuracy; hence, only the principal-component-based mixed-frequency error correction model was selected for analyses and predictions in the following part of the study.

Table 6. Comparisons of predictions based on different variables.

	beta	betaNN	Expalmon	Almon	Step	Same-Frequency
$K = 6$	0.7546	0.7203	0.7638	0.7728	0.7766	-
$K = 9$	0.6986	0.7331	0.764	0.6889	0.7667	-
$K = 12$	0.6913	0.6977	0.764	0.7265	0.7695	-
$K = 15$	0.6893	0.7849	0.764	0.7251	0.7593	-
$K = 18$	0.6882	0.697	0.764	0.6994	0.7474	-
$K = 21$	0.6875	0.7794	0.6935	0.7005	0.7484	-
$K = 24$	0.6721	0.6705	0.7461	0.6658	0.7464	-
$K = 27$	0.6681	0.6699	0.74	0.7362	0.753	-
$K = 30$	0.6706	0.7128	0.7461	0.699	0.7505	-
Averages of mixed- and same-frequency models based on DPC	0.6911	0.7184	0.7495	0.7127	0.7575	0.7923
Averages of mixed- and same-frequency models based on DC	1.1237	0.9486	1.1446	1.0888	1.5281	1.3383
Averages of mixed- and same-frequency models based on DI	0.7479	0.7518	0.7639	0.7480	0.7605	0.8334
Averages of mixed- and same-frequency models based on DT	0.7750	0.7965	0.8536	0.8484	0.8709	1.0841

4.3. Mixed-Frequency Prediction with Error Correction and Principal Component

4.3.1. Data Description

In this research, LNPC in the second month of each quarter and LNGDP in that quarter were selected to construct the ECM-MIDAS model; and LNPC in the second month of each quarter and LNGDP in the last quarter were used to build the CoMIDAS model. The time range of data within the mixed-frequency error correction model sample was unchanged, meaning it could predict the next eight quarters. Hereon, we denote the second month of each quarter as LNPC2. By utilising the ADF approach, with LNPC2 as a first-order single integral sequence, it could be tested for co-integration with LNGDP.

4.3.2. Co-Integration Test

First, we carried out the least squares regression to obtain the residual terms of two regression models with LNPC2 and LNGDP or with LNPC2 and LNGDP (−1), which are denoted as *ecm* and *ecmco*, respectively. Then, we evidenced their stability by using the ADF test. In Table 7, it can be seen that *ecm* and *ecmco* rejected the null hypothesis of the existence of a unit root at the significance levels of 5% and 10%. Hence, we could conclude that both the LNPC2 and LNGDP model and LNPC2 and LNGDP (−1) model had long-term co-integration relations, meeting the conditions for constructing the mixed-frequency error correction model.

Table 7. The unit root test of residual terms.

Variables	ADF Statistics	Test Forms	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)	p-Value
ecm	−2.0907 **	(0, 0, 4)	−2.6034	−1.9463	−1.6133	0.0361
ecmco	−1.9442 *	(0, 0, 3)	−2.6028	−1.9462	−1.6134	0.0502

Notes: ** and * denote the significance at the 5% and 10% levels.

4.3.3. Results of Mixed-Frequency Prediction

Based on *ecm* and *ecmco*, two principal-component-based mixed-frequency error correction models were constructed in combination with Equation (14). The predictions of the ECM-MIDAS and CoMIDAS models and comparisons are shown in Table 8.

Table 8. Comparisons of principal-component-based mixed-frequency error correction models.

	beta		**betaNN**		**Expalmon**		**Almon**		**Step**	
	ECM	Co	ECM	Co	ECM	Co	ECM	Co	ECM	Co
$K = 6$	0.7789	0.8061	0.6498	0.7056	0.7251	0.7568	0.7202	0.7501	0.7433	0.7436
$K = 9$	0.6441	0.7279	0.6495	0.676	0.7252	0.757	0.629	0.6603	0.7449	0.7452
$K = 12$	0.6413	0.7065	0.6354	0.7176	0.7252	0.757	0.6649	0.6917	0.7315	0.732
$K = 15$	0.6401	0.6699	0.7352	0.7362	0.7252	0.757	0.6745	0.7019	0.7279	0.7284
$K = 18$	0.6395	0.6694	0.6262	0.6954	0.7252	0.757	0.6593	0.6834	0.732	0.7327
$K = 21$	0.6392	0.669	0.7399	0.7203	0.6864	0.757	0.6669	0.6954	0.7355	0.7359
$K = 24$	0.6247	0.655	0.6283	0.6559	0.7127	0.7472	0.644	0.6751	0.7302	0.731
$K = 27$	**0.6212**	0.6507	0.6441	0.7143	0.7086	0.742	0.7138	0.7368	0.7499	0.7509
$K = 30$	0.6237	0.6524	0.6552	0.7696	0.7177	0.7483	0.6936	0.7293	0.7563	0.7582
Averages of mixed-frequency error correction model	0.6503	0.6897	0.6626	0.7101	0.7168	0.7533	0.6740	0.7027	0.7391	0.7398
Averages of mixed-frequency model without error correction term	0.6911		0.7184		0.7495		0.7127		0.7575	
Averages of same-frequency error correction model	0.7885									
Averages of the same-frequency model without error correction term	0.7923									

Notes: ECM and Co denote the ECM-MIDAS and CoMIDAS models, respectively.

As can be seen from Table 8, the predictive effectiveness of the principal-component-based mixed-frequency error correction model was the highest when K = 27 and the weight function was beta, and that model was the ECM-MIDAS system. We list these estimated results in Table 9. We found that the coefficient of the error correction term (λ) was significantly positive and belonged to 0 to 1, which conformed to the common practice; thus, constructing a principal-component-based mixed-frequency error correction model was appropriate.

Table 9. Estimated results of principal-component-based mixed-frequency error correction model.

Variables	β_0	β_1	θ_1	θ_2	γ_1	λ
Coefficients	0.9798 ***	0.5998 ***	1.0271 ***	13.2786 ***	0.4247 ***	0.0344 *

Notes: *** and * denote significance at the 1% and 10% levels.

5. Discussion

Previous studies confirmed the MIDAS model's effectiveness in predicting GDP [16–28], and this analysis produced further evidence of that conclusion by considering consumption, investment and trade. In Table 3, it can seen that no matter whether consumption, investment or trade was an explanatory variable, the RMSE of the mixed-frequency model was smaller than that of the same-frequency one. The main reason is that the MIDAS model could make full use of the information of high-frequency data and avoid the problem of losing data when averaging monthly data into quarterly sequences, which is conducive to improving prediction accuracy. In addition, the same- and mixed-frequency models based on investment had the highest predictive accuracy, followed by trade and consumption. Furthermore, among the five weight functions, the beta class and almon weight functions had better predictive abilities, while the step weight function had poorer predictive effectiveness. Thus, the beta class and almon weight functions can be given priority when constructing a mixed-frequency model. However, the research base on the MIDAS model with error correction and a principal component is insufficient, and this paper only begins to fill the gap.

On the one hand, we probed the predictive effectiveness of the principal-component-based MIDAS model. From Table 6, we can confirm that the accuracy of mixed-frequency prediction based on the principal component was better than that of the same-frequency prediction, highlighting that the MIDAS model is more effective in forecasting than the the ARDL model. More importantly, the same- and mixed-frequency models based on the principal component had better predictive effectiveness, mainly because the extracted principal component contained information on consumption, investment and trade. Thus, combining the PCA and MIDAS models can enhance predictive accuracy. Additionally, we could further ascertain that the beta weight function had the highest predictive effectiveness.

On the other hand, we explored the predictive effectiveness of the principal-component-based mixed-frequency error correction model. From Table 8, we can determine that the mixed-frequency model not only had better predictive effectiveness without an error correction term but also held better predictive accuracy than the same-frequency error correction model after adding an error correction term, and the prediction based on the beta weight function still had the best effectiveness. Furthermore, the predictive error was reduced when the error correction term was added to the mixed- and same-frequency models. The primary cause for this is that some vital information of the original data may be lost if the first-order difference is used to avoid "spurious regression" and the error correction term is missing, resulting in unsatisfactory predictive effectiveness. But adding the error correction term can allow for adjustment according to the long-term relationship, making the prediction more effective. Moreover, the predictive accuracy of the ECM-MIDAS model was slightly better than that of the CoMIDAS model in this case, mainly because the fitting effect when constructing error correction terms based on data of the same period was better than that of different periods. However, the actual selection of these two mixed-frequency error correction models should be analysed in detail in the future.

6. Conclusions

In this analysis, we selected consumption, investment and trade in order to construct MIDAS models to predict quarterly GDP in China. Furthermore, we utilised PCA to extract the principal component, and built a principal-component-based mixed-frequency error correction model, following which we probed the predictive effectiveness. Based on the RMSE, which measures the predictive accuracy, the effectiveness of different models in forecasting could be compared, and the following conclusions are drawn:

Firstly, the predictive accuracy of the mixed-frequency model is better than that of the same-frequency model. This conclusion can be observed in the MIDAS models based on consumption, investment, trade and principal component, and the ECM-MIDAS and CoMIDAS models. Thus, making predictions based on the mixed-frequency model is effective. In doing so, relevant policymakers could combine mixed-frequency data to predict quarterly GDP in China, supporting the real-time and accurate formulation of macroeconomic policies.

Secondly, consumption, investment and trade have different forecasting effects on GDP. The same- and mixed-frequency models show that investment has the best predictive effectiveness on GDP, followed by trade and consumption. In addition, the predicted value based on consumption is higher than the real value, while investment and trade are moderate, meaning that consumption plays a greater role in boosting GDP than investment and trade in China. Hence, China should not only implement relevant policies to stimulate investment and foreign trade but also give full play to the potential of promoting consumption to boost GDP growth, which is beneficial to promote economic development.

Thirdly, PCA is effective in mixed-frequency prediction. When applying the PCA technique here, the principal component that could reflect 96.93% of the information on the three variables was extracted to build the MIDAS model. When making a comparison, it can be observed that the predictive accuracy of the principal-component-based MIDAS model was significantly better than the MIDAS model based on consumption, investment and trade. This is because the principal-component-based MIDAS model not only makes full use of multiple variables but also overcomes problems such as inaccurate predictions and excessive parameters caused by multivariable collinearity.

Fourthly, combining the ECM and MIDAS models is effective in forecasting GDP. By constructing same- and mixed- frequency error correction models, it was found that adding the error correction term improved the predictive accuracy. In this case, the predictive effectiveness of the ECM-MIDAS model was better than that of the CoMIDAS system. Although the choice between these two mixed-frequency error correction models still needs to be analysed in detail, we can conclude that constructing the principal-component-based mixed-frequency error correction model is appropriate, and its prediction is effective.

Fifthly, the beta weight function has better predictive effectiveness. The beta weight function generally has the smallest predictive error and significant parameters; thus, in this study, we could directly select this weight function to simplify the analysis process when performing the mixed-frequency prediction.

In the future, research should focus on the following aspects: First, the construction of mixed-frequency error correction models should not be limited to same-frequency co-integration; instead, the theory and method of mixed-frequency co-integration need to be further explored. Second, consumption, investment and trade are not the only data used to predict GDP. Other predictors (e.g., money supply and inflation) should also be taken into account, and we would advise studying which predictor or combination possesses the most powerful predictive effect. Third, according to Ang et al. [49] and Evgenidis et al. [50], researchers should also consider the yield curve in order to predict GDP.

Author Contributions: Conceptualisation, C.-W.S. and Y.W.; methodology, Q.M.; software, Y.Z.; validation, O.-R.L., C.-W.S. and Y.W.; formal analysis, Q.M.; investigation, Y.Z.; resources, Y.W.; data curation, C.-W.S.; writing—original draft preparation, Q.M.; writing—review and editing, O.-R.L.; visualisation, Y.Z.; supervision, C.-W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data can be found here: https://data.stats.gov.cn/index.htm (accessed on 1 August 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chow, G.C.; Lin, A.L. Best linear unbiased estimation of missing observations in an economic time series. *J. Am. Stat. Assoc.* **1976**, *71*, 719–721. [CrossRef]
2. Harvey, A.C.; Pierse, R.G. Estimating missing observations in economic time series. *J. Am. Stat. Assoc.* **1984**, *79*, 125–131. [CrossRef]

3. Ghysels, E.; Santa-Clara, P.; Valkanov, R. *The MIDAS Touch: Mixed Data Sampling Regressions Models*; Working Paper; UNC: Chapel Hill, NC, USA; UCLA: Los Angeles, CA, USA, 2004.
4. Ghysels, E.; Sinko, A.; Valkanov, R. MIDAS regressions: Further results and new directions. *Economet. Rev.* **2007**, *26*, 53–90. [CrossRef]
5. Kumar, J.; Mishra, A.K. Modeling of sectoral GDPs for forecasting India GDP using ARMA models. *Indian J. Econ. Bus.* **2016**, *15*, 181–195.
6. Tripathi, P.K.; Ranjan, R.; Pant, R.; Upadhyay, S.K. An approximate bayes analysis of ARMA model for Indian GDP growth rate data. *J. Stat. Manage. Sys.* **2017**, *20*, 399–419. [CrossRef]
7. Li, G.D.; Masuda, S.; Yamaguchi, D.; Hayashi, M.; Nagai, M. A study on the development trends of GDP, population and primary energy consumption by grey-based dynamic model. *J. Grey Sys.* **2008**, *11*, 73–84.
8. Listra, O.A. Analysis of Indonesia's green GDP accounting with grey system theory type GM (1,1) for years 2000–2010. *J. Ilm. Mhs.* **2012**, *1*.
9. Abrishami, H.; Mehrara, M.; Ahrari, M.; Mirghasemi, S. Forecasting the GDP in Iran based on GMDH neural network. *Tahghighat-E-Eghtesadi* **2009**, *44*, 1–24.
10. Biau, O.; Elia, A.D. *Euro Area GDP Forecast Using Large Survey Dataset: A Random Forest Approach*; Bilgi University: Istanbul, Turkey, 2010.
11. Stundziene, A. Prediction of Lithuanian GDP: Are regression models or time series models better? *Econ. Manag.* **2014**, *18*, 721–734. [CrossRef]
12. León, A.; Nave, J.M.; Rubio, G. The relationship between risk and expected return in Europe. *J. Bank. Financ.* **2007**, *31*, 495–512. [CrossRef]
13. Alper, C.E.; Fendoglu, S.; Saltoğlu, B. *Forecasting Stock Market Volatilities Using MIDAS Regressions: An Application to the Emerging Markets*; University Library of Munich: Munich, Germany, 2008.
14. Engle, R.F.; Ghysels, E.; Sohn, B. *On the Economic Sources of Stock Market Volatility. Discussion Paper*; NYU: New York, NY, USA; UNC: Chapel Hill, NC, USA, 2009.
15. Colacito, R.; Engle, R.; Ghysels, E. A component model for dynamic correlations. *J. Econom.* **2011**, *164*, 45–59. [CrossRef]
16. Wu, X.Y.; Zhao, A.; Cheng, T.F. A real-time GARCH-MIDAS model. *Financ. Res. Lett.* **2023**, *56*, 104103. [CrossRef]
17. Li, D.X.; Zhang, L.; Li, L.H. Forecasting stock volatility with economic policy uncertainty: A smooth transition GARCH-MIDAS model. *Int. Rev. Financ. Anal.* **2023**, *88*, 102708. [CrossRef]
18. Yao, C.Z.; Li, M.J. GARCH-MIDAS-GAS-copula model for CoVaR and risk spillover in stock markets. *N. Am. J. Econ. Financ.* **2023**, *66*, 101910. [CrossRef]
19. Wang, L.; Wu, J.B.; Cao, Y.; Hong, Y.R. Forecasting renewable energy stock volatility using short and long-term Markov switching GARCH-MIDAS models: Either, neither or both? *Energ. Econ.* **2022**, *111*, 106056. [CrossRef]
20. Ersin, Ö.Ö.; Bildirici, M. Financial volatility modeling with the GARCH-MIDAS-LSTM approach: The effects of economic expectations, geopolitical risks and industrial production during COVID-19. *Mathematics* **2023**, *11*, 1785. [CrossRef]
21. Su, X.; Liu, Z. Sector volatility spillover and economic policy uncertainty: Evidence from China's stock market. *Mathematics* **2021**, *9*, 1411. [CrossRef]
22. Lin, A.J.; Chang, H.Y. Volatility transmission from equity, bulk shipping, and commodity markets to oil ETF and energy fund—A GARCH-MIDAS model. *Mathematics* **2020**, *8*, 1534. [CrossRef]
23. Clements, M.P.; Galvão, A.B. Macroeconomic forecasting with mixed-frequency data. *J. Bus. Econ. Stat.* **2008**, *26*, 546–554. [CrossRef]
24. Hogrefe, J. Forecasting data revisions of GDP: A mixed frequency approach. *AStA Adv. Stat. Anal.* **2008**, *92*, 271–296. [CrossRef]
25. Andreou, E.; Ghysels, E.; Kourtellos, A. Should macroeconomic forecasters use daily financial data and how? *J. Bus. Econ. Stat.* **2013**, *31*, 240–251. [CrossRef]
26. Aprigliano, V.; Foroni, C.; Marcellino, M.; Mazzi, G.; Venditti, F. A daily indicator of economic growth for the euro area. *Int. J. Comput. Econ. Ec.* **2017**, *7*, 43–63. [CrossRef]
27. Fu, R.; Xie, L.; Liu, T.; Huang, J.; Zheng, B. Chinese economic growth projections based on mixed data of carbon emissions under the COVID-19 pandemic. *Sustainability* **2022**, *14*, 16762. [CrossRef]
28. Mishra, P.; Alakkari, K.; Abotaleb, M.; Singh, P.K.; Singh, S.; Ray, M.; Das, S.S.; Rahman, U.H.; Othman, A.J.; Ibragimova, N.A.; et al. Nowcasting India economic growth using a mixed-data sampling (MIDAS) model (empirical study with economic policy uncertainty-consumer prices index). *Data* **2021**, *6*, 113. [CrossRef]
29. Chikamatsu, K.; Hirakata, N.; Kido, Y.; Otaka, K. Mixed-frequency approaches to nowcasting GDP: An application to Japan. *Jpn. World Econ.* **2021**, *57*, 101056. [CrossRef]
30. Pan, Z.Y.; Wang, Q.; Wang, Y.D.; Yang, L. Forecasting U.S. real GDP using oil prices: A time-varying parameter MIDAS model. *Energ. Econ.* **2018**, *72*, 177–187. [CrossRef]
31. Chernis, T.; Cheung, C.; Velasco, G. A three-frequency dynamic factor model for nowcasting Canadian provincial GDP growth. *Int. J. Forecast.* **2020**, *36*, 851–872. [CrossRef]
32. Xu, Q.F.; Zhuo, X.X.; Jiang, C.X.; Liu, X.; Liu, Y.Z. Group penalized unrestricted mixed data sampling model with application to forecasting US GDP growth. *Econ. Model.* **2018**, *75*, 221–236. [CrossRef]

33. Jiang, Y.; Guo, Y.J.; Zhang, Y.H. Forecasting China's GDP growth using dynamic factors and mixed-frequency data. *Econ. Model.* **2017**, *66*, 132–138. [CrossRef]
34. Pettenuzzo, D.; Timmermann, A.; Valkanov, R. A MIDAS approach to modeling first and second moment dynamics. *J. Econom.* **2016**, *193*, 315–334. [CrossRef]
35. Barsoum, F.; Stankiewicz, S. Forecasting GDP growth using mixed-frequency models with switching regimes. *Int. J. Forecast.* **2015**, *31*, 33–50. [CrossRef]
36. Degiannakis, S. The D-model for GDP nowcasting. *Swiss J. Econ. Stat.* **2023**, *159*, 7. [CrossRef]
37. Miller, J.I. Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *J. Financ. Economet.* **2012**, *12*, 584–614. [CrossRef]
38. Götz, T.B.; Hecq, A.; Urbain, J.P. Forecasting mixed-frequency time series with ECM-MIDAS models. *J. Forecast.* **2014**, *33*, 198–213. [CrossRef]
39. Findley, D.F.; Monsell, B.C.; Bell, W.R.; Otto, M.C.; Chen, S. New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *J. Bus. Econ. Stat.* **1998**, *16*, 127–152.
40. Karunasingha, D.S.K. Root mean square error or mean absolute error? Use their ratio as well. *J. Inf. Sci.* **2022**, *585*, 609–629. [CrossRef]
41. Schubert, A.L.; Hagemann, D.; Voss, A.; Bergmann, K. Evaluating the model fit of diffusion models with the root mean square error of approximation. *J. Math. Psychol.* **2017**, *77*, 29–45. [CrossRef]
42. Faber, N.M. Estimating the uncertainty in estimates of root mean square error of prediction: Application to determining the size of an adequate test set in multivariate calibration. *Chemom. Intell. Lab. Syst.* **1999**, *49*, 79–89. [CrossRef]
43. Mei, D.X.; Ma, F.; Liao, Y.; Wang, L. Geopolitical risk uncertainty and oil future volatility: Evidence from MIDAS models. *Energ. Econ.* **2020**, *86*, 104624. [CrossRef]
44. Wang, L.; Wu, R.; Ma, W.C.; Xu, W.J. Examining the volatility of soybean market in the MIDAS framework: The importance of bagging-based weather information. *Int. Rev. Financ. Anal.* **2023**, *89*, 102720. [CrossRef]
45. Li, X.; Shang, W.; Wang, S.Y.; Ma, J. A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data. *Electron. Commer. Res.* **2015**, *14*, 112–125. [CrossRef]
46. Granger, C.W.J. Some recent development in a concept of causality. *J. Econom.* **1988**, *39*, 199–211. [CrossRef]
47. Polyzos, E.; Siriopoulos, C. Autoregressive random forests: Machine learning and lag selection for financial research. *Comput. Econ.* **2023**, 1–38. [CrossRef]
48. Dickey, D.A.; Fuller, W.A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **1981**, *49*, 1057–1072. [CrossRef]
49. Ang, A.; Piazzesi, M.; Wei, M. What does the yield curve tell us about GDP growth? *J. Econom.* **2006**, *131*, 359–403. [CrossRef]
50. Evgenidis, A.; Papadamou, S.; Siriopoulos, C. The yield spread's ability to forecast economic activity: What have we learned after 30 years of studies? *J. Bus. Res.* **2020**, *106*, 221–232. [CrossRef]