INTERNATIONAL MONETARY FUND

Parameter Proliferation in Nowcasting: Issues and Approaches

An Application to Nowcasting China's Real GDP

Paul Cashin, Fei Han, Ivy Sabuga, Jing Xie, and Fan Zhang

WP/25/217

2025
OCT

WORKING PAPER

 WP/25/217

IMF Working Paper
Institute for Capacity Development

Parameter Proliferation in Nowcasting: Issues and Approaches—An Application to Nowcasting China's Real GDP
Prepared by Paul Cashin, Fei Han, Ivy Sabuga, Jing Xie, and Fan Zhang*

Authorized for distribution by Natan Epstein
October 2025

ABSTRACT: This paper evaluates three approaches to address parameter proliferation issue in nowcasting: (i) variable selection using adjusted stepwise autoregressive integrated moving average with exogenous variables (AS-ARIMAX); (ii) regularization in machine learning (ML); and (iii) dimensionality reduction via principal component analysis (PCA). Utilizing 166 variables, we estimate our models from 2007Q2 to 2019Q4 using rolling-window regression, while applying these three approaches. We then conduct a pseudo out-of-sample performance comparison of various nowcasting models—including Bridge, MIDAS, U-MIDAS, dynamic factor model (DFM), and machine learning techniques including Ridge Regression, LASSO, and Elastic Net to predict China's annualized real GDP growth rate from 2020Q1 to 2023Q1. Our findings suggest that the LASSO method outperform all other models, but only when guided by economic judgment and sign restrictions in variable selection. Notably, simpler models like Bridge with AS-ARIMAX variable selection yield reliable estimates nearly comparable to those from LASSO, underscoring the importance of effective variable selection in capturing strong signals.

JEL Classification Numbers:	C18, C53
Keywords:	China; GDP; Nowcasting

Authors' E-Mail Address:	PCashin@imf.org, FHan@imf.org, ISabuga@imf.org, JX3770@princeton.edu, FZhang@imf.org

WORKING PAPERS

Parameter Proliferation in Nowcasting: Issues and Approaches

An Application to Nowcasting China's Real GDP

Prepared by Paul Cashin, Fei Han, Ivy Sabuga, Jing Xie, and Fan Zhang[1]

[1] We are grateful for comments from Mr. Sam Ouliaris, Mr. Dimitre Milkov, and Ms. Hongbo Wang. The views expressed in this paper are those of the authors and not necessarily those of the International Monetary Fund, or its Executive Board.

Contents

I. Introduction

Monetary policy decisions in real time are typically based on assessments of current and future economic conditions using incomplete data. Since most data, particularly quarterly macroeconomic data such as Gross Domestic Products (GDP), are released with a lag and are subsequently revised, assessing the economic conditions in the current period becomes a challenging task for central banks. To address this issue, nowcasting techniques have been introduced that utilize high-frequency indicators to monitor real-time economic activity in the absence of up-to-date official GDP data.

Nowcasting is the practice of estimating the current or very recent state of an economic, financial, or other dynamic variable before official data are released, using partially available, real time, high-frequency data (see Giannone et al. 2008). Nowcasts help fill the information gap between the latest official statistics (often delayed by weeks or months) and real-time developments, thereby helping policymakers monitor the economy with minimal delays. Nowcasting uses statistical, econometric, or machine learning models to combine data from multiple sources (such as surveys, financial markets, or satellite imagery) to produce timely estimates of the target variable.

Current-quarter numbers produced by nowcasting models are crucial for policymaking as they serve as initial condition inputs for longer-term forecasting models, such as semi-structural models (including Quarterly Projection Models (QPM) and Dynamic Stochastic General Equilibrium (DSGE) models.[1] The accuracy of macroeconomic projections in these longer-horizon models heavily depends on these starting conditions. Empirical studies demonstrate that forecast errors tend to increase with the forecast horizon due to the accumulation of uncertainty. Inputs from nowcasting models can enhance the predictability of both near-term and longer-term forecasts (see, e.g., Giannone et al., 2004; Armstrong, 2002). A notable example of the usefulness of nowcasting emerged during the COVID-19 pandemic, which underscored the necessity for a systematic analysis of high-frequency indicators of economic activity. Crisis situations require higher-frequency information to enable swift decision-making by businesses, market analysts, and policymakers.

For instance, in China, the economy experienced its most significant contraction in several decades in the first quarter of 2020. The first official GDP growth estimates for this quarter were published only in mid-April. Although this release occurred earlier than in many other countries, it still provided limited and timely information to inform monetary policy responses. Consequently, policymakers relied on other readily available economic indicators, leading to significant monetary policy easing during February and March 2020 to mitigate the impact of the COVID-19 shocks, well before the official GDP data for Q1 2020 was released. A similar situation occurred in 2022 when the release of second-quarter GDP data was delayed due to lockdowns in various regions of China. In this context, a systematic approach to analyzing a variety of high-frequency data from different sources can provide timely and useful information on current economic conditions, informing policymaking.

Nowcasts for macroeconomic variables such as GDP are usually constructed by combining simple econometric models with qualitative judgment. These exercises involve analyzing a large set of time series data. In the case

[1] A DSGE model is a structural macro-economic model that is used to analyze co-movements of aggregate variables and predict how the economy behaves overtime. This is commonly used by policymakers to conduct policy scenario analysis, understand historical data, and conduct forecast.

of quarterly GDP, for instance, nowcasting provides a statistical framework that estimates the current state of the economy by incorporating the latest releases of high-frequency economic data, such as monthly or even higher-frequency data. As the narrative surrounding these high-frequency variables evolves over time, the framework is updated accordingly to reflect a more accurate assessment of economic activity (Banbura et al., 2012).

While the availability of several high-frequency indicators is beneficial for nowcasting, caution must be exercised in constructing a nowcasting model due to the issue of parameter proliferation, which can introduce more noise into the nowcast. A related concern arises when numerous variables exist within the information set, leading to what is known as the "curse of dimensionality." This phenomenon occurs when models include a large number of high-frequency predictors relative to the number of observations. As a result, parameter proliferation can lead to overfitting, where the model captures noise instead of the underlying patterns in the data, ultimately diminishing forecasting accuracy (Giannone et al., 2008).

Given that parameter proliferation is a significant issue, it is essential to explore approaches that can help distinguish informative variables from mere noise. This paper presents a systematic approach to addressing parameter proliferation in nowcasting, which combines expert knowledge, variable selection methods, statistical techniques, machine learning algorithms, and continuous monitoring.

This paper is structured as follows: Section II introduces the nowcasting process and commonly available techniques. Section III introduces three approaches to address the problem of parameter proliferation. Section IV applies these parameter proliferation approaches to nowcast China's real GDP during the COVID-19 pandemic crisis. It also conducts a pseudo out-of-sample forecast evaluation to compare the nowcast performance of various approaches in combination with different nowcasting models. Section V concludes.

II. Standard Nowcasting Techniques

Before discussing the approaches to parameter proliferation, it is essential to provide an overview of the commonly applied techniques and procedures in constructing a nowcasting tool. These nowcasting techniques range from standard models to more complex ones, as outlined below:

- **Bridge equation**: In a Bridge equation, higher-frequency indicators are converted into the target frequency (i.e., frequency of GDP, usually quarterly or annually) and an Ordinary Least Squares (OLS) regression is used to estimate the historical relationship between these higher-frequency indicators and GDP—which is typically assumed to be linear. Missing values of the high-frequency indicator are forecasted, generally by univariate time series model like an Autoregressive Integrated Moving Average (ARIMA). Then, the explanatory variables are aggregated to the target frequency using either the summation or the average method (Higgins, 2014). In other words, this approach utilizes linear regressions to bridge the information contained in one or a few key high-frequency data points with the quarterly or annual growth rate of GDP (e.g., Klein and Soji, 1989). This approach can be mathematically represented as follows:

$$y_{t_q} = \alpha + \sum_{i=1}^{j} \beta_i(L) x_{it_q} + u_{t_q}$$

Let y be the low frequency (LF) target variable (e.g., quarterly GDP growth), and x_i the high frequency (HF) indicators, $i = 1, \ldots, j$ (e.g., monthly industrial production, survey data, etc.), which are aggregated to LF according to their stock/flow nature. Here $t_q = 1, \ldots, T$ indicates time in quarters, tm in months. β_i(L) are polynomials in the lag operator L, where $\beta_i(L) = \beta_{0i} + \beta_{1i}L + \cdots + \beta_{p_i i}L^{p_i}$, one for each HF indicator, where p_i is the number of lags, and u_{t_q} is an i.i.d. error term. In example above, the target frequency t_q is quarterly, but the model can be re-estimated each month. A nowcast is constructed as:

$$\hat{y}_{T+1} = \widehat{\alpha} + \sum_{i=1}^{j} \hat{\beta}_i(L) x_{iT+1}$$

where $\widehat{\alpha}$ and $\hat{\beta}_{si}$, s= $0, \ldots, p$ are OLS estimators. When not all months of x_i in quarter $T+1$ are observable, we need to forecast the missing monthly observations prior to aggregation into x_{iT+1}using AR/ARMA models (with lag length chosen by an information criterion).

- **Mixed-frequency Data Sampling (MIDAS)**: The MIDAS approach can be regarded as a time-series regression tool that allows the dependent variable and regressors to be sampled at different frequencies while using distributed lag polynomials (DLPs) such as the exponential Almon lag to ensure parsimonious specifications. The methodology was originally proposed by Ghysels et al. (2004) and was initially used for financial applications (e.g., Ghysels et al., 2005, Ghysels et al., 2006), but has since been employed for nowcasting macroeconomic time series as well, in particular quarterly GDP using higher-frequency indicators (see, e.g., Clements and Galvão, 2008, Clements and Galvão, 2009; Marcellino and Schumacher, 2010, and Armesto et al., 2010). Unlike the Bridge equation approach, the MIDAS approach generally uses nonlinear least squares (NLS) as the estimation strategy due to the nonlinearities created by the DLPs. The basic MIDAS model for a single explanatory variable, and h_q-step-ahead forecasting, with h_q = h_m/m, is:

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

where $x_{t_m+w}^{(m)}$ is skip-sampled from the HF indicator x_{t_m} and $b(L_m;\theta)$ is a polynomial in the lag operator L for the high frequency variable.

- **Unrestricted MIDAS (or U-MIDAS)**: The U-MIDAS approach is a variant of the MIDAS approach where instead of using the DLPs for model dynamics to avoid parameter proliferation, the U-MIDAS does not impose such restrictions and instead uses unrestricted linear lag polynomials, which do not require NLS and can be estimated by OLS (Claudia et al., 2015). This could be a valid approach when discrepancies in sampling frequencies are not large, which is often the case in macroeconomic applications like nowcasting quarterly GDP using monthly indicators such as industrial production or consumer confidence survey outcomes. In its simplest form, the U-MIDAS model is a regression of y (real GDP at the quarterly frequency) on the three monthly HF skip-sampled variables $x_{1t_m}^{(m)}$:

$$y_{t_m} = \beta_0 + \beta_1 x_{1t_m}^{(m)} + \beta_2 x_{1t_m-1}^{(m)} + \beta_3 x_{1t_m-2}^{(m)} + \epsilon_{t_m}$$

U-MIDAS representations can be used when the number of skip-sampled HF variables on the right-hand side (RHS) is small. The unknown parameters can be estimated directly using OLS.

- **Factor Models (FMs):** Factor models are statistical tools used to reduce the dimensionality of data and extract meaningful information from a large set of variables. This task could be done using static and dynamic factor models. Static factor models assume that the relationships between observed variables and the underlying factors remain constant over time. Techniques like Principal Components Analysis (PCA) are often used for this purpose. Meanwhile, Dynamic Factor Models (DFMs) extend static factor models by allowing the factors to evolve over time, capturing temporal dependencies. A DFM model is an alternative way to handle mixed-frequency data in nowcasting by writing the model in a state-space form[2] and using the Kalman filter and maximum likelihood to undertake estimation. In this approach, the low-frequency (target) variable is considered a high-frequency variable with missing observations. For example, nowcasts for GDP can be produced after extracting an unobserved state of the economy and creating a new coincidence indicator (see, e.g., Mariano and Murasawa, 2003, 2010; Nunes, 2005; and Giannone et al., 2008). Compared to the MIDAS approach, which typically involves a single equation with only one dependent (target) variable, FMs generally include a system of equations, which therefore requires significant parametrization for the measurement equation, the state dynamics, and their error processesHence, FM are more likely to be subject to misspecification than alternative procedures. The DFM method can be represented by:

$$X_t = \lambda(\mathrm{L})f_t + e_t$$
$$f_t = \Psi(\mathrm{L})f_{t-1} + \eta_t$$

where the lag polynomial matrices $\lambda(\mathrm{L})$ and $\Psi(\mathrm{L})$ are N x q and q x q, respectively, and η_t is q x 1 vector of (serially uncorrelated) mean-zero innovations to the factors. The idiosyncratic disturbances are assumed to be uncorrelated with the factor innovations at all lead and lags, that is, $\mathrm{E}e_t\eta'_{t-k} = 0$ for all k. In general, e_t can be serially correlated. The i^{th} row of $\lambda(\mathrm{L})$, the lag polynomial $\lambda_i(\mathrm{L})$, is called the dynamic factor loading for the i^{th} series, $X_{it.}$ The $\lambda_i(\mathrm{L})f_t$ is the common component of the i^{th} series.

- **Machine Learning (ML) Models (Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), and Elastic Net)**: Machine learning models are also commonly used in nowcasting as they handle high-dimensional datasets effectively by performing variable selection and regularization, thus enhancing predictive accuracy. The Ridge regression model, developed in the 1960s to mitigate the problem of multicollinearity in linear regression, is a regularized linear regression technique that adds a penalty term to the sum of squared errors to shrink the regression coefficients towards zero (Hoerl and Kennard, 1970). The LASSO regression is also a regularization method used in linear regression to select a subset of relevant predictors and to reduce the impact of irrelevant indicators by imposing a penalty on their coefficient. Different from Ridge regression, LASSO could shrink irrelevant indicators equal to zero

[2] State-space equations are a mathematical framework used in DFM for nowcasting, containing state and observation equations. They represent the relationship between observed data and unobserved latent factors, which drive the dynamics of the system. State equation is also known as “transition equation” describes how the latent (unobserved) factors evolve over time. Observation equations, also known as “measurement equation” links the observed variables to the latent factors.

instead of towards zero (Tibshirani, 1996). Elastic Net combines both Ridge and LASSO regularization techniques to achieve a balance between variable selection and coefficient shrinkage (Zou and Hastie, 2005). While there is a growing body of literature exploring more advanced machine learning approaches, such as Random Forests, to manage large datasets and high-dimensional predictors (e.g., Breiman, 2001), this paper will focus exclusively on regularization techniques within ML models.

A standard nowcasting procedure includes four components: preparation, estimation, evaluation, and combination.

- **Preparation**: Determine a preliminary list of high-frequency indicators that are most relevant to the target variable (e.g., GDP). The selected indicators should be the same as or higher than the target frequency (e.g., quarterly or higher frequency for a quarterly target variable) and should have available data after the latest publication of the target variable to supplement any missing information. Once the list of indicators is compiled, nowcasters apply appropriate transformations to these indicators to ensure they are stationary and without seasonality.

- **Estimation**: Nowcasters need to determine the reduced-form baseline model, which includes explanatory variables that are meaningful for the nowcasting practice. In this step, nowcasters must ensure that the sign of the coefficient is consistent with the standard economic prior. After determining the baseline model, nowcasters can construct a nowcasting model using multiple approaches such as Bridge, MIDAS, U-MIDAS, DFM, and ML approaches.

- **Evaluation and Combination**: All the nowcasting models will then be evaluated using out-of-sample forecast evaluation, by assessing their forecast accuracy for data that was not used during model training. Based on judgment, a nowcast output will be produced either by using the model with the lowest forecast error (often using Root Mean Square Error (RMSE)) or a weighted nowcast combination. In undertaking out-of-sample evaluations, it has been demonstrated in several recent papers that in terms of real-time GDP nowcasting performance, simpler models deliver better nowcasting performance than more complex models, and traditional econometric models outperform machine learning models (see Akepanidtaworn and Akepanidtaworn 2025; Richardson et al. 2021).

III. Approaches to Parameter Proliferation

This section introduces a systematic approach to addressing parameter proliferation, which includes: (i) variable selection; (ii) regularization in machine learning (ML) models; and (iii) dimensionality reduction using principal component analysis.

A. Variable Selection

Without an objective algorithm, a judgment-based variable selection procedure can overlook superior specifications or introduce the researcher’s preconceived biases. Fernandez et al. (2001) find that statistically significant model specifications are not robust to alternative specifications. Similar conclusions are reported in several studies (e.g., Sala-i-Martin et al., 2004; Hoover & Perez, 2004; Hendry & Krolzig, 2004). Due to these concerns on non-systematic model selection, there is strong interest in automated model selection algorithms that can systematically select the “best” model specification from a large set of candidate variables (e.g., Oxley, 1995; Phillips, 2005; Castle et al., 2009). Castle et al. (2009) reviewed and compared twenty-one different model selection algorithms, in which they categorized into six groups:

- **Information Criteria such as AIC and SIC (McQuarrie & Tsai, 1998)**: involves choosing a single best baseline model using the information criterion such as Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC). These information criteria impose different penalties for including unnecessary explanatory variables and are more useful for measuring the goodness of fit for out-of-sample observations than R-squared or Adjusted R-squared. Smaller values are preferred.

- **Selection of a "portfolio" of the best subset of models (Poskitt and Tremayne, 1987)**: instead of selecting a single "best" model, this approach selects a set of models that are all "close" based on the information criteria. This approach selects models that lie within certain thresholds (Burnham and Anderson suggest a value of $\sqrt{10}$, Poskitt and Tremayne suggest a value of 2) for the minimum value model. Then, the "best" coefficient estimates were constructed using the portfolio of best subset models: for variables that never appear in the portfolio, coefficient estimates are set equal to zero; for variables that appear at least once, the respective coefficients are calculated as the arithmetic average of all nonzero coefficient estimates.

- **General-to-specific (AUTO) regression algorithm (Doornik, 2009)**: this approach is taken from Autometrics program available in PcGive 12. The Autometrics program begins with a comprehensive model that includes all potential variables and eliminates insignificant ones while ensuring the satisfaction of diagnostic tests.

- **Forward-stepwise (FW) algorithms**: through the FW algorithms, variables are added, one at a time, to a model based on the order of significance level. The procedure will stop until no further significant regressors are found. Specific included variables are removed from the model if they become insignificant as others are added. Similar to AUTO, FW algorithms will select a single-best model and assign a zero coefficient estimate for those variables excluded in the final model.

- **Bayesian Model Averaging (Hoeting, Madigan, Raftery, and Volinsky, 1999)**: In the Bayesian Model Averaging approach, a composite model is created with a weighted average of a series of models based on the posterior model probabilities. The coefficient in the composite model is equal to the weighted average of individual estimated coefficients of that variable.

- **Inclusion of all variables**: This approach creates the final model with the complete set of potential variables.

It should be noted that none of these approaches appears to dominate under all conditions. Castle et al. (2011) found that the AUTO method (Autometrics) outperforms others in over 90% of experiments under restricted conditions. One critical shortcoming of the current AUTO method is that it does not consider the overall forecast ability nor ensure the coefficient sign that matches economic priors. Moreover, the appropriate order of autoregressive (AR) and moving average (MA) terms to use in the baseline model is rarely discussed and is often set arbitrarily.

Xie (2023) proposes a simple automatic procedure: Adjusted Stepwise–ARIMAX (AS-ARIMAX) for selecting indicators, from a larger set of economic variables, that are economically meaningful, statistically significant, and effective in terms of improving the accuracy of the nowcast, while including the appropriate AR and MA terms. The AS-ARIMAX procedure uses a modified stepwise procedure, adding one variable at a time, to ensure each variable added satisfies the three conditions mentioned earlier. The effectiveness of the procedure was verified by comparing nowcasts of India and the real GDP of six other countries, achieving at least a 30% reduction in out-of-sample Root Mean Square Error (RMSE) compared to benchmark models (i.e., Random Walk and Combinatorial approach).

B. Regularization in Machine Learning (ML) Models

Another approach to addressing the parameter proliferation problem is the use of regularization techniques, such as the LASSO, Ridge regression, and Elastic Net, which impose penalties on coefficients that are less relevant to the target variable (e.g., Tibshirani, 1996; Hoerl & Kennard, 1970; Zoe & Hastie, 2005). Machine learning (ML) models, particularly those that emphasize regularization techniques, have significantly influenced statistical learning and predictive modeling, and are specifically designed to tackle issues such as multicollinearity, overfitting, and variable selection in high-dimensional datasets.

- **Ridge regression (Hoerl & Kennard, 1970):** This regularization technique addresses multicollinearity in linear regression by adding a penalty on the size of coefficients, improving prediction accuracy while retaining all variables. It adds the square of the coefficients as a penalty to the loss function and shrinks coefficients towards zero, but does not eliminate any variables. This means all predictors remain in the model. This method stabilizes estimates and reduces variance without significantly increasing bias, making it particularly useful for large datasets. Studies such as Feng, Giglio, and Xiu (2020) demonstrate its effectiveness in financial forecasting, while Kumar and Thenmozhi (2014) show its utility in predicting economic indicators like GDP and inflation. Ridge regression is very similar to OLS regression, except that the coefficients are estimated by minimizing a modified loss function using the squared magnitude of the model's coefficients. Formally:

$$\beta_{ridge} = argmin[\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + Penalty]$$

$$Penalty = \lambda \sum_{j=1}^{p} \beta_j^2$$

  In here, λ is a hyperparameter that determines the term's overall importance ($\lambda = 0$ would remove the penalty altogether and would be equivalent to fitting an OLS model, whereas increasing λ would result in a progressively less complex model, with coefficient estimates closer to zero. Note that the penalty hyperparameter of Ridge can shrink the size of the coefficient estimates, but does not reduce them all the way to zero.

- **LASSO (Tibshirani, 1996):** a regularization method that incorporates a penalty to the loss function, enabling both shrinkage and automatic variable selection by setting some coefficients exactly to zero, improving forecasting performance by reducing the added variance from irrelevant variables with small coefficients. LASSO has found applications in various domains, including economics, bioinformatics, and text mining for predictive modeling and variable selection. It gained prominence in the machine learning and statistical communities for its ability to perform variable selection and regularization simultaneously, particularly in the context of large datasets where overfitting and multicollinearity can be significant. Banbura et al. (2010) focused on Bayesian methods and incorporated LASSO as a comparative method for nowcasting GDP and inflation using large datasets of economic indicators. LASSO regression is very similar to Ridge regression, but uses a different penalty function. Instead of the squared magnitude, LASSO uses the absolute value of the same penalty term. Formally:

$$\beta_{LASSO} = argmin[\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + Penalty]$$

$$Penalty = \lambda \sum_{j=1}^{p} |\beta_j|$$

Note that, unlike Ridge, the penalty in LASSO can shrink specific coefficients to zero. Thus, LASSO can be used for feature selection, meaning it will automatically choose the subset of candidate variables that is most informative in predicting GDP. High λ suggests that more variables will be discarded.

- **Elastic Net (Zou & Hastie, 2005):** a regularization technique that combines the properties of both Ridge and Lasso regression. It adds penalties to the loss function, effectively addressing the limitations of LASSO with highly correlated variables. Elastic Net incorporates both penalties, making it suitable for scenarios where there are correlations among predictors. This model demonstrates its ability to select variables in the presence of multicollinearity and its grouping effect, where strongly correlated predictors are selected together. Schmidt and Møller (2016) compare Elastic Net and LASSO regression models for demand forecasting in the retail sector, using large datasets of sales data, and find that Elastic Net outperforms LASSO in scenarios where predictor variables are highly correlated, leading to more accurate and reliable demand forecasts. Elastic Net Regression is a combination of Ridge and LASSO, incorporating a hybrid of the two penalties. More formally:

$$\beta_{Elastic\ Net} = argmin[\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + Penalty]$$

$$Penalty = \lambda \left(\sum_{j=1}^{p} (1-\alpha)\beta_j^2 + \alpha|\beta_j|\right)$$

Here, both λ and α are the hyperparameters, with relative weights decided by the additional hyperparameter α. Similarly, the combination of λ and α are chosen via a cross-validation procedure to minimize expected out-of-sample performance.

C. Dimensionality Reduction using Principal Component Analysis (PCA)

The parameter proliferation problem can also be mitigated using Principal Component Analysis (PCA), a statistical technique that reduces the dimensionality of a dataset. It identifies the most important patterns and relationships among variables by transforming the original variables into a new set of uncorrelated variables, known as principal components. PCA can be used in nowcasting models to reduce the number of variables and capture the most relevant information for forecasting. Some applications of PCA in economics, specifically in constructing composite indicators, estimating latent factors, and improving forecasting accuracy, are discussed below.

Dimension reduction techniques, such as PCA, reduce the number of predictors by summarizing information across various variables (e.g., Stock & Watson, 2002; Forni et al., 2005). Model selection and shrinkage techniques, such as information criteria and stepwise regression, select the most important variables (e.g., Akaike, 1974; Draper & Smith, 1998).

- **Stock and Watson (1989)** introduced the use of PCA in constructing composite indexes of economic indicators. They applied PCA to a large dataset of economic variables. They identified the common factors driving business cycles, which led to the creation of popular economic indicators such as the Index of Leading Economic Indicators.

  PCA addresses the high-dimensional problem in macroeconomic forecasting by modeling the covariability of the series in terms of a relatively few unobserved latent factors. Stock and Watson (2002) proposed a two-step process to conduct forecasting using PCA. The first step is estimating a few time series of Principal Component Factors from the pool of selected predictors. The second step is to estimate the relationship between the target variable of the forecast and the factors using linear regression.

  Specifically, let y_t to be the scalar time series variable to be forecast and let X_t be a N-dimensional multiple time series of candidate predictors. It is assumed that (X_t, y_{t+h}) admit a factor model representation with r common latent factors F_t.

  $$X_t = \Lambda F_t + e_t$$

  and

  $$y_{t+h} = \beta_F' F_t + \beta_w' w_t + e_{t+h}$$

  where e_t is a N x 1 vector idiosyncratic disturbances, h is the forecast horizon, w_t is a m x 1 vector of observed variables, that together with F_t are useful for forecasting y_{t+h}, and e_{t+h} is the resulting forecasting error. Data are available for $\{y_t, X_t, w_t\}_{t=1}^{T}$, and the goal is to forecast y_{t+h}.

- **Giannone, Reichlin, and Small (2008)** applied PCA in nowcasting, where they combined high-frequency macroeconomic data with PCA to estimate real-time economic conditions. They showed that PCA-based models can provide accurate nowcasts of key macroeconomic variables, enabling timely decision-making. Bai and Ng (2002) developed statistical methods for determining the optimal number of factors in approximate factor models. They used PCA to estimate the number of common factors that explain the co-movement of economic variables and introduced a rigorous approach based on eigenvalues to determine the relevant factors

PCA has been combined with other methods, such as MIDAS and U-MIDAS, to enhance their performance in various applications. For instance:

- **Ahir, Loungani, and Stock (2019)** applied a MIDAS model combined with PCA to analyze inflation expectations. They used PCA to extract common factors from a large set of economic indicators. They incorporated them into a MIDAS framework to estimate inflation expectations, taking into account the heterogeneity of information contained in the dataset.

- **Mählmann and Gebauer (2016)** explored the combination of PCA with MIDAS models for nowcasting. They used PCA to extract common factors from a large panel of monthly indicators. They incorporated these factors into a MIDAS framework to enhance the accuracy of short-term economic forecasts.

- **Hallin, Liska, and van den Akker (2020)** employed a dynamic multilevel factor model, which combines principal component analysis (PCA) with multivariate integrated dynamic analysis (MIDAS), to analyze European macroeconomic survey data. They applied PCA to extract common factors from a panel of survey indicators. They incorporated these factors into a MIDAS framework to capture the dynamics and heterogeneity of the survey data.

IV. Application: Nowcasting China's Real GDP During COVID-19

Nowcasting China's real GDP is crucial, not only for timely and effective policymaking in China but also for assessing the global impact of China's growth[3]. As the world's second-largest economy, fluctuations in China's growth can disrupt supply chains, commodity prices, and financial markets. While China's GDP data are released relatively promptly compared to many peer countries, about two weeks after the end of each quarter, nowcasting remains important. Official GDP figures are published quarterly and often revised later, which can delay real-time economic assessment and policymaking. Meanwhile, nowcasting provides immediate estimates within the quarter using high-frequency indicators, which helps identify turning points or significant shifts in the economy's momentum, allowing for quicker policy responses (Giannone *et al*., 2008; Zhang *et al*., 2018).

The importance of nowcasting China's economy was especially clear during the COVID-19 pandemic, when China's economy faced sharp contractions and uneven recoveries. Traditional GDP reports lag, missing critical disruptions caused by lockdowns and supply chain bottlenecks. Nowcasting, on the other hand, detected shifts in industrial output, consumer activity, and freight movements far earlier than official statistics. As China continues to navigate post-pandemic recovery, real estate volatility, and geopolitical tensions, nowcasting is an essential tool for responding to economic changes and understanding their global ripple effects.

That said, nowcasting GDP has been particularly challenging in many emerging market countries, such as China, due to several factors, including the relatively short history of macroeconomic data, missing observations, variability in data quality, and various structural changes. In particular, the various structural changes and transformations may render some high-frequency variables, which are important for GDP nowcasting in certain periods, less significant in other periods. In the case of China, which was largely closed off before the 1970s, the country has undergone rapid integration into the global economy and has been rebalancing from an export- and investment-driven growth model to one focused increasingly on domestic consumption. These structural shifts could make nowcasting China's GDP, especially the selection of high-frequency indicators, particularly challenging.

[3] Previous IMF Working Papers on aspects of nowcasting national GDP include: Heng et al. (2024) for Cambodia; Heenan et al. (2025) for Samoa; and Ouliaris and Rochon (2023).

Our approach is similar to that in the literature, which nowcasts China's GDP growth (e.g., Giannone *et al*., 2013; Zhang *et al*., 2023). Some of the literature employs a single type of nowcasting model, such as the dynamic factor model, while others use multiple methods and compare their performance. For example, Zhang *et al*. (2023) conducted a "horse-race" comparison of various models for nowcasting China's GDP growth, including traditional nowcasting methods and machine learning (ML) approaches. They found that specific machine learning (ML) techniques, particularly ridge regression, outperformed the benchmark dynamic factor model. In contrast, our paper focuses more on variable selection for both traditional nowcasting and ML models, exploring whether the performance of traditional models can be improved to match, or even surpass, that of ML models through more effective variable selection. Additionally, we focus specifically on the ability of these models to predict economic downturns in the context of the COVID-19 pandemic, an area that has not been extensively explored in prior research.

A. Data Preparation

To nowcast China's real GDP (available quarterly), we collect **166** monthly indicators, ranging from November 1952 to April 2023. Our indicators list covers a wide range of macroeconomic sectors, including consumption, firm and production, prices, external environment, survey or forward-looking indicators, real estate, financial variables, government, foreign trade, transportation. Starting with a full candidate list, we conduct a preliminary filtration to exclude data with limited sampling (i.e., data that starts after 2011Q1). This procedure ensures a sufficiently balanced sample in the baseline regression.

All filtered indicators (a total of 132 variables, see Annex 2 for more details) are subject to five additional preparation steps: (i) Year-to-Date to flow conversion; (ii) nominal term deflation; (iii) seasonal adjustment; (iv) stationarity transformation; (v) frequency conversion. More specifically:

(i) Year-to-Date to Flow Conversion:

Some of China's economic indicators are only available in a Year-to-Date form, representing the value accumulated from the beginning of each year. Other indicators are available in a Flow form, indicating the value that occurred over a specific period within the year (e.g., each month or each quarter). To ensure consistency, we convert all year-to-date variables to flow format.

In our list, only one variable needs such conversion: "*China: Online Retail Sales of Goods and Services (YTD, NSA, 100 Mil. CNY)*". Because the timing of the Chinese New Year on the lunar calendar fluctuates between January and February, the National Bureau of Statistics of China (CNBS) combines the values for these two months. It reports them in February, with a missing January value. We first disaggregate the January and February data using the number of working days as the weighting factor. Let $w_{t,Jan}$ denotes weight for January at year t, $NWD_{t,Jan}$ represents the number of working days in January; Let $w_{t,Feb}$ denotes weight for February at year t, $NWD_{t,Feb}$ represents the number of working days in February.

$$w_{t,Jan} = \frac{NWD_{t,Jan}}{(NWD_{t,Jan} + NWD_{t,Feb})}$$

$$w_{t,Feb} = \frac{NWD_{t,Feb}}{(NWD_{t,Jan} + NWD_{t,Feb})}$$

We can then disentangle January and February value using the reported value in February (including production in both January and February) X_{Feb}^{YTD} and respective weights.

$$X_{Jan}^{Flow} = w_{t,Jan} * X_{Feb}^{YTD}$$

$$X_{Feb}^{Flow} = w_{t,Feb} * X_{Feb}^{YTD}$$

Where X_{Jan}^{Flow} and X_{Feb}^{Flow} denotes the disaggregated flow value of January and February.

For other months, we simply take the difference from the previous month to get the flow value, indicating the additional production/value generated during the specific month:

$$X_{m}^{Flow} = X_{m}^{YTD} - X_{m-1}^{YTD}$$

In which X_{m}^{Flow} denotes the flow value at month *m*; X_{m}^{YTD} and X_{m-1}^{YTD} denotes Year-to-Date value at month *m* and *m-1*.

(ii) Nominal Variables Deflation:

Since we aim to nowcast China's real GDP, we need to ensure all indicators are in real terms. We define nominal variables into five groups, based on the appropriate deflators to use:

a) ***Export Value Index***: appropriate for export-related indicators.

b) ***Import Value Index****:* suitable for import-related indicators

c) ***Purchasing Price Index****:* applicable for indicators relevant to investment, inventory, real estate, and capacity utilization.

d) ***Consumer Price Index- Consumer Goods****:* applicable for indicators relevant to retail sales.

e) ***Consumer Price Index****:* suitable for other nominal indicators

The formula used for deflation is simply:

$$X_{real} = 100 * \frac{X_{nominal}}{Deflator}$$

Where X_{real} is the deflated series and $X_{nominal}$ is the nominal original series.

(iii) Seasonal Adjustment

Most monthly macroeconomic indicators in China are subject to seasonal fluctuations. For example, agricultural production varies with the seasons, which causes seasonal adjustments to the overall volume and value of agricultural production. Additionally, consumer goods tend to spike in demand before public holidays, leading to fluctuating retail sales. Many traditional Chinese festivals, such as Lunar New Year and Mid-Autumn Festival, are dictated by the lunar calendar and therefore vary from year to year (Roberts and White, 2015).

To account for the impacts of shifting holidays to seasonal patterns, we utilize the X-13ARIMA-SEATS Seasonal Adjustment Program, proposed by the US Census Bureau, along with holiday regressors in EViews (US Census Bureau, 2023). We create Lunar New Year holiday regressors using “Win Genhol[4]” program, which counts for effects before, after, and during the holiday.

(iv) Stationarity

To ensure the stationarity of each variable in the regressions, we apply the appropriate first-difference or log-difference transformation. The first difference transformation is used for an index or series with non-positive observations. Log difference transformation applies to the remaining non-stationary series. No transformation is applied to most interest rates and the PMI index, which usually do not contain unit roots.

(v) Frequency Conversion

The final preparation required is to convert all monthly indicators into a target frequency (i.e., quarterly), which enables us to create the baseline nowcasting model (i.e., the bridge approach). We determine the aggregation method based on the “flow” and “stock/index” attributes of each indicator. For “flow” variables, we use the “sum” approach that adds up the monthly value for the quarterly value. For “stock/index” variables, we use the “average” approach that takes the average of the monthly values for the quarterly entry. For the indicators’ report at the end of the period, we take the value of the last month in the quarter.

B. Methodology

We applied three approaches in selecting the high-frequency variables for the nowcasting practices, while tackling parameter proliferation issues:

1) **AS-ARIMAX:** AS-ARIMAX (Xie, 2023) is a modified stepwise procedure that shifts the focus from statistical significance to the overall forecasting improvement that can be attributed to a specific exogenous variable. Starting with no exogenous variables in the model except for the constant term, the AS-ARIMAX procedure tests each variable separately and adds it to the baseline model if it has an estimated coefficient that is consistent with economic prior and yields superior forecasting performance.

   Specifically, the procedure decides whether a variable (X_t) is a suitable candidate based on three criteria: (i) the X_t decreases the Akaike Information Criteria (AIC) value[5], compared to the model without X_t. (ii) the coefficient sign of X_t matches economic priors. (iii) X_t is statistically significant at the 5% confidence level.

   The AS-ARIMAX variable selection procedure involves three steps:

   - **Step 1**: We add the first candidate indicator X_1 as an exogenous regressor to the automatic ARIMA procedure for the target variable (i.e., Model 1-A). Then we repeat the procedure without X_1 (i.e., Model 1-B). If Model 1-A has a lower AIC value than Model 1-B, X_1 passes condition 1. If the

[4] Win Genhol program creates holiday regressors using the same procedure as X-13ARIMA-SEATS uses to create regressors for the U..S holidays of Easter and Labor Day. More information available here: https://www.census.gov/data/software/x13as/genhol.html

[5] A lower AIC values indicate a better-fit model

coefficient sign of X_1 in Model 1-A matches economic priors, X_1 passes condition 2. Once X_1 passes both conditions 1 and 2, we keep X_1. Otherwise, it is discarded.

- **Step 2:** We repeat the same procedure as in Step 1 with all other selected variables, one by one. Then, we evaluate the p-value of each variable to evaluate the validity of condition 3. Meanwhile, we need to verify whether condition 2 remains valid as we adjust the model specifications.

- **Step 3:** After ensuring all independent variables meet the three previous conditions, we need to manually check the significance level of the selected ARIMA orders to ensure those orders are meaningful in the model. We may start by removing the ARIMA term with the highest non-significant t-statistics, until all ARIMA terms are statistically significant at 15% level and regressors are statistically significant at 5% level while having intuitive coefficient signs.[6]

2) **Machine Learning:** We first use LASSO regularization to select regressors that are the most predictive of the dependent variable under the LASSO's penalization framework. With the selected regressors, we then implement LASSO, Ridge, and Elastic Net to produce nowcast estimates of the Real GDP. In the process, we applied three types of sign restrictions to manually adjust the regression:

   - *Restriction #1: Sign Consistency with Aggressive Dropping:* Incorporate variables with reasonable signs based on economic intuition and drop variables with perceived incorrect signs at once.

   - *Restriction #2: Stepwise Sign-Based Dropping*: Removing variables with incorrect signs sequentially

   - *Restriction #3: Relaxed Sequential Sign-Based Restriction*: Applies the sequential dropping of variables with incorrect signs, but retains more variables compared to Restriction #2.

3) **PCA:** use PCA technique to reduce the number of parameters in the model. Instead of selecting variables, this approach weighs all pre-filtered high-frequency indicators by the PCA weights. We then bring the Principal Components generated to all four nowcasting techniques.

C. Results and Evaluation

Pseudo Out-of-Sample Experimental Design

We assess the performance of the nowcasting models using an out-of-sample evaluation. Specifically, we follow Zhang *et al*. (2023) and employ a rolling fixed window estimation method,[7] where the first M observations are used to estimate the parameters of each model, with M representing the number of in-sample periods and $T - M$ indicating the number of out-of-sample periods. The first nowcast is generated at time period $M + 1$. To compute the subsequent nowcast for period $M + 2$, we remove the first observation from the previous

[6] Economic judgement is also used in the variable selection according to the expected signs consistent with economic theory or intuition. For example, some of financial and prices variables tend to have "opposite" signs after controlling for the key variables and are hence dropped from the regressions (see Table 2).

[7] As pointed out by Zhang *et al*. (2023), one important advantage of using the fixed window approach is its ability to address model instability, which is an issue in China's nowcasting or forecasting. By excluding "old" data from the estimation process, the fixed window method prevents biases in parameter estimates. Additionally, fixed window estimates tend to be more stable than those obtained from expanding window methods, making more efficient use of the available data in situations where the parameters do not change significantly over time (Elliott and Timmermann, 2016).

estimation sample and add an additional observation to the end of it. The parameters of each model are re-estimated, and the nowcast is re-generated based on the updated parameters. This process—updating the estimation sample, re-estimating the model parameters, and generating a new nowcast—is repeated until we have computed $T - M$ out-of-sample nowcasts.

The first step in this exercise is to split the total sample between in-sample and out-of-sample periods to ensure a sufficient number of observations for model selection and evaluation. Considering the relatively limited size of the entire sample (2007Q2-2023Q1), we seek to maximize the use of available data in estimation. Since our focus is on the nowcasts for post-COVID periods, we specify the first in-sample period as the period from 2007Q2 (the starting sample period) to 2019Q4 and the first out-of-sample period as 2020Q1. In other words, the first out-of-sample nowcast of China's real GDP growth rate is produced for 2020Q1. To compute this nowcast, we deflate all nominal variables using the appropriate deflator indices, conduct seasonal adjustments to all series, perform the proper transformations to remove the unit root (see above, *A. Data Preparation*), and estimate the models using data only available for the first in-sample period. To compute the next nowcast for 2020Q2, we re-estimate the model parameters using data from 2007Q3 to 2020Q1.[8] This process repeats until the final simulated out-of-sample nowcast is made for 2023Q1.

Model Performance

In this section, we discuss the results of our analysis, examining the performance of various models for nowcasting China's real GDP growth. We compute the RMSE of different methods to identify which yields the lowest RMSE. Table 1 presents the RMSE results, along with other forecast evaluation statistics for each nowcasting method, over the simulated out-of-sample period from 2020Q1 to 2023Q1.

In terms of RMSE, the best model identified is based on the Machine Learning method, which utilizes Lasso regularization and sign restrictions guided by economic intuition (Lasso + Sign Restriction 3 – see Table 1.). This model achieved an RMSE of 0.01472, the lowest among those tested. However, this result is not significantly better than the model employing a simpler selection method, such as the AS-ARIMA selection method combined with a Bridge nowcast model (see Figure 1 (a)), which recorded an RMSE of 0.01655. Meanwhile, models that utilize the PCA for variable selection performed worse than the other two approaches.

We present the model performance of all the models we tested in Figure 1-3 below. Notably, it is evident that the model employing the regularization approach using LASSO for variable selection performed well only after being guided by a sign restriction filter. Prior to applying these sign restrictions, allowing the regularization to run did not yield strong forecast performance, particularly in capturing the turning points during the peak of the COVID pandemic and the subsequent period (see Figure 2 (a)). The ML Selected + Sign Restriction 1 Model enhances the results of the ML Selected Model by incorporating variables with reasonable signs based on economic intuition and dropping variables with perceived incorrect signs at once, thereby reflecting the expected direction of the variables in relation to real GDP growth (see Figure 2(b)). The ML + Sign Restriction 2 approach involves removing variables with incorrect signs sequentially, based on their correlation with GDP (see Figure 2(c)). This resulted in slightly poorer performance compared to the ML Selected + Sign Restriction 1. Meanwhile, the ML + Sign Restriction 3 model also applies the sequential dropping of variables with incorrect signs but retains more variables compared to ML + Sign Restriction 2. This model achieved the best

[8] To mitigate the potential structural change problem after COVID, we add back all the variables that are statistically significant with both pre- and post-COVID samples to the models before re-estimating the parameters.

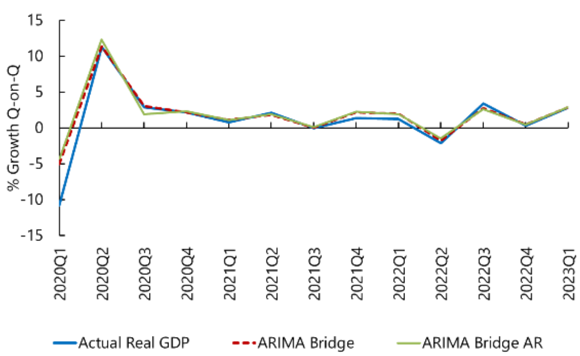
performance among the three ML models and even outperformed the ARIMA Bridge Model. This leads us to conclude that while regularization through Machine Learning is helpful in the selection of variables, not applying additional filters does not guarantee good forecasting performance.

We also applied a simple econometric method for selecting variables, specifically the AS-ARIMAX method proposed by Xie (2023), as discussed in Section 1. Despite its simplicity, this selection method performed well when combined with the Bridge Model and the Bridge model with an autoregressive form of some selected variables. This selection method performs well when combined with DFM nowcasting model (see Figure 1(d)). However, this variable selection approach did not yield strong results when integrated with the MIDAS, and U-MIDAS (see Figure 1 (b) and (c)). Based on the results, the AS-ARIMAX selection method demonstrates best forecasting performance for China's Real GDP when combined with the Bridge Model.

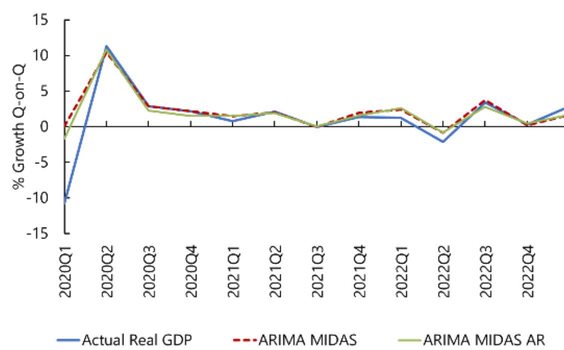
We further evaluated the third selection method, which is PCA. The selected principal components were then combined with the Bridge, MIDAS, U-MIDAS, and DFM nowcasting models. Unfortunately, this selection method performed the worst among the three variable selection methods tested.

Figure 1. Model Performance: Approach #1 AS-ARIMAX

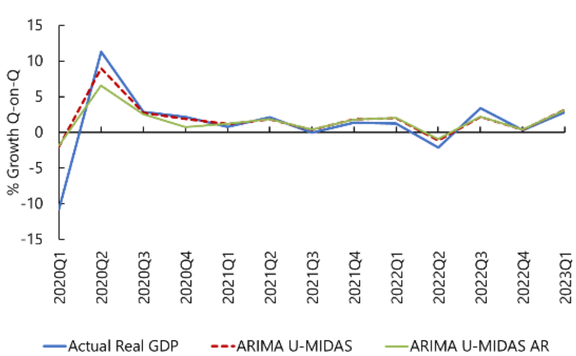
(a) Bridge



(b) MIDAS



(c) *U-MIDAS*



(d) DFM

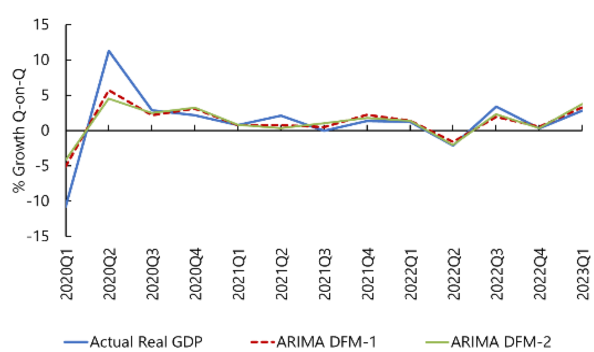
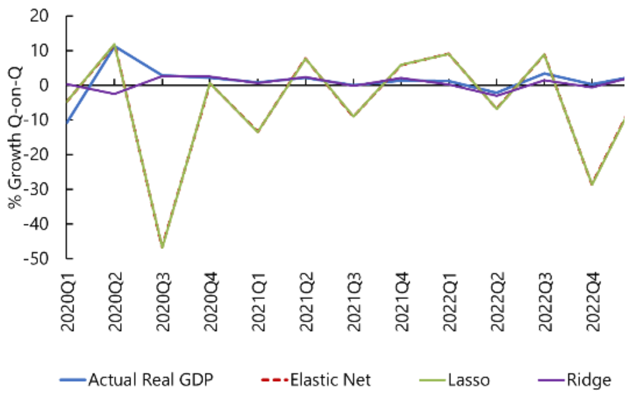
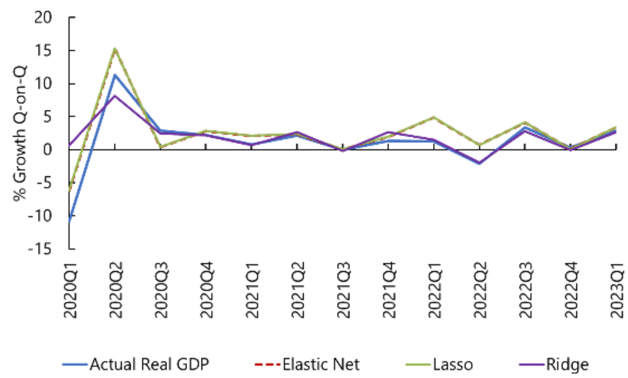


Figure 2. Model Performance: Approach #2 ML Regularization[1]

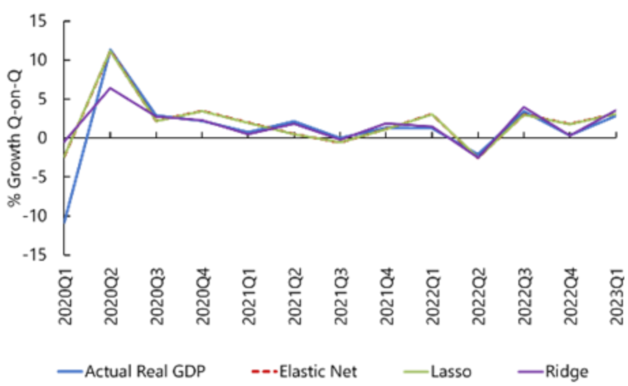
(a) ML Selected



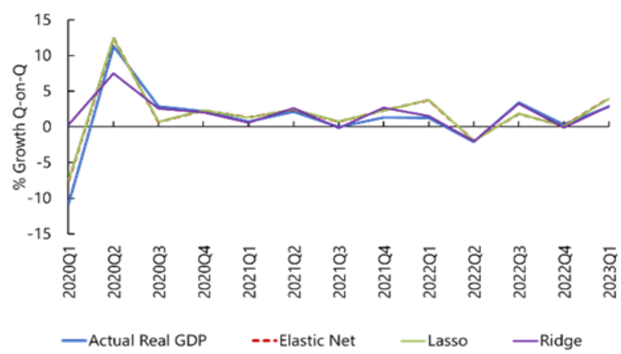
(b) ML Selected + Sign Restriction #1



(c) ML Selected + Sign Restriction #2



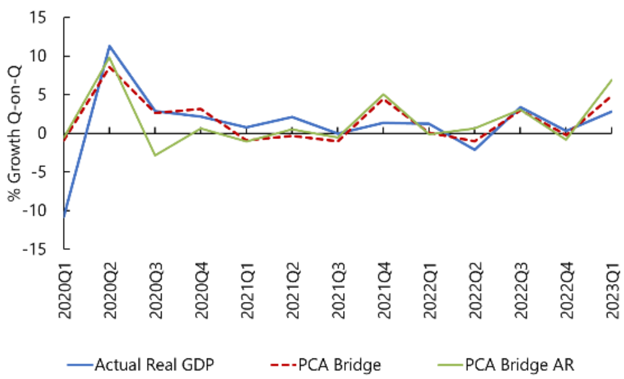
(d) ML Selected + Sign Restriction #3



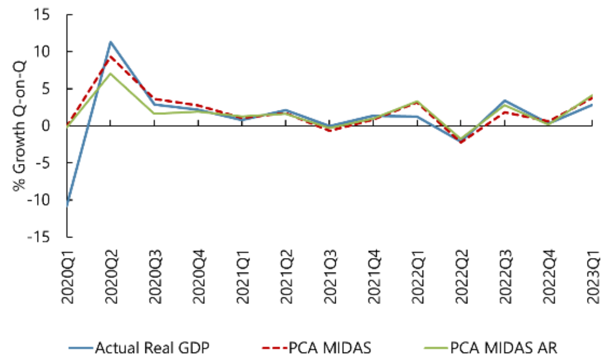
[1] Note: Restriction #1: Sign Consistency with Aggressive Dropping: Incorporate variables with reasonable signs based on economic intuition and drop variables with perceived incorrect signs at once. Restriction #2: Stepwise Sign-Based Dropping: Removing variables with incorrect signs sequentially. Restriction #3: Relaxed Sequential Sign-Based Restriction: Applies the sequential dropping of variables with incorrect signs but retain more variables compared to Restriction #2.

Figure 3. Model Performance: Approach #3 PCA

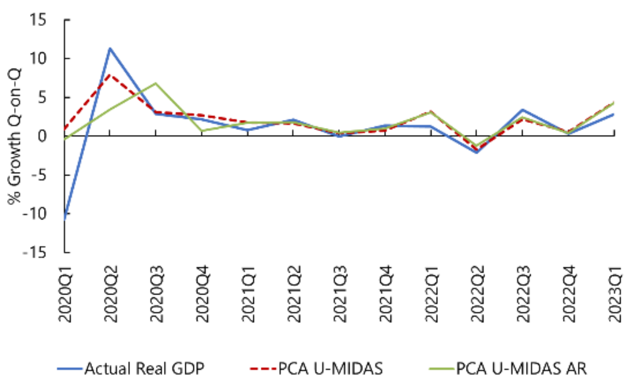
(a) Bridge



(b) MIDAS



(c) U-MIDAS



(d) DFM

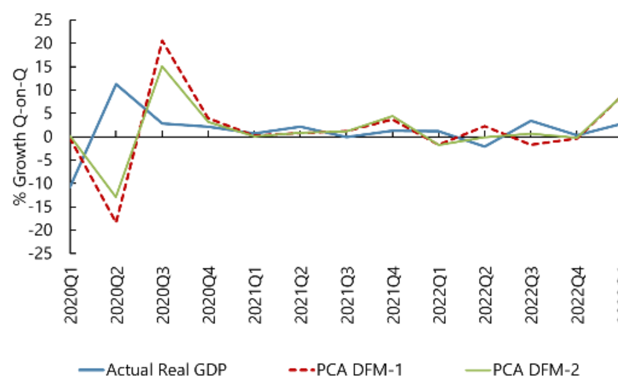


Table 1. Forecast Evaluation Statistics

Variable Selection Method	Forecast Method	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
AS-ARIMAX	ARIMA Bridge	0.01655	0.00738	39.79116	38.93664	0.19134	0.61225
	ARIMA Bridge AR	0.01889	0.00957	65.30593	43.64072	0.21580	0.65301
	ARIMA MIDAS	0.03079	0.01331	42.53706	47.79812	0.37692	0.42940
	ARIMA MIDAS AR	0.02639	0.01284	63.37614	58.11414	0.32362	0.18167
	ARIMA U-MIDAS	0.02576	0.01267	145.33050	49.63640	0.32968	0.32776
	ARIMA U-MIDAS AR	0.02902	0.01586	162.69102	60.51722	0.40148	0.31468
	ARIMA DFM-1	0.01667	0.00880	165.00059	47.16819	0.19355	0.30408
	ARIMA DFM-2	0.01927	0.00960	91.31718	55.98097	0.22699	0.18954
Machine Learning	Elastic Net + Sign Restrictions 3	0.01474	0.01140	270.62625	62.67548	0.15914	0.68731
	Lasso + Sign Restrictions 3	0.01472	0.01140	271.94160	62.91497	0.15894	0.68777
	Ridge + Sign Restrictions 3	0.03273	0.01427	75.16274	57.66361	0.43091	0.94864
PCA	PCA Bridge	0.03188	0.02096	374.91815	111.12632	0.39479	2.24488
	PCA Bridge AR	0.03828	0.02797	274.83789	139.85580	0.44807	2.65632
	PCA MIDAS	0.03134	0.01592	234.80063	60.07593	0.38852	0.36554
	PCA MIDAS AR	0.03282	0.01732	124.61489	61.55009	0.43766	0.25251
	PCA U-MIDAS	0.03483	0.01805	174.48336	66.29033	0.44612	0.43968
	PCA U-MIDAS AR	0.03877	0.02382	196.60818	79.62352	0.51671	0.26775
	PCA DFM-1	0.06324	0.03626	172.58405	110.91486	0.79160	1.57841
	PCA DFM-2	0.05065	0.03143	354.62378	123.77881	0.69896	1.66961
	Simple Mean	0.02654	0.01428	52.23770	58.20832	0.34593	0.64110

Note: DFM-1" uses one factor and "DFM-2" uses two factors in estimating the model.

Selected Variables and Performance

From the results and evaluation conducted, it is apparent that the performance of the model varies significantly, with variable selection playing a crucial role. Below, we present the lists of variables selected by each method. Notably, the AS-ARIMAX variable selection method identifies variables related to macroeconomics, firms production, and government indicators, while largely overlooking many financial and prices indicators, except for deposit rates and sector loans. This pattern is similar to the dimensionality reduction achieved through PCA. In contrast, the ML Lasso regularization selection method includes a wide range of variables from almost all categories, except for real estate indicators (see Table 2 below).

Table 2. Selected Variables by Different Approaches

LIST OF VARIABLES	TRANSFORMATION	AS-ARIMAX		Regularization in ML: LASSO				PCA	
		Bridge, MIDAS, UMIDAS, DFM 1, DFM 2	Bridge AR, MIDAS AR, UMIDAS AR	SELECTED: Ridge, Lasso, Elastic Net	SELECTED + SIGN RES 1: Ridge, Lasso, Elastic Net	SELECTED + SIGN RES 2: Ridge, Lasso, Elastic Net	SELECTED + SIGN RES 3: Ridge, Lasso, Elastic Net	Bridge, MIDAS, UMIDAS, DFM 1, DFM 2	Bridge AR, MIDAS AR, UMIDAS AR
China: Gross Domestic Product (SA, Bil.2020.Yuan)	DLOG_RGDP(-1)		✓						✓
China: Gross Domestic Product (SA, Bil.2020.Yuan)	DLOG_RGDP(-2)		✓						✓
Consumption:									
China: Retail Sales (SA, 100 Mil.Yuan)	DLOG(RETAIL)			✓	✓	✓			
Firm and Production:									
China: Index of Industrial Value Added (SA, 2005=100)	DLOG(II)	✓	✓	✓	✓			✓	✓
China: Fixed Investment: Manufacturing [Revised] (NSA, 100 Mil.Yuan)	DLOG(FI_MAN)	✓	✓					✓	✓
China: Industrial Product: Coal (NSA, 10,000 Metric Tons)	DLOG(IP_COAL)	✓	✓					✓	✓
China: IP: Gasoline (NSA, 10,000 Metric Tons)	DLOG(IP_GAS)	✓	✓					✓	✓
China: Output: Motor Vehicles (NSA, 10,000.Units)	DLOG(OUT_MV)	✓	✓	✓		✓	✓	✓	✓
China: Output: Cement (NSA, 10,000.Metric Tons)	DLOG(OUT_CMT)	✓	✓					✓	✓
China: Invest in Fixed Assets, ex Rural Households [Revised] (NSA, 100 Mil.Yuan)	DLOG(INV_FA)			✓	✓	✓	✓		
China: IP: Natural Gas (NSA, 100 Mil Cubic M)	DLOG(IP_NAT_GAS)			✓	✓	✓			
China: IP: Processed Crude Oil (NSA, 10,000 Metric Tons)	DLOG(IP_PC_OIL)								
Principal Component of: China PMI: Manufacturing (SA, 50+=Expansion) China PMI: Manufacturing Employment (SA, 50+=Expansion) China: PMI: Manufacturing (SA, 50+=Expansion) China: PMI: Manufacturing Employment (SA, 50+=Expansion)	PMI_M_PC1			✓	✓	✓	✓		
China: Output: Steel Products (NSA, 10,000.Metric Tons)	DLOG(OUT_SP)			✓	✓	✓	✓		
China: Output: Plate Glass (NSA, 10,000.Weight.Cases)	DLOG(OUT_PG)			✓	✓	✓	✓		
China: Output: Alternating Current Motors (NSA, 10,000.Kwatt)	DLOG(OUT_AL_CM)			✓	✓	✓	✓		
China: IP: Coke (NSA, 10,000 Metric Tons)	DLOG(IP_COKE)			✓	✓	✓	✓		
China: IP: Electricity Production (SA, 100 Mil.KWH)	DLOG(IP_EP)			✓	✓				
China: Freight Traffic [Cargo] (SA, 100 Mil.Metric Tons)	DLOG(FT)			✓	✓	✓	✓		
China: Output: Steel (NSA, 10,000.Metric Tons)	DLOG(OUT_S)			✓	✓	✓	✓		
China: IP: Aluminum Oxide (NSA, 10,000 Metric Tons)	DLOG(IP_ALU)			✓	✓	✓	✓		
China: Output: Internal Combustion Engines (NSA, 10,000.KWatt)	DLOG(OUT_ICE)			✓	✓	✓	✓		
China: Output: Yarn (NSA, 10,000.Metric Tons)	DLOG(OUT_Y)			✓	✓	✓			
China: IP: Diesel Oil (NSA, 10,000 Metric Tons)	DLOG(IP_DIESEL)			✓	✓				
China: Output: Cotton Cloth (NSA, 100 Mil.Meters)	DLOG(OUT_COTTON)			✓	✓	✓	✓		
China: Output: Household Refrigerators (NSA, 10,000.Units)	DLOG(OUT_REF)			✓					
China: Industrial Product: Coal (NSA, 10,000 Metric Tons)	DLOG(IP_COAL)			✓	✓				
China: Real Estate: Newly Started Construction (SA, Mil.Sq.Meters)	DLOG(RR_NEW_CST)			✓	✓				
China: Output: Cement (NSA, 10,000.Metric Tons)	DLOG(OUT_CMT)			✓	✓				
Government:									
China: General Govt Revenue: Domestic Consumption Tax (Bil.Yuan)	DLOG(GGV_REV_DCT)	✓	✓	✓	✓			✓	✓
China: General Govt Revenue (SA, Bil.Yuan)	DLOG(GGV_REV)			✓	✓				
Foreign Trade:									
China: Imports: Crude Petroleum Oil (NSA, Mil.Metric Tons)	DLOG(IM_PETRO)	✓	✓					✓	✓
China: Exports: Mechanical & Electrical Products (NSA, Bil.USD)	DLOG(EX_MECH_ELEC)			✓	✓	✓	✓		
China: Import Volume Index (NSA, 2010=100, no missing)	DLOG(IM_VOL)			✓	✓	✓	✓		
China: Export Volume Index (NSA, 2010=100, no missing)	DLOG(EX_VOL)			✓	✓	✓	✓		
China: JP Morgan Broad Nominal Effective Exchange Rate (2010=100)	DLOG(NEER)			✓	✓			✓	✓
Survey or Forward-Looking Indicators:									
China PMI: Services Business Activity (SA, 50+=Expansion)	PMI_SERV_CAIXIN	✓	✓	✓	✓			✓	✓
Financial Variables:									
China: Uses of Funds by Sector Loans (EOP,SA, 100 Mil Yuan)	DLOG(UF_SL)	✓	✓					✓	✓
Principal Component of: China: Deposit Rates: 3-Month Certificates of Deposit (% per annum) China: Deposit Rates: 6-Month Certificates of Deposit (% per annum)	DEP_R_PC1	✓	✓		✓			✓	✓
Principal Component of: China: Shanghai-Shenzhen-300 Stock Price Index (EOP, Dec-31-2004=1000) China: Shanghai Stock Price Index [SSE] (EOP, Dec-19-90=100) China: Shenzhen Stock Price Index: Components (EOP, Jul-20-94=1000)	SP_PC1			✓	✓	✓	✓		
China: Prime Lending Rate (AVG, % per annum)	PLR			✓	✓				
Hong Kong SAR: Stock Price Index: Hang Seng Bank (Jul-31-64=100)	DLOG(HK_HSB)			✓	✓	✓	✓		
Principal Component of: China: Shibor: 1-Week (%) China: Shibor: 3-Month (%)	SHIBOR_PC1			✓					
Principal Component of: Nominal Lending Rate: Within 1 Year (Including 1 Year) CN: Nominal Lending Rate: 1-5 Year (Including 5 Year) Nominal Lending Rate: Over 5 Year	NLR_PC1			✓	✓	✓			
China: Required Reserve Ratio (EOP, %)	RRR			✓		✓			
Prices:									
China: PPI: Consumer Goods (SA, 2020=100)	DLOG(PPI_CG)			✓	✓				
China: CPI: Transportation and Communications (SA, 2020=100)	DLOG(CPI_T)								
China: CPI: Recreation, Education & Cultural Services (SA, 2020=100)	DLOG(CPI_RECS)			✓	✓				
China: CPI: Food (SA, 2020=100)	DLOG(CPI_F)			✓	✓				
China: CPI: Housing (SA, 2020=100)	DLOG(CPI_H)			✓	✓				
China: Producer Prices: All Industry Products (SA, 2020=100)	DLOG(PP_ALL)			✓	✓				
China: PPI: Producer Goods (SA, 2020=100)	DLOG(PPI_PG)			✓	✓				
China: CPI: Clothing (SA, 2020=100)	DLOG(CPI_C)			✓	✓				
China: CPI: Medicine (SA, 2020=100)	DLOG(CPI_M)			✓	✓				
External Environment:									
Developed Markets PMI: Composite (SA, 50+=Expansion)	PMI_AE			✓	✓	✓			
CBOE VIX Volatility Index [VVIX] (AVG, Index)	VVIX_USA			✓		✓	✓		
Global PMI: Services Business Activity (SA, 50+=Expansion)	GPMI_SERV			✓	✓	✓			
Global PMI: Composite Output (SA, 50+=Expansion)	GPMI			✓	✓	✓			
Global Manufacturing PMI Using Markit Mfg for U.S. (SA, 50+=Expansion)	GPMI_MFG			✓	✓				
Real Estate:									
Principal Component of: China: Tier-2 Cities: Price of Existing Residential Buildings (NSA, 2020=100) China: Tier-3 Cities: Price of Existing Residential Buildings (NSA, 2020=100)	HP_23_PC1								
Transportation:									
China: Volume of Transportation in Costal Ports (SA, 100 Mil.Metric Tons)	DLOG(AT_CP)			✓	✓				
China: Passenger Traffic (SA, 100 Mil.Persons)	DLOG(PASS_TFC)			✓	✓				

V. Conclusion

The availability of several high-frequency indicators is advantageous for nowcasting; however, caution must be exercised when constructing nowcasting models due to the issue of parameter proliferation, which can introduce excessive noise into forecasts. A related concern arises when an extensive number of variables are present within the information set, leading to what is commonly referred to as the "curse of dimensionality." This phenomenon occurs when models include many high-frequency predictors relative to the number of observations. Consequently, parameter proliferation can lead to overfitting, where the model captures noise instead of the underlying patterns, ultimately diminishing forecasting accuracy. Given the significance of the parameter proliferation issue in nowcasting, it is essential to explore approaches that can effectively distinguish informative variables from mere noise.

We evaluate three approaches to mitigate parameter proliferation: (i) variable selection using AS-ARIMAX; (ii) regularization techniques in ML models; and (iii) dimensionality reduction through PCA using Chinese data. Utilizing 166 variables, we estimate our models from 2007Q2 to 2019Q4 using rolling-window regression. Subsequently, we conduct an out-of-sample performance comparison of various nowcasting models, including Bridge, MIDAS, U-MIDAS, DFMs, and ML techniques such as Ridge Regression, LASSO, and Elastic Net, to nowcast China's real GDP from 2020Q1 to 2023Q1. Our findings indicate that the LASSO method outperforms all other models, but only when guided by economic judgment and sign restrictions in variable selection. Notably, simpler models like Bridge provide reliable estimates that are nearly comparable to those from LASSO, underscoring the importance of effective variable selection in capturing strong signals.

The significance of variable selection using AS-ARIMAX cannot be overstated. Simple models, such as the Bridge model, which employs the AS-ARIMAX variable selection approach along with economic judgment and sign restrictions, can yield forecasting results comparable to (and sometimes superior to) those derived from more complex parametric regularization methods. In contrast, the PCA approach is less desirable as it treats all variables as a weighted average, potentially diluting the impact of more relevant predictors. Regularization methods like LASSO can enhance forecast accuracy and effectively complement variable selection when combined with economic judgment and sign restrictions.

Annex I. Data Description

No.	Description	Deflation Method	Seasonal adjustment	Aggregation	Start Date	End Date	Source
1	CBOE Market Volatility Index: VIX (Index)	No	Yes--regular X13	Average	Jan-1990	Mar-2023	Wall Street Journal
2	CBOE VIX Volatility Index [VVIX] (AVG, Index)	No	Yes--regular X13	Average	Jan-2007	Mar-2023	Chicago Board Options Exchange
3	China PMI: Manufacturing (SA, 50+=Expansion)	No	No	Average	Apr-2004	Mar-2023	Caixin/S&P Global
4	China: Acquisition of Land by Real Estate Developers (SA, Mil.Sq.Meters)	No	No	Sum	Jan-2001	Dec-2022	China National Bureau of Statistics/Haver Analytics
5	China: Aggregate Social Financing to the Real Economy (EOP, SA, Tril.Yuan)	CPI	No	End of Period	Dec-2002	Mar-2023	People's Bank of China/Haver Analytics
6	China: Composite PMI Output Index (SA, 50+=Expansion)	No	No	Average	Jan-2017	Mar-2023	China Federation of Logistics & Purchasing/CNBS
7	China: Consumer Confidence (SA, 100+=Optimistic)	No	No	Average	Jan-1990	Feb-2023	China National Bureau of Statistics/Haver Analytics
8	China: Consumer Price Index (SA, 2020=100)	No	No	Average	Jan-1984	Mar-2023	China National Bureau of Statistics
9	China: CPI Excluding Food and Energy (SA, 2020=100)	No	No	Average	Jan-2005	Mar-2023	China National Bureau of Statistics
10	China: CPI: Clothing (SA, 2020=100)	No	No	Average	Dec-2000	Mar-2023	China National Bureau of Statistics
11	China: CPI: Food (SA, 2020=100)	No	No	Average	Dec-2000	Mar-2023	China National Bureau of Statistics
12	China: CPI: Housing (SA, 2020=100)	No	No	Average	Dec-2000	Mar-2023	China National Bureau of Statistics
13	China: CPI: Medicine (SA, 2020=100)	No	No	Average	Dec-2000	Mar-2023	China National Bureau of Statistics
14	China: CPI: Recreation, Education & Cultural Services (SA, 2020=100)	No	No	Average	Dec-2000	Mar-2023	China National Bureau of Statistics
15	China: CPI: Transportation and Communications (SA, 2020=100)	No	No	Average	Dec-2000	Mar-2023	China National Bureau of Statistics
16	China: Deposit Rates: 3-Month Certificates of Deposit (% per annum)	No	No	Average	May-1990	Mar-2023	People's Bank of China
17	China: Deposit Rates: 6-Month Certificates of Deposit (% per annum)	No	No	Average	May-1990	Mar-2023	People's Bank of China
18	China: Electricity Consumption (SA, Bil.KWH)	No	No	Sum	Jul-2008	Mar-2023	China Electricity Council
19	China: Export Volume Index (NSA, 2010=100)	No	Yes--CHN SA	Average	Jan-2005	Feb-2023	General Administration of Customs, China
20	China: Exports: Mechanical & Electrical Products (NSA, Bil.USD)	Export price index	Yes--CHN SA	Sum	Jan-2001	Mar-2023	General Administration of Customs, China
21	China: Exports: Steel Products (NSA, Bil.USD)	Export price index	Yes--CHN SA	Sum	Jan-2000	Mar-2023	General Administration of Customs, China
22	China: Financial Conditions Index (0+=Tightening)	No	No	Average	Sep-2008	Feb-2023	Yicai Research Institute
23	China: Fix Inv: Farming/For/Animal Husbandry/Fishery [Rev] (NSA, 100 Mil.Yuan)	PPI	Yes--CHN SA	Sum	Jan-1999	Feb-2023	China National Bureau of Statistics
24	China: Fixed Investment: Manufacturing [Revised] (NSA, 100 Mil.Yuan)	PPI	Yes--CHN SA	Sum	Jan-2003	Feb-2023	China National Bureau of Statistics
25	China: Fixed Investment: Real Estate [Revised] (NSA, 100 Mil.Yuan)	PPI	Yes--CHN SA	Sum	Jan-2003	Feb-2023	China National Bureau of Statistics
26	China: Freight Traffic [Cargo] (SA, 100 Mil.Metric Tons)	No	No	Sum	Jan-2002	Feb-2023	China National Bureau of Statistics/Haver Analytics
27	China: General Govt Expenditure (SA, Bil.Yuan)	CPI	No	Sum	Jan-1998	Feb-2023	Ministry of Finance of China/Haver Analytics
28	China: General Govt Revenue (SA, Bil.Yuan)	CPI	No	Sum	Jan-1998	Feb-2023	Ministry of Finance of China/Haver Analytics
29	China: General Govt Revenue: Domestic Consumption Tax (Bil.Yuan)	CPI	Yes--CHN SA	Sum	Jan-2007	Feb-2023	Ministry of Finance of China/Haver Analytics
30	China: General Govt Revenue: Domestic Value Added Tax (Bil.Yuan)	CPI	Yes--CHN SA	Sum	Jan-2007	Feb-2023	Ministry of Finance of China/Haver Analytics
31	China: General Retail Price Index (SA, 2020=100)	No	No	Average	Jan-1985	Dec-2022	China National Bureau of Statistics/Haver Analytics
32	China: Government Bond Yield: 1 Year (AVG, % p.a)	No	No	Average	Mar-2006	Mar-2023	The People's Bank of China
33	China: Government Bond Yield: 10 Year (AVG, % p.a)	No	No	Average	Mar-2006	Mar-2023	The People's Bank of China
34	China: Government Bond Yield: 3 Month (AVG, % p.a)	No	No	Average	Mar-2006	Mar-2023	The People's Bank of China
35	China: Government Bond Yield: 3 Year (AVG, % p.a)	No	No	Average	Mar-2006	Mar-2023	The People's Bank of China
36	China: Government Bond Yield: 30 Year (AVG, % p.a)	No	No	Average	Mar-2006	Mar-2023	The People's Bank of China
37	China: Government Bond Yield: 5 Year (AVG, % p.a)	No	No	Average	Mar-2006	Mar-2023	The People's Bank of China
38	China: Government Bond Yield: 6 Month (AVG, % p.a)	No	No	Average	Mar-2006	Mar-2023	The People's Bank of China
39	China: Government Bond Yield: 7 Year (AVG, % p.a)	No	No	Average	Mar-2006	Mar-2023	The People's Bank of China
40	China: Import Volume Index (NSA, 2010=100)	No	Yes--CHN SA	Average	Jan-2005	Feb-2023	General Administration of Customs, China
41	China: Imports: Crude Petroleum Oil (NSA, Mil.Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Mar-2023	General Administration of Customs, China
42	China: Index of Industrial Value Added (SA, 2005=100)	No	No	Average	Jan-1997	Feb-2023	China National Bureau of Statistics/Haver Analytics
43	China: Industrial Product: Coal (NSA, 10,000 Metric Tons)	No	Yes--CHN SA	Sum	Jan-1989	Feb-2023	China National Bureau of Statistics
44	China: Industrial Production: Sales Ratio (NSA, %)	No	Yes--CHN SA	Average	Jul-2011	Dec-2022	China National Bureau of Statistics
45	China: Inv in Fixed Assets: Infrastructure [Revised] (NSA, 100 Mil.Yuan)	PPI	Yes--CHN SA	Sum	May-2013	Feb-2023	China National Bureau of Statistics
46	China: Inventory of Industrial Enterprises (EOP, NSA,Mil.Yuan)	PPI	Yes--CHN SA	End of Period	Feb-2010	Feb-2023	China National Bureau of Statistics
47	China: Invest in Fixed Assets, ex Rural Households [Revised] (NSA, 100 Mil.Yuan)	PPI	Yes--CHN SA	Sum	Jan-1997	Feb-2023	China National Bureau of Statistics/Haver Analytics
48	China: Investment Completed in Real Estate Development (SA, Bil.Yuan)	PPI	No	Sum	Jan-2000	Feb-2023	China National Bureau of Statistics/Haver Analytics
49	China: IP: Aluminum Oxide (NSA, 10,000 Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
50	China: IP: Coke (NSA, 10,000 Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
51	China: IP: Crude Oil (NSA, 10,000 Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
52	China: IP: Diesel Oil (NSA, 10,000 Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
53	China: IP: Electricity Production (SA, 100 Mil.KWH)	No	No	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics/Haver Analytics
54	China: IP: Gasoline (NSA, 10,000 Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
55	China: IP: Natural Gas (NSA, 100 Mil Cubic M)	No	Yes--CHN SA	Sum	Jan-2007	Feb-2023	China National Bureau of Statistics
56	China: IP: Processed Crude Oil (NSA, 10,000 Metric Tons)	No	Yes--CHN SA	Sum	Mar-1998	Feb-2023	China National Bureau of Statistics
57	China: JP Morgan Broad Nominal Effective Exchange Rate (2010=100)	No	No	Average	Jan-1970	Mar-2023	JP Morgan
58	China: JPMorgan Real Broad Effective Exchange Rate Index, CPI Based (2010=100)	No	No	Average	Jan-1990	Mar-2023	JP Morgan
59	China: JPMorgan Real Broad Effective Exchange Rate Index, PPI Based (2010=100)	No	No	Average	Jan-1990	Mar-2023	JP Morgan
60	China: Max Supply Price: Gas # 89 [II] for Military/Other Dept (AVG, RMB/Ton)	No	Yes--CHN SA	Average	Jan-2009	Mar-2023	National Development and Reform Commission

No.	Description	Deflation Method	Seasonal adjustment	Aggregation	Start Date	End Date	Source
61	China: Money Supply: M1 (EOP,SA, Bil.Yuan)	CPI	No	End of Period	Dec-1998	Mar-2023	People's Bank of China/Haver Analytics
62	China: Money Supply: M2 (EOP,SA, Bil.Yuan)	CPI	No	End of Period	Dec-1998	Mar-2023	People's Bank of China/Haver Analytics
63	China: Monthly ECRI Peaks/Troughs (+1 or -1)	No	No	Not Allowed	Jan-1984	Jan-2023	Economic Cycle Research Institute/Haver Analytics
64	China: Monthly ECRI Recession/Expansion (+1 or 0)	No	No	Not Allowed	Jan-1984	Jan-2023	Economic Cycle Research Institute/Haver Analytics
65	China: News-Based Econ Policy Uncertainty Index [Mainland](Jan.00-Dec.18=100)	No	No	Average	Jan-2000	Mar-2023	PolicyUncertainty.com
66	China: News-Based Economic Policy Uncertainty Index (Mean=100)	No	No	Average	Jan-1995	Mar-2023	PolicyUncertainty.com
67	China: Online Retail Sales of Goods and Services (YTD, NSA, 100 Mil.CNY)	CPI of goods	Yes--CHN SA	End of Period	Feb-2014	Feb-2023	China National Bureau of Statistics
68	China: Output: Alternating Current Motors (NSA, 10,000.Kwatt)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
69	China: Output: Cement (NSA, 10,000.Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
70	China: Output: Chemical Fertilizers (NSA, 10,000.Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
71	China: Output: Cotton Cloth (NSA, 100 Mil.Meters)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
72	China: Output: Household Refrigerators (NSA, 10,000.Units)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
73	China: Output: Household Washing Machines (NSA, 10,000.Units)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
74	China: Output: Internal Combustion Engines (NSA, 10,000.KWatt)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
75	China: Output: Motor Vehicles (NSA, 10,000.Units)	No	Yes--CHN SA	Sum	Jan-1997	Feb-2023	China National Bureau of Statistics
76	China: Output: Plate Glass (NSA, 10,000.Weight.Cases)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
77	China: Output: Steel (NSA, 10,000.Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
78	China: Output: Steel Products (NSA, 10,000.Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
79	China: Output: Yarn (NSA, 10,000.Metric Tons)	No	Yes--CHN SA	Sum	Feb-1998	Feb-2023	China National Bureau of Statistics
80	China: Passenger Traffic (SA, 100 Mil.Persons)	No	No	Sum	Jan-2002	Feb-2023	China National Bureau of Statistics/Haver Analytics
81	China: PMI: Manufacturing (SA, 50+=Expansion)	No	No	Average	Jan-2005	Mar-2023	China Federation of Logistics & Purchasing/CNBS
82	China: PMI: Manufacturing Employment (SA, 50+=Expansion)	No	No	Average	Jan-2005	Mar-2023	China Federation of Logistics & Purchasing/CNBS
83	China: PMI: Nonmanufacturing Business Activity: Services (SA, 50+=Expansion)	No	No	Average	May-2012	Mar-2023	China National Bureau of Statistics
84	China: PPI: Consumer Goods (SA, 2020=100)	No	No	Average	Jan-1996	Mar-2023	China National Bureau of Statistics/Haver Analytics
85	China: PPI: Producer Goods (SA, 2020=100)	No	No	Average	Jan-1996	Mar-2023	China National Bureau of Statistics/Haver Analytics
86	China: Prime Lending Rate (AVG, % per annum)	No	No	Average	Sep-1972	Mar-2023	People's Bank of China
87	China: Private Investment in Fixed Assets [Revised] (NSA, 100 Mil.CNY)	PPI	Yes--CHN SA	Sum	Apr-2011	Feb-2023	China National Bureau of Statistics
88	China: Producer Prices: All Industry Products (SA, 2020=100)	No	No	Average	Jan-1996	Mar-2023	China National Bureau of Statistics/Haver Analytics
89	China: Profits (SA, Bil.Yuan)	PPI	No	Sum	Oct-2011	Feb-2023	China National Bureau of Statistics/Haver Analytics
90	China: Purchasing Price Index of Raw Materials (NSA, M/M %Chg)	No	Yes--CHN SA	Not Allowed	Jan-2011	Mar-2023	China National Bureau of Statistics
91	China: Real Estate Investment: Land Acquisition Costs (NSA, Bil.Yuan)	PPI	Yes--CHN SA	Sum	Jan-2002	Feb-2023	China National Bureau of Statistics
92	China: Real Estate: Buildings Sold (SA, Mil.Sq.Meters)	No	No	Sum	Jan-2000	Feb-2023	China National Bureau of Statistics/Haver Analytics
93	China: Real Estate: Newly Started Construction (SA, Mil.Sq.Meters)	No	No	Sum	Jan-2000	Feb-2023	China National Bureau of Statistics/Haver Analytics
94	China: Real Industrial Value Added (SA, M/M %Chg)	No	No	Not Allowed	Feb-2011	Feb-2023	China National Bureau of Statistics
95	China: Required Reserve Ratio (EOP, %)	No	No	End of Period	Jan-1985	Mar-2023	People's Bank of China
96	China: Required Reserve Ratio: Medium Depository Institution (EOP, %)	No	No	End of Period	Mar-2018	Mar-2023	People's Bank of China
97	China: Required Reserve Ratio: Small Depository Institution (EOP, %)	No	No	End of Period	Apr-2018	Mar-2023	People's Bank of China
98	China: Reserve Requirement Ratio: Deposits in Yuan: Large Dep Insts (EOP, %)	No	No	End of Period	Aug-2008	Mar-2023	People's Bank of China
99	China: Reserve Requirement Ratio: Small/Medium Depositary Insts (EOP, %)	No	No	End of Period	Jan-1985	Apr-2019	People's Bank of China
100	China: Reserve Requirement: Interest Paid on Excess Reserves (EOP, %)	No	No	End of Period	May-1993	Mar-2023	People's Bank of China
101	China: Reserve Requirement: Interest Paid on Reserves (EOP, %)	No	No	End of Period	May-1993	Mar-2023	People's Bank of China
102	China: Retail Sales (SA, 100 Mil.Yuan)	CPI of goods	No	Sum	Oct-1983	Feb-2023	China National Bureau of Statistics/Haver Analytics
103	China: RMB Exchange Rate: U.S. (Average, Yuan/100 US$)	No	No	Average	Jan-1972	Mar-2023	State Administration of Foreign Exchange
104	China: Savings Deposit Rate (% per annum)	No	No	Average	May-1990	Mar-2023	People's Bank of China
105	China: Services Sector Production Index (NSA, Y/Y %Chg)	No	No	Not Allowed	Mar-2016	Dec-2022	China National Bureau of Statistics
106	China: Shanghai Stock Price Index [SSE] (EOP, Dec-19-90=100)	No	Yes--CHN SA	End of Period	Dec-1990	Mar-2023	Shanghai Stock Exchange
107	China: Shanghai-Shenzhen-300 Stock Price Index (EOP, Dec-31-2004=1000)	No	Yes--CHN SA	End of Period	Apr-2005	Mar-2023	Shanghai Stock Exchange
108	China: Shenzhen Stock Price Index: Components (EOP, Jul-20-94=1000)	No	Yes--CHN SA	End of Period	Jan-2004	Mar-2023	Shenzhen Stock Exchange
109	China: Shibor: 1-Week (%)	No	Yes--CHN SA	Average	Oct-2006	Mar-2023	National Interbank Funding Center
110	China: Shibor: 3-Month (%)	No	Yes--CHN SA	Average	Oct-2006	Mar-2023	National Interbank Funding Center
111	China: Social Financing: RMB Bank Loans (EOP, NSA, Tril.Yuan)	CPI	Yes--CHN SA	End of Period	Dec-2002	Mar-2023	People's Bank of China
112	China: Sources by Sector: Deposits of NonFin Enterprises(EOP,NSA,100 Mil. Yuan)	CPI	Yes--CHN SA	End of Period	Jan-2008	Feb-2023	People's Bank of China
113	China: Sources of Funds Deposits (EOP,NSA,100 Mil. Yuan)	CPI	Yes--CHN SA	End of Period	Jan-2006	Mar-2023	People's Bank of China
114	China: Steel Industry PMI (SA, 50+=Expansion)	No	No	Average	Mar-2012	Mar-2023	CFLP Steel Logistics Professional Committee
115	China: Tier-1 Cities: Price of Existing Residential Buildings (NSA, 2020=100)	No	Yes--CHN SA	Average	Jan-2010	Mar-2023	China National Bureau of Statistics/Haver Analytics
116	China: Tier-2 Cities: Price of Existing Residential Buildings (NSA, 2020=100)	No	Yes--CHN SA	Average	Jan-2010	Mar-2023	China National Bureau of Statistics/Haver Analytics
117	China: Tier-3 Cities: Price of Existing Residential Buildings (NSA, 2020=100)	No	Yes--CHN SA	Average	Jan-2010	Mar-2023	China National Bureau of Statistics/Haver Analytics
118	China: Trade Policy Uncertainty [Mainland] (Jan.00-Dec.18=100)	No	No	Average	Jan-2000	Mar-2023	PolicyUncertainty.com
119	China: Turnover of Trading (100 Mil Yuan)	CPI	Yes--CHN SA	Average	Jan-2002	Feb-2023	People's Bank of China
120	China: Uses of Funds by Sector Loans (EOP,SA, 100 Mil Yuan)	CPI	No	End of Period	Jan-2003	Mar-2023	People's Bank of China

No.	Description	Deflation Method	Seasonal adjustment	Aggregation	Start Date	End Date	Source
121	China: Value of Buildings Sold (SA, Bil.Yuan)	CPI	No	Sum	Jan-2000	Feb-2023	China National Bureau of Statistics/Haver Analytics
122	China: Value of Foreign Direct Investment Actually Utilized (YTD, NSA, Mil.US$)	PPI	Yes--CHN SA	End of Period	Dec-1983	Feb-2023	Ministry of Commerce/Haver Analytics
123	China: Volume of Transportation in Costal Ports (SA, 100 Mil.Metric Tons)	No	No	Sum	Jan-2002	Feb-2023	China National Bureau of Statistics/Haver Analytics
124	China: Wholesale Price 200 Index: Agricultural Products (NSA, 2015=100)	No	Yes--CHN SA	Average	Feb-2015	Mar-2023	Ministry of Agriculture of China
125	China: Yicai Chief Economists Confidence Index (NSA, 50+=Positive View)	No	Yes--CHN SA	Average	Oct-2010	Apr-2023	Yicai Research Institute
126	Hong Kong: Stock Price Index: Hang Seng Bank (Jul-31-64=100)	No	Yes--CHN SA	Average	Dec-1969	Mar-2023	Wall Street Journal
127	University of Michigan: Consumer Sentiment (NSA, Q1-66=100)	No	Yes--CHN SA	Average	Nov-1952	Apr-2023	University of Michigan
128	US PMI: Composite Output [Flash] (SA,50+=Expansion)	No	No	Average	Oct-2009	Mar-2023	S&P Global
129	Global Manufacturing PMI Using Markit Mfg for U.S. (SA, 50+=Expansion)	No	No	Average	Jan-1998	Mar-2023	JP Morgan/S&P Global
130	China: Number of Working Days (NSA, Days)	No	Yes--CHN SA	Sum	Jan-1981	Dec-2023	State Council of the People's Republic of China
131	China: CPI: Consumer Goods (SA, 2020=100)	No	No	Average	Dec-2004	Mar-2023	China National Bureau of Statistics
132	China: Online Retail Sales of Goods and Services (YTD, NSA, 100 Mil.CNY)	No	Yes--CHN SA	End of Period	Feb-2014	Feb-2023	China National Bureau of Statistics

References

Ahir, H., Loungani, P., and Stock, J. H. "Inflation Expectations and the Missing Disinflation." Journal of Money, Credit and Banking, 51(7), (2019): 1831-1855.

Akaike, H. "A new look at the statistical model identification," in IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716-723, December 1974, doi: 10.1109/TAC.1974.1100705.

Akepanidtaworn, K. and Akepanidtaworn, K. (2025). GDP Nowcasting Performance of Machine Learning versus non-Machine Learning Models. IMF Working Paper (forthcoming).

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression Models with Mixed Sampling Frequencies. Journal of Econometrics, 158(2), 246-261.

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting: A handbook for researchers and practitioners* (p. 564). Kluwer Academic Publishers.

Bai, J., and Ng, S. "Determining the Number of Factors in Approximate Factor Models." Econometrica, 70(1), (2002): 191-221.

Banbura, M., Giannone, D., and Reichlin, L. (2010). "Large Bayesian Vector Auto Regressions." Journal of Applied Econometrics, 25(1), 71-92.

BańBura, M., Giannone, D., and Reichlin, L. (2012) 'Nowcasting', in Michael P. Clements, and David F. Hendry (eds), The Oxford Handbook of Economic Forecasting, Oxford Handbooks (2011); online edn, Oxford Academic, 18 Sept. 2012), https://doi.org/10.1093/oxfordhb/9780195398649.013.0008, accessed 10 Jan. 2025.

Breiman, L. (2001). Random Forests. *Machine Learning* **45**, 5–32. https://doi.org/10.1023/A:1010933404324

Castle, J., Doornik, J., and Hendry, D. (2011). Evaluating Automatic Selection. Journal of Time Series Econometrics 3(1):8-8.

Castle, J.L., Qin, X., and Reed, R. W. (2009). "How To Pick The Best Regression Equation: A Review And Comparison Of Model Selection Algorithms," Working Papers in Economics 09/13, University of Canterbury, Department of Economics and Finance.

Chan, J., Koop, G., Poirier, D. J., and Tobias, J. L. (2019). "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure." Journal of Econometrics, 212(1), 97-115.

Clements, M. P., and Galvão, A. B. (2008). Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States. Journal of Business & Economic Statistics, 26(4), 546-554.

Doz, C., Giannone, D., and Reichlin, L. (2011). "A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models." Review of Economics and Statistics, 94(4), 1014-1024.

Draper, N. R., and Smith, H. (1998). *Applied regression analysis* (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781118625590

Feng, G., Giglio, S., and Xiu, D. (2020). Predicting Stock and Portfolio Returns with Lasso and Ridge: Evidence from the US Equity Market. Journal of Financial Economics, 135(3), 599-622.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association*, *100*(471), 830–840. http://www.jstor.org/stable/27590616

Foroni, C., and Marcellino, M. (2013). A Survey of Econometric Methods for Mixed-Frequency Data. Norges Bank Research Working Paper 2013/06.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The Real Time Informational Content of Macroeconomic Data. Journal of Monetary Economics, 55: 665-676.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies. Journal of Econometrics, 131(1-2), 59-95.

Ghysels, E., and Wright, J. H. (2009). Forecasting Professional Forecasters. Journal of Business & Economic Statistics, 27(4), 504-516.

Giannone, D., Reichlin, L., and Small, D. (2008). "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." Journal of Monetary Economics, 55(4), 665-676.

Hallin, M., Liska, R., & van den Akker, R. "Dynamic Multilevel Factor Models with an Application to European Macroeconomic Survey Data." Journal of Applied Econometrics, 35(7), (2020): 812-832.

Hendry, David F. and Hans-Martin Krolzig, "We Ran One Regression." Oxford Bulletin of Economics and Statistics, Vol. 66, No. 5 (2004): 799-810.

Heenan, G.M., Lui, K., Nield, I., Muaulu, E., Reupena, V., Sabuga, I., and Tolovaa, A. (2025). Nowcasting Real GDP in Samoa. IMF Working Paper 25/92, International Monetary Fund.

Heng, D., Han, F., Chey, S., Uch, R., Kuchsa, D., and Phork, P. (2024). Nowcasting and Near-Term Forecasting Cambodia's Economy. IMF Working Paper 24/247, International Monetary Fund.

Higgins, P.C.: GDPNow: a model for GDP "Nowcasting". SSRN Electron. J. (2014). https://doi.org/10.2139/ssrn.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55-67.

Hoover, Kevin D. and Stephen J. Perez, "Truth and Robustness in Cross-country Growth Regressions." Oxford Bulletin of Economics and Statistics_Vol. 66, No. 5 (2004): 765-798.

Klein, L.R. and E. Sojo (1989) "Combinations of High and Low Frequency Data in Macroeconometric Models" In Klein and Marquez (eds), Economics in Theory and Practice: An Eclectic Approach. Dordrecht: Kluwer, pp. 3-16.

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2019). A Bayesian Multivariate Functional Dynamic Linear Model. Journal of the American Statistical Association, 114(527), 1258-1271.

Kumar, M., and Thenmozhi, M. (2014). Ridge Regression and its Application to Economic Forecasting. Journal of Quantitative Economics, 12(2), 151-166.

Kuzin, V., Marcellino, M., and Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. International Journal of Forecasting, 27(2), 529-542.

Li, S., Zhang, Y., and Xu, J. (2019). Nowcasting Air Quality Indices using Ridge Regression and Satellite Data. Atmospheric Environment, 209, 1-8.

Mählmann, T., and Gebauer, S. "Nowcasting with MIDAS: Model Choice and Forecast Combination." Journal of Business Cycle Research, 12(2), (2016):189-218.

Mariano, R.S., and Murasawa, Y. (2003). "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series." Journal of Applied Econometrics, 18(4), 427-443.

Matteson, D. S., Tsay, R. S., and Preve, D. (2019). "Factor Modeling for High-Dimensional Time Series: Inference and Model Selection." Journal of Time Series Analysis, 40(2), 241-264.

Miller, J. I., and Gencay, R. (2012). Adaptive U-MIDAS models for high-frequency predictions. International Journal of Forecasting, 28(2), 500-511.

Ouliaris, S., and Rochon, C. (2023). Assessing the Impact of Policy Changes on a Nowcast. IMF Working Paper 23/153, International Monetary Fund.

Oxley, L. T. “An Expert Systems Approach to Econometric Modelling.” Mathematics and Computers in Simulation, Vol. 39 (1995): 379-383.

Phillips, Peter C. B. “Automated Discovery in Econometrics.” Econometric Theory Vol. 21, No. 1 (2005): 3-20.

Roberts, I. and White, G. (2015). “Seasonal Adjustment of Chinese Economic Statistics,” Research Discussion Paper, RDP 2015-13, Reserve Bank of Australia.

Richardson, A., van Florenstein Mulder, T., and Vehbi, T. (2021). Nowcasting GDP Using Machine Learning Algorithms: A Real Time Assessment. International Journal of Forecasting, 37: 941-48.

Sala-i-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller, “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach.” American Economic Review Vol. 94, No. 4 (2004): 813-835.

Schmidt, T., and Møller, J. K. (2016). Demand Forecasting in Retail: A Comparison Between Lasso and Elastic Net Regression Models. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 123-134.

Siliverstovs, B. (2011). Multivariate MIDAS models: A unified approach to forecasting with mixed-frequency data. KOF Swiss Economic Institute.

Stock, J. H., and Watson, M. W. (1989). "New Indexes of Coincident and Leading Economic Indicators." NBER Macroeconomics Annual, 4 (1989): 351-409.

Stock, J. H., and Watson, M. W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. https://doi.org/10.1198/073500102317351921

Stock, J.H., and Watson, M.W. (2005). "Understanding Changes in International Business Cycle Dynamics." Journal of the European Economic Association, 3(5), 968-1006.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1), 267-288.

United States Census Bureau, 2023. “X-13ARIMA-SEATS Seasonal Adjustment Program,” retrieved from https://www.census.gov/data/software/x13as.html.

Zhang, Q., Ni, H. and Xu, H. (2023). Nowcasting Chinese GDP in a Data-Rich Environment: Lessons from Machine Learning Algorithms. Economic Modelling, Vol 122, 106-204.

Zou, H., and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.

PUBLICATIONS

Parameter Proliferation in Nowcasting: Issues and Approaches - An Application to Nowcasting China's Real GDP
Working Paper No. WP/2025/217