Forecasting models for the Chinese macroeconomy: the simpler the better?

Chris Heaton[1] · **Natalia Ponomareva**[1] · **Qin Zhang**[1]

Abstract

We consider the problem of macroeconomic forecasting for China. Our objective is to determine whether well-established forecasting models that are commonly used to compute forecasts for Western macroeconomies are also useful for China. Our study includes 19 different forecasting models, ranging from simple approaches such as the naive forecast to more sophisticated techniques such as ARMA, Bayesian VAR, and factor models. We use these models to forecast two different measures of price inflation and two different measures of real activity, with forecast horizons ranging from 1 to 12 months, over a period that stretches from March 2005 to December 2018. We test null hypotheses of equal mean squared forecasting error between each candidate model and a simple benchmark. We find evidence that AR, ARMA, VAR, and Bayesian VAR models provide superior 1-month-ahead forecasts of the producer price index when compared to simple benchmarks, but find no evidence of superiority over simple benchmarks at longer horizons, or for any of our other variables.

✉ Chris Heaton
chris.heaton@mq.edu.au

Natalia Ponomareva
natalia.ponomareva@mq.edu.au

Qin Zhang
qin.zhang1@students.mq.edu.au

[1] Department of Economics, Macquarie University, Sydney, Australia

1 Introduction

With an annual GDP of around US$13.6 trillion[1] in 2018, the Chinese economy is now the second largest in the world and, if current growth rates continue, it will become the largest within the next few years. From being a predominantly closed country in the 1970s, China is now well integrated into the world economy. As examples, in 2017–2018, it accounted for 33.7% of Australia's exports,[2] in February 2019 it held approximately 18% of US Treasury Securities,[3] and in 2018 it provided 20% of the imports of the European Union.[4]

Surprisingly, despite the importance of the Chinese economy, there is a paucity of published academic research on forecasting the Chinese macroeconomy. This stands in stark contrast to the published literature on forecasting the macroeconomies of Western countries, which covers a wide range of forecasting models over long periods of time. Examples include Ang et al. (2007), Artis et al. (2005), Atkeson and Ohanian (2001), Bandt et al. (2007), Camba-Mendez et al. (2001), Ciccarelli and Mojon (2010), Faust and Wright (2013), Forni et al. (2003), Marcellino et al. (2003), Pincheira and Medel (2015), den Reijer (2005), Schneider and Spitzer (2004), Schumacher (2007), Stock and Watson (1999), Stock and Watson (2002b), and Stock and Watson (2008).

In this paper, we consider the following question: Do the models that are well established for forecasting the Western macroeconomies also work well for China? Although one might hope that this is the case, the relatively short time span of available Chinese macroeconomic data, variations in data quality and the rapid pace of ongoing structural change could limit the effectiveness of forecasts obtained using established methods. We are particularly interested in determining whether more sophisticated forecasting models that require parameter estimation (e.g. ARMA models, VAR models, and factor models) and forecasting models with fixed tuning parameters (e.g. moving averages and exponential smoothing models) provide superior forecasts to simple parameter-free methods (e.g. the naive forecast and the mean forecast).

The scholarly journals have published a small number of papers on forecasting Chinese macroeconomic variables, but none are sufficient to answer the question that we pose above. Mehrotra and Sánchez-Fung (2008), Lin and Wang (2013), Kamal (2013), Zhou et al. (2013), and He and Fan (2015) all conduct comparisons of alternative forecasting models for Chinese macroeconomic variables, and between them cover a broad range of models. However, these studies consider only a very small number of forecasts. Mehrotra and Sánchez-Fung (2008) and Lin and Wang (2013) provide forecasts for a maximum of 12 months only; Kamal (2013) generates annual out-of-sample forecasts for each of 12 years; Zhou et al. (2013) consider only six out-of-sample months, and He and Fan (2015) produce forecasts over only five quarters. Stekler and Zhang (2013) consider forecasts produced by the IMF, OECD, and private sector forecasters, but their analysis covers only 10–12 time periods. All of the

---

[1] https://research.stlouisfed.org/fred2/series/MKTGDPCNA646NWDB.

[2] https://dfat.gov.au/trade/resources/Documents/chin.pdf.

[3] http://ticdata.treasury.gov/Publish/mfh.txt.

[4] https://ec.europa.eu/eurostat/statistics-explained/index.php/China-EU_-_international_trade_in_goods_statistics\#Both_exports_to_and_imports_from_China_rose_between_2008_and_2018.

above papers provide rankings of forecasting models, but none provide any measures of sampling variability or hypothesis tests. In the light of the very small number of forecasts, it is reasonable to suspect that these rankings would not be robust to changes in the sampling periods considered. In a more recent paper, Higgins et al. (2016) compute monthly forecasts of Chinese GDP and inflation from 2011 to 2015 with forecast horizons ranging from 1 year to 4 years. This provides 60 1-year-ahead forecasts and 24 4-year-ahead forecasts. However, as is the case in the papers cited above, they provide no measures of sampling variability, so it is not possible for the reader to deduce whether the differences in the mean squared forecasting errors reported for different forecasting models are statistically significantly different from zero.

In this paper, we present the results of an out-of-sample forecasting exercise for the Chinese macroeconomy. We forecast two different measures of real economic activity and two different measures of price inflation in each of at least 155 months. Our measures of price inflation are constructed from the Consumer Price Index and the Producer Price Index, and we measure real activity using data on Industrial Production and Electricity Production. As is the case for most of the studies we cite in the second paragraph of this section, we use seasonally adjusted data since the nonseasonal components of variables are generally of most interest in macroeconomic analysis.

We use 19 different forecasting models, each of which has proved itself useful for forecasting macroeconomic variables in Western economies. Our set of models includes simple approaches such as the naive forecast and the mean forecast, approaches based on smoothing, classical time series methods and factor-based forecasts. We use a rolling window of 100 observations for parameter estimation. We calculate the relative mean squared forecasting errors, test null hypotheses of mean squared error equality using the conditional and unconditional tests of Giacomini and White (2006) and control the familywise error rate using the sequentially rejective Bonferroni procedure of Holm (1979).

To the best of our knowledge, this is the first paper on forecasting the Chinese macroeconomy that uses such a wide range of different forecasting models, considers such a long time span of forecasts, and formally tests null hypotheses of equal forecasting power across alternative models instead of simply presenting point estimates of the magnitude of forecast errors.

The remainder of this paper is structured as follows. In Sect. 2, we briefly describe each of the forecasting techniques that we employ and the statistical methodology that we use. In Sect. 3, we describe the data. In Sect. 4, we provide plots of the data and the results of unit root tests. In Sect. 5, we present our empirical results. In Sect. 6, we discuss our findings, and in Sect. 7, we make some concluding comments.

2 Methodology

Let T denote the total number of observations available on the variables of interest. We use the first R observations to estimate model parameters. We generate forecasts for the time period $R + h$ using several competing forecasting models, where we set h equal to 1, 3, 6, 9, and 12 months. We then drop the first observation from the estimation sample, add the $(R+1)^{st}$ observation to the estimation sample, re-estimate parameters,

and generate forecasts for the time period $R + h + 1$. We continue this process—updating the estimation sample, re-estimating the parameters, and computing forecasts for the next time period—until we have computed out-of-sample forecasts for P time periods using each forecasting model with a rolling sample of R observations. Thus, $P = T - R - h + 1$.

In this study, we assess forecast accuracy using the mean squared forecasting error (MSE). Let y_t denote the variable to be forecast. If y_t is integrated of order zero (I(0)), then we fit a forecasting model to the level of the variable and, at time t, generate the h-months-ahead forecast, which we denote $f_{t+h|t}$. If y_t is integrated of order one (I(1)), then we fit a forecasting model to the first difference of the variable and, at time t, for $s = 1, \ldots, h$, generate the s-months-ahead forecast of the first difference of the variable, which we denote $f_{t+s|t}^{\Delta}$. The h-month-ahead forecast of the level of the variable is then $f_{t+h|t} = y_t + \sum_{s=1}^{h} f_{t+s|t}^{\Delta}$.

Our choice of forecasting models was guided by two considerations. Firstly, in this paper our interest is focused on models that are well established and widely used for forecasting macroeconomic variables in Western countries. Thus, we wish to include in our study the types of models that might commonly be used for forecasting by central banks, national treasuries, and by forecasters in industry, rather than models taken from the recent research literature. Similarly, it is not our intention in this paper to propose novel forecasting methods. Secondly, we wish to consider a range of models which includes some very simple techniques, standard smoothing methods, classical time series methods, and some more sophisticated approaches such as Bayesian vector autoregression and large-scale factor models. We settled on the following set of models:

Mean For an I(0) variable, for all forecast horizons $h \in \mathbb{N}$, the h-period-ahead simple mean forecast computed in time period R is the mean of all the available observations up to and including period R, i.e. $f_{R+h|R} = \frac{1}{R} \sum_{t=1}^{R} y_{R-t+1}$. For an I(1) variable, the corresponding forecast computed in time period R is $f_{R+h|R} = y_R + \frac{h}{R-1} \sum_{t=1}^{R-1} \Delta y_{R-t+1}$. If y_t is a serially uncorrelated process or a random walk with drift, then the appropriate mean forecast is a mean-square optimal estimator of the mean-square optimal forecast.

Naive For all forecast horizons $h \in \mathbb{N}$, the h-period-ahead naive forecast of an I(0) variable is equal to the last available observation in the data set, i.e. $f_{R+h|R} = y_R$. If the variable is I(1), then we compute the naive forecast as $f_{R+h|R} = y_R + h\Delta y_R$.

Moving average (MA) For an I(0) variable, for all forecast horizons $h \in \mathbb{N}$, the h-period-ahead moving average forecast of order m computed in time period R is the mean of the last m observations in the data set, i.e $f_{R+h|R} = \frac{1}{m} \sum_{t=1}^{m} y_{R-t+1}$. For an I(1) variable, the corresponding forecast is $f_{R+h|R} = y_R + \frac{h}{m} \sum_{t=1}^{m} \Delta y_{R-t+1}$. We consider two moving average models. For the first (MA(4)), we (arbitrarily) set the order of the moving average to 4. For the second (MA-opt), we choose the order of the moving average in each time period to minimize the mean squared error of the MA forecast up until that point in time. Thus, the order of the moving average potentially changes for every new forecast in the sequence of forecasts.

Simple Exponential smoothing (SES) For an I(0) variable y_t, the simple exponential smoothing model produces the smoothed series $s_t = \alpha y_{t-1} + (1 - \alpha)s_{t-1}$,

where α is a smoothing parameter for which $0 \leqslant \alpha \leqslant 1$. For all forecast horizons $h \in \mathbb{N}$, the h-period-ahead simple exponential smoothing forecast is defined as $f_{R+h|R} = s_R$. For an I(1) variable, the corresponding forecast is $f_{R+h|R} = y_R + h\Delta s_R$. We consider two implementations of this model. For the first (SES(0.5)), the smoothing parameter α is (arbitrarily) set equal to 0.5. For the second (SES-opt), for each forecast, α is set equal to the value that minimizes the mean squared historical one-step-ahead forecasting error up until the time that the forecast is made. We compute the SES forecasts using the ses command in the forecast[5] package in the R programming language.[6]

Direct (dAR) and multistep iterated (AR) autoregression We compute both multistep iterated and direct autoregressive forecasts.[7] For an I(0) variable and an autoregression of order p, $y_t = \beta_0 + \sum_{j=1}^{p} \beta_j y_{t-j} + \varepsilon_t$, the h-step-ahead multistep iterated forecast computed in time period R is $f_{R+h|R} = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j \hat{y}_{R+h-j}$ where $\hat{y}_t = \begin{cases} y_t \text{ if } t \leqslant R \\ f_{t|R} \text{ if } t > R \end{cases}$ and $\hat{\beta}_j$ is the ordinary least squares (OLS) estimator of β_j computed using observations $y_1, \ldots, y_R$. Similarly, for an I(1) variable and an autoregression of order p in first differences, $\Delta y_t = \beta_0 + \sum_{j=1}^{p} \beta_j \Delta y_{t-j} + \varepsilon_t$, the h-step-ahead multistep iterated forecast computed in time period R is $f_{R+h|R} = y_R + h\hat{\beta}_0 + \sum_{k=1}^{h} \sum_{j=1}^{p} \hat{\beta}_j \Delta \hat{y}_{R+k-j}$.

For an I(0) variable, a direct h-step-ahead forecast at time period R is calculated by estimating by OLS the equation $y_{t+h} = \beta_0 + \sum_{j=1}^{p} \beta_j y_{t-j} + \varepsilon_t$ using observations $y_1, \ldots, y_{R-1}$ and computing $f_{R+h|R} = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j y_{R+1-j}$. For an I(1) variable, the direct forecast is computed by estimating the equation $y_{t+h} - y_t = \beta_0 + \sum_{j=1}^{p} \beta_j \Delta y_{t-j} + \varepsilon_t$ using observations $y_1, \ldots, y_{R-h}$ and computing $f_{R+h|R} = y_R + \beta_0 + \sum_{j=1}^{p} \hat{\beta}_j \Delta y_{R+1-j}$.

We consider two implementations of the AR and ARd models. For the first (AR(1) and ARd(1)), the order of the model is (arbitrarily) set equal to 1. For the second (AR(p) and ARd(p)), for each forecast generated, the order is chosen using Akaike Information Criterion (AIC) with a maximum allowed order of 24.

Autoregressive moving average (ARMA) The ARMA(p,q) model for an I(0) variable, $y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$, may be written in state-space form and estimated using the maximum likelihood method (see Gardner et al. (1980)). Forecasts may then be computed using the Kalman filter (see Section 5.5 of Box et al. (2015)). For an I(1) variable, the model is instead $\Delta y_t = \mu + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$. We consider two implementations of the ARMA model. For the first (ARMA(1,1)), we (arbitrarily) set both orders of the model to 1. For the second (ARMA(p,q)), for each forecast generated, the order is chosen using the Akaike Information Criterion (AIC) with maximum allowed orders of 2 and 2. We use the Arima and auto.arima functions in the forecast package of the R programming language to compute the ARMA forecasts.

---

[5] Hyndman et al. (2018), Hyndman and Khandakar (2008).

[6] R Core Team (2018).

[7] See Marcellino et al. (2006).

Vector autoregression (VAR)   We follow Stock and Watson (2002b) by specifying a vector to contain a price inflation variable, a variable measuring real activity, and an interest rate variable. To forecast a price inflation variable, the vector Y_t consists of the variable to be forecast, Industrial Production, and the interest rate. To forecast a real activity variable, the vector consists of the variable being forecast, the CPI inflation rate, and the interest rate. The VAR model for an I(0) variable y_t is then $Y_t = A_0 + \sum_{j=1}^{p} A_j Y_{t-j} + \eta_t$, where y_t is the first element of the vector Y_t. The parameters are estimated using the OLS method, and the h-step-head forecast of the vector Y_t computed in time period R is $f_{R+h|R} = \hat{A}_0 + \sum_{j=1}^{p} \hat{A}_j \hat{Y}_{R+h-j}$ where $\hat{Y}_t = \begin{cases} Y_t & \text{if } t \leqslant R \\ f_{t|R} & \text{if } t > R \end{cases}$. For an I(1) variable, the VAR is specified in first difference form, $\Delta Y_t = A_0 + \sum_{j=1}^{p} A_j \Delta Y_{t-j} + \eta_t$, and the h-step-head forecast of the vector Y_t computed in time period R is $f_{R+h|R} = Y_R + h \hat{A}_0 + \sum_{k=1}^{h} \sum_{j=1}^{p} \hat{A}_j \Delta \hat{Y}_{R+k-j}$. In each case, the forecast of y_{R+h} may be recovered from the forecast of the vector Y_{R+h}.

   We consider two implementations of the VAR model. For the first (VAR(1)), the order of the model is (arbitrarily) set equal to 1. For the second (VAR(p)), for each forecast generated, the order is chosen using the Akaike Information Criterion (AIC) with a maximum allowed order of 4. We compute all the VAR estimates and forecasts using the vars package[8] in the R programming language.

Bayesian vector autoregression (BVAR)   We also estimate a Bayesian vector autoregression using the Minnesota prior of Doan et al. (1984) and Litterman (1986). We refer the reader to Canova (2007) Chapter 10 for technical details of the method.[9] The construction of the vectors and forecasts is as for the frequentist VAR described above. We set the lag order to 4 for all BVAR models. We compute the BVAR forecasts using the BMR package[10] in the R programming language.

Factor models   Let x_t be a $N \times 1$ vector of macroeconomic variables that have been suitably transformed to become I(0). Let s_t be a $k \times 1$ vector of unobservable factors such that $x_t = Bs_t + \eta_t$, where η_t is a $N \times 1$ vector of unobservable errors, B is a $N \times k$ matrix of unknown coefficients, and $k << N$. An estimator of the factor vector $\hat{s}_t$ may be constructed as the principal components of the sample covariance matrix of x_t corresponding to the largest k eigenvalues. We standardize all the elements of x_t prior to computing the principal components. For an I(0) variable y_t in time period R, the h-steps-ahead forecast is computed as $f_{R+h|R} = \hat{\beta}' \hat{s}_R$ where $\hat{\beta}$ is the least squares estimator of β for the equation $y_{t+h} = \beta' \hat{s}_t + \varepsilon_t$ computed using observations from $t = 1$ to $t = R - h$. For an I(1) variable, $\hat{\beta}$ is the least squares estimator of β for the equation $y_{t+h} - y_t = \beta' \hat{s}_t + \varepsilon_t$, but the rest of the procedure is identical to the I(0) case. Stock and Watson (2002a) establish conditions under which $\hat{\beta} - \beta \xrightarrow{p} 0$ and $\hat{s}_t - s_t \xrightarrow{p} 0$ as N and R jointly grow. We consider two implementations of this forecasting model. In the first (F(2)), the

---

[8] Pfaff (2008).

[9] Using the notation of Canova (2007) Section 10.2.2, we set $\phi_0 = \phi_1 = 0.5$, $\phi_2 = 100$, $\phi_3 = 2$ and specify $h(\ell)$ to be geometric decay.

[10] O'Hara (2018).

number of factors is set (arbitrarily) to $k = 2$. In the second (F(k)), for each time period in the forecasting exercise, the number of factors is estimated by minimizing the $PC_{p1}(k)$ criterion[11] of Bai and Ng (2002) with the maximum allowed number of factors set to 10.

Factor-VAR   We also consider VAR models in which the vector consists of the variable being forecast and two factors, estimated in the way described above. Note that this is the special case of the FAVAR model of Bernanke et al. (2005). We consider two implementations of this model. In the first (F(2)VAR(1)), the VAR order is (arbitrarily) set to 1. In the second (F(2)VAR(p)), the VAR order is chosen to minimize the AIC with a maximum allowable lag of 4.

Our estimator of the mean squared error of the ith forecast is $\frac{1}{P} \sum_{t=R}^{T-h} (f_{t+h|t}^i - y_{t+h})^2$. For all but one forecasting model, we also test the null hypotheses

$$H_0^u : E\left((f_{t+h|t}^i - y_{t+h})^2\right) - E\left((f_{t+h|t}^b - y_{t+h})^2\right) = 0$$
$$H_0^c : E\left((f_{t+h|t}^i - y_{t+h})^2|\mathscr{F}_t\right) - E\left((f_{t+h|t}^b - y_{t+h})^2\right)|\mathscr{F}_t) = 0$$

where forecasting model b is a benchmark model that we choose. H_0^u is the hypothesis that model i has the same mean squared forecasting error as the benchmark model. Note that this is a hypothesis about the model with estimated parameters, so sampling uncertainty contributes to the forecasting error. It is possible that a correctly specified forecasting model could have a larger mean squared forecasting error than an incorrectly specified benchmark model if the variance of the sampling error is sufficiently large—as could be the case if the sample is quite small. H_0^c is the hypothesis of equality of the mean squared error conditional on $\mathscr{F}_t$, an information set available at time t. Giacomini and White (2006) provide the method and theory for testing both hypotheses, and we refer the reader to their paper for more details of the two tests. We follow Giacomini and White (2006) closely by using the Newey–West estimator of the variance with a bandwidth of $h - 1$ to construct the test statistics, and for H_0^c, we condition on the lagged value of the difference of the squared forecast errors from the pair of models under consideration. Since our study involves individual tests for hypotheses for many forecasting models, several forecasting horizons, and multiple variables, it is likely to yield some small p values purely by chance, even if all the null hypotheses are true. In order to control the risk of false discovery, in addition to reporting the p values for each hypothesis test, we also report the results of the sequentially rejective Bonferroni procedure proposed by Holm (1979). For a set of m null hypotheses, this consists of rejecting the null hypothesis with the smallest p value if its p value is smaller than α/m. If this hypothesis is rejected, then the hypothesis corresponding to the second smallest p value is rejected if its p value is smaller than $\alpha/(m - 1)$. The process continues until no more hypotheses are rejected, with the p value for the jth null hypotheses being compared to $\alpha/(m - j + 1)$. This procedure

---

[11] Bai and Ng (2002) propose several alternative criteria for estimating the number of factors and find many that are both consistent and perform well in finite sample simulations. Our choice from these of $PC_{p1}(k)$ is arbitrary. We have also used $PC_{p2}(k)$, $IC_{p1}(k)$, and $IC_{p2}(k)$ and found that the choice of criterion makes very little difference to our results.

provides strong control of the familywise error rate—that is to say, it controls the probability of rejecting at least one true null hypothesis from a collection of multiple hypotheses, for any combination of true and false hypotheses.

3 Data

Empirical research in economics is often constrained by data availability. This is particularly the case for research on the Chinese macroeconomy. Many thousands of economic time series are available from China's National Bureau of Statistics (NBS),[12] the People's Bank of China,[13] and other government sources, but only a small proportion are of sufficient length to be suitable for a study such as this. Moreover, many series have missing observations and/or implausible outliers. A common concern with Chinese data is the integrity of the data collection process. China is a large and complex country which raises technical challenges for its statistical agencies. Furthermore, questions are sometimes raised about the extent to which political pressure might be such that the published data reflect production targets rather than actual outcomes. We do not explore these issues here, but instead refer the interested reader to Crabbe (2016), Holz (2014), Orlik (2011), Holz (2003), and Holz (2008).

The four variables that we forecast are the Consumer Price Index (CPI) month-on-month inflation rate, the Producer Price Index (PPI) year-on-year inflation rate, the year-on-year growth rate of industrial production (IP), and the month-on-month growth rate of the production of electricity (EP). All our variables are monthly final revisions and are measured from October 1996 to December 2018. We do not consider Gross Domestic Product (GDP) in this study since the data are available only at the quarterly and annual frequencies, which would greatly reduce the number of observations in our sample.[14] The CPI data are supplied by the NBS as month-on-month percentage changes. The PPI and IP data are supplied as year-on-year percentage changes, and the EP data are supplied in billions of kilowatt/hours which we convert to a month-on-month growth rate. We downloaded these data sets from Thomson Reuters Datastream[15] since this was easier to work with than the NBS Website. The IP and EP data sets have a number of missing values and outliers. We cleaned the data sets by removing observations for which the first difference deviates from the median

---

[12] http://www.stats.gov.cn/english/.

[13] http://www.pbc.gov.cn/en/3688006/index.html.

[14] Two approaches to dealing with this problem are to interpolate the quarterly GDP data to create a monthly series (see, e.g. Higgins et al. (2016)) or to use mixed-frequency modelling techniques (see, e.g. Ghysels (2016)). We are sceptical of the use of interpolation to create monthly series from quarterly series in forecasting studies since it results in two-thirds of the forecasted data points being the outcomes of the interpolation procedure, rather than actual data points. While the presence of outliers and missing values dictates that a certain amount of interpolation is practically unavoidable when analysing Chinese macroeconomic data, our preference is that this is kept to the minimum possible. Mixed-frequency modelling techniques are interesting, but are beyond the scope of the present paper, which focuses on forecasting models that are well established in Western economies.

[15] The Datastream codes for our variables IP, EP, CPI and PPI are CHIPTOT.H, CHPBRENTP, CHCPINATR, and CHPXGPRDF, respectively.

of the first differences by more than five times the interquartile range.[16] All missing values were then replaced by cubic interpolation.[17] All four variables were seasonally adjusted using the X-13 method with a deterministic adjustment made for the Chinese New Year holiday.[18]

Our VAR models also include the Prime Lending Rate published by the People's Bank of China.[19] For the factor models, we constructed factor estimates from 50 different monthly macroeconomic variables sourced from Datastream. Outliers and missing values were dealt with as described above, and with the exception of financial variables, all variables were seasonally adjusted. All 50 variables were either differenced or log-differenced prior to the application of the principal components estimator. "Appendix 1" includes tables which list the variable names, their Datastream codes, whether they have missing values, and a code which indicates whether they were differenced or log-differenced. All our code and data are available online.[20]

Readers familiar with the literature will notice that our cross-sectional dimension of 50 variables is much smaller than that of, as examples, Stock and Watson (2002b) (215 variables), Schumacher (2007) (124 variables) or Artis et al. (2005) (81 variables). This is an important consideration since the asymptotic justification for the principal components estimator requires the number of variables to grow simultaneously with the number of observations. However, Boivin and Ng (2006) argue that "...as few as 40 pre-screened series often yield satisfactory or even better results than using all 147 series..." which may provide us with some hope that our 50 variables may suffice.[21] Nonetheless, we must concede that our variables are 'pre-screened' only in the sense that we used all the variables for which we had data available at the monthly frequency over the entire time period of interest.

4 Preliminary analysis

Plots of the four variables that we will forecast appear in Fig. 1. Note that these plots are of the data *after* outliers have been removed, missing values interpolated, and seasonal adjustment applied.

---

[16] We have also conducted our forecasting exercise using data without any outliers removed. The results do not change the main conclusions of the paper, so we do not report them here.

[17] The cubic interpolation was done using the na.spline function in the zoo package in R. See Zeileis and Grothendieck (2005).

[18] The seasonal adjustment was carried out using the seasonal package in R (Sax and Eddelbuettel (2018)), which is an interface for the X-13 software written by the US Bureau of Census (https://www.census.gov/srd/www/x13as/).

[19] We downloaded these data from Datastream. The variable code is CHBANKR.

[20] https://github.com/cheaton/China-Forecasting.

[21] Precedents include Pincheira and Gatty (2016) who construct two sets of factors using inflation of 18 Latin American countries and inflation of 30 OECD countries, respectively.
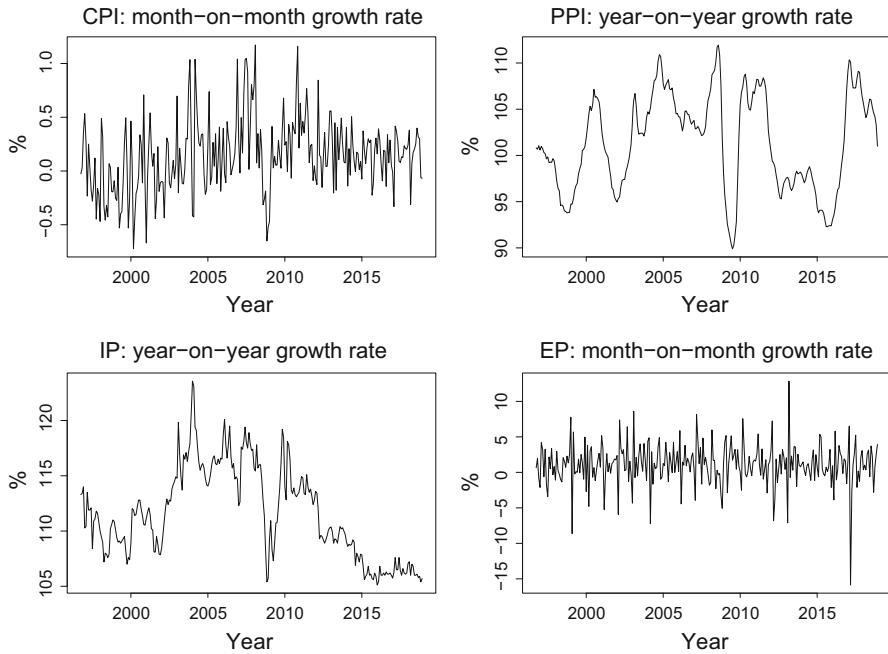
Fig. 1 Forecasted variables

Table 1 Augmented Dickey–Fuller test statistics

	CPI	PPI	IP	EP
Lag	6.000	6.000	6.000	6.000
Statistic	−3.868	−3.943	−2.566	−6.433
p value	0.016	0.012	0.338	0.010

The results of Augmented Dickey–Fuller (ADF) tests for unit roots[22] are presented for each variable in Table 1. In each case, the maximum lag was set equal to the largest integer smaller than $4(T/100)^{\frac{2}{9}}$. If a 5% significance level is used, then evidence against a unit root exists for the CPI, PPI, and EP growth rates, but not for the growth of IP.[23]

We also estimated each of the frequentist models with fixed model orders using all the available observations for both the levels and first differences of the variables. We note that, when judged by measures such as the R^2 and the F-statistic for overall significance, most of the models appear to fit the variables reasonably well. In the

---

[22] The ADF test statistics were computed using the adf.test command in the tseries package in R. See Trapletti and Hornik (2018).

[23] Varying the lag order from 2 to 12 did not change the conclusions of the ADF test. Similar results were also found using the Phillips-Perron unit root test.

interests of saving space, these results are not presented here, but are available in an online supplement.[24]

5 The forecasting results

The first step of the forecasting exercise is to take the first $R = 100$ observations from the raw[25] data set—that is, observations from October 1996 to January 2005—remove outliers, interpolate missing values, seasonally adjust, compute principal components, estimate the model orders and all the model parameters using only the data in this subsample, and then generate forecasts for the time periods h months after the end of the subsample, where $h = 1, 3, 6, 9, 12$. After generating forecasts from all models at each horizon, we drop the first observation from the subsample and add an extra observation of raw data to the end of the subsample of raw data. We repeat the same steps for the second iteration of the forecasting exercise, with the subsample used for estimation now running from November 1996 to February 2005. We continue in this fashion, dropping the first observation, appending an extra observation of raw data to the end of the subsample of raw data, adjusting for outliers, missing values[26] and seasonality, estimating principal components, model orders, and all the model parameters using only the data in the subsample, and computing forecasts, until we have used the entire data set. This produces a sequence of 167 1-month-ahead forecasts from February 2005 to December 2018. The sequences of h-month-ahead forecasts for $h = 3, 6, 9, 12$ start $h - 1$ months later than the corresponding 1-month-ahead forecast. These forecasts are compared to the data sets generated by removing outliers, interpolating missing values, and applying seasonal adjustment to the entire span of data available for each variable[27] (i.e. from October 1996 to December 2018). We measure the deviations of the forecasts from these variables using the mean squared error, and we test null hypotheses of the equality of the mean squared error of each forecast to that of a suitable benchmark. For each variable forecast, we use as a benchmark model whichever of the naive forecast and the simple mean forecast generates the smallest estimated mean squared error.

Estimates of the mean squared forecasting errors for inflation produced by the models are presented in Table 2. The mean squared errors have been standardized so that the mean forecast has a standardized mean squared error of 1 for the CPI

---

[24] https://cheaton.github.io/China-Forecasting/online_supplement.pdf.

[25] By 'raw,' we mean the data that have not had outliers removed, missing values interpolated, or seasonal adjustment applied.

[26] For cases in which a missing value occurs at the beginning or end of the subsample, cubic extrapolation is used to replace the missing value instead of cubic interpolation.

[27] The more conventional approach, followed by (e.g.) Ang et al. (2007), Artis et al. (2005), Bandt et al. (2007), Ciccarelli and Mojon (2010), Faust and Wright (2013), Marcellino et al. (2003), Marcellino et al. (2006), den Reijer (2005), Schneider and Spitzer (2004), Schumacher (2007), Stock and Watson (1999), Stock and Watson (2002b), and Stock and Watson (2008), is to make a single adjustment for outliers and seasonality to the entire data set, or to source data that are already seasonally adjusted, and then use subsets of these data in the forecasting exercise. As pointed out by an anonymous referee, this approach pollutes the forecasts with information from the future of each subsample. We have also used this approach and found that it produced results that are broadly similar to those we report in this paper.

forecasts and the naive forecast has a standardized mean squared error of 1 for the PPI forecasts since, in each case, these are the benchmarks with the smallest mean squared error. A number of features of Table 2 are worth noting. For the CPI inflation rate, the factor-VAR(1) model with two factors produced the best 1-month-ahead forecasts within-sample and, in general, the more sophisticated forecasting models outperform the simple models. For all other forecast horizons, the more sophisticated techniques either perform worse than the simple mean forecast or provide only a slight improvement. Thus, judged purely by point forecasts, there appears to be little to gain from using sophisticated methods to forecast Chinese CPI inflation at horizons longer than one month.

The results for the PPI inflation rate are markedly different. In this case, the BVAR model generates the smallest sample mean squared forecasting error for horizons of 1, 3 and 6 months, and the improvement over the benchmark naive forecast is substantial. The ARMA, AR(p), and ARd(p) models also perform relatively well at short horizons, but the relative performance declines substantially as the forecast horizon grows. In contrast, the factor models perform extremely poorly relative to the naive forecast at short horizons, but perform the best at horizons of 9 and 12 months. The MA and SES forecasts are generally outperformed by the naive forecast, and the simple mean forecast performs poorly at horizons of 1, 3, and 6 months, but performs quite well at horizons of 9 and 12 months.

The above comments are based purely on the sample estimates of the mean squared forecast errors and take no account of sampling variability. Tables 3 and 4 provide the p values for the conditional and unconditional Giacomini and White (2006) tests, using the mean forecast as the benchmark for the CPI and the naive forecast as the benchmark for the PPI. For cases in which the p value is less than 0.05, if the corresponding relative mean squared error is greater than 1 (i.e. the forecasting model is inferior to the benchmark), the p value is presented in italics. If the p value is less than 0.05 and the relative MSE is less than 1 (i.e. the forecasting model is superior to the benchmark), then the p value is displayed with a bold font. We have also applied the sequentially rejective Bonferroni adjustment of Holm (1979) to each table of p values to control the probability that at least one true null hypothesis in each table is rejected to be less than 0.05. The p values of any such rejected hypotheses are underlined. For example, in Table 3(a) in the row marked 'Factor(2)' and the column headed '$h = 6$' appears the figure 0.022. This indicates that the p value for the null hypothesis that the conditional expected value of the mean squared error of the factor model with two factors, when used to forecast 6 months ahead, is equal to that of the mean forecast, is 0.022 when tested against a two-sided alternative hypothesis. This number is printed in a bold font to indicate that, if a Neyman–Pearson hypothesis test were conducted with a 5% significance level, the null hypothesis would be rejected and that the estimated mean squared error of the Factor(2) forecast is less than that of the mean forecast (i.e. the Factor(2) model is superior). Note, however, that the null would not be rejected if a 1% significance level was used. Furthermore, the fact that the number 0.022 is not underlined indicates that, when the sequentially rejective Bonferroni adjustment of Holm (1979) is applied to restrict the probability that at least one true null hypothesis in Table 3(a) is rejected to be less than 5%, the null hypothesis for the Factor(2) model at a horizon of 6 months is not rejected. For this reason, we do not consider the fact

Table 2 Inflation: standardized mean squared forecast errors by horizon

	(a) CPI					(b) PPI				
	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Mean	1.000	1.000	1.000	1.000	1.000	23.159	3.453	1.290	0.821	0.641
Naive	1.231	1.521	1.490	1.689	1.958	1.000	1.000	1.000	1.000	1.000
MA	0.997	1.031	1.098	1.401	1.474	4.553	1.729	1.257	1.131	1.027
MA-opt	0.934	1.054	1.172	1.269	1.324	13.318	2.865	1.494	1.164	1.005
SES	0.966	1.113	1.156	1.395	1.531	2.761	1.385	1.132	1.059	0.999
SES-opt	0.919	1.043	1.132	1.231	1.355	1.000	1.000	1.000	1.000	1.000
AR(1)	0.885	0.975	0.997	0.993	1.000	1.026	1.027	1.031	1.036	1.016
AR(p)	0.925	1.050	1.073	1.117	1.070	0.393	0.684	0.843	0.832	0.729
ARd(1)	0.885	0.990	0.973	1.007	0.966	1.026	1.043	0.944	0.827	0.748
ARd(p)	0.904	1.078	1.119	1.136	1.117	0.386	0.692	0.918	1.011	1.027
ARMA(1,1)	0.913	1.003	0.992	0.992	0.998	0.508	0.814	0.866	0.828	0.781
ARMA(p,q)	0.956	1.061	1.084	1.102	1.053	0.378	0.654	0.818	0.782	0.706
VAR(1)	0.839	1.100	1.222	1.376	1.404	1.081	1.065	1.033	1.014	0.933
VAR(p)	0.830	1.101	1.259	1.443	1.489	0.404	0.687	0.837	0.823	0.824
BVAR	0.877	1.184	1.342	1.660	1.901	0.377	0.605	0.720	0.725	0.749
Factor(2)	0.892	0.986	0.978	1.040	0.988	22.927	3.129	1.044	0.646	0.535
Factor(k)	1.022	1.101	1.160	1.107	1.168	25.366	3.564	1.166	0.710	0.560
F(2)VAR(1)	0.822	1.094	1.218	1.408	1.472	0.843	0.888	0.901	0.885	0.832
F(2)VAR(p)	0.865	1.191	1.360	1.537	1.548	0.504	0.800	0.894	0.778	0.757

The rows are the forecasting models described in Sect. 2. The columns are forecast horizons. The table elements are the point estimates of the mean squared forecasting error expressed as a ratio of the mean squared forecasting error of the benchmark model for the same forecast horizon. Thus, a number less than 1 indicates that the model produced a smaller sample mean squared forecasting error than the benchmark model at that forecast horizon. The benchmark models are the simple mean forecast for the CPI and the naive forecast for the PPI

Table 3 Inflation: *p* values for conditional test by horizon

(a) CPI

	h = 1	h = 3	h = 6	h = 9	h = 12
Naive	0.218	*0.018*	0.176	0.055	*0.035*
MA	0.795	0.962	0.773	0.273	0.284
MA-opt	0.393	0.28	0.581	0.508	0.316
SES	0.971	0.641	0.814	0.297	0.221
SES-opt	0.633	0.498	0.734	0.519	0.242
AR(1)	0.326	0.211	0.387	0.063	0.241
AR(p)	0.603	0.792	0.725	0.548	0.24
ARd(1)	0.326	0.825	0.305	0.871	0.425
ARd(p)	0.621	0.8	0.781	0.521	0.18
ARMA(1,1)	0.471	0.706	0.134	0.204	0.337
ARMA(p,q)	0.922	0.494	0.681	0.474	0.534
VAR(1)	0.371	0.65	0.555	0.3	0.358
VAR(p)	0.37	0.583	0.431	0.357	0.406
BVAR	0.646	0.309	0.398	0.247	0.332
Factor(2)	0.247	0.194	**0.022**	0.801	0.936
Factor(k)	0.3	0.498	*0.033*	0.121	*0*
F(2)VAR(1)	0.311	0.742	0.569	0.284	0.4
F(2)VAR(p)	0.613	0.424	0.347	0.211	0.38

(b) PPI

	h = 1	h = 3	h = 6	h = 9	h = 12
Mean	0	*0*	0.102	0.809	0.433
MA	0	*0.013*	0.08	0.092	0.478
MA-opt	0	*0.004*	0.081	0.581	0.888
SES	0	*0.015*	0.085	0.215	0.554
SES-opt	0	*0.017*	0.096	0.103	0.368
AR(1)	0.11	0.245	0.297	0.276	0.752
AR(p)	**0.001**	0.16	0.526	0.738	0.523
ARd(1)	0.11	0.747	0.259	0.8	0.656
ARd(p)	**0.001**	0.231	0.676	0.325	0.25
ARMA(1,1)	**0**	**0.045**	0.15	0.1	0.089
ARMA(p,q)	**0**	0.117	0.637	0.634	0.501
VAR(1)	0	0.301	0.953	0.908	0.438
VAR(p)	**0.001**	0.171	0.397	0.751	0.648
BVAR	**0.001**	0.111	0.245	0.401	0.47
Factor(2)	0	0	0.194	0.473	0.226
Factor(k)	0	0	0.105	0.61	0.322
F(2)VAR(1)	**0.049**	0.223	0.732	0.814	0.551
F(2)VAR(p)	**0.006**	0.44	0.264	0.673	0.623

The rows are the forecasting models described in Sect. 2. The columns are forecast horizons. The table elements are the *p* values for the Giacomini and White (2006) conditional test that the mean squared forecasting error is equal to that of the benchmark model. If a *p* value is less than 0.05, and the corresponding within-sample relative mean squared error in Table 2 is less than 1, then the *p* value is in a bold font, indicating that the model generates significantly better forecasts at the 5% significance level. If the *p* value is less than 0.05, and the corresponding within-sample relative mean squared error in Table 2 is greater than 1, then the *p* value is in italics, indicating that the model generates significantly worse forecasts. *p* values for hypotheses that are rejected by the sequentially rejective Bonferroni procedure of Holm (1979), applied to restrict the familywise error rate of each table to be less than 5%, are underlined. The benchmark model for the CPI forecasts is the naive forecast. The benchmark for the PPI forecasts is the mean forecast

that the individual p value for this hypothesis is less than 0.05 to constitute strong evidence in favour of this particular forecasting model.

Overall, Tables 3(a) and 4(a) show no strong evidence that the forecasting models considered are able to outperform the simple mean forecast when applied to the CPI inflation rate for China. Furthermore, when restricting the familywise error rate to be less than 5%, we find evidence only that the k-factor model with a horizon of 12 months has a worse performance than the simple mean forecast.

For the PPI with a one-month forecast horizon, when we restrict the familywise error rate to be less than 5%, the conditional test finds evidence that ARMA models provide superior forecasts to the naive model. The unconditional test also finds evidence in favour of the direct AR(p) model, the VAR(p) model, and the BVAR model. Some individual p values are also less than 0.05 for longer forecasting horizons but in each case the p values are reasonably close to 0.05 and the hypotheses are not rejected when controlling the familywise error rate. For these reasons, we do not consider these statistics to provide evidence of superior forecasting power at longer horizons.

We now consider the results for the forecasts of real activity. The relative mean squared forecasting errors are presented in Table 5. For IP at a forecasting horizon of 1 month, all the models considered are inferior to the simple mean forecast. For a forecasting horizon of 3 months, only the AR(p) and ARd(p) models provide smaller mean squared forecasting errors. At longer horizons, the relative performance of the AR, VAR, BVAR, and factor models improves. Interestingly, however, ARMA models are inferior at all forecasting horizons. Similarly, the naive, MA, and SES models always perform poorly.

For EP, the smallest relative mean squared error at any horizon is 0.974 (the Factor(2) model with a horizon of 1 month). Thus, none of the models was able to produce an economically meaningful improvement over the forecasting power of the simple mean forecast.

The p values for the conditional Giacomini and White (2006) test appear in Table 6, and for the unconditional test, in Table 7. The lack of evidence that any of the forecasting models considered are able to consistently produce mean squared forecasting errors that are less than those of the simple mean forecast is striking.

6 Discussion

The forecasting models that we consider in this paper are all well established, and it may be somewhat surprising that we find so little evidence of their utility in the Chinese context. However, from a theoretical perspective, the mean squared optimal forecast of a stochastic process y_t relative to its past values is $E(y_{R+h}|y_R, y_{R-1}, y_{R-2}, \ldots)$ (see, e.g. Hamilton (1994), Section 4.1). Furthermore, for any process y_t that has a stable ARMA representation, $E(y_{R+h}|y_R, y_{R-1}, y_{R-2}, \ldots) - E(y_{R+h}) \to 0$ as $h \to \infty$. Consequently, we should expect our models to be less likely to produce forecasts superior to the mean forecast as the forecast horizon grows. As a simple example, consider the AR(1) model $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$. The mean squared optimal forecast is $f_{R+h} = \beta_0 \sum_{j=0}^{h-1} \beta_1^j + \beta_1^h y_R$. Plots of the values of the OLS estimator of β_1 for each

Table 4 Inflation: p values for unconditional test by horizon

(a) CPI

	h = 1	h = 3	h = 6	h = 9	h = 12
Naive	0.226	*0.005*	0.076	*0.036*	*0.011*
MA	0.979	0.813	0.673	0.173	0.133
MA-opt	0.488	0.696	0.435	0.28	0.206
SES	0.806	0.407	0.53	0.179	0.102
SES-opt	0.407	0.722	0.522	0.325	0.147
AR(1)	0.131	0.064	0.658	0.161	0.922
AR(p)	0.565	0.712	0.676	0.28	0.127
ARd(1)	0.131	0.641	0.745	0.834	0.242
ARd(p)	0.469	0.62	0.525	0.298	0.314
ARMA(1,1)	0.254	0.742	0.051	0.071	0.252
ARMA(p,q)	0.648	0.242	0.449	0.311	0.366
VAR(1)	0.165	0.39	0.315	0.234	0.267
VAR(p)	0.151	0.387	0.237	0.171	0.224
BVAR	0.305	0.196	0.21	0.147	0.228
Factor(2)	0.102	0.78	0.391	0.392	0.679
Factor(k)	0.827	0.342	*0.005*	0.129	*0.001*
F(2)VAR(1)	0.142	0.419	0.332	0.22	0.259
F(2)VAR(p)	0.287	0.197	0.155	0.121	0.209

(b) PPI

	h = 1	h = 3	h = 6	h = 9	h = 12
Mean	0	0	0.377	0.527	0.195
MA	0	*0.004*	*0.029*	*0.025*	0.594
MA-opt	0	*0.001*	0.065	0.352	0.978
SES	0	*0.004*	*0.042*	0.075	0.978
SES-opt	0	*0.005*	*0.041*	*0.031*	0.435
AR(1)	0.105	0.216	0.224	0.299	0.641
AR(p)	**0**	0.088	0.404	0.413	0.244
ARd(1)	0.105	0.504	0.735	0.503	0.395
ARd(p)	**0**	0.176	0.747	0.969	0.942
ARMA(1,1)	**0**	**0.016**	**0.047**	**0.026**	**0.024**
ARMA(p,q)	**0**	0.051	0.35	0.322	0.226
VAR(1)	0.361	0.613	0.814	0.916	0.635
VAR(p)	**0**	0.077	0.423	0.442	0.456
BVAR	**0**	**0.048**	0.199	0.19	0.224
Factor(2)	0	0	0.894	0.229	0.09
Factor(k)	0	0	0.624	0.355	0.141
F(2)VAR(1)	0.212	0.464	0.571	0.524	0.388
F(2)VAR(p)	**0.002**	0.263	0.616	0.353	0.348

The rows are the forecasting models described in Sect. 2. The columns are forecast horizons. The table elements are the p values for the Giacomini and White (2006) unconditional test that the mean squared forecasting error is equal to that of the benchmark model. If a p value is less than 0.05, and the corresponding within-sample relative mean squared error in Table 2 is less than 1, then the p value is in a bold font, indicating that the model generates significantly better forecasts at the 5% significance level. If the p value is less than 0.05, and the corresponding within-sample relative mean squared error in Table 2 is greater than 1, then the p value is in italics, indicating that the model generates significantly worse forecasts. p values for hypotheses that are rejected by the sequentially rejective Bonferroni procedure of Holm (1979), applied to restrict the familywise error rate of each table to be less than 5%, are underlined. The benchmark model for the CPI forecasts is the mean forecast. The benchmark for the PPI forecasts is the naive forecast

Table 5 Real activity: standardized mean squared forecast errors by horizon

	(a) IP					(b) EP				
	h = 1	h = 3	h = 6	h = 9	h = 12	h = 1	h = 3	h = 6	h = 9	h = 12
Mean	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Naive	3.387	9.488	18.002	27.118	35.691	2.028	2.092	2.163	2.103	2.221
MA	1.347	1.936	3.009	3.968	5.139	1.334	1.247	1.308	1.251	1.237
MA-opt	1.117	1.397	1.711	2.144	2.559	1.105	1.107	1.107	1.078	1.090
SES	1.667	3.259	5.224	7.486	9.813	1.395	1.354	1.393	1.339	1.430
SES-opt	1.088	1.478	1.836	2.189	2.525	1.000	1.000	1.000	1.000	1.000
AR(1)	1.053	1.046	0.964	0.941	0.922	1.054	1.001	1.001	0.998	1.001
AR(p)	1.094	0.961	0.913	0.863	0.850	1.147	1.079	1.044	1.015	0.988
ARd(1)	1.063	1.018	0.958	0.889	0.861	1.072	1.020	1.009	1.010	1.072
ARd(p)	1.107	0.930	0.842	0.722	0.658	1.236	1.157	1.154	1.163	1.175
ARMA(1,1)	1.233	1.703	2.247	2.712	3.073	1.115	0.992	1.005	1.000	0.999
ARMA(p,q)	1.214	1.656	1.793	1.874	1.856	1.109	1.023	1.007	0.991	0.982
VAR(1)	1.165	1.066	0.996	0.920	0.857	1.140	1.032	1.014	1.002	1.033
VAR(p)	1.348	1.205	1.024	0.907	0.809	1.224	1.052	1.022	1.003	1.029
BVAR	1.189	1.004	0.981	0.958	0.916	1.168	1.043	1.018	1.001	1.019
Factor(2)	1.085	1.132	0.988	0.879	0.934	0.974	1.031	1.044	0.985	1.077
Factor(k)	1.271	1.332	1.069	0.877	0.938	1.126	1.071	1.103	1.023	1.133
F(2)VAR(1)	1.200	1.188	1.121	1.165	1.243	1.107	1.029	1.019	1.006	1.031
F(2)VAR(p)	1.267	1.091	1.055	1.063	1.113	1.217	1.095	1.034	0.995	1.027

The rows are the forecasting models described in Sect. 2. The columns are forecast horizons. The table elements are the point estimates of the mean squared forecasting error expressed as a ratio of the mean squared forecasting error of the benchmark model for the same forecast horizon. Thus, a number less than 1 indicates that the model produced a smaller sample mean squared forecasting error than the benchmark model at that forecast horizon. The benchmark model is the simple mean forecast for both IP and EP

Table 6 Real activity: p values for conditional test by horizon

(a) IP	h = 1	h = 3	h = 6	h = 9	h = 12	h = 1 (b) EP	h = 3	h = 6	h = 9	h = 12
Naive	*0*	*0.004*	*0.001*	*0.007*	*0.01*	*0.002*	*0*	*0*	*0.001*	*0.001*
MA	*0.041*	0.061	*0.004*	*0.025*	*0.01*	*0.002*	0.052	*0.004*	0.059	0.09
MA-opt	0.062	0.057	*0.043*	0.054	0.055	*0.035*	*0.002*	*0.014*	0.083	0.054
SES	*0.004*	*0.014*	*0.001*	*0.004*	*0.009*	*0.003*	*0.025*	*0*	*0.015*	*0.004*
SES-opt	0.433	0.315	0.41	0.099	0.132	0.264	0.71	0.706	0.932	0.537
AR(1)	0.198	0.169	0.608	0.326	0.366	0.535	0.471	0.301	0.634	0.528
AR(p)	*0.039*	0.687	0.154	0.171	0.154	0.27	0.119	0.522	0.577	0.086
ARd(1)	0.199	0.742	0.141	**0.048**	0.053	0.412	0.125	0.42	0.42	0.115
ARd(p)	0.178	0.869	0.32	0.186	0.138	0.106	0.262	0.256	0.071	0.232
ARMA(1,1)	0.197	0.516	0.301	0.475	0.501	0.384	0.715	0.6	0.601	0.173
ARMA(p,q)	0.423	0.371	0.619	0.36	0.067	0.372	0.113	0.246	0.248	0.313
VAR(1)	*0.015*	0.162	0.376	0.203	0.277	0.12	0.305	0.318	0.949	0.106
VAR(p)	*0.028*	0.381	0.955	0.112	0.087	0.07	0.094	0.314	0.965	0.208
BVAR	0.051	0.798	0.569	0.544	0.523	0.126	0.092	0.118	0.975	*0.02*
Factor(2)	0.562	0.432	0.697	0.396	0.216	0.74	0.517	0.101	0.185	0.293
Factor(k)	0.143	0.305	0.765	0.617	**0.01**	0.35	0.763	0.278	0.51	0.333
F(2)VAR(1)	0.067	0.305	0.297	0.641	0.264	0.409	0.395	0.194	0.933	0.063
F(2)VAR(p)	*0.013*	0.309	0.686	0.597	0.29	0.071	0.326	0.086	0.904	0.097

The rows are the forecasting models described in Sect. 2. The columns are forecast horizons. The table elements are the p values for the Giacomini and White (2006) conditional test that the mean squared forecasting error is equal to that of the benchmark model. If a p value is less than 0.05, and the corresponding within-sample relative mean squared error in Table 5 is less than 1, then the p value is in a bold font, indicating that the model generates significantly better forecasts at the 5% significance level. If the p value is less than 0.05, and the corresponding within-sample relative mean squared error in Table 5 is greater than 1, then the p value is in italics, indicating that the model generates significantly worse forecasts. p values for hypotheses that are rejected by the sequentially rejective Bonferroni procedure of Holm (1979), applied to restrict the familywise error rate of each table to be less than 5%, are underlined. The benchmark model is the mean forecast

Table 7 Real activity: p values for unconditional test by horizon

(a) IP	h = 1	h = 3	h = 6	h = 9	h = 12	(b) EP h = 1	h = 3	h = 6	h = 9	h = 12
Naive	*0*	*0.001*	*0.002*	*0.002*	*0.002*	*0.001*	*0.003*	*0*	*0.002*	*0*
MA	*0.011*	*0.013*	*0.001*	*0.003*	*0.003*	*0.001*	*0.017*	*0.001*	*0.02*	*0.046*
MA-opt	0.055	0.01	0.007	0.01	0.009	*0.014*	*0.001*	*0.005*	*0.023*	*0.049*
SES	*0.001*	*0.002*	*0.001*	*0.001*	*0.001*	*0.001*	*0.03*	*0*	*0.018*	*0.001*
SES-opt	0.383	0.261	0.277	0.264	0.241	0.438	0.764	0.739	0.708	0.781
AR(1)	0.107	0.145	0.323	0.162	0.186	0.325	0.827	0.161	0.304	0.264
AR(p)	0.097	0.679	0.312	0.079	0.087	0.107	*0.046*	0.283	0.766	0.656
ARd(1)	0.082	0.785	0.459	0.057	0.076	0.195	0.091	0.394	0.182	0.407
ARd(p)	0.294	0.631	0.279	0.063	**0.041**	*0.045*	0.098	0.111	0.058	0.149
ARMA(1,1)	0.202	0.285	0.299	0.305	0.311	0.166	0.452	0.302	0.91	0.782
ARMA(p,q)	0.193	0.219	0.342	0.377	0.399	0.173	0.051	0.113	0.096	0.274
VAR(1)	*0.004*	0.242	0.913	0.102	0.121	0.054	0.121	0.41	0.913	*0.045*
VAR(p)	*0.024*	0.174	0.793	0.152	**0.037**	*0.021*	0.1	0.291	0.885	0.088
BVAR	*0.015*	0.952	0.886	0.717	0.378	*0.045*	*0.04*	0.226	0.949	0.129
Factor(2)	0.355	0.288	0.875	0.248	0.528	0.569	0.123	0.053	0.56	0.174
Factor(k)	*0.047*	0.078	0.509	0.36	0.59	0.228	0.318	0.069	0.445	0.164
F(2)VAR(1)	*0.022*	0.109	0.272	0.357	0.351	0.2	0.162	0.274	0.727	*0.029*
F(2)VAR(p)	*0.004*	0.143	0.38	0.591	0.526	*0.029*	0.137	0.206	0.748	0.088

The rows are the forecasting models described in Sect. 2. The columns are forecast horizons. The table elements are the p values for the Giacomini and White (2006) unconditional test that the mean squared forecasting error is equal to that of the benchmark model. If a p value is less than 0.05, and the corresponding within-sample relative mean squared error in Table 5 is less than 1, then the p value is in a bold font, indicating that the model generates significantly better forecasts at the 5% significance level. If the p value is less than 0.05, and the corresponding within-sample relative mean squared error in Table 5 is greater than 1, then the p value is in italics, indicating that the model generates significantly worse forecasts. p values for hypotheses that are rejected by the sequentially rejective Bonferroni procedure of Holm (1979), applied to restrict the familywise error rate of each table to be less than 5%, are underlined. The benchmark model is the mean forecast
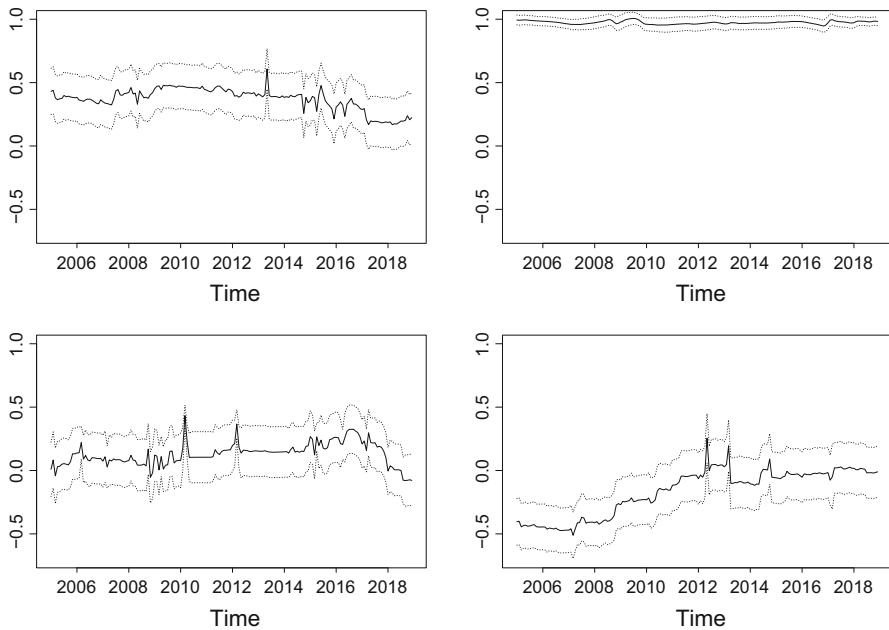
Fig. 2 AR(1) coefficient estimates: rolling 100-month window

Table 8 AR(1) estimates: full sample

	Dependent variable:			
	CPI_t	PPI_t	ΔIP_t	EP_t
Lagged dependent variable	0.402∗∗∗	0.978∗∗∗	0.069	0.162∗∗∗
	(0.056)	(0.013)	(0.062)	(0.061)
Constant	0.085∗∗∗	2.215∗	−0.027	1.489∗∗∗
	(0.021)	(1.298)	(0.080)	(0.378)

p values are in parentheses. ∗,∗∗ and ∗∗∗ indicate $p < 0.1$, $p < 0.05$ and $p < 0.01$

of the subsamples used in the forecasting exercises reported in Sect. 5 are presented in Fig. 2, and estimates of both coefficients computed using the full sample[28] are in Table 8.

Note that the coefficient estimates for the 'slope' term are fairly small for all variables except the PPI. The forecast equations for each horizon, computed using the coefficient estimates from Table 8 obtained using the whole sample, are presented in Table 9. With the exception of the PPI forecasts, the forecast equations all produce forecasts that are very close in value to the estimated unconditional means of each series for forecast horizons greater than 1. Thus, the main reason why the AR(1) model fails to produce forecasts superior to the simple mean for these variables is that

---

[28] Coefficient estimates for all the models computed using the entire sample are available in the online supplement to the paper.

Table 9 AR(1) forecast equations estimated using whole sample

	CPI	PPI
$h = 1$	$E(y_{R+1}|y_R) = 0.085 + 0.402 y_R$	$E(y_{R+1}|y_R) = 2.215 + 0.978 y_R$
$h = 3$	$E(y_{R+3}|y_R) = 0.133 + 0.065 y_R$	$E(y_{R+3}|y_R) = 6.5 + 0.935 y_R$
$h = 6$	$E(y_{R+6}|y_R) = 0.142 + 0.004 y_R$	$E(y_{R+6}|y_R) = 12.58 + 0.875 y_R$
$h = 9$	$E(y_{R+9}|y_R) = 0.142 + 0.000 y_R$	$E(y_{R+9}|y_R) = 18.268 + 0.819 y_R$
$h = 12$	$E(y_{R+12}|y_R) = 0.142 + 0.000 y_R$	$E(y_{R+12}|y_R) = 23.588 + 0.766 y_R$
$h = \infty$	$E(y_{R+h}) = 0.142$	$E(y_{R+h}) = 100.682$
	Δ IP	EP
$h = 1$	$E(y_{R+1}|y_R) = -0.027 + 0.069 y_R$	$E(y_{R+1}|y_R) = 1.489 + 0.162 y_R$
$h = 3$	$E(y_{R+3}|y_R) = -0.029 + 0.000 y_R$	$E(y_{R+3}|y_R) = 1.769 + 0.004 y_R$
$h = 6$	$E(y_{R+6}|y_R) = -0.029 + 0.000 y_R$	$E(y_{R+6}|y_R) = 1.777 + 0.000 y_R$
$h = 9$	$E(y_{R+9}|y_R) = -0.029 + 0.000 y_R$	$E(y_{R+9}|y_R) = 1.777 + 0.000 y_R$
$h = 12$	$E(y_{R+12}|y_R) = -0.029 + 0.000 y_R$	$E(y_{R+12}|y_R) = 1.777 + 0.000 y_R$
$h = \infty$	$E(y_{R+h}) = -0.029$	$E(y_{R+h}) = 1.777$

the estimated autocorrelation at a lag of 1 month is small. In contrast, the estimated 'slope' coefficient for PPI is consistently close to 1 which explains why the AR(1) model produces a similar forecasting performance to the naive model, and why both are markedly superior to the mean forecast.

More generally, there are a number of potential explanations for the relatively poor performance of standard forecasting models in the Chinese context.

Firstly, it might simply be that the available macroeconomic time series for China are not sufficiently long to provide statistically significant results. Our analytical approach effectively requires two data sets—one to estimate the model parameters and another to produce a series of out-of-sample forecasts. Western macroeconomic time series are generally available over much longer time spans than their Chinese counterparts. For example, Giacomini and White (2006) have 468 observations for their study of forecasting models for US industrial production, personal income, CPI inflation, and PPI inflation and generate forecasts for 318 months. In contrast, we have a total of 267 observations in total, yielding a maximum of 167 forecasts and raising the possibility that statistically significant results might emerge from a larger sample of forecasts, if it were available. Having said this, we note that only 31% of our CPI forecasts produce point estimates of mean squared errors that are smaller than those of the benchmark forecast. For the IP forecasts, this figure is 30%, and for EP it is 12%. Furthermore, as seen in Tables 3, 4, 6, and 7, in many cases, the benchmark model provides forecasts that are statistically significantly better than some of the other models considered. Consequently, we are not convinced that a larger sample would significantly change the conclusions of our study.

A second potential explanation for the failure of our more sophisticated models to consistently outperform simple benchmarks is nonstationarity. Forecasting models in the (vector) ARIMA tradition are well suited to variables that are weakly stationary

(perhaps after differencing). In cases where the marginal distributions of variables are slowly and continually changing over time, or where the marginal distributions change more suddenly, the rationale for such models is less clear. It should be noted that there can be little doubt that nonstationarity of these types exists. The question is whether they are so extensive in the Chinese context as to render traditional forecasting methods impotent. One particular potential source of nonstationarity is the financial crisis that commenced in late 2008. In the online supplement[29] to this paper, we have provided plots of the mean squared forecasting errors over 24-month widows for every forecasting model, forecasting horizon, and variable. We have also provided plots of the period-by-period difference in the squared forecasting errors generated by each forecasting model, and the squared forecasting errors generated by the benchmark model. It is clear from these plots that the period immediately following the onset of the crisis was one of great volatility in the forecasting performance of many models relative to the benchmarks. However, with the mean squared errors and p values recomputed using only forecasts from January 2010 to December 2018,[30] we find little difference from the results generated using the full samples that we report in Sect. 5. Consequently, it does not appear to be the case that our results are being driven primarily by the financial crisis. More generally, the impact of nonstationarity may vary across our forecasting models. Trivially, the naive forecast is unaffected by changes in the marginal distribution that occurred more than one month previously, whereas the estimation of parameters in other forecasting models usually places equal weight on all past observations. How nonstationarity would affect the ranking of forecasting models depends on the particular changes that occur in the marginal distribution over time. Therefore, it is difficult to make general comments about these effects. Further investigation into the issue of nonstationarity might proceed by comparing models that explicitly incorporate parameter drift and structural change to the benchmark models that we have used in this paper. This is an interesting but extensive task that we leave for future research.

Finally, it is possible that traditional forecasting methods perform relatively poorly for China because of low data quality. Much has been written about the quality of Chinese macroeconomic data (see, for example, Crabbe (2016), Holz (2014), Orlik (2011), Holz (2003) and Holz (2008)), and we do not have anything to add to this literature. To understand the potential impact of poor quality data, consider the following simple model of measurement error: $y_t = y_t^* + e_t$ where y_t^* is a weakly stationary stochastic process that we wish to forecast, and e_t is a measurement error that we assume to be serially uncorrelated and uncorrelated with y_t^* at all leads and lags. It is simple to show that $Corr(y_t, y_{t-j}) \leqslant Corr(y_t^*, y_{t-j}^*)$ for all $j \in \mathbb{N}$ and that the difference between the two correlations increases with $Var(e_t)$. Thus, measurement error is a possible cause of the small estimates of the AR(1) slope coefficients reported in Fig. 2 and Table 8 and may contribute to the relatively poor forecasting power of models that exploit correlation structure. Ultimately, if the published data do not reliably measure the variables that they are claimed to represent, then forecasting those variables may be a hopeless task and, in any case, might be argued to be pointless.

---

[29] https://cheaton.github.io/China-Forecasting/online_supplement.pdf.

[30] These tables are presented in the online supplement.

In such a situation, the way forward is to seek better ways to measure the aspects of Chinese economic activity that are of interest, and to forecast these measures directly, rather than relying on the official data collection.

7 Conclusion

For analysts interested in forecasting real activity in China, the results in our paper provide clear guidance. We find no evidence that any of the forecasting models that we consider provide smaller mean squared forecasting errors than the simple mean forecast. For inflation, the results are less clear-cut. For PPI inflation, we found statistically significant evidence that AR, ARMA, VAR, and BVAR models produce superior forecasts to the naive model at a forecasting horizon of 1 month. For each of these models, the estimated mean squared forecasting error is approximately half that generated by the benchmark naive forecasting model. However, we do not find strong evidence that any of the models that we consider produce better forecasts of the PPI than the naive model at horizons longer than 1 month, and we find no evidence that any of the models we consider produce superior forecasts of the CPI than the simple mean model at any horizon.

An important contribution of our paper is that (to our knowledge) it is the first paper on forecasting the Chinese macroeconomy that has formally tested hypotheses that the forecasting models under consideration have forecasting power equal to that of simple benchmark models. Our samples of forecasts range in size from 156 (for 12-month-ahead forecasts) to 167 (for 1-month-ahead forecasts). These are much larger samples than have been generated in the prior published literature, which range from five observations to 60. The fact that we find so little statistically significant evidence that sophisticated forecasting models provide forecasts superior to those available from simple benchmark models raises considerable doubt about the generality of previously published results that suggest otherwise, beyond the particular sample periods considered.

While our results might create an impression that, for the most part, Chinese macroeconomic variables *cannot* be forecast, such a conclusion should be resisted. Instead, our main finding is that, with the exception of the PPI, for the variables we consider there does not exist sufficiently strong serial correlation for the widely used conventional forecasting techniques to consistently outperform a forecast based on an estimator of the mean, in the sense of producing a smaller mean squared forecasting error. Therefore, for CPI, industrial production, and electricity production, we recommend the mean forecast due to its simplicity. Note, however, that our results do not necessarily imply that the more sophisticated models we consider would not perform relatively well when judged against criteria other than the mean squared forecasting error (for example, for sign forecasting). Our findings also do not preclude the possibility that using nonsample information might improve the forecasts, that less conventional forecasting models might perform better than the models we have considered, or that alternative measures of inflation and real activity might have more forecastable structure than the ones used in this paper. Thus, while our results have

important implications for those interested in forecasting Chinese macroeconomic variables, considerable scope remains for future research.

Appendix

The following tables list the variables that were used to estimate the factors used in the factor forecasts (Tables 10, 11, 12, 13). All variables were sourced from Thomson Reuters Datastream and are monthly variables dating from October 1996 to December 2018. The tables provide the Datastream variable name and code in the first two columns. The third column contains NA if the variable has missing observations. The fourth column contains a 2 if the variable was used in first difference form, and a 3 if logged first differences were used.

Table 10 Real activity

Variable	Code	NAs	Trans.
CH INDUSTRIAL PRODUCTION INDEX VOLN	CHIPTOT.H	NA	3
CH OUTPUT OF INDUSTRIAL PRODUCTS—CRUDE STEEL VOLN	CHVALSTLH	NA	3
CH OUTPUT OF INDUSTRIAL PRODUCTS—MOTOR VEHICLES VOLN	CHVALMVHH	NA	3
CH OUTPUT OF PRIMARY ENERGY—HYDRO-ELECTRIC POWER VOLN	CHYDENRGR	NA	3
CH INDUSTRIAL PRODUCTION: ELECTRICITY VOLN	CHPBRENTP	NA	3
CH CRUDE OIL PRODUCTION—CHINA VOLN	CHPCOBD.P	NA	2
CH INVESTMENT: FIXED ASSETS—PRIMARY INDUSTRY(CMLV)(%YOY) NADJ	CHIFARI%A	NA	2
CH INVESTMENT: FXD.ASS.—SECONDARY INDUSTRY(CMLV) (%YOY) NADJ	CHIFARS%A	NA	2
CH INVESTMENT: FXD.ASS.—TERTIARY INDUSTRY (CMLV) (%YOY) NADJ	CHIFART%A	NA	2
CH GOVERNMENT EXPENDITURE (BUDGETARY) CURN	CHEXPENDA		3
CH GOVERNMENT REVENUE—TAXES CURN	CHTAXREVA	NA	3
CH EXPORTS CURN	CHEXPGDSA		3
CH IMPORTS CURN	CHIMPGDSA		3
CH CARGO HANDLED AT MAJOR SEAPORTS VOLN	CHCGOTOTP	NA	3
CH EXTERNAL TRADE BALANCE CURN	CHVISGDSA		2
CH FREIGHT TRAFFIC VOLN	CHFRETOTP	NA	3
CH RAILWAY CARGO TURNOVER VOLN	CHCNXNO.M		3
CH RETAIL SALES: CONSUMER GOODS (UNREVISED) CURN	CHRETUOTA	NA	3
CH CONSUMER CONFIDENCE INDEX NADJ	CHCNFCONR		3
CH CONSUMER EXPECTATION INDEX NADJ	CHNBSCXIR		3
CH CONSUMER SATISFACTION INDEX NADJ	CHNBSCSIR		3

Table 11 Prices

Variable	Code	NAs	Trans.
CH CPI (CPPY=100) NADJ	CHCONPRCF		2
OPEC Oil Basket Price U$/Bbl	T42463		3
CH RPI—RURAL (CPPY=100) NADJ	CHGRPINRF		2
CH RPI (CPPY=100) NADJ	CHORPOVEF		2
CH RPI: URBAN (CPPY=100) NADJ	CHGRPINUF		2

Table 12 Financial variables

Variable	Code	NAs	Trans.
CH 20 DAY RELOAN RATE NADJ	CHRELOAN		2
CH DISCOUNT RATE (EP) NADJ	CHDISCRT		2
CH PRIME LENDING RATE (EP)	CHBANKR.		2
SHANGHAI SE COMPOSITE—PRICE INDEX	CHSCOMP		2
SHENZHEN SE COMPOSITE—PRICE INDEX	CHZCOMP		3
CH CHINESE YUAN TO 100 JAPANESE YEN—WEIGHTED AVERAGE CURN	CHXRWJPY		2
CH CHINESE YUAN TO US DOLLAR (AVERAGE AMOUNT) NADJ	CHXRUSD.		2
CH NOMINAL EFFECTIVE FX RATE(NEER)BASED ON CONSUMER PRICE INDEX	CHI..NECE		2
CH MONEY SUPPLY—CURRENCY IN CIRCULATION CURN	CHM0....A		3
CH MONEY SUPPLY—M1 CURN	CHM1....A		3
CH MONEY SUPPLY—M2 (METHODBREAK OCT2011) CURN	CHM2....A		3
CH FINANCIAL INST.—DEPOSITS (RMB)(METHO BREAK IN 2011) CURN	CHFISD.RA		3
CH FINANCIAL INST-FINANCIAL BONDS(RMB)(METHO BREAK IN 2011) CURN	CHKFIABFA		3
CH GOLD AND FOREIGN RESERVES—FOREIGN RESERVE CURN	CHRESERVA		2

Table 13 External trade

Variable	Code	NAs	Trans.
EX EU—IMPORTS FROM CHINA, (CIF) CURN	EXI7D1CHA		3
UK UNITED KINGDOM—IMPORTS FROM CHINA, (CIF) CURN	UKI7D1CHA		3
US UNITED STATES—IMPORTS FROM CHINA, (CIF) CURN	USI7D1CHA		3
JP JAPAN—IMPORTS FROM CHINA, (CIF) CURN	JPI7D1CHA		3
JP JAPAN—EXPORTS TO CHINA CURN	JPI7D0CHA		3
UK UNITED KINGDOM—EXPORTS TO CHINA CURN	UKI7D0CHA		3
US UNITED STATES—EXPORTS TO CHINA CURN	USI7D0CHA		3
EX EU—EXPORTS TO CHINA CURN	EXI7D0CHA		3
AU AUSTRALIA—EXPORTS TO CHINA CURN	AUI7D0CHA		3
CN CANADA—EXPORTS TO CHINA CURN	CNI7D0CHA		3

References

Ang A, Bekaert G, Wei M (2007) Do macro variables, asset markets, or surveys forecast inflation better? J Monet Econ 54:1163–1212

Artis MJ, Banerjee A, Marcellino M (2005) Factor forecasts for the UK. J Forecast 24(4):279–298

Atkeson A, Ohanian LE (2001) Are Phillips curves useful for forecasting inflation? Fed Reserve Bank Minneap Q Rev 25(1):2–11

Bai J, Ng S (2002) Determining the number of factors in approximate factor models. Econometrica 70(1):191–221

Bandt OD, Michaux E, Bruneau C, Flageollet A (2007) Forecasting inflation using economic indicators: the case of France. J Forecast 26(1):1–22

Bernanke BS, Boivin J, Eliasz P (2005) Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. Q J Econ 120(1):387–422

Boivin J, Ng S (2006) Are more data always better for factor analysis? J Econ 132(1):169–194

Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time series analysis: forecasting and control. Wiley, Hoboken

Camba-Mendez G, Kapetanios G, Smith RJ, Weale MR (2001) An automatic leading indicator of economic activity: forecasting GDP growth for European countries. Econom J 4(1):S56–S90

Canova F (2007) Methods for applied macroeconomic research, vol 13. Princeton University Press, Princeton

Ciccarelli M, Mojon B (2010) Global Inflation. Rev Econ Stat 92(3):524–535

Crabbe M (2016) Myth-busting China's numbers: understanding and using China's statistics. Springer, Berlin

den Reijer A (2005) Forecasting Dutch GDP using large scale factor models. DNB Working Papers 028, Netherlands Central Bank, Research Department

Doan T, Litterman R, Sims C (1984) Forecasting and conditional projection using realistic prior distributions. Econom Rev 3(1):1–100

Faust J, Wright JH (2013) Forecasting inflation. In: Handbook of economic forecasting, vol 2. Elsevier, pp 2–56

Forni M, Hallin M, Lippi M, Reichlin L (2003) Do financial variables help forecasting inflation and real activity in the euro area? J Monet Econ 50:1243–1255

Gardner G, Harvey A C, Phillips G D (1980) Algorithm AS 154: an algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of kalman filtering. J Roy Stat Soc: Ser C (Appl Stat) 29(3):311–322

Ghysels E (2016) Macroeconomics and the reality of mixed frequency data. J Econom 193(2):294–314

Giacomini R, White H (2006) Tests of conditional predictive ability. Econometrica 74(6):1545–1578

Hamilton J (1994) Time Series Analysis. Princeton University Press, Princeton

He Q, Fan C (2015) Forecasting inflation in China. Emerg Mark Finance Trade 51(4):689–700

Higgins P, Zha T, Zhong W (2016) Forecasting China's economic growth and inflation. China Econ Rev 41:46–61

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Holz CA (2003) Fast, clear and accurate": how reliable are Chinese output and economic growth statistics? China Q 173:122–163

Holz CA (2008) China's 2004 economic census and 2006 benchmark revision of GDP statistics: more questions than answers? China Q 193:150–163

Holz CA (2014) The quality of China's GDP statistics. China Econ Rev 30:309–338

Hyndman RJ, Khandakar Y (2008) Automatic time series forecasting: the forecast package for R. J Stat Softw 26(3):1–22

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2018) forecast: forecasting functions for time series and linear models. R package version 8:4

Kamal L (2013) Forecasting inflation in China. J Appl Bus Res (JABR) 29(6):1825–1832

Lin C-Y, Wang C (2013) Forecasting China's inflation in a data-rich environment. Appl Econ 45(21):3049–3057

Litterman RB (1986) Forecasting with Bayesian vector autoregressions—five years of experience. J Bus Econ Stat 4(1):25–38

Marcellino M, Stock JH, Watson MW (2003) Macroeconomic forecasting in the euro area: country specific versus area-wide information. Eur Econ Rev 47(1):1–18

Marcellino M, Stock JH, Watson MW (2006) A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. J Econom 135(1):499–526

Mehrotra A, Sánchez-Fung JR (2008) Forecasting inflation in China. China Econ J 1(3):317–322

O'Hara K (2018) BMR: Bayesian Macroeconometrics in R. R package version 0.11.0

Orlik T (2011) Understanding China's economic indicators: translating the data into investment opportunities, 1st edn. FT Press, Upper Saddle River

Pfaff B (2008) VAR, SVAR and SVEC models: implementation within R package vars. J Stat Softw 27(4):1–32

Pincheira P, Gatty A (2016) Forecasting Chilean inflation with international factors. Emp Econ 51:981–1010

Pincheira PM (2015) Medel (2015) Forecasting inflation with a simple and accurate benchmark: the case of the us and a set of inflation targeting countries. Financ Uver: Czech J Econ 65(1):29

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Sax C, Eddelbuettel D (2018) Seasonal adjustment by X-13ARIMA-SEATS in R. J Stat Softw 87(11):1–17

Schneider M, Spitzer M (2004) Forecasting Austrian GDP using the generalized dynamic factor model. Working Papers 89, Oesterreichische Nationalbank (Austrian Central Bank)

Schumacher C (2007) Forecasting German GDP using alternative factor models based on large datasets. J Forecast 26(4):271–302

Stekler HO, Zhang H (2013) An evaluation of Chinese economic forecasts. J Chin Econ Bus Stud 11(4):251–259

Stock JH, Watson MW (1999) Forecasting inflation. J Monet Econ 44(2):293–335

Stock JH, Watson MW (2002a) Forecasting using principal components from a large number of predictors. J Am Stat Assoc 97(460):1167–79

Stock JH, Watson MW (2002b) Macroeconomic forecasting using diffusion indexes. J Bus Econ Stat 20(2):147–162

Stock JH, Watson MW (2008) Phillips curve inflation forecasts. Technical report, National Bureau of Economic Research

Trapletti A, Hornik K (2018) tseries: Time Series Analysis and Computational Finance. R package version 0.10-45

Zeileis A, Grothendieck G (2005) zoo: S3 infrastructure for regular and irregular time series. J Stat Softw 14(6):1–27

Zhou R, Wang X, Tong G (2013) Forecasting macroeconomy based on the term structure of credit spreads: Evidence from China. Appl Econ Lett 20(15):1363–1367

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.