BBVA Research

Creating Opportunities

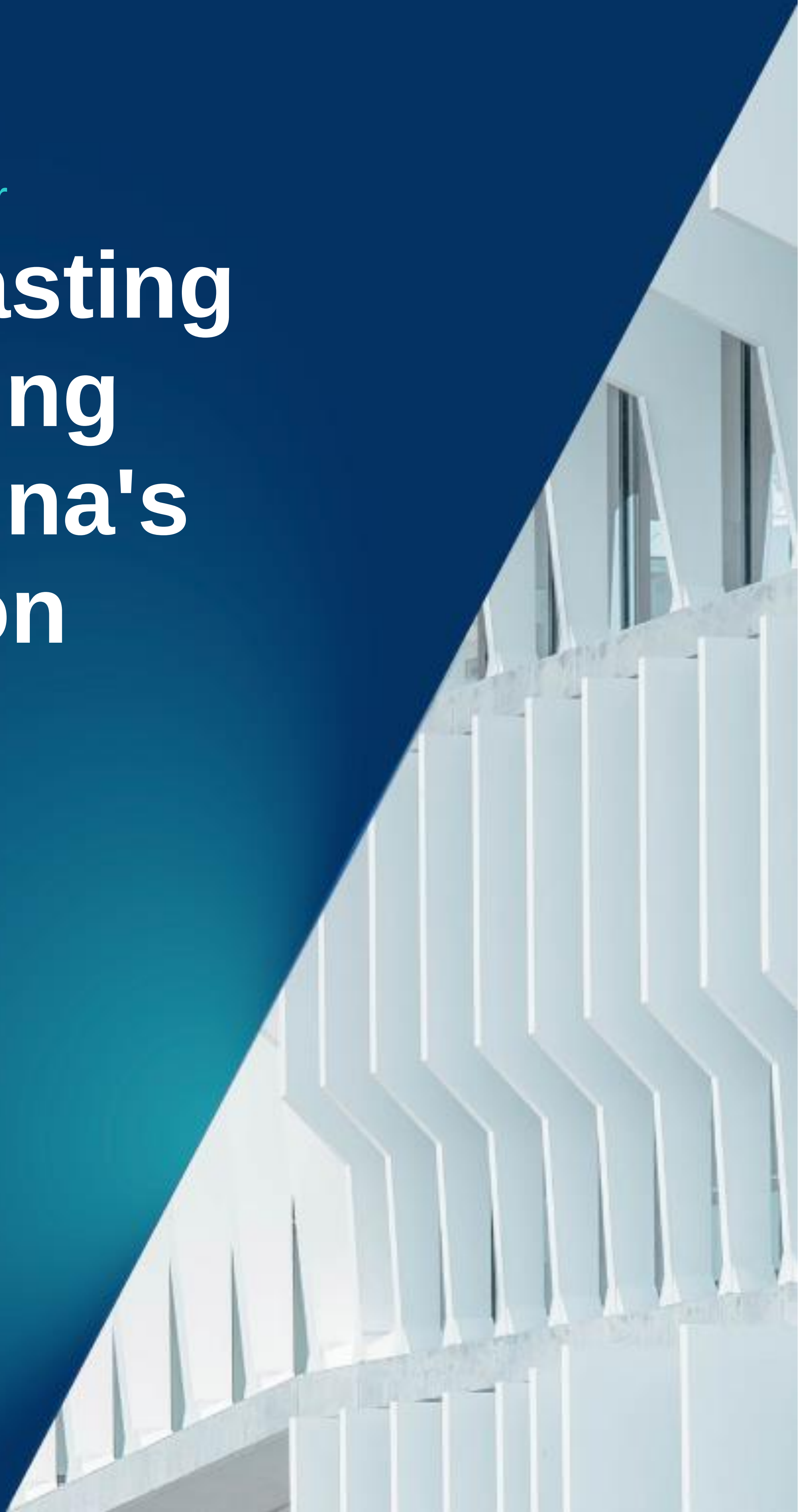
N.º 20/05

Working paper

Forecasting modeling for China's inflation

Jinyue Dong

March 2020



Forecasting modeling for China's inflation

Jinyue Dong

Abstract

Against the background of African Swine Flu outbreak in China, COVID-19 and global oil price dipping, we are trying to assess a large group of forecasting models' performance in predicting China's inflation. Both linear and structural forecasting models are discussed, estimated and evaluated based on some standard criteria such as RMSE, MAE and Theil-U. Our empirical results suggest that unlike widely found in the literature that complicated models are difficult to significantly beat the naïve random walk model, by adding monetary and economic indicators into the linear model, as well as by adding the structural forms, some models are found to significantly out-perform random walk model in China's inflation forecasting practice. Among all of these models, multi-variate model performs best, and the VECM follows. That means, by including the error-correction term to solve the non-stationary issue, VECM could out-perform the basic reduced-form VAR to predict China's inflation.

Keywords: inflation forecasting, model comparison, linear models, VAR, VECM, China

JEL classification: C22, C32, C52, C53

Some stylized facts on China's rocketing inflation in 2019-2020

China's CPI inflation has experienced a significant pick-up starting from Q2 2019 due to the outbreak of Africa Swine Flu (ASF) in mainland China. In particular, headline CPI growth increased by 3.3% y/y in Q2-Q4 2019, much higher than its reading of 2018 at 2.1%. Moreover, CPI inflation of February 2020 came out at 5.2% y/y, almost in line with the previous month's high growth at 5.4% y/y, which is the highest record since October 2011. The persistently high CPI was primarily driven by the impact of African Swine Flu which has lasted throughout 2019 and is expected to continue in several years that led to the substantial supply reduction in pork and thereby contributed to the significant rise of food prices. Except for the ASF outbreak, the recent globally spread COVID-19 and global oil price dipping might add more uncertainties to China's inflation which motivate us to develop and assess the forecasting models to predict China's inflation.

Against this background, forecasting CPI in China becomes an important issue for Chinese authorities. By applying a large group of forecasting models to Chinese inflation prediction, this project tries to answer the following questions: (i) Among the six typical forecasting models, which model could provide the best forecasting for Chinese inflation (i.e. to produce the least forecasting error)? (ii) Amid large amount of forecasting literature which indicates that complicated econometrics models cannot significantly beat the naïve random walk model, is this also true in Chinese inflation case? (iii) By adding the monetary and economic indicators as well as adding structural form into the model, could it out-perform the random walk and univariate model? (iv) By adding the error-correction term into the reduced-form VAR, does the VECM have better performance than the basic VAR? (v) In a more economic sense, how will the high CPI due to the Africa Swine Flu in China affect the monetary policy room amid the slowdown business cycle? And ultimately will it be a constraint of the authorities' counter-cyclical measures?

To answer these questions, we are trying to apply two groups of typical forecasting models to predict Chinese CPI in 2020 and compare these models' forecasting performance based on the standard criteria. In particular, one group of the models are linear model group which includes random walk model, univariate model, multi-variate (two-pillar Philips Curve) model, ARIMA; and the other group is structural form based, mainly include VAR and VECM. On top of the estimation and forecasting, we also compare all of the six models' forecasting performance based on the standard criteria such as RMSE, MAE and Theil-U.

As we all know, all econometrics forecasting models are the abstract of the reality and no model could capture every perspective of it. Thus, there does not exist the "true" model or "perfect" model for any forecasting work. However, we could still rank different forecasting models by some standard criteria based on calculating their forecasting errors etc. to find what the best-performing model is and to identify what is the implication for us. Nevertheless, for academic as well as policy reasons, we are still interested in knowing the "driving forces" of CPI inflation in China, which may not be directly observable. A comparison of model performance would shed light on those driving forces.

The second section below provides the details of a large set of linear and non-linear forecasting models. The third section provides the formal econometric evidence on the estimation and prediction performance of them. Model comparison results are displayed in the fourth section. The forecast evaluation results suggest that, unlike what has been found widely in the literature for many other countries' empirical evidence which indicates it is difficult to find econometric models that significantly outperform a simple random walk model (Stock and Watson 2007, Arratibel, Kamps, Leiner-Killinger, 2009, etc.), in China's inflation case, the multi-variate model including the monetary and economic variables, as well as the VECM by adding the error-correction terms for VAR could out-perform the random walk model significantly.

Models:

In this section, we briefly introduce two groups of models which include six models in our arsenal. In the first group, we mainly discuss the liner and single-equation forecasting models, as they have been found in the forecast evaluation literature that outperform some large multi-equation models such as vector auto-regressions or traditional structural DSGE type of macro-econometric models. The first group includes the naïve random walk model (without drift), the simple univariate autoregressive model, a multi-variate (two-pillar Phillips curve) model as well as ARIMA.

In the second group, we discuss about some VAR-type models for forecasting practice in order to capture the simultaneous effect of the endogenous macro variables, which mainly include the reduced-form VAR and Vector Error Correction Model (VECM) which solves the potential non-stationary of the standard VAR model.

1.1 The linear and single-equation model group:

The linear and single-equation model group mostly builds on the empirical framework first applied by Stock and Watson (1999) to U.S. data and subsequently applied by Nicoletti-Altimari (2001), Carstensen (2007) and Hofmann (2008) and Arratibel, Kamps, Leiner-Killinger (2009) to Euro Area data.

As it is common in the literature, we implement single-equation forecasting models in this section with a few independent variables for the following reasons. First, actually in our practice, we have tried a very large number of endogenous macro variables to be included in our single equation models, however, we only display the finally chosen ones which are significant in the statistic sense and are not subject to multicollinearity problem. Second, we concentrate on models with a few of independent variables in order to account for our short data samples. Increasing the number of indicator variables included in the models would quickly exhaust the degrees of freedom.

In the below notations, π is defined as the year on year growth rate of CPI. It has become common practice in the literature to focus on average inflation over the coming h quarters.

First, the benchmark model has been routinely used in the forecasting exercise which is the "naïve" random walk model. It states that the forecast of h-quarter ahead inflation at time t is simply the last observable value of h-quarter inflation:

(i) The naïve random walk model and the random walk with drift model:

$$\hat{\pi}_t = \pi_{t-1}$$
$$\hat{\pi}_t = \pi_{t-1} + \alpha$$

However, in our practice, we drop the random walk with drift model because our CPI growth data is observed to be "not trending with a drift". Thus, applying the random walk with drift model seems irrelevant. Here, we assume as is the case in practice that the inflation rate at time t, is not part of the time-t information set.

(ii) The simple univariate autoregressive model (hereafter AR).

Autoregressive model is based on the assumption that each value of the time series depends only on the weighed sum of the product of the previous values and the regression coefficient, plus the residual.

Again, we assume that the inflation rate at time t, is not part of the time-t information set. We follow Carstensen (2007) and use a stepwise procedure to determine which lagged values of inflation enter the model. In practice, we apply Akaike information criterion (AIC) to determine the optimal lag value.

$$\pi_t = \beta_0 + \beta_1(L)\pi_{t-1} + u_t$$

where $\beta_1(L)$ is a finite polynomial of order p in the lag operator.

$$\beta_1(L) = \beta_{11}L + \ldots + \beta_{1p}L^p \ .$$

The Autoregressive model is capable in a wide variety of time series forecasting by adjusting the regression coefficients. The difference between the Autoregressive models and other conventional regression models is with respect to the assumption of the independence of the error term. Since the independent variables are time-lagged values for the dependent variable, the assumption of uncorrelated error is easily violated.

Second, we also apply multi-variate forecasting models. The first set of forecasting models used in the forecasting exercise extends the univariate autoregressive model discussed above by both monetary and economic indicators:

(iii) The multi-variate linear forecasting model includes both monetary and economic indicators.

Gerlach (2003, 2004) proposed a simple tri-variate (two-pillar Phillips curve) forecasting model intended to capture at the same time the information of both monetary and economic indicators. Such tri-variate forecasting models specify h-quarter ahead average inflation as a function of its own lags, lags of broad money growth, as well as lags of non-monetary indicator variables.

A stepwise procedure is used to determine which lagged values of inflation, of trend money growth and of the respective indicator variable enter the model. Our monetary indicators include (all are year on year growth): M2, RMB bank loans; the economic indicators include (all are year on year growth): unemployment, GDP.

$$\pi_t = \beta_0 + \beta_1(L)\pi_{t-1} + \beta_2(L)x_{t-1} + \beta_3(L)m_{t-1} + u_{t-1}$$

Third, we provide the standard ARIMA model for analyzing and building a forecasting model which best represents a time series by modeling the correlations in the data. In the empirical research, many advantages of the ARIMA model were found and support the ARIMA as a proper way in especially short-term time series forecasting (Box, 1970; Jarrett, 1990). Taking advantage of its strictly statistical approach, the ARIMA method only requires the prior data of a time series to generalize the forecast. Hence, the ARIMA method can increase the forecast accuracy while keeping the number of parameters to a minimum.

On the other hand, its proved main disadvantage is that the ARIMA models, as all forecasting methods, are essentially “backward looking”. Such that, the long term forecast eventually goes to be straight line and poor at predicting series with turning points. Fortunately, based on the structure of China’s CPI historical data, this disadvantage does not influence our CPI forecast much.

(iv) ARIMA(p,d,q):

If we combine differencing with auto-regression (AR) and a moving average (MA) model, we obtain a non-seasonal ARIMA model. Since AR model was introduced in the previous section, here we only introduce MA model first and then the ARIMA model.

The basic idea of Moving-Average model is firstly finding the mean for a specified set of values and then using it to forecast the next period and correcting for any mistakes made in the last few forecasts. It takes this form:

$$\pi_t' = w_0 + \varepsilon_t - w_1\varepsilon_{t-1} - \dots - w_q\varepsilon_{q-1}$$

To specify a Moving-Average, the number and the value of the q moving average parameter have to be decided subject to the certain restrictions in value in order for the process to be stationary. The Moving-Average model works well with stationary data, a type of time series without trend or seasonality.

Altogether, to combine the AR and MA model, ARIMA is an acronym for Auto Regressive Integrated Moving Average (in this context, “integration” is the reverse of differencing). The full model can be written as:

$$\pi_t' = c + \beta_1\pi_{t-1}' + \dots + \beta_p\pi_{t-p}' + \theta_1\varepsilon_{t-1} + \dots + \theta_q\varepsilon_{q-1} + \varepsilon_t$$

where π_t' is the differenced series (it may have been differenced more than once). The “predictors” on the right-hand side include both lagged values of π_t' and lagged errors. In the notations above:

p=order of the autoregressive part;

d=degree of first differencing involved;

q=order of the moving average part.

Thus, the ARIMA model uses combinations of past values and past forecasting errors and offer a potential for fitting models that could not be adequately fitted by using an AR or an MA model alone. Furthermore, the addition of the differencing eliminates most non-stationarity in the series.

1.2 The VAR-type multi-equation forecast models

In the second model group, the VAR-type of models are discussed. It mainly includes the reduced-form VAR and VECM.

(i) Reduced-form VAR model:

A pth order Vector Autoregression (denoted as a VAR(p)) for N variables is defined as a set of equations where each variable depends on p lags of the all N variables included in the model. In general, the VAR(p) model could be written as:

$$Y_t = c + B_1 Y_{t-1} + B_2 Y_{t-2} ... + B_p Y_{t-p} + v_t$$
$$E(v_t' v_s) = \Sigma(\text{if } t = s)$$
$$E(v_t' v_s) = 0(\text{if t} \neq \text{s})$$
$$E(v_t) = 0$$

Where Y_t is the data matrix containing the N variables in the VAR. The matrices B_1 to B_p represent N byN matrices of coefficients on each lag of Y_t and c denotes the N*1 vector of constants.

The coefficients of the VAR model in equation above can be estimated using OLS for each equation of the VAR.

The above equation can be written compactly as:

$$Y^i = XB_i + v^i$$

where $X = \{c_i, Y_{it-1}, Y_{it-2} ... Y_{it-p}\}$ denotes the right-hand side of the ith equation. Note that each equation in the VAR has identical regressors. Therefore, we can re-write all equations of the VAR as the following system:

$$Y = (I_N \otimes X)B + v$$

where $\otimes$ denotes the Kronecker product.

With identical regressors, Generalized Least Squares (GLS) is equivalent to OLS. This is in-turn equivalent to estimating each equation of the VAR via OLS:

$$B_i = (X'X)^{-1}(X'Y^i)$$

Or the coefficients for all equations in the VAR can be estimated by using:

$$B = [I_N \otimes (X'X)^{-1} X']Y$$

The ijth element of the covariance matrix of the residuals can be estimated using:

$$\Sigma_{ij} = \frac{(Y^i - XB_i)'(Y^j - XB_j)}{T - k}$$

where T is total sample size and k = NP + 1 are the total number of parameters in each equation of the VAR model. The standard errors of the VAR coefficients are obtained by calculating the covariance matrix: This is given by:

$$\text{cov}(B) = \Sigma \otimes (\text{X}'\text{X})^{-1}$$

The standard errors of the B matrix can be calculated as the square root of the diagonal of the covariance matrix in the above equation.

(ii) Vector Error Correction Model (VECM):

The reason that we also include Vector Error Correction Model (VECM) is to deal with the potential co-integration problem. The above vector autoregressive (VAR) model is a general framework used to describe the dynamic interrelationship among stationary variables. So, the first step in time-series analysis should be to determine whether the levels of the data are stationary. If not, take the first differences of the series. Usually, if the levels (or log-levels) of your time series are not stationary, the first differences will be.

If the time series are not stationary then the VAR framework needs to be modified to allow consistent estimation of the relationships among the series. The VECM is just a special case of the VAR for variables that are stationary in their differences (i.e., I(1)). The VECM can also take into account of any co-integrating relationships among the variables.

The reason we want to try the VECM to deal with the non-stationary problem is based on our data structure. We decide to use the vector error correction model because: (1) the time series are not stationary in their levels but seems like stationary in their differences and (2) the variables are co-integrated. Our initial impressions are gained from looking at the plots of the data series. The graph below shows that our CPI line at the Q1 1992 to Q1 1998 period displays some trend structure while the rest of the time looks more likely to be stationary and this pattern also happens in other variables. (Figure 1)

Suppose x and y and I(1) and co-integrated. Then ε_t is I(0) in the co-integrating equation:

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

These equations often are interpreted as long-run or equilibrium relationships between x and y. A researcher will also be interested in the short-run dynamics - the way that x and y fluctuate around this long-run relationship, as in a business cycle. This is done by estimating an error correction model, which contains first differences of x and y, their lags, and an error correction term. An ECM is:

$$Dy_t = \mu + \gamma_1 Dy_{t-1} + \ldots + \gamma_p Dy_{p-1} + \omega_0 Dx_{t-1} + \ldots + \omega_r Dx_{t-r} + \lambda EC_{t-1} + u_t$$

where $EC_t = y_t - (a + bx_t)$, the error correction term, is the lagged OLS residual from the cointegrating equation. The lag orders p and r are chosen in the usual ways discussed earlier.

Since the OLS estimate of β is super consistent, the sampling error from estimating it in the cointegrating equation is less important than the sampling error of the ECM coefficient estimates asymptotically. This justifies a two-step procedure where the cointegrating equation is estimated first, followed by an ECM with the lagged OLS residual (EC_t−1) from the estimated cointegrating equation serving as the error correction term in the ECM.

Step 1: OLS: $y_t = a + bx_t + EC_t$ (cointegrating equation)

Step 2: OLS: $Dy_t = \mu^* + \omega_1 Dx_t + \lambda EC_{t-1} + u_t$ (error correlation model)

Unit root tests are performed using the augmented Dickey-Fuller regressions, which require some judgment about specification.

The method of comparing models' performance

Obviously, all models are abstract of the reality and hence no model can capture every aspect of the reality. The truth is that there is no “true” or “perfect” model in the econometrics modeling. Nevertheless, for academic as well as policy reasons, we are still interested in knowing the “driving forces” of the asset markets, which may not be directly observable. A comparison of model performance would shed light on those driving forces. (Leung, Kwan and Dong, 2014)

In order to compare the performance of different model's forecasting ability, we perform the standard RSME and MAE criteria to calculate their forecasting error.

We first compute the model-generated one step forecasting of CPI inflation of each model. We then compare the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of each model, based on the comparison of one-step forecasting aggregate inflation data and the observed actual return data.

To calculate the RMSE and MAE based on the comparison of model-generated CPI inflation and the actual data, the RMSE and MAE formula is provided as follows:

$$RMSE(i) = \sqrt{\frac{1}{N}\left(\sum_{j=1}^{N}\left(e^{i}\right)^{2}\right)}$$

$$MAE(i) = \frac{1}{N}\sum_{j=1}^{N}\left|e^{i}\right|$$

We also provide the Theil's U criteria which is a relative accuracy measure that compares the forecasted results with the results of forecasting with minimal historical data. It also squares the deviations to give more weight to large errors and to exaggerate errors, which can help eliminate methods with large errors. The Theil's U criteria could directly tell us whether the model is outperforming or underperforming the naive random walk model. The formula for calculating Theil's U statistic:

$$U = \sqrt{\frac{\sum_{t=1}^{n-1}\left(\frac{f_{t+1}-a_{t+1}}{a_t}\right)^2}{\sum_{t=1}^{n-1}\left(\frac{a_{t+1}-a_t}{a_t}\right)^2}}$$

where f is the forecasting value and a is the actual value. If U = 1, there is no difference between a naïve forecast and the technique used; if U < 1 the technique is better than a naïve forecast; and if U > 1 the technique is no better than a naïve forecast.

Data and Empirical results

This sector shows the data we use in the empirical work and present the RATS results of different models' estimation and forecasting. We use the data from 1992 Q1 to 2019 Q4, all are quarterly series and year on year % changes for the stationary purpose. The main variables include: CPI inflation, GDP, aggregate loan growth rate, unemployment, M2 growth. They are all from CEIC data base. We also applied some other variables such as unemployment, VAT, domestic and global oil price etc., but due to the worse-than-expected fitting, we only present the variables above.

The estimation results of six different models are summarized in Table 1. Some analysis of the forecasting results of different models is also discussed. Moreover, we provide the model comparison result of their forecasting ability based on RMSE, MAE and Theil's U criteria which is summarized in the Table 2. Separately, we discuss about VAR and VECM in more details.

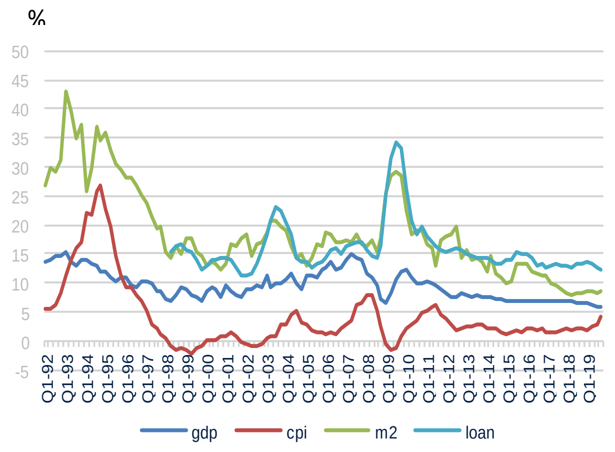
Table 1. **ESTIMATION RESULTS OF THE SIX MODELS**

	Random Walk Model	Univariate Model	Multivariate Model	ARIMA	VAR	VECM
β1	.	0.9696	0.7957	0.8888	0.7976	0.7957
(coefficient of CPI_t-1)		(41.5)***	(17.4874)***	(19.2311)***	(2.8313)***	(17.4874)***
β2			-0.1656		-0.1514	-0.1655
(coeffieicnt of M2_t-1)			(-4.3159)***		(-3.6402)***	(0.0384)***
β3			0.1574		0.1051	0.1574
(coeffieicnt of LOAN_t-1)			(4.6472)***		(3.0864)***	(4.6472)***
β4			0.2427		0.1326	0.2427
(coeffieicnt of GDP_t-1)			(4.7599)***		(2.8313)***	(4.7599)***
MA_t-1				0.5877		
				(7.3858)***		
MA_t-2				0.6675		
				(9.1995)***		
MA_t-3				0.6198		
				(7.8023)***		
constant		0.1158	-1.6602	3.7465		-1.6601
		(0.6947)	(-4.0235)***	(1.5705)		(-4.0234)***
EC1_t-1(error correction term)						-0.5251
						(-6.2097)***
Number of lags		1	1	AR=1, MA=3	1	1
Centered R^2	.	0.9405	0.8701	0.9711	.	.
Uncerntered R^2	.	0.9604	0.9318	0.9808	.	.

Notes: (i) The figures in the brackets are the t-statistic for the estimated coefficient. (ii)*: 10% significant level; **: 5% significant level; ***: 1% significant level. (iii) In VAR and VECM column, the reported figures are especially for the equation that to regress CPI on other endogenous variables.

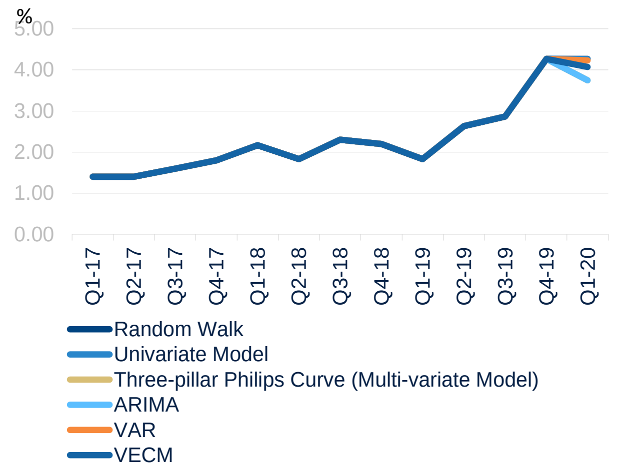
To summarize the results for different models' forecasts, we provide the prediction of 2020 Q1 CPI as an example by different models in Figure 2. We could observe that these models' forecasting range of 2020 Q1 CPI growth rate is around 3.75% to 4.27%. (Figure 2)

Figure 1. **THE UNDERLYING DATA STRUCTURE OF ENDOGENOUS VARIABLES**



Source: BBVA Research and RATS

Figure 2. **FORECASTING RESULTS OF DIFFERENT MODELS**



Source: BBVA Research and RATS

We also display the results in the following Table 2 for the comparison of all linear forecasting models together with VAR and VECM based on the criteria of the RMSE, MAE and Theil-U which capture the in-sample forecasting error of each model from different perspectives. This helps to analyze the models' forecasting ability by different criteria.

Table 2. **MODEL COMPARISON BY DIFFERENT CRITERIA**

	eg. Forecast of Q1 2020 (%)	RMSE	MAE	Theil-U
Random Walk	4.27	1.44	0.98	1
Univariate Model	4.26	0.97	0.72	0.98
Two-pillar Philips Curve (Multi-variate Model)	3.75	0.751	0.56	0.79
ARIMA	3.75	3.09	2.56	3.13
VAR	4.23	0.82	0.62	N.A
VECM	4.07	0.754	0.57	N.A

Notes: (i) The figures in the brackets are the t-statistic for the estimated coefficient. (ii)*: 10% significant level; **: 5% significant level; ***: 1% significant level. (iii) In VAR and VECM column, the reported figures are especially for the equation that to regress CPI on other endogenous variables.

The findings in the table above show that the rank of the models' forecasting performance is:

Two-pillar Philips Curve (Multi-variate model) $\succ$ *VECM* $\succ$ *VAR* $\succ$ *Univariate model* $\succ$ *random walk model* $\succ$ *ARIMA*

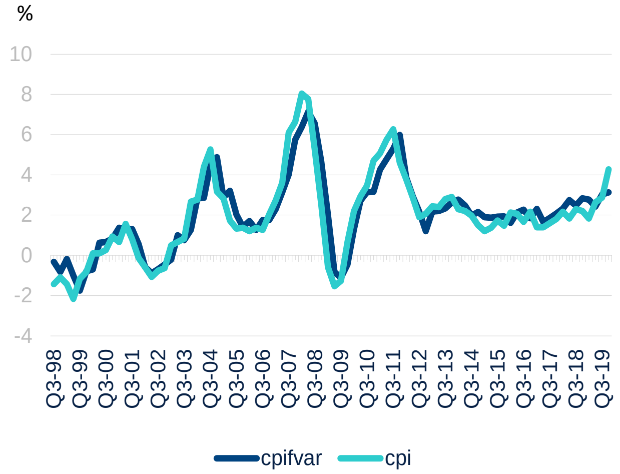
Our model comparison results could be summarized as follows: (i) Actually, many research papers find that the complicated forecasting models, either by adding more endogenous variables or by adding the structural forms of the model cannot significantly beat the random walk model. This phenomenon not only exists in the inflation forecasting practice, but also in various asset classes, such as stock and FX. (Kilian, L, and M. Taylor, 2001; Ca'Zorzi, M., J Muck and M.Rubaszek, 2015, etc.) In particular, the findings of Stock and Watson's (2008) literature review suggests that inflation is hard to forecast in the U.S. and other OECD economies and that it is difficult to improve upon simple benchmark models. Arratibel, Kamps, Leiner-Killinger (2009) also find this applying the inflation forecasting models in the new members of EU countries. Unlike many research papers in the literature that indicate forecasting models cannot be significantly found to beat the naive random walk model for other countries or regions' inflation prediction (Stock and Watson, 2007 and 2008; Arratibel, Kamps, Leiner-Killinger, 2009, etc.), our finding shows that in China's CPI inflation forecasting case, the multi-variate model, VECM, VAR and univariate model, by including more information into the forecasting model, especially the monetary and economic variables, actually have better performance than the naive random walk model. (ii) Among these four models, multi-variate model performs best. (iii) In addition, the VECM model, by correcting the partially trending structure of the data, could out-perform the basic VAR model from the perspective of both RMSE and MAE criteria.

Regarding our two structural models, the VECM model and VAR, there are more details to display and to discuss here.

First, as the test prior to VECM, the Dickey-Fuller Unit Root Test shows that, in each case for four of our endogenous variables, the null hypothesis of non-stationarity cannot be rejected at any reasonable level of significance. Actually, in Figure 1, we could find that our variables indeed have some spike or trend before 1999, although the series after 1999 look like more stationary.

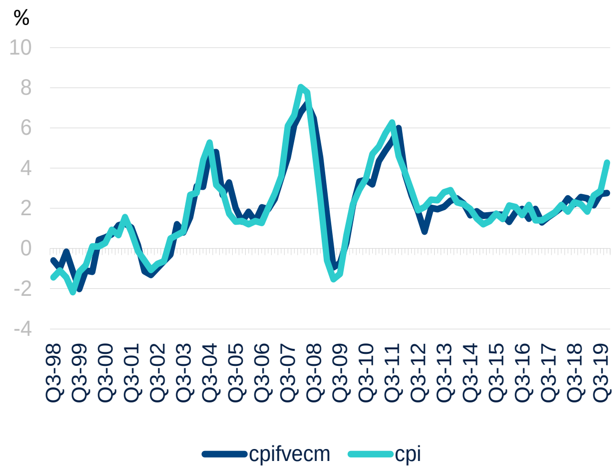
Second, both VAR and VECM could provide good in-sample forecasting performance, while VECM, by correcting the trending structure of the data, outperforms VAR in terms of the RMSE and MAE criteria. Figure 3 and 4 show the in-sample forecasted CPI and the actual CPI by VAR and VECM separately. (Figure 3 and 4).

Figure 3. **VAR IN-SAMPLE FORECASTING: PREDICTED CPI VS. ACTUAL CPI**



Source: BBVA Research and RATS

Figure 4. **VECM IN-SAMPLE FORECASTING: PREDICTED CPI VS. ACTUAL CPI**



Source: BBVA Research and RATS

Third, regarding the out-of-sample forecasting of VAR and VECM, below is the 4-step VAR and VECM out-of-sample forecasting for all the endogenous variables. It seems like both VAR and VECM predict a gradual declining CPI in 2020. Actually, this matches our own forecast that in the second half of 2020, due to the very high base effect of 2H 2019, CPI will gradually decline.

Table 3. **VAR 4-STEP OUT-OF-SAMPLE FORECAST**

Entry	GDP	CPI	M2	LOAN
2020:01	5.456306220	4.226418860	8.302824160	11.40030532
2020:02	4.945848157	4.035832671	8.120899761	10.60386813
2020:03	4.486943141	3.759956474	7.901441342	9.95472847
2020:04	4.087742352	3.444056383	7.658759242	9.42121920

Table 4. **VECM 4-STEP MODEL COMPARISON BY DIFFERENT CRITERIA**

Entry	GDP	CPI	M2	LOAN
2020:01	6.029025762	4.067092301	8.453758163	12.62293232
2020:02	6.048175465	3.933224244	8.469432577	12.78320656
2020:03	6.060809454	3.844904990	8.479773748	12.88894726
2020:04	6.069144709	3.786636493	8.486596321	12.95870953

Conclusion

Chinese inflation has been rising significantly since 2H 2019 due to the outbreak of African Swine Flu (ASF) in mainland China. This adverse supply shock is expected to last for several years. Thus, how long and how high of the inflation will last is an important policy question to Chinese authorities. Beyond the ASF, the global spread of COVD-19, together with global oil price dipping also add more uncertainties for China's inflation forecasting practice.

Against this background, this paper is trying to compare different typical econometric models' forecasting power for China's inflation. Six models are discussed which are standard and widely used in the forecasting literature: random walk model, univariate model, multi-variate (two-pillar Philips Curve) model, ARIMA, VAR and VECM. We presented and discussed model estimation and forecasting results, as well as compared these models' performance based on some standard criteria, such as RMSE, MAE and Theil-U statistic.

Our findings suggest that unlike many research papers in the literature that indicate complicated forecasting models cannot be significantly found to beat the naive random walk model, our finding shows that in China's CPI inflation forecasting case, the multi-variate model, VECM, VAR and univariate model, by including more information into the forecasting model (the monetary and economic indicators and some structural forms), actually have better performance than the naive random walk model. Among the six models, multi-variate model performs best, with VECM following it. That means, the VECM model, by correcting the partially trending structure of the data, could out-perform the basic reduced-form VAR model from the perspective of both RMSE and MAE criteria.

These findings indicate that monetary and economic indicators are found to contain useful information for predicting inflation both in-sample and out-of-sample. This result seems to echo what Friedman (1963) indicated, "Inflation is always and everywhere a monetary phenomenon". In addition, adding structural forms, together with error-correction term in VECM, could improve models' forecasting performance for Chinese inflation.

Reference:

- Arratibel, O., C. Kamps and N. Leier-Killinger (2009), “Inflation forecasting in the new EU member states”, European Central Bank working paper series, No. 1015, Feb 2009
- Box, G.E.P. and G.M. Jenkins (1970), “Time series analysis: Forecasting and control”, San Francisco: Holden-Day.
- Carstensen, K. (2007), “Is Core Money Growth a Good and Stable Inflation Predictor in the Euro Area?” Kiel Working Paper 1318.
- Ca’Zorzi, M., J Muck and M.Rubaszek, (2015), “Real Exchange Rate Forecasting and PPP: This Time the Random Walk Loses”, Federal Reserve Bank of Dallas Globalization and Monetary Policy Institute Working Paper No. 229, March 2015.
- Friedman, M. (1963), “Inflation: Causes and Consequences”, Asia Publishing House, New York
- Gerlach, S. (2003), “The ECB’s Two Pillars”, CEPR Discussion Paper 3689.
- Gerlach, S. (2004), “The Two Pillars of the European Central Bank”, Economic Policy 40: 389-439.
- Hofmann, B. (2008), “Do Monetary Indicators Lead Euro Area Inflation?” ECB Working Paper 867.
- Jarrett, J (1990), “Business Forecasting Methods”, Second Edition, Basil Blackwell (1990).
- Kilian, L, and M. Taylor (2001), “Why is it so difficult to beat the random walk forecast of exchange rates??”, European Central Bank working paper, No. 88, Nov. 2001
- Leung C., Fred Y. Kwan and Jinyue Dong (2015), “Comparing Consumption-based Asset Pricing Models: The Case of an Asian City”, Journal of Housing Economics, Vol. 28, June 2015, 18-41.
- Nicoletti-Altimari, S. (2001), “Does Money Lead Inflation in the Euro Area?”, ECB Working Paper 63.
- Stock, J., and M. Watson (1999), “Forecasting Inflation”, Journal of Monetary Economics 44: 293-335.
- Stock, J., and M. Watson (2007), “Why Has U.S. Inflation Become Harder to Forecast?” Journal of Money, Credit and Banking 39: 3-33.
- Stock, J., and M. Watson (2008), “Phillips Curve Inflation Forecasts”, NBER Working Paper 14322, Cambridge, MA.

Working Papers

2020

20/05 **Jinyue Dong:** Forecasting modeling for China's inflation.

20/04 **Le Xia:** Lessons from China's past banking bailouts.

20/03 **Ángel de la Fuente y Pep Ruiz:** Series largas de VAB y empleo regional por sectores, 1955-2018. RegData_Sect FEDEA-BBVA (v5.0_1955-2018).

20/02 **Luis Antonio Espinosa y Juan José Li Ng:**

ESP/ El riesgo del sargazo para la economía y turismo de Quintana Roo y México.

ING / The risk of sargassum to the economy and tourism of Quintana Roo and Mexico.

20/01 **Ángel de la Fuente:** La dinámica territorial de la renta en España, 1955-2018. Los determinantes directos de la renta relativa: productividad, ocupación y demografía.

2019

19/13 **Filip Blazheski and Adrian Casillas:** ripping over the edge of the cliff - business and credit cycles, financial instability and recession triggers.

19/12 **Ángel de la Fuente:** La evolución de la financiación de las comunidades autónomas de régimen común, 2002-2017.

19/11 **Alfonso Arellano, Noelia Cámara and Diana Mejía:**

ING / Disentangling Vulnerability through Consumer Behavior: The Role of Financial Health.

ESP / Desentrañar la vulnerabilidad mediante el comportamiento del consumidor: el papel de la salud financiera.

19/10 **Ángel de la Fuente:** La liquidación de 2017 del sistema de financiación de las comunidades autónomas de régimen común.

19/09 **Guillermo Jr. Cárdenas, Luis Antonio Espinosa, Juan José Li Ng y Carlos Serrano:** La crisis por escasez de gasolina en México: un análisis de Big Data.

19/08 **Ángel de la Fuente:** Series largas de algunos agregados económicos y demográficos regionales: Actualización de RegData hasta 2018.

19/07 **Ángel de la Fuente**: La dinámica territorial de la renta en España, 1955-2016: una primera aproximación.

19/06 **Ángel de la Fuente:** Las finanzas autonómicas en 2018 y entre 2003 y 2018.

19/05 **J. E. Boscá, R. Doménech, J. Ferri and J. Rubio-Ramírez:** Macroeconomic Effects of Taxes on Banking.

19/04 **Santiago Fernández de Lis and Olga Gouveia:** Central Bank digital currencies: features, options, pros and cons.

Creating Opportunities

19/03 **Santiago Fernández de Lis y Olga Gouveia:** Monedas digitales emitidas por bancos centrales: características, opciones, ventajas y desventajas.

19/02 **Santiago Fernández de Lis and Pablo Urbiola Ortún:** Digital transformation and competition in the financial sector.

19/01 **Santiago Fernández de Lis y Pablo Urbiola Ortún:** Transformación digital y competencia en el sector financiero.

CLICK HERE TO ACCESS THE WORKING DOCUMENTS PUBLISHED IN
Spanish and English

BBVA Research publications can be viewed at : http://www.bbvaresearch.com

ENQUIRIES TO:
BBVA Research: Azul Street. 4. La Vela Building – 4th and 5th floor. 28050 Madrid (Spain).
Tel. +34 91 374 60 00 y +34 91 537 70 00 / Fax (+34) 91 374 30 25
bbvaresearch@bbva.com www.bbvaresearch.com