Forecasting Inflation in Russia Using Dynamic Model Averaging

Konstantin Styrin, Bank of Russia and New Economic School
styrinka@cbr.ru

In this study, I forecast CPI inflation in Russia using the method of Dynamic Model Averaging pseudo out-of-sample on historical data. This method can be viewed as an extension of Bayesian Model Averaging, where the identity of the model that generates data is allowed to change over time, as are the model parameters. DMA is shown not to produce forecasts superior to simpler benchmarks, even if a subset of individual predictors is pre-selected ‘with the benefit of hindsight’ from the full sample. The two groups of predictors that give the highest average values for the posterior inclusion probability are loans to non-financial firms and individuals, along with actual and anticipated wages.

Keywords: Bayesian model averaging, model uncertainty, econometric modelling, high-dimension model, inflation forecast

JEL Codes: C5, C53, E37


1. Introduction

In this study, I apply the method of Dynamic Model Averaging (DMA) to forecast monthly CPI inflation in Russia one through six months ahead. This method was developed by Raftery et al. (2010) and illustrated using an engineering application. Koop and Korobilis (2012) was the first paper to apply this methodology to forecasting US inflation. Koop and Korobilis (2011) use the method to forecast UK macroeconomic variables, comparing its performance with that of other data-rich models. Byrne et al. (2018) is an application of DMA to forecasting exchange rates, while Dangl and Halling (2012) apply the method to forecasting stock returns.

1 I thank my discussants Anna Pestova and Valery Charnavoki, as well as participants of the seminars at the National Research University – Higher School of Economics and at the Bank of Russia for their helpful suggestions, and Vlad Abramov and Petr Milyutin for excellent research assistance. The views expressed in the paper are solely those of the author and do not necessarily represent the official position of the Bank of Russia. The Bank of Russia is not responsible for the contents of the paper. All remaining errors are mine.
DMA assumes that the data-generating process for inflation is not known with certainty. It states instead that, with some probability, inflation data can be generated by any of $K$ alternative models, which differ in terms of the set of predictors employed. The DMA forecast on date $t$ is the weighted average of the forecasts produced by all alternative models with weights proportional to the predictive density of the respective individual model $k \in \{1, \ldots, K\}$ on date $t$. In that respect, the DMA is a generalisation of Bayesian Model Averaging (BMA), a conventional approach to dealing with model uncertainty in Bayesian econometrics (see e.g. ch. 11 in Koop (2003) for a general discussion and Geweke and Whiteman (2006) for application of the method to forecasting). Wright (2009) applies BMA to forecast inflation in the US. Unlike BMA, the DMA method allows the identity of the model that generates data to randomly change over time. DMA also allows for time-varying parameters and stochastic volatility.

I conduct a standard pseudo out-of-sample forecasting exercise for the DMA forecast and a set of benchmark forecasts on the data sample 2002M1–2017M9, using the first 60 months of observations as an initial estimation sample. My findings suggest that DMA does not produce forecasts superior to simpler benchmarks, even if a subset of individual predictors is pre-selected ‘with the benefit of hindsight’ from the full sample. The two groups of predictors that give the highest average values for the posterior inclusion probability are loans to non-financial firms and individuals, along with actual and anticipated wages.

The rest of the paper is organised as follows. Section 2 presents the methodology of the DMA and provides details of the pseudo out-of-sample forecasting exercise. Section 3 describes the data. Section 4 sets out and discusses the empirical findings, and Section 5 concludes.

2. Methodology

2.1. Dynamic Model Averaging

This description of the DMA method closely follows Koop and Korobilis (2012). By $y_t$ I denote the variable to be forecast. In this exercise, this variable is the monthly rate of Consumer Price Index (CPI) inflation. If the forecast horizon is $h$, then

$$y_{t+h} = 1200 \log \left( \frac{CPI_{t+h}}{CPI_{t+h-1}} \right),$$

i.e. $y_t$ is the annualised CPI growth rate between the dates $t + h - 1$ and $t + h$. The vector of all the potential predictors is $z_t$. It contains lagged values for monthly CPI inflation and other macroeconomic variables that potentially have some forecasting value with regard to future inflation. I consider all alternative forecasting models, each using a different subset of $z_t$ as predictors. If the dimensionality of $z_t$ is $M$, then the overall number of all subsets of $z_t$ is $K = 2^M$, as is the number of all alternative
forecasting models that can be constructed from $z_t$. There is no prior knowledge of which particular combination of predictors yields the best forecast. Furthermore, the performance of each individual forecasting model can change over time.

Suppose that model $k$ uses a subset $z_t^{(k)} \subseteq z_t$, $k = 1, \ldots, K$ as predictors. It can then be written

$$ y_t = z_t^{(k)' \theta_t^{(k)} + \epsilon_t^{(k)}, } $$(1)

where $\theta_t^{(k)}$ is a vector of time-varying unobserved parameters that follows a vector random walk process

$$ \theta_{t+1}^{(k)} = \theta_t^{(k)} + \eta_{t+1}^{(k)} $$

with the disturbance term being multivariate normal:

$$ \eta_t^{(k)} \sim \text{i.i.d. } \mathcal{N}(0, Q_t^{(k)}). $$

In the language of state-space models, equation (2) is known as the law of motion for the unobserved state $\theta_t^{(k)}$, or simply the state equation, while equation (1) is known as the measurement equation. The error term in the measurement equation (1) is assumed to be normal:

$$ \epsilon_t^{(k)} \sim \text{i.i.d. } \mathcal{N}(0, H_t^{(k)}). $$

Denote by $y^t$ the history of observations of CPI inflation up to date $t$, $y^t \equiv (y_t, y_{t-1}, \ldots, y_0)$, and by $L_t$ the identity of the model that generates data on date $t$. Conditional on model $k$ being the data-generating process (DGP) on date $t - 1$, and given the history of observations up to date $t - 1$, the vector of parameters has a multivariate normal distribution

$$ \theta_{t-1} | L_{t-1} = k, y^{t-1} \sim \mathcal{N}\left(\hat{\theta}_{t-1}^{(k)}, \Sigma_{t-1|t-1}^{(k)}\right), $$

where, with some abuse of notation,

$$ \hat{\theta}_t^{(k)} \equiv \mathbb{E}(\theta_t | L_t = k, y^t), $$

$$ \Sigma_{t|t}^{(k)} \equiv \mathbb{E}\left[(\theta_t - \hat{\theta}_t^{(k)})(\theta_t - \hat{\theta}_t^{(k)})' | L_t = k, y^t\right]. $$

Conditional on model $k$ being the DGP on the next date, date $t$, a one-step-ahead forecast of $\theta_t$ is

$$ \theta_t | L_t = k, y^{t-1} \sim \mathcal{N}\left(\hat{\theta}_{t-1}^{(k)}, \Sigma_{t|t-1}^{(k)}\right), $$
where
\[ \Sigma_{t|t-1}^{(k)} = \mathbb{E} \left[ (\theta_t - \hat{\theta}_{t-1}^{(k)})(\theta_t - \hat{\theta}_{t-1}^{(k)})' | L_t = k, y^{t-1} \right]. \]

Finally, an updated estimate for \( \theta_t \) that exploits the information available on date \( t \) and is conditional on model \( k \) being the DGP on date \( t \), i.e. the nowcast of \( \theta_t \), is
\[ \theta_t | L_t = k, y^t \sim \mathcal{N} \left( \hat{\theta}_t^{(k)}, \Sigma_{t|t}^{(k)} \right). \]

Given the law of motion for \( \theta_t^{(k)} \), equation (2), the conditional variances of the one-step-ahead forecast and nowcast, both conditional on model \( k \) being the DGP on date \( t \), are related as follows:
\[ \Sigma_{t|t-1}^{(k)} = \Sigma_{t-1|t-1}^{(k)} + Q_t^{(k)}, \]
where the covariance matrix of the disturbance term in state equation (2), \( Q_t^{(k)} \), is unobserved and therefore has to be estimated. Following Raftery et al. (2010) and Koop and Korobilis (2012), I use an approximation by assuming that
\[ \Sigma_{t|t-1}^{(k)} = \frac{1}{\lambda} \Sigma_{t-1|t-1}^{(k)}, \quad (3) \]
where the smoothing parameter \( \lambda \) is set very close to one from below, \( 0 < \lambda \leq 1 \).

An application of standard Kalman filter formulas yields an updating equation for \( \theta_t^{(k)} \) and \( \Sigma_{t|t}^{(k)} \):
\[ \hat{\theta}_t^{(k)} = \hat{\theta}_{t-1}^{(k)} + \Sigma_{t|t-1}^{(k)} z_t^{(k)} \left( H_t^{(k)} + z_t^{(k)}' \Sigma_{t|t-1}^{(k)} z_t^{(k)} \right)^{-1} (y_t - z_t^{(k)}' \hat{\theta}_{t-1}^{(k)}), \]
\[ \Sigma_{t|t}^{(k)} = \Sigma_{t|t-1}^{(k)} - \Sigma_{t|t-1}^{(k)} z_t^{(k)} \left( H_t^{(k)} + z_t^{(k)}' \Sigma_{t|t-1}^{(k)} z_t^{(k)} \right)^{-1} z_t^{(k)}' \Sigma_{t|t-1}^{(k)}. \]

The probability density of \( \theta_t \), conditional on the history of observations up to date \( t - 1 \), is the weighted average of model-specific nowcast densities:
\[ p(\theta_{t-1} | y^{t-1}) = \sum_{k=1}^{K} p(\theta_{t-1} | L_{t-1} = k, y^{t-1}) \Pr(L_{t-1} = k, y^{t-1}), \]
where
\[ p(\theta_{t-1} | L_{t-1} = k, y^{t-1}) \sim \mathcal{N} \left( \hat{\theta}_{t-1}^{(k)}, \Sigma_{t-1|t-1}^{(k)} \right). \]

I now introduce some new notation for the probability of model \( l \) being the DGP on date \( t \) conditional on the history of observations up to date \( s \), \( s \leq t \):
\[ \pi_{t|s,l} \equiv \Pr(L_t = l | y^s). \]

It follows that
\[ \Pr(L_{t-1} = k | y^{t-1}) = \pi_{t-1|t-1,k}. \]
The DMA framework assumes that the identity of the data-generating model can be different on different dates. In general, the process of switching between alternative data-generating models can be characterised by an unrestricted matrix of transition probabilities

$$ P = (p_{kl})_{k,l=1}^{K}, $$

where $p_{kl}$ is the probability of model $k$ being the DGP on date $t$ conditional on model $l$ being the DGP on date $t-1$, and $\sum_{k=1}^{K} p_{kl} = 1$. The model prediction equation can then be written

$$ \pi_{t|t-1,k} = \sum_{l=1}^{K} \pi_{t-1|t-1,l} p_{kl}. \quad (4) $$

The difficulty with the unrestricted matrix $P$ is that it introduces numerous additional parameters to be estimated. These parameters are transition probabilities $p_{kl}$. As a result, the estimation problem is difficult to handle even when the number of predictors in $z_t$ is moderate. In order to circumvent this obstacle, Raftery et al. (2010) replace equation (4) with an approximation:

$$ \pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}}{\sum_{l=1}^{K} \pi_{t-1|t-1,l}} \alpha, $$

where $\alpha$ is a forgetting factor that is set very close to one from below, $0 < \alpha \leq 1$.

Given $\pi_{t|t-1,k}$ and $y^{t-1}$, the updating equation for $\pi_{t,k}$ is

$$ \pi_{t|t,k} = \frac{\pi_{t-1|t-k} p_k(y_t|y^{t-1})}{\sum_{l=1}^{K} \pi_{t-1|t-l} p_l(y_t|y^{t-1})}, $$

where

$$ y_t|L_t = k, y^{t-1} \sim p_k(y_t|y^{t-1}) \equiv \mathcal{N}(z_t^{(k)}', \theta_t^{(k)}') + z_t^{(k)} \Sigma_{t|t-1}^{(k)} z_t^{(k)}. $$

The DMA recursive point forecast is then obtained by averaging the point forecasts of all individual models with weights equal to the conditional probabilities of the respective model being the DGP on the date for which the forecast is made:

$$ \mathbb{E}(y_t|y^{t-1}) = \sum_{k=1}^{K} \pi_{t-1|t-k} z_t^{(k)}', \hat{\theta}_t^{(k)}. \quad (5) $$

Along with the DMA forecast, I also compute the DMS point forecast, which is the point forecast of the model with the highest conditional probability of being the DGP on the date for which the forecast is made:

$$ k^* = \arg\max_k \pi_{t-1,k}, $$

$$ \mathbb{E}(y_t|y^{t-1}) = z_t^{(k^*)}', \hat{\theta}_t^{(k^*)}. $$
As should be clear, given \( H_t^{(k)}, \pi_{0|0,k}, \) and \( \theta_0^{(k)}, k = 1, \ldots, K, \) derived analytical formulas can be applied to compute forecasts with no need to run MCMC posterior simulations, which are likely to be computationally demanding and time-consuming in this case. This reduces the computational burden significantly.

Following Koop and Korobilis (2012), the variance of the error term in the measurement equation (1) is estimated as

\[
\hat{\sigma}_t^{(k)} = \kappa \hat{\sigma}_{t-1}^{(k)} + (1 - \kappa) \left( y_t - z_t^{(k)'} \hat{\theta}_t^{(k)} \right)^2,
\]

where \( \kappa \) is a decay factor, \( 0 < \kappa < 1. \)

Dangl and Halling (2012) suggest one more layer of averaging: by the forgetting factor \( \lambda \) in equation (3). A higher value of \( \lambda \) implies that the uncertainty introduced by an innovation in the error term of state equation (2), \( \eta_{t+1} \), is small. This might be true in tranquil times, but not in times of turbulence. In order to account for time variation in uncertainty caused by innovations to the vector of coefficients \( \theta_t \), Dangl and Halling (2012) allow the forgetting factor \( \lambda \) to vary over time. This is implemented as follows. They consider a grid of values for \( \lambda, \) say, \( \lambda_j \in \{0.90, 0.91, \ldots, 1.00\} \). Conditional on \( \lambda_j, j = 1, 2, \ldots, J, \) a forecast is obtained using formula (5). Finally, all forecasts \( \mathbb{E}(y_t|y_t^{-1}, \lambda_j) \) are averaged with weights equal to posterior probabilities \( \Pr(\lambda_j|y_t^{-1}) \), to yield an ultimate forecast:

\[
\mathbb{E}(y_t|y_t^{-1}) = \sum_{j=1}^{J} \Pr(\lambda_j|y_t^{-1}) \mathbb{E}(y_t|y_t^{-1}, \lambda_j).
\]

### 2.2. Pseudo out-of-sample forecasting

With the data described in the next section, I evaluate the retrospective performance of the forecast based on DMA and DMS, and compare it with the performance of a few benchmark forecasts. The benchmark forecasts that I employ are those produced by Bayesian Model Averaging and Bayesian Model Selection (Koop and Potter, 2004; Wright, 2009); the Unobserved Components – Stochastic Volatility model (Stock and Watson, 2007); Bayesian autoregression of order 2, with time-invariant and (separately) time-varying coefficients; DMA applied to a minimalist set of predictors that includes an intercept and two lags of inflation; and, finally, the ‘kitchen sink’ model with time-varying parameters.

Taken literally, the DMA considers as an individual model any possible mix of predictors from the pool, with the number varying from one (a time-varying intercept and one predictor with a time-varying coefficient) to \( M, \)
where all predictors from the pool along with the intercept are included (at least in the implementation by Koop and Korobilis, 2012). DMA then averages out these individual models with weights equal to the posterior probabilities of a respective model being the data-generating process on a given date. The overall number of individual models is $K = 2^M$. A data set containing $M = 40$ time series is quite moderate by modern standards. For example, a popular data set composed by Stock and Watson (2012) contains about 200 quarterly US macro series. If $M = 40$, i.e. five times smaller than Stock and Watson (2012)’s dataset, the number of models $K = 2^M$ is of the order of $10^{12}$. Handling such a large number of models requires considerable computational power and a prohibitive amount of time for computation on a desktop intended for research or a laptop computer with a typical technical specification. For this project, I have employed the $R$ package called eDMA (Catania and Nonejad, 2018), which employs efficient algorithms and parallel computation. If $M = 20$ and the number of time observations is 189, my ASUS laptop with Intel Core i9 12-core processor and 64 GB of RAM completes all computations for the main exercise within about ten minutes. Each extra predictor added doubles the number of models, and hence the computation time. If the number of predictors increases by 20, i.e. from $M = 20$ to $M = 40$, then the computation time increases by a factor of $2^{20} \approx 10^{6}$, approaching $1/6 \times 10^{6} \approx 19$ years, which is prohibitively high.

Several approaches which aim to keep the number of individual models manageable are known in the literature. In an engineering application, Raftery et al. (2010) employ five predictors and consider only seventeen combinations of these predictors as candidate DGPs, the selection being based on external information about the physical processes involved. In his inflation forecasting exercise using BMA, Wright (2009) considers only individual models, each containing an intercept, lagged inflation, and one predictor, thus limiting the number of models to $K = M$. Other papers impose an upper limit on the number of exogenous predictors: $M = 14$ in Koop and Korobilis (2012), and $M = 15$ in Groen et al. (2013). Both papers forecast inflation in the US. Onorante and Raftery (2016) nowcast GDP for the Euro area, employing 30 predictors. The curse of model space dimensionality is circumvented by introducing stochastic search over the space of the models. An important assumption needed for the approximation to be well-grounded is that the DGP smoothly transitions from one model to another (Catania and Nonejad, 2018). Inspection of the posterior inclusion probabilities in Onorante and Raftery (2016) and other papers, which feature occasional swings, suggests that this is unlikely to be the case. The interim conclusion seems to be that, given the present availability of computational power, the number of predictors that DMA is able to accommodate for practical purposes is unlikely to exceed 20–25.
In this study, I use a data set of 97 macro series for Russia that cover the time period from January 2002 through September 2017, 189 time observations in total. In order to make the estimation manageable, I give my DMA forecast a certain ‘benefit of hindsight’: for each forecast horizon, I pre-select a set of top-performing predictors based on the so-called ‘hard thresholding’ selection procedure (Ng, 2013). Following this method, I run 97 predictive regressions on the full sample with monthly inflation as the dependent variable and one variable at a time from the list of predictors as a regressor, in addition to an intercept. I then rank all predictors according to the absolute value of their $t$-statistic in descending order. For two different forecast horizons $h = 1, 2, \ldots, 6$, the respective rankings need not coincide. I then pick nineteen top predictors based on the absolute value of the $t$-statistic for each $h$, add the contemporaneous and lagged inflation $y_{t-1}$ to it and consider only $K = 2^{19}$, each containing an intercept, contemporaneous inflation $y_t$, and up to nineteen other predictors, the dependent variable being $y_{t+h}$. For this set of models, I then perform a pseudo out-of-sample forecasting evaluation. The whole exercise is by no means purely pseudo out-of-sample forecasting since, as I mentioned above, the pre-selection of nineteen out of 97 predictors is performed on the full sample. If it turns out that, even with the benefit of hindsight, DMA does not yield considerable improvements over simpler benchmarks (which appears to be true), then this will make the case against DMA even stronger.

The design of the (post-pre-selection part of) the recursive pseudo out-of-sample experiment is standard. Firstly, 60 time observations are reserved for initial estimation of the model. The first one-month-ahead forecast is thus made for March 2007, the two-month-ahead forecast for April 2007, etc. The last forecast is made for September 2017. Forecast horizons are $h = 1, 2, \ldots, 6$ months. After an $h$-month-ahead forecast from date $t$ for date $t + h$ is obtained, the date $t + 1$ observation is added to the estimation sample, the parameters of the forecasting model in hand are updated, and forecasts for date $t + 1 + h$ are produced from the perspective of date $t + 1$. The procedure continues until the end of the sample period is reached. In order to avoid complications, I seasonally adjust all series on the full sample and do not address the ‘ragged edge’ problem. As far as the latter is concerned, it is well known that different series are updated with different delays. Most recent readings of financial series are available immediately, whereas in the case of data collected by national statistical agencies (Rosstat in Russia) or survey companies, it might take up to two months for the data to be prepared for publication. For the purposes of simplicity, I pretend that February 2007 readings of all predictors are available when I am about to make a one-month-ahead forecast for March 2007, although, in reality, Rosstat data updated in February is released only around March 23. Seasonal adjustment on the full sample and ignorance of the ‘ragged edge’
problem are thus more ‘benefits of hindsight’ granted to my DMA forecast. Last but perhaps not least, my data set contains revised macro series that are different from historical real-time data.

2.3. Alternative forecasting models

Below is the complete list of models that I employ in this study. In addition to DMA/DMS, it also includes benchmark forecasts against which I compare the out-of-sample forecast performance of DMA/DMS.

**Dynamic Model Averaging (DMA).** Following the literature, I set $\alpha = 0.99$ and $\kappa = 0.96$, while $\lambda$ takes its values on the grid $\{0.90, 0.91, ..., 1.00\}$.

**Dynamic Model Selection (DMS).** This is the same as DMA except that, instead of averaging over all models, only one model, that with the highest posterior likelihood, is chosen for the forecast. The parametrisation is similar to that of DMA.

**Bayesian Model Averaging (BMA).** DMA can be viewed as an extension of BMA. Like DMA, BMA assumes that there is uncertainty with regard to the model that generates data, and therefore considers several candidate models. Unlike DMA, however, BMA postulates that the identity of the data-generating process is fixed but unknown and that model parameters do not change over time. BMA can be understood as a special case of DMA with $\alpha = 1$, $\kappa = 1$, and $\lambda = 1$.

**Bayesian Model Selection (BMS).** Similar to BMA except that, instead of averaging over all models, only one model, that with the highest posterior likelihood, is chosen for the forecast. The parametrisation is similar to that of BMA.

**Unobserved Components – Stochastic Volatility Model (UC-SV).** This model acquired a good reputation for forecasting inflation in the US (Stock and Watson, 2007). It states that inflation consists of persistent and transitory components:

$$ y_t = \tau_t + \eta_t, $$

$$ \tau_t = \tau_{t-1} + \epsilon_t, $$

where $\eta_t$ and $\epsilon_t$ are two i.i.d. Gaussian processes with time-varying covariance matrices, $Q_t$ and $R_t$ respectively, where

$$ \log(Q_t) = \log(Q_{t-1}) + u_t, $$

$$ \log(R_t) = \log(R_{t-1}) + u_t, $$

with $u_t$ and $v_t$ being two i.i.d. Gaussian processes. The UC-SV forecast is $\mathbb{E}_t y_{t+h} = \mathbb{E}_t \tau_t$. 
AR(2). Bayesian autoregression of order 2.

TVP-AR(2). Time-varying autoregression of order 2.

DMA-AR(2). This is a special case of the DMA model where only two lags of inflation are available as predictors.

TVP-KS. This is a ‘kitchen sink’ regression that contains all predictors from the list, with parameters allowed to change over time.

3. Data

In this study I employ monthly data covering the period from January 2002 through September 2017. The variable to be forecast is consumer price inflation. There are 97 exogenous predictors (as labelled in Koop and Korobilis, 2012, with the sense being that neither of them is a lag of the variable to be forecast), all listed and described in Table 1 (see Appendix). As additional predictors, I also include a time-varying intercept and two lags of inflation. An intercept and two lags of inflation are set to be included in each individual model within DMA.

The macroeconomic variables that I consider as potential predictors are
- domestic prices of goods and services: PPI, cargo tariffs, etc.
- money and credit: monetary aggregates and the Bank of Russia’s international reserves, credit to individuals and firms, etc.
- labour market indicators: unemployment rate, number of people in employment
- real economic activity indicators: retail sales, retail services, wholesale sales, investment, real disposable income, real wages, construction of new houses, cargo shipments, etc.
- financial market indicators: real and nominal exchange rates, stock market index, interest rates, etc.
- survey indicators: various versions (composite, manufacturing, services, etc.) of the Purchase Management Institute Index (PMI) produced by Markit, a London-based consultancy firm; the Russian Economic Barometer (REB) industry survey: current and anticipated prices for output and inputs, planned purchases of equipment, etc.
- commodity market indicators: international prices of oil, aluminium, wheat, etc.
- indicators of economic activity in systemic economies such as the US and the Euro area.

The rationale for considering non-consumer prices as predictors is straightforward. Producer prices tend to be more flexible (or less sticky) empirically. They respond faster to macroeconomic shocks than consumer prices and can therefore, in theory at least, serve as a leading indicator with respect to the latter. A rise in transportation costs leads to an increase in all
consumer prices, perhaps with some delay. Empirically, lagged CPI inflation is a good predictor of future CPI inflation. Furthermore, such a forecast, which is consistent with CPI inflation following a random walk, is very difficult to outperform with a more sophisticated model for the US and other advanced economies (Atkeson and Ohanian, 2001; Stock and Watson, 2007; Faust and Wright, 2013). Activity variables convey information about the state of aggregate demand. Unusually strong growth in economic activity may signal an ongoing rise in production costs and the resulting accumulation of inflationary pressures on prices. Exchange rates are potentially informative because of the incomplete and gradual pass-through of changes in exchange rates into the retail prices of imported consumer goods, which is documented empirically (Burstein and Gopinath, 2014). Furthermore, along with other asset prices, exchange rates are essentially forward-looking variables that should respond to shifts in the expected time path of future inflation and the anticipated response to it from monetary policy.

All variables except CPI inflation and financial market indicators (interest rates, stock market index, and exchange rates) were seasonally adjusted using the U.S. Bureau of Census X-13-ARIMA-SEATS seasonal filter as implemented in the R package seasonal (http://www.seasonal.website). The CPI inflation series was seasonally adjusted using the methodology adopted by the Bank of Russia (Sapova et al., 2018). As mentioned above, seasonal adjustment was performed on the full sample.

All predictors containing a unit root (quantities and prices in levels) were transformed to be approximately stationary by log differencing.

4. Findings

To evaluate the performance of forecasts, I employ two standard metrics: root mean squared forecast error (RMSFE) and mean absolute forecast error (MAFE). Tables 2 and 3 (see Appendix) give the RMSFE and MAFE for DMA/DMS forecasts and competitors.

Inspection of Table 2 (see Appendix) suggests that the quality of DMA/DMS is not superior to benchmark forecasts. These two models outperform their rivals only at a horizon of three months, with RMSFE equal to 5.1 percentage points (p.p.). For the forecast horizon $h = 1$ month, the best performer is the BMA method, with RMSFE = 3.5 p.p. The precision of the DMA forecast is almost as high as that of BMA, with RMSFE = 3.6 p.p. Surprisingly, for $h = 2$, the best forecast is produced by the TVP-KS model, where all predictors from the list are involved. This is unusual since, generally, kitchen sink regressions tend to produce inferior forecasts as so many parameters must be estimated. As in the previous case, the DMA/DMS forecast is roughly on par with the top performer, yielding an RMSFE equal to 4.7 and 4.8 p.p., respectively, against the 4.6 yielded
by TVP-KS. For \( h = 4, 5, 6 \) the best forecast is generated by DMA-AR(2), which lags behind DFM/DMS for \( h = 1, 2, 3 \). Its RMSFE is a consistent 5.3 p.p. against 5.4 p.p. for DMA and 5.7 to 6.0 p.p. for DMS. Overall, DMA/DMS does not demonstrate a systematic advantage over other methods. Even for \( h = 3 \), where it outdoes its rivals, its advantage is marginal: just 0.3 p.p. relative to the closest runner-up. It is worth recalling that, in this exercise, DMA/DMS is given an important benefit of hindsight: for each \( h \), the set of predictors involved are pre-selected on the full sample based on the hard thresholding method (Ng, 2013). Overall, it must be admitted that the precision of both DMA/DMS and their rivals is quite poor in absolute terms: the width of a 67% forecast interval is about 7 to 11 p.p., depending on the forecast horizon, which is far from satisfactory for practical purposes.

The rankings produced using the MAFE metrics are practically identical. DMA outperforms the alternatives only for the forecast horizon \( h = 3 \). TVP-AR(2) is the best performer for \( h = 1 \) month, TVP-KS for \( h = 2 \) months, and DMA-AR(2) for \( h = 4, 5, 6 \) months. Again, DMA/DMS is almost on par with the winner for each forecast horizon, with differences in MAFE being marginal.

Tables 4 to 9 and Figures 1 to 6 (see Appendix) summarise findings related to the informativeness of each individual pre-selected predictor.

For each pre-selected predictor, Tables 4 to 9 (see Appendix) report the following statistics: (i) the sample mean of the coefficient on this predictor, (ii) the standard deviation of the coefficient, (iii) the sample mean of the posterior probability of the inclusion of this predictor in the data-generating process, and (iv) the sample standard deviation of the inclusion probability. Depending on the forecast horizon, three to nine pre-selected predictors have a sample mean posterior inclusion probability of 0.25 or above.

As shown in Table 4 (see Appendix), for \( h = 1 \) month, the most informative predictors are the nominal effective exchange rate (f10), the output-to-input price ratio from the REB industrial survey (s36), the growth rate in credit to non-financial enterprises with maturity up to one year (f6), and the international price of wheat (f21). According to Figure 1 (see Appendix), the informativeness of the top four predictors was not uniform over the sample period. At the beginning of the sample period, the inclusion probability for each of these predictors was close to zero. The inclusion probability for credit growth spikes around 2010 and then quite soon dips below 0.30, whereas the other three predictors gain importance, with the inclusion probability remaining around 0.9 for the nominal effective exchange rate after 2009, between 0.6 and 0.9 for the price of wheat starting from 2011, and between 0.3 and 0.6 for the output-to-input price ratio in industry after mid-2013.

As Table 5 shows (see Appendix), for \( h = 2 \), the top predictors are the output-to-input price ratio (s36) and the anticipated growth rate of wages (s41) from the REB industrial survey, along with credit growth to non-financial enterprises
(f6 and f7). According to Figure 2 (see Appendix), the posterior probability of inclusion for loans, f6 and f7, tends to be high, 0.4 to 0.6, in the middle of the sample and declines toward the end of the sample. The inclusion probability of the two survey indicators, s36 and s41, instead grows over time, reaching one in the case of the output-to-input price ratio (s36).

Table 6 and Figure 3 (see Appendix) summarise the findings for $h = 3$. The most informative predictors are loans to individuals with maturity beyond one year (f9), overall loans to individuals (w4), monetary aggregate M0 (f13), and nominal wages (r24). The posterior inclusion probability for the two loan variables is moderate, below 0.4, between 2007 and 2015, but reaches 1.0 very soon afterwards. The predictive power of M0 and nominal wages is quite high, 0.6–0.7 and 0.5–0.6 respectively, before 2011 and then gradually declines.

At the horizon $h = 4$ months, as Table 7 and Figure 4 (see Appendix) illustrate, the most valuable predictors appear to be loans to non-financial enterprises and individuals with maturity beyond one year (f7 and f9 respectively), overall loans to individuals (w4), and expected wage inflation in manufacturing (s41). The posterior inclusion probability behaves quite unevenly for the three loan variables, featuring spikes and abrupt declines. The expected growth of wages is uninformative before 2010, but then steadily gains predictive power: its posterior inclusion probability fluctuates between 0.5 and 0.9 in the second half of the sample period.

Out-of-sample forecasting results for $h = 5$ are reported in Table 8 and Figure 5 (see Appendix). The top predictors are loans to non-financial enterprises with maturity of up to one year (f6), monetary aggregate M0 (f13), nominal wages (r24), and the financial health of firms in manufacturing (s45). The predictive content of loans and nominal wages decreases over time, with the inclusion probability declining from 0.4 at the beginning of the sample to 0.2 toward the end of the sample. M0 appears to be highly informative only occasionally, in 2015–2016, when its posterior inclusion probability experiences a spike to 0.6–0.7 and then drops to 0.3–0.4. The forecasting value of nominal wages also proves to be temporary: the inclusion probability is close to zero at the beginning of the sample period, grows rapidly after 2010, reaching 0.6 in 2012, and then begins a steady descent, approaching 0.1 at the end of the sample.

Finally, Table 9 and Figure 6 (see Appendix) contain forecast evaluation results for the longest forecast horizon we consider, $h = 6$ months. The leading predictors are nominal and real wages (r24 and r25), the PMI index of input prices in services (s27), and expected purchases of equipment from the REB industrial survey (s44). The posterior inclusion probability of wages fluctuates wildly over time, between 0.2 and 0.5. The PMI index remains almost uninformative until 2015, when its inclusion probability grows rapidly
from 0.1 to almost 1.0 within a couple of months and then declines gradually to 0.6 during 2017.

To summarise, two groups of predictors tend to receive high posterior weights quite frequently. The first group is loans to non-financial firms and individuals. For all values of the forecast horizon, at least one member of this group is among the four top-performing predictors. Second, wages, either actual or anticipated, are among the most informative predictors for forecast horizons of two months and longer.

The following are occasional top predictors. Monetary aggregate M0 appears to be helpful in forecasting inflation three and five months ahead. The output-to-input price ratio in manufacturing from the REB industrial survey is among the best performers at horizons of one and two months. The international price of wheat and nominal effective exchange rate demonstrate high predictivity for forecasting one month ahead. The latter might reflect a pass-through of changes in the rouble exchange rate into the retail prices of imported goods, although, as documented in the literature, the exchange rate pass-through effect is more pronounced at horizons longer than one month (Burstein and Gopinath, 2014). Finally, the financial health of firms in manufacturing is a highly informative predictor at the horizon of five months, whereas the PMI index of input prices in services along with expected purchases of equipment from the REB industrial survey are particularly informative at the horizon of six months.

Table 10 (see Appendix) breaks down the variance of the forecast error into four components (Dangl and Halling, 2012). The first component, labelled Observations, is related to the variability of data. The second, Coefficients, arises because the parameters of the model are estimated with error. The third component, labelled Model, accounts for uncertainty about the data-generating process. The final component, Time-Varying Parameters, reflects the time variability of the coefficients in DMA. The variance decomposition is reported separately for DMA and BMA for six values of the forecast horizon. Table 10 (see Appendix) suggests that the variability of the data-generating process is an important factor in forecast errors, accounting for ten to seventeen per cent of the forecast error variance. For BMA, the importance of model uncertainty is much lower, with a contribution not exceeding four per cent. This is by design, however: unlike DMA, BMA assumes that the identity of the individual model that generates data is fixed but unknown. DMA, on the contrary, explicitly allows different models to generate data at different times, with random switches between them. BMA also assumes that the parameters of the model are fixed, and therefore that the last component of the variance decomposition, Time-Varying Parameters, is always zero. For a similar reason, the contribution of the estimation error for coefficients to the forecast error variance is much smaller for BMA than for DMA.
5. Conclusion

This paper applies the Dynamic Model Averaging method to forecasting monthly CPI inflation in Russia out of sample. Unexpectedly, given the superior performance of DMA in other studies, e.g. Koop and Korobilis (2012) for US inflation, this method does not yield forecasts that would systematically outperform simpler benchmarks in the case of Russia. Two groups of predictors give the highest average values for the posterior inclusion probability. These are, first, loans to non-financial firms and individuals and, second, actual wages or anticipated wages as established by surveys. The former are likely to reflect inflation pressures originating from aggregate demand, whereas the latter are related to the cost of production. Among the best performers, there is no single variable that remains evenly informative over the full sample period 2007–2018. A common pattern is that the posterior inclusion of a predictor is high over one subperiod and moderate to low over the rest of the sample. Breakdown of the forecast error variance suggests that time variation in the identity of a model that generates data is a non-negligible source of forecast uncertainty.

Appendix is available at http://www.cbr.ru/eng/money-and-finance; dx.doi.org/10.31477/rjmf.201901.03

References


