Usage of online-prices high-frequency data
for purposes of forecasting of official inflation

Modern opportunities of collecting and processing information make it possible to use high-frequency data for solving various research problems. An example of such a data is a set of prices of online retailers, which is available at a daily frequency for a wide range of products.

According to current research, prices of online retailers can improve forecasting of official inflation (Macias, Stemasiak, Szafranek, 2022), (Hull et al., 2017). Online data is measured at a high frequency and therefore contains more information (Cavallo, Rigobon, 2016). The data is available in real time and contains information that will be reflected in the official statistics with a lag (Aparicio, Bertolotto, 2020). Also, online prices are less rigid (Bozhechkova, Evseev 2020), so gradual price changes form a price trend earlier than the official CPI.

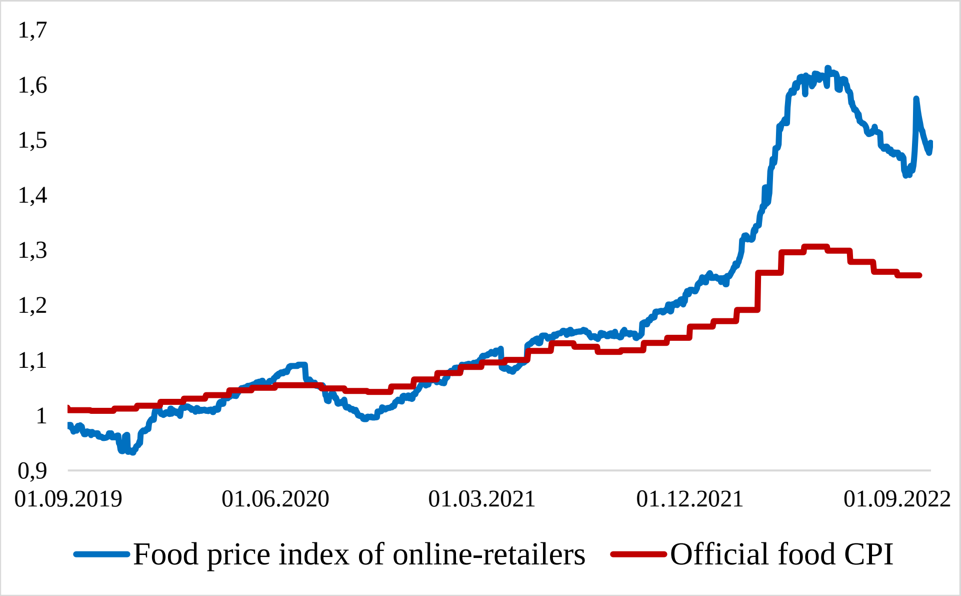
Inflation forecast models usually don’t take online price data into account, or use it in an aggregated form (for example, single-frequency VAR models). High-frequency information about prices can taken into account in in mixed-frequency data models, in particular, MIDAS models and MF-VAR models. A summary of the advantages and disadvantages of these methods is presented in Table 1.

The author is testing the hypothesis that online-prices high-frequency data (both in aggregate form and in high-frequency form) can improve the forecast of official inflation. We use the official food CPI data (in Moscow) from February 2019 to September 2022 as a dependent variable (44 monthly observations). As a variable of interest we use the food price index of Moscow online-retailers, calculated by RANEPA staff from February 2019 to present (November 2022) with a daily frequency (1360 observations at the time of calculations). The dynamics of these indicators during the period under review is shown in Figure 1.

Table 1 – Econometric methods of forecasting of official inflation using online-price data (summary)

Method	Article	Advantages of the method	Disadvantages of the method
ARMA-model	(Vicente, Pereira, 2022)	The results can be used as a benchmark for comparison with results obtained by other methods	The model doesn't contain online-price data
VAR-model	(Aparicio, Bertolotto, 2020)	The model allows to take into account aggregated online CPI data and other factors of the dependent variable. The model takes endogeneity problem into account	There is a loss of information due to aggregation of the high-frequency regressor to the level of frequency of other variables
MIDAS-model	(Chysels, Santa-Clara, Valkanov, 2004)	The model allows to use high frequency data to predict low-frequency variable without loss of information	Unconstrained model has a high risk of multicollinearity. The result of forecast is sensitive to the constraint function. The problem of endogeneity exists
MF-VAR-model	(Schorfheide, Song, 2013)	The model allows to use high frequency data to predict low-frequency variable without loss of information. The model takes endogeneity problem into account	The model may require reduction of frequency of the high-frequency variable

Souse: complied by the author.



Source: RANEPA staff data and Rosstat data.

Figure 1 – The data set of food price index of online-retailers (daily data) and official food CPI (monthly data) in 2019-2022.

Based on a summary of methods (Table 1), the author constructed the following models for the official food CPI:

A) ARMA-model (used as a benchmark):

$$p_t = \gamma + \sum_{t=1}^{p} \alpha_i p_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t, \tag{1}$$

where p_t – official inflation in period t;

p – number of lags of AR-process;

q – number of lags of MA-process.

B) SVAR-X-model with standard inflation factors (Russian ruble exchange rate; MIACR interest rate; FAO food price index is used as an exogenous variable).

C) SVAR-X-model with standard inflation factors and online price index data:

$$B_o Y_t = A_1 X_{t-1} + \cdots + A_p X_{t-p} + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + u_t, \tag{2}$$

$$\varepsilon_t = B_0^{-1} u_t, \tag{3}$$

where X_t – exogenous macroeconomic variables vector;

Y_t – endogenous macroeconomic variables vector;

p – number of lags.

Models (A) and (B) don't contain information about online prices and considered as standard inflation forecasting models; model (C) contains information about online prices in an aggregated form (monthly frequency).

D) MIDAS-models with four standard weighted functions (PDL, step function, beta function and exponential Almon function). These models (D) contain information about online prices in the original form (daily frequency):

$$A(L)\pi_t = c + B\left(L^{1/m}\right) x_t^{(m)} + \varepsilon_t\ , \tag{4}$$

where π_t – official inflation, measured with monthly frequency $\{\pi_t, t \in Z\}$;

$x_t -$ explanatory regressor (online inflation), measured with m higher frequency $\{x_t^{(m)}, t \in Z\}$.

E) MF-VAR-model. This model (E) contains information about online prices in a partially aggregated form (weekly frequency):

$$y_t = Ax_t + \varepsilon_t, \tag{5}$$

where y_t, x_t – vectors of variables, measures with monthly frequency and with weekly frequency.

As a result, the author obtains out-of-sample forecasts of the official food CPI in Moscow for 1 month and 3 months ahead. The forecasts are compared by RMSE. Figure 2 presents a summary of RMSE gained in different models.

Forecas-ting date	1 month ahead forecast			
	MIDAS group	VAR group (online)	VAR	SARIMA
Oct.21	0,0095	0,0032	0,0031	0,01
Nov.21	0,0088	0,001	0,0013	0,01
Dec.21	0,0015	0,0039	0,0028	0,01
Jan.22	0,0087	0,0084	0,0096	0,01
Feb.22	0,0061	0,0038	0,007	0,0063
Mar.22	0,0093	0,015	0,017	0,01
Apr.22	0,0402	0,0163	0,0648	0,0559
May.22	0,042	0,0484	0,0276	0,04
Jun.22	0,0223	0,0211	0,0139	0,0248
Jul.22	0,0276	0,0167	0,0171	0,0178
Aug.22	0,0109	0,0096	0,0087	0,0144
Sep.22	0,0035	0,0071	0,0024	0,011
Oct.22	0,0027	0,0034	0,0028	0,0111

Forecas-ting data	3 month ahead forecast			
	MIDAS group	VAR group (online)	VAR	SARIMA
Oct.21	0,0162	0,0206	0,0231	0,0082
Nov.21	0,0074	0,0061	0,0045	0,0129
Dec.21	0,0109	0,0074	0,0041	0,0173
Jan.22	0,0104	0,007	0,0073	0,01
Feb.22	0,0062	0,0077	0,0079	0,007
Mar.22	0,0103	0,0116	0,0146	0,0058
Apr.22	0,0222	0,0332	0,0353	0,0452
May.22	0,0306	0,0237	0,047	0,0727
Jun.22	0,0297	0,0629	0,0639	0,0757
Jul.22	0,0571	0,0371	0,0499	0,0264
Aug.22	0,0238	0,0503	0,043	0,0508
Sep.22	0,0218	0,0331	0,0371	0,0318
Oct.22	0,01	0,0098	0,0049	0,0117

Figure – RMSE of 1 month and 3months ahead forecasts gained in groups of models

Note – the forecasting date is the date on which the forecast was made. The forecasts with the largest error are marked in red, the forecasts with the smallest error are marked in blue. Source: calculated by the author.

According to the results, SARIMA-models have the largest forecast error. On the average, VAR-models which contain information about online prices have lower RMSE rather than VAR-model which doesn't use this information. MIDAS-models take into account changes in price dynamics on a daily basis, which may be helpful for forecasting in crisis conditions. MF-VAR-models take into account endogeneity of macroeconomic variables and information from high-frequency variables, also they have a better interpretability. The author concludes that inclusion of high-frequency data in the forecasting model of official inflation can improve the quality of the short-term forecast, so such a forecasting technique can be useful for the operational purposes of economic policy in Russia, particularly in crisis conditions.