SHORT-RUN FORECASTING OF CORE INFLATION IN UKRAINE: A COMBINED ARMA APPROACH

DMYTRO KRUKOVETS[a], OLESIA VERCHENKO[b]

[a]National Bank of Ukraine, Kyiv, Ukraine
E-mail: Dmytro.Krukovets@bank.gov.ua
[b]Kyiv School of Economics, Kyiv, Ukraine
E-mail: verchenko@kse.org.ua

Abstract The ability to produce high-quality inflation forecasts is crucial for modern central banks. Inflation forecasts are needed for understanding current and forthcoming inflation trends, evaluating the effectiveness of previous policy actions, making new policy decisions, and building the credibility of a central bank in the eyes of the public. This motivates a constant search for new approaches to producing inflation forecasts. This paper analyses the empirical performance of several alternative inflation forecasting models based on structural vs. data-driven approaches, as well as aggregated vs. disaggregated data. It demonstrates that a combined ARMA model with data-based dummies that uses the disaggregated core inflation data for Ukraine allows to considerably improve the quality of an inflation forecast as compared to the core structural model based on aggregated data.

JEL Codes C53, E31, E37

Keywords short-run forecasting, core inflation, ARMA, disaggregation

1. INTRODUCTION

A high-quality inflation forecast is a must-have for a central bank as it provides the foundation for many of its decisions and policy actions. Besides, an accurate forecast boosts the credibility of a central bank by enhancing its reputation as a trustworthy analytical center and a force to reckon with, which in turn could help to influence the public's expectations, which are among the fundamental determinants of economic behavior.

For this reason, many central banks develop and use a wide range of econometric models, starting from small univariate models for separate macroeconomic series, to big structural models, which contain complex relationships between various parts of the economy.

Small data-driven models can be particularly useful for short-horizon (up to six months) forecasting, due to their ability to work with a huge amount of data without the need to impose strong relationships between economic variables. On the other hand, structural models, which are frequently based on microfoundations, can serve well in describing how the economy works and how shocks are transmitted between its different parts, but can also be cumbersome and have low forecasting ability, in particular over relatively short time horizons.

Short-run inflation forecasts provide information on the dynamics of inflation in the nearest future. The data on the current level of inflation is revealed only with a lag from seven to ten days.[1] Therefore, a central bank is keen on getting constant updates on where the economy currently is, where it is heading and whether the current monetary policy strategy is still in line with the set targets.[2]

This paper focuses on data-driven inflation forecasting models. Our main model is based on the Combined ARMA (CARMA) framework developed by Huwiler and Kaufmann (2013) and currently used by the Swiss National Bank. Each of the inflation components is first modeled individually and then their forecasts are combined back together into a single core inflation estimate. The disaggregation approach allows using a rich structure of data on various inflation components. Our goal is to evaluate its performance relative to several alternative statistical models, which use both aggregated and disaggregated data, and to the NBU baseline forecasts, based on the structural Quarterly Projection Model (QPM).

[1] From Ukrstat data and reports release calendar on the official State Statistics Service of Ukraine (SSSU) website, ukrstat.gov.ua, “Express reports” section.
[2] At the same time, there is general agreement in the literature and among policymakers that monetary policy actions affect inflation with a lag of at least half a year. This coincides with the findings of Gruen et al. (1997) and Batini and Nelson (2001), who report about 4-6 quarter lag in the monetary policy effect in the US, UK and Australia. This suggests that at any point in time, inflation is already predetermined for the next 6+ months.

Our main specification contains dummies, which capture periods of excessive volatility and thus help to improve both the in-sample fit of the model and its forecasting quality.

One of the most important questions discussed in the relevant literature is whether data-driven models can outperform structural ones. While the latter are built to investigate complex links between different parts of the economy, their short-term forecasting abilities are typically quite poor (see Grui and Lepushynskyi, 2016). There is also no consensus in the literature on how the microfoundation-based (DSGE) models perform in this regard: while one part of the literature shows that such models can produce quite good forecasts (see Yau and Hueng, 2019), other authors reach the opposite conclusions (see Edge and Gurkaynak, 2010). This suggests that further comparison of alternative models for different data sets is needed to reach more definite conclusions.[3]

Numerous authors demonstrate that data-driven models can produce positive results in the context of emerging economies. Frequently, standard models are extended to reflect peculiarities of the data from these markets due to their excessive volatility, structural breaks or other non-standard data patterns.[4]

There is also no agreement in the literature about the advantages of the disaggregated (vs. aggregated) approach both from the theoretical and the empirical points of view. There are two main camps of authors: those who strongly support the effectiveness of disaggregation for improving the forecast quality, and those who oppose this view. The first camp includes, for example, Hendry and Hubrich (2011) and Zellner and Tobias (1999). Bermingham and D'Agostino (2011) also conclude that the disaggregation technique improves forecasting performance. These conclusions are based on various autoregressive-type models on the US and EU datasets.

On the other hand, Benalal et al., (2004) demonstrate that disaggregation has limited usefulness. This ambiguity in the literature indicates that further investigation of this question is required.

This study contributes to the existing literature in several ways. First, to the best of our knowledge, there is little empirical evidence on the relative forecasting performance of ARMA-based models for inflation in developing economies. Second, we suggest several specifications of dummy variables to capture periods of excessive volatility, and show that they can significantly improve the quality of the forecasting model. Third, this study is the first attempt to investigate empirically the disaggregated Ukrainian inflation data in terms of how much forecasting power it has relative to the aggregated inflation series. Therefore, this paper will contribute to the discussion on the usefulness of disaggregated vs. aggregated models by providing new empirical evidence.

In addition, we will analyze the statistical features of the inflation components, which have a heterogeneous nature. The aggregated series make these peculiarities invisible, though they can be potentially exploited to improve our understanding of the inflation dynamics and forecast.[5]

The paper is structured as follows.

- The data description section discusses the main features of the data, as well as issues related to changes in definitions and data collection methodologies.

- The methodology section describes what the models consist of, how these models are estimated, how the forecasts are produced and how they are formally compared to each other.

- The results section contains discussion on the comparative empirical performance of the models.

- The last section concludes and delineates directions for future research.

2. DATA DESCRIPTION

The data used in this paper was provided by the NBU.[6] The data contains monthly observations for core inflation components from the beginning of 2007 to the end of 2018 (144 time points in total). Core inflation is calculated based on these components of the Consumer Price Index (CPI), which have relatively low volatility, experience low influence from global prices and are not subject to administrative controls.[7]

Figure 1 presents the core inflation dynamics over the sample period. As we can see, there was a spike in inflation in March of 2015, caused by the economic crisis, which started in February and resulted in a drastic (more than threefold) devaluation of the national currency in the first quarter of 2015.

Two hundred forty components in core inflation are divided into four main categories: processed food, services, clothes and other. Processed food and clothes include most of the goods, that might be purchased in retail stores, excluding raw food (meat, fruits, vegetables), administratively regulated items (alcohol, cigarettes), and low-weight items in the basket (exotic foods, rare services).

The number of components in each category and their weights in the consumption basket are shown in Figure 2. Even though all four categories contain more or less the same number of components, their weights in the consumption

[3] There is a vast range of other tools to predict inflation: VAR and its Bayesian version, VECM, GARCH, factor models etc. Koop and Korobilis (2012) have considered a Dynamic Model Averaging approach to inflation forecasting and have shown that their forecasts are better than the Greenbook forecasts by the Federal Reserve Board of Governors. A MIDAS approach makes it possible to work with mixed-frequency data: Schorfheide and Song (2013) have shown that using dozens of macroeconomic variables on a quarterly basis, mixed with so-called “real-time,” outperforms a VAR benchmark.

[4] For example, Huwiler and Kaufmann (2013) have shown that a combination of data-driven models (Vector Error Correction Model (VECM) for oil and the disaggregated Autoregressive Moving Average (ARMA) model for other inflation components) outperform structural models and expert judgment for predicting inflation in Switzerland. Stelmasiak and Szafranski (2016) use two different Bayesian Vector AutoRegression (BVAR) approaches for inflation forecasting in Poland, paying particular attention to the issue of shifting seasonality (seasonal spikes might appear in 11 or 13 months after the previous one, and cannot be captured well by means of simple seasonal adjustment).

[5] For example, while during the sample period the total core inflation in Ukraine reached its peak in 2015m03 and the biggest contribution was from the exchange rate side (see Faryna, 2016), this was not true for every component, which suggests that the nature of the rapid increase in prices of different goods is also an interesting topic for investigation.

[6] The NBU obtains its inflation component data from the SSSU. This data is similar to the open-access data available from the SSSU website, www.ukrstat.gov.ua, but is more detailed and disaggregated.

[7] The other constituents of the CPI are raw food, energy and administratively regulated prices.

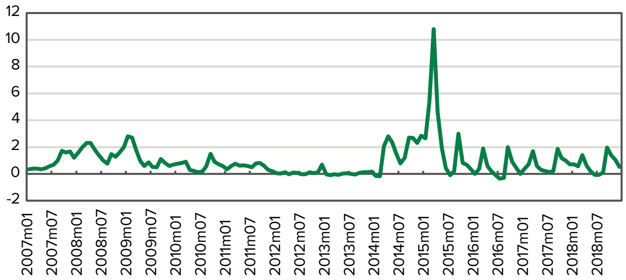

Figure 1. Core Inflation, monthly changes.

basket are quite different: the weight of the food category is much higher. This is consistent with the data from other emerging markets, where people tend to spend higher shares of their income on food rather than other goods.

Figures 3 to 6 visualize the most commonly encountered data patterns and present inflation dynamics for selected components and categories of core inflation. In particular, as Figure 3 demonstrates, component 31 (sausages) has relatively uniform dynamics over the entire data period (without much seasonality, spikes or drops), while in Figure 4, component 301 (higher education) exhibits many very distinct movements that occurred in September.

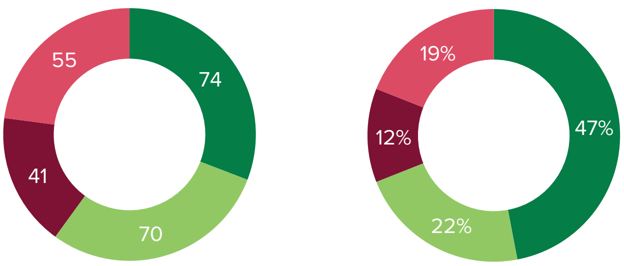

Figure 2. Left: the Number of Series in Each Category.
Right: Relative Weight of Categories in the Core Inflation Basket.

A similar conclusion can be drawn for category 5 (food) and category 7 (clothes) in Figures 5 and 6 respectively. The former has a much more distinct seasonality pattern in the earlier periods than in more recent ones, while the latter exhibits a strong seasonality pattern after 2014, which was not observed in earlier periods. This can be attributed to the changes in the data collection methodology.[8]

Not all 240 components have recorded prices starting from 2007 due to changes in CPI methodology. Seven components have data starting only from 2016. These series are too short to produce any meaningful coefficient estimates and therefore have been dropped from the sample.[9] There are also 32 series that start in 2012, which have enough observations for model estimations.[10] The resulting sample includes 31,632 observations for 233 components.

Table 1 contains a basic statistical description of the aggregated core inflation series, as well as pooled component data (233 series pooled together). The last two columns of the table contain a summary of individual component means and their standard deviations to shed some light on the differences in the dynamics of various components.

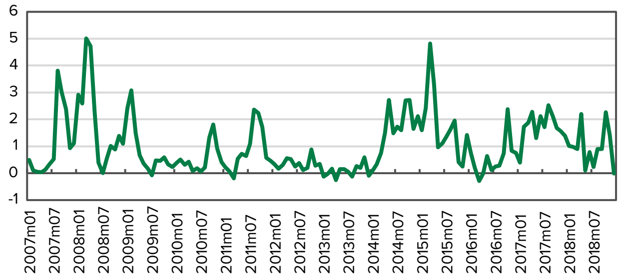

Figure 3. Monthly Inflation for Component #31 – Sausages.

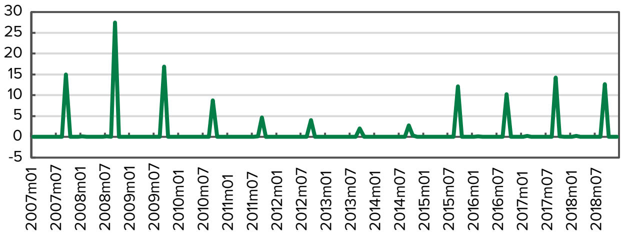

Figure 4. Monthly Inflation for Component #301 - Higher Education.

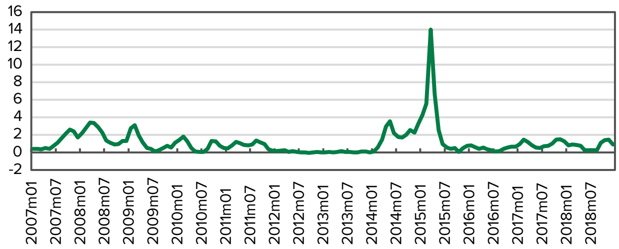

Figure 5. Monthly Inflation for Category #5 - Food.

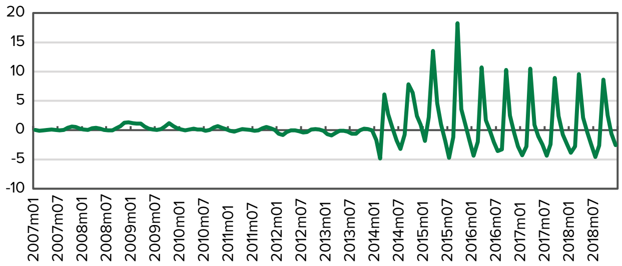

Figure 6. Monthly Inflation for Category #7 – Clothes.

The table suggests that the unweighted average inflation of all the components is around 0.9% per month, the series of means are expectedly much less volatile than the pooled data, and the standard deviation of the pooled data is more than twice as high as the pooled component inflation. This indicates that there is a lot of variability in individual components. Also, the mean of the pooled series is much higher than its median. This suggests that the inflation levels of individual components are typically quite low, and the average statistics are driven by relatively infrequent large shocks, which most likely happened during the crisis period of 2015.

Since we work with monthly data, there is a visible seasonality pattern in many of them, including the core

[8] In particular, starting from 2014, the prices of clothes are recorded with seasonal sale discounts, while such discounts were not included in the official statistics in previous years.
[9] These observations constitute 0.7% of the entire sample of data. The total weight of these series in the core inflation basket is around 2%.
[10] The total weight of these series in the core inflation basket is 12.6%.

Table 1. Descriptive Statistics for Core Inflation and its Components.

	Core inflation	Pooled component series	Means of components	Standard deviations of components
Mean	0.93	0.88	0.88	2.20
Standard deviation	1.25	2.55	0.32	1.30
Minimum	-0.36	-22.08	-0.08	0.40
Median	0.60	0.40	0.93	1.81
Maximum	10.80	46.26	1.75	6.89
Observations	144	31,632	233	233

inflation itself (see Figure 1). This seasonality will be taken care of by including 12 to 13 seasonal dummies in the models.[11]

For some series, such as clothes (see Figure 6), the seasonality has become much more pronounced starting from 2014.[12] To deal with this structural break in the data, we have evaluated all model coefficients for the clothes components in using the post-break period only.

To produce an aggregated inflation estimate, these disaggregated components must be combined into categories and then into one total core inflation indicator. To do this, weights should be assigned to each of them. There are official weights that are used by the SSSU to calculate the core inflation. However, these weights change constantly and are not known in advance. Our approach is to use a set of weights, produced by NBU statisticians, for internal inflation estimation and forecasting purposes. These weights are updated on a much less frequent basis that the SSSU weight and they track the latter closely. Therefore, in our forecasting exercise, we use the most recently available values of these "static" weights.

To investigate how important is the resulting "aggregation bias" due to differences in official and static weights, we have plotted actual core inflation and constructed core inflation (based on static weights) in Figure 7. The differences between these two series in most cases are quite small[13], especially in the most recent period, and since we use the same weights for all models, the relevance of our general conclusions should not be affected by the weights error issue.

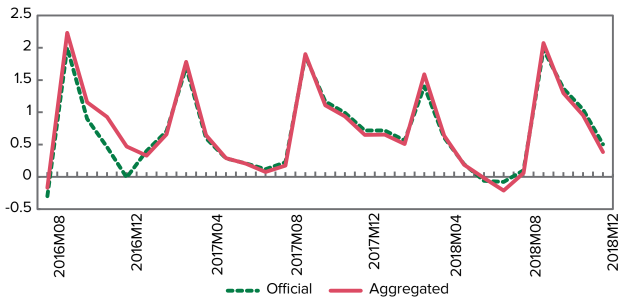

Figure 7. Official vs Aggregated Core Inflation, monthly.

3. METHODOLOGY

The empirical methodology of this paper is based on three core elements:

- The use of the disaggregated inflation component series;
- ARMA modelling framework;
- Dummies to capture periods with unusually large shocks.

The key feature of our approach is the use of disaggregated series, which means that instead of direct core inflation forecasting, its components are predicted first and then reaggregated back into core inflation. This allows for using all available information on individual inflation components. Also, it captures co-movements of components, which are due to the complementarity and the substitution effects.

The predicted core inflation $\hat{y}$ in period τ is calculated as:

$$\hat{y}_\tau = \sum_{k=1}^{p} w_k * \hat{y}_\tau^k, \quad (1)$$

where k is the index of a component, w_k is its weight in the basket, p is the total number of components, and $\hat{y}_\tau^k$ is the forecasted inflation of component k for period τ.

Equation (1) is generally referred to as the CARMA model in the results section of this paper.

To forecast individual inflation components (and core inflation itself as one of the benchmarks in performance evaluation exercises), ARMA-type models are employed. These models are widely used in modelling time series data since many economic variables strongly depend on their previous values.

The classical ARMA model has the following structure:

$$\hat{y}_t = \beta_0 + \sum_{i=1}^{m} \beta_i * y_{t-i} + \sum_{j=1}^{n} \gamma_j * \varepsilon_{t-j}, \quad (2)$$

[11] The thirteenth lag allows for capturing a floating seasonal pattern, such as a shifting harvest.
[12] As mentioned in footnote 5, before 2014 it was common to observe hikes in reported prices just before sales started, so the actual changes in consumer prices could be lower than indicated in the sales price. After 2014, the new methodology with the inclusion of discounts brought a visible seasonality pattern to inflation, with the source being mostly in the clothes category.
[13] The root mean squared error (RMSE) between the two series is about 0.09, which is less than 1/10th of the average core inflation in the sample period.

where y_t is the value of a component/core inflation in period *t*; β_0 is a slope coefficient; β_i's and γ_j's are the coefficients corresponding to autoregressive and moving-average factors respectively; and ε_{t-j} is the model residual in period *t-j*.

We identify the number of AR and MA terms for each series using the Schwarz (Bayesian) Information Criterion (Schwartz, 1978). The classical ARMA model is extended by adding dummy variables to account for excessive market movements. An ARMA model with a dummy has the following structure:

$$\hat{y}_t = \beta_0 + \sum_{i=1}^{m} \beta_i * y_{t-i} + \sum_{j=1}^{n} \gamma_j * \varepsilon_{t-j} + \alpha * D_t, \tag{3}$$

where D_t is a dummy variable.

Once a dummy variable is added to a classical ARMA model, it essentially turns into an ARMAX (ARMA with exogenous variables) model. Kongcharoen and Kruangpradit (2013) used such a model to forecast exports in Thailand. Their results show that an ARMAX-type model significantly outperforms a simple ARMA approach in most exercises. Bos, Franses and Ooms (2001) also demonstrated, using ARMAX models, superior results in forecasting the post-war core inflation in the US.

This paper uses two alternative approaches to defining dummy variables: the non-zero dummies are assigned to 1) periods in which component inflation levels have the highest deviations from their means, or 2) periods in which no-dummy model errors are the highest.

To illustrate the importance of the first dummy type, let's assume that the data contains a single, but big shock at some point in time. With the quadratic optimization function, the outlier will have a strong impact on coefficients and, therefore, predicted values. The dummy captures these spikes and prevents systemic shifts in forecasts, smoothing out the effect of the outliers.

On the other hand, a data series might have a predictably volatile structure, for example, if there is a strong seasonal pattern. At the same time, there might also be some other truly unpredictable large shocks ("extreme events"), the effect of which can be quite distortive but requires a different treatment than the one offered above. The residual-based approach to a dummy is better suited in handling such a situation.

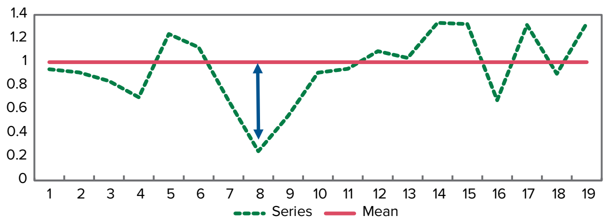

Figure 8. Example of "Deviation from the Mean" Dummy.

We have considered five possible sub-definitions for both types of dummies: the dummy takes the value of one when *the highest or two highest* or three highest deviations from the mean are observed (see Figure 8), or the dummy takes the value of one whenever an observation is located further from the mean than *three* or *four* standard deviations (see Figure 9). The first three definitions work best for cases in which there are very few strong spikes in the data (e.g., the effect of crisis). However, if the spikes are much more common, this approach will be powerless in improving the models' fit to the data.

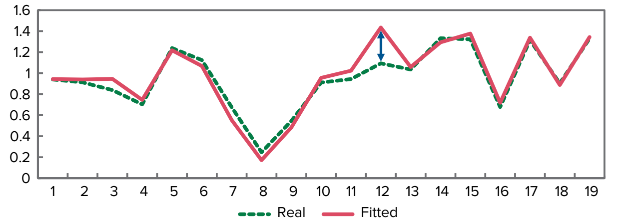

Figure 9. Example of "Deviation in Residuals" Dummy.

To illustrate the implications of the first and the third sub-definition for a dummy (i.e. the dummy takes the value of one when *the highest* deviation or *three highest* deviations from the mean are observed), Figures 10 and 11 plot the number of non-zero values for respective dummies for all inflation components. As we can see, the most turbulent period is March-April 2015, when many inflation components exhibit extremely high deviations from their means.

The last two sub-definitions (i.e., the dummy takes the value of one whenever an observation is located further from the mean than *three* or *four* standard deviations) allow different series to have a different number of associated non-zero dummy values. Series characterized by occasional spikes are treated differently from series with no big spikes. Therefore, this approach is more flexible.

To identify which dummy works best for each series, we once again calculate SIC coefficients for each definition of a dummy and choose the specification with the lowest value of the criterion.[14]

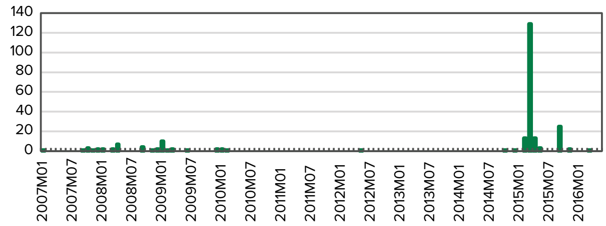

Figure 10. The Number of Inflation Components with Non-Zero Dummies of the First Type (the dummy is equal to one when the data point corresponds to the highest deviation from the component's mean).

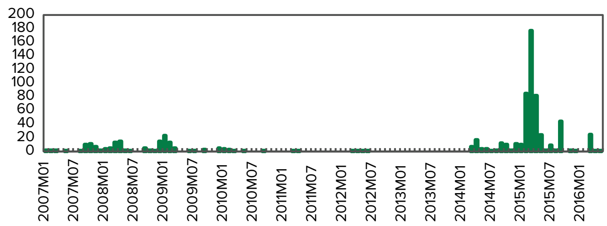

Figure 11. The Number of Inflation Components with Non-Zero Dummies of the Third Type (the dummy is equal to one when the data point corresponds to the highest, the second-highest or the third-highest deviation from the component's mean).

[14] In theory, a more appropriate approach to selecting the best model specification is to consider all possible combinations of AR/MA lags and dummy definitions and then choose the one with the lowest SIC. However, this requires considerable computational power, which the authors currently have no access to.

When building forecasts, we assume that the dummy variables for the forecasted periods are all equal to zero (no abnormal shocks).

To evaluate the forecasting performance of alternative model specifications, we calculate pseudo out-of-sample rolling-window forecasts for each of them and then construct two summary statistics for these forecasts: 1) their RMSEs, and 2) Diebold-Mariano-West (DMW) statistics for the relative forecasting performance test.

Overall, each model produces 19 forecasts starting from 2017m1. We chose this starting point for the forecasting exercise since it allows to focus on a relatively calm period (at least one year after the crisis of 2015), which is consistent with setting the predicted values for dummy variables to zeroes.

The Diebold-Mariano-West test (Diebold and Mariano, 1995, and West, 1996) is a classical test for this. It determines whether the difference between forecast errors (for different forecasts) is significant. The algorithm calculates the quadratic (to be consistent with RMSE) difference between predicted and actual values.

This test suffers strongly if the forecast horizon is small, which is the case for this exercise. It tends to give high p-values and does not reject the hypothesis about forecasts' similarity. Therefore, if the results are not significantly different, it says little about the real relationship between two predictions. However, a positive result is evidence of the very strong diversity.

The model proposed in this paper aims to enhance the forecasting toolbox of the NBU, so we consider the NBU's official inflation forecasts in 2017-2018 as a benchmark.[15] These forecasts are made public only on a quarterly basis; however, monthly forecasts are also generated for internal use, and they were made available to us to be used within this study.[16] The official forecasts may incorporate inputs from various models and expert judgements but, generally, are based on simulations of the NBU's core Quarterly Projection Model (QPM).

The QPM is a semi-structural New Keynesian small open economy model[17], in which different parts of the economy are connected via a so-called transmission mechanism. The model is widely used for explanatory purposes, policy analysis and medium-term forecasting.[18]

4. RESULTS

As explained in the methodology section, overall we estimate 33 models that produce forecasts: 11 ARMA models for core inflation (one without dummies and 10 for alternative dummy specifications), 11 CARMA models for categories, and 11 CARMA models for components. The rolling-window, one- to six-month forecasts produced by these models are then compared to benchmark forecasts, which come from the NBU baseline model.

Table 2 reports root mean squared prediction errors (RMSPEs) for 10 selected models: the benchmark model (NBU), three no-dummy models (one for each level of disaggregation), two best models with dummies for aggregated core inflation (one from the cohort of five mean-based dummy specifications and one from the cohort of five residual-based specifications), two best models for disaggregated category-level data, and two best models for disaggregated component-level data.[19] The criterion for choosing the "best" models for each of the cohorts was the lowest RMSE-based in-sample fit to the training data (the part of the sample used to estimate model parameters).

The table suggests that component-based CARMA models have the lowest RMSEs among all the models, and adding dummies helps to reduce the forecasting errors considerably. This suggests that the disaggregation approach is indeed effective in terms of increasing forecasting accuracy, and the precision level increases with the level of disaggregation. Interestingly, the semi-structural model shows lower RMSEs than the data-driven ARMA model for core inflation (both with and without dummies). Therefore, it is the disaggregation feature of CARMA, which more than compensates for the drop in performance of the aggregated statistical vs. structural model. Also, the disaggregated models with dummies are the only models that consistently produce lower predicted RMSEs for all forecasting horizons and all dummy specifications, while other models outperform the NBU forecast for only some of the horizons (see Appendix A).

The next step is to formally test for the difference in the forecasting performance of the models. Table 3 contains p-values of the DMW test for model forecasting abilities against the benchmark (NBU model).

The table shows that in all but a few cases the p-values of the test are quite high (above 10%). Formally, this indicates that there are no significant differences between the benchmarks and selected model forecasts. However, since there are only 19 observations in the sample of forecasts, the power of the test is expected to be quite low. Still, there is evidence that the components-based CARMA with dummy produces better forecasts for short horizons (one month ahead) than the semi-structural model.[20]

Overall, taking into account data limitations and therefore expectedly low power of tests, we believe that these results support the claim that the disaggregated data analysis can considerably improve inflation forecasting.

[15] Another benchmark that we considered was a random walk model. However, its performance was so poor that we decided to exclude it from the paper entirely.

[16] Another option is to transform the results of other models from monthly to quarterly, but then we encounter the problem of an extremely low number of observations (about six in total).

[17] The four main equations in the model are Aggregate Demand, Price Phillips curve, Hybrid Uncovered Interest Rate Parity and the Monetary Policy rule. Equations are given in gaps form, built via the Kalman filter. All coefficients are calibrated to incorporate expert judgements on the reaction of the Ukrainian economy to shocks, and to be consistent with other similar models for world economies. Monetary policy and the economy are linked through the interest rate and exchange rate transmission channels.

[18] More details about the model architecture, methodology, data, calibration, analysis and forecasting procedures might be found in Grui and Vdovychenko (2019).

[19] Appendix A contains the results for all considered model specifications.

[20] In addition to the DMW test, we also followed Diebold and Mariano (1995) and did sign and Wilcoxon small-sample tests. The results are similar to the ones presented in Table 3. However, we have also found some, albeit weak, evidence that disaggregated models with dummies produce better long-term (five and six month ahead) forecasts that the benchmark.

Table 2. RMSPEs of Forecasts for Selected Models.

Forecast horizon (months ahead)	Benchmarks				CARMA without dummies		CARMA with dummies			
	NBU model	ARMA for a total core	ARMA for the total core with dummy		Compo-nents	Cate-gories	Components		Categories	
			1 highest, mean	2 highest, mean			1 highest, mean	3 highest, residuals	2 highest, mean	2 highest, residuals
1	0.329	0.332	0.337	0.334	0.219	0.249	0.180	0.201	0.228	0.229
2	0.394	0.448	0.450	0.436	0.302	0.340	0.241	0.245	0.310	0.319
3	0.365	0.515	0.500	0.503	0.337	0.409	0.253	0.261	0.360	0.369
4	0.370	0.521	0.505	0.520	0.349	0.445	0.275	0.269	0.391	0.402
5	0.429	0.507	0.493	0.518	0.342	0.439	0.276	0.263	0.393	0.403
6	0.444	0.495	0.481	0.500	0.334	0.414	0.263	0.254	0.376	0.388

Table 3. DMW Test for Different Models Compared to the NBU Semi-Structural Model Benchmark, p-value.

Forecast horizon (months ahead)	Benchmarks			CARMA without dummies		CARMA with dummies			
	ARMA for a total core	ARMA for a total core with dummy		Compo-nents	Cate-gories	Components		Categories	
		1 highest, mean	2 highest, mean			1 highest, mean	3 highest, residuals	2 highest, mean	2 highest, residuals
1	0.11	0.82	0.80	0.10	0.13	0.06	0.09	0.11	0.10
2	0.36	0.97	0.96	0.30	0.39	0.16	0.20	0.25	0.31
3	0.53	0.88	0.89	0.40	0.59	0.13	0.21	0.40	0.39
4	0.60	0.77	0.83	0.39	0.65	0.10	0.16	0.50	0.48
5	0.57	0.71	0.79	0.39	0.57	0.15	0.18	0.49	0.51
6	0.49	0.64	0.72	0.35	0.38	0.15	0.18	0.40	0.45

5. CONCLUSIONS

The existing demand for well-performing short-run forecasting data-driven models is partially satisfied by the model developed in this paper. It performs well on Ukrainian data and can enhance the NBU forecasting toolbox. It also outperforms some benchmarks, such as univariate ARMA for core inflation and Combined ARMA for components without dummies, which is in line with the results of Huwiler and Kaufmann (2013). Also, the results show that disaggregation improves model performance. So, the paper contributes to this discussion as well.

The data used in this study contain several issues that complicate the estimation of any model. These issues can be attributed to the transitional nature of the Ukrainian economy. Among them are strong structural shocks in its recent history. However, as the paper demonstrates, the suggested model is flexible enough to deal with such problems and to produce reasonable forecasts.

There are several directions for further model development. For example, some clustering techniques might be used over space with distances between inflation series. Such an approach can assign series with similar dynamics into clusters and extract additional information on links between them, which can potentially improve model performance even further.

In addition, some exogenous variables can potentially be included in the models. These could improve prediction quality because inflation is likely to be driven by other economic variables as well. However, in this case, the model would face the problem of obtaining forecasts of these exogenous variables to be used as inputs for the inflation forecasting exercise.

REFERENCES

Batini, N., Nelson, E. (2001). The lag from monetary policy actions to inflation: Friedman revisited. Discussion Paper, 6. Bank of England. Retrieved from https://www.lancaster.ac.uk/staff/ecajt/inflation%20lags%20money%20supply.pdf

Benalal, N., Hoyo, J., Landau, B., Roma, M., Skudelny, F. (2004). To aggregate or not to aggregate? Euro-area inflation forecasting. Working Paper Series, 374. European Central Bank. Retrieved from https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp374.pdf

Bermingham, C., D'Agostino, A. (2011). Understanding and forecasting aggregate and disaggregate price dynamics. Working Paper Series, 1365. European Central Bank. Retrieved from https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1365.pdf

Bos, C., Franses, P., Ooms, M. (2002). Inflation, forecast intervals and long memory regression models. International Journal of Forecasting, 18(2), 243-264. https://doi.org/10.1016/S0169-2070(01)00156-X

Diebold, F., Mariano, R. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics, 13(3), 253-263. https://doi.org/10.1080/07350015.1995.10524599

Edge, R., Gurkaynak, R. (2010). How useful are estimated DSGE model forecasts for central bankers? Brookings Papers on Economic Activity, 2. Retrieved from https://www.phil.frb.org/-/media/research-and-data/events/2012/data-revision/papers/Edge_Gurkaynak.pdf

Faryna, O. (2016). Nonlinear exchange rate pass-through to domestic prices in Ukraine. Visnyk of the National Bank of Ukraine, 236, 30-42. https://doi.org/10.26531/vnbu2016.236.030

Gruen, D., Romalis, J., Chandra, N. (1997). The lags of monetary policy. Retrieved from https://www.bis.org/publ/confp04l.pdf

Grui, A., Lepushynskyi, V. (2016). Applying foreign exchange interventions as an additional instrument under inflation targeting: the case of Ukraine. Visnyk of the National Bank of Ukraine, 2016, 238, 39-56. https://doi.org/10.26531/vnbu2016.238.039

Grui, A., Vdovychenko, A. (2019). Quarterly projection model for Ukraine. NBU Working Papers, 3/2019. Kyiv: National Bank of Ukraine. Retrieved from https://bank.gov.ua/news/all/kvartalna-proektsiyna-model-dlya-ukrayini

Hendry, D., Hubrich K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. Journal of Business & Economic Statistics, 29(2), 216-227. https://doi.org/10.1198/jbes.2009.07112

Huwiler, M., Kaufmann, D. (2013). Combining disaggregate forecasts for inflation: The SNB's ARIMA model. Swiss National Bank Economic Studies, 7. Retrieved from https://www.snb.ch/n/mmr/reference/economic_studies_2013_07/source/economic_studies_2013_07.n.pdf

Kongcharoen, C., Kruangpradit, T. (2013). Autoregressive integrated moving average with explanatory variable (ARIMAX) model for Thailand export. 33rd International Symposium on Forecasting. Seoul.

Koop, G., Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. International Economic Review, 53(3), 867-886. https://doi.org/10.1111/j.1468-2354.2012.00704.x

Schorfheide, F., Song, D. (2013). Real-time forecasting with a mixed-frequency VAR. Working Paper, 19712. National Bureau of Economic Research. https://doi.org/10.3386/w19712

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464. https://doi.org/10.1214/aos/1176344136

Stelmasiak, D., Szafranski, G. (2016). Forecasting the Polish inflation using Bayesian VAR models with seasonality. Central European Journal of Economic Modelling and Econometrics, CEJEME, 8(1), 21-42. Retrieved from: http://cejeme.org/publishedarticles/2016-24-25-635945306981718750-3327.pdf

West, K. (1996). Asymptotic inference about predictive ability. Econometrica, 64, 1067-1084. https://doi.org/10.24425/cejeme.2016.119185

Yau, R., Hueng, C .J. (2019). Nowcasting gdp growth for small open economies with a Mixed-Frequency Structural Model. Computational Economics, 54, 177-198. https://doi.org/10.1007/s10614-017-9697-1

Zellner, A., Tobias, J. (1999). A note on aggregation, disaggregation and forecasting performance. Journal of Forecasting 19(5). https://doi.org/10.1002/1099-131X(200009)19:5%3C457::AID-FOR761%3E3.0.CO;2-6

APPENDIX A

Table 4. Table with RMSE for all Possible Architectures with and without a Dummy, for Random Walk and NBU Forecasts.

Months ahead	Components mean				
	1_highest	2_highest	3_highest	3stdev	4stdev
1	0.180	0.216	0.264	0.239	0.195
2	0.241	0.271	0.298	0.283	0.249
3	0.253	0.262	0.283	0.270	0.250
4	0.275	0.284	0.301	0.294	0.279
5	0.276	0.279	0.288	0.276	0.272
6	0.263	0.264	0.278	0.262	0.255
	Components residuals				
1	0.172	0.196	0.201	0.327	0.377
2	0.239	0.253	0.245	0.353	0.382
3	0.258	0.266	0.261	0.393	0.424
4	0.280	0.282	0.269	0.401	0.428
5	0.277	0.277	0.263	0.407	0.419
6	0.264	0.273	0.254	0.380	0.414
	Categories mean				
1	0.248	0.228	0.290	0.241	0.238
2	0.339	0.310	0.350	0.335	0.334
3	0.391	0.360	0.386	0.388	0.386
4	0.426	0.391	0.408	0.418	0.421
5	0.433	0.393	0.399	0.428	0.428
6	0.426	0.376	0.396	0.412	0.420
	Categories residuals				
1	0.247	0.229	0.285	0.248	0.484
2	0.341	0.319	0.344	0.295	0.509
3	0.397	0.369	0.389	0.327	0.592
4	0.431	0.402	0.409	0.354	0.576
5	0.435	0.403	0.405	0.350	0.569
6	0.427	0.388	0.398	0.338	0.549
	Core mean				
1	0.337	0.334	0.423	0.334	0.337
2	0.450	0.436	0.459	0.436	0.450
3	0.500	0.503	0.535	0.503	0.500
4	0.505	0.520	0.549	0.520	0.505
5	0.493	0.518	0.552	0.518	0.493
6	0.481	0.500	0.524	0.500	0.481
	Core residuals				
1	0.337	0.352	0.423	0.347	0.423
2	0.450	0.458	0.459	0.459	0.459
3	0.500	0.531	0.535	0.525	0.535
4	0.505	0.536	0.549	0.530	0.549
5	0.493	0.521	0.552	0.513	0.552
6	0.481	0.505	0.524	0.501	0.524
	Simple CARMA	Simple cat	Simple core	Random walk	Official
1	0.219	0.249	0.332	0.541	0.329
2	0.302	0.340	0.448	0.783	0.394
3	0.337	0.409	0.515	1.017	0.365
4	0.349	0.445	0.521	0.978	0.370
5	0.342	0.439	0.507	0.960	0.429
6	0.334	0.414	0.495	0.892	0.444

Table 5. Table with p-values for the Relative Performance of all Above-Mentioned Models Compared to the NBU Benchmark, according to the Diebold-Mariano-West test.

Months ahead	Components mean				
	1_highest	2_highest	3_highest	3stdev	4stdev
1	0.060	0.100	0.500	0.120	0.070
2	0.160	0.280	0.500	0.350	0.200
3	0.130	0.120	0.170	0.110	0.110
4	0.100	0.090	0.090	0.070	0.090
5	0.150	0.140	0.150	0.120	0.140
6	0.150	0.150	0.140	0.130	0.140
	Components residuals				
1	0.060	0.090	0.080	0.060	0.060
2	0.180	0.200	0.160	0.160	0.150
3	0.140	0.210	0.160	0.130	0.120
4	0.140	0.160	0.120	0.100	0.080
5	0.190	0.180	0.150	0.150	0.130
6	0.170	0.180	0.150	0.150	0.140
	Categories mean				
1	0.120	0.110	0.490	0.080	0.090
2	0.360	0.250	0.470	0.340	0.350
3	0.460	0.400	0.260	0.460	0.430
4	0.530	0.500	0.350	0.530	0.510
5	0.520	0.490	0.390	0.530	0.500
6	0.430	0.400	0.330	0.430	0.410
	Categories residuals				
1	0.120	0.100	0.480	0.160	0.110
2	0.360	0.310	0.530	0.280	0.380
3	0.430	0.390	0.390	0.380	0.490
4	0.440	0.480	0.420	0.440	0.530
5	0.440	0.510	0.450	0.440	0.520
6	0.360	0.450	0.370	0.380	0.440
	Core mean				
1	0.820	0.800	0.970	0.800	0.820
2	0.970	0.960	0.980	0.960	0.970
3	0.880	0.890	0.900	0.890	0.880
4	0.770	0.830	0.840	0.830	0.770
5	0.710	0.790	0.810	0.790	0.710
6	0.640	0.720	0.720	0.720	0.640
	Core residuals				
1	0.820	0.820	0.970	0.970	0.820
2	0.970	0.930	0.980	0.980	0.930
3	0.880	0.870	0.900	0.900	0.870
4	0.770	0.820	0.840	0.840	0.820
5	0.710	0.770	0.810	0.810	0.770
6	0.640	0.700	0.720	0.720	0.700
	Simple CARMA	Simple cat	Simple core		
1	0.100	0.130	0.110		
2	0.300	0.390	0.360		
3	0.400	0.590	0.530		
4	0.390	0.650	0.600		
5	0.390	0.570	0.570		
6	0.350	0.380	0.490		