Comparison of Inflation Forecasting Models: The Case of Slovakia

By

Tomas Sivak

Submitted to

Central European University

Department of Economics and Business

In partial fulfillment of the requirements for the degree of Master of Arts in Economics

Supervisor: Professor Robert Lieli

Vienna, Austria

2022

Abstract

I perform forecasting accuracy tests comparing six different inflation-forecasting models for Slovakia. Using an RMSE objective function, I find that on average the performance of the models is in the following order (from best to worst): i) the average of all individual methods, ii) the stationary VAR model, iii) the nonstationary VAR model, iv) the disaggregated ARMA model, v) the aggregated ARMA model, and vi) the factor model. Another contribution of this exercise is to show that univariate ARMA forecasts of aggregate inflation can be improved by forecasting the components of HICP inflation separately and then taking the appropriate weighted average of the component forecasts. While there is a large literature documenting the improved performance of combined forecasts, this is usually in the context of averaging across models rather than data components.

Acknowledgements

I would like to thank my supervisor Professor Robert Lieli for his valuable advices, suggestions, and guidance throughout my work on the thesis, as well as introducing me to time series analysis.

Table of Contents

List of Abbreviations

ACF	Autocorrelation Function
ARMA	Autoregressive-Moving-Average
BIC	Bayesian Information Criterion
CPI	Consumer Price Index
DSGE	Dynamic Stochastic General Equilibrium
ECM	Error Correction Model
HICP	Harmonized Index of Consumer Prices
IMF	International Monetary Fund
NBS	National Bank of Slovakia
OLS	Ordinary Least Squares
PACF	Partial Autocorrelation Function
PCA	Principal Component Analysis
PCE	Personal Consumption Expenditure
PPI	Producer Price Index
RMSE	Root Mean Squared Error
SO SR	Statistical Office of Slovak Republic
VAR	Vector Autoregression
VWN	Vector White Noise
WB	World Bank

Chapter 1: Introduction

Forecasts of the inflation rate are an important input for many institutions. The chief among these is a country's central bank, whose explicit mandate is to control inflation. Also, institutions such as ministries may incorporate forecasts of inflation rate into their analyses and decision-making processes. For example, if social transfers are indexed to inflation, inflation forecasts facilitate the planning of government expenditures. Many items on the revenue side of the government budget are similarly tied to inflation.

The quality of inflation forecasts is therefore very important. Monetary policy has direct impact on aggregate demand and hence the short-term performance of the economy, as well as on long-term price stability necessary for a well-functioning economy. Central banks, especially those which adopted inflation targeting or similar regimes focusing on the inflation rate, make their decisions based partly on inflation forecasts.

While inflation has not been at the forefront of economic policy for the last decade or so, times seem to be changing. Inflation is making a comeback and has recently reached levels not seen since 1980s. Initially, the current high inflation rate was considered temporary, just a consequence of a sudden increase in aggregate demand following the lifting of covid-related measures. It was expected that inflation would go back to its low levels after the initial positive demand shock would dissipate and supply chains disruptions would be resolved or mitigated. Later, especially because of the outbreak of the Russia-Ukraine war, the high inflation rate started to be considered a longer-lasting issue, especially because of negative aggregate supply shocks on the commodity market. Good inflation forecasts that try to account for these processes are important so that central banks can decide whether to tighten their policies or leave it loose. This topic is more important now and any time in recent memory. And central banks have indeed started taking action now, based also on inflation forecasts.

The current elevated inflation has a rather negative impact on regular consumers, and a positive impact on those benefiting from higher prices (e.g., sellers of commodities such as oil or natural gas). People are responsive to it, and are quite dissatisfied, urging their governments to adopt inflation compensating measures.

This is also the case of Slovakia, which is the country of interest in my thesis. I focus on comparing different time series methods used for forecasting inflation, and evaluating their accuracy. Different institutions of the Slovak government are of course already making inflation rate forecasts. The Ministry of Finance uses a structural model with error-correction equations, which is not publicly available. However, they make available a short-run factor model (Tóth, 2014). The Council for Budget Responsibility uses a dynamic stochastic general equilibrium (DSGE) model (Múčka & Horváth, 2015) for long-run forecasts and a factor model (Kľúčik, 2019) for short-run forecasts. The forecasting model of the National Bank of Slovakia is not publicly available.

The field of forecasting has been neglected over the past ten years. Current trends in inflation are likely to change this course. Moreover, much of the forecast accuracy evaluation in the literature has been done on U.S. data, where the economy has different dynamics and the approach to monetary policy is somewhat different from that of the Eurozone, or Slovakia especially.

In conducting the inflation forecasting exercises, I use the harmonized index of consumer prices (HICP) as a measure of inflation in Slovakia, published monthly from January 1997 till December 2021. I forecast the inflation one-month ahead, one-quarter ahead, and one-year ahead. For forecast comparison I use rolling window strategy.

The first contribution of this thesis is comparing the accuracy of different inflation forecasting methods. I construct five different models and forecasts based on them, plus a simple average of the forecasts. The models are as follows: i) a univariate autoregressive-moving-average (ARMA) model for aggregate inflation, ii) combined ARMA models for the disaggregated inflation components, iii) a vector autoregression (VAR) model with stationary variables, iv) a VAR model with nonstationary variables, and v) a factor model. The last method is a combined forecast that simply takes the average of the five individual forecasts.

The ARMA family is a univariate time series model, which can be used to forecast the aggregate inflation rate of HICP directly (I will henceforth refer to this model as "aggregate ARMA"). Alternatively, HICP can be disaggregated on four levels. On the first level there are twelve categories (e.g., food, housing, etc.). My "disaggregated ARMA" approach consists of forecasting each category itself (based on a suitable ARMA model), and then taking a weighted average of these forecasts, where the weights are the weights of each category within the HICP.

Comparing aggregated and disaggregated inflation forecasts is an important contribution of the thesis. In the forecasting literature it is a common finding that averages of several forecasts work better than the single-model forecasts themselves. By contrast, here I propose a combination approach that is rarely considered in the literature – making forecasts for the various components of inflation separately and then taking a *weighted* average of the forecasts of the twelve inflation categories. I find that this disaggregated approach using weighted averages outperforms the aggregated approach at all forecast horizons.

VAR is a multivariate time series model. As I am forecasting inflation in the short run, I use the New-Keynesian Phillips curve approach. That is, as explanatory variables I use expected inflation (assuming adaptive expectations), real activity measure (output gap), and a nominal supply shock (commodity index). I perform forecasts with both stationary variables (all three first-differenced) and nonstationary variables (keeping all three in their levels where all of them have a unit root). I find that the stationary VAR model outperforms the nonstationary VAR model. Moreover, both VAR models outperform all of the other models, except the simple average of all forecasts.

The factor model is also a multivariate time series model. However, instead of using observable time series as explanatory variables, it rather makes forecasts based on principal components. The idea behind it is that inflation is driven by several “factors” (e.g., monetary policy, fiscal policy, productivity changes, etc.), which are either unobservable or hard to measure. Yet we can extract these “factors” from tens or hundreds of time series using principal component analysis. I find that this model gives the worst results. This is a surprising result, as it is a very popular model. I will discuss it in more detail, but the main downside can be that a lot of variables that I originally planned to use in the model have a lot of missing observations over my estimation period, so I needed to drop them from the analysis. Ultimately, I use only 87 variables and many potentially important variables are left out.

I compare forecast accuracy using root mean squared error (RMSE). The estimation period is from 1997 to 2012. The forecasting period is from 2013 to 2021. I use rolling window of size of 192 months. I find that on average the performance of the models is in this order (from best to worst): “average” model[1], stationary VAR model, nonstationary VAR model, disaggregated ARMA model, aggregated ARMA model, and factor model.

The rest of the thesis is organized as follows. In chapter 2 I provide a brief literature review of the current state of forecasting inflation. In chapter 3 I describe the basic settings for all forecasts, describe all of the models one by one, and describe the respective data used for their estimation. In chapter 4 I describe my results and the quality of different forecasting methods. In chapter 5 I conclude.

[1] “average” model is a simple average of five estimated models

Chapter 2: Literature Review

There is quite a large literature on inflation forecasting. Some models are used more often, some less often, and there is no one size fits all model. It usually depends on the institution which model it chooses. Usually several models are used, and a simple average is done. Moreover, educated guess plays role in publishing actual inflation forecasts.

For short-term forecasts, usually ARMA, VAR, ECM, or factor models are popular. For long-term forecasts, DSGE models perform better. James Stock and Mark Watson did some work in 1990s on forecasting the inflation. Stock & Watson (1999) use Phillips curve to forecast inflation and use an index of aggregate activity based on sixty-one real economic indicators. They present a view on the prediction ability of different variations of Phillips curve as well as other models of inflation forecasting. They analyze monthly data of inflation in USA in years 1959-1997 and test the quality of predictions of different models on 12-month horizon. They start with a general model of Phillips curve depending on the unemployment rate. Then they test the alterations. Explanatory variable is always a real factor, for example industrial production or capacity utilization. Eventually, they test also prediction power of nominal variables such as money supply, however, their inclusion does not improve the forecasts.

Together they test 189 economic indicators. They construct an index of sixty-one measures of aggregate economic activity, which significantly improves the predictions of traditional Phillips curve. They do not focus on supply shocks (according to their preliminary calculations they worsen the prediction ability of the models). Best predictions are made by one-factor models (lags including if necessary). The main conclusion of the study is that the Phillips curve, interpreted as a relation between current real economic activity and future inflation, produces the most reliable and most precise short-term predictions.

Cecchetti, et al. (2000) forecast inflation on a two-year horizon using nineteen real economic activity measures. They find that these forecasts are inferior to their benchmark model, which is a simple AR process. Some other papers have similar findings using Phillips curve exploiting output gap as a measure of real activity, where forecasts perform worse than the AR benchmark, for example Camba-Mendez & Rodriguez-Palenzuela (2001). To summarize the period of 1990s and early 2000s, real economic activity-based models provided small or no improvement over the AR benchmark. However, the results are very sensitive to the choice of estimation period and specification of the particular model.

In the early 2000s comes an important contribution by Atkeson & Ohanian (2001) of introducing a new benchmark, which beats AR benchmark. This was four-quarter inflation in the current period predicting four-quarter inflation in the next period. They show on the data that over the sample period of 1984-1999 the US inflation was better predicted by this new benchmark. Subsequent papers try to verify this finding. Fisher, et al. (2002) indeed, verify Phillips curve approach but on the sample period of 1977-1984, and in this case the new benchmark is beaten. They demonstrate the same also at some periods after 1984. Their other finding is that the Phillips curve models forecast inflation quite badly in times of low inflation volatility and after regime shift.

Stock & Watson (2003) add other predictors to the Phillips curve model and estimate the model in period 1985-1999. In this period, again, new benchmark of Atkeson & Ohanian (2001) provide better forecasts than the extended Phillips curve model. Ang, et al. (2007) use more indicators of inflation rate, such as PCE deflator, CPI core inflation measure, CPI excluding housing, fifteen different horizons, ARMA(1,1), AR(1), and the simple average of forecasts. Their findings confirm the Atkeson & Ohanian (2001) message, that Phillips curve models do not beat the univariate models during the periods 1985-2002 and 1995-2002. One unexpected result is that simple average of forecasts does not beat other forecasts, which is usually found in inflation forecasting practice. One of the most interesting results of the paper is that survey-based measures of inflation almost always beat simple ARMA(1,1) predictions.

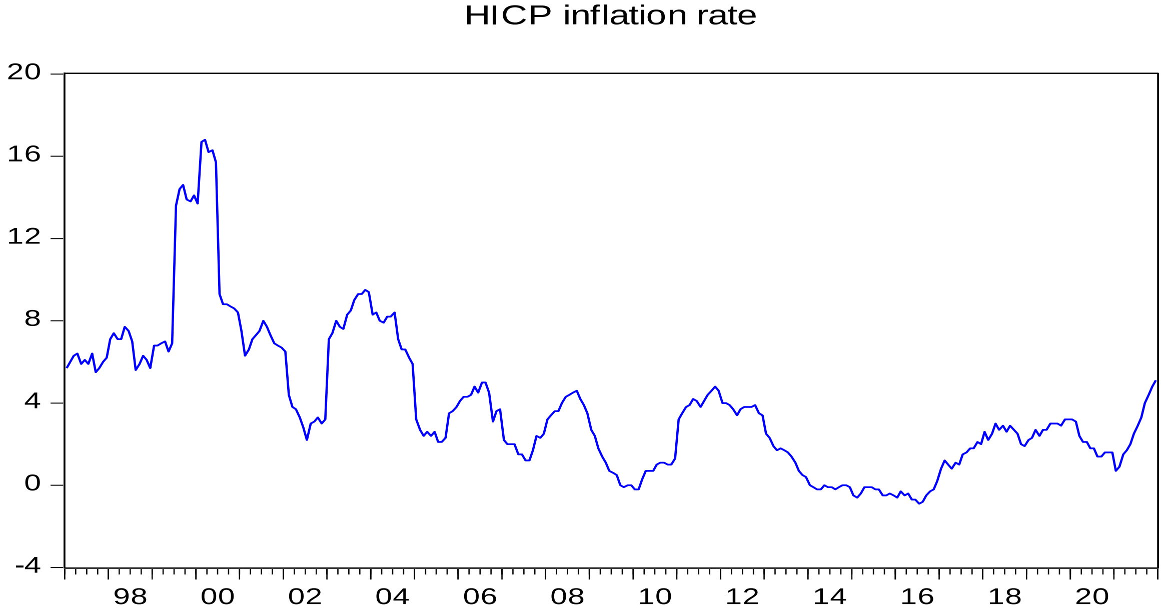
The response to poor performance of Phillips curve forecasts of inflation in the early 2000s was to improve the multivariate models. Stock & Watson (2002) propose using dynamic factor models based on principal component analysis. Their proposal is based on the fact that there is no single predictor of inflation, but many predictors can be used to measure the economic activity. They demonstrate that dynamic factor model performs better than both simple Phillips curve models and simple Atkeson & Ohanian (2001) benchmark in both 1970-1983 and 1984-1996 periods. Despite these results, Stock & Watson (2008) say that backward-looking Phillips curve remains the workhorse of many macroeconomic forecasting models.

In Slovakia, public institutions make their own inflation rate forecasts. Ministry of finance uses a structural model with ECM equations, which is not publicly available. However, they make available a short-run factor model (Tóth, 2014). Council for Budget Responsibility uses a DSGE model (Múčka & Horváth, 2015) for long-run forecasts. The model is used also for policy analysis. For the short-run forecasts, a factor model (Kľúčik, 2019) is used. The forecasting model of National Bank of Slovakia is not publicly available.

Chapter 3: Data and Models

"It's tough to make predictions, especially about the future", said famously Yogi Berra. This tautology is deeply true. Despite that, forecasts are not useless. To the contrary, without forecasts we would not be prepared for maybe negatively surprising future. Inflation forecasting is not an exception. Good inflation forecasts are valuable for a lot of stakeholders. Those include central banks, government, firms, commercial banks, or other institutions.

Figure 1: HICP inflation rate in Slovakia (1997 - 2021, not seasonally adjusted)



Source: Eurostat

After the Great recession, inflation rate in Slovakia was mostly low, as can be seen in Figure 1. There was even mild deflation during 2014-2016. Conditions changed with covid-crisis and current war. Inflation started to increase steeply. New discussion started; will the inflation rate be transitory or long-term? Based on the actions of central banks it seems that the latter is forecasted.

Here come inflation forecasts and their quality. In my thesis I focus on comparing different time series methods used for forecasting inflation, and evaluating their accuracy. There are several motivations for this.

First, this area seemed to be neglected in academia over the last ten years. Second, I focus particularly on Slovakia. Based on my readings of the Slovak literature, an exercise of comparing different forecasting methods has not yet been done, at least not published. Most institutions have probably internal analyses, and several models are being used. I will try to contribute with my own part. Third, most of the published forecasting accuracy measuring papers are focused on US, where the longest time series are available. Fourth, in order to improve forecasts, usually a simple average of several forecasts is made. I do also a weighted average, besides a simple average.

I construct five different models and forecasts based on them, plus a simple average of the forecasts. Those models are namely aggregated ARMA model, disaggregated ARMA model, VAR model with stationary variables, VAR model with nonstationary variables, factor model, and “average” model.

Inflation can be measured in a lot of ways. National CPIs, HICP (in Europe), PCE deflator, GDP deflator, PPI index, etc. I choose HICP, which is harmonized at the EU level. It differs from the national CPIs in what it includes in the consumption basket. Compared to national CPIs, HICP excludes imputed rents and investment goods (such as plastic window). I choose HICP because of its comparability among EU countries, in case somebody else does a similar exercise as me in a different EU country.

HICP can be disaggregated at four levels. First level of disaggregation consists of twelve components. Examples are food (23 % of the basket), housing (18 % of the basket), transport (7 % of the basket), etc. The weights themselves (changing each year) will be later important in the disaggregated ARMA approach. The source of the HICP as well as its weights is Eurostat. The data are not seasonally adjusted.

My working horizon is from 1997 to 2021. I use monthly data. This gives me 300 observations. I divide the period into two parts, and then use the rolling window strategy to make 1-step ahead, 3-step ahead, and 12-step ahead forecasts (one step is one month). First part is from 1997 to 2012. This period is used purely for estimation. So, the size of my window is 192 months. The period from 2013 to 2021 is the part which is used for comparing forecasted values to true values. I choose the division point based on the fact, that the recession in Eurozone following the Great recession ended approximately at this time, according to the Euro Area Business Cycle Dating Committee[2].

As Figure 1 shows, inflation had downward (though uneven) trend, reaching even mild deflation in years 2014-2016. At the end of the figure, we see an upward jump, which is the currently observed elevated inflation rate. Table 1 shows some summary statistics about the inflation in Slovakia over 1997-2021 period.

Table 1: Summary statistics of HICP inflation rate in Slovakia (1997 - 2021, not seasonally adjusted)

Mean	3.86
Median	3.15
Maximum	16.80
Minimum	-0.90
Standard deviation	3.49
p-value of Jarque-Bera test	0.00
Number of observations	300

Source: Own calculations

The highest point was reached during the start of the reform period in Slovakia of Dzurindas' governments. Minimum of -0.90 was reached during the period of mild deflation

[2] https://eabcn.org/dc/chronology-euro-area-business-cycles

of 2014-2016. At this time, several countries in euro area had a similar problem. The rest of the chapter describes forecasting models and data used to estimate them.

3.1 ARMA model

ARMA model is in essence a sophisticated method of extrapolation (Kennedy, 2008). To explain the current value it uses past values, and current and present shocks. Explicitly,

$$\varphi(L)y_t = c + \theta(L)\varepsilon_t$$

where y_t is time series at time t, ε_t is shock at time t, L is a lag operator, c is a constant, and φ and θ are autoregressive (AR) and moving average (MA) coefficients, respectively. y_t is observed time series, in my case inflation rate. ε_t shock can be recovered from y_t by lag polynomial inversion. The model is estimated by conditional maximum likelihood estimation method.

The model is univariate and seems simple, but the problematic part is the specification of ARMA, choosing the right order. I use autocorrelation function (ACF), partial autocorrelation function (PACF), Bayesian Information Criterion (BIC), t-statistics, and as an ultimate check white noise in residuals. Fitting ARMA is partly science, and partly art.

3.1.1 Aggregated ARMA model

Estimation of the aggregated ARMA model requires finding the right order based on criteria mentioned above. I take the time series of HICP in Slovakia, fit the model, fix the order of the model in the first window from January 1997 until December 2012, and use this model to create time series of forecasts, rolling the window always by one period ahead.

3.1.2 Disaggregated ARMA model

The disaggregated model is much more challenging to estimate. It is necessary to estimate each of the twelve components of the HICP index by ARMA, and then making a weighted average of the forecasted time series, based on the weights in the basket in the respective year.

Important is to correctly specify the order of each ARMA, so weighted average of forecasts will be based on the best models. Specifying ARMA model for each component requires the same procedure as described above. After getting twelve time series of forecasts, I create a weighted average. The weights are differing from year to year, but not from month to month. This approach of averaging by weights is not usually used in the literature.

3.2 VAR model

VAR model is just the generalization of an AR model to more than one variable. Y_t here is the vector of time series. In VAR model, each variable depends on the lags of all the other variables, explicitly

$$\Phi(L)Y_t = c + \varepsilon_t$$

where Y_t is a vector of time series at time t, ε_t is a vector of shocks at time t, L is a lag operator, c is a vector of constants, and Φ is a matrix of the coefficients. $\varepsilon_t \sim VWN(0, \Omega)$, where variance-covariance matrix Ω does not need to be diagonal. This intuitively means, that there is allowed contemporaneous correlation between the shocks.

Important is to choose the correct order of the VAR model, i.e., the number of lags of Y_t included. This is done using BIC. The model can be estimated consistently by OLS equation by equation, as there cannot be correlation between the current shock ε_t and the lags of Y_t.

My VAR model contains three variables. As I am making short-run prediction about the inflation rate, I use the New Keynesian Phillips curve approach. This model explains current inflation rate as a function of expected inflation rate, some demand-side or real measure, and a supply shock.

For expected inflation rate I use just adaptive expectation assumption. This means that people expect inflation to be the same in the next period as it was in the current period. As a measure of real activity, I use the output gap. This was calculated by National Bank of Slovakia (NBS) by production function approach. As a supply measure I use a broad commodity index used by International Monetary Fund (IMF) containing all important commodities traded on the market.

Here I do two kinds of estimations. All three variables are non-stationary, by visual inspection and by augmented Dickey-Fuller test (ADF test). The integration of all three variables is of order one, I(1). In the first estimation, I first-difference all of the variables and estimate VAR model with stationary variables. In the second estimation, I keep all of the variables in their level form.

3.2.1 Stationary VAR model

Stationary VAR model uses stationary variables created by first-differencing all three variables, which are in level I(1). The order of VAR chosen based on BIC is VAR(1), where BIC is used from the regression, where the inflation rate is the dependent variable.

3.2.2 Nonstationary VAR model

As another modification of VAR model, I try one with nonstationary data. All the three time series are used in their levels. The order of VAR is again using just VAR(1), as this was preferred to other orders of the model.

3.3 Factor model

Factor model is a multivariate model but the model is estimated based on observable variables only implicitly. Factor model does not use other time series to directly explain the inflation rate. It rather uses so-called factors. The intuition behind the model is that the drivers of inflation are not always measurable or even observable. It admits that as macroeconomists we do not know everything about the economy. As factors driving inflation we can imagine the monetary policy, the fiscal policy, or productivity improvements. These factors are used as explanatory variables.

In order to be able to explain the inflation rate by factors, first we need to estimate them. This is done by principal component analysis (PCA). Principal components obtained by this estimation are used as factors in the factor model. They can be estimated using tens or even hundreds of variables, which most probably drive the inflation rate. As it is a short-term forecast, the intuition for choosing variables is the same as with the VAR model. The model should contain measures of expected inflation, measures of real activity, and supply-side measures, such as commodity prices.

When we find these variables, we can estimate principal components. This is done by maximizing the variance of linear combination of all variables, subject to the constraint that the sum of squared weights should be equal to one. This condition is necessary, as otherwise we would be able to increase variance just by increasing the weights (in this case, scalars in linear combination).

In particular, first principal component should maximize the variance of linear combination of the variables, such that sum of the squares of the weights is equal to one. Second principal component should maximize the variance of linear combination of the variables, and at the same time being uncorrelated with the first principal component, subject to the same constraint. In general, j-th principal component should maximize the variance of linear combination of the variables, and at the same time being uncorrelated with $j-1$ principal components.

$$PC_j = \sum_{i=1}^{n} w_i X_i$$

is the linear combination maximizing the variance of linear combination, always subject to constraint

$$\sum_{i=1}^{n} w_i^2 = 1$$

The number of principal components that can be estimated is the lower number of number of observations and number of variables. However, we are interested only in the first few principal components, as they usually capture most of the variation.

Factor model uses inflation rate as the dependent variable. Independent variables are first few factors (three in my case) just estimated by PCA, and lags of the dependent variable. Explicitly

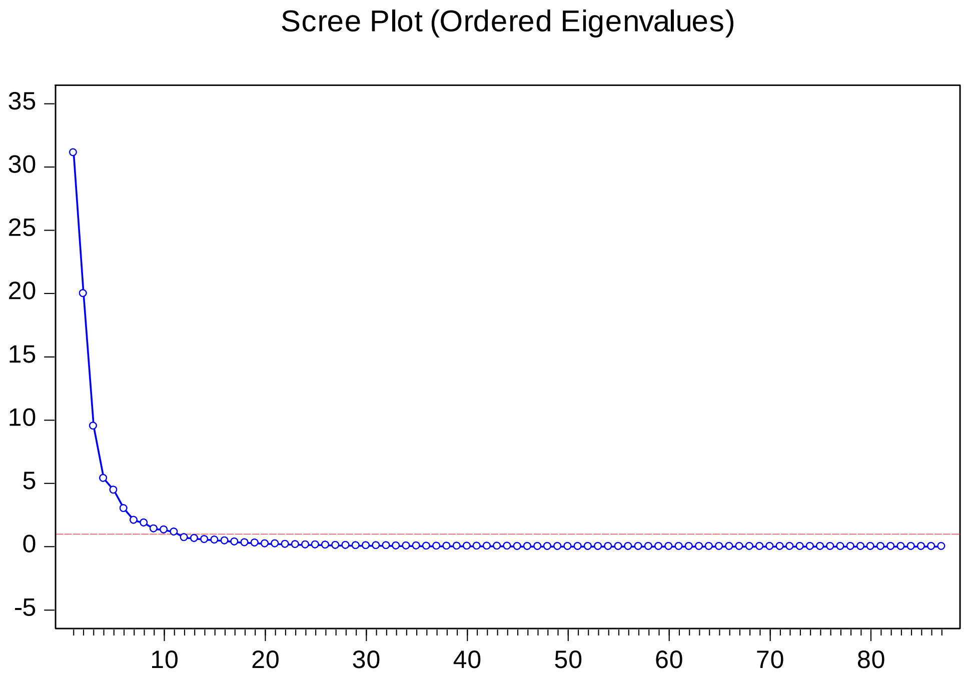
$$Y_t = \delta_0 + \delta_1 \hat{F}_{1t-h} + \cdots + \delta_r \hat{F}_{rt-h} + \delta_{r+1} Y_{t-h} + \cdots + \delta_{r+p} Y_{t-h-p} + u_t$$

where Y_t is the inflation rate at time t, $\hat{F}_{it-h}$ is the i-th factor estimated by PCA at time $t-h$, Y_{t-h-p} is the lagged inflation rate at time $t-h-p$, and δ_j are the coefficients estimated by the factor model by OLS.

3.3.1 Factor model

I use 87 variables in order to estimate principal components. Those include measures such as different commodity prices, economic sentiment indicators, interest rates, employment and other labor market measures, or new orders. List of the variables is included in Appendix 1. From these I estimate principal components. Figure *2* presents the amount of variation that is explained by principal components, starting with first principal component, here using eigenvalues.

Figure 2: Scree plot of the principal components



Source: Own calculations

From Figure 2 we see, that the first three principal components explain most of the variation in the variables. Hence, in my estimation I will use first three principal components as input to the factor model. I use two lags of the dependent variable, as this model gives me the lowest BIC.

3.4 "Average" model

The "average" model is just a simple average of all forecasts. Justification for this is that empirically a combination of forecasts has higher predictive power than just using only single method.

Chapter 4: Forecasting and Results

All the models were estimated by the methods described in chapter 3. For forecasting purposes, I use the rolling window strategy. This means that first I use data from the period of January 1997 until December 2012 and estimate the model based on this data. Then I make 1-step ahead, 3-step ahead, and 12-step ahead forecasts. Hence, I get the forecasts for January 2013, April 2013, and December 2013. Rolling window means that I keep the size of the window fixed, which is in my case 192 months. Then I move the window one-step ahead, which is from February 1997 until January 2013. Using the same model, I again make 1-step ahead, 3-step ahead, and 12-step ahead forecasts. Hence, I get the forecasts for February 2013, May 2013, and January 2014. I continue in this fashion, until I reach the last possible window.

Forecasting has two steps. The first step is iterating the model h-step ahead, and expressing it as a function of the information set and future shocks. Information set contains all the variables we know at the current time. The second step is setting the future shocks to zero. Using this method I get optimal linear forecasts, which are linear projection of Y_{t+h} on the information set.

I forecast on three horizons, 1-step ahead, 3-step ahead, and 12-step ahead. Each step is one month, so effectively it is a month ahead, a quarter ahead, and a year ahead forecast. In order to evaluate accuracy of different forecasting models I use RMSE. In essence it compares forecasted time series to the actual time series and evaluates the difference between them. RMSEs of different models for all forecasting horizons are shown in Table 2.

Table 2: RMSE’s of forecasts of models at different forecasting horizons

	1-step ahead	3-step ahead	12-step ahead
ARMA - aggregated model	0.333	0.322	0.346
ARMA - disaggregated model	0.317	0.306	0.328
VAR - stationary model	0.296	0.287	0.296
VAR - nonstationary model	0.305	0.297	0.308
Factor model	0.350	0.367	0.520
“Average” model	0.285	0.275	0.295

Source: Own calculations

I find that on average the performance of the models is in this order (from best to worst): “average” model, stationary VAR model, nonstationary VAR model, disaggregated ARMA model, aggregated ARMA model, and factor model.

The “mistake” in forecasts is usually in the size of 0.3 percentage points. This can be considered as a quite precise forecast. Even at 12-step ahead forecast the accuracy is quite well. The outlier is the factor model (RMSE of 0.520), but this is the worst-performing model. What is striking about the results is that in all the models (besides the factor model) the 3-step ahead forecasts outperform the 1-step ahead forecasts.

Interesting observation is that at each forecast horizon, the order of accuracy of the models is the same. Usually, some models perform better at short-term forecasting and some models perform better at long-term forecasting. However, in case of the six models presented, the horizon does not matter. Yet we need to take into account, that even the longest horizon (12-months ahead) can be considered a short-term forecast. Time-series methods usually perform better at short horizons, while DSGE models perform better at longer horizons. Nevertheless, the consistency of accuracy ordering is worth mentioning.

Another interesting observation is that disaggregated ARMA outperforms aggregated ARMA at each horizon. This is the second contribution of my thesis. It follows, that extra effort of making twelve forecasts (one for each component of HICP) pays off, taking into account that we ultimately make a weighted average based on current years' weights in the consumption basket.

Factor model gives the worst results of all the estimated models. This is an unusual result, because this model is very popular, both in short-term forecasting and in nowcasting.

There can be several problems with the factor model. Probably the main is that the number of variables is quite low. I use only 87 variables, which does not need to be sufficient in the setting of factor model. There were several reasons why I dropped more than 300 variables. The main issue was that my estimation period starts in 1997, but the other time series start later or much later. Other time series had breaks in them. And some of the time series ended too soon, current data were not available.

A connected problem can be that important variables were dropped, those capturing most of the variation in the inflation rate. This can be tackled by using different sample period, e.g., starting in year 2000 or later. Yet because of the comparability of the forecast accuracy to the other models I have to preserve the sample period of 1997-2021.

In order to formally asses if two RMSEs are equal, I use Diebold-Mariano test. The null hypothesis here is that "Both forecasts have the same accuracy". Table 3 shows the p-values for the test for 1-step ahead. Most p-values are close to zero, which indicates statistically significant difference in RMSEs.

Table 3: Diebold-Mariano test p-values for 1-step ahead forecast

	Aggregated ARMA	Disaggregated ARMA	Stationary VAR	Nonstationary VAR	Factor model	"Average" model
Aggregated ARMA	1.00					
Disaggregated ARMA	0.00	1.00				
Stationary VAR	0.00	0.19	1.00			
Nonstationary VAR	0.00	0.01	0.24	1.00		
Factor model	0.00	0.02	0.01	0.25	1.00	
"Average" model	0.00	0.04	0.00	0.00	0.00	1.00

Source: Own calculations

Table 4 shows outcomes of Diebold-Mariano test for 3-step ahead forecasts.

Table 4: Diebold-Mariano test p-values for 3-step ahead forecast

	Aggregated ARMA	Disaggregated ARMA	Stationary VAR	Nonstationary VAR	Factor model	"Average" model
Aggregated ARMA	1.00					
Disaggregated ARMA	0.00	1.00				
Stationary VAR	0.00	0.21	1.00			
Nonstationary VAR	0.00	0.01	0.20	1.00		
Factor model	0.00	0.01	0.00	0.09	1.00	
"Average" model	0.00	0.04	0.00	0.00	0.00	1.00

Source: Own calculations

Table 5 shows outcomes of Diebold-Mariano test for 12-step ahead. Results of the tests almost consistently show that there is not a statistically significant difference in RMSEs between these pair of models: disaggregated ARMA vs. stationary VAR, nonstationary VAR vs. stationary VAR, and nonstationary VAR vs. factor model.

Table 5: Diebold-Mariano test p-values for 12-step ahead forecast

	Aggregated ARMA	Disaggregated ARMA	Stationary VAR	Nonstationary VAR	Factor model	“Average” model
Aggregated ARMA	1.00					
Disaggregated ARMA	0.00	1.00				
Stationary VAR	0.00	0.49	1.00			
Nonstationary VAR	0.00	0.03	0.11	1.00		
Factor model	0.00	0.00	0.00	0.00	1.00	
“Average” model	0.03	0.01	0.00	0.00	0.00	1.00

Source: Own calculations

Important to say is that RMSE for disaggregated ARMA is significantly different from aggregated ARMA. This formal test provides reassurance for the already obvious difference in the numbers and implies, that first forecasting each of the twelve components separately and then making a weighted average of them is worth the effort to get more precise forecasts of the inflation rate.

Chapter 5: Conclusion

I perform accuracy tests comparing six different inflation-forecasting models for Slovakia. I find that on average the performance of the models is in this order (from best to worst): “average” model, stationary VAR model, nonstationary VAR model, disaggregated ARMA model, aggregated ARMA model, and factor model.

My contribution is twofold; first, I assess which of the models are more suitable for forecasting inflation in Slovakia, as mentioned above. The usual practice of making a simple average of several forecasts works the best. My results just reassure this. VAR models have superior forecasting ability compared to ARMA models. Higher complexity of VARs can probably explain its better performance as compared to univariate ARMA models. The inferiority of factor model is somehow disappointing, but some issues have already been discussed. In general, it should perform the best as in normal settings it sometimes includes even hundreds of time series. But specification plays a significant role in building any model. So, the performance can be and is explained by correct choice of variables and their functional form. Here is the possible pitfall of poorer performance of factor model, especially at 12-step ahead horizon. Data availability for particular horizon should also be taken into account.

Second, I compare the accuracy of aggregated ARMA and disaggregated ARMA. Aggregated ARMA forecasts HICP by itself. In disaggregated ARMA, I first forecast each of the twelve components of HICP, and then make a weighted average, where weights are the weights in the consumption basket. My contribution is that I make specifically a *weighted* average of several forecasts (twelve time series), as opposed to making a *simple* average done in any other setting (i.e., not necessarily ARMA). I find that making a weighted average (disaggregated ARMA) outperforms just simple forecast based on HICP itself (aggregated ARMA) at each horizon.

An interesting question is how well would the same models perform in country other than Slovakia. Both ARMA and VAR models have a reputation of delivering very good forecasting results in general. Factor model is a very popular model currently. Assessing external validity would need to take into account the specification of the models. Even very well performing models can give poor results if they are not well-specified. Results presented here depend a lot on specification and could be quite different, if not enough time is allocated to this issue. Also, the choice of explanatory variables for VAR and factor models plays a big role. Factor model results presented here are somewhat disappointing despite their popularity in many countries. However, mainly data constraints and choice of estimation period are most probably responsible for poorer performance in my setting. A simple change of the starting estimation period from 1997 to 2002 could make a lot of difference, and the performance would come closer to the international results, where factor models do a very good job of forecasting.

There is space for further research. Comparing different forecasting models and constantly improving forecasting methods only helps getting better forecasts of the inflation rate. However, there is much broader space in trying to make weighted-average forecasts. HICP can be disaggregated on the first level to twelve components, which is not that many to forecast. Models other than ARMA can be assessed. For some components it can be quite easy to find variables explaining them. Literature on food inflation forecasting is not scarce, mentioning just one component.

Appendices

Appendix 1

Table 6: List of time series used in factor model

Indicator	Frequency	Source
3m EURIBOR	m	NBS
Actual individual consumption	q	Eurostat
Aluminum price	m	WB
Banana price (EU)	m	WB
Banana price (Africa)	m	WB
Base metals price	m	IMF
Beef price	m	WB
Consumption of government	q	Eurostat
Consumption of NPISH	q	Eurostat
Chicken price	m	WB
Coal price (Europe)	m	WB
Coal price index	m	IMF
Coal price (Russia)	m	WB
Cocoa price	m	WB
Coconut oil price	m	WB
Coffee arabica price	m	WB
Coffee robusta price	m	WB
Consumption of households	q	Eurostat
Copper price	m	WB

Cotton price index	m	IMF
Crude oil Brent price	m	WB
Crude oil Dubai price	m	WB
Crude oil price index	m	IMF
Crude petrol price	m	WB
Crude oil WTI price	m	WB
Economic sentiment indicator	m	SO SR
Exports	q	Eurostat
Fish meal price	m	WB
Gross fixed capital formation	q	Eurostat
Gold price	m	WB
Groundnuts price	m	WB
Natural gas index	m	WB
Industrial inputs price index	m	IMF
Industrial production	m	SO SR
Interest rate on deposits - nonfinancial companies	m	NBS
Interest rate on household deposits	m	NBS
Lead price	m	WB
Logs price (Europe)	m	WB
Logs price (Canada)	m	WB
Money supply M1	m	NBS
Money supply M2	m	NBS
Money supply M3	m	NBS
Maize price	m	WB

Natural gas price index	m	IMF
Natural gas (Russia)	m	WB
Natural gas (US)	m	WB
Nickel price	m	WB
Net exports	q	Eurostat
Orange price	m	WB
Output gap	m	NBS
Palm oil price	m	WB
Participation rate (employment indicator)	q	Eurostat
Platinum price	m	WB
Palm kernel oil price	m	WB
Plywood price	m	WB
Potash price	m	WB
Precious metals price index	m	IMF
Rice price (Thailand)	m	WB
Rice price A1 (Thailand)	m	WB
Rubber price	m	WB
Sawn wood (Cameroon) price	m	WB
Sawn wood (Malaysia) price	m	WB
Shrimp price	m	WB
Silver price	m	WB
Soybean meal price	m	WB
Soybean oil price	m	WB
Soybeans index	m	IMF

Sugar EU price	m	WB
Sugar US price	m	WB
Sugar world price	m	WB
Tea average price	m	WB
Tea Colombo price	m	WB
Tea Kolkata price	m	WB
Tea Mombasa price	m	WB
Total factor productivity	q	NBS
Tin price	m	WB
Tobacco price	m	WB
Unemployed (Labor force survey)	q	Eurostat
Unemployment rate (Labor force survey)	q	Eurostat
Urea price	m	WB
Wages (Compensation of employees)	q	Eurostat
Wheat hrw price	m	WB
Wheat srw price	m	WB
Zinc price	m	WB
m - monthly frequency, q - quarterly frequency		
Eurostat - European Statistical Office, IMF - International Monetary Fund, NBS - National Bank of Slovakia, SO SR - Statistical office of Slovak Republic, WB - World Bank		

References

Ang, A., Bekaert, G. & Wei, M., 2007. Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?.

Atkenson, A. & Ohanian, L. E., 2001. Are Phillips Curves Useful for Forecasting.

Atkeson, A. & Ohanian, L. E., 2001. Are Phillips Curves Useful for Forecasting.

Camba-Mendez, G. & Rodriguez-Palenzuela, D., 2001. Assessment Criteria.

Cecchetti, S., Chu, R. & Steindel, C., 2000. The Unreliability of.

Fisher, J., Liu, C. T. & Zhou, R., 2002. When Can We Forecast Inflation?.

Kennedy, P., 2008. *A Guide to Econometrics.* s.l.:s.n.

Kľúčik, M., 2019. *Tracking the Course of the Economy - Nowcasting of basic macroeconomic indicators of Slovakia.* s.l.:Council for Budget Responsibility.

Múčka, Z. & Horváth, M., 2015. *Fiscal Policy Matters - A New DSGE Model for Slovakia.* s.l.:Council for Budget Responsibility.

Stock, J. & Watson, M., 1999. *Forecasting Inflation.* s.l.:s.n.

Stock, J. & Watson, M., 2002. Forecasting Using Principal Components from a Large Number of Predictors.

Stock, J. & Watson, M., 2003. Forecasting output and inflation: The role of asset prices.

Stock, J. & Watson, M., 2008. Inflation Forecasting. *Journal of Monetary Economics.*

Tóth, P., 2014. *Model pre krátkodobý výhľad ekonomickej aktivity (MRKVA).* s.l.:Ministry of Finance.