Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and Operational Research

Inflation Forecast Combination: The Effectiveness Of Shapley Values Using ARMA Models

Shang Dong (547762)

Erasmus Universiteit Rotterdam

Supervisor:	Philip Hans Franses
Second assessor:	Wendun Wang
Date final version:	2nd July 2023

Abstract

Forecasting inflation remains one of the topics that is still sought after. The search for the best forecasting model for inflation forecasting is still ongoing. However, Bates & Granger (1969) have shown that the combination of forecasts improves upon individual forecast models.

This study dives deeper into the search for optimizing inflation forecasting using forecast combinations. This is done by conducting a simulation- and empirical study for the newfound method by Franses (2023), namely, forecast combinations with Shapley value-based weights. The study uses different ARMA models and found that forecast combination based on weights determined by Shapley values performs better than the equal-weighted average and weights determined by OLS. However, the difference in values is very minimal.

Contents

1 Introduction

The interest in inflation forecasts continues to grow, and the search for better forecast models is still ongoing. This is mainly due to the importance of inflation in many sectors. One of them is for policy making as central banks conduct monetary policies based on inflation forecasts. Besides, it is a big factor in government fiscal policy decisions such as the tax rate or government expenditures. Furthermore, accurate inflation expectations are crucial for businesses and households to make informed decisions regarding investments, production, pricing, and finances.

One of the models that have been widely used for inflation forecasting, is the ARMA model. Many studies have proven that it performed relatively well in capturing the dynamics of inflation (such as Oseni et al. (2017) and Stovicek (2007)), due to its flexibility and for capturing both short and long-term trends (Stock & Watson (2007)).

However, individual ARMA models also have shortcomings, such as data instability, model misspecification, or overfitting. To mitigate these problems, combining multiple forecasts can improve accuracy according to R. T. J. Elliott G. & Stock (2005). Recently, a new method has been introduced to combine forecasts with Shapley values. This method was proposed by Franses (2023), where the study proved that assigning the weight by use of Shapley values, improved the accuracy of the forecast combinations in comparison with the equal-weights combination. Since this is quite a new concept, there are not many studies about this subject yet. This leads to the following research question:

Does Shapley value-based combination improves upon an equally-weighted combination when forecasting inflation with ARMA models?

This study contributes to the literature by building on the existing literature on forecast combinations with Shapley values. This paper investigates the newfound combination technique by the use of different ARMA models for inflation prediction. This study uses the CPI values of the monthly variables from the FRED-MD (Federal Reserve Economic Data - Monthly Database) proposed by McCracken & Ng (2016) which includes 127 variables. The full sample data includes all observations from January 2010 (2010M1) up until December 2022 (2022M12), resulting in 156 observations. The sample period is split into three equal parts.

Five different ARMA models are used to forecast inflation individually, namely: ARMA(1,0), ARMA(0,1), ARMA(1,1), ARMA(2,1), ARMA(1,2). Hereafter, the individual forecasts are combined using three different methods, namely: weights determined by Ordinary Least Squares (OLS), weights determined by Shapley values(SV), and the equal-weighted average(WA). These combinations are then compared using two loss functions, the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE).

The results show that the forecast combination based on weights determined by Shapley values performs better than the equal-weighted average, as mentioned by Franses (2023), and weights determined by OLS. However, the difference in values is very minimal, which is in line with Smith & Wallis (2009). They concluded that estimating weights does not lead to significant improvements when compared to the equal-weighted average.

The remainder of this study is structured as follows. In Section 2, the literature on inflation forecasting is discussed. In Section 3, the dataset that is used in this study is discussed. In Section 4, the considered models and methods are discussed and explained. In Section 5, the simulation study is presented and discussed. In Section 6, the results are presented and Section 7 provides a conclusion, and further research is discussed.

2 Literature

It can be seen in practice that a particular forecast model performs better in certain periods while in other periods other forecast models perform better. According to Hendry (2002), forecast models often lack certain properties that are required to track changes in the targeted variable. Moreover, Hendry (2002) also concludes that forecast models are not always accurately specified. Bates & Granger (1969) suggested a new method to this, namely, the combination of forecasts to improve individual forecast models. They showed that the lowest variance of an individual forecast is always higher or equal to an equal-weighted average of two unbiased forecasts. Based on their findings multiple other researches have been conducted, where it was shown that forecast combinations can improve forecast accuracy over an individual model (Newbold & Granger (1974), Granger & Newbold (1986), Granger & Jeon (2004), and Yang (2004)). Moreover, Smith & Wallis (2009) concluded that the equally-weighted combination is more accurate, in terms of Mean Squared Forecasting Error (MSFE) than a combination of forecasts where the weights are estimated. They also conducted the example from Stock & Watson (2007) where they found that the gain in MSFE had no practical significance.

However, another study conducted by G. Elliott (2011) showed that the average weight method is optimal if and only if the unit vector is a scalar multiple of an eigenvector of its variance-covariance matrix. Franses (2023) added to the ongoing research of forecast combinations. He proposed a new method to combine forecasts using Shapley values and proved that it indeed performed better than the equally-weighted combination, where the original idea of Shapley values can be seen in Shapley (1953).

To test whether this is the case for inflation forecasting, ARMA models are used in this study. ARMA models have become one of the main models to forecast and model time series since the study of Box & Jenkins (1970). One of the studies that have been conducted is by Stovicek (2007)). He revealed that ARMA models outperform AR models when the same degrees of freedom are allowed. Moreover, Nyoni & Nathaniel (2019) conducted a study where the study concluded that the ARMA(1,2) model is the best model to forecast inflation in Nigeria. The used ARMA models in this study are based on previous studies and a more detailed description of the used ARMA models is described in section 4.1.

3 Data

The data used in this study is from the FRED-MD (Federal Reserve Economic Data - Monthly Database) proposed by McCracken & Ng (2016) which includes 127 monthly macroeconomic variables. The dataset is updated in real-time using the FRED database and can be found on the webpage of McCracken & Ng (2016).

In this study the data which includes all observations from January 2010 (2010M1) up until December 2022 (2022M12) is used, resulting in 156 observations, where the sample period is split into three equal parts. The first sub-sample is used to estimate the models, whereas the second sub-sample is used to find the weights of the combined forecasts and the third sample is used for out-of-sample evaluation of forecast accuracy. Furthermore, the inflation in month t, denoted by π_t, is computed as $\pi_t = \log(P_t) - \log(P_{t-1})$, where P_t is the CPI (Consumer Price Index) in period t.

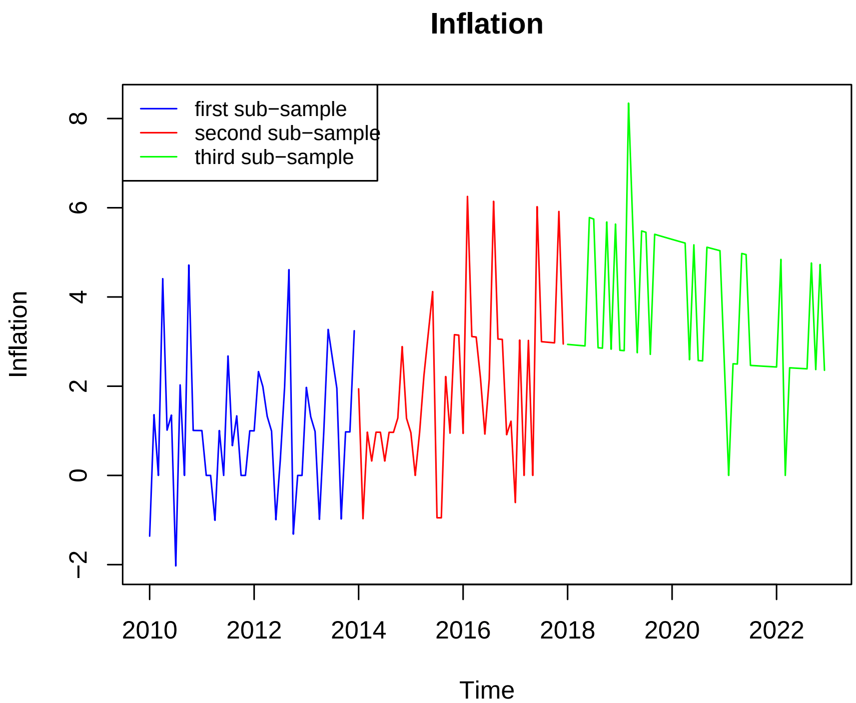

Figure 1: Plot of CPI inflation

Table 1: Summary characteristics of CPI inflation

Minimum	**Median**	**Mean**	**Maximum**
-2.0284	2.4213	2.3601	8.3450

Figure 1 shows the monthly CPI inflation in percentages over the whole sample period where for example, 2 stands for 2 percent and Table 1 shows the key characteristics of inflation over the whole sample period. It can be seen that the inflation is quite volatile in this period. However, based on Federal Reserve's vision, where the target for inflation is 2 percent, it can be considered as normal inflation since the mean and the median are both around 2 percent. Moreover, this paper uses three different sub-samples, which is different from the commonly used method which uses only two sub-samples. Where the commonly used method uses the second sub-sample for forecasting, this paper uses it to find the weights of the Shapley values and the third sub-sample for forecasting. This is illustrated in Figure 1 using various colors. Section 5 dives deeper into this.

4 Methodology

4.1 ARMA-model

To forecast inflation, a well-known model in time series forecasting is used, namely, the Auto Regressive Moving Average (ARMA) model. The ARMA model is derived from the Auto-Regressive Integrated Moving Average (ARIMA) model. The ARIMA models, first introduced by Box & Jenkins (1970), belong to the group of linear methods, where it integrates into Auto-Regressive (AR) and Moving Average (MA) models. The model is used to forecast based on historic data and has three components: p (number of auto-regressive terms), d (number of non-seasonal differences), and q (number of lagged forecast errors). This model generates time series of the form:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \tag{1}$$

where y_t represents the actual values at time t and ϵ_t represents the error terms at time t. ϕ_0 is the constant term and ϕ_i and θ_j stands for the model parameters where i = 1,2,...,p and j = 0,1,...,q. Furthermore, the error terms are assumed to be random independent and identically distributed with mean zero and variance σ^2.

In this study d is set to 0, making it an ARIMA(p,0,q) model, this is equivalent to ARMA(p,q). Armstrong (2001) suggests that using at least five models for forecast combination can lead to substantial error reduction, while additional models provide marginal improvements. Therefore, this study considers five different ARMA models.

The chosen ARMA models are based on their performance in previous studies on inflation forecasting. Nyoni (2019) concluded that the optimal model for forecasting inflation rates in Senegal is the ARMA (1, 0) model and Moffat & David (2016) found that ARMA (0,1) is the best model to forecast the rate of inflation in Nigeria. Furthermore, an often used ARIMA benchmark model is the ARMA(1,1) model, as seen in Stock & Watson (2008). Moreover, based on Guerrón-Quintana & Zhong (2017) the ARMA(2,1) model was the optimally selected linear model for forecasts in times of crisis, and Nyoni & Nathaniel (2019) compared three different models (ARMA(1, 2) model, ARIMA(1, 1, 1) model and the AR(3)–GARCH(1, 1) model), where it was clear that the ARMA(1,2) model performed the best. Therefore the ARMA models that are used are: ARMA(1,0), ARMA(0,1), ARMA(1,1), ARMA(2,1), ARMA(1,2).

4.2 Forecast combinations

4.2.1 Equally-weighted combination

According to Smith & Wallis (2009), estimating weights for forecast combinations is unlikely to lead to any significant improvements. Therefore, our benchmark model is the equally-weighted forecast combination, which is first introduced by Bates & Granger (1969). The equally-weighted combination is given by the following formula:

$$\hat{y}_{t+h} = \sum_{i=1}^{n} \frac{1}{n} f_{i,t}, \tag{2}$$

Here, n is the number of models used, and $f_{i,t}$ is the forecast for model i at time t.

4.2.2 Shapley based combination

Contrary to Smith & Wallis (2009), Franses (2023) has proven that assigning weights can indeed lead to forecast performance improvement compared to the equal-weighted average. To test if this is the case for inflation, by use of ARMA models, the following formula to calculate the Shapley values is used:

$$SH_j = \sum_{S \subseteq K \setminus \{j\}} \frac{\mid S \mid ! (\mid K \mid - \mid S \mid -1)!}{\mid F \mid !} [R^2(S \cup \{j\}) - R^2(S)] \tag{3}$$

Where K is the set with all models and $R^2(S)$ is the R^2 of the model with all forecasts in a set S $\subseteq$ K. Moreover, the Shapley weights in the new combined forecast:

$$\hat{y}_{t+h} = s_1 f_{1,t} + s_2 f_{2,t} + .. + s_K f_{K,t} \tag{4}$$

are found as:

$$s_j = \frac{SH_j}{R^2_{12..k}} \tag{5}$$

where $R^2_{12..k}$ is the R^2 with all forecasts.

4.2.3 OLS combination

Another method is the Ordinary Least Squares (OLS)-based combination, first introduced by Granger & Ramanathan (1984). This method regresses the actual value of the forecasts and is as follows:

$$y_t = \beta_1 + \beta_2 f_{1,t} + \beta_3 f_{2,t} + ... + \beta_K + f_{K,t} + \varepsilon_t \tag{6}$$

where β_1 is the constant, β_r, $r = 2, ..., K$ are the weights and ε_t is the error term.

4.3 Loss functions

To measure the accuracy of the forecast combinations the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are used. The two measurements can take on values $[0, \infty)$, where the lower the error measurement the better the accuracy of a forecast.

4.3.1 RMSE

The RMSE is a useful measurement when large predictions need to be penalized more and is done by squaring the error. The RMSE is given by

$$RMSE = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(\hat{y}_t - y_t)^2} \tag{7}$$

and measures the mean squared error, where more importance is given to larger errors.

4.3.2 MAE

The MAE is a useful measurement because the marginal contribution of the error is constant. The MAE is given by

$$MAE = \frac{1}{m}\sum_{t=1}^{m}|\hat{y}_t - y_t| \tag{8}$$

and measures the mean of absolute errors.

5 Simulation study

A simulation study is conducted to compare and evaluate the performance of forecast combination methods. The simulation is performed by use of an R code which focuses on three different methods, namely: weights determined by Ordinary Least Squares (OLS), weights determined by Shapley values, and the equal-weighted average.

5.1 Data Generated Process

To simulate a realistic forecasting scenario a Data Generated Process(DGP) is used. The simulation is repeated three times, each time with a different DGP. The DGP is generated using three different models, namely: ARMA(5,5), ARMA(4,4), and ARMA(3,3), and consists of 1500 observations. These models use the following coefficients where the first x numbers are coherent to the ARMA(x,x) model: AR= (0.5, -0.2, 0.3, 0.1, -0.4) and MA= (0.4, -0.1, 0.2, 0.2, -0.3), where the coefficients are arbitrary chosen. For example, the ARMA(4,4) model has the following coefficients: AR= (0.5, -0.2, 0.3, 0.1) and MA= (0.4, -0.1, 0.2, 0.2). Furthermore, the generated sample period is split into three equal sub-samples, each containing (1500/3=) 500 observations.

5.2 First sub-sample

The first sub-sample is used to estimate the models, where five different ARMA models are used. The five ARMA models that are used are: ARMA(1,0), ARMA(0,1), ARMA(1,1), ARMA(2,1), ARMA(1,2).

5.3 Second sub-sample

The second sub-sample is used to find the weights of the Shapley values. The data of the first sub-sample is fit onto the ARMA models and then the recursive one-step-ahead forecasts are generated for each model for the second sub-sample. Hereafter, the R^2 value is calculated for each possible combination of the five forecasts, resulting in 31 combinations. With these values, the Shapley weights are calculated for each forecast. This process is repeated 1000 times. This concludes the second sub-sample.

5.4 Third sub-sample

The third sub-sample is used for out-of-sample evaluation of forecast accuracy. The data of the second sub-sample is fit onto the ARMA models and then the recursive one-step-ahead forecasts are generated for each model for the third sub-sample. For the OLS-based forecast combination, an OLS regression is conducted where the dependent variable is the third sub-sample based on DGP and the independent variables are the forecasts. The found coefficients are then multiplied by their corresponding forecasts. For the Shapley-based forecast combination, the forecasts are multiplied by their corresponding Shapley weights. For the equal-weighted average, all the forecasts are combined and divided by 5 (total number of forecasts).

5.5 Simulation results

To assess the performance of these methods, two loss functions are used. The loss functions are the root mean squared error (RMSE) and the mean absolute error (MAE). The results are shown in Table 2, where SV stands for the Shapley-based combination, WA stands for the equal-weighted average combination and OLS stand for the OLS-based combination.

Table 2: Loss function values of the different DGPs

	ARMA(5,5)	**ARMA(4,4)**	**ARMA(3,3)**
RMSE(SV)	1.828103	<u>1.756940</u>	<u>1.512374</u>
RMSE(WA)	<u>1.828101</u>	1.757257	1.512424
RMSE(OLS)	1.845220	1.803618	1.623429
MAE(SV)	<u>1.460097</u>	<u>1.404627</u>	<u>1.207737</u>
MAE(WA)	1.460099	1.404899	1.207778
MAE(OLS)	1.477029	1.422533	1.264674

The best-performing model is underlined in Table 2. The results differ based on the RMSE values, WA performed the best for ARMA(5,5) and SV performed the best for ARMA(4,4) and ARMA(3,3). SV performed the best across all DGPs based on the MAE values. OLS performed the worst across all measurements. However, it needs to be mentioned that the difference between the values for SV and WA is minimal, with emphasis on ARMA(5,5) where the difference for both the RMSE and MAE is 0.000002.

Moreover, how many times (out of 1000 times) SV performed better than WA and OLS is another interesting thing to take into account. In Table 3 the number of times for each measure where SV performed better than WA (SV<WA) or OLS (SV<OLS) can be seen. A thing to notice is that WA performed more often better than SV for ARMA(5,5) based on RMSE, this explains why the RMSE for WA was lower than the RMSE for SV (Table 2).

What's more, is that every value that compares SV with WA is close to 500. This means that there is not a big difference in the amount of which model performed more often better, this explains the minimal difference between the RMSE and MAE values for SV and WA (Table 2). When compared to OLS it can be seen that SV performed approximately 60% of the time better than OLS, which can also be linked back to the RMSE and MAE values.

Table 3: Amount of times SV performs better than WA and OLS of the different DGPs

	ARMA(5,5)	**ARMA(4,4)**	**ARMA(3,3)**
#RMSE(SV<WA)	494	547	511
#MAE(SV<WA)	508	538	518
#RMSE(SV<OLS)	579	577	563
#MAE(SV<OLS)	592	577	576

6 Results

The same process has been conducted, as shown in 5, with real data. This results in the following weights for each method, where again SV stands for the Shapley-based combination, WA stands for the equal-weighted average combination and OLS stands for the OLS-based combination:

Table 4: Weights for the different forecasts

	ARMA(1,0)	**ARMA(0,1)**	**ARMA(1,1)**	**ARMA(2,1)**	**ARMA(1,2)**
SV	0.07670405	0.07482136	0.33955242	0.22432683	0.28459533
WA	0.20	0.20	0.20	0.20	0,20
OLS	0.119861758	0.921210961	0.007964881	-0.053156154	0.004118555

It is quite interesting to look at Table 4, where the difference in weights, obtained by SV and OLS can be seen. SV assigns its weights based on individual contributions and fairness, while OLS assigns its weights based on minimizing the sum of squared errors (SSE). Whereas SV assigns the lowest weight to the ARMA(0,1) forecast, OLS assigns the highest weight to it. Meaning that a greater weight of the ARMA(0,1) forecast minimizes the SSE but has the lowest contribution in the forecast combination.

To look further into the forecast combinations, the RMSE and MAE are used to evaluate the performance of each model. The results are presented in Table 5, where it can be seen that SV performs the best based on both RMSE and MAE and OLS performs the worst. However, just like in the simulation study, the difference between the values is minimal. These results are in line with Smith & Wallis (2009), who concluded that estimating weights for forecast combinations have no practical significance.

Table 5: Loss function values for the forecast combinations

	RMSE	**MAE**
SV	0.2071613	0.1480711
WA	0.2097413	0.1507285
OLS	0.2134280	0.1553116

7 Conclusion

Inflation forecasting is important since it can influence economic decision-making. By predicting inflation, informed decisions can be made to manage financial situations and advance economic stability. Bates & Granger (1969) showed that combining forecasts improve forecast accuracy over an individual model. Previous research has demonstrated that assigning weights in forecast combinations has little to no effect compared to the equal-weighted average (Smith & Wallis (2009).

However, a newfound study (Franses (2023)) proposed a new method for forecast combination based on Shapley-based weights, where he proved that it performed better than the equal-weighted combination. Since this is a newfound concept, not much research has been conducted about this subject. Therefore, the focus of this study is to compare this claim by evaluating the Shapley-based combination (SV) with the equal-weighted Average combination (WA) and the OLS-based combination (OLS). Where the study answers the research question: "Does Shapley-values- based combination improves upon an equal-weights combination and OLS-based combination when forecasting inflation with ARMA models?". This is done using a simulation- and empirical study.

The main results for the simulation study show that in general the SV indeed performs better than WA, whereas OLS performs the worst, when using the loss functions RMSE and MAE. However, the difference between the values in the loss function is minimal. The same results hold when applied to the real data. Moreover, in the simulation study, it can be seen that SV and WA almost equally perform better than the other method when looking at the number of times SV performed better than WA (close to the average). When compared to OLS it can be seen that SV performed approximately 60% of the time better than OLS.

When it comes to inflation, time can play a major role. The findings of this study contribute to seeing whether it is worth the time to assign specific weights to predictions. This is especially useful in high-pressure work environments. Therefore, making it a valuable study in the literature.

Nevertheless, there are potential limitations to this study that must be considered for future research. One of them is that this study uses January 2010 up until December 2022 as the full sample dataset. Where the third period is characterized by high and volatile inflation where the relationship between economic variables tends to be nonlinear. This makes the ARMA model possibly not suitable for this period since it has the assumption that there are linear relationships between the variables. Therefore, SV may be significantly better than WA and OLS for other models besides ARMA when forecasting inflation. Moreover, the average values of the loss functions are used in the simulation study where outliers are not taken into account, which can lead to different results and conclusions.

Furthermore, for further research, it can be interesting to evaluate forecasting performance and pick the best model using the Diebold Mariano test or the Model Confidence Set (MCS) introduced by Diebold & Mariano (1995) and Hansen (2011) respectively. Another thing to take into consideration is that the performance of forecasting models varies by forecasting horizon Makridakis & Hibon (2000). Since this study only considers 1-step-ahead forecasts, it can also be fascinating to look further into this. In addition to this, conducting the simulation multiple times with other parameters is another thing to take into account. This is important for establishing robustness, which is not considered due to time limits.

References

Armstrong, J. S. (2001). Combining forecasts. *Principles of Forecasting: A Handbook for Researchers and Practitioners*, 417–439.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, *20*(4), 451–468.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253-263.

Elliott, G. (2011). *Averaging and the optimal combination of forecasts*. Oxford: University of California, San Diego.

Elliott, R. T. J., G., & Stock, J. H. (2005). Efficient tests for an autoregressive unit root. *Econometrica*, *64*(4), 813–836.

Franses, P. (2023). Shapley-values-based forecast combination. *Econometric Institute Erasmus School of Economics*, 1–26.

Granger, C. W. J., & Jeon, Y. (2004). Forecasting performance of information criteria with many macroeconomic time series. *Empirical Economics*, *29*(2), 231–251.

Granger, C. W. J., & Newbold, P. (1986). Forecasting economic time series. In C. W. J. Granger & P. Newbold (Eds.), *Forecasting economic time series* (pp. 165–190). Academic Press.

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*(2), 197–204.

Guerrón-Quintana, P., & Zhong, M. (2017). Macroeconomic forecasting in times of crises. *Working Paper*(17-21).

Hansen, L. P. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hendry, D. F. (2002). Forecast failure, expectations formation, and the lucas critique. *Econometric Reviews*, *21*(4), 331–349.

Makridakis, S., & Hibon, M. (2000). The m3-competition: Results, conclusions, and implications. *International Journal of Forecasting*, *16*(4), 451–476.

McCracken, M. W., & Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, *34*(4), 574–589.

Moffat, I. U., & David, A. E. (2016). Modeling inflation rates in nigeria: Box-jenkins' approach. *International Journal of Mathematics and Statistics Studies*, *4*(2), 20–27.

Newbold, P., & Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society. Series A (General)*, *137*(2), 131–165.

Nyoni, T. (2019, 02). Predicting inflation in senegal: An arma approach.

Nyoni, T., & Nathaniel, S. P. (2019). *Modeling rates of inflation in nigeria: An application of arma, arima, and garch models* (Tech. Rep. No. Paper No. 91351). Munich University Library – Munich Personal RePEc Archive (MPRA).

Oseni, I., Olasunkanmi, Salihu, M., & Olalekan. (2017, 12). Inflation rate forecasting: An application of autoregressive moving average model (arma).

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, *2*, 307–317.

Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, *71*(3), 331–355.

Stock, J. H., & Watson, M. W. (2007). Why has u.s. inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*(1), 3–33.

Stock, J. H., & Watson, M. W. (2008). Phillips curve inflation forecasts. *NBER Working Paper Series*(14322).

Stovicek, K. (2007). Forecasting with arma models: The case of slovenia inflation. *Bank of Slovenia*, 23–56.

Yang, Y. (2004). Combining forecast combinations. *International Journal of Forecasting*, *20*(1), 69–80.

A Programming code

This paper uses two different R codes, which are specifically designed for this research, named "SimulationARMA" and "RealARMA".

A.1 SimulationARMA

The first code that is used is the SimulationARMA code. This code is used for Section 5 where it first creates a function of the process that is described in 5.2, 5.3, and 5.4 which repeats itself 1000 times using a different seed. In the function, a data-generating process is created and split into three parts. Part 1 is used to fit 5 different ARMA models. Hereafter, the forecasts are generated on this. These forecasts are used to determine the Shapley weights. To find these weights the code first finds all possible combinations of the forecasts and assigns its R^2 values and creates the subsets $R^2(S \cup \{j\})$ and $R^2(S)$, which is used in the formula 3. Hereafter, the forecasts of weights determined by Ordinary Least Squares (OLS), weights determined by Shapley values, and the equal-weighted average are established with part 2, and their corresponding loss functions with part 3. Where the mean of the loss functions of the 1000 repeats is used for the end result. Moreover, the code also provides the number of times SV performed better than WA and OLS. It has to be noted that the code provides the values for the loss functions: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Deviation (MAD). However, MAD is not used in this paper due to RMSE and MAE being more suitable for the purpose of this paper.

A.2 RealARMA

The second code that is used is the RealARMA code. This code is used for Section 6 and is similar to the SimulationARMA code. However, instead of using a data-generating process, it uses real data. Therefore, the code is only repeated once and does not use the mean of loss functions.