Combining forecast with the use of the bias from the models

by

Rogier van de Kamp

Student ID: 426029

Thesis supervisor: J.A. van Oorschot

Second assessor: prof. dr. P.H.B.F Franses

A thesis submitted in partial fulfilment for the degree of

Bachelor of Science in Econometrics & Operations Research

at the

Erasmus School of Economics

Erasmus Universiteit Rotterdam

Date: 6 july 2019

Abstract

This paper will take an extensive look at the conditionally optimal weights with shrinkage (COWS) and the bias corrected optimal weights with shrinkage (BC-OWS) introduced by Gibbs and Vasnev (2018). This paper will select a new way to calculate the variables for the weights and it will compare the weights with the use of the US inflation and the Mexican inflation which has gone trough hyperinflation to proof that these weights will not outperform the simple equal weights when there is hyperinflation.

Contents

1 Introduction **3**

2 Data **4**

3 Methodology **6**

3.1 Weights 6

3.2 Models 8

4 Results **10**

4.1 US 10

4.2 Mexican 13

4.3 Comparison 16

5 Conclusion and discussion **17**

A Appendix **19**

1 Introduction

Researchers have always tried to find the best way to forecast the performance of an economy. They do so because if you can forecast an economy correctly you can foreshadow when the next economic crisis will appear or when an economy will restore from a crisis and you will be able to make the most profit. That is also why a lot of different models have been introduced. Hibon and Evgeniou (2005) noted that it differs per situation which individual model performs best. Since it is hard to predict which forecast will perform best it seems that combining the forecasts could be benificial.

The first combination of models was introduced by Bates and Granger (1969). Since then multiple papers have tried to find the best way to make a combination of the forecasts. Timmermann (2006) noted that most of the time simple forecast combinations, like equal weights, are better at forecasting than the more advanced methods. Since this is still poorly understood people still try to find more advanced methods to give weights to the forecasts. With these combinations they try to outperform the equal weights. Gibbs and Vasnev (2018) found that most of these more advanced methods do not take into account that most models are assumed to be unbiased which cannot be applied in a lot of cases. That is why they propose weights which will include the bias.

One of these weights is the conditional optimal weight with shrinkage (COWS). In principle the bias of each individual forecast is calculated and is then used to construct the weights. These weight will then shrink towards equal weights via a parameter named α, this will be further discussed in section 3.1.

Besides the COWS there will also be looked at a different weight which is also introduced by Gibbs and Vasnev (2018). This is the bias-corrected with optimal weight and shrinkage (BC-OWS). Gibbs and Vasnev (2018) have shown that using bias correction for individual models does not seem to work. However, combining them does seem to give a reasonable forecast, but the variable used to calculate the bias is fixed. But it is not necessarily the best variable to calculate the bias. The bias could be different for different models and it could even be different over time. This is why I will calculate the bias for each model in a different way. I will use the general-to-specific method, which has been developed over a couple of decades, largely by David Henry and his coworkers according to Lütkepohl (2007), to choose which variable will be used to construct the bias.

Timmermann (2006) mention that most weights are being calculated via a backward-looking approach. Gibbs and Vasnev (2018) notice that this approach is empirically not very effective. Which brings them to use a forward-looking approach. This is why this paper will also use a

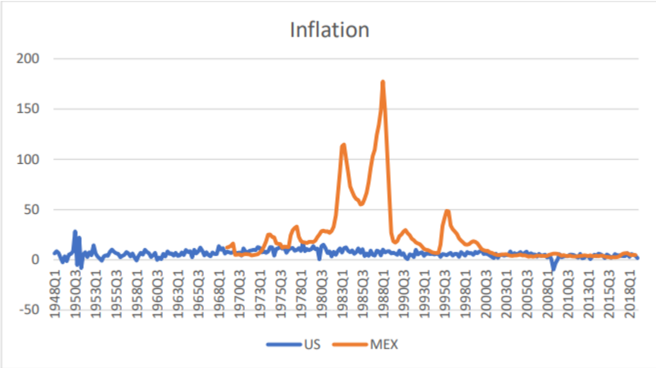

Figure 1: US and Mexico Inflation historical data

forward looking approach to calculate the weights, this will be further discussed in section 3.1. Equal weights have been a benchmark for testing different weights in most of the literature. That is why I will also compare the combined forecasts and the individual forecasts with equal weights.

The performance of an economy is normally measured through the GDP. However, there are more measurements to look at the performance of an economy. An interesting measurement for the economy is inflation. Since historically inflation has been used a lot to test the performance of models (e.g. Bollerslev (1986), Stock and Watson (1999)) and it is very hard to predict, according to Gibbs and Vasnev (2018). That is why they use inflation to test their weights. They use the US inflation from the second quarter of 1947 until the first quarter of 2017. During this period the inflation has been pretty steady except for a small time period around 1950 where it exceeded 25 percent. After this rise it never got above 20 percent again[1]. That is why I will use a data set which contains hyperinflation to compare the results of the COWS and the BC-OWS and see if it still outperforms equal weights. To do so the inflation in Mexico will be used, since Mexico went through hyperinflation during most of the 1980s. The information about inflation from Mexico has been measured from the first quarter of 1969 so there should be enough data available to construct the models and have a long enough out of sample period left to compare the weights.

2 Data

This paper will use data of the inflation of the US and Mexico on a quarterly basis. The data of the US will be used because this is a data set with enough variables to investigate the models properly (Gibbs and Vasnev (2018)). The data of Mexico will be used since there seems to be enough data to get decent models and it has gone through hyperinflation which has been most

[1]Information gotten from https://tradingeconomics.com/united-states/inflation-cpi

extreme in the early 1980's and the end of the 1980's[2]. The difference between the US inflation and the Mexican inflation can be seen in Figure 1. In this graph it can be seen that although Mexico has gone through hyperinflation during the 1980's it has never had deflation like the US had during the financial crisis in 2009.

The first data set will be of the US and is retrieved from the Federal Reserve Bank of Philadelpia real time macroeconomic data set which is described in Croushore and Stark (2001). This data set start in the first quarter of 1948 and ends in the first quarter of 2019. In this data set the same measures used by Gibbs and Vasnev (2018) will be used. These are the quarterly nominal personal consumption expenditures, the total real GDP and unemployment rate. The quarterly nominal personal consumption expenditures will be used to construct the inflation according to equation 1. In which π_t is the inflation in period t and p_t the consumption expenditures.

$$\pi_t = ln\left(\frac{p_t}{p_{t-1}}\right) \cdot 400 \tag{1}$$

The total real GDP and unemployment rate will be used to construct 5 different explanatory variables. Three variables will be constructed with the real GDP and two with the unemployment rate. The first variable will be constructed as the GDP growth. This will be constructed by taking the log-difference of the GDP. The second variable will be an output gap. Which will be constructed with the use of the standard Hodrick Prescott (HP) filter introduced by Hodrick and Prescott (1997)[3]. The HP will be made with the use of eviews, where a λ of 1600 will be used, since this is the appropriate amount for quarterly observations[4]. The third variable will be a GDP growth gap. This growth gap will be constructed by taking the difference between the GDP growth and the maximum GDP growth over the past twelve observations. The fourth variable will be the unemployment rate in levels. The fifth variable will be a unemployment rate gap. This will be the difference between the unemployment rate and the maximum unemployment rate over the past 12 observations. The bias for the models will also be evaluated with each other, that is why I will normalize[5] the data with the use of equation 2 with the exception of the inflation.

[2]Most of the research towards the Mexican hyperinflation has been done why it wasn't recovering after the early 1980's (for example Dornbusch et al. (1988)). I have not been able to find much literature after the hyperinflation at the end of the 1980's.

[3]This paper is originally from 1981 but finally came out in its final version in 1997 the only difference is the tables added in the appendix.

[4]*http://www.eviews.com/help/helpintro.html#page/content/series-Hodrick-Prescott_Filter.html*

[5]https://medium.com/@rrfd/standardize-or-normalize-examples-in-python-e3f174b65dfc

	Mean	St. Dev.	Min	Max
US	6.235	3.841	-9.79	28.31
MEX	23.045	31.072	2.27	177.44

Table 1: Properties of the data

$$x_{new} = \frac{x - min(x)}{max(x) - min(x)} \tag{2}$$

The second data set is from Mexico and is obtained via the Organization for Economic Cooperation and Development (OECD). This data set goes from the first quarter of 1969 until the fourth quarter of 2018. In this data set the Consumer Price Index (CPI) will be used as the inflation[6]. The quarterly GDP growth provided by the OECD will be used to construct the GDP growth gap. The real GDP is not available for every quarter, only for the entire year. Therefore, the GDP growth per quarter will be used to construct the real GDP per quarter. However, there are some differences between the calculated values, so the real annual values. That is why the real GDP will be annually adjusted.

The Mexican inflation has some different properties in comparison with the US inflation. In Table 1 it can be seen that the standard deviation of the Mexican inflation is much higher than the standard deviation of the US inflation. Therefore, it may be expected that the models with the Mexican data set will have a higher Root Mean Squared Forecast Error (RMSFE) and a higher bias in comparison to the RMSFE and the bias from the models which use the US data set.

3 Methodology

In the first section of this chapter the weights and how they are constructed will be explained. In the second section of the chapter it will be explained which models are being used and how to construct the models and how to retrieve the values for the weights.

3.1 Weights

Optimal weights have been constructed since Bates and Granger (1969) started constructing them. They argue that the model with a lower MSE should be given more weight. Throughout the literature constructing the weights is usually done with the use of the unconditional variance

[6] A explanation for the use of CPI as inflation https://data.oecd.org/price/inflation-cpi.htm

of the errors of the forecast models. In these cases it is assumed that the model which is being used is unbiased. However, in a lot of cases this is not true. That is why this paper assumes that there is a bias. When this bias is calculated it can be used to construct optimal weights which are conditioned on some extra information.

The weights which are being used in this paper are the equal weights, the conditional optimal weights with shrinkage (COWS) and bias-corrected optimal weights with shrinkage (BC-OWS). The COWS and BC-OWS both need the bias of the individual models to construct the weight. The bias will be constructed via equation 3. With $e_{i,t}$ the error of the i'th model in time period t, c_i a constant, x_t the explanatory variable and ξ the error term.

$$e_{i,t+4} = c_i + \beta_i x_t + \xi_{i,t+4} \tag{3}$$

The explanatory variable will be chosen via the General-to-specific approach. I will use this approach in the following manner. First start by doing a regression on all the explanatory variables in x_t. In this regression there will be searched for the variable with the highest p-value and it will be removed from x_t. Then, a regression with the reduced x_t will be done and again the variable with the highest p-value will be removed. This will be done until there is only one variable left. This will be done for every individual observation since the variable which will explain the most of the bias of the models might change over time. After this procedure a new $e_{i,t+4}$ will be estimated which will be used to construct the COWS and the BC-OWS.

The conditional optimal weight (COW) is introduced by Gibbs and Vasnev (2018) and is written down in equation 4. The bias $\widehat{\mathbf{b}}_T$ that is used here is the bias which is calculated in equation 3 and can be formulated as $\widehat{\mathbf{b}}_{T+4} = E[e_{T+4}|I_T]$ where I_T is an information matrix consisting of all the explanatory variables up to time period T. The $\widetilde{\Sigma}_\xi$ that will be used is the variance co-variance matrix from the realized errors of each model their bias which can be formulated as $\widetilde{\Sigma}_\xi = var(\xi_{T+4}|I_T)$.

$$\widehat{\mathbf{w}}_{COS}(I_T) = \frac{[\widetilde{\Sigma}_\xi + \widehat{\mathbf{b}}_{T+4}\widehat{\mathbf{b}}'_{T+4}]^{-1}\iota}{\iota'[\widetilde{\Sigma}_\xi + \widehat{\mathbf{b}}_{T+4}\widehat{\mathbf{b}}'_{T+4}]^{-1}\iota} \tag{4}$$

Equation 4 shows how the COW is calculated. However, I want to use the COWS which means that a shrinkage method will be used. The shrinkage method that will be used is shown in equation 5. Where Σ_0 is an identity matrix to pull the weights towards equal weights. There are several possibilities to chose for $\widehat{\Sigma}_\xi$ to shrink. I will use the same part to shrink as Gibbs and Vasnev (2018) did, which is $\widetilde{\Sigma}_\xi + \widehat{\mathbf{b}}_{T+4}\widehat{\mathbf{b}}'_{T+4}$. Shrinking this entire part will have the most effect on pulling the weights towards equal weights.

$$\widetilde{\Sigma}_{\xi} = \alpha\Sigma_0 + (1-\alpha)\widehat{\Sigma}_{\xi} \tag{5}$$

The bias correction that will be used is shown in equation 6. This is the same approach to correct for the bias as in Gibbs and Vasnev (2018). In equation 6, $\widehat{y}$ is the forecast of the model and $\mathbf{b}_T$ will be the bias e_{t+4} as calculated for that model in the same way as described above.

$$\widetilde{y}_{t+4} = \widehat{y}_{t+4} + \mathbf{e}_{t+4} \tag{6}$$

When the forecasts have been corrected for the bias I will use the classical result for the optimal weights as discussed in Elliott (2011) and is shown in equation 7 where $\Sigma_e = var(\mathbf{e})$. I will also shrink this equation towards the equal weights by using Σ_e as $\widehat{\Sigma}_{\xi}$ in equation 5.

$$\mathbf{w}_{OW} = \frac{\Sigma_e^{-1}\iota}{\iota'\Sigma_e^{-1}\iota} \tag{7}$$

3.2 Models

The models that will be used in this paper are all the models used in Gibbs and Vasnev (2018) which are shown in Table 2. The AO model will be usedas a benchmark which is based upon the naïve random walk model from Atkeson and Ohanian (2001). Stock and Watson (2007) show that the model has a decent forecast for a longer horizon after 1984 but not as well with a shorter horizon and before 1984. This model simply takes the average of the previous four quarters of inflation to make a prediction for the next four quarters. The standard ARMA(p,q) model is shown in equation 8. To get an AR model we put q=0 and p to degree of the AR model. Since a four quarter ahead forecasts is required the forecast is made iteratively.

$$y_t = c + \sum_{i=1}^{p} \beta_i y_{t-i} + e_t + \sum_{i=1}^{q} \theta_i e_{t-i} \tag{8}$$

The Phillips curve models are constructed in the same way as in Gibbs and Vasnev (2018), which means they are bi-variate and have two lags for inflation and two lags for an explanatory variable. This is shown in equation 9 with y_t as the inflation in period t and x_t as an explanatory variable. The forecasts for the Phillips curves also have to be made iteratively.

$$y_t = c + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 x_{t-1} + \beta_4 x_{t-2} + e_t \tag{9}$$

The Direct Forecasts (DF) are also constructed in the same way as in Gibbs and Vasnev (2018), which means it is an OLS regression with an explanatory variable being regressed on inflation which is four-quarter-ahead. This model is shown in equation 10. In which $\pi_t + 4$ is the inflation in period t+4, x_t the explanatory variable and ϵ the error term.

$$\pi_{t+4} = c + \beta x_t + \epsilon_{t+4} \tag{10}$$

Univariate	Phillips Curve	Direct Forecasts
AR(1)	PC Output Gap	DF Output Gap
AR(2)	PC Unemployment Gap	DF Unemployment Gap
AR(4)	PC GDP Growth	DF GDP Growth
ARMA(1, 1)	PC Growth Gap	DF Growth Gap
ARMA(4, 4)	PC Unemployment Rate	DF Unemployment Rate
AO	VAR Al	

Table 2: Models

All the models mentioned above will be programmed in Python with the help of Statsmodels from Seabold and Perktold (2010).
The weights for the combined forecasts have to be constructed for each time period, this will be done in accordance with Gibbs and Vasnev (2018) with the following procedure.

1. Each of the 17 models is estimated
2. Each of the 17 models is used to construct a four-quarter-ahead forecast.
3. The bias will be estimated according to equation 3 with the available real-time forecast errors for each of the 17 models.
4. Equation 3 is used to predict the expected forecast errors for the four quarter ahead forecast. To construct a bias $\widehat{e}_{i,t+4}$ for each of the 17 models.
5. These errors and biases are used to construct the weights for the combined forecast methods.

To create a four quarter ahead forecast we need to estimate the models on a sufficient amount of data. That is why for the US the first models will be estimated on data from 1948Q1 until 1965Q1. So that the first four quarter ahead forecast can be made for 1966Q1. In the Mexican data set the models will be estimated on the data from 1969Q1 until 1979Q2 so the first forecast will be in 1980Q2.

There is a second period needed to estimate the bias for every model. Since estimating it on one variable will not result in a reliable estimation. That is why for the US data set the errors from 1966Q1 until 1968Q4 will be used to get a proper estimation for the first bias model. The first bias estimation will be made for 1969Q4. In the Mexican data set the first bias model will be estimated on the errors from 1980Q2 until 1983Q1 so the first estimation will be made in 1984Q1.

This results in an out of sample period from 1969Q4 until 2019Q1 for the US data and from 1984Q1 until 2018Q1 for the Mexican data set. The out of sample period for the Mexican data set will also be divided into 2 parts one from 1985Q1 until 1999Q4 and the other from 2000Q1 until 2018Q4. This is to include hyperinflation in the first time frame and exclude it from the second. Since the hyperinflation makes the data more volatile it could be expected that the models will perform worse in comparison to the part without hyperinflation in terms of RMSFE. Since the errors are expected to be larger it can also be expected that the bias will be greater and the variance of the bias its errors to be greater. This most likely has the most effect on the BC-OWS since the individual bias will be added. In the COWS this will have less of an effect since the bias will most likely be bigger for all the models and this may cancel each other out while constructing the weights.

4 Results

In the first section of this chapter the results of how the forecasts perform using the data from the US will be discussed. The second section will be used to discuss the forecasts made with the Mexican data. Finally, in the third section both the results of the forecasts with the US data and the results of the forecasts with the Mexican data will be compared.

4.1 US

The results of all the individual models and the combination of models are shown in Table 3. The RMSFE are calculated for the individual models in the period 1966Q1-2019Q1 and for the combined forecasts 1970Q4-2019Q1. This is because the first weights will only be available in 1970Q4 since the first time they are being calculated is in 1969Q4. With the equal weights it was possible to take the sample from 1966Q1-2019Q1 that is why I wrote that number down in the parenthesis so it could be used for calculating the Relative RMSFE for the individual models. The average bias is the average bias over the calculated future bias for individual models in the time frame 1970Q4 until 2020Q1 (so the first value is calculated at 1969Q4 but is being used for 1970Q4) and for the combined forecasts it is from 1976Q1 until 2020Q1 this is taken from

Model	Most common Bias predictor	Average Predicted bias	RMSFE	Rel. RMSFE (AO)	Rel. RMSFE (Equal weights)
Phillips Curve					
PC GDP gr.	GDP gr. gap	1.19	2.443	1.245	1.033
PC GDP gr. gap	GDP gr. gap	1.25	2.439	1.123	1.031
PC Unemp.	Unemp. rate	0.72	2.569	1.183	1.086
PC Unemp. gap	GDP gr. gap	0.99	2.390	1.100	1.011
PC Output gap	GDP gr.	1.37	2.621	1.207	1.108
PC VAR all	GDP gr.	0.87	2.591	1.193	1.096
Direct Forcast					
DF GDP gr.	GDP gr.	1.04	2.655	1.222	1.123
DF GDP gr. gap	GDP g.	1.13	2.573	1.185	1.088
DF Unemp.	GDP gr.	0.62	2.626	1.209	1.110
DF Unemp. gap	GDP gr.	0.84	2.625	1.209	1.110
DF Output gap	GDP gr.	1.40	2.874	1.323	1.215
Univariate					
AO	OUTPUT gap	0.20	2.172	1.000	0.918
AR(1)	GDP gr.	1.31	2.657	1.223	1.123
AR(2)	GDP gr.	1.09	2.394	1.102	1.012
AR(4)	GDP gr.	1.72	2.698	1.242	1.141
ARMA(1,1)	Output gap	1.23	2.151	0.990	0.910
ARMA(4,4)	Output gap	0.77	2.251	1.036	0.952
Combined					
Equal weights	GDP gr. gap	0.93	2.381(2.365)	1.089	1.000
COWS	GDP gr. gap	-0.06	2.142	0.986	0.900
BC-OWS	GDP gr. gap	0.51	2.106	0.970	0.885

Table 3: The results over the full sample for the individual and combined weights on the US data set.

1976Q1 so that there is some historical data on which the bias model can be calculated.

In Table 3 we can see that almost none of the individual models seems to outperform the AO, except for the ARMA(1,1). However, the ARMA(1,1) improvement is not a statistically different, it also seems to perform worse on 53% of the forecasts. The AO, ARMA(1,1) and the ARMA(4,4) are all performing better than the equal weights. Hibon and Evgeniou (2005) noted that combined forecasts will never outperform the best single forecast for each observation. Since there is quite some difference in the performance in terms of RMSFE of the individual forecasts it can be expected that some of the forecasts will perform better in comparisson to equal weights. Here also lies the problem of combining forecasts chosing the best model for each observation. It seems that the COWS and the BC-OWS are doing a better job at chosing the right individual model since they seem to outperform all the individual models. Note that none of these results seem to be significantly different since the standard deviation of the RMSFE of

	OUTPUT gap	GDP gr. gap	GDP gr.	Unempl.	Unemp. gap
Philips Curve					
PC GDP gr.	27	129	20	6	16
PC GDP gr. gap	18	141	24	0	15
PC Unemp.	8	59	0	130	1
PC Unemp. gap	16	136	8	15	23
PC OUTPUT gap	59	44	85	4	6
PC VAR all	26	0	100	72	0
Total PC	154	509	237	227	61
Direct Forecast					
DF GDP gr.	0	7	191	0	0
DF GDP gr. gap	0	8	190	0	0
DF Unemp.	0	5	193	0	0
DF Unemp. gap	0	8	190	0	0
DF OUTPUT gap	34	0	164	0	0
Total DF	34	28	928	0	0
Univariate					
AO	166	10	22	0	0
AR(1)	22	123	7	27	19
AR(2)	26	130	20	6	16
AR(4)	21	126	6	25	20
ARMA(1,1)	136	31	10	2	19
ARMA(4,4)	99	16	18	64	1
Total Univariate	470	436	83	124	75
Combined					
Equal weights	4	129	1	31	12
COWS	46	55	26	50	0
BC-OWS	1	110	66	0	0
Total Combined	51	294	93	81	12

Table 4: Times bias used

most models is around 2. The COWS and BC-OWS also seem to outperform the equal weights. The COWS perform better in 60% of the forecasts and the BC-OWS perform better in 58% of the forecasts. The Average bias that is being used is positive for all the individual models and none of them seem to be significantly different from 0. This could be due to the fact that for every prediction there is another variable used for the prediction of the bias

Table 4 shows how many times each variable is used for predicting the bias. Here we can see that it mostly depends on which model we use instead of which variable is being used in each model. For the Philips Curves we can see that the GDP growth gap is the most commonly used even when the GDP growth gap is used for predicting the model. Also the Unemployment rate is used the most when it is used to predict the model. The Unemployment rate and Unemployment rate gap are least likely to be chosen as bias. The Direct Forecasts seem to only have the GDP growth as a recurring option for estimating the bias. The Univariate models seem to be more spread out with the main focus on the Output gap and the GDP growth gap. With the Output gap being used for the two models with the lowest RMSFE. In the combined forecasts it seems that the GDP growth is also the most common. However if we look at the COWS we see that the bias for this model seems to be spread out mostly among the Output gap, GDP growth gap and the Unemployment rate.

4.2 Mexican

Model	Most common Bias predictor	Average Predicted bias	RMSFE	Rel. RMSFE (AO)	Rel. RMSFE (Equal weights)
Philips Curve					
PC GDP gr.	GDP gr.	-1.70	20.179	1.370	1.098
PC GDP gr. gap	GDP gr.	-1.63	20.688	1.404	1.125
PC Output gap	GDP gr.	1.01	18.655	1.266	1.015
PC VAR all	GDP gr. gap	0.14	21.176	1.437	1.152
Direct Forecast					
DF GDP gr.	GDP gr.	6.36	27.323	1.855	1.486
DF GDP gr. gap	GDP gr.	0.75	25.317	1.718	1.377
DF Output gap	GDP gr. gap	15.60	27.757	1.884	1.510
Univariate					
AO	Output gap	0.55	14.733	1.000	0.801
AR(1)	Output gap	-1.55	13.741	0.933	0.747
AR(2)	GDP gr.	1.88	23.489	1.594	1.278
AR(4)	Output gap	-4.26	16.735	1.136	0.910
ARMA(1,1)	Output gap	-1.90	14.367	0.975	0.781
ARMA(4,4)	Output gap	3.92	12.857	0.873	0.699
Combined					
Equal weights	GDP gr.	-4.16	16.931(18.385)	1.248	1.000
COWS	GDP gr.	-1.30	17.430	1.183	1.029
BC-OWS	Output gap	3.70	17.870	1.213	1.055

Table 5: The results over the full sample for the individual and combined weights on the Mexican data set.

The results of all the individual models and the combination of models of the Mexican data set are shown in Table 5. The RMSFE are calculated for the individual models in the period 1980Q2-2018Q4 and for the combined forecasts 1985Q1-2018Q4. This is because the first weights will only be available in 1985Q1 since the first time they are being calculated is in 1984Q1. With the equal weights it was possible to take the sample from 1980Q2-2018Q4, which is why I wrote that number down in the parenthesis so it could be used for calculating the Relative RMSFE for the individual models. The average bias is the average bias over the calculated future bias for individual models in the time frame 1985Q1 until 2019Q4 (so first value is calculated at 1984Q1 but is being used for 1985Q1) and for the combined forecasts it is from 1990Q2 until 2019Q4 this is calculated from 1990Q2 so that there is some historical data on which the bias model can be calculated.

	Output gap	GDP gr.	GDP gr. gap
Philips Curve			
PC GDP gr.	8	132	0
PC GDP gr. gap	3	134	3
PC Output gap	16	124	0
PC VAR all	14	49	77
Total PC	41	439	80
Direct Forecast			
DF GDP gr.	34	100	6
DF GDP gr. gap	35	105	0
DF Ouput gap	55	0	85
Total DF	124	205	91
Univariate			
AO	134	6	0
AR(1)	140	0	0
AR(2)	19	121	0
AR(4)	140	0	0
ARMA(1,1)	140	0	0
ARMA(4,4)	140	0	0
Total Univariate	713	127	0
Combined			
Equal weights	20	99	0
COWS	43	10	66
BC-OWS	119	0	0
Total Combined	182	108	66

Table 6: Times bias used

In Table 5 we can see that the Philips curves and the direct forecasts seems like terrible options to predict the inflation in comparison to the simple AO model. Only the AR(1), ARMA(1,1) and ARMA(4,4) seem to perform better. Since the standard deviation of RMSFE of the AO is 25.47 those models are not significantly better. The AR(1) model only outperforms the AO in 35% of the time which suggest that the AO performs really badly in a couple of observations. Which seems logical when looking at Figure 1. When the inflation is really high it suddenly drops a lot which makes the RMSFE of the AO really big. The same is true for the ARMA(1,1) model which performs worse than the AO in 71% of the time and for the ARMA(4,4) who performs worse than the AO in 55% of the time. When looking at the combined forecasts neither of the forecasts seem to outperform the equal weights based on the RMSFE. The BC-OWS perform worse than the equal weights in 55% of the time and the COWS perform worse on 58% of the time. However, the COWS do outperform the BC-OWS on 58% of the time. The average predicted bias seems to be pretty close towards 0 for all of the models. However, most of the time the bias which is projected is very positive or very negative. This can be seen in the variance which shows that for most of the individual models it is around 200 and for the DF Output gap it is even over 1000. For the equal weights and the COWS the variance is about 60 and for the BC-OWS it is about 250.

Table 6 shows how often each bias is being chosen as a predictor via the general-to-specific approach. Here we can see that the predictor depends on what kind of model it is. For the Philips curves it is mostly the GDP growth the same is true for the Direct forecasts. The univariate on the other hand mostly uses the Output gap to construct the bias. We can see that for the individual models the GDP growth gap is barely being used. It is just the main predictor in 2 (PC VAR all and DF Output gap) out of the 13 models and in 9 of the models it isn't even used once. If we look at the models which are performing the best, according to the RMSFE in Table 5, all of them have the Output gap as the most or even only used bias predictor. Looking at the combined forecasts we see that every one of the combinations has another preferred bias the equal weights has GDP growth the COWS has GDP growth gap and the BC-OWS has the Output gap.

Table 7 shows us the RMSFE and the percentage that the weights have a better prediction than the equal weights prediction. This data had been divided from 1985Q1 until 1999Q4 and from 2000Q1 until 2018Q1. In Table 7 we can see that the BC-OWS performs worse than the equal weights in both of the time frames for both the percentage and the RMSFE. The COWS does perform better than the equal weights but only in the time frame where the inflation is more stable.

	1985Q1-1999Q4		2000Q1-2018Q4	
	RMSFE	% Outperform vs EW	RMSFE	% Outperform vs EW
Equal Weights	24.096		11.274	
Cows	26.138	23%	10.555	57%
BC-OWS	24.825	43%	12.379	46%

Table 7: Comparing two time frames of the Mexican data set

4.3 Comparison

In Table 8 we see the results of all the combined forecasts for the US database and the Mexican database with only the use of the GDP growth, GDP growth gap and the Output gap. The selection for these three variables is just so we can compare the data sets in the most appropriate way. The RMSFE is calculated from 1986Q1 until 2018Q4 and the bias is the average bias calculated from these models in the time period of 1989Q2 until 2018Q4.

If we compare these results we can see that the RMSFE of the Mexican data set is much higher which was expected by the higher variance of the Mexican inflation as shown in Table 1. Furthermore, we can see in Table 8 that in both data sets the RMSFE is the lowest for the Equal weights. However, if we look at Table 3 the RMSFE for BC-OWS and COWS are lower when all the variables are included. If we look at the full US sample with only the GDP growth, GDP growth gap and output gap included the RMSFE for equal weights is 2.457, for COWS 2.101 and for BC-OWS is 2.095. Note that the RMSFE is a lot lower for COWS and BC-OWS this shows that COWS and BC-OWS perform better in the time frame 1970Q4 until 1985Q4 with only three variables. If all the variables are included in the time frame from 1986Q1 until 2019Q1 the RMSFE of the equal weights is 2.056. This seems to be much higher than the 1.592 from the equal weights with only 3 variables. The COWS have a RMSFE of 1.889 with all the variables in the time frame from 1986Q1 until 2019Q1 and for the BC-OWS the RMSFE is 1.860. This shows us that when the unemployment rate is included the equal weights seem to perform much worse in the time frame from 1986Q1 until 2019Q1 but the COWS and BC-OWS do not differ that much. This most likely comes due to the fact that the volatility of inflation the GDP growth and the GDP growth gap in the period from 1986Q1 until 2019Q1 is lower and volatility the unemployment rate is higher in this period. Which most likely have a negative effect on the models where it is being used as a variable for estimating the inflation. Which seems to have more effect on the equal weights than on the COWS and BC-OWS.

	US		MEX	
	Bias	RMSFE	Bias	RMSFE
Equal Weights	1.794	1.592	-4.576	17.060
Cows	0.178	1.847	-1.297	17.356
BC-OWS	0.546	1.832	3.702	17.425

Table 8: Comparing the US results with the Mexican results

5 Conclusion and discussion

If the results from the COWS and BC-OWS are compared to the results from Gibbs and Vasnev (2018) it does seem that the general-to-specific approach for calculating the bies for each model by each observation does provide better forecasts since they have a lower, relative to equal weights, RMSFE.

Furthermore, it can be seen that the COWS and BC-OWS do not seem to outperform the equal weights when there is hyperinflation since the RMSFE seems to be higher. However, if we look at the results from the US data set, it shows us that when extra models are added the equal weights seem to performing worse and the COWS and BC-OWS seem to perform about the same. Which would indicate that a lack of variables and models could be the reason why the COWS and the BC-OWS do not outperform the equal weights with the use of the Mexican data set. Although all the weights perform worse in the time frame from 1986Q1 until 2019Q1 when the unemployment variables are included, the equal weights perform better over the entire period when they are included. Which shows that these variables do have explanatory value on inflation when the inflation is more volatile. It could be investigated if the COWS and BC-OWS will outperform the equal weights when there is hyperinflation when an unemployment measurement is used.

References

Atkeson, A. and Ohanian, L. E. (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, 25(1):2–11.

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of econometrics*, 105(1):111–130.

Dornbusch, R., Fischer, S., et al. (1988). *Inflation Stabilization: The Experience of Israel, Argentina, Brazil, Bolivia, and Mexico.* mit Press.

Elliott, G. (2011). Averaging and the optimal combination of forecasts. *Manuscript, Department of Economics, UCSD.*

Gibbs, C. and Vasnev, A. L. (2018). Conditionally optimal weights and forward-looking approaches to combining forecasts. *Available at SSRN 2919117.*

Hibon, M. and Evgeniou, T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, 21(1):15–24.

Hodrick, R. J. and Prescott, E. C. (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, pages 1–16.

Lütkepohl, H. (2007). General-to-specific or specific-to-general modelling? an opinion on current econometric terminology. *Journal of Econometrics*, 136(1):319–324.

Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference.*

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2):293–335.

Stock, J. H. and Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, 39:3–33.

Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting*, 1:135–196.

A Appendix

The codes in the ZIP file are:

AO The calculation of the AO.

ARMA The calculation of the AR and ARMA models.

BC-OWS The calculation of the BC-OWS.

bias The calculating the average bias.

countbias Counting how many times each variable is used for calculating the bias.

COWS The calculation of the COWS.

DF The calculation of the Direct Forecasts.

equal The calculation of the equal weights.

PC The calculation for the phillips curves.

PCALL The calculation for the phillips curves with all the variables.

Note that these are the codes used for the US data set the only difference with the Mexican data set is the values when to start calculating and which variables to use.