ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrie en Operationele Research

Conditionally Optimal Weights and Forward-Looking Approaches to Combining Forecasts

Author: A. van der Wal

Student number: 380344

Supervisor: J.A. Oorschot

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

Date final version: July 7, 2019

Abstract

As forecasting remains a major topic of interest in economics, combining forecasts has been described extensively in literature. Gibbs and Vasnev (2017) use a forward looking approach to combining forecasts, in which predictions of future forecast errors are made which are used to construct combination weights. In this paper, this method is further explored. Besides unconditional optimal weights and bias-corrected forecasting models, two combination models using conditionally optimal weights are constructed. Specifically, we investigate whether it is beneficial for forecasting to use these conditionally optimal weights. US inflation data is used to examine the forecast performance of the different individual models, bias-corrected models and the forecast combination models. The forecast performances are compared with those of two parsimonious benchmark models, namely the Naive forecasting model as described by Atkeson et al. (2001) and the Equal Weights forecasting model, which both have relatively good forecast performance. In contrast to the promising results in Gibbs and Vasnev (2017), we do not find that the forecast combining models outperform the parsimonious forecasting models.

Contents

1 Introduction and problem description

For governments and policymakers, knowing what to expect of macro-economic variables such as GDP growth or inflation is important. However, forecasting such variables remains a puzzle, and although different individual econometric models have been proposed, none of them seem to produce good forecasts consistently. Instead of searching for the individual model with the best forecasting performance, combining forecasts is also topic of research. Although research in combining forecast strategies has been promising, methods using a backward-looking approach have a relatively poor performance. This is partially due to biases that exist in individual forecasts, which makes it difficult to exploit predictable information in conditional forecast errors, which leads to optimal weights that are misspecified. Therefore, such methods do not consistently outperform simple strategies like the Equal Weights strategy of combining forecasts.

It seems logical to investigate whether correcting the biases that are present in the individual forecasts is beneficial for the forecast performance of both the individual models and the forecast combinations. Indeed, when the forecasts are corrected with accurate estimates of the biases, this should leave us with more accurate forecasts, as proven in Gibbs and Vasnev (2017). However, this is exactly the problem: estimates of biases are often noisy, and correcting the forecasts with a contaminated bias results in worse forecasts.

An approach different from the backward-looking method is described in Gibbs and Vasnev (2017), where it is suggested to use a forward-looking approach. In the methods used, predictions of the future forecast errors are made and subsequently combination weights are constructed using these predictions. Rather than using the past forecast performance to construct the combination weights as done in the backward-looking approach, the forward-looking approach uses the expected forecast performance. In practice, they predict the forecast errors of each model and they assign weights such that the expected squared errors of the forecast combination models are minimized.

The paper of Gibbs and Vasnev (2017) will be used as a base for our research. Five different forecast combinations of seventeen different individual model forecasts will be compared, including Equal Weights, unconditional optimal weights, bias-corrected forecasts with unconditional optimal weights and conditionally optimal weights. The weights of the forecast combination of conditionally optimal weights are constructed using a covariance matrix of the bias-corrected forecast errors. In Gibbs and Vasnev (2017), this covariance matrix is assumed to be fixed over time. In addition to Gibbs and Vasnev (2017), the forecasting performance of a second forecast combination strategy with conditionally optimal weights is added. However, in this strategy we allow the covariance matrix of the bias-corrected forecast errors to be time-varying. The variance matrix will be updated using the DCC-GARCH model, as described in Engle (2002).

We use quarterly US inflation rates to compose forecasts. The forecasts of the different models will be compared based on the Root Mean Squared Forecast Error (RMSFE) and the Mean Forecast Error (MFE). The forecast performance of the models proposed will be tested in three different time periods to be able to compare the performance for different volatility levels of inflation. This allows us to compare the models on robustness as well. Specifically, we want to answer the question whether or not the forecast combination strategies of conditionally optimal weights (with or without time-varying variance matrix) outperform other forecast combinations in inflation forecasting, when compared based RMSFE and MFE.

We find that the forecasting accuracy differs greatly among three different subsamples used, which are characterised by differences in volatility levels of the inflation rates. In periods with more volatile inflation rates, predictive accuracy drops which results in higher RMSFE. Also, when inflation is more volatile, the MFE tend to be more biased. Furthermore, as expected, parsimonious benchmark models are still hard to beat, regardless the volatility levels of inflation. Applying bias-correction to forecasts also does not lead to better forecasts, as it remains a puzzle to accurately estimate biasedness. Furthermore, the extended model of conditionally optimal weights with time-varying covariance matrix does not perform better than the conditionally optimal weights with fixed covariance matrix. In contrast to Gibbs and Vasnev (2017), we find that none of the individual or combined forecasting models is able to consistently beat the benchmark models.

Next section contains the data descriptions. Section 3 introduces the econometric models used to create forecasts, the combined forecast strategies and some performance measures. In Section 4 the results will be given, and Section 5 concludes.

2 Data

To investigate the forecast performance of the different models, we use quarterly US inflation rates similar to Gibbs and Vasnev (2017), which we have obtained from the Consumer Price Index (CPI) from the International Monetary Fund (IMF). In their research however, real-time vintage data is used which allows them to know for sure that the data they use at the time a forecast is made, was actually available to the public at that time. For simplicity, we only use the last vintage data, in contrast to their paper. However, to still account for lagged information availability of inflation rates, we always forecast $h = 4$ quarters ahead. We denote time T as the moment the last known inflation rate is available to the public and usable for forecasting, and construct forecasts for time $T + h$. This difference in data usage causes that forecasting results in our paper and the paper of Gibbs and Vasnev (2017) cannot be compared directly.

The time series of quarterly US inflation rates used in our research starts from 1960Q2 until 2019Q1, which gives us a total of 236 observations. In Figure 1, the US inflation rates are shown.

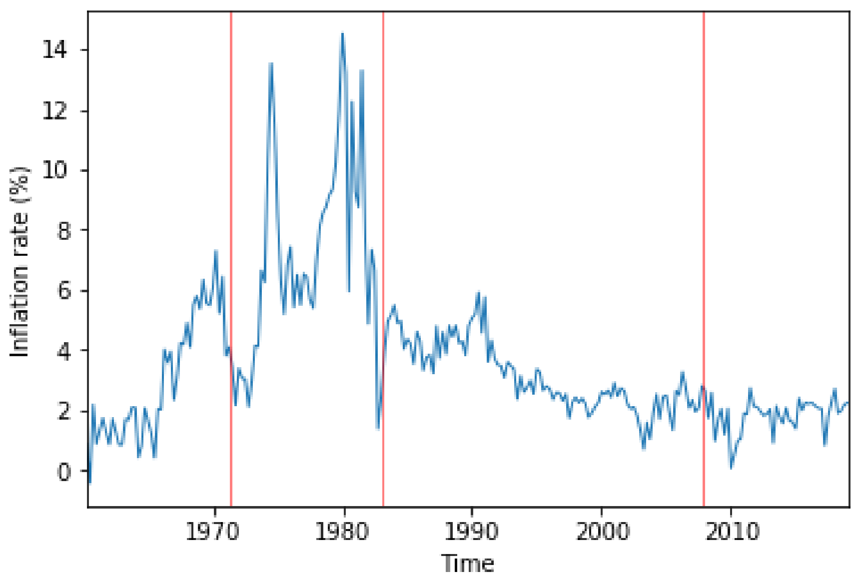

Figure 1: US quarterly inflation rate from 1960Q2 until 2019Q1, split up into the training- and hold-in sample and three forecasting samples

In the inflation time series Figure 1 it can be seen that there are great differences in general inflation patterns over different periods in time, for example in volatility levels. Therefore, for the model estimation in later periods, it might be beneficial to only use recent past observations which show the same patterns to estimate the models and construct forecasts, because when the behaviour of variables change over time, the last observations might give us more relevant information for future behaviour. Therefore, for the estimation of the models, we use a rolling window consisting of $n = 40$ observation. This window contains both the training sample for the estimation of parameters and the hold-in-sample to obtain the initial forecast errors for the calculation of biases. This is in contrast to Gibbs and Vasnev (2017), where an expanding window is used which causes the training- and hold-in sample to grow at each iteration, and which causes the estimation window to also contain observations from periods in which the inflation rates showed different patterns.

Besides the training- and hold-in sample, there is the hold-out sample of length $n = 196$ used for forecasting. This hold-out sample is split up into three different time periods, as can be seen in Figure 1. In this figure, the first period consists of the training and the hold-in sample. The remainder of the sample is split up in three forecasting samples to be able to compare forecast performance in different periods. As inflation rates were more volatile and showed higher peaks between roughly 1965 and 1983 than in the years after, we set the first forecasting sample from 1970Q2 up to 1982Q4. Because we forecast $h = 4$ quarters ahead, the first forecasts are made for 1971Q2, such that in Section 4, the performance measures are given from 1971Q2. The remaining data corresponds to the forecasting sample from Gibbs and Vasnev (2017), and for simplicity we use the same breakpoint as they use, namely at 2007Q3. This gives us a second forecasting sample from 1983Q1 until 2007Q3, and a third forecasting sample from 2007Q4 and 2019Q1. However, because of some differences in the methods used in this paper and the methods in Gibbs and Vasnev (2017), we cannot directly compare the results. In Table 1, the descriptive statistics of the inflation data in the four different samples can be found. This supports our observation that inflation rates were more volatile in the first forecasting sample. Besides the volatility, also the inflation rates itself gradually decreased over the last three decades, ranging between 0.06% and 2.81% in the last sample, close to the target inflation rate of 2% of most central banks.

Table 1: Descriptive statistics of quarterly US inflation data (in percentages)

Sample	Training and hold-in sample	First forecasting sample	Second forecasting sample	Third forecasting sample
Timeframe	1960Q2 - 1970Q1	1970Q2 - 1982Q4	1983Q1 - 2007Q3	2007Q4 - 2019Q1
Mean	2.56	7.13	3.19	1.85
St. dev.	1.81	3.19	1.26	0.58
Min	-0.44	2.10	0.69	0.06
Max	6.34	14.52	7.33	2.81
Observations	40	51	99	46

The descriptive statistics of US inflation in percentages for four different time periods, namely the training and hold-in sample and the three forecasting samples.

Besides the time series for inflation, also time series of the GDP growth rates and unemployment rates are used in the forecasting models. These rates are also obtained from the IMF database. From these variables, the GDP growth gap and unemployment gap are created by taking the difference between the maximum value over the last twelve quarters and the current observation. Furthermore, the output gap is created using a Hodrick-Prescott Filter, as described by Hodrick and Prescott

(1980). For this, the statsmodels function in Python was used. In the next section, the forecasting models will be described.

3 Methodology

First, in Section 3.1, the estimation procedure of the models is explained in detail. In Section 3.2, the econometric models which are used to create individual forecasts are described. The procedure for obtaining the bias-corrected forecasts is described in Section 3.3. The description of the five forecast combination strategies are provided in Section 3.4, and lastly, the combined forecasts are compared using performance measures described in Section 3.5.

3.1 Estimation procedure

For each model, in each iteration, the estimation and forecasting procedure at time T goes as follows:

1. All 40 observations in the rolling window ($T - 39$ up to T) are used to obtain parameter estimates for the models. This is in contrast to the method in Gibbs and Vasnev (2017), where they use all past observations except the last 20 observations in the estimation window. The last 20 observations are used in the next step for the estimation of the bias. However, due to poor forecasting performance when using only the first 20 observations from the rolling window for the estimation of the parameters, we use the whole rolling window for this.

2. When the parameter estimates are obtained, a forecast of the bias is constructed using the last 20 observations of the rolling window, in line with the procedure in Gibbs and Vasnev (2017). The estimation procedure of the bias will be discussed in detail in Section 3.3.

3. Using the parameters estimates and estimated biases, regular and bias-corrected forecasts are constructed for $T + h$ using only the information available at time T.

4. For each forecast, the forecast error and the squared forecast error are calculated for the MFE and the RMSFE. The performance measures of the different models are compared using the tests as described in Section 3.5.

3.2 Individual forecasting models

To create individual h-step-ahead forecasts, Gibbs and Vasnev (2017) propose 17 different econometric forecasting models which we adopt in our research. The models are summarized in Table 2.

An extensive description of the models can be found below.

Table 2: Forecasting models

Univariate	Phillips Curve	Direct Forecasts
Naive model	PC GDP Growth	DF GDP Growth
AR(1)	PC GDP Growth Gap	DF GDP Growth Gap
AR(2)	PC Output Gap	DF Output Gap
AR(4)	PC Unemployment Rate	DF Unemployment Rate
ARMA(1,1)	PC Unemployment Gap	DF Unemployment Gap
ARMA(4,4)	All Variables	

The different models used to make forecasts of US inflation. The models consist of univariate models, models based on the Phillips curve and Direct forecasting models, based on OLS regression.

Naive forecasting model A parsimonious model which is used in Gibbs and Vasnev (2017), is the Naive forecasting model as proposed by Atkeson et al. (2001). The inflation rate of a coming year is assumed to be equal to the average inflation rate of the past year: $y_{T+1} = \frac{y_T + y_{T-1} + y_{T-2} + y_{T-3}}{4} + \varepsilon_{T+1}$. Because of the lagged information availability of inflation, the inflation forecast at time T is estimated as the average of the last four known quarters:

$$f_{T+h} = \frac{y_T + y_{T-1} + y_{T-2} + y_{T-3}}{4} \tag{1}$$

As this simple model proves to be quite a good estimate, this model will be used as a benchmark model for the comparison of the individual forecasting models.

Autoregressive Moving Average forecasting model The Autoregressive Moving Average (ARMA(p,q)) models of order p and q are relatively simple univariate models that are widely used to forecast time series. Forecasts are based on the past p observations and the past q shocks ε. Thus, the model is given as follows:

$$y_{T+h} = \phi_1 y_T + ... + \phi_p y_{T-p+1} + \theta_1 \varepsilon_T + \cdots + \theta_q \varepsilon_{T-q+1} + \varepsilon_{T+h}. \tag{2}$$

For the estimation of the ARMA(p,q) models, the Statsmodels module in Python is used. The Autoregressive (AR(p)) models of order p are restricted ARMA(p,q) models and produce forecasts based on a linear combination of the last p observation. In Equation 2, the parameters $\theta_i = 0$, for $i = 1, \ldots q$ and the parameters ϕ_i, for $i = 1, \ldots, p$, are unknown and can be estimated using OLS.

Phillips Curve model Furthermore, bivariate Vector Autoregressive (VAR) models are included based on the Phillips Curve. For these forecasts the GDP growth, GDP growth gap, the output gap, the unemployment rate and unemployment gap are used, forming five individual models, each consisting of inflation and one of the variables. Besides this, one model is added using the three variables which are good predictors of forecast bias. In this VAR model, the GDP growth, output gap and unemployment gap are used to forecast inflation. Forecasts created with the bivariate VAR model are based on past observation of the two (or four in the model using all variables) included variables. The bivariate VAR model is described as follows:

$$\boldsymbol{Y}_{T+h} = \begin{bmatrix} y_{T+h} \\ z_{T+h} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{1,1} \\ b_{1,1} & b_{1,2} \end{bmatrix} \begin{bmatrix} y_T \\ z_T \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} y_{T-1} \\ z_{T-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{T+h} \\ \varepsilon_{T+h} \end{bmatrix} \tag{3}$$

$$= \boldsymbol{a} + \boldsymbol{B}_1 \boldsymbol{Y}_T + \boldsymbol{B}_2 \boldsymbol{Y}_{T-1} + \boldsymbol{\varepsilon}_{T+h}, \tag{4}$$

where Equation 4 is the model in matrix form. The parameters can be estimated using System OLS.

Direct forecasting models Lastly, five Direct Forecasting models using OLS regressions are included, in which the forecast of inflation is based on the last observation of some other variable z_T. This model is described as follows:

$$y_{T+h} = \alpha + \beta z_T + \varepsilon_{T+h}. \tag{5}$$

The choice of variable for z_T is, just like the previous models, the five variables as described before.

3.3 Bias corrected forecasts

Besides investigating the forecasts of the individual models as described above, we also construct forecasts which are corrected with estimated biases. This strategy builds on the idea that biasedness that is present in the forecasts of individual forecasting models can be corrected to obtain more accurate forecasts. This can be shown by observing the forecast error $\tilde{e}_{T+h}$ that remains when we use forecasts that are corrected with the actual bias, $\tilde{f}_{T+h} = f_{T+h} + b_T$, as also described in Gibbs and Vasnev (2017):

$$\tilde{e}_{T+h} = y_{T+h} - \tilde{f}_{T+h} \tag{6}$$

$$= e_{T+h} - b_T = \xi_{T+h}, \tag{7}$$

which is the bias-corrected forecast error. In practice however, when the noise in the estimated bias is large, the bias-correction might lead to forecasts that are less accurate than the uncorrected forecasts.

To obtain an estimate of the bias that is contained in the individual forecast, we regress the h-quarter-ahead forecast error of model i, $e_{i,T+h} = y_{T+h} - f_{i,T+h}$, on the observation at time T of a specific variable z_T at time T:

$$e_{i,T+h} = \alpha_i + \beta_i z_T + \xi_{i,T+h}. \tag{8}$$

At each time T, with this model an estimate of the bias can be made which is used for the bias-correction using $\tilde{f}_{T+h} = f_{T+h} + b_T$. For the bias-corrected individual forecast models, the variable output gap is used as regression variable z_T for the estimation of the bias.

3.4 Forecast combination strategies

Forecasts made with the econometric models as described above, are combined using weights in order to create combined forecasts. Consider the h-step-ahead forecasts vector of individual models $\boldsymbol{f}_{T+h}$:

$$\boldsymbol{f}_{T+h} = (f_{1,T+h}, \dots, f_{k,T+h})' \in \mathbb{R}^k, \tag{9}$$

with k the amount of individual forecasting models. The combined inflation forecast at time $T+h$ is then expressed as $f_{c,T+h} = \boldsymbol{w}'_{T+h}\boldsymbol{f}_{T+h}$ with weights $\boldsymbol{w}_{T+h} = (w_{1,T+h}, \dots, w_{k,T+h})'$. To evaluate the forecasts, we define the vector of forecasting errors as described in Gibbs and Vasnev (2017):

$$\boldsymbol{e}_{T+h} = y_{T+h}\mathbf{1} - \boldsymbol{f}_{T+h} \tag{10}$$

$$= \boldsymbol{b}_T + \boldsymbol{\xi}_{T+h}, \tag{11}$$

with y_{T+h} the real inflation value, $\boldsymbol{b}_T = \mathbb{E}[\boldsymbol{e}_{T+h}|\mathcal{I}_T]$ the forecast bias vector and $\boldsymbol{\xi}_{T+h}$ the vector of bias-corrected forecast errors with $\mathbb{E}[\boldsymbol{\xi}_{T+h}|\mathcal{I}_T] = 0$. The errors of the combined forecasts are given by

$$e_{c,T+h} = y_{T+h} - f_{c,T+h} = \boldsymbol{w}'_{T+h}\boldsymbol{e}_{T+h}. \tag{12}$$

Five different forecast combination strategies will be used: the benchmark strategy of Equal Weights, unconditional optimal weights, bias-corrected forecasts combined using unconditional optimal weights and the conditionally optimal weights strategy with respectively fixed and time-varying covariance matrix of the bias-corrected forecast errors $\boldsymbol{\xi}_{T+h}$.

Equal Weight forecasts This simple forecast combination strategy corresponds to averaging the individual forecasts in order to form the combined forecasts. Hence, the weights are defined as $\boldsymbol{w}_{T+h} = \boldsymbol{w} = \frac{1}{n}\mathbf{1}$. This strategy will be used as a benchmark forecast combination.

Unconditional optimal weights The unconditionally optimal weights $\boldsymbol{w}^*$ are calculated using

$$\boldsymbol{w}^* = \frac{\Sigma_e^{-1}\mathbf{1}}{\mathbf{1}'\Sigma_e^{-1}\mathbf{1}}, \tag{13}$$

where Σ_e corresponds to the unconditional variance of the forecast errors. This strategy boils down to using no information to predict the errors.

Bias-corrected forecasts combined with unconditional optimal weights This forecast combining strategy uses bias-corrected forecasts as constructed in Section 3.3. The individual forecasts are then combined using unconditional optimal weights, calculated using Equation 13. We use the five variables as described earlier as regression variable z_T to make estimates for the biases. Theoretically, if the forecasts can be corrected perfectly by removing the contaminating bias from the forecasts before combining, a more optimal combined forecasts can be obtained. Gibbs and Vasnev (2017) already described that when the biases are corrected perfectly, combining these corrected forecasts with unconditional optimal weights leads to forecasts with a lower MSE. However, because the estimated bias can be quite noisy, in practice this noise could also pollute the forecast, especially when the noise is relatively big compared to the bias.

Conditionally optimal weights with fixed $\boldsymbol{\Sigma}_\xi$ For this strategy, the conditionally optimal weights $\boldsymbol{w}^*(\mathcal{I}_T)$ are chosen such that the mean squared error (MSE) of the combined forecast errors, $e^2_{c,T+h}$ is minimised. This optimisation problem can be written as:

$$\boldsymbol{w}^*(\mathcal{I}_T) = \arg\min_{\boldsymbol{w}} \mathbb{E}[e^2_{c,T+h}|\mathcal{I}_T]. \tag{14}$$

The MSE of this optimization problem is $\text{MSE}(\boldsymbol{w}) = \boldsymbol{w}'(\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T')\boldsymbol{w}$ and is minimised by the conditionally optimal weights,

$$\boldsymbol{w}^*(\mathcal{I}_T) = \frac{[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\mathbf{1}}{\mathbf{1}'[\Sigma_\xi + \boldsymbol{b}_T\boldsymbol{b}_T']^{-1}\mathbf{1}}, \tag{15}$$

with fixed covariance matrix $\boldsymbol{\Sigma}_\xi$. In this strategy, a shrinkage component $\alpha \in (0, 1)$ is included such that the covariance matrix can be stabilised, $\tilde{\boldsymbol{\Sigma}}_\xi = \alpha\boldsymbol{\Sigma}_0 + (1-\alpha)\hat{\boldsymbol{\Sigma}}_\xi$. In our research, we use $\alpha = 0.5$ and for $\boldsymbol{\Sigma}_0$ we use the identity matrix.

Conditionally optimal weights with time-varying $\boldsymbol{\Sigma}_{\xi,T}$ This strategy is similar to the strategy above, however, we allow the covariance matrix $\boldsymbol{\Sigma}_{\xi,T}$ to vary over time. To achieve this, we use the GARCH composition as introduced in Bollerslev (1986). Specifically, we use the multivariate Dynamic Conditional Correlation (DCC) GARCH(1,1) model, as proposed in Engle (2002) and described in Orskaug (2009). The idea of the model is that the covariance matrix of the bias-corrected forecast errors $\boldsymbol{\xi}_T$, $\boldsymbol{\Sigma}_{\xi,T}$, can be decomposed as $\boldsymbol{\Sigma}_{\xi,T} = \boldsymbol{D}_T\boldsymbol{R}_T\boldsymbol{D}_T$, where $\boldsymbol{R}_T$ is a correlation matrix of the bias-corrected forecast errors and $\boldsymbol{D}_T$ a diagonal matrix of conditional standard deviations. The correlation matrix $\boldsymbol{R}_T$ can be decomposed using $\boldsymbol{R}_T = \boldsymbol{Q}_T^{*-1}\boldsymbol{Q}\boldsymbol{Q}_T^{*-1}$, where $\boldsymbol{Q}_T^{*-1}$ has the square root elements of $\boldsymbol{Q}_T$ on the diagonal, and zeros on the off-diagonal. The DCC-GARCH model is updated using Equation 16, where $\bar{\boldsymbol{Q}}$ is estimated using the bias-corrected forecast errors from the rolling window: $\bar{\boldsymbol{Q}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\xi}_t\boldsymbol{\xi}_t'$.

$$\boldsymbol{Q}_T = (1-a-b)\bar{\boldsymbol{Q}} + a\boldsymbol{\xi}_{T-1}\boldsymbol{\xi}_{T-1}' + b\boldsymbol{Q}_{T-1}. \tag{16}$$

For the initial $\boldsymbol{Q}_0$, the matrix $\bar{\boldsymbol{Q}}$ is chosen. The covariance matrix $\boldsymbol{\Sigma}_{\xi,T}$ is positive semidefinite if $a \geq 0$, $b \geq 0$ and $a+b < 1$. The optimal values for the parameters a and b are estimated in the training sample, using the generalised optimisation method, choosing the a and b such that the expected RMSFE is as low as possible. Just like in the strategy with the fixed $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$, the shrinkage component $\alpha = 0.5$ is added to stabilise the covariance matrix such that $\tilde{\boldsymbol{\Sigma}}_{\xi,T} = \alpha\boldsymbol{\Sigma}_0 + (1-\alpha)\hat{\boldsymbol{\Sigma}}_{\xi,T}$, with $\boldsymbol{\Sigma}_0$ the identity matrix.

3.5 Performance Measures

To measure the overall forecasting performance of the forecasting models, the Root Mean Squared Forecasting Error (RMSFE) for each model i is calculated. The RMSFE of model i is defined as

$$\text{RMSFE}_i = \sqrt{\frac{1}{T}\sum_{T=i}^{N}(y_T - f_{i,T})^2}, \tag{17}$$

with y_T the real inflation rate at time T and $f_{i,T}$ the inflation forecast of model i at time T. To be able to compare the forecasting performance of each model with the benchmark models (for the individual models and the combined forecasting models respectively the Naive forecasting model and the combination model of Equal Weights), in the results the relative RMSFE will be given. To test whether a particular model outperforms the benchmark model, the Diebold-Mariano test for the comparison of predicting accuracy is used, as proposed by Diebold (2015). The test compares the RMSFE of the particular model and the benchmark model in a certain period, with the null hypothesis that both models have equal forecast accuracy. In the results, for each model only the significance level of the model itself compared to the benchmark will be given.

Furthermore, to measure if there is any biasedness present in the forecasts of model i, the Mean Forecast Error (MFE) is calculated for each model. The MFE is defined as

$$\mathrm{MFE}_i = \frac{1}{T}\sum_{j=i}^{N}(y_T - f_{i,T}). \tag{18}$$

To test whether the MFE is significantly different from zero, we use the two-sided t-test with Newey-West Heteroskedasticity and Autocorrelation Corrected (HAC) standard errors, as described in Newey and West (1986). The standard errors are obtained using Eviews 10+. In next section, the results will be described.

4 Results

In this section, we will describe the results of benchmark models, the individual forecasting models and the bias-corrected individual forecasting models. After this, the results of the combined forecasts will be described.

4.1 Results benchmark models

First, we will describe the results of the benchmark models, namely the Naive forecasting model and the combination strategy of Equal Weights. In Table 3 the RMSFE and the MFE of both models over the total forecasting sample and the three separate forecasting samples are given. For comparison, also the bias-corrected Naive forecasting model is added. Notable is that the forecasting performance of the three models differ greatly among the three forecasting samples. The RMSFE is almost twice as big in the first subsample compared to the RMSFE of the total sample. Furthermore, the RMSFE drops in the second sample, and becomes even lower in the third forecasting subsample. In Figure 2, the two forecasts together with the true inflation rates are shown. It can be seen that the highly volatile inflation rates during the first forecasting sample gives less accurate forecasts of the benchmark models and it is notable that for the Naive forecasting model, the short but large peaks in inflation rates are hard to predict because of the lagged forecasts. Also, the Equal Weights strategy does not manage to predict the inflation well in the first sample. This is in contrast to the second forecasting sample and (even more) third forecasting sample, when inflation rates are less volatile and gradually drop to around 2%. Both forecasting models obtain more accurate forecasts when the inflation rate is less volatile.

Table 3: The MFE and RMSFE of the inflation forecasts of the benchmark forecasting models: the (bias-corrected) Naive forecasting model and the forecasting combination strategy of Equal Weights

	Total Sample		1971Q2 - 1982Q4		1983Q1 - 2007Q3		2007Q4 - 2019Q1	
Benchmark model	MFE	RMSFE	MFE	RMSFE	MFE	RMSFE	MFE	RMSFE
Naive	$0.12^{\dagger}$	2.150	$-0.16^{\dagger}$	3.906	$0.29^{\dagger}$	1.213	$0.04^{\dagger}$	0.733
Bias-corrected Naive	$0.04^{\dagger}$	2.496	$0.13^{\dagger}$	4.399	$0.02^{\dagger}$	1.594	$0.02^{\dagger}$	0.877
Equal Weights	$0.40^{\dagger}$	1.965	-1.61	3.657	1.12	1.679	$0.02^{\dagger}$	0.715
p-value DM-test		0.354		0.914		0.049		0.771

For the benchmark models and the bias-corrected Naive forecasting model, the MFE and RMSFE are shown. MFEs which are not significantly different from zero at the 10 percent level are indicated with an †. Besides this, the p-value of the Diebold-Mariano test for equal forecast accuracy between the two benchmark models is added.

When we compare the Equal Weights strategy and the Naive forecasting models in more detail, it stands out that in the first and third forecasting sample, the Equal Weights strategy has the lowest RMSFE. This is in contrast to the second forecasting sample, where the Naive forecasting model has the lowest RMSFE. Also the p-values of the Diebold-Mariano test for equal forecasting accuracy of the two benchmark models for the total sample and the three forecasting samples can be found in Table 3. These indicate that in the total sample and in the first and third forecasting sample, we cannot reject the null hypothesis that the forecasting accuracy of both models is equal. However, in the second forecasting sample, we reject this null hypothesis at the 5% level. As the RMSFE of the Naive forecasting model is lower in this sample, the forecasting accuracy of the Naive forecasting model is better in the second sample.

4.2 Individual forecasting models

In Table 4, the MFE and RMSFE of the individual forecasting models of the complete sample and the three forecasting samples are shown. The RMSFE is given relative to the RMSFE of the benchmark, the Naive forecasting model.

As can be seen, over the complete sample, the Naive forecasting model turns out to be hard to beat. Only the AR(1) and AR(2) models have a lower RMSFE, however, these are both not significantly lower than the RMSFE of the benchmark. Furthermore, most models have a RMSFE which is relatively close to the RMSFE of the Naive forecasting model. Besides this, almost none

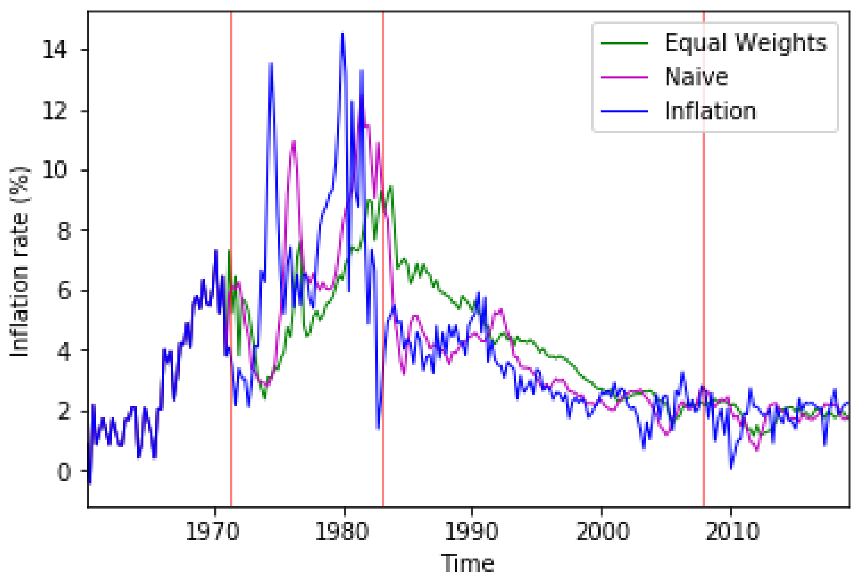

Figure 2: US inflation rates with forecasts of the benchmarks, the Naive forecasting model and the forecast combining model of Equal Weights

of the biases are significantly different from zero.

When the forecasting performance of the different samples are compared, it turns out that all models have worse forecasting performance in the first forecasting sample. This can be due to the high volatility levels of inflation which are present in this sample, as can be seen from Table 1. Besides the higher RMSFE, most individual forecasting models show a relatively large negative MFE, although not all are significantly different from zero at the 10% level. Especially the Direct Forecasts show negative biasedness in the forecast errors. In the second forecasting sample, the forecasting performance of all the individual models is better than in the first sample, as seen by the lower RMSFE. Also, for most models, the forecasting performance seems to be better in this sample than in the total sample. However, none of the individual models produces more accurate forecasts than the benchmark in the same sample. Besides this, all forecasting models except the benchmark have positively biased forecast errors significantly different from zero. In the third sample, the MFE and RMSFE again show different characteristics than the performance measures in the other samples. All RMSFE are lower in this sample than in the previous samples, and almost all MFE are unbiased. However, again not a single model is able to significantly outperform the Naive forecasting model in this sample. These results shows that in general, inflation seems to be easier to forecast in periods when inflation has low volatility (like in the second and third forecasting sample). When the MFE of the three different forecast samples are compared, it seems like forecasts of most models are less biased when the inflation is stable, and that the bias in the MFE inflates when there is more volatility. However, in the second forecasting sample, when the volatility is already relatively low compared to the first sample, there is still biasedness in the MFE present.

Table 4: MFE and relative RMSFE of individual forecasting models

	Total Sample		1971Q2 - 1982Q4		1983Q1 - 2007Q3		2007Q4 - 2019Q1	
Forecasting model	MFE	Rel. RMSFE	MFE	Rel. RMSFE	MFE	Rel. RMSFE	MFE	Rel. RMSFE
Univariate								
Naive	0.12†	1.000	-0.16†	1.817	0.29†	0.564	0.04†	0.341
AR(1)	0.22†	0.957	-1.00†	1.646	0.85	0.667	0.11†	0.306
AR(2)	0.18†	0.993	-1.00†	1.717	0.78	0.683	0.07†	0.316
AR(4)	0.09†	1.064	-1.32†	1.859	0.78	0.711	0.05†	0.322
ARMA(1,1)	-0.78	1.034	-1.20†	1.883	-0.71	0.570	-0.51	0.378
ARMA(4,4)	0.07†	1.070	-0.34†	1.908	0.33	0.661	-0.02†	0.345
Phillips curve								
GDP Growth	0.18†	1.001	-0.91†	1.730	0.75	0.688	0.06†	0.321
GDP Growth Gap	0.17†	1.018	-0.98†	1.730	0.78	0.736	0.06†	0.318
Output Gap	0.09†	1.169	-1.15†	2.071	0.65	0.742	0.18†	0.370
Unemployment	0.08†	1.029	-1.17	1.657**	0.79	0.827	-0.16†	0.373
Unemployment Gap	0.16†	1.030	-1.28	1.609	0.92	0.872	-0.03†	0.386
Var All	0.64	1.394	0.94†	2.355	0.83	0.951	-0.05†	0.706
Direct Forecasts								
GDP Growth	0.25†	1.060	-2.02	1.701	1.40	0.872	0.12†	0.308
GDP Growth Gap	0.25†	1.063	-2.04	1.704	1.40	0.877	0.11†	0.306
Output Gap	0.25†	1.098	-2.08	1.790	1.41	0.877	0.13†	0.318
Unemployment	0.00†	1.176	-2.86	1.843	1.46	1.006	-0.21†	0.349
Unemployment Gap	0.11†	1.171	-2.75	1.742	1.56	1.081	-0.08†	0.330

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

For each model, the MFE are shown. MFEs which are not significantly different from zero at the 10 percent level are indicated with an †. Furthermore, for each model the relative RMSFE and the one-sided significance of the Diebold-Mariano test compared to the benchmark model (the Naive forecasting model) at three different significance levels are shown (indicated with the asterices).

4.3 Bias corrected individual forecasting models

In Table 5, the results of the bias-corrected individual forecasting models can be found. For each model, the forecasts are corrected using biases as estimated with Equation 8, using the variable output gap as z_T. Again, the MFE of each model is shown, together with the RMSFE relative to the RMSFE of the uncorrected Naive forecasting model. Over the complete sample, none of the bias-corrected individual models manage to significantly outperform the uncorrected Naive forecasting model in forecasting accuracy when tested for using the Diebold-Mariano test.

In the remaining three forecasting sample, there are also no models which have significantly more accurate forecasting accuracy than the uncorrected Naive forecasting model. However, just like the results of the uncorrected individual forecasts, the forecasts in the third forecasting sample have the lowest RMSFE of the three samples, followed by the second and third forecasting sample, suggesting that models can forecast less volatile inflation better.

Table 5: MFE and relative RMSFE of bias-corrected individual forecasting models

	Total Sample		1971Q2 - 1982Q4		1983Q1 - 2007Q3		2007Q4 - 2019Q1	
Forecasting model	MFE	Rel. RMSFE	MFE	Rel. RMSFE	MFE	Rel. RMSFE	MFE	Rel. RMSFE
Univariate								
Naive	0.04†	1.161	0.12†	2.046	0.02†	0.741	0.02†	0.408
AR(1)	1.14	1.265	1.51†	2.046	1.48	1.024	0.02†	0.378
AR(2)	0.17†	1.101	-0.50†	1.821	0.54	0.841	0.07†	0.390
AR(4)	0.11†	1.182	-0.70†	2.010	0.51	0.846	0.07†	0.403
ARMA(1,1)	-0.03†	1.199	0.25†	2.164	-0.15†	0.701	-0.04†	0.397
ARMA(4,4)	0.25†	1.424	1.08†	2.339	-0.08†	1.107	0.14†	0.483
Phillips curve								
GDP Growth	0.17†	1.127	-0.44†	1.872	0.50†	0.852	0.06†	0.395
GDP Growth Gap	0.16†	1.141	-0.49†	1.855	0.53†	0.906	0.04†	0.390
Output Gap	0.14†	1.239	-0.53†	2.056	0.46†	0.941	0.12†	0.424
Unemployment	0.17†	1.015	-0.66†	1.719	0.70	0.738	-0.10†	0.328
Unemployment Gap	0.24†	1.008	-0.57†	1.601	0.79	0.830	-0.09†	0.374
Var All	0.62	1.389	0.92†	2.350	0.78	0.947	-0.02†	0.698
Direct Forecasts								
GDP Growth	0.22†	1.067	-0.88†	1.738	0.80	0.840	0.10†	0.391
GDP Growth Gap	0.23†	1.098	-0.94†	1.754	0.89	0.897	0.00†	0.402
Output Gap	0.20†	1.154	-0.95†	1.971	0.79	0.820	0.11†	0.385
Unemployment	0.10†	1.058	-1.78	1.694	1.05†	0.871	-0.12†	0.331
Unemployment Gap	0.18†	0.967	-1.38	1.441	1.05	0.885	-0.10†	0.312

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

For each model, the MFE are shown. MFEs which are not significantly different from zero at the 10 percent level are indicated with an †. Furthermore, for each model the relative RMSFE and the one-sided significance of the Diebold-Mariano test compared to the benchmark model (the uncorrected Naive forecasting model) at three different significance levels are shown (indicated with the asterices).

4.4 Combined forecasting models

The results of the combined forecasting models can be found in Table 6. In this table, again the MFE of each combined forecasting model is given, together with the RMSFE relative to the benchmark model in the complete sample, the forecast combining strategy of Equal Weights. Also, the results of the Naive forecasting model without bias-correction are given for comparison.

When comparing the results of the total sample, the benchmark model as well as the Naive forecasting model both turn out to be hard to beat. Also, the forecast combination model with unconditional weights performs surprisingly well.

Some of the bias-corrected forecasting models have similar RMSFE as the Equal Weight strategy and the Naive forecasting model, depending on the variable used to forecast the bias although the models do not outperform the benchmark. Also, the conditional optimal weights with fixed covariance matrix of the bias-corrected forecast errors $\mathbf{\Sigma}_\xi$ do not outperform the benchmark model. The combination model with conditional optimal weights and a time-varying $\mathbf{\Sigma}_\xi, T$ matrix even has a higher RMSFE than the unconditionally optimal weights with fixed covariance matrix. Overall, there is not a single model that significantly outperforms the benchmark models. The forecast combination strategy of Equal Weights therefore seems a good choice. Also, some of the forecast combinations with conditionally optimal weights produce biased forecasts which are significantly different from zero, especially when the covariance matrix is time-varying.

When we compare the models in the first forecasting sample with highly volatile inflation rates, again the performance of the models drops, which is shown by the higher RMSFE, which is for each model around 1.4 to 2.6 times as high as the benchmark model in the complete sample. There is some difference between the bias-corrected combination models depending on the variable used to correct the bias, but in general these models do not outperform the benchmark. The combination models with conditionally optimal weights with both fixed and time-varying $\mathbf{\Sigma}_{\xi,T}$ matrix do have lower RMSFE in this sample, however when tested for equal forecasting accuracy with the Diebold-Mariano test, the models do not significantly outperform the Equal Weights strategy in the same forecasting sample. Notable is however, that all of the forecast combination with conditionally optimal weights with time-varying covariance matrix of the bias-corrected forecasts errors, and some of those with fixed covariance matrix, do produce unbiased forecast errors in this sample.

When we compare the models in the last two forecasting samples we see that, just like in the previous results, the RMSFE improves as the volatility of the inflation drops. However, there is not one model which is able to outperform the benchmark. In the second forecasting sample, all forecasting combinations have biased forecast errors, while in the third forecasting sample, they are all unbiased.

Table 6: MFE and relative RMSFE of combined forecasting models

	Total Sample		1971Q2 - 1982Q4		1983Q1 - 2007Q3		2007Q4 - 2019Q1	
Forecasting model	MFE	Rel. RMSFE	MFE	Rel. RMSFE	MFE	Rel. RMSFE	MFE	Rel. RMSFE
Benchmark Models								
EW	0.40†	1.000	-1.61	1.861	1.12	0.855	0.03†	0.364
Naive	0.12†	1.094	-0.16†	1.988	0.29†	0.618	0.04†	0.373
Uncond	-0.06†	1.039	-2.68	2.145	0.61	0.744	0.02†	0.376
Bias corrected forecasts with unconditional optimal weights								
GDP Growth	0.21†	0.978	-1.47	1.782	0.78	0.847	0.02†	0.411
GDP Growth Gap	0.19†	0.968	-1.47	1.776	0.73	0.830	-0.02†	0.411
Output Gap	0.23†	1.056	-1.46†	2.003	0.79	0.879	0.00†	0.391
Unemployment	0.20†	1.164	-1.62	2.212	0.76	0.964	0.05†	0.442
Unemployment Gap	-0.25†	1.316	-2.97	2.586	0.49	0.931	-0.22†	0.827
Conditional optimal weights with fixed covariance matrix Σ_{ξ}.								
GDP Growth	0.35†	1.005	-1.61	1.724	1.03	0.926	0.04†	0.433
GDP Growth Gap	0.37†	1.026	-1.59	1.731	1.05	0.960	0.04†	0.443
Output Gap	0.36†	1.024	-1.58	1.620	1.05	0.979	0.01†	0.562
Unemployment	0.49	1.034	-1.15†	1.617	1.11	0.972	0.12†	0.656
Unemployment Gap	0.50	1.089	-0.28†	1.685	0.91	1.021	0.10†	0.724
Conditional optimal weights with time-varying covariance matrix $\Sigma_{\xi,T}$.								
GDP Growth	0.56	1.018	-0.41†	1.470	1.07	1.058	0.04†	0.442
GDP Growth Gap	0.59	1.044	-0.34†	1.485	1.10	1.094	0.05†	0.451
Output Gap	0.55	1.052	-0.48†	1.474	1.08	1.085	0.01†	0.575
Unemployment	0.66	1.054	-0.05†	1.554	1.11	1.037	0.13†	0.652
Unemployment Gap	0.71	1.171	0.96†	1.880	0.94	1.086	0.08†	0.716

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

For each model, the MFE are shown. MFEs which are not significantly different from zero at the 10 percent level are indicated with an †. Furthermore, for each model the relative RMSFE and the one-sided significance of the Diebold-Mariano test compared to the benchmark model (combining forecasts with Equal Weights) at three different significance levels are shown (indicated with the asterices). For the models with conditional optimal weights, a shrinkage component of $\alpha = 0.5$ is used to stabilise the covariance matrix.

5 Conclusion

In this research, we replicated the paper of Gibbs and Vasnev (2017) about the performance of forecasts combination models. Specifically, Gibbs and Vasnev (2017) put effort in examining the forecasting performance of forecast combinations with conditionally optimal weights with fixed covariance matrix of the bias-corrected forecast errors. To extent their research, we added a similar model and allowed the covariance matrix to be time-varying, using a DCC-GARCH(1,1) model. In line with Gibbs and Vasnev (2017), the individual forecasting models do not outperform the parsimonious Naive forecasting model, which is used as benchmark model for the individual models.

When we correct the individual forecasting models with an estimated bias, this also does not deliver more accurate forecasting results. Seemingly, the estimated biases used to correct the forecasts are noisy estimates, and therefore are not able to improve the forecasts significantly.

Furthermore, in contrast to the results in Gibbs and Vasnev (2017), the forecast combining strategies do not outperform the benchmark strategy of the combined forecasting models, the forecast combination of Equal Weights. None of the models manages to significantly outperform the parsimonious model in one of the forecasting subsamples at all. There are a few factors which could cause these differences.

First, one major difference between this paper and the paper of Gibbs and Vasnev (2017) is that we use the last vintage data for US inflation, in contrast to using all vintages. It could be that using the data at the time that it is available to the public is beneficial for some models in forecasting inflation.

A second reason for this could be the differences between the estimation windows. In Gibbs and Vasnev (2017), an expanding window is used, which enlarges the training sample at every iteration. In this research, a rolling window of length $n = 40$ observations is used. It might be the case that the training sample is too short to provide accurate parameter estimates. However, as mentioned earlier, the inflation data from roughly 1965 until 1983 shows very different patterns compared to the last two forecasting samples, characterised by high volatility levels. This rises the question whether keeping observations from past periods with different characteristics in the training sample by using an expanding window is beneficial for the forecasting accuracy.

A third difference between the methods used in this paper and the methods used in Gibbs and Vasnev (2017), is that the estimation of the biases is done by using forecast errors of the last 20 observations in the rolling window. However, because we estimate the model parameters using the entire rolling window, using the forecast errors from the same sample to obtain the bias estimates is not entirely correct. Therefore, this could also lead to different outcomes compared to Gibbs and Vasnev (2017). However, because of the poor forecast outcomes when only using the first 20 observations for the estimation of the parameters, this choice seemed the right way to go.

One secure conclusion that we can draw from this research is that forecasting inflation remains a puzzle, especially when inflation rates are highly volatile. Although in our research, we were not able to produce a model that significantly outperforms the parsimonious benchmark forecasting models, methods of combining forecast strategies stays promising. More research on this topic, and specifically, on the topic of conditionally optimal weights might be beneficial in finding more accurate forecasting methods.

References

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

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

F. X. Diebold. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–1, 2015.

R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

C. Gibbs and A. L. Vasnev. Conditionally optimal weights and forward-looking approaches to combining forecasts, 2017.

R. Hodrick and E. C. Prescott. Post-war us business cycles. *An Empirical*, 1980.

W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix, 1986.

E. Orskaug. Multivariate dcc-garch model:-with various error distributions. Master's thesis, Institutt for matematiske fag, 2009.

A Appendix

The python code used for this thesis is provided separately in the ZIP-file, BachelorThesisCodes. This file contains the following codes, followed by a brief description of the code:

HodrickPrescott.ipynb

Python code for obtaining the output gap from the GDP growth using the Hodrick-Prescott filter.

weightsCalculator.ipynb

Calculates all the models used in this paper, including the weights and other output. Also, graphs found in this paper are plotted using this code. In the code, at the start of each model, the name of the model can be found.