Forecasting US Inflation

Udacity Data Scientist Nanodegree Capstone project

November 28, 2023

Contents

1 Project Definition

1.1 Project Overview

This is the Udacity Data Scientist Nanodegree Capstone project. The aim is to demonstrate the concepts learned in the programme. To do this, I have set up a project that includes

- Collecting the data (web scrap and using APIs)
- Cleaning and storing the data in a database
- Writing modular, documented code
- Analysing different learning algorithms
- Drawing conclusions and communicating them

Because of my background, I decided to work on an economic problem, namely forecasting inflation. Inflation measures the general increase in the price of goods and services over time. Inflation affects the purchasing power of consumers and is an important economic variable. Inflation forecasting is particularly relevant in the current context of high inflation rates following the COVID pandemic. The current discussion is whether and how fast inflation will return to pre-COVID levels.

The aim of this project is to analyse different inflation forecasting methods and to compare their accuracy and reliability. The project will focus on the US case. Macroeconomic variables will be retrieved from the Federal Reserve Economic Data API and by webscraping the BLS homepage.

1.2 Problem Statement

The main research question of this project is: Which inflation forecast method is the most accurate and reliable? What is the current prediction in the high inflation environment?

To answer this we compare different inflation forecast methods for the US case. We compare the forecasting power among the following dimensions:

- Is it helpful to forecast the components of the consumer price index individually and aggregating them?
- Do time series models gain forecasting power by adding economic variables (macroeconomic variables, such as output, unemployment, and exchange rates.)?

1.3 Metrics

The main measure used to evaluate inflation forecasting methods is the root mean squared error (RMSE), which measures the average deviation of the forecast values from the actual values. The lower the RMSE, the better the forecasting method. The RMSE of a naive forecasting method (such as a random walk or a no-change forecast) is shown as a benchmark. In addition to the RMSE, we also look at the sample mean (out of sample in the test data). As I will explain later, it makes the exercise more useful to split the data into test and training using a time stamp (rather than random shuffling). It is not guaranteed that the out of sample error will be zero on average. Therefore, it is also useful to look at the average error.

2 Analysis

2.1 Data Exploration

Table 1 shows the subcomponents of the CPI (the parts that make up the CPI) together with the date when the item was added to the CPI and the distribution of the monthly (seasonally adjusted, not annualised) CPI changes. The "All items" component of the CPI (the one we are interested in forecasting) has been available since 1947. The average monthly increase has been 0.29%, but the increases range from -1.77% to 1.96%.

We see that while some items are available for the whole period (such as food), others are only available since 1994 (video and audio) or even 1999 (personal care). This reflects the fact that products available to the public change over time. Not only do they change the weight in the CPI, but also completely new products may appear over time.

The most volatile components of the CPI are:

- Tobacco and smoking products
- Private and public transportation
- Fuels and utilities
- Food
- Apparel (mens and boys, womens and girls, infants)

The least volatile components of the CPI are:

- Education
- Medical care
- Personal care
- Housing

Table 2 shows the weights that the subcomponents (introduced in table 1) have in the CPI for some selected years. Most of the years are omitted to save space. Observations:

- The most important component has always been housing and its weight has increased over time from 30% to 40%.
- Food is also an important component of the CPI, but its importance has declined over time from 30% to around 15%.
- The next important category is private transport, which has increased from 10% to 16% over time.

Table 3 shows the additional variables used in the analysis which are not the components of the CPI itself. The additional variables are motivated by economic considerations and include concepts that can reasonably be expected to have an impact on the price level. The variables include

- Unemployment rates
- Credit data
- Money supply (monetary aggregates)
- Interest rates
- Exchange rates

The reasons for using the variables are summarised in the table.

	available since	motivation
WTI	1946-01-31	The oil price is an important driver of the overall price level of the economy. A higher oil price contemporeanously affects CPI. It can be used to nowcast the inflation. Note, however, that the oil price can only be used to forecast CPI if it affects certain price categories with a *lag*.
Consumer_Loans	1947-01-31	A decrease in consumer loans can lead to a decrease in consumer spending, which can cause a decrease in demand for goods and services.
Loans_Leases	1947-01-31	Similar to consumer loans, but here also loans to enterprises are included.
Unemployment_Rate	1948-01-31	When unemployment is high, there's little need for employers to "bid" for the services of employees by paying them higher wages. In times of high unemployment, wages typically remain stagnant, and wage inflation is non-existent.
10Y_Rate	1953-04-30	Interest rate on government debt with 10 year maturity. When interest rates are high, borrowing becomes more expensive, and people tend to spend less money. This decrease in spending can lead to a decrease in demand for goods and services, which can lead to a decrease in prices and a decrease in the CPI .

FFER	1954-07-31	Fed Funds effective rate: Interest rate on short term loans between banks. Same motiviation as for 10 Y rates.
Real_M1	1959-01-31	M1, M2, and M3 are monetary aggregates that represent different measures of the money supply in an economy. M1 is the narrowest measure of the money supply and includes currency, demand deposits, and other liquid assets. M2 and M3 include other liquid assets. (An increase in the money supply (M1 to M3) can lead to inflation, which can cause the CPI to rise.)
M1	1959-01-31	see Real_M1
M2	1959-01-31	see Real_M1
Real_M2	1959-01-31	see Real_M1
M3	1960-01-31	see Real_M1
JPY	1971-01-31	Exchange rate (Japan), import prices may rise, triggering increases in the CPI of the home country.
CAD	1971-01-31	Exchange rate (Canada), see JPY for motivation
GBP	1971-01-31	Exchange rate (United Kingdom), see JPY for motivation
Real_Borad_Effective_Exchange	1994-01-31	Exchange rate (aggregate, real), see JPY for motivation
Inflation_Exp_Market_10YR	1982-01-31	Market expectation of inflation (10Y expectatin) derived from prices of inflation protected bonds.
Inflation_Exp_Market_1YR	1982-01-31	Market expectation of inflation (1Y expectatin) derived from prices of inflation protected bonds.

Table 3: Additional variables used to forecast CPI

2.2 Data Visualization

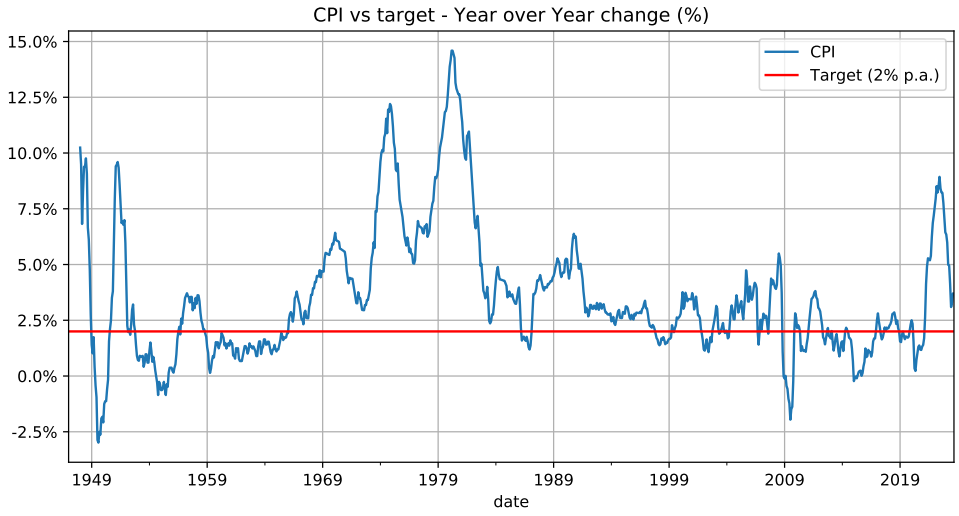

Figure 1: Year over Year changes of CPI (seasonally adjusted) vs target

Figure 1 shows how the CPI is rising over time. The FED's inflation target is 2% per annum. We can see that the FED has missed the target in the long run. However, this is due to specific periods of inflation overshooting (such as 1970-1984) or the most recent episode. There have been longer periods where inflation has been in line with the target (1989-2020). It is important to note that the CPI is often commented on in terms of year-on-year (YoY) changes. However, YoYs are easy to forecast ahead of the data release because 11 of the 12 months that make up a YoY data point are already known ahead of the release (so-called base effects). It is much harder to forecast the month on month (MoM) changes shown in the figure 2. Below we forecast the year-on-year changes in the CPI, but with a forecast horizon of 12 months. This means that we are forecasting year-on-year inflation in one year's time, so we are not making the exercise any easier by including base effects. Figure 3 shows the weights of the items in the CPI basket over time. As already mentioned,

FRED Name	first date	mean	std	min	max
All items	1947-01-01	0.29%	0.34%	-1.77%	1.96%
Food and beverages	1967-01-01	0.33%	0.41%	-0.96%	5.17%
Food	1947-01-01	0.29%	0.56%	-2.32%	5.87%
Housing	1967-01-01	0.35%	0.31%	-0.81%	1.86%
Shelter	1953-01-01	0.34%	0.35%	-1.43%	2.24%
Fuels and utilities	1953-01-01	0.31%	0.73%	-2.40%	5.61%
Household furnishings and operations	1967-01-01	0.19%	0.33%	-0.64%	2.02%
Apparel	1947-01-01	0.13%	0.47%	-3.66%	1.87%
Mens and boys apparel	1947-01-01	0.13%	0.58%	-4.42%	2.61%
Womens and girls apparel	1947-01-01	0.08%	0.75%	-4.09%	2.92%
Footwear	1947-01-01	0.19%	0.57%	-2.58%	3.10%
Infants and toddlers apparel	1989-01-01	0.02%	1.39%	-4.13%	9.57%
Transportation	1947-01-01	0.30%	1.03%	-10.28%	5.87%
Private transportation	1947-01-01	0.29%	1.07%	-10.80%	6.22%
Public transportation	1989-01-01	0.19%	1.80%	-11.62%	11.09%
Medical care	1947-01-01	0.41%	0.31%	-0.68%	1.82%
Medical care commodities	1967-01-01	0.32%	0.34%	-0.85%	1.45%
Medical care services	1956-01-01	0.45%	0.31%	-0.70%	2.12%
Recreation	1993-01-01	0.11%	0.23%	-0.60%	0.87%
Video and audio	1994-01-01	0.05%	0.33%	-1.25%	1.16%
Education and communication	1993-01-01	0.15%	0.23%	-1.72%	1.12%
Education	1993-01-01	0.37%	0.17%	-0.24%	1.19%
Communication	1998-01-01	-0.09%	0.41%	-3.31%	1.81%
Other goods and services	1967-01-01	0.41%	0.40%	-0.98%	4.07%
Tobacco and smoking products	1986-01-01	0.56%	1.60%	-5.19%	17.74%
Personal care	1999-01-01	0.19%	0.21%	-0.28%	1.16%

Table 1: Overview of the used CPI items

	1952	1972	1992	2012
FRED Name				
All items	100.00%	100.00%	100.00%	100.00%
Food and beverages	NA	NA	17.40%	15.26%
Food	29.84%	22.49%	15.78%	14.31%
Housing	32.18%	33.86%	41.40%	41.02%
Shelter	17.46%	21.83%	27.88%	31.68%
Fuels and utilities	NA	4.71%	7.28%	5.30%
Household furnishings and operations	6.45%	7.32%	6.24%	4.04%
Apparel	9.42%	10.37%	6.00%	3.56%
Mens and boys apparel	3.00%	2.80%	1.42%	0.86%
Womens and girls apparel	4.16%	3.98%	2.46%	1.50%
Footwear	1.44%	1.57%	0.80%	0.70%
Infants and toddlers apparel	NA	NA	0.19%	0.20%
Transportation	11.33%	13.13%	17.01%	16.85%
Private transportation	10.11%	11.66%	15.48%	15.66%
Public transportation	1.22%	1.47%	1.53%	1.19%
Medical care	4.78%	6.45%	6.93%	7.16%
Medical care commodities	NA	NA	1.28%	1.71%
Medical care services	3.99%	5.58%	5.65%	5.45%
Recreation	NA	3.77%	NA	5.99%
Video and audio	NA	NA	NA	1.90%
Education and communication	NA	NA	NA	6.78%
Education	NA	NA	NA	3.28%
Communication	NA	NA	NA	3.50%
Other goods and services	5.01%	5.09%	6.90%	3.38%
Tobacco and smoking products	NA	NA	1.75%	0.80%
Personal care	2.12%	2.57%	1.19%	2.57%

Table 2: Weights of the CPI items in some selected years. Not all weights are available at all time.

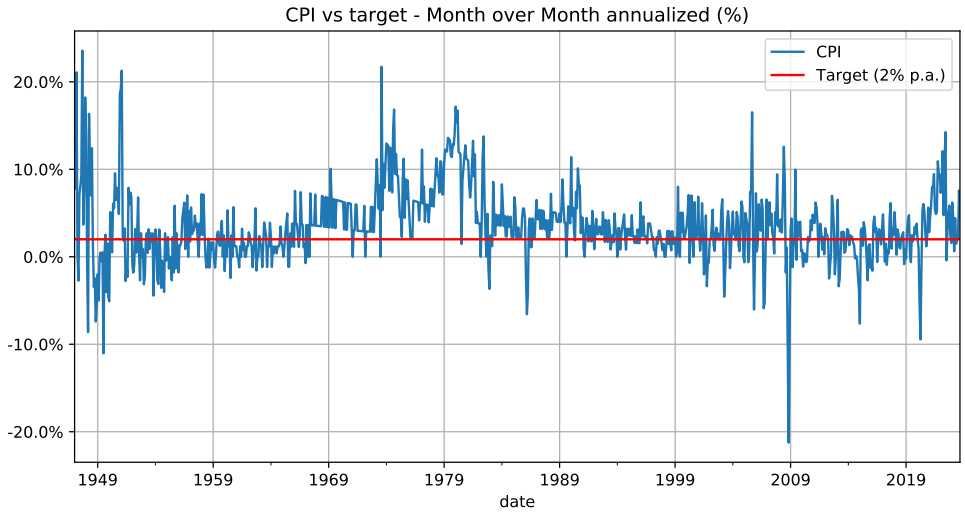

Figure 2: annualized month over over month changes of CPI vs target

shelter is one of the most important components. Note also that the weights do not add up to 100%. This is due to missing data and the fact that some items that are part of the CPI today did not exist 40 years ago (e.g. video and audio). Missing data is an issue to be dealt with separately.

3 Methodology

3.1 Data Preprocessing

The FRED API requires registering and getting an API key. Registering at FRED is free and you get an API key within minutes. However, we should never hardcode authentication credentials like tokens, keys, or app-related secrets into code published to Github. I checked several ways to solve the problem here: Keeping your API credentials secure (Github post). I decided to use the Python dotenv package, because it is a straightforward solution. Instructions on how to write the API key into the environment is provided in the submission comment for the project.

The overall goal of the CPI-U index is to use consumer spending from as recent a period as possible, and hold the set (or more precisely, the quantity mix) of goods and services purchased fixed over time until new spending weights can be introduced. In general, estimates of current period inflation calculated with outdated spending

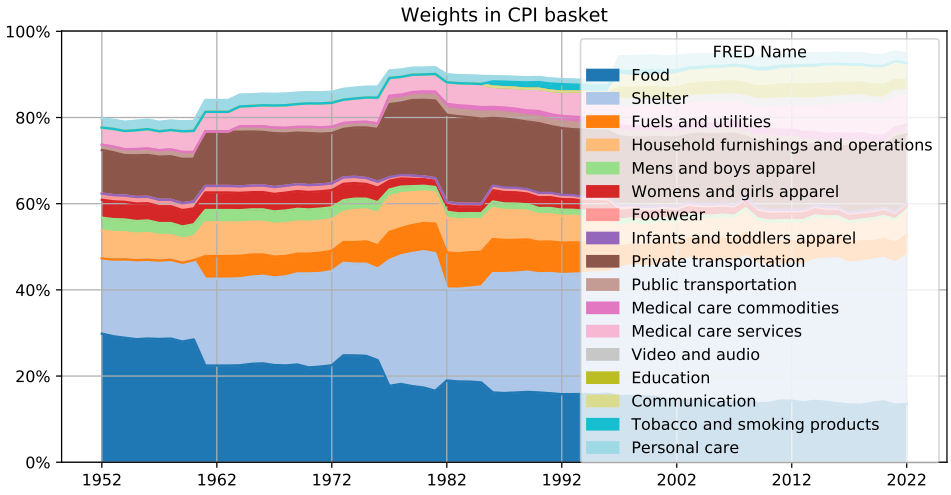

Figure 3: Weights in the cpi basket

weights tend to be higher than inflation estimates calculated with more current spending weights. This is because consumers change, or substitute, what they buy over time, often shifting purchases away from items that are becoming relatively more expensive to alternatives whose prices are not rising as fast.

The CPI weights are unfortunately not available directly through an API as far as I have seen. The weights can however be downloaded by scraping the website of the U.S. Bureau of Labor Statistics (BLS). Building the historical weights is however not straightforward, as the names change slightly over time. The subbaskets of the CPI are aggregated to categories with different granularity levels. As shown in Figure 4, the "food and beverages" component (identation level 1) is further divided into "Food" (identation level 2) which can be furtther divided into "Food at home" (level 3).. Combining the weights (scraped from the U.S. Labor Statistics webpage) with the Index series downloaded from FRED requires some extra effort, because 1) the names do not match exaclty 2) not all CPI baskets reported on the BLS webpage can be downloaded from FRED. To match the data I followed the following steps:

- Download all series available from FRED until identation level 2 (i.e. "Food at home" in figure 4)
- Match them by name with the weighting data from BLS. In cases where the matching is not possible (because of slightly differnt spelling) a manual mapping

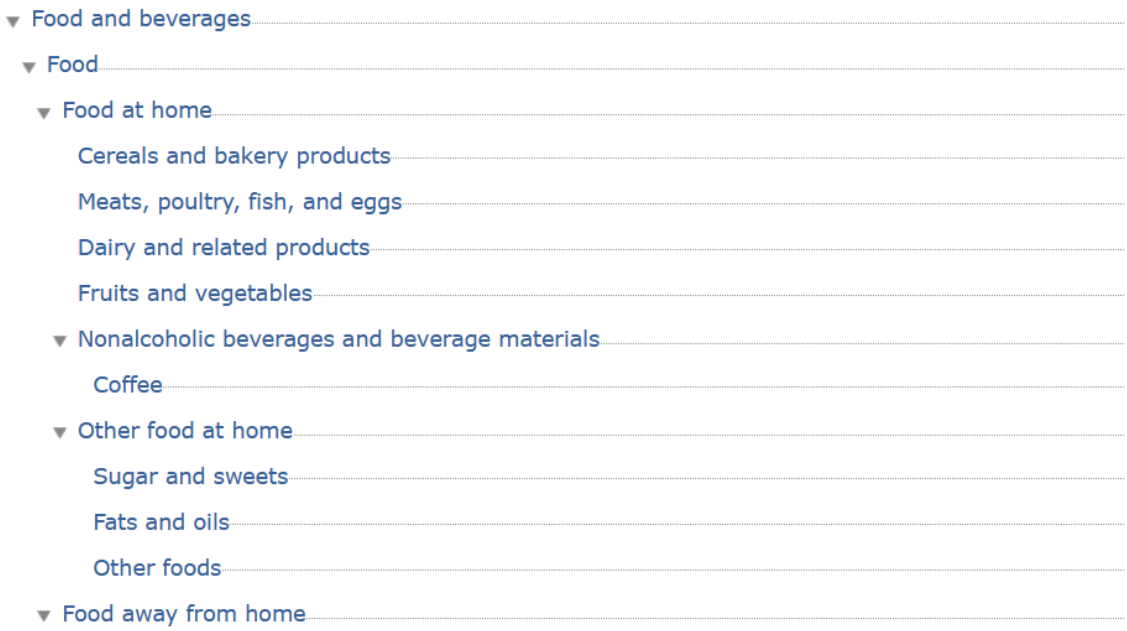

Figure 4: Example of the subcompents of the CPI

is used. Construction of a manual mapping is necessary since the names of the items change over time. For instance the item "fuels and utilities" has over 4 different names over time ("Fuels and utilities", "HOUSEHOLD FUELS AND OTHER UTILITIES", "Fuel and other utilities" and "Fuel and utilities")

The weighting data from the BLS war hard to clean. The history does not come through an API but rather requires the download of Excel, ZIP and text files. The format of the data has changed 5 times in total:

- The data between 1952 and 1986 comes in one Excel File
- Between 1952 and 1981 the data comes in a sheet of an Excel (several years in one sheet, 1-2 sheets per decade).
- Between 1982 and 1986 the data comes in several sheets (one sheet per year).
- Between 1987 and 2018 the data comes in individual txt files (one file per year). The format has changed considerably over time and is not machine friendly. The relevant items need to be extracted with regular expressions.
- Between 2018 and 2022 the data comes in individual txt files (one file per year).

3.2 Implementation

I decide to split the data as follows

- Train set: until December 2014
- Test set: January 2015 - now

Note that in computer science, test and train sets are often computed by random splitting. Here I do not want to use this approach because the data points are autocorrelated. I am afraid that if the model sees part of the corona inflation shock, it will adapt to it.

The data set has a lot of missing values as discussed in the previous sections. The standard approach is to drop the data so that we have a data set with no missing values. However, this simple approach is not appropriate here. By dropping the data, we miss periods that are very interesting from an economic point of view (such as the inflationary period in the 1970s). Moreover, the missing data should not be a big problem, as it often only concerns items that have a low weight in the CPI basket (such as audio equipment). So I decided to fill in the missing values as follows:

- Weight of the items in the CPI basket: Fill with 0 (realistic since the items simply did not exist). Rescale the weights so that they add up to one (see figure 3)
- CPI items: Fill with the value of the total CPI. This means that if an item is missing/absent, it is assumed to have simply moved in line with the overall price level.
- Other economic data: Fill with the insample mean of the data. Here I make sure not to use data that belong to the test set (i.e. data after January 2015).

As we can see in the table 3, some variables come in percentage points and can be assumed to be stationary, while other variables come in levels (i.e. exchange rates). For variables that come in levels, we take pct differences before passing them to the model. It does not make sense to forecast inflation using the absolute level, i.e. the GBP/USD exchange rate. It is more realistic to assume that the pct change in the exchange rate (i.e. the USD has gained 4% against the GBP) will affect the CPI.

3.2.1 Benchmark models

I start with simple (naive) models that can be used as benchmarks. Figure 5 shows the out-of-sample results of a model that simply uses 2% as the inflation forecast.

2% is the inflation forecast that the FED uses as its long-term target. The figure 6 shows the out-of-sample results of a model that simply uses the historical mean as the forecast. Since the historical mean was above 2% and the covid shock caused inflation to spike above 2%, the mean model outperforms the 2% model. Both models have the same RMSE (as they predict constant values), but the mean model has a lower average error.

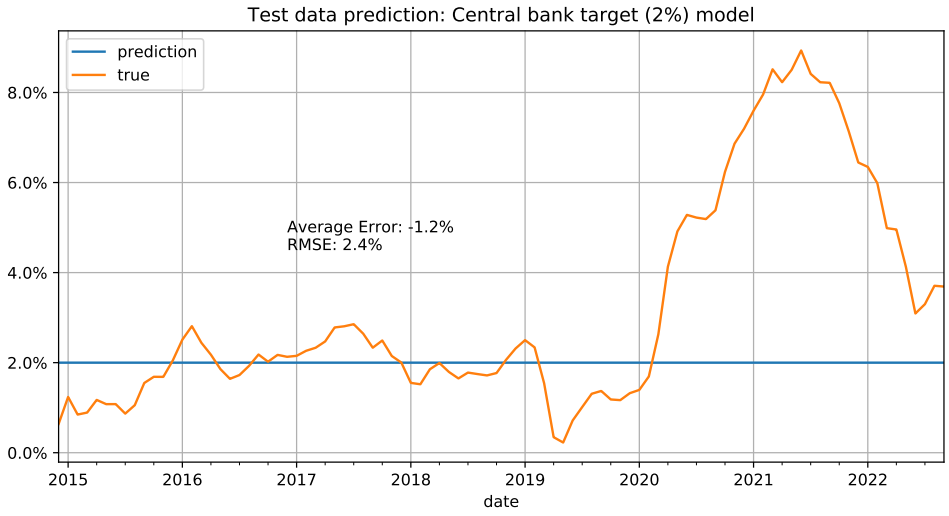

Figure 5: Out of sample precdiction of a benchmark model that predicts inflation to be at the central bank target (2%)

3.2.2 Univariate time series models

The next models I would like to test is to forecast inflation as follows:

$$CPI_{YoY,t,t+12} = \alpha + \beta \cdot MA(CPI_{MoM,t-6,t}) + \epsilon \tag{1}$$

I.e. these models look at predicting inflation using a historical average and a moving average (the average of the MoM changes over the last 6 months) to predict future inflation. I estimate the model using OLS and get the result shown in 7. The model has a better out-of-sample RMSE than the naive one. The beta coefficient is estimated to be positive. This is interesting and means that the model does not capture the mean reversion properties of inflation, but rather predicts that high

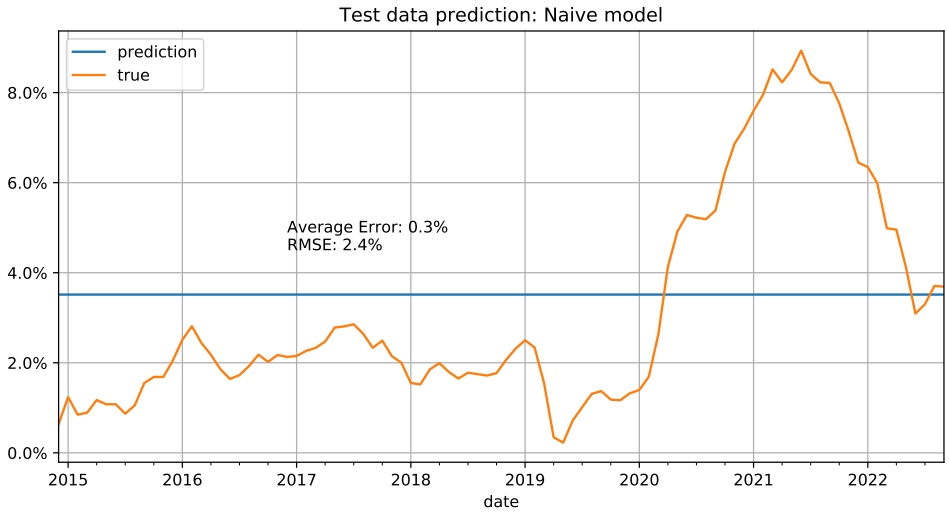

Figure 6: Out of sample precdiction of a benchmark model that predicts inflation to be at the historical (-2014) mean

inflation in the past is a sign of high inflation in the future. We can also see this clearly in the figure 7.

An extension of the model would be to use several moving averages (such as the 2-month and 6-month averages). The motivation is that we might want the model to react quickly to short-term CPI developments.

$$CPI_{YoY,t,t+12} = \alpha + \beta_1 \cdot MA(CPI_{MoM,t-6,t}) + \beta_2 \cdot MA(CPI_{MoM,t-2,t}) + \epsilon \qquad (2)$$

Figure 8 shows the results of such a model. The RMSE is lowered compared to the model in figure 7.

In section 3.3 we will use a cross-validation approach to find the best moving averages.

3.2.3 Bottom up aggregation models

In this section I use models that estimate each component of the CPI separately and then aggregate them using CPI weights. The motivation is that some parts of the CPI may behave differently from others (i.e. be more or less sticky/mean reverting). Each component of the CPI is modelled as in equation 1.

The results are shown in 9. Interestingly, the model gives a worse prediction than the simple model shown in 7. So it seems that it is not worth predicting each

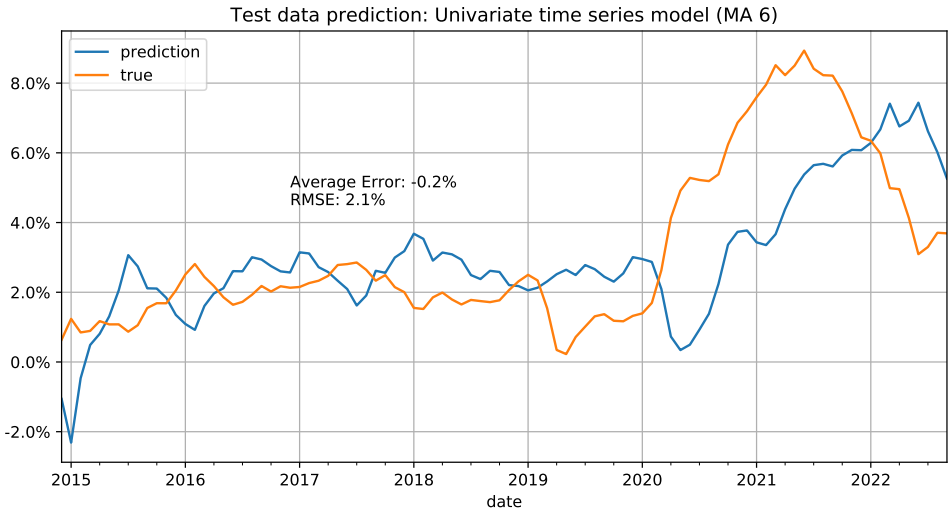

Figure 7: Out of sample precdiction of a model that uses the 6 month realized inflation to forecast the future inflation.

variable independently. In the table 4 I show the in-sample R2 I get for forecasting each component, as well as the average weight they have in the out-of-sample period. We see huge differences in the forecasting power of the individual components of the CPI. For example, food is very difficult to forecast and accounts for 14.4% of the CPI. Private transport is also not very predictable, but accounts for 15.9% of the CPI. In section 3.3 I will test whether it makes sense to only forecast some parts of the CPI and leave those that cannot be forecast.

3.2.4 Using additional economic variables

In this section I estimate the model presented in 7 but I add as independent variables the additional economic variables introduced in 3. The results are shown in 10. The model shows an improvement relative to the simple MA models outlined above. We will see in the section 3.3 if it makes sense to remove some variables.

3.3 Refinement

So far we have seen the following results.

	R2	avg. weight (2015-2023)
Private transportation	0.7%	15.9%
Public transportation	1.5%	1.1%
Womens and girls apparel	4.3%	1.2%
Footwear	9.3%	0.7%
Fuels and utilities	12.6%	4.9%
Tobacco and smoking products	14.0%	0.6%
Infants and toddlers apparel	15.8%	0.1%
Mens and boys apparel	18.6%	0.7%
Food	19.3%	14.4%
Household furnishings and operations	40.0%	4.7%
Shelter	48.3%	35.1%
Education	49.7%	3.1%
Personal care	51.8%	2.6%
Medical care services	53.6%	7.3%
Medical care commodities	54.4%	1.8%
Video and audio	56.3%	1.7%
Communication	60.8%	3.9%

Table 4: Forcasting power on the indivual CPI items, as well as average weight.

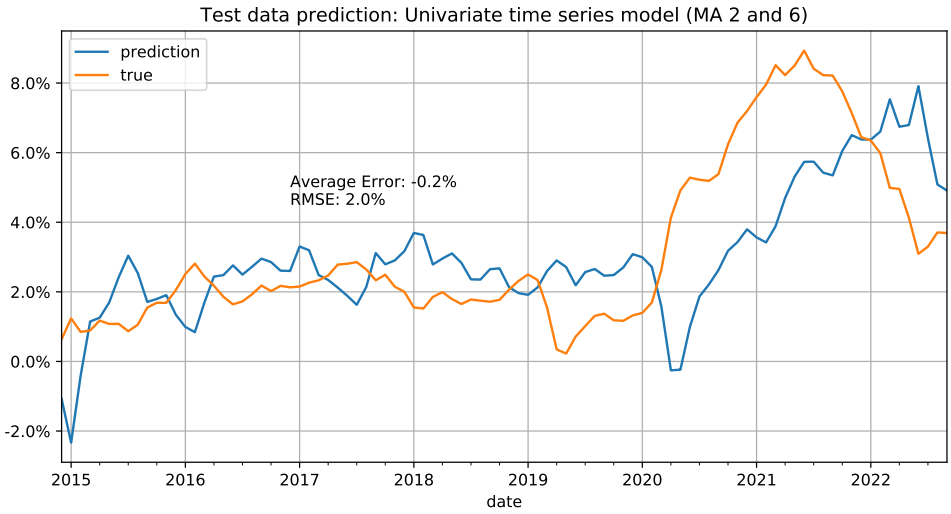

Figure 8: Out of sample precdiction of a model that uses the 2 and 6 month realized inflation to forecast the future inflation.

- Moving average models, which use 2-6 months of realised inflation to forecast future inflation, work reasonably well.
- Bottom-up aggregation models do not work. The components of the CPI basket differ in their difficulty of forecasting.
- The use of additional economic variables shows promising results. But so far we use too many of them.

In this section I use cross-validation techniques to fine-tune the parameters. The parameters are fitted to the training data (i.e. up to 2014). For the models in section3.2.2, I investigate whether other moving averages do a better job. I test all 1-18 month lags and up to 3 moving averages. So I test (using a grid search) 833 parameter combinations. The best model is one that uses the 1 and 7 month moving averages. It is shown in figure 11. The RMSE is further reduced compared to the previous models, although the difference is not large.

Regarding the bottom-up aggregation models presented in section 3.2.3, I test whether we get better results by simply aggregating some subcomponents. I do a grid search to select up to 5 subcomponents that we forecast individually and then aggregate them. The selected variables are 'shelter', 'medical services', 'video and

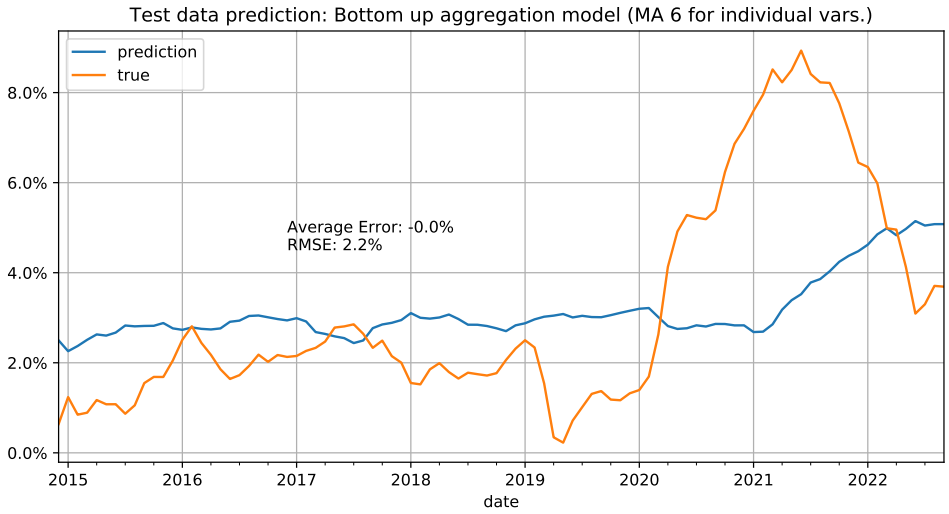

Figure 9: Out of sample precdiction of a model that predicts inflation components individually and sums them up using basket weights.

audio', The models perform worse than the simple moving average models described above. We therefore conclude that it is not worth aggregating the individual forecasts.

With regard to the models presented in section 3.2.4, I test whether we get better results by using only some of the economic varialbes outlined in table 3. I do a grid search to select up to 7 varialbes that can be used to improve the MA 1, 7 model. The motivation is to find a parsimonious model that can be expected to perform well out of sample. The result is shown in figure 13. We see a small out of sample improvement compared to the model 10. In addition, the model uses only 7 economic variables instead of 20 in model 10.

4 Results

4.1 Model Evaluation and Validation

The preferred model is 13. The model uses both a 1-month and a 7-month moving average of realised inflation and the 7 most important economic variables for forecasting inflation. The out-of-sample RMSE on the test data is 1.2%, clearly superior to the benchmark models presented in section 3.2.1. The linear model is shown in

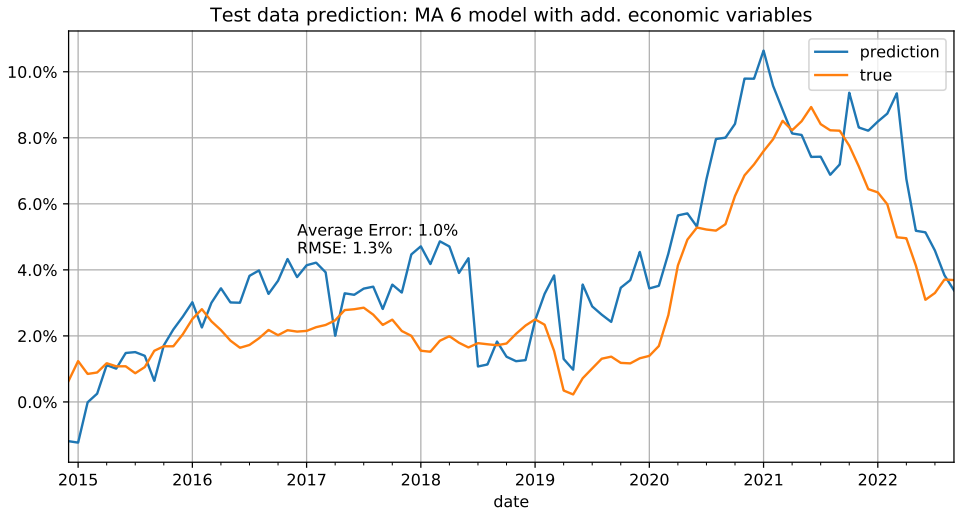

Figure 10: Out of sample precdiction of a model that uses the 6 month realized inflation and the variables outlined in table 3 to forecast the future inflation.

table 5. The variables chosen are reasonable and are motivated in table 3. It turns out that the money supply (money aggregate, M1, M2,..) is an important determinant of future inflation. In addition, monetary policy, as indicated by the FFER, the exchange rate and the volume of credit are chosen.

4.2 Justification

The chosen model is shown in 13. I chose a simple moving average model, enlarged with some economic variables, as a linear model. The reasons for the choice are:

- It is simple and easy to implement and interpret. It only requires the historical data of the CPI and the economic variables. It uses a linear regression method that is well-known and widely used in economics.
- It provides accurate and reliable forecasts that minimize the error and uncertainty.
- It uses a simple moving average technique that smooths out the fluctuations and outliers in the data, and it captures the momentum and direction of inflation

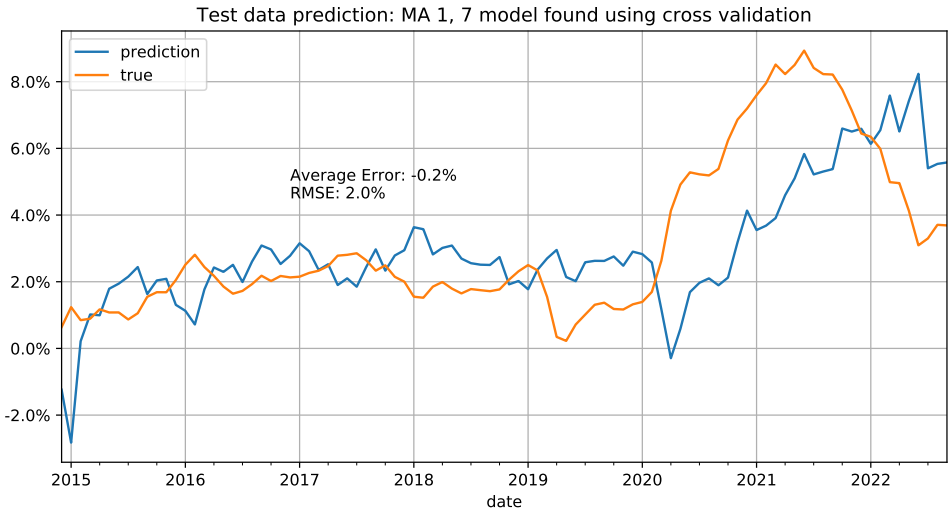

Figure 11: Out of sample precdiction of a model that uses the 1 and 7 month realized inflation to forecast the future inflation. This is the best parametrization found using cross validation.

changes. It uses a combination of a short (1 month) and a longer term moving average (7 month) to forecast inflation, which can capture both the short-term and the long-term dynamics of inflation.

- It captures the relevant information and factors that affect inflation, and it reflects the economic theory and intuition. It uses economic variables that are theoretically and empirically related to inflation, such as money supply, interest rates and exchange rates. It also uses money aggregates as the dominant variable.

As we have seen in previous sections it was tested on a out of sample period that is very special (2015-2023). In this period the chosen model has the lowest RMSE of all tested models as was shown on the different evaluations in the previous sections.

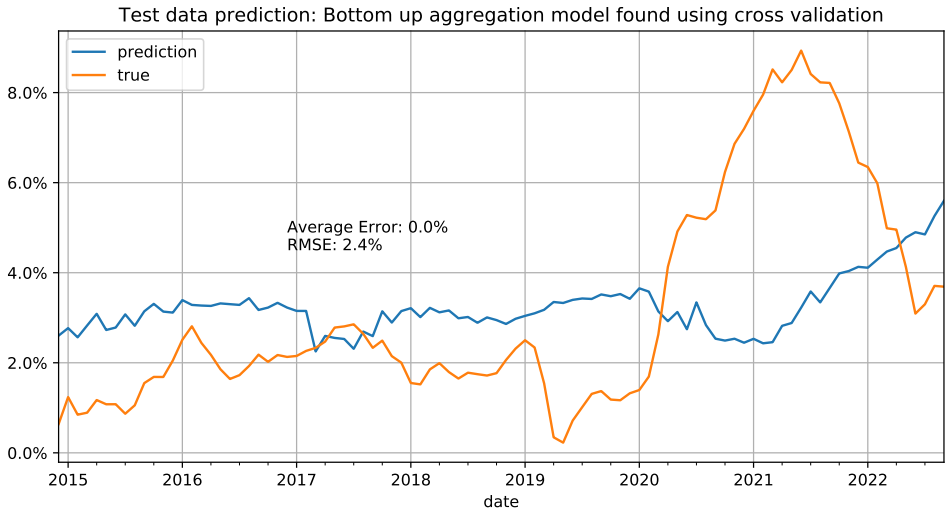

Figure 12: Out of sample precdiction of a model that predicts inflation components individually and sums them up using basket weights. Using cross validation we try to select up to 5 components that we forecast individually.

5 Conclusion

5.1 Reflection

In this project, I explored different methods of forecasting inflation using various data sources and models. I compared the performance of time series models, models that aggregate inflation forecasts based on individual components, and models enhanced with economic variables. I used cross-validation to select the best model and variables for each method.

- Time series models do reasonably well in forecasting inflation, and they are simple and easy to implement. They do not require much data cleaning or manipulation, and they can capture the trends and seasonality of inflation.
- Using a short (1 month) and a longer term moving average (7 month) to forecast inflation is a useful technique that can smooth out the fluctuations and outliers in the data. It can also capture the momentum and direction of inflation changes.

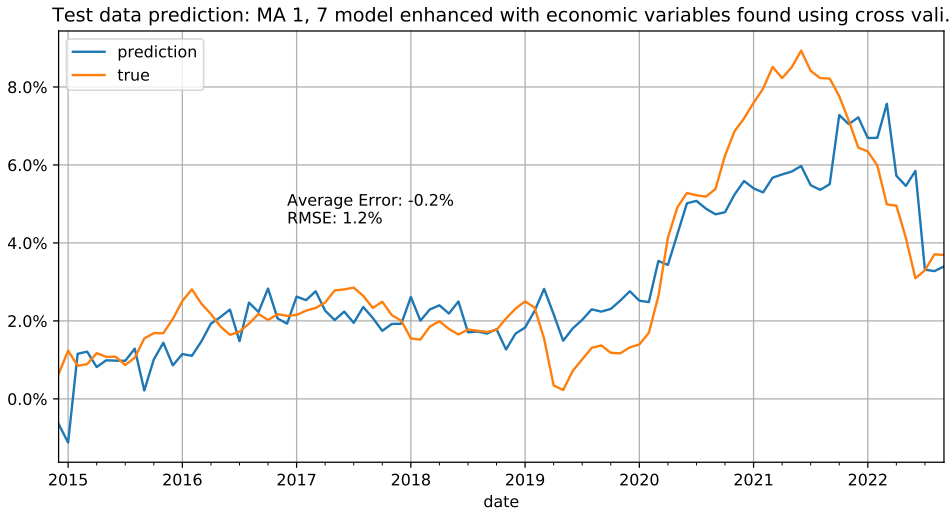

Figure 13: Out of sample precdiction of a model that uses the 1 and 7 month realized inflation as well as a subset (found by cross validation) of the variables outlined in table 3.

- Aggregating inflation forecasts based on individual components does not improve the accuracy of the forecasts. It is not worth estimating individual components of the CPI and aggregating them up, as this approach introduces more noise and complexity to the model. It is better to use the overall CPI as the target variable for forecasting inflation.

- Economic variables enhance the model and provide more information about the underlying factors that affect inflation. Interestingly, money aggregates play a dominant role in the selected variables, which suggests that the quantity theory of money still holds in the long run. Other variables that are relevant for inflation forecasting include interest rates, exchange rates. Not used for inflation prediction are oil prices and unemployment.

This project shows how hard it is to get a clean dataset that one can use and how easy it is then to play around with the models. It also demonstrates how the techniques learnt in the course can be well used in economics, such as scraping to find data, clean coding, and cross-validation to select variables. Although in economics advanced forecasting models are not used simply because there are not many datapoints (i.e. neural networks require much more data), the concepts learned in the course are still

Dep. Variable:	All items	**R-squared:**		0.569			
Model:	OLS	**Adj. R-squared:**		0.564			
Method:	Least Squares	**F-statistic:**		130.9			
		coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Moving average 1 month		1.3388	0.283	4.726	0.000	0.783	1.895
Moving average 7 months		2.7379	0.491	5.578	0.000	1.774	3.701
Inflation_Exp_Market_1YR		-0.0134	0.006	-2.400	0.017	-0.024	-0.002
Loans_Leases		0.0785	0.024	3.213	0.001	0.031	0.126
Real_M1		-0.4810	0.071	-6.733	0.000	-0.621	-0.341
M1		0.6235	0.068	9.230	0.000	0.491	0.756
Real_M2		0.1037	0.043	2.430	0.015	0.020	0.187
Real_Borad_Effective_Exchange		-5.186e-07	1.35e-06	-0.385	0.700	-3.16e-06	2.13e-06
FFER		0.0039	0.001	6.726	0.000	0.003	0.005
constant		0.0060	0.016	0.385	0.700	-0.025	0.037

Table 5: Preferred model. Estimation details. The linear model uses a 1 and 7 month moving average of realized inflation as well as 7 economic variables to forecast inflation.

very useful.

5.2 Improvement

For me it hard to understand why the bottom up aggregation approach does not work.

Some possible improvements for future work are:

- Exploring why the aggregation model does not work and how to individually forecast subcomponents of the CPI. For instance, shelter inflation could be forecasted using housing market specific factors or food inflation using wholesale prices on the stock exchanges. This can provide more granular and accurate forecasts for each component of inflation.

- Experimenting with different ways of aggregating the individual components using the CPI weights. For example, instead of using the basket weights (as done in the analysis here), we could try to see if there are some parts of the CPI that are more predictive for the overall CPI than others. We could also

use a dynamic weighting scheme that adjusts to the changes in the relative importance of each component.

- Apply nonlinear models that can handle different inflation dynamics if inflation is very high or very low.