This will need a VAR review: Decoding the Battle of Time Series Models in the United States Economy

Tools for Macroeconometrics Term Paper

Petr Čala

SS 2022/2023

May 13, 2023

This will need a VAR review: Decoding the Battle of Time Series Models in the United States Economy

1 Introduction

In this term paper, I will attempt to model the behavior of the United States economy using various macroeconomic models. Using real world data, I shall try to construct systems of models that will analyze the influence that different macroeconomic indicators have on each other, and after understanding this relationship, will suggest forecasts of these series for the future.

First I will fetch the data from a reliable online source, preprocess these data to account for any possible inconsistencies, orders of integration, structural breaks, etc. After cleaning the data, I will consolidate them into a single data frame, which I will attempt to use for multiple methods. This approach should give my estimations a degree of credibility and replicatibility.

As far as the models themselves are concerned, I will first attempt to model the system using simple autoregressive models such as ARIMA, or VAR. After gaining an initial picture of the data behavior, I will move onto estimation of more complex models, such as Structural VAR, VAR in levels, Vector Error Correction Model, or Bayesian VAR. Using these advanced methods, I hope to account for deeper potential problems of the series such as cointegration, non-linearity, etc.

The term paper is structured as follows. Section 2 deals with data acquisition and preprocessing. Section 3 focuses on simple linear methods and forecasting. Section 4 delves deeper by exploring the field of Vector Autoregression Models, their impulse response functions, and subsequent forecasts. Section 5 brings additional validation checks in the form of the more complex models outlined above. Section 6 concludes this paper.

2 Data description

Obtaining data

All data that we use for this paper come from an external source. Out of the numerous credible sources from which the United States data can be obtained, we opted for the FRED database. From here, we fetched the data using the *pdfetch* R package, and in total, acceseds the data of four time series. For the first part of the task, those are be Gross Domestic Product (GDP), and Consumer Price Index (CPI), while for the latter part, the data of the oil prices, and the Economic Policy Uncertainty Index.

When it comes to range, we always try to obtain data for the longest time span possible. The data of the first two series (GDP & CPI) range from the beginning of 1948 to the end of 2022, while for the latter two series, only data starting from January 1987 are available. Depending on the task, we will always use the conjunction of these ranges, e.g. the full range for the first part, and the shorter range for the second part.

As for the granularity of data, all series apart from GDP are available in monthly intervals, whereas GDP can only be obtained in quarterly form.

The last matter regarding fetching is the issue of seasonality. While for GDP and CPI we were able to obtain the already seasonally adjusted data, for oil prices and the uncertainty index, these series were available only in their unadjusted form. Consequently, we transformed the latter two series to account for seasonality using the *seas* package.

Data preprocessing

Data preprocessing consists of two major steps - logarithmization and transformation to percentage changes.

In order to ensure the procedure is directly comparable for all series, we opt to transform all of them to logarithmic form as suggested by the setup.

Then we address the issue of data interpretation. While in the first part of the task we work with period-over-period changes of the series (quarter-over-quarter for GDP, month-over-month for CPI), in the latter parts, we uniformly work with year-over-year changes to allow comparability. These are all percentage changes, meaning we obtain them as the first differences using function *diff* with the adequate difference period, and then multiply by a hundred. We also discard any missing observations.

Data validation

To verify that further procedures are indeed valid, we need to make sure our data behaves properly. This includes checking for stationarity, structural breaks, orders of integration, etc.

We employ a variety of tests to accomplish this - first we use the augmented Dickey-Fuller test to check for stationarity. Then we run the *breakpoints* function combined with *statcheck* to spot any structural breaks, and lastly we observe the periodogram, spectrum, as well as the autocorrelation and partial autocorrelation functions. This we do for all four of the series. In Figure 1 we include the plots of the transformed series. For results of the structural validation tests, see the appendix.

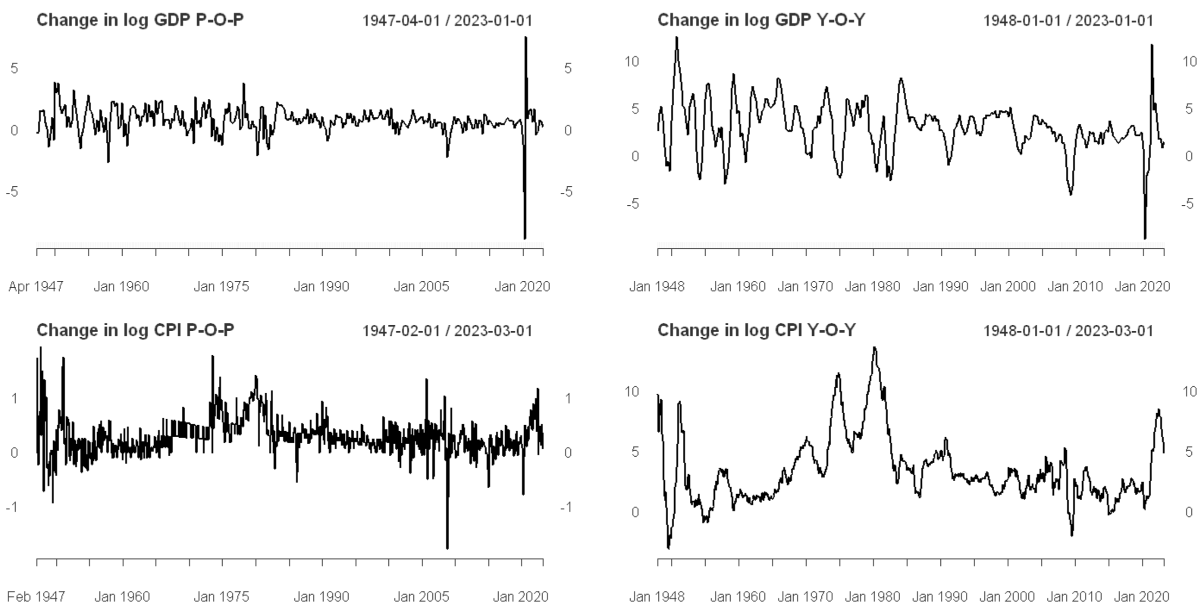


Figure 1: Percentage changes in GDP and CPI appear stationary

In summary, all our series behave very orderly, appearing strongly stationary by a glance as well as under more quantitative scrutiny. We observe no major structural breaks, nor is there a need to take second differences. The periodograms also suggest no obvious seasonality in either of the series.

3 Analysing Inflation and real GDP

In this section, we will try to estimate several linear models, to see which fit the two series the best.

Linear Model Estimation

Both of the series should be stationary, so linear models should work just fine in our case. We will model on the period-over-period changes in the data, which allow for the largest granularity. Further, we will supply year-over-year changes as a validity check.

For modelling, we will use the simple ARMA model. We do not consider any order of integration (ARIMA) for these series, as their behavior appears quite proper, and does not suggest any need for further differencing.

The ACF and PACF plots in the appendix suggest a simple model should suffice for both series, say ARMA(2,0) in the case of GDP, and ARMA(1,0) in the case of CPI.

These seem like a rather arbitrary choice, however, so in order to determine the best model in a more rigorous way, we ran a simulation for all 25 combinations of ARMA models of order 5 or below with various orders of the lag. This means ARMA(1,1), ARMA(2,1), ARMA(1,2), and so on, all the way until ARMA(5,5), all ran several times with different lags. We do not include all the results here, only point out the best candidates, which were *ARMA(3,1)* in the case of real GDP, and *ARMA(1,2)* for CPI.

To verify that these models are indeed suitable, we include more specification checks in the appendix. The residuals in the ACF/PCF plots of these functions seem well behaved in both cases, although there is the peculiar persistence in the 12th lag in the CPI residuals. We did a bit of exploration to see what could cause this, but the biggest suspect, seasonality trend, should be gone, so we suspect this may be caused by some outliers in the series.

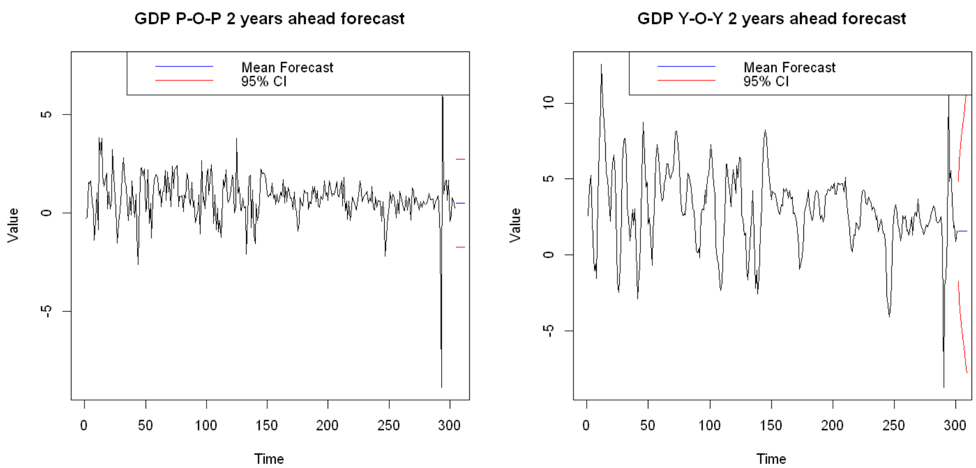
Linear Model Forecasting

Now that we have selected the most suitable models (*ARMA(3,1)* for real GDP, and *ARMA(1,2)* for CPI), we can run some forecasts.

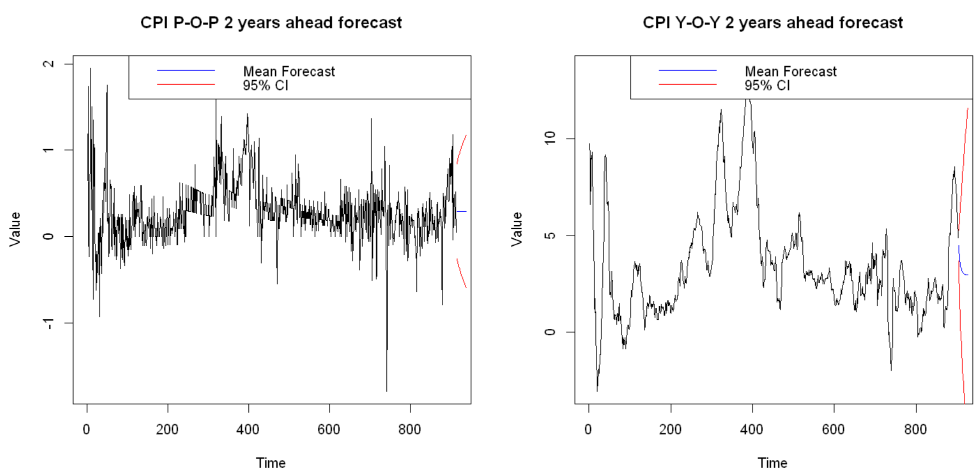
As mentioned before, we will be running 2 forecasts for each series - first, it's monthly for CPI, and quarterly for GDP, and then we have yearly for each of these two. This should be most sufficient, and in the case of GDP, we do not have the data to do more.

As for the forecasts themselves we choose to run a very simple form, using the *forecast* package. A rolling or expanding forecast could also make sense, to account for the new, forecasted data, but we may consider doing that in the future.

Figure 2: Univariate forecasts



(a) Percentage changes in real GDP



(b) Percentage changes in CPI

There is not much to be said about the results. It seems that in the P-O-P case, the forecasts just predict a linear continuation of the series in both cases, without much change, while in the Y-O-Y case, the suggested forecasts tend toward mean reversal. These claims also hold for both of the series.

4 Identification and Estimation Using VAR Models

Motivating the Choice of VARiables

For this task, given that the country in question is the United States, we have been explicitly asked not to use monetary policy shock due to the extensive exploration of this data during the exercises. As such, we chose to observe these two shocks - shock to oil prices (as it has had a significant impact on the U.S. economy historically), and the economic policy uncertainty shock.

When it comes to the process of obtaining and preprocessing the data, please refer to Section 2, where all should be explained. In Figure 3, you can observe the behavior of the two new series.

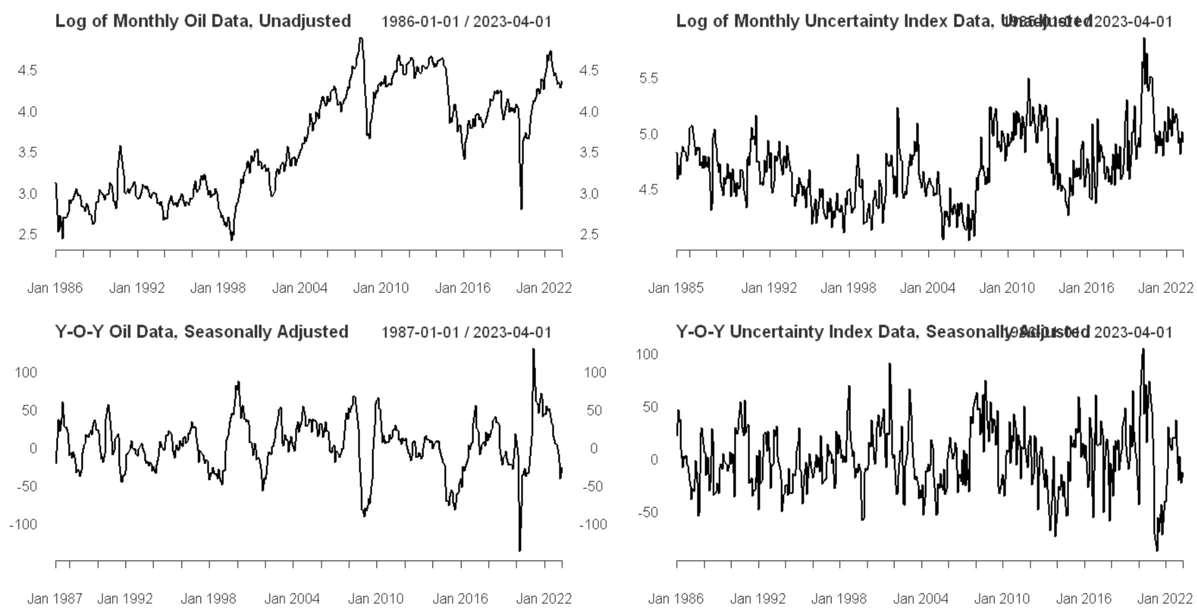


Figure 3: Log levels and percentage changes in the Uncertainty Index and Oil Prices

Ordering of VARiables and Structural Integrity

To ease the process of identification/estimation, we construct a single data set with all four series (we will refer to these interchangably as *variables* too. Here, the order in which we put the variables matters greatly, as it will determine the causal relationship among these variables further on. Let us assume that the following is true:

- GDP afftects all other variables contemporaneously
- Inflation affects all variables but GDP growth
- Uncertainty index affects oil prices only

As for the contemporaneous relationship restrictions in the SVAR model that will help us with identification, we will leave those for later. For now, let us assume this simple, ordered contemporaneous causal relationship that we outlined, and order the variables as follows:

1. GDP growth
2. Inflation
3. Change in uncertainty index
4. Change in oil prices

Interestingly, the new data seems way more volatile in terms of scale. While most percentage changes for GDP growth or inflation stay in the low single digits, sudden doubling or complete dwindling of the oil prices of the uncertainty index over the year are not uncommon. The difference in scales could have a negative impact on the identification, regression, and forecasting, so we decided to standardize the data, so that the output is more clear, and easily readable. Both unstandardized and standardized data can be viewed in Table 1.

We also ran all the structural checks that we did in Section 3 and found both the series well behaved.

Identification Using Cholesky Decomposition

We first make use of the *VARselect* function for lag length selection. Afterwards, we will construct the Vector Autoregression model using the suggested number of lags, and using this model, we will discuss the issue of identification.

The suggested lag number is split equally between 1 and 8, as can be seen in Table 2. Although 8 is the maximum allowed number of lags we permitted for suggestion ($2*n$, where

Table 1: Sample of unstandardized/standardized data for VAR

		Panel A: Unstandardized data		
Date	**GDP Growth**	**Inflation**	**Uncertainty Index**	**Oil Prices**
1987-01-01	2.67	1.35	-24.44	-20.652
1987-04-01	3.30	3.61	-30.48	37.48
1987-07-01	3.21	3.85	-21.69	61.04
1987-10-01	4.38	4.26	25.29	28.73
⋮	⋮	⋮	⋮	⋮
2022-01-01	3.61	7.32	-40.09	47.02
2022-04-01	1.78	7.90	21.38	50.02
2022-07-01	1.92	8.07	20.08	33.77
2022-10-01	0.87	7.47	37.49	7.18
		Panel B: Standardized data		
Date	**GDP Growth**	**Inflation**	**Uncertainty Index**	**Oil Prices**
1987-01-01	0.10	-0.86	-0.79	-0.7
1987-04-01	0.40	0.57	-0.98	0.91
1987-07-01	0.36	0.72	-0.70	1.57
1987-10-01	0.92	0.98	0.78	0 .66
⋮	⋮	⋮	⋮	⋮
2022-01-01	0.55	2.92	-1.28	1.18
2022-04-01	-0.33	3.29	0.66	1.26
2022-07-01	-0.26	3.40	0.62	0.80
2022-10-01	-0.77	3.02	1.17	0.05

Note: This table shows the panel data for four series for the simple Vector Autoregression Models (VAR). In Panel A, unstandardized, raw data for four series is displayed, while in Panel B, the data is standardized to a comparable scale. GDP Growth = Percentage changes in real Gross Domestic Product. Inflation = Percentage changes in real Consumer Price Index. Uncertainty Index = Percentage changes in the Economic Uncertainty Index. Oil Prices = Percentage changes in Oil Prices. Both panels contain 144 observations.

Table 2: Results of the $VARselect$ function

	AIC(n)	**HQ(n)**	**SC(n)**	**FPE(n)**
Suggested lags:	8	1	1	8

Note: This table shows the results of running the $VARselect$ function on the standardized data. Displayed are the suggested number of lags to use in the VAR estimation. AIC = Akaike Information Criterion. HQ = Harley-Quinn Information Criterion. SC = The Schwarz Criterion. FPE = Akaike's Final Prediction Error.

$n = 4$ in our case), we find the higher number of lags more suitable than using just a single one.

We then run the Cholesky decomposition. Here, the identification is handled implicitly through the ordering the variables. The variables unaffected by others will come first in the system, while the most endogenous variables will come last. This will, after constructing the lower-triangular matrix, determine the identification of structural shocks. The impulse response functions of the Cholesky decomposition can be found in Figure 4.[1]

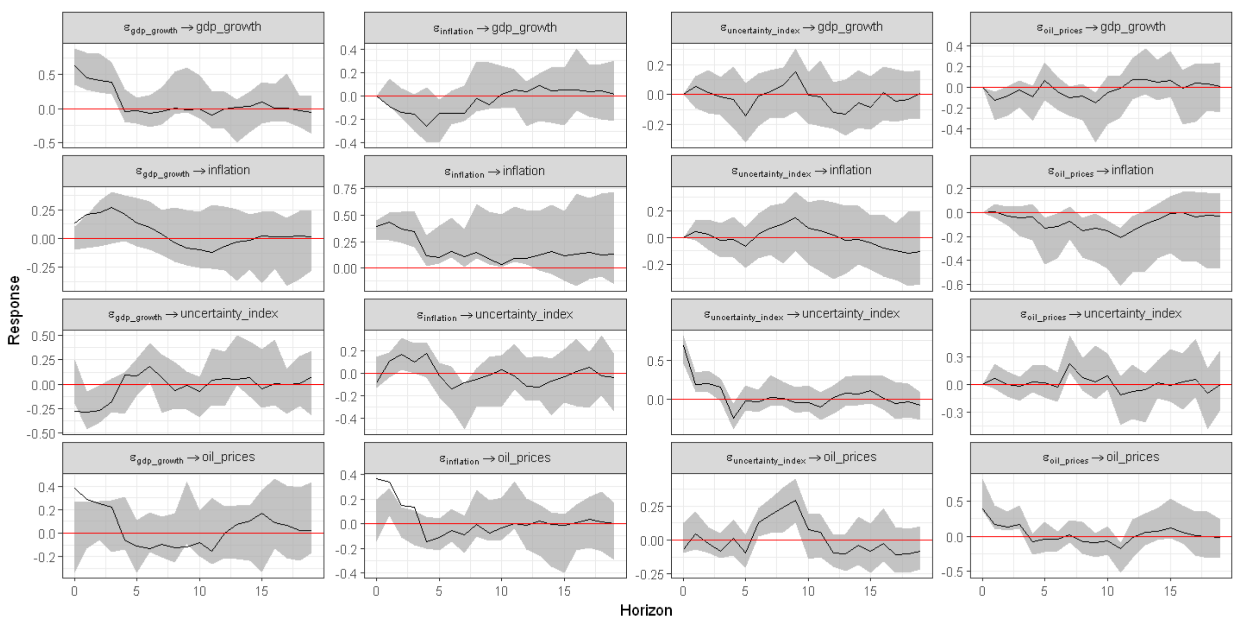


Figure 4: Cholesky decomposition Impulse Response Functions

With the Cholesky decomposition, we were able to achieve identification quite easily, and the VAR Impulse Response Functions can be seen in the plots above. For most of the shock resopnses, we can not claim a clear/signification impact. All of the variables have, however, a clear, positive reponse to shocks in themselves, in the short run. This makes sense, as any shock to a variable can be expected to come with a bit of friction before it reverts back to the mean, whether the shock is in this or that direction.

There are some obvious relationships, such as shock in the inflation making oil prices rise, or growth in GDP causing slight, short term rise in inflation. Also it is good to note, that the uncertainty index and oil prices should not be interpreted as such, because we transformed

[1] In case the figure descriptions are clearly visible, the variable order is GDP growth, Inflation, Uncertainty Index, Oil Prices from left to right, top to bottom.

them into percentage year-over-year changes. So the representation would be something like "change in oil prices", or "change in the uncertainty index".

Identification Using Structural VAR

The Cholesky decomposition procedure itself is simple, as it requires no additional assumptions from our side. The same is not true for the SVAR model, which requires imposing restrictions on the contemporanous causal relationships among variables. Let us define the structural matrix, that will serve us for this imposition.

As for the choice of restrictions, our reasoning is as follows - we would assume that GDP growth and inflation might be simultaneously influenced by each other and oil prices, while the uncertainty index is assumed to have no direct contemporaneous effect on GDP growth or inflation. These stem from intuition, as well as macroeconomic theory - (Bloom 2009; Barsky & Kilian 2002; Blanchard & Gali 2007).

To explain a bit more in detail, we would guess that there is a contemporaneous effect of GDP growth on inflation or vice versa. This captures the possibility that, in the short run, changes in GDP growth could impact inflation or that inflation could impact GDP growth. For example, an unexpected increase in demand could cause both GDP growth and inflation to rise simultaneously.

Furthermore, changes in oil prices can have an immediate impact on inflation due to their effect on production costs and consumer prices. Conversely, changes in inflation could impact oil prices through changes in the overall price level or shifts in demand for oil.

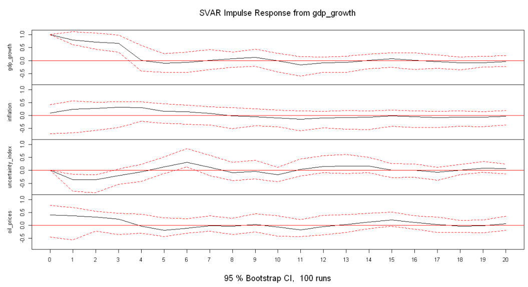
As a last thing to keep in mind, we need to keep the number of restrictions to the number of coefficients under the diagonal, which in this case is 6.

Consequently, the structural identification matrix looks as follows:

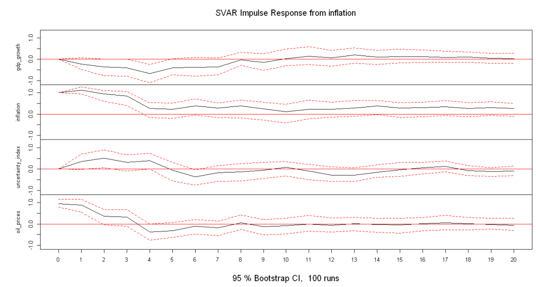
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \text{NA} & 1 & 0 & \text{NA} \\ 0 & 0 & 1 & \text{NA} \\ \text{NA} & \text{NA} & \text{NA} & 1 \end{pmatrix}$$

With this matrix, we can run the Structural VAR, and obtain the impulse response functions, which can be seen below in figure Figure 5.

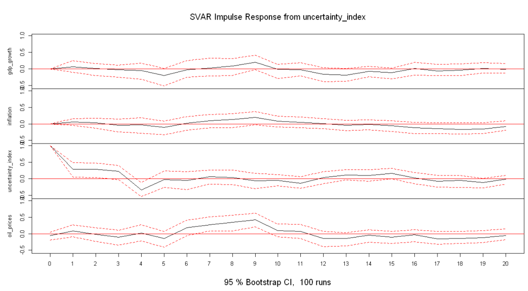
Figure 5: Impulse Response Functions of the Structural VAR



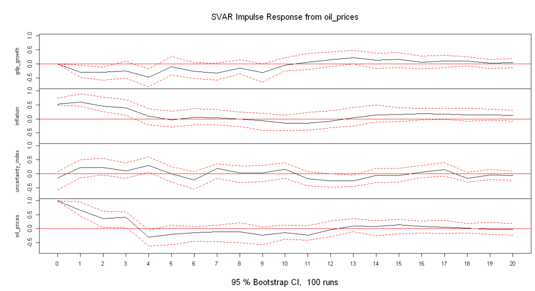
(a) IR from GDP



(b) IR from Inflation



(c) IR from Uncertainty Index



(d) IR from Oil Prices

From a glance, there appear to be no major *structural* changes in the results. It may be worth using the impulse response functions as a baseline for further comparison in the next section, but on its own, we observe no major differences compared to the Cholesky decomposition Impulse Response Functions.

Forecast Comparison

Let us now use the models we constructed in this section to run more forecasts, and compare these new forecasts with the ones we have obtained during the univariate modelling.

First of all, a comment about the results. They seem stable, depicting mostly mean-reverting behavior. Most of the confidence bands also appear quite narrow, which is a sign that the model is fairly confident when it comes to the predictions. It is difficult to say just from this simple forecast how much the volatility during the 2020 crisis affected the forecasts, we would have to dive in a bit deeper to claim anything with certainty.

In any case, the forecasts seem immediately *smarter* than their univariate counterpart, where

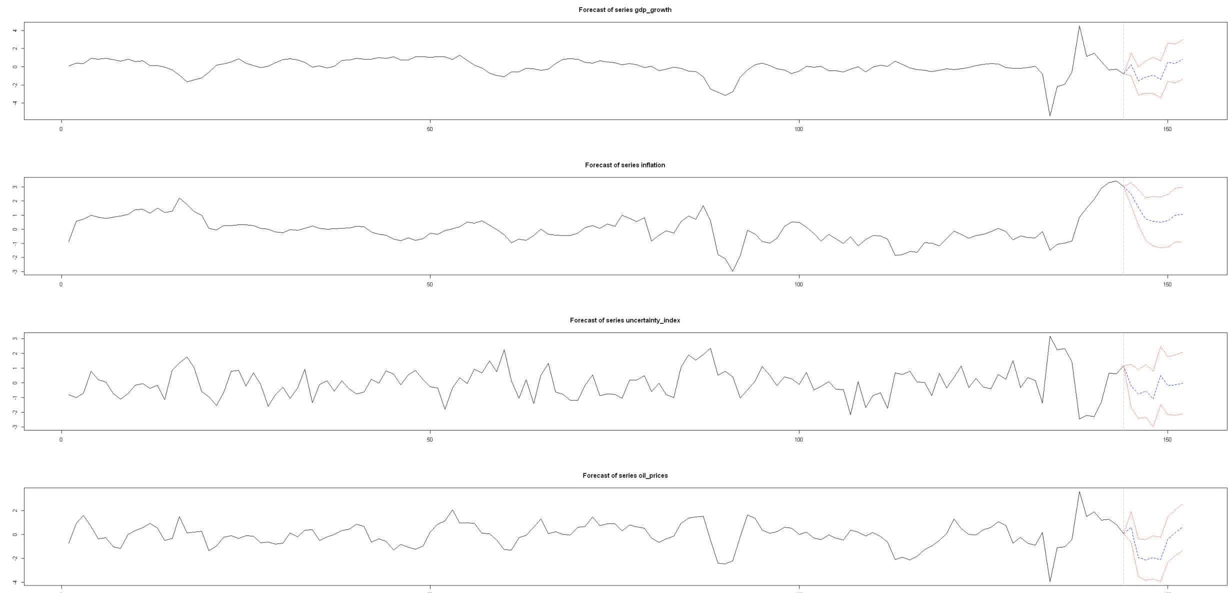


Figure 6: VAR forecasts appear almost unanimously optimistic

the predicted behavior was but a straight line copying the latest trend. Furthermore, these are almost unanimously optimistic, suggesting growth in GDP, and fall in inflation and in oil prices. They will serve as a good benchmark for the last section of this paper.

5 Taking It to the Next Level

Motivation for Extensions

In the last section of this paper, I would like to extend the scope of the analysis by adding several new methods - VAR in Levels, Vector Error Correction Model (VECM), Bayesian VAR (BVAR), and Local Projections. Relying solely on basic Vector Autoregression models might not capture all necessary information, leading to potentially inaccurate forecasts. The proposed extensions may come in several ways. For example, VAR in levels allows for the inclusion of non-stationary series, while controlling for cointegration is crucial to avoid spurious regression results and false interpretations. A Vector Error Correction Model can be employed to handle cointegration among variables and to help explain long-term relationships. Further, Bayesian VAR (BVAR) introduces prior knowledge into the forecasting model, which can be particularly beneficial when dealing with small sample sizes or when robust external information is available. Lastly, local projections can offer a flexible and straightforward way to estimate impulse responses without the need for specifying a particular dynamic system. With this reasoning, I believe adding the methods to the analysis makes a lot of sense as an additional robustness check.

However, before proceeding any further, I must address the issue of data validity and variable selection.

Addressing data validity and variable selection

In order to provide a more robust set of validation checks, I make the following changes to the data set:

- Add interest rate to the list of VAR variables
- Transform the variables based on an *aggregate first, then take first differences* approach, contrary to the *difference right away with 4 times more lags* approach I chose in the previous section

As such, I will first be making the necessary data transformations, adding the interest rate data onto the third position (below inflation and above uncertainty index). This stems from the assumption that changes in inflation will motivate the central bank to policy adjustments, while the index and consequent market prices are but a (partial) product of the fiscal policy.

These new data should serve us as a robustness check against Section 4, and so I will apply the changes to this part and this part and this part only. Sample of the transformed data (standardized in percentage changes, as well as in levels) can be found in the appendix.

VAR in levels

For this section, we will employ data from Panel B of Table 6, which come in log levels. As previously, we will first run the *VARselect* function to see what the suggested optimal number of lags is (see Table 3).

Table 3: Results of the $VARselect$ function on the log level data

	AIC(n)	**HQ(n)**	**SC(n)**	**FPE(n)**
Suggested lags:	2	2	1	2

Note: This table shows the results of running the $VARselect$ function on the log level data. Displayed are the suggested number of lags to use in the VAR estimation. AIC = Akaike Information Criterion. HQ = Harley-Quinn Information Criterion. SC = The Schwarz Criterion. FPE = Akaike's Final Prediction Error.

Right away, we can see that the number of suggested lags is significantly lower - 2 instead of 8. We run the VAR in levels using this number of lags, and present the impulse response functions from the model in Figure 7.

Most of the responses are as expected. Shock in GDP and in the uncertainty causes price growth, although in the latter case the response is positive and statistically significant only for but a single period across all observed periods. The response virtually disappears right after this.

Shock in the uncertainty index leads to a negative GDP growth, which is also a good sign and is in line with macroeconomic theory.

When it comes to predictions using the VAR levels model, these can be found in Figure 8. The GDP and prices predictions generally appear very stable, as opposed to the wide confidence intervals of the remaining 3 series. Most interestingly, the forecast for oil prices appears anything but positive, despite the series having a possible long-term growth trend. Otherwise, no surprising forecasts are present.

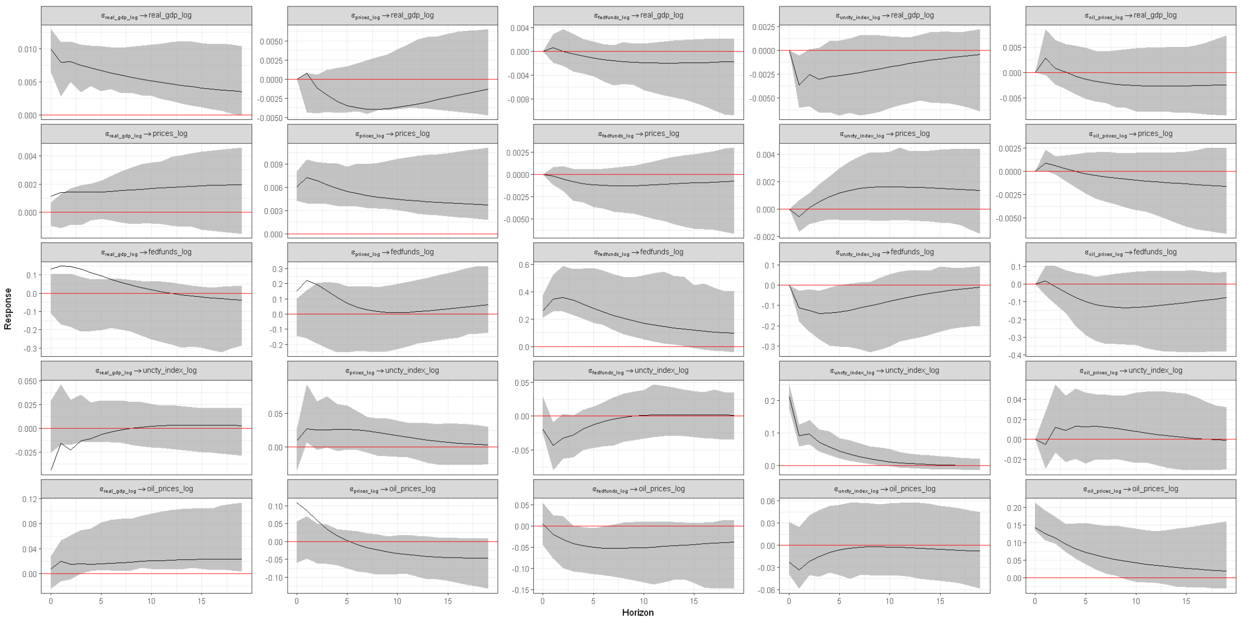


Figure 7: Impulse Response Functions for VAR in levels

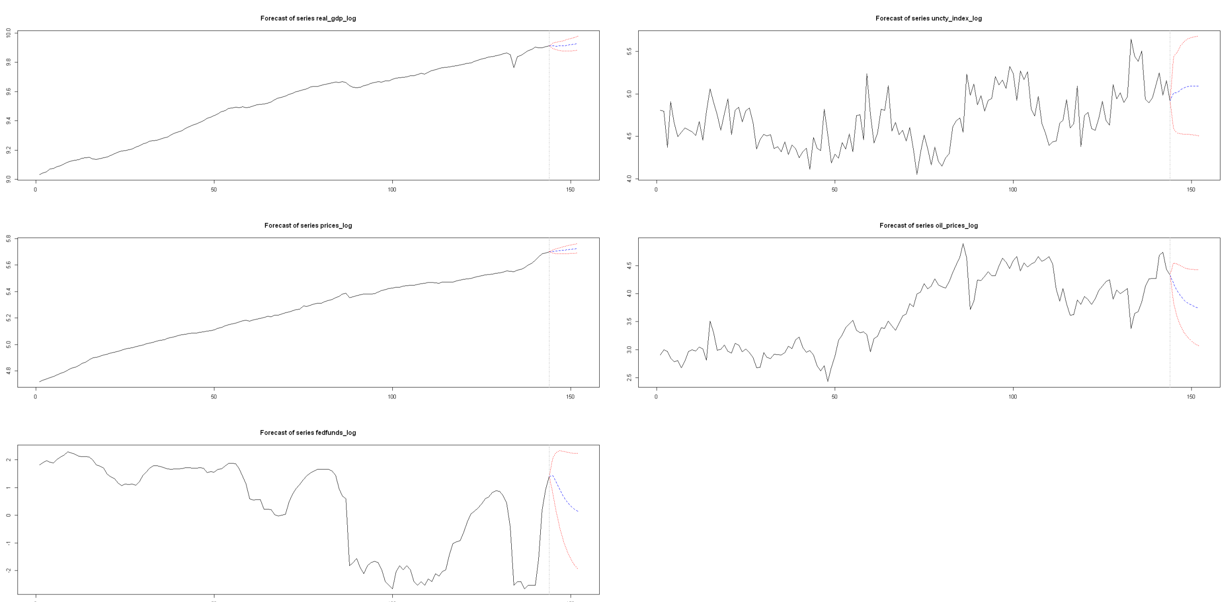


Figure 8: Forecasts for VAR in levels

Vector Error Correction Model

Now I test whether cointegration is present in our data. We can use the Johansen procedure for this (Hjalmarsson & Österholm 2007). Note that I still use the data in levels.

Table 4: Results of the Johansen procedure

	test	**10%**	**5%**	**1%**
$r \leq 4$	1.84	7.52	9.24	12.97
$r \leq 3$	7.45	13.75	15.67	20.20
$r \leq 2$	16.96	19.77	22.00	26.81
$r \leq 1$	33.22	25.56	28.14	33.24
$r = 0$	90.44	31.66	34.40	39.79

Note: This table shows the results the Johansen procedure. The first column displays the tested hypotheses, where r is the number of hypotheses cointegration vectors present in the data. test = Test statistic in the hypothesis testing. 10/5/1% = Significance values and their corresponding test statistics. If these are higher than the test statistic in the *test* column, the hypothesis can be rejected at this level of significance.

Based on the results of the Johansen procedure in Table 4, we can claim with near certainty that there are 2 cointegration vectors in our data. Possibly, one could claim the presence of but one vector, as the $r <= 1$ hypothesis gets rejected at 1% significance level given the test statistic 33.22, but given that the test statistic for $r <= 2$ is enough to reject all 3 significance levels, I assume the presence of no more than 2 cointegration vectors in our data.

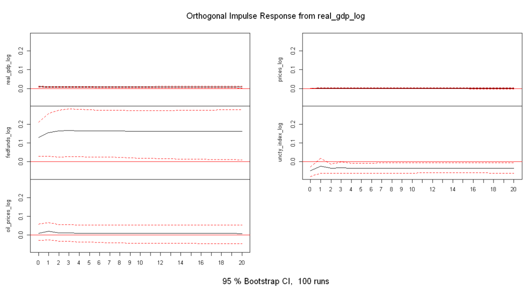
The actual Vector Error Correction Model itself is not of the most importance, so I do not display it here. If need may be, please refer to the source script of this paper.

The impulse response functions for the VECM can be found in Figure 9.

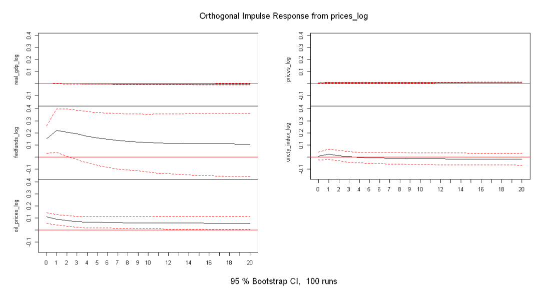
The response of real GDP to an oil price shock is positive for the first few periods, with a peak at an early period (probably around 2-3). This suggests that an increase in oil prices initially has a positive impact on real GDP, but the effect diminishes over time.

The response of prices to an oil price shock is also positive and peaks very early, which should mean higher oil prices lead to an increase in the general price level, which is consistent with the theory that oil price shocks can cause inflation.

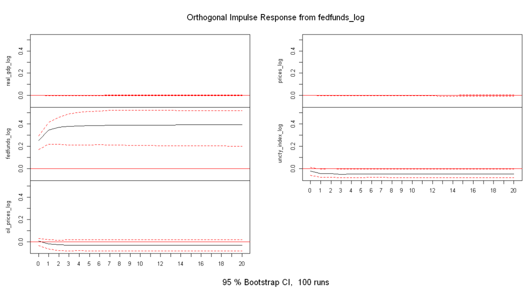
Figure 9: Impulse Response Functions of the VECM



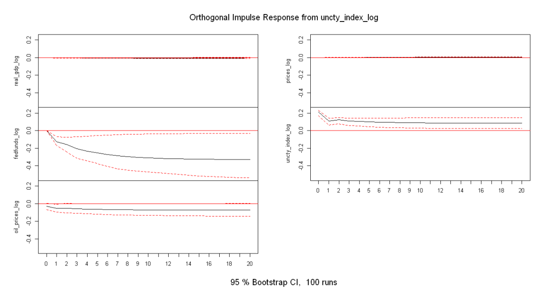
(a) IR from GDP



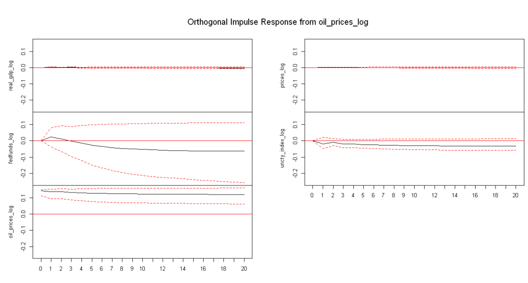
(b) IR from Inflation



(c) IR from the Interest Rate



(d) IR from Oil Prices



(e) IR from Uncertainty Index

As for the new variable in the model (compared to the last part), the federal funds rate, its response to an oil price shock is negative in the first few periods. This suggests that the central bank initially lowers the interest rate in response to higher oil prices, perhaps in an attempt to mitigate the negative impact on economic growth.

Lastly, we can see an interesting response of the uncertainty index to an oil price shock. The uncertainty index is negative in the first few periods, which could mean that an increase in oil prices tends to reduce economic uncertainty in the short run. However, the effect reverses and becomes positive after roughly 2 years, suggesting that the initial reduction in uncertainty may be temporary.

In order to display at least some of the results of this section, in Figure 10, I present the forecasts using the VECM model.

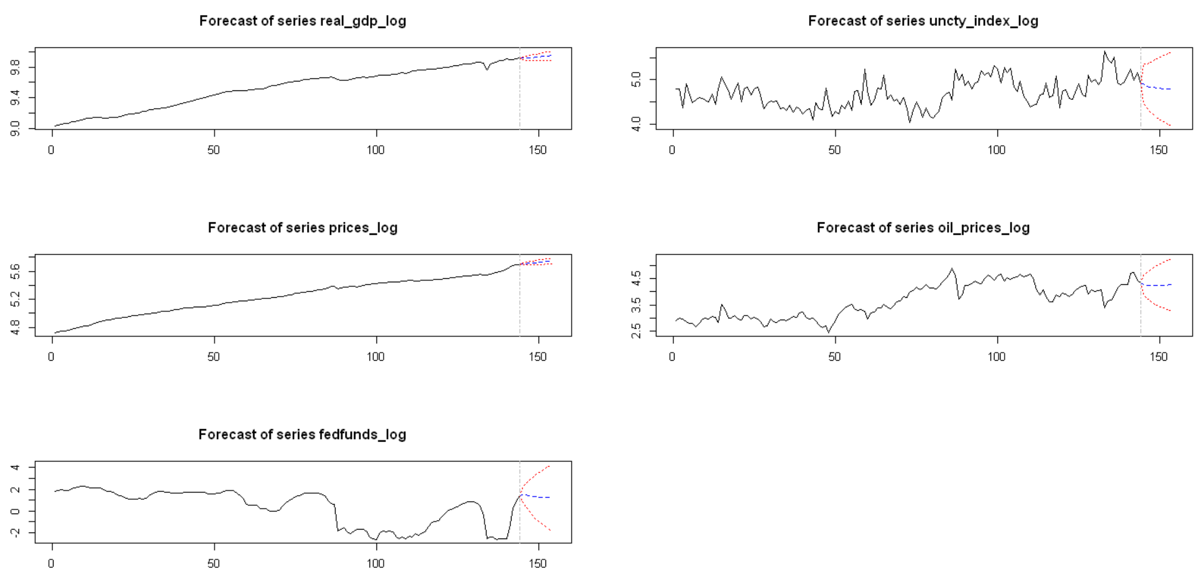


Figure 10: Forecasts for VECM

The one and immediately only visible change in the forecasts compared to the simple VAR in levels is the direction of the oil prices forecast. While in the previous section the series was predicted to have a uniformly downwards trend, in this case, the prediction is (at least from an unqualified perspective) more in line with the series' overall behavior - slight, but constant growth over long periods of time.

All in all, we could say that this model brings to the table more robust results than its alternatives.

Bayesian VAR

For the next section, I will try to estimate the series' behavior using Bayesian VAR (BVAR). Given that we are no longer dealing with VAR in levels, I will be working with the differenced data (but not from the stage 2 as suggested, but the new, more robust data, with the interest rate series). Again, see the the appendix for these, panel A of the corresponding table.

When it comes to the setup, I choose to emply simply the Minnesota prior (Sevinç & Erguen 2009). As for the number of lags for the estimation, I went for 6, more so due to a hunch than anything else. In Figure 11 and Figure 12, you can find the impulse response functions and forecast results from the Bayesian VAR.

The results seems quite stable at first glance. We will not delve deep into the individual coefficients from the BVAR, and will instead focus more on the comparison of the IRFs to the ones we obtained during the VAR in levels and VECM estimation.

The response of GDP to oil prices behaves virtually the same as it did with the VECM, as does the response of inflation.

On the other hand, the impact of an oil prices shock on both federal funds rate and uncertainty index is quite different. BVAR predicts an strong positive early response of the central bank, which is a little counter intuitive. Similarly, the sudden peak in the uncertainty index reponse around period 5 is hard to explain intuitively, and may just be caused by fluctuations in the data or a setup specification.

Generally speaking, the results are quite similar to the ones we obtained during VECM estimation. We can thus claim that these conclusions are the most robust out of the ones we have made thus far.

Given that we are not dealing with level data anymore (as we have been explicitely asked to estimate the procedure on data/model from the second part of the task), it is difficult to make direct comparisons between the forecasts of this model and the previous ones.

Nonetheless, the forecast for inflation is, even though only slightly, predicted to stay below 0 for most of the time, as is the case with GDP growth. Oil prices, uncertainty index, and fed funds are expected all to change very little (prediction near 0).

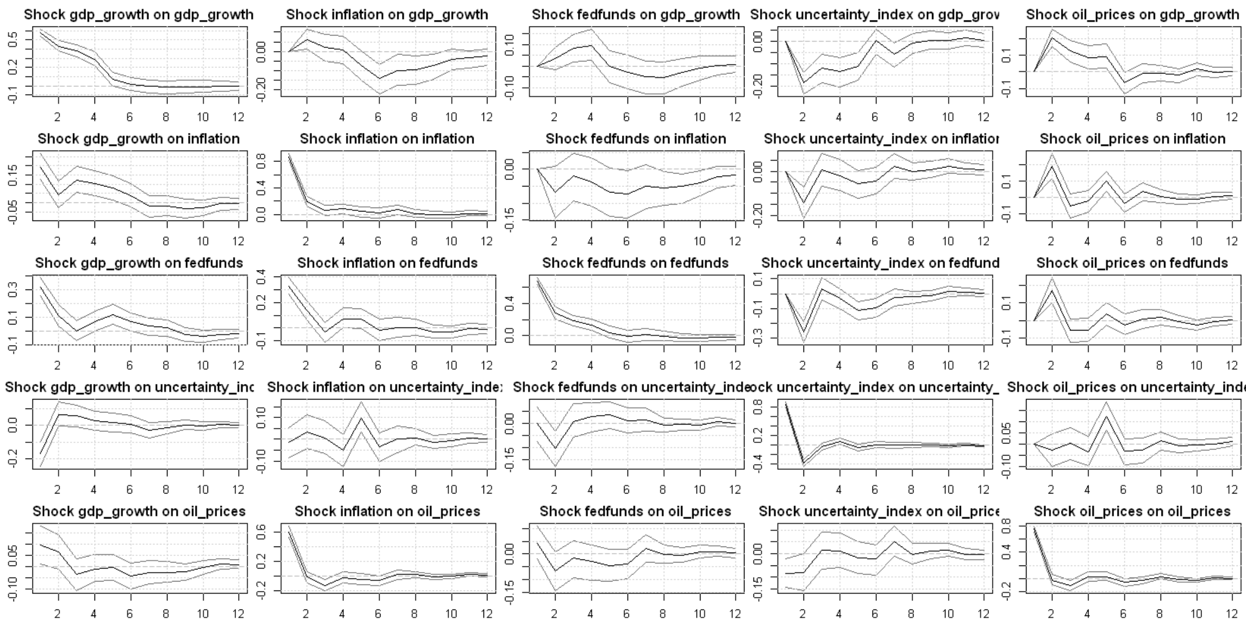


Figure 11: Bayesian VAR Impulse Response Functions

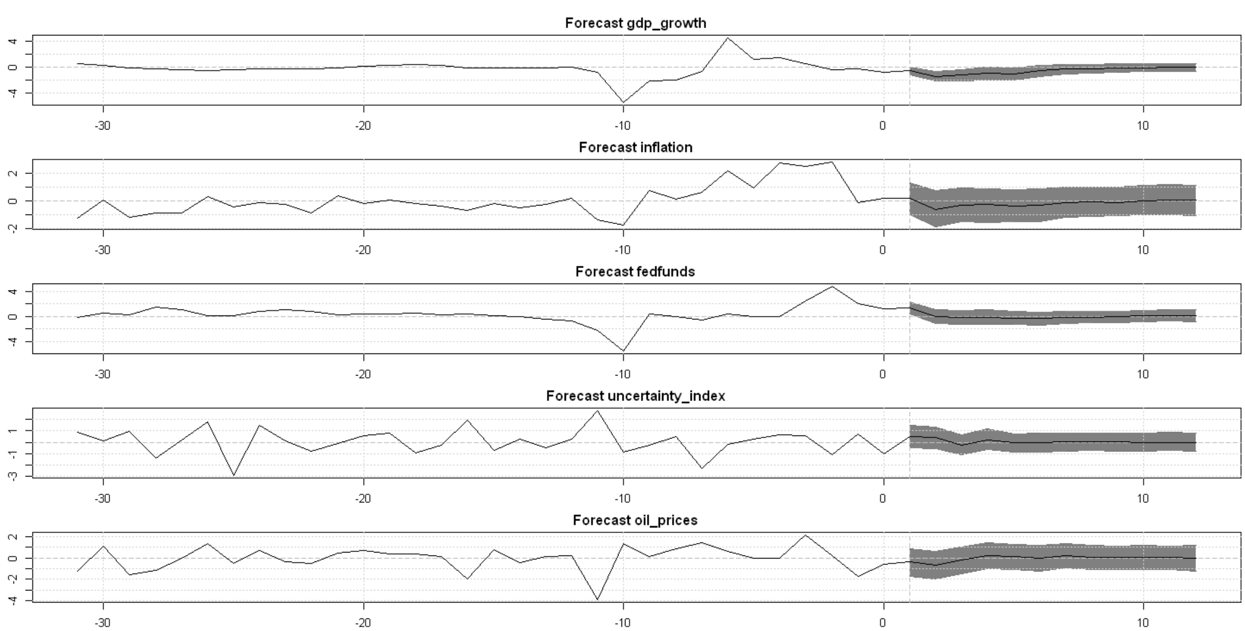


Figure 12: Bayeasian VAR Forecasts

Local Projections

As a last part of the exercise, we will construct several local projections. Again, we will be using the data set differenced data set, which does not contain log level data, but rather percentages. Not much is to be said about the setup for the estimation; the projections can be ran using a single R function *lp_lin*. For more details, refer to the source R script of the paper.

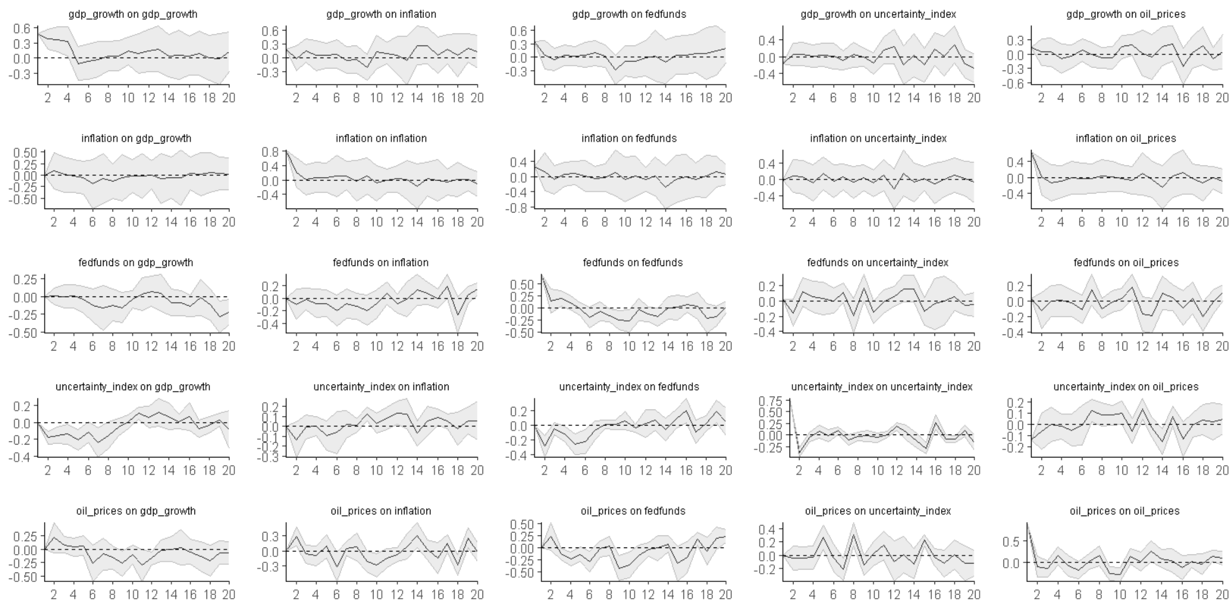


Figure 13: Impulse Response Functions of Local Projections

The results from Figure 13 tell us quite little. Most of the trends we observed with the previous methods are gone, and we can make out no clear picture. In the vast majority of the cases, the confidence bands are also too wide to claim any change of IRF behavior with certainty. As such, we consider this method only as an extra validity check.

6 Conclusion

In this term paper, I attempted to model the behavior of the United States economy using advanced macroeconomic models. Using models such Autoregressive Models, Vector Autoregressions, Bayesian Vector Autoregressions, or Structural Vector Autoregressions, I forecasted the behavior of five different indicators of the economy performance. Further, using these models, I infered the propagation of shocks within the model system, and compared the results within different models to see whether the propagation differs across time or across different methods.

Using real world data with observations spanning over 35 years, I focused on these indicators - Gross Domestic Product, Consumer Price Index, Federal Funds Rate, Economic Uncertainty Index, Oil Prices. I discovered that the simple models such as ARIMA or Vector Autoregressions did not hold much predicting power, and failed to detect empirically established relationship between variables. On the other hand, more robust models such as Structural VAR, VAR in levels, Vector Error Correction Model, or Bayesian VAR all displayed similar and stable results. By studying their impulse response functions and predictions, I was able to infer several strong variable relationships within the system, which coincide with macroeconomic theory.

All of these procedures served to give me an idea about how the economy of the United States behaves behind the obvious trends, and helped me explore a little futher into what may be causing the market to behave the way it does. They should, nevertheless, be taken with a grain of salt, as no estimation is without its flaws, and its validity may be disproven or at least questioned in the future.

References

Barsky, R. B. & Kilian, L. (2002). Oil and the macroeconomy since the 1970s. *Journal of Economic Perspectives*, *18*(4), 115–134.

Blanchard, O. J. & Gali, J. (2007). The macroeconomic effects of oil shocks: Why are the 2000s so different from the 1970s?

Bloom, N. (2009). The impact of uncertainty shocks. *econometrica*, *77*(3), 623–685.

Hjalmarsson, E. & Österholm, P. (2007). Testing for cointegration using the johansen methodology when variables are near-integrated.

Sevinç, V. & Erguen, G. (2009). Usage of different prior distributions in bayesian vector autoregressive models. *Hacettepe Journal of Mathematics and Statistics*, *38*(1), 85–93.

Appendix

Tables

Table 5: Box-Ljung test results for percentage changes in real GDP and CPI.

Lags	**GDP (X^2, p-value)**	**CPI (X^2, p-value)**
4	0.010786, 0.9173	0.84973, 0.9317
8	0.011105, 0.9945	1.9273, 0.9832
12	0.41463, 0.9372	47.682, 0.000003551
Model	*ARIMA(3,0,1)*	*ARIMA(1,0,2)*

Note: This table shows the Box-Ljung test results for the residuals of the Autoregressive Moving Average (ARMA) models on the percentage changes in real Gross Domestic Product (GDP) and Consumer Price Index (CPI). Each row corresponds to a different number of lags (4, 8, 12) used in the test. The X^2 value (Chi-square statistic) and the p-value for each test are reported in the corresponding cell.

Table 6: New data with added interest rate

Date	GDP	Inflation	Fed Funds	Uncertainty	Oil Prices
Panel A: Standardized data with Interest Rate					
1987-01-01	0.10	0.88	-0.30	0.45	0.62
1987-04-01	0.40	0.72	0.25	-0.03	0.43
1987-07-01	0.36	0.56	0.19	-1.74	-0.20
1987-10-01	0.92	0.14	-0.16	2.18	-0.71
⋮	⋮	⋮	⋮	⋮	
2020-07-01 -	2.18	0.76	0.32	-0.24	0.12
2020-10-01 -	1.93	0.12	0.00	0.48	0.85
2021-01-01 -	0.62	0.62	-0.65	-2.30	1.48
2021-04-01	4.49	2.21	0.36	-0.19	0.66
Panel B: Data in levels					
Date	**GDP**	**Inflation**	**Fed Funds**	**Uncertainty**	**Oil Prices**
1987-01-01	9.03	4.72	1.81	4.80	2.90
1987-04-01	9.04	4.73	1.90	4.79	2.99
1987-07-01	9.05	4.74	1.97	4.36	2.97
1987-10-01	9.06	4.75	1.91	4.90	2.84
⋮	⋮	⋮	⋮	⋮	
2020-07-01	9.83	5.56	-2.40	5.38	3.67
2020-10-01	9.84	5.56	-2.40	5.50	3.85
2021-01-01	9.86	5.57	-2.65	4.94	4.13
2021-04-01	9.88	5.60	-2.52	4.89	4.26

Note: This table shows the panel data for four series for the last section of the paper. In Panel A, standardizaed data with added information about the interest rate is displayed, while in Panel B, the data presented in log levels. GDP = Percentage changes in real Gross Domestic Product. Inflation = Percentage changes in real Consumer Price Index. Fed Funds = Federal Funds Interest Rate. Uncertainty = Percentage changes in the Economic Uncertainty Index. Oil Prices = Percentage changes in Oil Prices. Both panels contain 144 observations.

Correlation functions

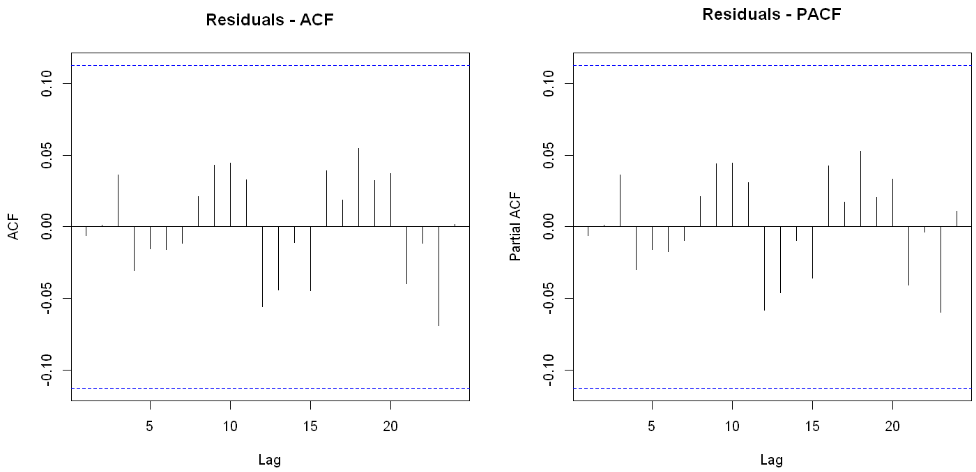


Figure 14: Correlation functions - real GDP - ARIMA(3,0,1)

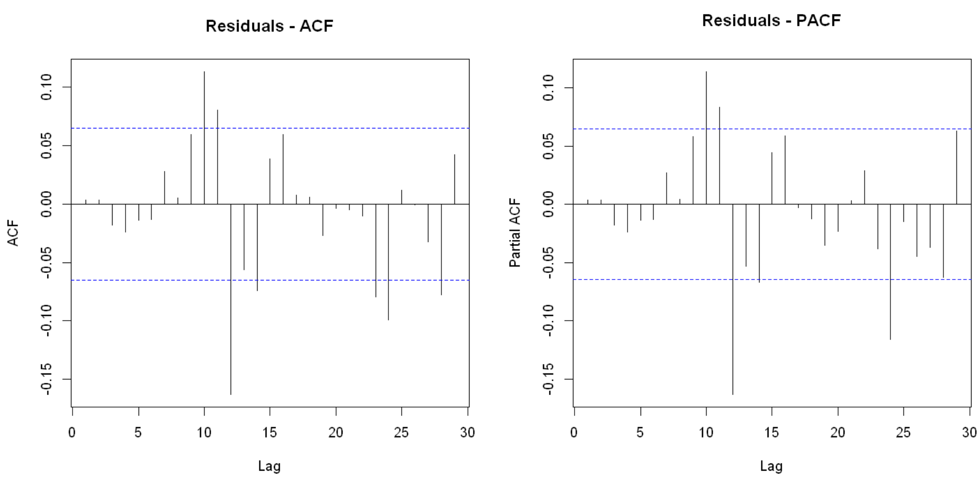


Figure 15: Correlation functions - CPI - ARIMA(1,0,2)