1.

```
> data_var=cbind(Yt,OIL_P,OIL_PROD,OIL_STOCKS,WORLD_IP,US_CPI)

> Reduced_VAR12 = VAR(data_var, p = 12, type = "const")
```

```
> VARselect(data_var, lag.max=25, type = "const")$selection

AIC(n)  HQ(n)  SC(n) FPE(n)
     4      2      2      4

$criteria

                1          2           3          4          5          6          7           8          9         10         11
AIC(n) -0.3964581 -0.8967478 -0.95954493 -0.9820788 -0.9744037 -0.9510328 -0.9696509 -0.91438635 -0.8687685 -0.8293339 -0.8226177
HQ(n)  -0.2593044 -0.6420337 -0.58727048 -0.4922440 -0.3670086 -0.2260773 -0.1271351  0.04568987  0.2088681  0.3658631  0.4901396
SC(n)  -0.0466939 -0.2471857 -0.01018492  0.2670791  0.5745521  0.8977209  1.1789007  1.53396315  1.8793789  2.2186114  2.5251255
FPE(n)  0.6727055  0.4079214  0.38314741  0.3747106  0.3777613  0.3869413  0.3801404  0.40221179  0.4216178  0.4393989  0.4433730

               12         13         14         15         16         17         18         19         20
AIC(n) -0.8070262 -0.8388341 -0.7885336 -0.7400193 -0.7525792 -0.6995828 -0.6315342 -0.6136871 -0.5629803
HQ(n)   0.6232915  0.7090439  0.8769048  1.0429794  1.1479798  1.3185366  1.5041456  1.6395531  1.8078201
SC(n)   2.8405149  3.1085049  3.4586033  3.8069155  4.0941535  4.4469478  4.8147943  5.1324393  5.4829439
FPE(n)  0.4515780  0.4388628  0.4632566  0.4884231  0.4847576  0.5140708  0.5538332  0.5679050  0.6022945
```

```
> residuals_VAR12 <- resid(Reduced_VAR12)

> acf(residuals_VAR12)
> pacf(residuals_VAR12)
> plot(Reduced_VAR12)
```

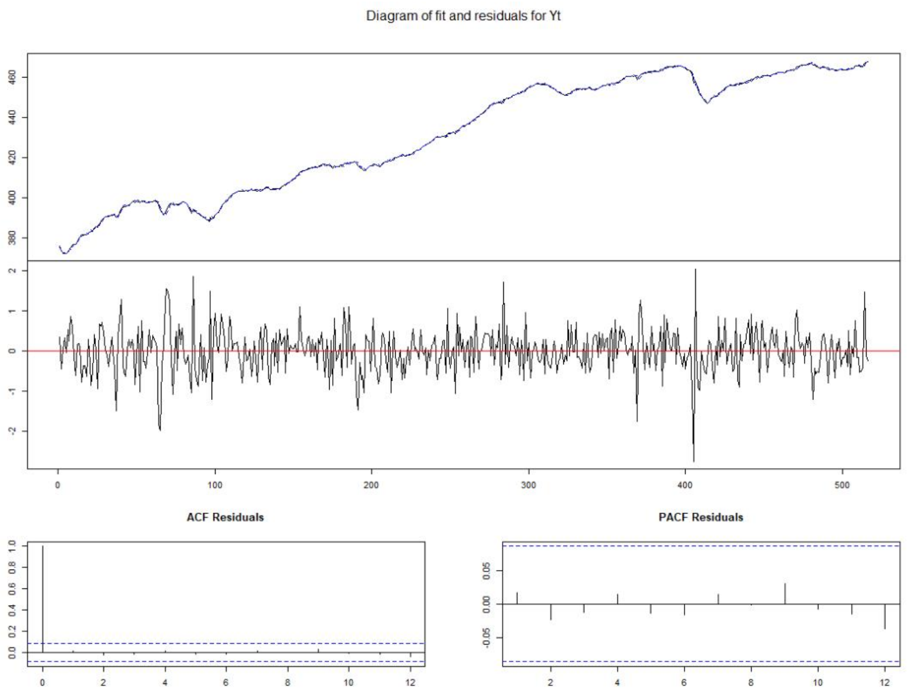


Diagram of fit and residuals for OIL_P

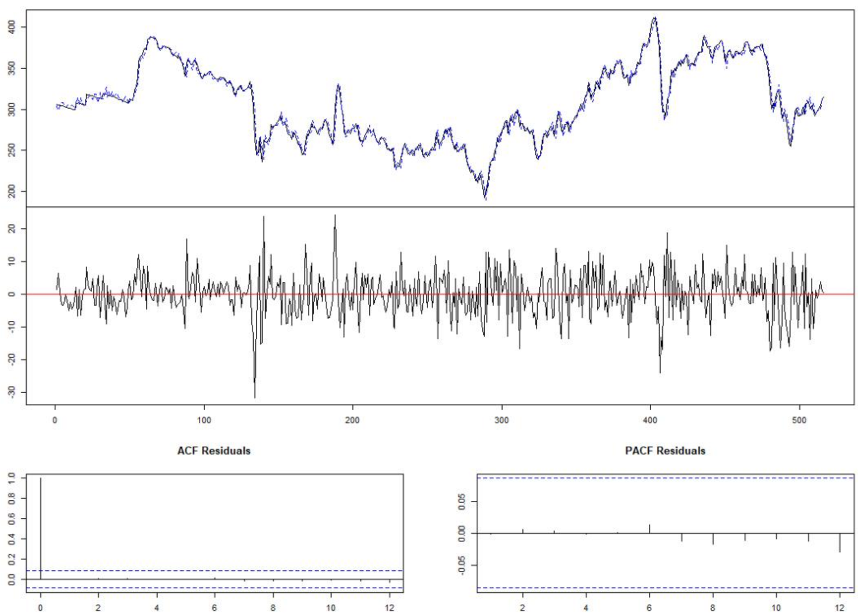


Diagram of fit and residuals for OIL_PROD

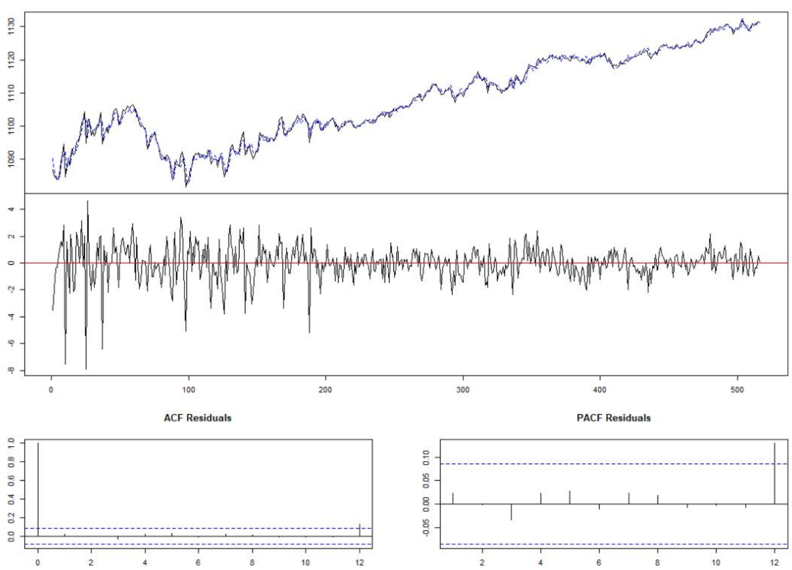


Diagram of fit and residuals for OIL_STOCKS

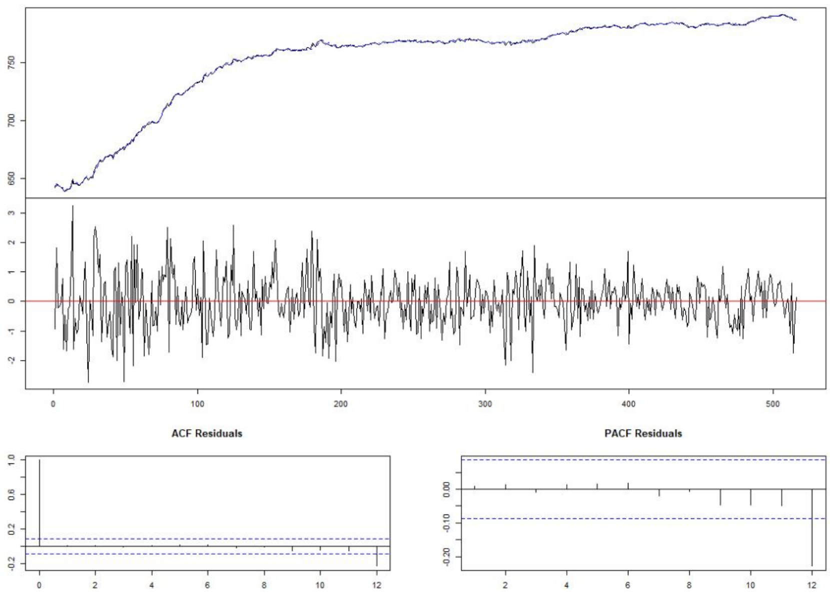
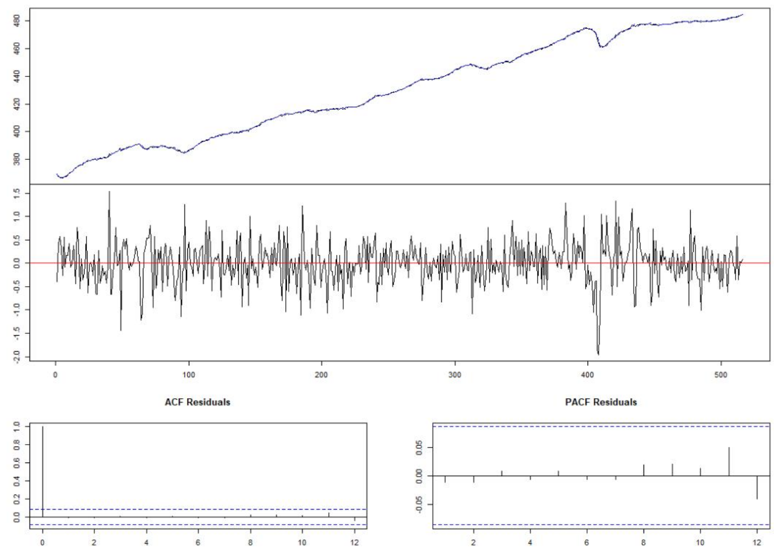
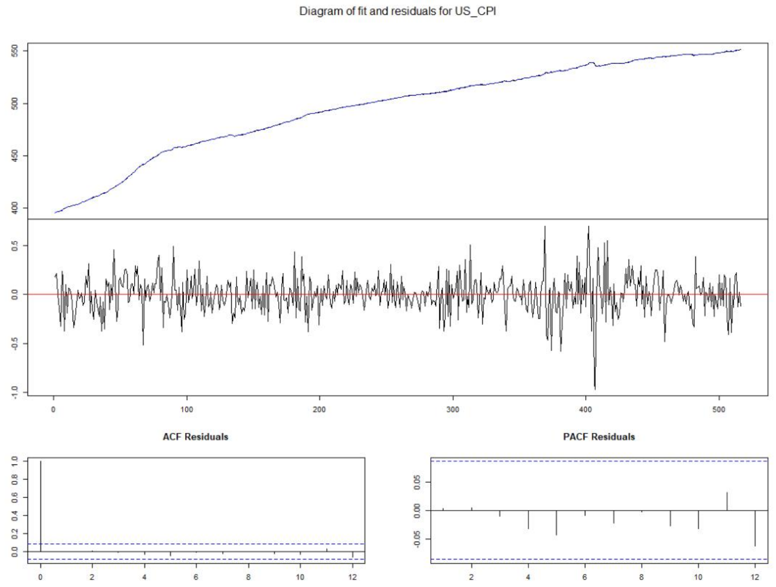


Diagram of fit and residuals for WORLD_IP





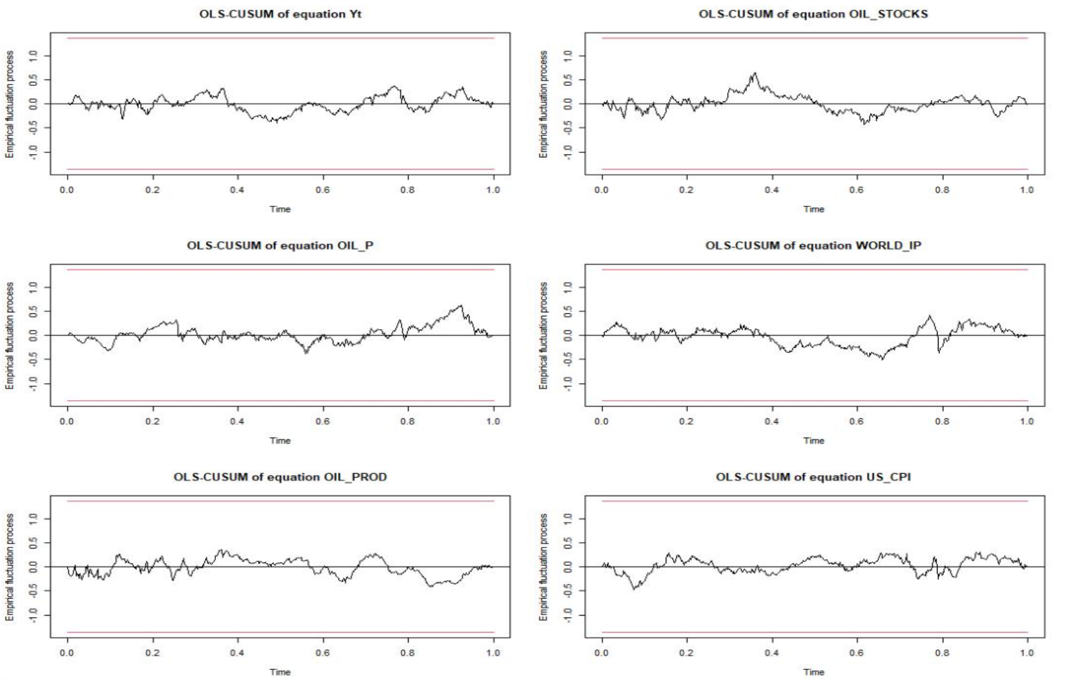

```
> serialcorrelation(VARmodel = Reduced_VAR12,nlag=25)
```

lag		Portm. stat	Portm. p-value	adj Portm. stat	adj Portm. p-value	BG-LM stat	BG LM-p-value
1	1	1.3708	NaN	1.3735	NaN	63.8122	0.0029
2	2	4.8664	NaN	4.8827	NaN	141.2874	0.0000
3	3	8.6478	NaN	8.6862	NaN	185.8766	0.0000
4	4	13.4323	NaN	13.5081	NaN	245.8107	0.0000
5	5	19.3868	NaN	19.5208	NaN	293.0318	0.0000
6	6	26.6193	NaN	26.8384	NaN	323.9861	0.0000
7	7	30.0562	NaN	30.3225	NaN	383.5362	0.0000
8	8	35.3055	NaN	35.6545	NaN	418.3753	0.0000
9	9	47.1452	NaN	47.7044	NaN	460.3834	0.0000
10	10	54.9013	NaN	55.6137	NaN	519.9594	0.0000
11	11	69.9511	NaN	70.9914	NaN	565.2669	0.0000
12	12	135.8679	NaN	138.4777	NaN	611.1380	0.0000
13	13	171.0530	0	174.5721	0	714.5417	0.0000
14	14	218.7288	0	223.5775	0	775.8595	0.0000
15	15	260.0793	0	266.1661	0	816.6645	0.0000
16	16	287.9043	0	294.8814	0	872.5984	0.0000
17	17	308.6594	0	316.3437	0	900.6508	0.0000
18	18	357.2990	0	366.7413	0	953.1479	0.0000
19	19	387.7885	0	398.3964	0	1003.9242	0.0000
20	20	413.6835	0	425.3355	0	1055.5668	0.0000
21	21	440.3220	0	453.1042	0	1099.0246	0.0000
22	22	477.7466	0	492.1954	0	1152.1972	0.0000
23	23	522.9076	0	539.4633	0	1192.3130	0.0000
24	24	597.8207	0	618.0307	0	1232.6144	0.0000
25	25	637.3679	0	659.5916	0	1268.2195	0.0000

Lag	BG-LM stat	BG_LM p-value
1	63.8122	0.0029
2	141.2874	0.0000
3	185.8766	0.0000
4	245.8107	0.0000
5	293.0318	0.0000
6	323.9861	0.0000
7	383.5362	0.0000
8	418.3753	0.0000
9	460.3834	0.0000
10	519.9594	0.0000
11	565.2669	0.0000
12	611.1380	0.0000
13	714.5417	0.0000
14	775.8595	0.0000
15	816.6645	0.0000
16	872.5984	0.0000
17	900.6508	0.0000
18	953.1479	0.0000
19	1003.9242	0.0000

```
20  1055.5668         0.0000
21  1099.0246         0.0000
22  1152.1972         0.0000
23  1192.3130         0.0000
24  1232.6144         0.0000
25  1268.2195         0.0000
```

```
> var12.stab=stability(Reduced_VAR12,type = "OLS-CUSUM")
> plot(var12.stab, alpha=0.05)
```



What is the advantage of a VAR model over an ADL model ?

By the VAR model, it didn't ignore the endogenous problem. ADL model can only allow the unidirectional causal relationship, and now VAR can identify the multi directional causal relat ionship.

Do you agree with the choices made by Känzig ?

On average, if stick to a VAR model, I agree with Känzig.

(1) Why use a VAR in levels ?

By using VAR in levels , **the estimation will be biased but consistent** because the real data generating process will be nested in our model. **the only price has to pay is no can do standard inference before test to know there is cointegration.**

If choose to estimate this in first differences, **the estimation will be inconsistent because of the misspecification if there is cointegration.**

And the cointegration relationships is not easy to test to be sure, **there may be type 1 error and type 2 error.** so choose VAR in levels would be the most "safe" way.

(2) What is the intuition for setting the lag length to 12

Lag length is a month and 12 lag length is a year and there may be macroeconomic cycle in a year.

(3) Is the lag length justified when considering information criteria and residual diagnostics ?

No, by the residual diagnostics, there is no autocorrelation in the error term of each equation if just looking at the ACF and PACF. **but by the result of the joint test of the autocorrelation in the error term, it strongly reject the Hh0 of there is no autocorrelation in the error term.** so there is still pattern in the error term even with such high orders.

And the information criteria choose VAR(2) and VAR(4), relatively much less dynamics model.

so to be concluded, there may be patterns that can not be captured by simply adding autoregressive terms like moving average parts or structural break. so it may be better to change the specification to a more flexible way like local projection.

2.

```
  Granger causality H0: OIL_P OIL_PROD OIL_STOCKS WORLD_IP US_CPI do not Granger-cause Yt

data:  VAR object Reduced_VAR12
F-Test = 2.1707, df1 = 60, df2 = 2658, p-value = 6.706e-07


  Granger causality H0: Yt OIL_PROD OIL_STOCKS WORLD_IP US_CPI do not Granger-cause OIL_P

data:  VAR object Reduced_VAR12
F-Test = 1.3992, df1 = 60, df2 = 2658, p-value = 0.02392


  Granger causality H0: Yt OIL_P OIL_STOCKS WORLD_IP US_CPI do not Granger-cause OIL_PROD

data:  VAR object Reduced_VAR12
F-Test = 2.1014, df1 = 60, df2 = 2658, p-value = 2.033e-06
```

```
Granger causality H0: Yt OIL_P OIL_PROD WORLD_IP US_CPI do not Granger-cause OIL_STOCKS

data:  VAR object Reduced_VAR12
F-Test = 2.5774, df1 = 60, df2 = 2658, p-value = 6.336e-10


Granger causality H0: Yt OIL_P OIL_PROD OIL_STOCKS US_CPI do not Granger-cause WORLD_IP

data:  VAR object Reduced_VAR12
F-Test = 2.1508, df1 = 60, df2 = 2658, p-value = 9.237e-07


Granger causality H0: Yt OIL_P OIL_PROD OIL_STOCKS WORLD_IP do not Granger-cause US_CPI

data:  VAR object Reduced_VAR12
F-Test = 3.2726, df1 = 60, df2 = 2658, p-value = 1.221e-15


Granger causality H0: Yt do not Granger-cause OIL_P OIL_PROD OIL_STOCKS WORLD_IP US_CPI

data:  VAR object Reduced_VAR12
F-Test = 1.492, df1 = 60, df2 = 2658, p-value = 0.008885


Granger causality H0: OIL_P do not Granger-cause Yt OIL_PROD OIL_STOCKS WORLD_IP US_CPI

data:  VAR object Reduced_VAR12
F-Test = 2.3002, df1 = 60, df2 = 2658, p-value = 7.891e-08


Granger causality H0: OIL_PROD do not Granger-cause Yt OIL_P OIL_STOCKS WORLD_IP US_CPI

data:  VAR object Reduced_VAR12
F-Test = 1.8261, df1 = 60, df2 = 2658, p-value = 0.0001266


Granger causality H0: OIL_STOCKS do not Granger-cause Yt OIL_P OIL_PROD WORLD_IP US_CPI

data:  VAR object Reduced_VAR12
F-Test = 1.9387, df1 = 60, df2 = 2658, p-value = 2.475e-05


Granger causality H0: WORLD_IP do not Granger-cause Yt OIL_P OIL_PROD OIL_STOCKS US_CPI

data:  VAR object Reduced_VAR12
F-Test = 2.2614, df1 = 60, df2 = 2658, p-value = 1.511e-07


Granger causality H0: US_CPI do not Granger-cause Yt OIL_P OIL_PROD OIL_STOCKS WORLD_IP

data:  VAR object Reduced_VAR12
F-Test = 2.1163, df1 = 60, df2 = 2658, p-value = 1.606e-06
```

Do oil prices Granger cause the US macro variables (and vice versa) ?

From the Granger univariate test, it reject the H0 of oil prices do not Granger cause all the oth er variables. so oil prices Granger cause the US macro variables.

From the Granger joint test, it also reject the H0 of all the other US macro variables Granger c ause the oil prices. So other macro variables also Granger cause the oil price.

Can the results give you guidance to decide on the ord ering of the variables in your Cholesky decomposition (in q3 below) ?

No, because Granger cause analysis is only based on the reduced form VAR, and only reflect whether the lag of one or several variable(s) affect other current or future variable. It can not r eflect the contemporaneous effect. but use cholesky decomposition to identify the strutural V AR is imposing recursive restrictions on contemporaneous effect which need to know the “or der of the exogenity” of these variables.

3.

```
> residCorr = VARsum$corres
> stargazer(residCorr, type = "text")

============================================================
             Yt   OIL_P  OIL_PROD OIL_STOCKS WORLD_IP US_CPI
------------------------------------------------------------
Yt            1   0.012  0.111     0.036      0.480   -0.023
OIL_P      0.012    1    -0.058    -0.093     0.109   0.381
OIL_PROD   0.111 -0.058    1       -0.022     0.003   0.007
OIL_STOCKS 0.036 -0.093 -0.022       1        0.014   -0.022
WORLD_IP   0.480 0.109   0.003     0.014        1     0.057
US_CPI    -0.023 0.381   0.007     -0.022     0.057     1
------------------------------------------------------------
```

```
> causality(Reduced_VAR12, cause = c("OIL_P", "OIL_PROD","OIL_
STOCKS", "WORLD_IP","US_CPI"))$Instant
  H0: No instantaneous causality between: OIL_P OIL_PROD OIL_STOCKS WORLD_IP US_CPI and Yt

data:  VAR object Reduced_VAR12
Chi-squared = 101.98, df = 5, p-value < 2.2e-16
```

```
> causality(Reduced_VAR12, cause = c("Yt","OIL_PROD","OIL_STOC
KS", "WORLD_IP","US_CPI"))$Instant

  H0: No instantaneous causality between: Yt OIL_PROD OIL_STOCKS WORLD_IP US_CPI and OIL_P

data:  VAR object Reduced_VAR12
Chi-squared = 72.781, df = 5, p-value = 2.698e-14
```

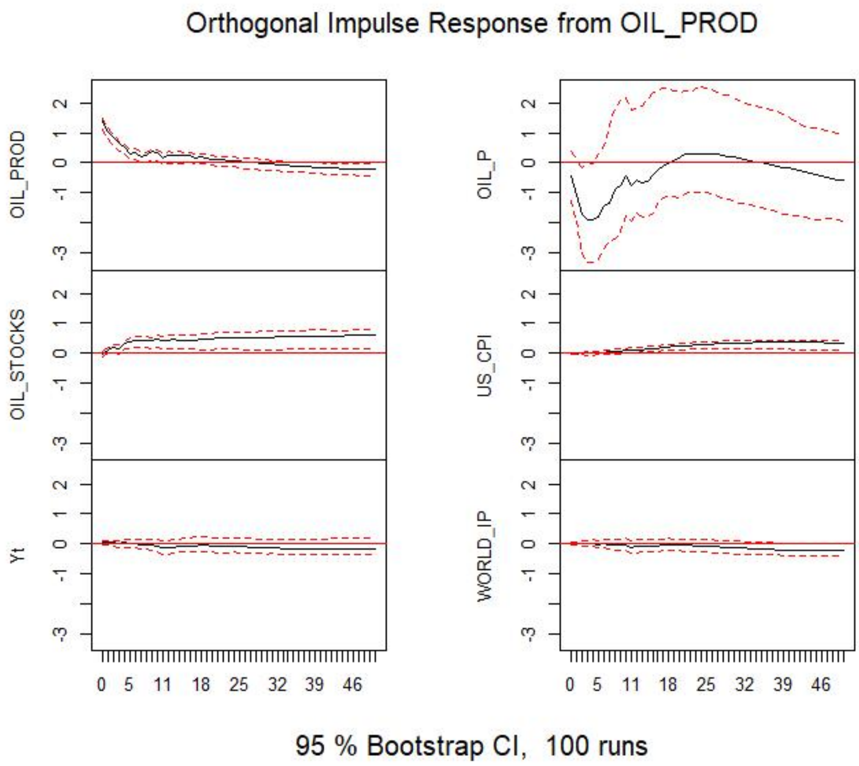
```
> causality(Reduced_VAR12, cause = c("Yt","OIL_P","OIL_STOCKS",
 "WORLD_IP","US_CPI"))$Instant

  H0: No instantaneous causality between: Yt OIL_P OIL_STOCKS WORLD_IP US_CPI and OIL_PROD

data:  VAR object Reduced_VAR12
Chi-squared = 10.522, df = 5, p-value = 0.06173
```

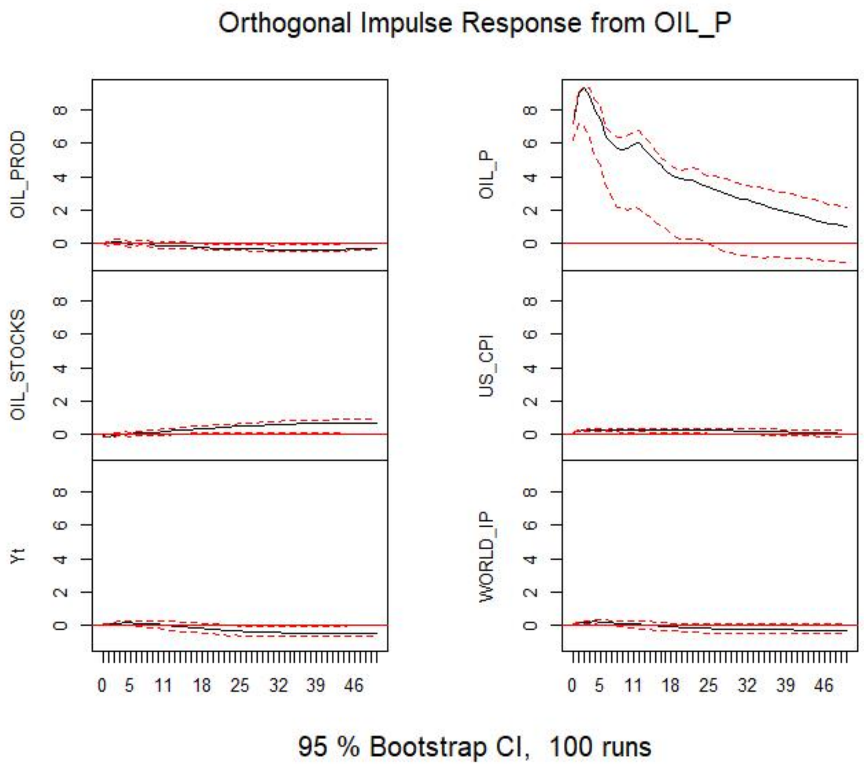
```
> causality(Reduced_VAR12, cause = c("Yt","OIL_P","OIL_PROD",
"WORLD_IP","US_CPI"))$Instant

  H0: No instantaneous causality between: Yt OIL_P OIL_PROD WORLD_IP US_CPI and OIL_STOCKS

data:  VAR object Reduced_VAR12
Chi-squared = 5.7406, df = 5, p-value = 0.3323
```

```
> causality(Reduced_VAR12, cause = c("Yt","OIL_P","OIL_PROD",
"OIL_STOCKS","US_CPI"))$Instant

  H0: No instantaneous causality between: Yt OIL_P OIL_PROD OIL_STOCKS US_CPI and WORLD_IP

data:  VAR object Reduced_VAR12
Chi-squared = 101.33, df = 5, p-value < 2.2e-16
```

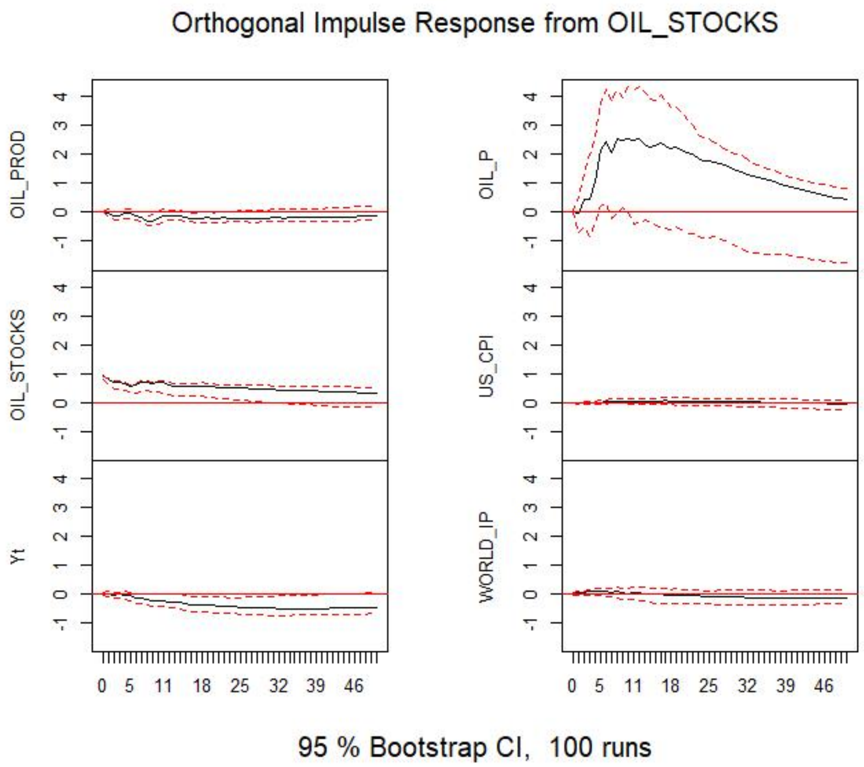
```
> causality(Reduced_VAR12, cause = c("Yt","OIL_P","OIL_PROD",
"OIL_STOCKS","WORLD_IP"))$Instant

  H0: No instantaneous causality between: Yt OIL_P OIL_PROD OIL_STOCKS WORLD_IP and US_CPI

data:  VAR object Reduced_VAR12
Chi-squared = 66.594, df = 5, p-value = 5.232e-13
```

```
VARord = cbind(OIL_PROD, OIL_P, OIL_STOCKS, US_CPI, Yt, WORLD_
IP)

SVAR = VAR(VARord, p = 12, type = "const")

IRF_SVAR = irf(SVAR, n.ahead = 50, boot='TRUE',runs=100)

plot(IRF_SVAR)
```

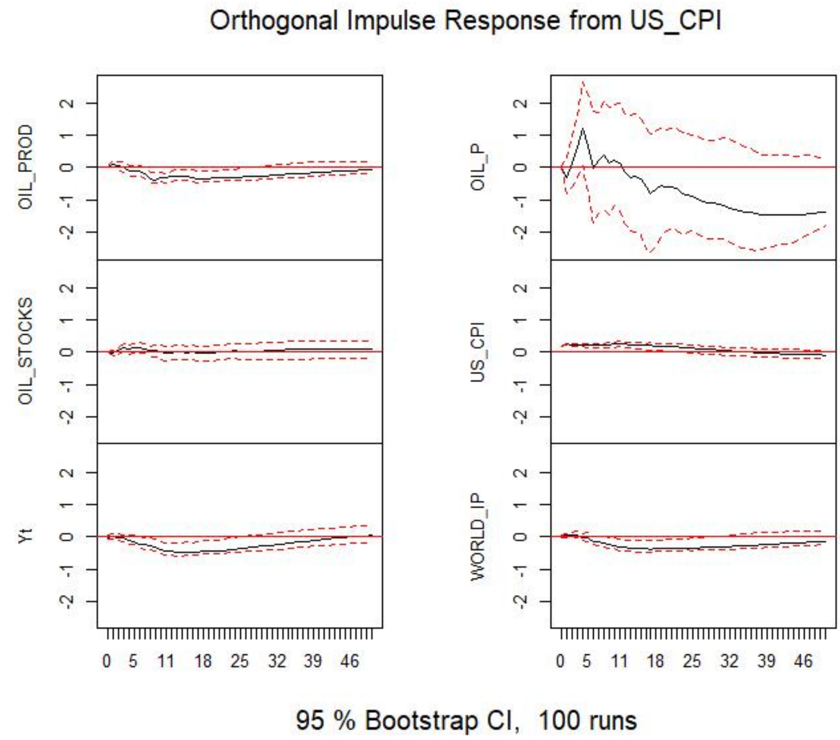




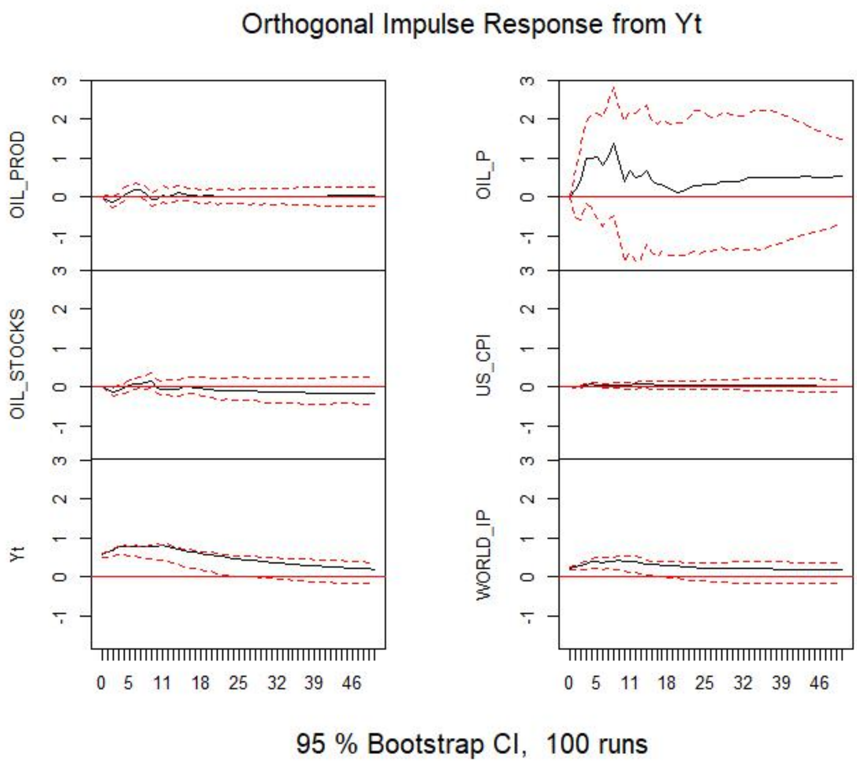
Orthogonal Impulse Response from OIL_P
OIL_PROD
OIL_STOCKS
Yt
OIL_P
US_CPI
WORLD_IP
0 5 11 18 25 32 39 46
95 % Bootstrap CI, 100 runs



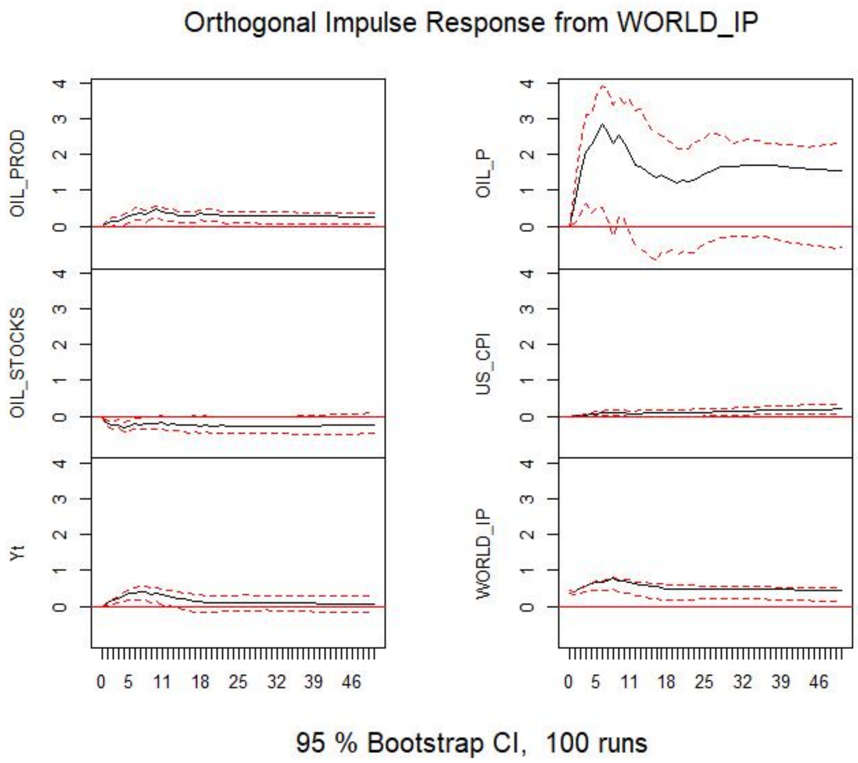
Orthogonal Impulse Response from OIL_STOCKS
OIL_PROD
OIL_STOCKS
Yt
OIL_P
US_CPI
WORLD_IP
0 5 11 18 25 32 39 46
95 % Bootstrap CI, 100 runs



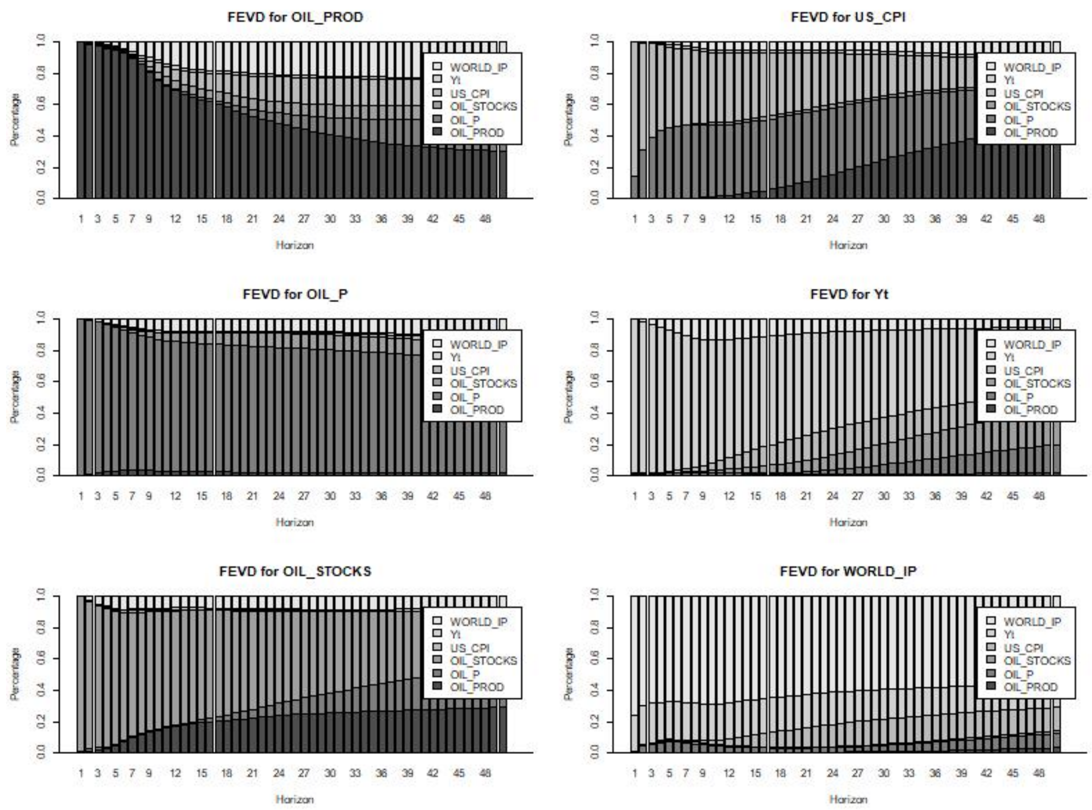
Orthogonal Impulse Response from US_CPI
OIL_PROD
OIL_P
OIL_STOCKS
US_CPI
Yt
WORLD_IP
0 5 11 18 25 32 39 46
95 % Bootstrap CI, 100 runs



Orthogonal Impulse Response from Yt
OIL_PROD
OIL_P
OIL_STOCKS
US_CPI
Yt
WORLD_IP
0 5 11 18 25 32 39 46
95 % Bootstrap CI, 100 runs




```
VD_SVAR<-fevd(SVAR, n.ahead = 50)
plot(VD_SVAR)
```



Goal: We want to identify the impact of an oil supply shock. How are you going to order the variables to ac hieve this goal ? Clear motivate your choice !

The order came from the economic logic and intuitions, the “most exogenous” variable would be the oil production, because the production mainly determined by the previous conditions a nd factors, it is not really likely to be determined contemporaneously within a month.

Also the same reason for the oil price as it would be adjusted contemporaneously because of t he production, but to other variables, it is more likely to be determined by the past shocks. so i t could be the second exogenous variable.

The oil stock is placed in the third order, it is for sure will be impact contemporaneously by th e shock of the production and price, but there would be the lag effect in oil stock with the mac ro demand factors of CPI of the United states of America, industrial production of the United States of America, and the industrial production of the world.

The forth and fifth exogenous variables would be the CPI of the United States of America, an d industrial production of the United States of America. this two variables are the indicators o f the America’s economy.

And the last exogenous variable would be the world industrial production. It would be the mo st endogenous variable as it is the most complex system.

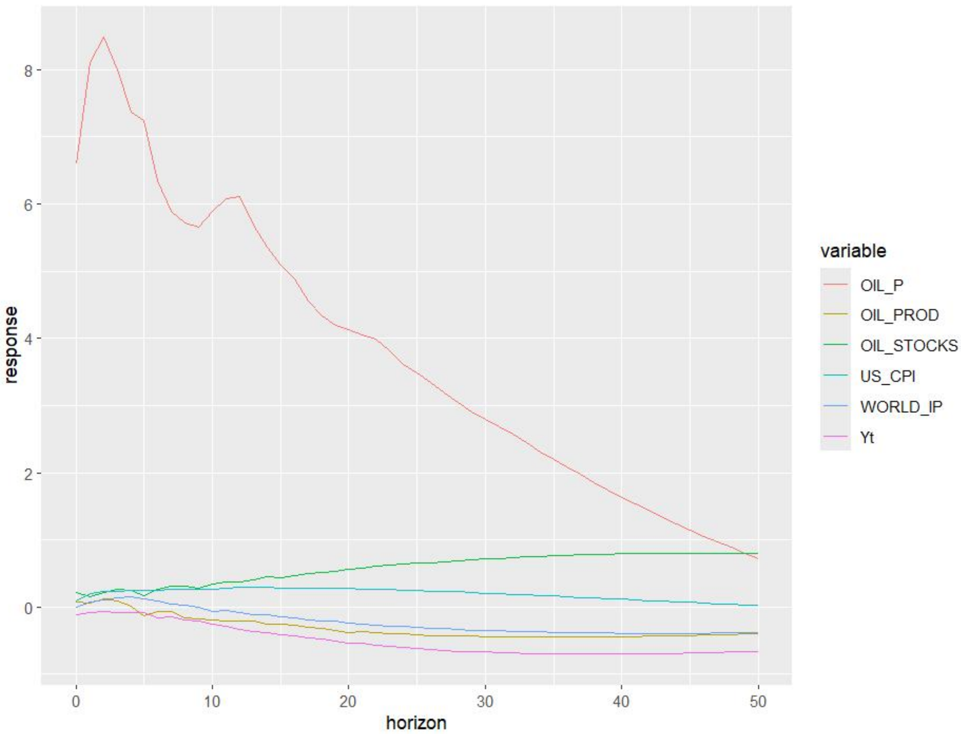
Then I create the correlation matrix from the residuals of the reduced form VAR and run a “in stantaneous causality analysis. **in principle, these can only reflect the correlation relations hips with the residuals and can not reflect the contemporaneous causality. But it can hel p me to see if it is in line with the economic logic and intuition. And it is in line with what is imposed**

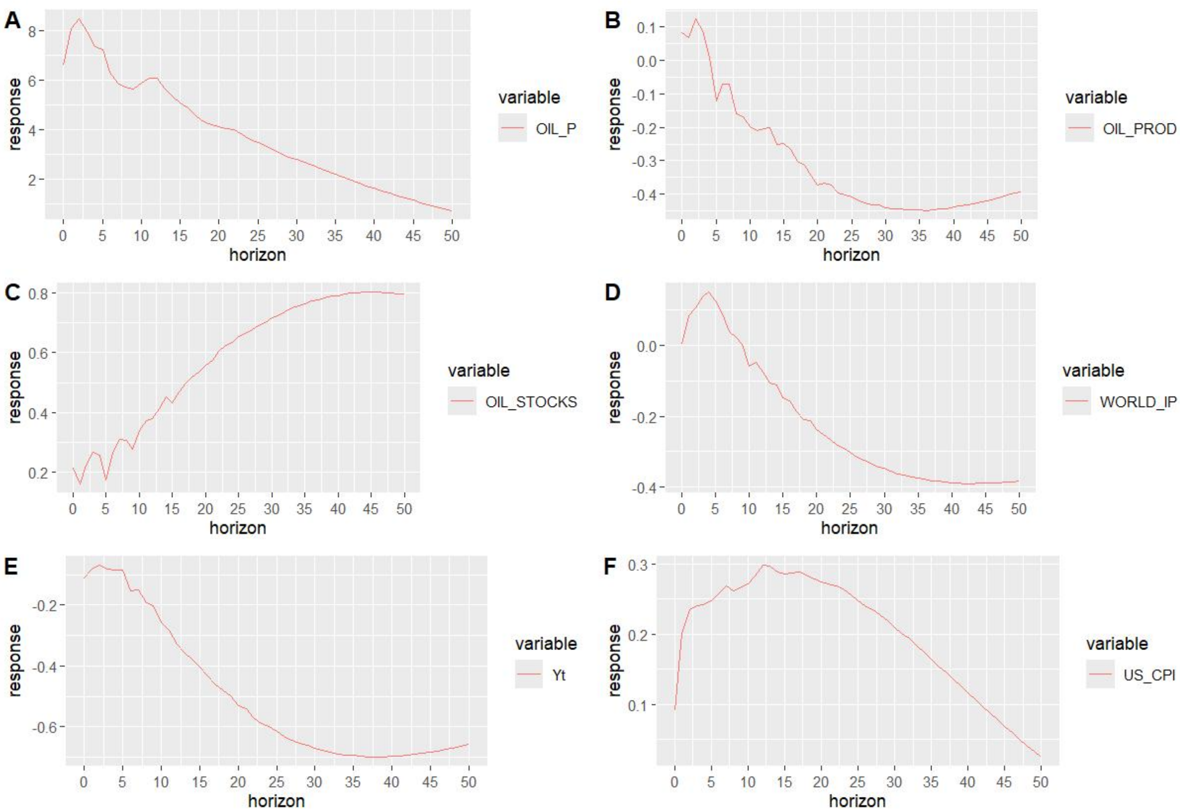
Are the results in line with figure 3 of Känzig (2021) ? If you see differences, what can be the (intuitive) reas on(s) ?

No, it is not line with Känzig, the identifying strategies are different so can get different stru ctural VAR. Känzig use the high frequency data of the oil supply news shock as an external in strument to identify the structural VAR. But we impose recursive restrictions on the contemp oraneous effect to get the structural VAR. So it is clearly different imposing and different intu ition behind. and the results is different with the different identifying strategies

4.

```
ggplot(irfs, aes(x=horizon, y=response, group= variable, color
= variable))+geom_line()
```


variable
OIL_P
OIL_PROD
OIL_STOCKS
US_CPI
WORLD_IP
Yt
response
horizon


A
B
C
D
E
F
variable
OIL_P
OIL_PROD
OIL_STOCKS
WORLD_IP
Yt
US_CPI
response
horizon

Interpret the response of each of the variables to an oi l surprise (i.e. make sure you can read the IRs correc tly)

To a oil supply news shock, the oil price first increase then it goes down continually to 0.

the oil production firstly increase a little bit, then goes down continually to arrive -0.5%.

The oil stock could increase from 0.3% to 1.2% gradually.

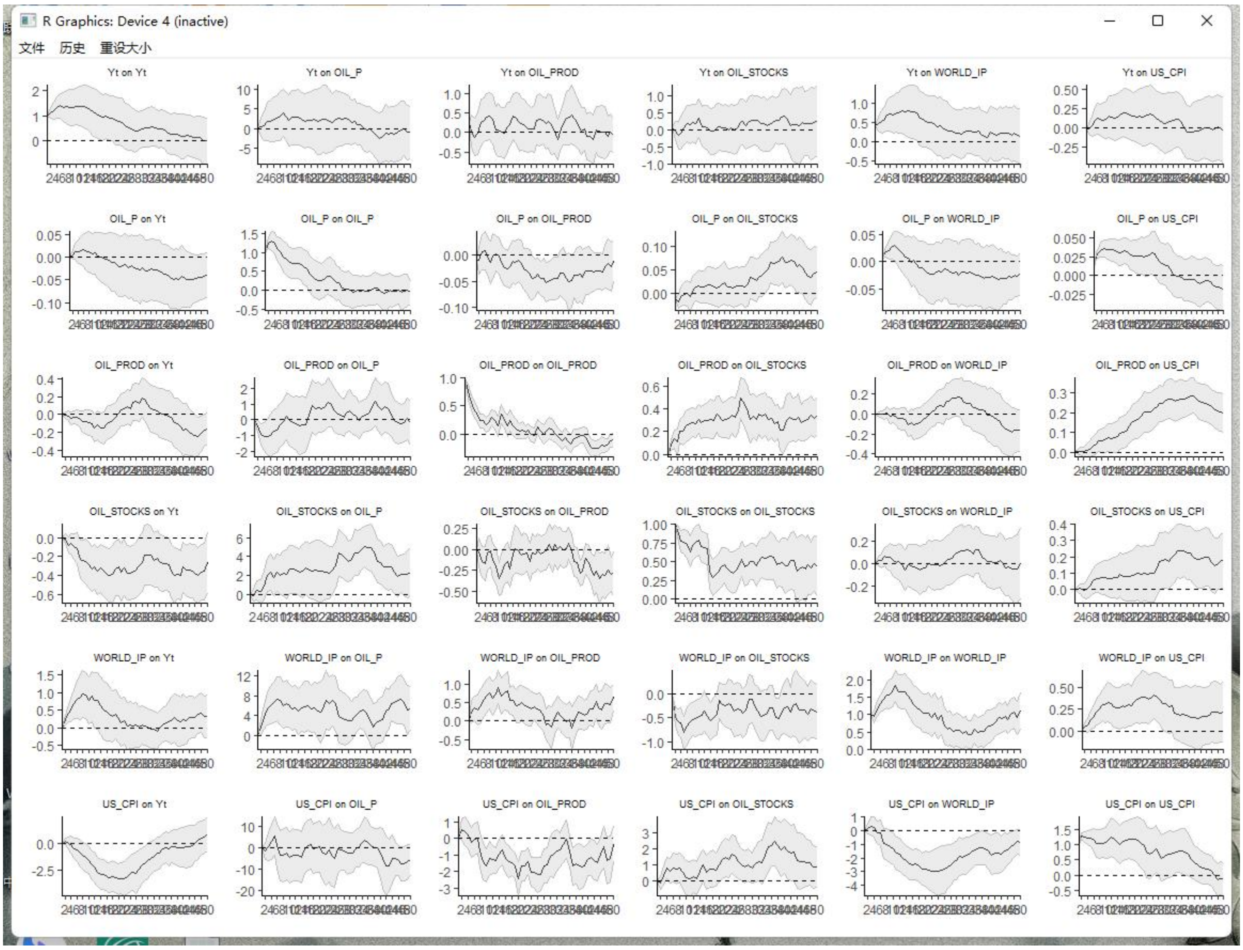
The world industrial production could increase a little and then decreases from the period 10 t o arrive the level of -0.5%.

The industrial production of the United States of America could decrease slowly from the peri od 1 to 10, the speed more faster after period 10. Finally, the industrial production of the Unit ed States of America will fall 1%.

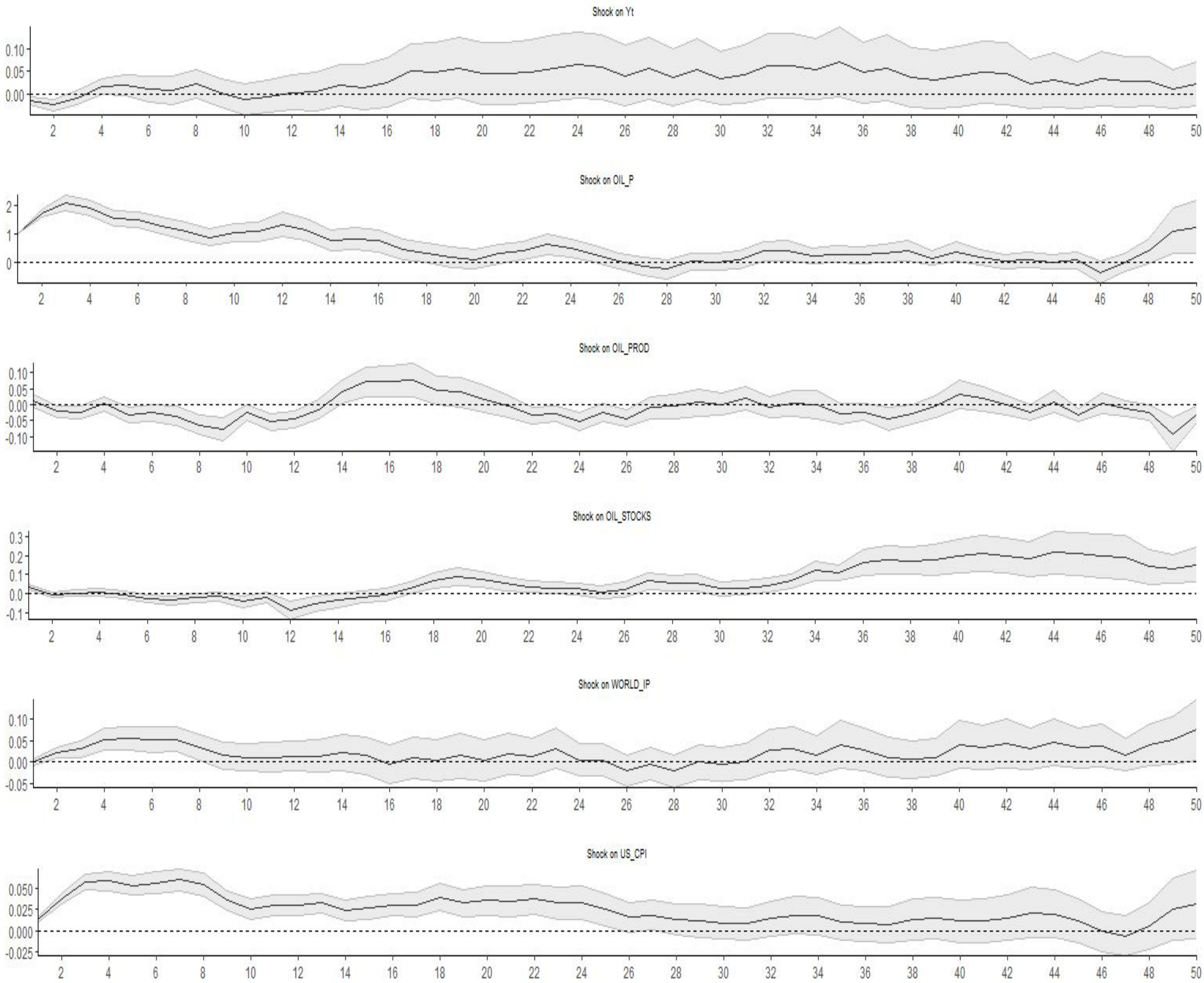
The CPI of the United States of America will increase 0.4% from the period 2 to 20 and then decreases to the original level after 5.

5.

```
> colnames(y) <- c("Yt","OIL_P", "OIL_PROD", "OIL_STOCKS", "WO
RLD_IP", "US_CPI")

> y_lp <-as.data.frame(y)

> LP = lp_lin(y_lp, lags_endog_lin = 12, shock_type = 1, trend
 = 0, confint = 1.96, hor = 50)> plot(LP)
```




```
> shock <- as.data.frame(new_data$OIL_P)

> instrum <-as.data.frame(new_data$Surprises)

> LP_iv = lp_lin_iv(y_lp, lags_endog_lin = 12, shock = shock,i
nstrum = instrum, use_twosls = TRUE, trend = 0, confint = 1.96,
 hor= 50)
```



Compare the results to those obtained from the IV VAR, If you see differences, what can be the (intuitive) reason(s) ?

VAR is more smooth, it's using several parameters to measure the entire horizions. Local Projection is estimating every point on the impulse response function directly and individually, so it has more structure.

6.

The whole analysis was done using log levels of the variables.

(1) Is it justified to do the analysis in first difference, what is a potential advantage of the first difference approach?

For the local projection, yes, by using local projection in first difference, all the variables are stationary, and the local projection is robust to misspecification, **so even there is cointegration, the model is still consistent. and asymptotically normal distributed, so we can do standard inference.**

By using local projection in level, the estimation will deteriorate as the h increases, even there is cointegration. It may bring the "spurious regression like" error accumulating in the error term as the h increases if you do not add addition lags.

(2) Is the argument different for a VAR versus local projection ?

Yes it is different, for the VAR, stick to the VAR in levels is better. because by using the VAR in levels, it can be remaramepterization to the VAR in first difference and vector error correction term. So the real DGP is nested in our model no matter whether there is cointegration. So the estimation is biased but consistent no matter whether there is cointegration. all the price need to pay is no can do the standard inferences now because do not know if there is error correction term or not.

But by using VAR in first difference, if it is sure data is non stationary and no cointegration, VAR in first difference would be consistent and asymptotically normal distributed, but in practice, this is very hard goal to achieve, because multi cointegrating relationships is hard to test and there also very possible to make type I error or type II error. And the model will be misspecified if using VAR in first difference and there is cointegration. and the model would be inconsistent and not asymptotically normal distributed.

(3) How do impulse response function computed using first difference to those using level.

The impulse response function of level is the accumulated impulse response function of first difference. accumulating each point in the impulse response function in first difference can obtain the impulse response function in level.