

# Econometrics II - Problem Set 3

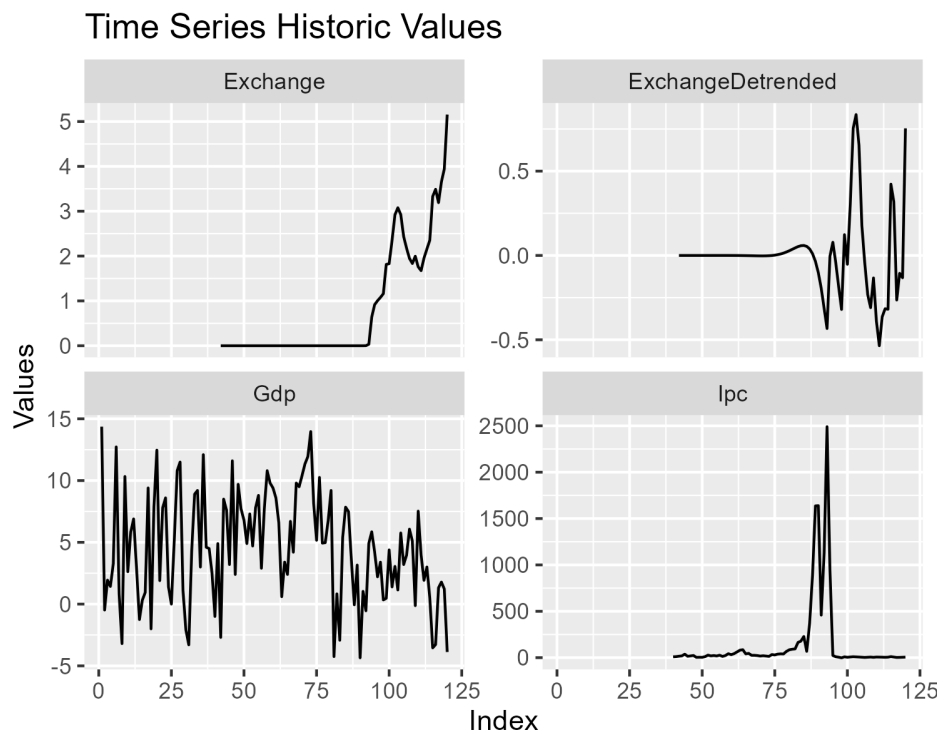
Ricardo Semião e Castro

05/2024

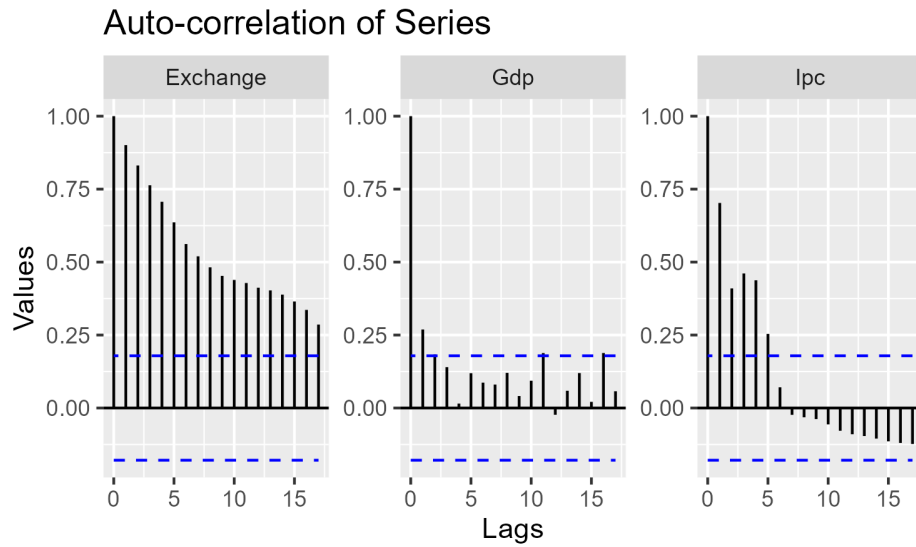
## Exploratory Data Analysis

This section is dedicated to the exploratory analysis made for questions 1. and 4.

We can see the time series values below. Exchange had a clear trend, so it was detrended using a HPfilter with  $\lambda = 100$  for yearly data.



It is also interesting to consider the plots of the ACF of the series.



We can see that Exchange presents a trend, and strong memory. With this motivation, I ran the ADF test for the series. The results can be seen below.

Table 1: ADF Test - Exchange

Lag	Type 1	Type 2	Type 3
0	3.8 (0.99)	2.9 (0.99)	0.82 (0.99)
1	2.53 (0.99)	1.93 (0.99)	0.21 (0.99)
2	2.28 (0.99)	1.73 (0.99)	0.07 (0.99)

Table 2: ADF Test - GDP

Lag	Type 1	Type 2	Type 3
0	-5.25 (0.01)	-8.18 (0.01)	-8.32 (0.01)
1	-3.21 (0.01)	-5.46 (0.01)	-5.67 (0.01)
2	-2.46 (0.02)	-4.41 (0.01)	-4.65 (0.01)

Table 3: ADF Test - IPC

Lag	Type 1	Type 2	Type 3
0	-3.45 (0.01)	-3.68 (0.01)	-3.67 (0.03)
1	-3.67 (0.01)	-3.97 (0.01)	-3.96 (0.02)
2	-1.91 (0.06)	-2.07 (0.3)	-2.03 (0.56)

Table 1 indicates that the series Exchange probably is non-stationary, while the other two are stationary. So, it was probably more suited to do a differencing, instead of a detrending. For simplicity, the variable was included with trend in the models.

## Question 1

The results can be seen below. The row 'predictions' and 'MSE' are related to the prediction for 2020.

Table 4:

	<i>Dependent variable:</i>			
	GDP Growth			
	(1)	(2)	(3)	(4)
lag(Gdp, 1)	0.323*** (0.114)	0.368*** (0.111)	0.282** (0.121)	0.373*** (0.109)
lag(Gdp, 2)	0.230** (0.110)	0.264** (0.110)	0.195* (0.117)	0.273** (0.107)
lag(Exchange, 1)	−0.591 (0.401)		−1.497 (2.051)	
lag(Exchange, 2)			0.751 (2.122)	
lag(Ipc, 1)		0.0003 (0.001)	−0.0004 (0.001)	
lag(Ipc, 2)		−0.001 (0.001)	−0.001 (0.001)	
Constant	2.490*** (0.883)	1.777** (0.746)	3.200*** (1.085)	1.625** (0.665)
Predictions	0.96	2.7	0.72	2.57
MSE	23.43	43.24	21.19	41.55
Observations	76	76	76	76
R <sup>2</sup>	0.333	0.318	0.349	0.313
Adjusted R <sup>2</sup>	0.305	0.279	0.292	0.294
Residual Std. Error	3.382 (df = 72)	3.444 (df = 71)	3.414 (df = 69)	3.409 (df = 73)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Via the MSE, we can see that the model generates the best prediction is from model 3. This makes sense, given that it is the most complex model, while seeming to be not too complex, not suffering from overfitting.

## Question 2

Let  $\phi_j$  denote the vector of coefficients of each lag  $j$  in the  $VAR(p)$  model, and  $\epsilon_t$  the error term. Then, we can write the model and take the expectancy of both sides:

$$E[Y_t] = c + \phi_1 E[Y_{t-1}] + \cdots + \phi_p E[Y_{t-p}] + E[\epsilon_t]$$

As  $\{Y_t\}$  is covariance-stationary,  $E[Y_t] = E[Y_{t-1}] = \dots = E[Y_{t-p}]$  Denote this value by  $\mu$ . Also  $E[\epsilon_t] = 0$ . So, we have:

$$\begin{aligned}\mu &= c + \phi_1\mu + \dots + \phi_p\mu \\ (I - \phi_1 - \dots - \phi_p)\mu &= c \\ \mu &= (I - \phi_1 - \dots - \phi_p)^{-1}c\end{aligned}$$

### Question 3

First, note that  $\Gamma_{-j} = E[(Y_t - \mu)(Y_{t+j} - \mu)']$ .

Then:

$$\begin{aligned}\Gamma'_j &= E[(Y_{t-j} - \mu)(Y_t - \mu)'] \\ &= E[(Y_{t-j+j} - \mu)(Y_{t+j} - \mu)'] \\ &= \Gamma_{-j}\end{aligned}$$

The second line step came from  $Y_t$  being covariance-stationary, "the window doesn't matter", i.e.:

$$E[(Y_{t-j+k} - \mu)(Y_{t+k} - \mu)'] = E[(Y_{t-j} - \mu)(Y_t - \mu)'], \quad \forall k \in \mathbb{N}$$

### Question 6

#### Item 1.

The results, for each model, can be seen below. I used the detrended version of Exchange.

Table 5: VAR(1)

	<i>Dependent variable:</i>		
	Gdp	Ipc	Exchange
Gdp.l1	0.292** (0.111)	-17.725* (9.499)	-0.001 (0.007)
Ipc.l1	-0.002* (0.001)	0.632*** (0.087)	0.0001* (0.0001)
Exchange.l1	-1.164*** (0.411)	-61.510* (35.044)	1.041*** (0.024)
const	4.478*** (0.874)	182.530** (74.463)	0.011 (0.052)
Predictions	0.23	-78.95	4.12
MSE	16.92	7152.84	1.08
Observations	77	77	77
R <sup>2</sup>	0.301	0.522	0.971
Adjusted R <sup>2</sup>	0.273	0.503	0.970
Residual Std. Error (df = 73)	3.457	294.664	0.204

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 6: VAR(2)

	<i>Dependent variable:</i>		
	Gdp	Ipc	Exchange
Gdp.l1	0.282** (0.121)	-18.671* (10.306)	0.001 (0.007)
Ipc.l1	-0.0004 (0.001)	0.707*** (0.120)	0.0001* (0.0001)
Exchange.l1	-1.497 (2.051)	-194.682 (175.155)	1.334*** (0.121)
Gdp.l2	0.195* (0.117)	-13.218 (9.982)	-0.0002 (0.007)
Ipc.l2	-0.001 (0.001)	-0.173 (0.117)	-0.00005 (0.0001)
Exchange.l2	0.751 (2.122)	112.603 (181.280)	-0.309** (0.125)
const	3.200*** (1.085)	288.045*** (92.642)	0.0001 (0.064)
Predictions	0.72	-112.25	4.13
MSE	21.19	13893.94	1.04
Observations	76	76	76
R <sup>2</sup>	0.349	0.557	0.973
Adjusted R <sup>2</sup>	0.292	0.519	0.971
Residual Std. Error (df = 69)	3.414	291.590	0.201

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 7: VAR(3)

	<i>Dependent variable:</i>		
	Gdp	Ipc	Exchange
Gdp.l1	0.269** (0.128)	-12.038 (9.569)	0.003 (0.007)
Ipc.l1	-0.001 (0.001)	0.800*** (0.111)	0.0001 (0.0001)
Exchange.l1	-1.557 (2.181)	-97.002 (162.545)	1.342*** (0.127)
Gdp.l2	0.235* (0.131)	-15.233 (9.755)	0.001 (0.008)
Ipc.l2	-0.001 (0.002)	-0.568*** (0.136)	-0.00004 (0.0001)
Exchange.l2	0.941 (3.544)	-199.978 (264.062)	-0.341 (0.206)
Gdp.l3	-0.043 (0.123)	1.963 (9.176)	-0.012 (0.007)
Ipc.l3	-0.0004 (0.001)	0.495*** (0.108)	-0.00003 (0.0001)
Exchange.l3	-0.141 (2.294)	239.065 (170.953)	0.008 (0.133)
const	3.269** (1.292)	217.662** (96.271)	0.058 (0.075)
Predictions	0.8	-168.99	4.12
MSE	21.88	30489.6	1.07
Observations	75	75	75
R <sup>2</sup>	0.352	0.667	0.974
Adjusted R <sup>2</sup>	0.262	0.621	0.970
Residual Std. Error (df = 65)	3.496	260.490	0.203

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Comparing the predictions,  $VAR(1)$  yielded 0.23 (MSE of 16.92),  $VAR(2)$  yielded 0.72 (MSE of 21.19), and  $VAR(3)$  yielded 0.8 (MSE of 21.88).  $VAR(1)$  was the best.

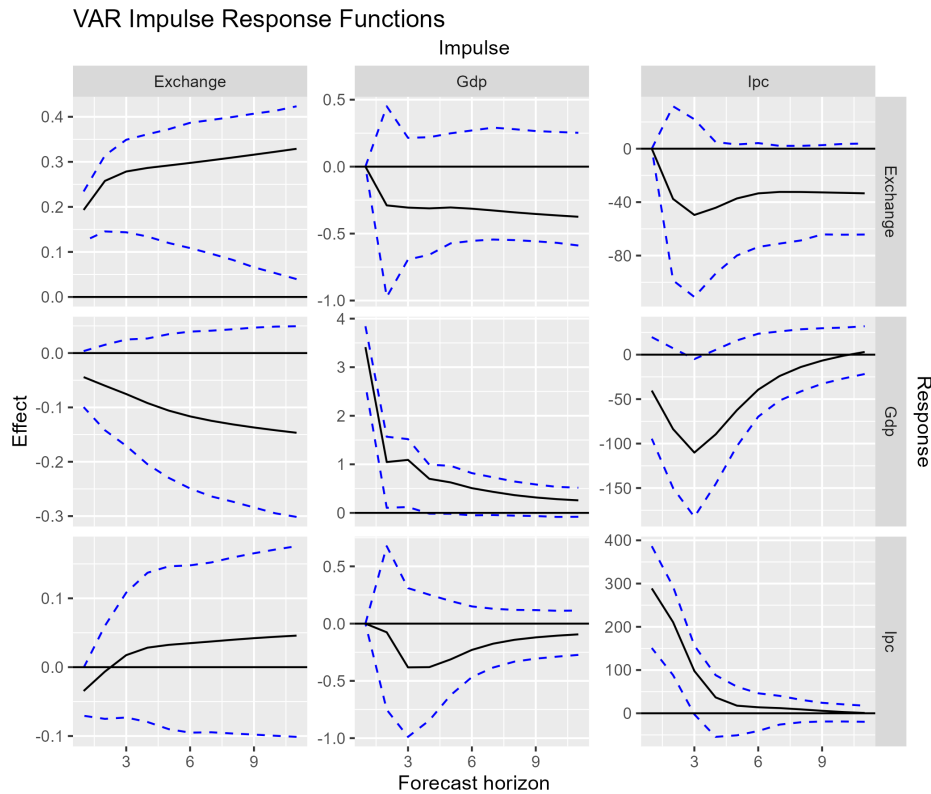
For yearly data, we don't normally need to use too many lags to capture the dynamic of the problem, specially when using multiple series. This might be why  $VAR(1)$  had the better prevision.

## Item 2.

The chosen order was (Ipc, Gdp, Exchange). The choice lies in the analysis of the level of rigidity of each variable. We have lots of empirical evidence to consider price rigidities, such that GDP should respond to IPC. The inflation could be set by the BC given the GDP, or react to the state of the economy, but this shouldn't be too intense, and one could argue relatively

exogenous shocks on inflation. Lastly, the exchange rate should be the most endogenous variable, as the value of the currency is set by expectations, which relate to both GDP and IPC, and propagate quickly, given the high liquidity of the exchange market.

The IRFs with the Cholesky decomposition in the order above can be seen below.



The results are not really credible. We have no statistically significant response of GDP given a shock of inflation, which contradicts our assumption of price rigidities. The exchange rate also doesn't present significant responses to IPC and GDP.

These poor results are not surprising, given the low quality of the data and the decomposition assumptions. For the data, there have been many structural breaks in Brazil's economy since 1949, as the exploratory analysis showed. There have been the military dictatorship, hyperinflation, "the economic miracle" the Real plan, etc. All of this makes a fixed-coefficient VAR very unreliable.

As for the assumptions, as stated before, are very ambiguous, and there are ways to justify IPC also responding to GDP via BC's choice and other market mechanisms. Also, not always the exchange market is so liquid.