UNIVERSITY OF BIRMINGHAM | BIRMINGHAM BUSINESS SCHOOL

Assessment and Feedback: Student Template

Student ID Number(s): 2244195

Programme: Economics

Module: Advanced Econometrics: Theory & Applications

Name of Tutor: Dr Marco Barassi & Dr Lívia Menezes

Assignment Title: Advanced Econometrics - Coursework

Date and Time of Submission: 16/12/2021 11 am

Actual Word Count: 1983

Extension: N **Extension Due Date:**

I do not wish my assignment to be considered for including as an exemplar in the **School Bank of Assessed Work. ***

delete as appropriate

The purpose of this template is to ensure you receive targeted feedback that will support your learning. It is a requirement to complete to complete all 3 sections, and to include this completed template as the first page of every assignment that is submitted for marking (your School will advise on exceptions).
Section One: Reflecting on the feedback that I have received on previous assessments, the following issues/topics have been identified as areas for improvement: (add 3 bullet points). *NB – for first year students/PGTs in the first term, this refers to assessments in your previous institution* • Detail • Clarity • citation
Section Two: In this assignment, I have attempted to act on previous feedback in the following ways (3 bullet points) • detail • clarity • citation

Section Three: Feedback on the following aspects of this assignment (i.e. content/style/approach) would be particularly helpful to me: (3 bullet points)

- Whether models are appropriate
- detail
- logic

Please ensure that you complete and attach this template to the front of all work that is submitted.

By submitting your work online you are confirming that your work is your own and that you understand and have read the University's rules regarding authorship and plagiarism and the consequences that will arise should you submit work not complying with University's Code of Practice on Academic Integrity.

I confirm that I have / have not used a proof-reader(s) (delete as appropriate). If I have used a proof-reader(s) I confirm that I understand the guidance on use of a proof-reader, as specified in the Code of Practice and School guidance.

Ad Econometrics assignment

ID: 2244195

December 2021

Contents

Part A

1 Introduction to Methodology

1.1 Introduction

Crime rate is one of the hot topics in rule of law. Researchers have been looking at various variables to explain it. Levitt (1998) found a deterrent effect of arrest rates on crime rates. Phillips and Land (2012) discovered a delayed incentive effect of unemployment on various crime rates at the county, state, and national levels. Coccia (2017) demonstrated that income inequality is a competing explanation for intentional homicide. Anderson and Bushman (2018) showed that exposure to media violence can increase aggressive behavior. This paper delves into some new variables as well as studied ones - the arrest rate, conviction rate, prison sentence rate, police per capita and percentage of minority - and examines whether they can influence crime rates in order to build on past research and further explain crime rates.

1.2 Dataset and variable choice

The dataset covers 59 variables and 630 records, which includes annual data for 90 counties in US for the years 1981 to 1987. Our explanatory variables of interest are listed below.

lprbarr: log of 'probability' of arrest
lprbconv: log of 'probability' of conviction
lprbpris: log of 'probability' of prison sentence
lpolpc: log of police per capita
lpctmin: log of percentage of minority in 1980

To control for time effects, we also added

d82: =1 if in 1982
d83: =1 if in 1983
d84: =1 if in 1984
d85: =1 if in 1985
d86: =1 if in 1986
d87: =1 if in 1987

The explained variable is

lcrmrte: log of crimes committed per person

For all variables of interest, we chose the logarithmic form, as this made their distribution more natural and expanded their range.

We summarized the data and found that the values of *lprbarr* and *lprbconv* exceed 0, which was not very reasonable. Therefore, we performed data cleaning to change these positive values to 0.

1.3 Methodology

After choosing variables, we got our model as follows:

$$\begin{aligned} lcrmrte_{it} = \beta_0 + \beta_1 lprbarr_{it} + \beta_2 lprbconv_{it} + \beta_3 lprbpris_{it} + \beta_4 lpolpc_{it} + \beta_5 lpctmin_i \\ + \beta_{82} d82_t + \beta_{83} d83_t + \beta_{84} d84_t + \beta_{85} d85_t + \beta_{86} d86_t + \beta_{87} d87_t + a_i + u_{it} \end{aligned}$$

where a_i represents the county characteristics and u_{it} is the error term. i is the number of the 90 counties and t ranges from 81 to 87.

Regression should first be performed using pooled OLS. However, since the dataset is penal data, we wanted to get rid of the county characteristics to avoid serial correlation and heterogeneity bias. Therefore, we also used first difference (FD), fixed effects (FE) and random effects (RE) models in this paper. We then compared the results of the models and conducted relevant tests to figure out the most appropriate one.

2 Result

Table 1: regression results (excluding time dummy variable results)

Variable	**lcrmrte_POLS**	**lcrmrte_FD**	**lcrmrte_FE**	**lcrmrte_RE**
lprbarr	-0.735***		-0.240***	-0.368***
lprbconv	-0.699***		-0.242***	-0.337***
lprbpris	0.079		-0.136***	-0.146***
lpolpc	0.228***		0.288***	0.285***
lpctmin	0.240***			0.244***
D.lprbarr		-0.211***		
D.lprbconv		-0.174***		
D.lprbpris		-0.111***		
D.lpolpc		0.295***		
_cons	-4.214***	0.015	-2.283***	-3.253***
N	630	540	630	630
r2	0.675	0.312	0.352	

Legend: * p<.1; ** p<.05; *** p<.01

We see that all variables of interest are significant at the 1% level in any model. lprbarr and lprbconv have much larger absolute coefficients in the POLS model than in the other models, suggesting that county characteristics other than percentage of minority may play a role in crime rates and that POLS overestimates the effects of arrest and conviction rates. After conducting relevent tests, we found that there is heteroskedasticity, autocorrelation, and cross-sectional correlation. We also performed Hausman test and found significant differences between FE and RE. Considering all these, since FE is more effective than FD and RE, we choose FE as our model for further analysis.

The results show that a 1% increase in arrest rates can reduce crimes committed per person by 0.24%. This supports Levitt's argument about the deterrent effect of arrest rates (1998). Conviction and prison sentence rates are roughly similar, with a one percentage point increase resulting in a 0.24% and 0.14% decrease in crime rates, respectively.

However, a one percent increase in police per capita will increase crime by 0.29 percent, which is contrary to one's intuition, since an increase in police strength should deter crime. This may be because counties with high crime rates must have a stronger police force to face violence and disorder. When crime rates rise, more police must also be deployed. Thus, crime rates and police per capita are positively correlated, both in the time series and in the cross-section. In order to find the true effect of police forces, random assignment should be done, which is not possible in this dataset. Also, if we have higher frequency time series data, we should use a causality test, such as Granger causality, to see which variable is the cause, since the response of police deployment may be too fast to show up in annual data.

Although FE does not include *lpctmin* as it is a county characteristic, we can still be sure that an increase in minority leads to an increase in crime because it has a positive coefficient in both POLS and RE. A 1% increase in the percentage of minorities leads to about 0.24% increase in the crime rate.

3 Discussion

The results show that increasing arrest, conviction, and prison sentence rates can reduce crime rates and benefit social welfare. It is recommended that legal technology be developed to help catch criminals and find more evidence to convict them. Laws could be made stricter to increase the rate of prison sentences to deter potential crimes. Besides, we may seriously consider the problem of minorities.

The study has some drawbacks. The causal relationships in the study are not clear and should be further investigated to demonstrate the deterrent effect of arrest, conviction, and prison sentence rates and whether increased policing reduces crime rates. In addition, the low frequency of the data prevented us from conducting tests of causality. Moreover, the data cover only 90 counties and may not be generalizable.

Part B

1 Introduction to Methodology

1.1 Introduction

The relationship between unemployment and economic growth has been studied extensively in almost all countries using various methods. Zagler (2003) used vector error correction model to show that unemployment and economic growth in European countries are negatively related in the short run and positively related in the long run. Kreishan (2011) used the standard version of the Okun's law to show that Jordan does not conform to the law. Shahid (2014) used autoregressive distributed lag (ARDL) model to show their negative relationship in Pakistan. Few studies have used VAR, SVAR and VECM to study the UK, however, this paper will open up this possibility.

1.2 Dataset and variable choice

We used data obtained from the World Bank for the annual unemployment rate and GDP in the United Kingdom under the names "Unemployment, total (% of total labor force) (national estimate)" and "GDP (constant 2015 US$)", respectively. It contains 46 records ranging from 1971 to 2019 after we conducted trimming. After logging and first differencing the GDP, we obtained the GDP growth rate. These variables are as follows.

unempl_rte: unemployment rate
c_ln_GDP: growth rate of GDP

1.3 Methodology

After performing ADF tests on the variables, we found that *unempl_rte* is I(1), but *c_ln_GDP* is I(0). However, the Johansen test with a lag of 2 years showed that the two series are cointegrated and that there is more than one cointegration relationship, indicating that both series should be stationary. We also did not want to differentiate them to retain their economic meaning. Faced with this contradictory result, we chose to perform both VAR and VECM models and compare their results. We also used the SVAR model to assess the contemporanous impact of the variables.

Then we should choose the best lag length. We calculated some information criteria, including LR test, final prediction error (FPE), Akaike information criterion (AIC) test, Hannan-Quinn information criterion (HQIC), and Schwarz Bayes information criterion (SBIC) test, for different lags, and the results are as follows.

Table 2: choosing lags

Lag	**LL**	**LR**	**FPE**	**AIC**	**HQIC**	**SBIC**
0	220.698		0	-9.72	-9.69	-9.64
1	287.78	134.16	0	-12.524	-12.434	-12.2827*
2	294.451	13.341*	1.1e-08*	-12.6423*	-12.4926*	-12.241
3	296.216	3.53	0	-12.543	-12.333	-11.981

* indicates the best lag chosen by the criteria

Most test results indicate a lag of 2 and we should choose that.

Therefore, the VAR model is as follows:

$$y_t = \Gamma_0 + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + u_t$$

where

$$\Gamma_0 = \begin{pmatrix} \gamma_{10}^{(0)} \\ \gamma_{20}^{(0)} \end{pmatrix}; \Gamma_1 = \begin{pmatrix} \gamma_{11}^{(1)} & \gamma_{12}^{(1)} \\ \gamma_{21}^{(1)} & \gamma_{22}^{(1)} \end{pmatrix}; \Gamma_2 = \begin{pmatrix} \gamma_{11}^{(2)} & \gamma_{12}^{(2)} \\ \gamma_{21}^{(2)} & \gamma_{22}^{(2)} \end{pmatrix}; u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix};$$

$$y_t = \begin{pmatrix} unempl\_rte_t \\ c\_ln\_GDP_t \end{pmatrix}, t = 1, 2, ..., 46$$

The SVAR model is as follows:

$$C_0 y_t = \Gamma_0 + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + u_t$$

where

$$C_0 = \begin{pmatrix} 1 & -c_{12} \\ -c_{21} & 1 \end{pmatrix}; \Gamma_0 = \begin{pmatrix} \gamma_{10}^{(0)} \\ \gamma_{20}^{(0)} \end{pmatrix}; \Gamma_1 = \begin{pmatrix} \gamma_{11}^{(1)} & \gamma_{12}^{(1)} \\ \gamma_{21}^{(1)} & \gamma_{22}^{(1)} \end{pmatrix}; \Gamma_2 = \begin{pmatrix} \gamma_{11}^{(2)} & \gamma_{12}^{(2)} \\ \gamma_{21}^{(2)} & \gamma_{22}^{(2)} \end{pmatrix}; u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix};$$

$$y_t = \begin{pmatrix} unempl\_rte_t \\ c\_ln\_GDP_t \end{pmatrix}, t = 1, 2, ..., 46$$

In this case, when estimating c_{21}, we assume $c_{12} = 0$ and when estimating c_{12}, we assume $c_{21} = 0$.

The VECM model is as follows:

$$\Delta y_t = \Gamma_0 + \Gamma_1 \Delta y_{t-1} - \Pi u_{t-1} + e_t$$

where

$$\Gamma_0 = \begin{pmatrix} \gamma_{10}^{(0)} \\ \gamma_{20}^{(0)} \end{pmatrix}; \Gamma_1 = \begin{pmatrix} \gamma_{11}^{(1)} & \gamma_{12}^{(1)} \\ \gamma_{21}^{(1)} & \gamma_{22}^{(1)} \end{pmatrix}; \Pi = \begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{pmatrix}; \hat{u}_{t-1} = \begin{pmatrix} \hat{u}_{1,t-1} \\ \hat{u}_{2,t-1} \end{pmatrix}; e_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix};$$

$$\Delta y_t = \begin{pmatrix} \Delta unempl\_rte_t \\ \Delta c\_ln\_GDP_t \end{pmatrix}, t = 1, 2, ..., 46$$

with $-\Pi u_{t-1}$ the adjustment term.

2 Result

Table 3: VAR, SVAR and VECM result

Dependent Variable	Independent Variable	VAR	SVAR	VECM
unempl_rte	c_ln_GDP		-0.248***	
	L1.unempl_rte	1.363***		
	L2.unempl_rte	-0.459***		
	L1.c_ln_GDP	-0.187***		
	L2.c_ln_GDP	0.129**		
	_cons	0.008**		
	r2	0.92		
c_ln_GDP	unempl_rte		-1.418***	
	L1.unempl_rte	0.347		
	L2.unempl_rte	-0.027		
	L1.c_ln_GDP	0.317**		
	L2.c_ln_GDP	-0.153		
	_cons	-0.006		
	r2	0.31		
D.unempl_rte	LD.unempl_rte			0.493***
	LD.c_ln_GDP			-0.183***
	adj			0
	cons			0
	r2			0.45
D.c_ln_GDP	LD.unempl_rte			-0.021
	LD.c_ln_GDP			0.23
	adj			0.185***
	cons			0
	r2			0.50
N		46	46	46

Legend: * p<.1; ** p<.05; *** p<.01

We can see from the VAR model that the lagged terms can explain the unemployment rate well. Unemployment rate with one period lag has a positive effect on the unemployment rate, while unemployment rate with two periods lag has a negative effect on it. The GDP growth rate with one lag has the expected negative impact on the unemployment rate, but the one with two lags is positive. On the other hand, the lag terms do not explain the GDP growth rate well, except that the past GDP growth rate can be partially carried over to the next period. The stability test shows that the model is stable. The residual autocorrelation is small and the errors are normally distributed. The Grangar causality test shows that each variable is the cause of the other.

The impulse diagram1 shows that an initial increase in GDP growth rate leads to a decrease in the unemployment rate, and when GDP growth rate recovers, the unemployment rate also picks up. The initial increase in unemployment leads to a sudden drop in GDP growth rate, indicating that GDP growth rate is vulnerable to shocks. Unemployment rate deviates even further when the shock still persists, but finally both variables regain stability. The variance decomposition plot2 shows that shocks from GDP growth rate and unemployment each explains about half of the variance in GDP growth rate, while the variance of unemployment rate comes almost entirely from its own shocks.

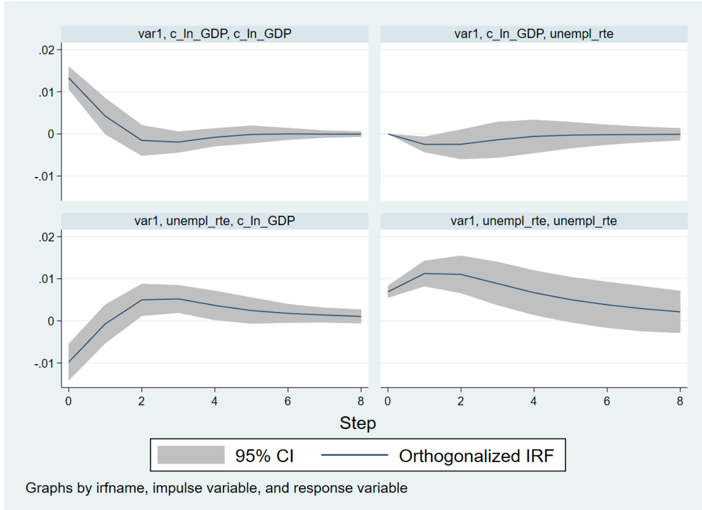

Figure 1: impulse graph

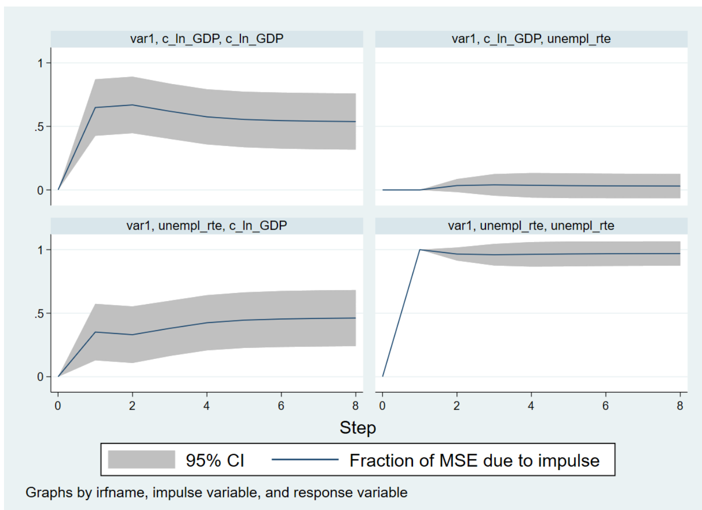

Figure 2: variance decomposition graph

The SVAR model shows the contemporanous effects of these two variables on each other and shows that both have a negative impact on the other. Specifically, a 1 percentage point increase in GDP implies a 0.25 percentage point decrease in the unemployment rate, while a 1 percentage point increase in unemployment implies a 1.42% decrease in GDP growth, though the results may be biased because of the restrictions on estimating the parameters.

The VECM model also supports a positive effect of lagged unemployment and a negative effect of lagged GDP growth on unemployment. However, the adjustment term does not have an effect. On the other hand, the adjustment term explains much of the GDP growth rate. When the unemployment rate reaches a high level, the GDP growth rate follows suit and supports their positive long-run correlation. The covariance equation shows that $(unempl\_rte - 4.96 * c\_ln\_GDP)$ is stable, implying that they are positively correlated in the long run. Subsequent tests show that there is no residual autocorrelation and the errors are normally distributed. The model is also stable using stability tests.

3 Discussion

This study shows that the unemployment rate and GDP growth rate negatively affect each other in the short run, especially the effect of GDP growth on the unemployment rate. However, they are positively cointegrated in the long run. This reaffirms Zagler's study in 2003 and the short-run relationship is similar to Okun's law.

The study has some drawbacks. The most obvious one is that the data we use do not exactly match the underlying assumptions, leading to questionable results. Since GDP and unemployment rate are objective and cannot be changed, there is not a lot of room for us to make progress. What we can do further is to obtain data at higher frequencies or longer time ranges to increase the sample points and see if they meet the presupposition. If both time series data are I(0) then we should focus on the VAR and SVAR models and if they are I(1) and cointegrated then the VECM should be chosen. If the problems still exist, more advanced models like LA-VAR can be considered. In addition, the study is limited to the UK and the results may not be appropriate for other countries.

Appendix (Stata code)

****

*** Advanced Econometrics Assignment Stata Code ***

****

Part A

```
clear all

cd ""


***Step 1: Install relevent packages***

ssc install xttest3

ssc install xtcsd


***Step 2: Prepare data***

use crime


***Step 3: Set county as cross-section variable and year as time variable***

tsset county year


***Step 4: Keep only the variables needed***

keep county year lcrmrte lprbarr lprbconv lprbpris lpolpc lpctmin d8*


***Step 5: Data Cleaning***

summarize

/* We find that there are unreasonable values. */


***Step 5.1: Set the maximum value a variable can get***

scalar Maximum=0

/* Log of a probability can not exceed 0. */


***Step 5.2: Change the extreme values to the maximum value***

replace lprbarr = Maximum if lprbarr >Maximum
```

replace lprbconv = Maximum if lprbconv >Maximum

Step 6: Conduct pooled OLS and save the result

reg lcrmrte lprbarr lprbconv lprbpris lpolpc lpctmin d82 d83 d84 d85 d86 d87, robust

est store lcrmrte_POLS

Step 7: Conduct FD and save the result

reg D.lcrmrte D.lprbarr D.lprbconv D.lprbpris D.lpolpc D.lpctmin D.d83 D.d84 D.d85 D.d86 D.d87, robust

est store lcrmrte_FD

Step 8: Conduct FE and save the result

xtreg lcrmrte lprbarr lprbconv lprbpris lpolpc lpctmin d82 d83 d84 d85 d86 d87, fe

est store lcrmrte_FE

Step 9: Heteroskadasticity test

xttest3

/* There is heteroskadasticity.*/

Step 10: Conduct RE and save the result

xtreg lcrmrte lprbarr lprbconv lprbpris lpolpc lpctmin d82 d83 d84 d85 d86 d87, re

est store lcrmrte_RE

Step 11: Cross-section correlation test

qui xtreg lcrmrte lprbarr lprbconv lprbpris lpolpc lpctmin d82 d83 d84 d85 d86 d87, re

xtcsd,fre

/* There is cross-section correlation. */

Step 12: Autocorrelation test (should use "search xtserial" to install xtserial)

tab county,gen(county)

xtserial lcrmrte lprbarr lprbconv lprbpris lpolpc lpctmin county2-county90 d82 d83 d84 d85 d86 d87

/* There is autocorrelation. */

Step 13: Compare the results

estimates table lcrmrte*, b(%9.3f) star(.01 .05 .10) stats(N r2) drop(d82 d83 D1.d83 d84 D1.d84 d85 D1.d85 d86 D1.d86 d87 D1.d87)

Step 14: Conduct Hausman test

hausman lcrmrte_FE lcrmrte_RE, sigmamore

/* There is significant difference between RE and FE */

Part B

clear all

cd ""

Step 1: Install relevent packages

ssc install wbopendata

ssc install sxpose

Step 2: Get and preprocessing data

Step 2.1: Get data for the UK

wbopendata, language(en - English) country(GBR;) topics() indicator()

Step 2.2: Keep only the data we want

keep if indicatorname == "GDP (constant 2015 US$)" | indicatorname == "Unemployment, total (% of total labor force) (national estimate)"

Step 2.3: Drop irrelavent information

drop countrycode countryname region regionname adminregion adminregionname incomelevel incomelevelname lendingtype lendingtypename indicatorname indicatorcode

Step 2.4: Transpose the data

```
sxpose, clear force

***Step 2.5: Rename the variables***
rename (_var1 _var2)(unempl_rte_temp GDP)

***Step 2.6: Change the data from string type to float***
destring unempl_rte_temp GDP, replace

***Step 2.7: Delete rows with null***
drop if unempl_rte_temp==. | GDP == .

***Step 2.8: Add time variable***
gen year = _n+1970

***Step 2.9: Change the unit of unemployment rate from % to unity***
gen unempl_rte = unempl_rte_temp/100

***Step 2.10: Drop original unemployment rate data***
drop unempl_rte_temp

***Step 2.11: Change GDP to ln form***
gen ln_GDP = ln(GDP)

***Step 2.12: Take first difference of ln(GDP) so that we get the GDP growth rate***
gen c_ln_GDP = ln_GDP - ln_GDP[_n-1]

***Step 3: Set year as time variable***
tsset year

***Step 4: Perform stationarity test***
dfuller unempl_rte
```

```
/* unempl_rte is not stationary. */

dfuller d.unempl_rte

/* d.unempl_rte is stationary. */

dfuller c_ln_GDP

/* c_ln_GDP is stationary. */


***Step 5: Determine the optimal lag***

varsoc unempl_rte c_ln_GDP, maxlag(3)


***Step 6: Perform Johansen cointegration test***

vecrank unempl_rte c_ln_GDP, lag(2)

/* The result indicates two cointegration relationship. */


***Step 7: VAR model***


***Step 7.1: Run the unrestricted VAR model***

var unempl_rte c_ln_GDP, lags(1/2)


***Step 7.2: Run the Granger Test***

vargranger

/* The two variables affects each other. */


***Step 7.3: Run the autocorrelation test***

varlmar

/* It indicates little autocorrelation. */


***Step 7.4: Run the normality test***

varnorm

/* Errors are normally distributed. */


***Step 7.5: Run the stability test and get the graph***
```

```
varstable, graph

/* The model is stable. */


***Step 7.6: Conduct impulse response analysis and plot the graph***

irf create var1, set(var_irf) replace

irf graph oirf, irf(var1)


***Step 7.7: Conduct forecast-error variance decompositions and plot the graph***

irf graph fevd, irf(var1) impulse(unempl_rte c_ln_GDP) response(unempl_rte c_ln_GDP)


***Step 8: SVAR model***


***Step 8.1: Set the restrictive matrix***

matrix A=(1,0\.,1)

matrix B=(.,0\0,.)


***Step 8.2: Conduct SVAR model***

svar unempl_rte c_ln_GDP, aeq(A) beq(B)

svar c_ln_GDP unempl_rte, aeq(A) beq(B)


***Step 9: VECM model***


***Step 9.1: Conduct VECM model***

vec unempl_rte c_ln_GDP, lags(2)


***Step 9.2: Run the autocorrelation test***

veclmar

/* It indicates little autocorrelation. */


***Step 9.3: Run the normality test***

vecnorm
```

/* Errors are normally distributed. */

Step 9.4: Run the stability test and get the graph

vecstable, graph

/* The model is stable. */

Reference

Anderson, C.A. and Bushman, B.J. (2018) ‘Media violence and the general aggression model’, *Journal of Social Issues*, 74(2), pp. 386-413.

Coccia, M. (2017) ‘A Theory of general causes of violent crime: Homicides, income inequality and deficiencies of the heat hypothesis and of the model of CLASH’, *Aggression and Violent Behavior*, 37, pp. 190-200.

Kreishan, F.M. (2011) ‘Economic growth and unemployment: An empirical analysis’, *Journal of Social Sciences*, 7(2), pp. 228-231.

Levitt, S.D. (1998) ‘Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error?’, *Economic inquiry*, 36(3), pp. 353-372.

Phillips, J. and Land, K.C. (2012) ‘The link between unemployment and crime rate fluctuations: An analysis at the county, state, and national levels’, *Social science research*, 41(3), pp.681-694.

Shahid, M. (2014) ‘Effect of Inflation and Unemployment on Economic Growth in Pakistan’, *Journal of Economics and Sustainable Development*, 5 (15), pp. 103-106.

Zagler, M. (2003) ‘A Vector Correction Model of Economic Growth and Unemployment in Major European Countries and an Analysis of Okun’s Law’, *Applied Econometrics and International Development*, 3(3), pp. 93-118.