Lecture 10. Case Studies with R Demonstration

Muyi LI

WISE&SOE, Xiamen University

Outline

1. Monthly Global Temperature
2. Financial Asset Volatility
3. Quaterly GDP Growth Rate in UK, Canada and US

Motivation

- Demonstrate applications of methods discussed in previous Chapters;
- To show usefulness and limitations of time series models;
- To gain further experience in analyzing time series data with R.

General guidelines

- A main difficulty for the beginners of time series analysis is finding an adequate model for a given series. This is particularly so when the dynamic dependence of the data is complex or when many models seem to fit the data well.
- "*All models are wrong, but some are useful*"—George Box's dictum (1976)
- Our goal is to find an appropriate model that is useful to the objective of data analysis. It would not be surprising that a reader can find alternative models from these data sets.

General guidelines

- First, data are only part of information available in an application. We may have some prior knowledge about the problem at hand. In this situation, it is important to make use of substantive information in model selection. Combination and cross-validation between prior knowledge and data can improve model selection.
- Second, in some cases, many models are available and the distinction between these competing models is small. The issue of model selection then becomes less important and one can comfortably use one of the models.
- Third, we may combine several competing models for pooling or combining forecasts;
- Fourth, a general principle is to start with a simple model, that is "*Keeping it sophistically simple (KISS)*" (Arnold Zellner)

Case 1: Global Temperature Anomalies

There are several data sets available for global temperature anomalies:

- Goddard Institute for Space Studies (GISS), National Aeronautics and Space Administration(NASA): https://data.www.giss.nasa.gov/gistemp/
- National Climatic Data Center (NCDC), National Oceanic and Atmospheric Adminstration(NOAA): https://www.giss.nasa.gov/
- We employ the series of monthly means based on land-surface air temperature anomalies of GISS, NASA.
- However, we obtained similar results from the data of NOAA. The same models apply to both series.
- Reference: Chapter 3 of *An introduction to analysis of financial data with R. by Ruey S. Tsay, 2012.*

Case 1: Global Temperature Anomalies

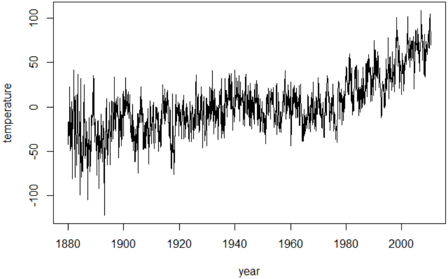
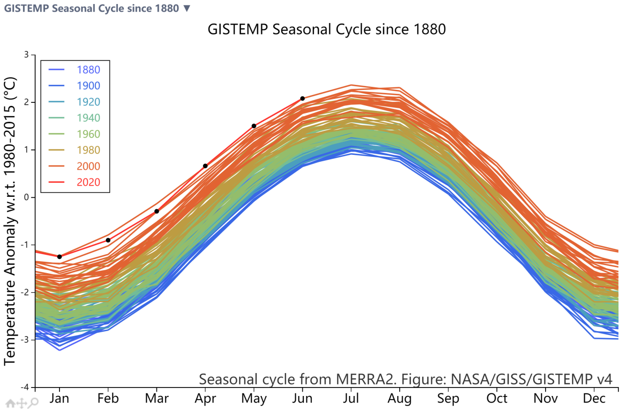


Figure 1: Monthly global temperatures from January 1880 to August 2010 (denoted by G_t), sample size=1568.

An upward trend is clearly seen from the plot. In particular, the slope of the trend seems to increase in the early 1980s. On the other hand, the variability of the temperature is relatively stable over the 131 years.

Case 1: Global Temperature Anomalies



Case 1: Global Temperature Anomalies

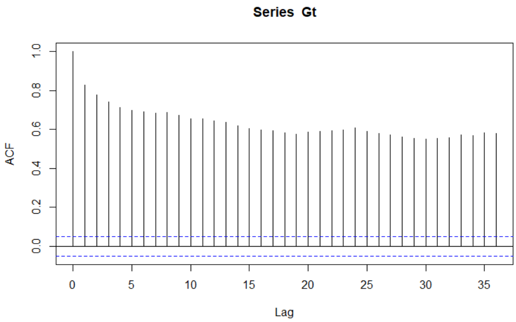


Figure 2: ACF of G_t.

Model 1:Unit-Root Stationarity

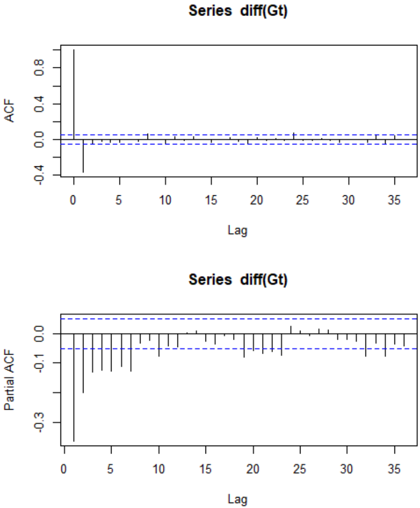


Figure 3: ACF of differenced G_t.

Model 1: Unit-Root Stationarity

- Model 1: ARIMA(1,1,2):

$$(1 - \phi_1 B)(1 - B) G_t = (1 + \theta_1 B + \theta_2 B^2) e_t,$$

- The significance of ACF at lag 24 is understandable because of the seasonal nature of temperature. On the other hand, it is not easy to explain the serial correlation at lag 8.
- Model 1: Seasonal ARIMA$(1,1,2) \times (0,0,2)_{12}$ or Seasonal ARIMA$(1,1,2) \times (0,0,1)_{24}$

$$(1 - \phi_1 B)(1 - B) G_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \theta_{24} B^{24}) e_t \quad (1)$$

Model 2: Trend-Stationarity

- In the literature, some analysts and scientists use time trend to model the global temperature anomalies. By time trend, we mean using time index as an explanatory variable. Consider the model

$$G_t = \beta_0 + \beta_1 t + Z_t$$

  where Z_t is an innovation series denoting the deviation of the global temperature anomalies from a time trend.

Model 2: Trend-Stationarity

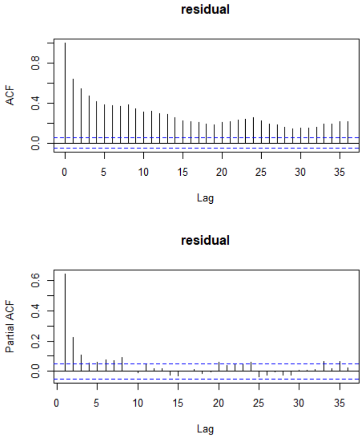


Figure 4: Sample autocorrelations and partial autocorrelations of the innovation series Z_t for monthly global temperature anomalies.

- Putting information together and keeping the order simple, we start with an ARMA(2,1) model for Z_t, Then, the model for G_t becomes

$$(1-\phi_1 B-\phi_2 B^2)(G_t-\beta_0-\beta_1 t)=(1+\theta_1 B)e_t$$

Trend-Stationarity

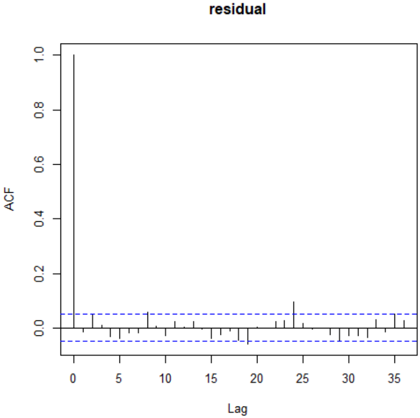


Figure 5: ACF of residual

Trend-Stationarity

- However, the residual ACFs of the fitted model show a significant value at lag 24. Consequently, we further refine the model and obtain
- $$(1-\phi_1 B-\phi_2 B^2)(G_t-\beta_0-\beta_1 t)=(1+\theta_1 B)(1+\theta_{24}B^{24})e_t \quad (2)$$

Trend-Stationarity

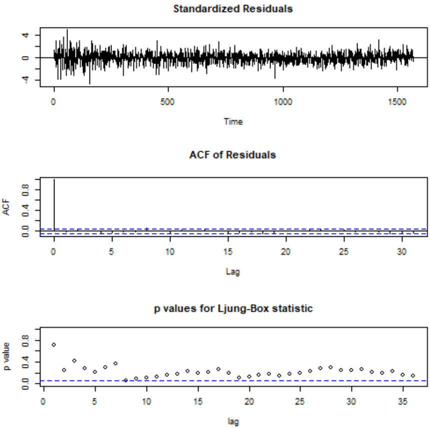


Figure 6: Diagnostic checking of the fitted model in Equation (2). (a) Standardized residuals, (b) ACF of residuals, (c) p-values for Ljung-Box statistic.

Trend-Stationarity

- On the basis of model in Eq.(2), the global temperature increases on an average $0.0529/100^o$ C per month. That is, the global temperature increases $0.00635^o C$ per year.
- This is very significant because it implies that global temperature on an average will increase $1^o C$ every 157 year.

Out-of-Sample Comparison

- We divide the sample into modeling and forecasting subsamples with the latter consisting of the last 200 observations.
- We compute the 1-step ahead prediction of the two competing models in Equations (1) and (2).

Table 1:

Model	RMSFE	MAFE
Model 1 (Unit root-stationary)	14.526	11.167
Model 2 (Trend-stationary)	15.341	11.966

- The difference-stationary model is preferred based on the 1-step ahead prediction. The drop in RMSFE is about $(15.341 - 14.526)/14.526 = 5.6\%$.

Long-Term Prediction

- Using August 2010, which gives the last data point, as the forecast origin, we compute 1-step to 1200-step ahead predictions of the monthly global temperature anomalies.
- In other words, we use the models built based on data of the past 131 years to predict the global temperatures for the next 100 years.

Long-Term Prediction on Unit-Root Stationary Model 1

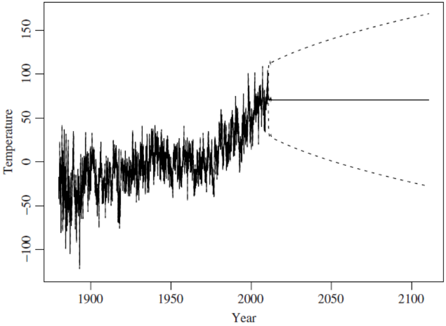


Figure 7: Long-term point and interval forecasts of the monthly global temperature anomalies based on the difference-stationary model in Equation (1). The forecast origin is August 2010 and the forecast horizon is 100 years.

Long-Term Prediction On Unit-root Stationary Model 1

- First, similar to other unit-root models, the long-term forecasts converge to a constant represented by a horizontal line in the plot. The level of this horizontal line depends on the forecast origin. Second, the length of the 95% interval forecasts continues to grow with the forecast horizon. In fact, the length of the interval diverges to infinity eventually.
- These two features have important implications in forecasting. First, they indicate that the long-term forecasts are rather uncertain. This makes intuitive sense because long-term predictions of the model are dominated by its random walk component and for a random walk the current value contains little information about the future. Second, they demonstrate clearly that the model is only informative in short-term prediction.

Long-Term Prediction on Trend-Stationary Model 2

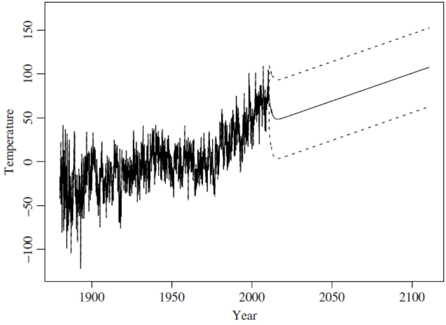
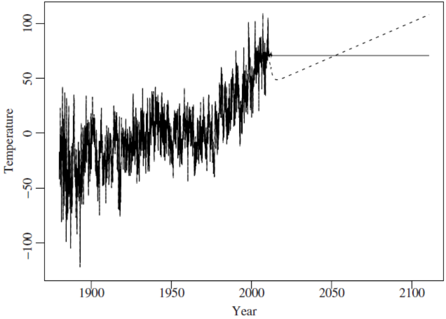


Figure 8: Long-term point and interval forecasts of the monthly global temperature anomalies based on the trend-stationary Model 2. The forecast origin is August 2010 and the forecast horizon is 100 years.

Long-Term Prediction Using Trend-Stationary Model 2

- First, because of the positive time slope, the predictions grow with the forecast horizon.
- Second, the lengths of the interval forecasts are stable over time. In fact, the lengths quickly converge to a constant with the constant being approximately $4\sigma_z$, where σ_z is the sample standard error of the innovation series z_t .
- The innovation series z_t in Eq. (1) is stationary. As such the variances of the forecasts of z_t converge to its variance, σ_z^2 , when the forecast horizon increases.
- For the trend-stationary model, in Equation (2), the prediction of the time trend is certain conditioned on the coefficients β_0 and β_1 being fixed. The uncertainty in forecasts is determined by that of z_t . Consequently, the variances of forecast errors of the model in Equation (2) converge to that of z_t.

Point forecasts of the monthly global temperature anomalies based on two competing models. The forecast origin is August 2010 and the forecast horizon is 100 years.



- There are many volatility models available in the literature.
- The univariate models include: ARCH (Engle ,1982), GARCH (Bollerslev,1986), EGARCH(Nelson ,1991), TGARCH(Glosten et al., 1993; Zakoian, 1994), the nonsymmetric GARCH(NGARCH)(Engle and Ng,1993; Duan, 1995);
- Besides GARCH family, the stochastic volatility (SV) models of Melino and Turnbull (1990), Taylor (1994), Harvey et al. (1994), and Jacquier et al. (1994) are also of interest.
- The modeling of high frequency data will not discussed here.
- *Reference:*
  - Chapter 4 of An introduction to analysis of financial data with R. by Ruey S. Tsay, 2012.
  - Chapter 3 of Elements of Financial Econometrics. by Jianqing Fan and Qiwei Yao, 2015.

Case 2: GARCH Modeling

Building a volatility model for an asset return series consists of four steps:

1. Specify a mean equation by testing for serial dependence in the data and, if necessary, building an ARMA model for the return series to remove any linear dependence.
2. Use the residuals of the mean equation to test for ARCH effects.
3. Specify a volatility model if ARCH effects are statistically significant, and perform a joint estimation of the mean and volatility equations.
4. Check the fitted model carefully and refine it if necessary.

References

- Cont, R. (2001). Empirical properties of assets returns: stylized facts and statistical issues. Quantitative Finance, 2001, 223-236.
- Hansen P. R. and Lundea, A. (2005). Forecast comparison of volatility models: does anything beat a GARCH (1,1)? Journal of Applied Econometrics, 20: 873-889.

Case 3: GDP of UK,Canada,and US

Reference: Chapter 2 of *Multivariate Time Series Analysis with R and Financial Applications. by Ruey S. Tsay, 2014.*

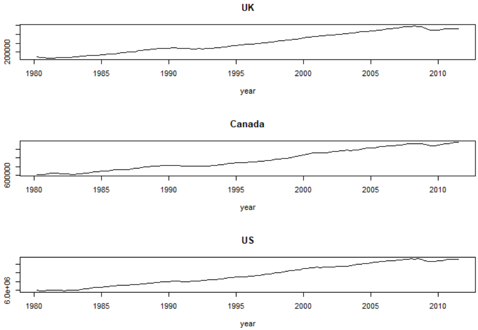


Figure 9: Quarterly Real Gross Domestic Products of United Kingdom, Canada, and United States from the second quarter of 1980 to the first quarter of 2011. The data are from Federal Reserve Bank at St. Louis.

Growth Rate of GDP of UK,Canada,and US

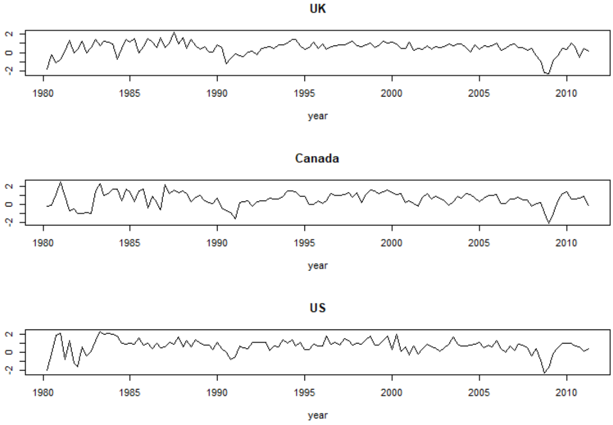


Figure 10: Quarterly growth rates, in percentages, of Real Gross Domestic Products of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011.

Cross-correlation matrices

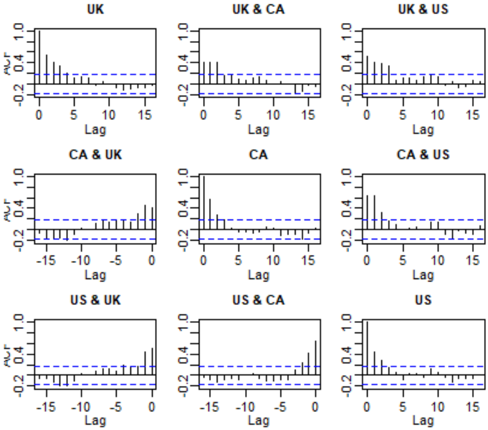


Figure 11: Cross-correlation matrices.

Select Order

- package:MTS; function:VARorder
- Based on the plots, there exits strong serial or cross-correlation in each (i,j) position, $1 \leq i,j \leq 3$.
- We apply all three information criteria to the data.

$$
\begin{aligned}
AIC(p) &= ln|\hat{\Sigma}_p| + \frac{2}{T}pd^2 \\
BIC(p) &= ln|\hat{\Sigma}_p| + \frac{ln(T)}{T}pd^2 \\
HQ(p) &= ln|\hat{\Sigma}_p| + \frac{2ln(ln(T))}{T}pd^2
\end{aligned}
$$

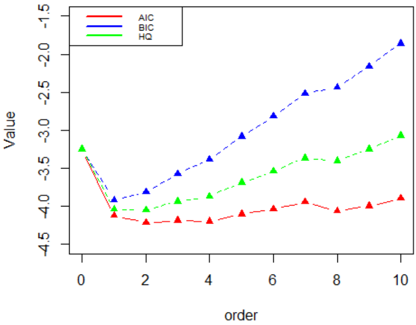


Figure 12: Information criteria.

According to the Figure, a VAR(1) or VAR(2) model may serve as a starting model for the three-dimensional GDP series ($d = 3$).

Estimate Coeffients of VAR(2)

- function: VAR.
- The fitted VAR(2) model for the percentage growth rates of quarterly GDP of UK,Canada, and US is

$$z_t = \begin{pmatrix} 0.13 \\ 0.12 \\ 0.29 \end{pmatrix} + \begin{pmatrix} 0.38 & 0.10 & 0.05 \\ 0.35 & 0.34 & 0.47 \\ 0.49 & 0.24 & 0.24 \end{pmatrix} z_{t-1} + \begin{pmatrix} 0.06 & 0.11 & 0.02 \\ -0.19 & -0.18 & -0.01 \\ -0.31 & -0.13 & 0.09 \end{pmatrix} z_{t-2} + \epsilon_t$$

- The standard errors of the coefficient estimates are

$$\begin{pmatrix} 0.07 \\ 0.07 \\ 0.08 \end{pmatrix}, \begin{pmatrix} 0.09 & 0.09 & 0.09 \\ 0.09 & 0.10 & 0.09 \\ 0.11 & 0.11 & 0.10 \end{pmatrix}, \begin{pmatrix} 0.09 & 0.08 & 0.09 \\ 0.09 & 0.09 & 0.09 \\ 0.11 & 0.09 & 0.11 \end{pmatrix}.$$

- There are some of the estimates are not statistically significant at the usual 5% level.

Refined VAR(2)

- package: MTS; function: refVAR
- The refined model is

$$z_t = \begin{pmatrix} 0.16 \\ 0 \\ 0.28 \end{pmatrix} + \begin{pmatrix} 0.46 & 0.21 & 0 \\ 0.33 & 0.27 & 0.49 \\ 0.47 & 0.23 & 0.23 \end{pmatrix} z_{t-1} + \begin{pmatrix} 0 & 0 & 0 \\ -0.19 & 0 & 0 \\ 0.30 & 0 & 0 \end{pmatrix} z_{t-2} + \epsilon_t$$

- The standard errors of the coefficient estimates are

$$\begin{pmatrix} 0.07 \\ 0 \\ 0.08 \end{pmatrix}, \begin{pmatrix} 0.08 & 0.07 & 0 \\ 0.09 & 0.09 & 0.09 \\ 0.10 & 0.09 & 0.10 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0.09 & 0 & 0 \\ 0.10 & 0 & 0 \end{pmatrix}.$$

- AIC=-3.53, BIC=-3.30, HQ=-3.43.

- function: MTSdiag

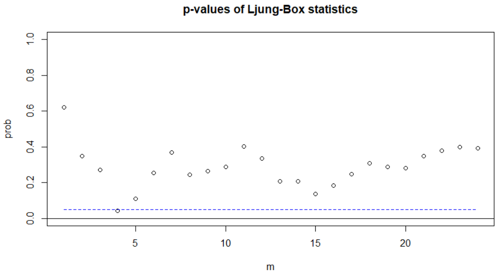


Figure 13: p-values of the $Q_k(m)$ statistics applied to the simplified VAR(2) model.

Diagnostic Checking

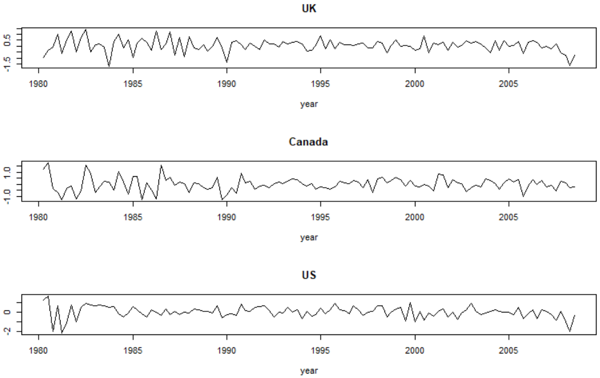


Figure 14: Plot of residuals of VAR(2) model.

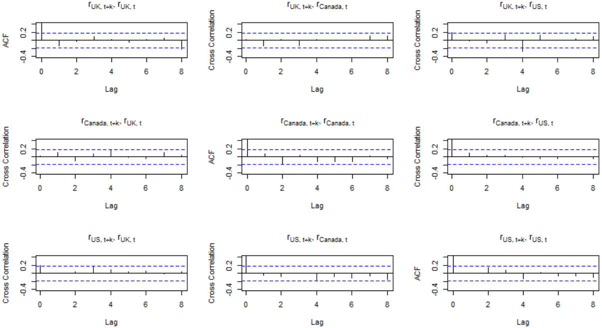


Figure 15: Autocorrelation and crosscorrelation of residuals of VAR(2) model.

Diagnostic Checking

- In conclusion, the simplified VAR(2) model is adequate for the GDP growth rate series. The model can be written as
  UK:$z_{1t} = 0.16 + 0.47z_{1,t-1} + 0.21z_{2,t-1} + \epsilon_{1,t}$
  CA:$z_{2t} = 0.33z_{1,t-1} + 0.27z_{2,t-1} + 0.5z_{3,t-1} - 0.2z_{1,t-2} + \epsilon_{2,t}$
  US:$z_{3t} = 0.28 + 0.47z_{1,t-1} + 0.23z_{2,t-1} + 0.23z_{3,t-1} - 0.3z_{1,t-2} + \epsilon_{3,t}$
- From the fitted model,
  - The GDP growth rate of UK does not depend on the lagged growth rates of the US in the presence of lagged Canadian GDP growth rates, but the UK growth rate depends on the past growth rate of Canada.
  - The GDP growth rate of Canada is dynamically related to the growth rates of UK and US. Similarly, the GDP growth rate of US depends on the lagged growth rates of UK and Canada.
  - In summary, the simplified VAR(2) model indicates that the GDP growth rate of UK is conditionally independent of the growth rate of the US given the Canadian growth rate.

Impluse Response Function

function: VARirf

- There is another approach to explore the relation between variables. As a matter of fact, we are often interested in knowing the effect of changes in one variable on another variable in multivariate time series analysis.
- For instance, suppose that the bivariate time series z_t consists of monthly income and expenditure of a household, we might be interested in knowing the effect on expenditure if the monthly income of the household is increased or decreased by a certain amount, for example, 5%.
- This type of study is referred to as the impulse response function in the statistical literature or multiplier analysis in economics.

Impluse Response Function

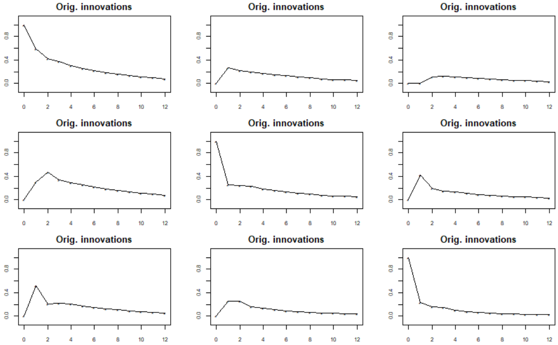


Figure 16: Impulse response functions of the simplified VAR(2) model for the quarterly growth rates of real gross domestic products of UK, Canada, and US from the second quarter of 1980 to the second quarter of 2011.

Impulse Response Function

- From the plots, the impulse response functions decay to 0 quickly. This is expected for a stationary series.
- The upper-right plot shows that there is a delayed effect on the U.K. GDP growth rate if one changed the U.S. growth rate by 1. This delayed effect is due to the fact that a change in U.S. rate at time t affects the Canadian rate at time t + 1, which in turn affects the U.K. rate at time $t + 2$.
- This plot thus shows that the impulse response functions show the marginal effects, not the conditional effects. Recall that conditional on the lagged Canadian rate, the growth rate of U.K. GDP does not depend on lagged values of U.S. growth rate.

Forecast Error Variance Decomposition

function: FEVdec

Variable	Step	United Kingdom	Canada	United States
United Kingdom	1	1.0000	0.0000	0.0000
	2	0.9645	0.0355	0.0000
	3	0.9327	0.0612	0.0071
	4	0.9095	0.0775	0.0130
	5	0.8956	0.0875	0.0170
Canada	1	0.0036	0.9964	0.0000
	2	0.1267	0.7400	0.1333
	3	0.1674	0.6918	0.1407
	4	0.1722	0.6815	0.1462
	5	0.1738	0.6767	0.1495
United States	1	0.0473	0.1801	07726
	2	0.2044	0.1999	0.5956
	3	0.2022	0.2320	0.5658
	4	0.2028	0.2416	0.5556
	5	0.2028	0.2460	0.5512

Figure 17: Forecast error variance decomposition for 1-step to 5-step ahead predictions based on the fitted simplified VAR(2) model. The forecast origin is the second quarter of 2011.