Econometric Methods II Lecture Note 01

Junyong Kim

January 23, 2017

Database: FRED (economic research, GDPs)

1 Time series basics

1.1 Stochastic process

- A random variable indexed with time is called a stochastic process.
- Stochastic process $\Rightarrow$ chance and time
- $\{Y_t\}_{t=-\infty}^{\infty} = \{\dots, Y_1, \dots, Y_t, \dots\}$ is a stochastic process.

1.2 Strict stationary

- Definition: $\{Y_t\}$ is strictly stationary if for any finite integer r and any set of subscripts $\{t_1, \dots, t_r\}$, pdf$(Y_1, \dots, Y_{t_r})$ depends on $\{t_1 - t, \dots, t_r - t\}$ and not on t.
- Example: pdf(Y_2, Y_6)=pdf(Y_1, Y_5), pdf(Y_5, Y_{15})=pdf(Y_4, Y_{14})
- Result: If $\{Y_t\}$ is strictly stationary and $f(\cdot)$ is any function, then $\{f(Y_t)\}$ is also strictly stationary.
- If $\{Y_t\}$ is strictly stationary, then $\left\{Y_t^2\right\}$, $\left\{Y_t^3\right\}$ are also strictly stationary.

1.3 Weak stationary

- Definition: $\{Y_t\}$ is weakly stationary if mean, variance and covariance of this time-series do not depend on time.

$$\begin{aligned} \mathrm{E}[Y_t] &= \mu \\ \mathrm{Var}[Y_t] &= \sigma^2 \\ \mathrm{Cov}[Y_t, Y_{t-j}] &= \mathrm{E}[(Y_t - \mu)(Y_{t-j} - \mu)] \\ &= \gamma_j \end{aligned}$$

- Why we need "stationarity?"

- Without stationarity, one must estimate infinite parameters (e.g. mean, variance and covariance of all observations) and unfortunately this is impossible due to the finite numbers of degrees of freedom.

- Classic example: White noise process $\varepsilon_t \sim \text{wn}\left(0, \sigma^2\right)$. It can be easily shown that the white noise process $\{\varepsilon_t\}$ is stationary.

$$\begin{aligned} \mathrm{E}\left[\varepsilon_t\right] &= 0 \\ \mathrm{Var}\left[\varepsilon_t\right] &= \sigma^2 \\ \mathrm{Cov}\left[\varepsilon_t, \varepsilon_{t-j}\right] &= \gamma_j \\ &= 0 \end{aligned}$$

- Mean, variance and covariance are independent of time. Therefore, $\{\varepsilon_t\}$ is stationary.

1.4 Examples of non-stationary processes

1.4.1 Trend stationary process

- If a time-series can be made stationary by detrending it, we call it trend stationary process.

- Example: $Y_t = \alpha + \beta t + \varepsilon_t$, where t is time-trend.

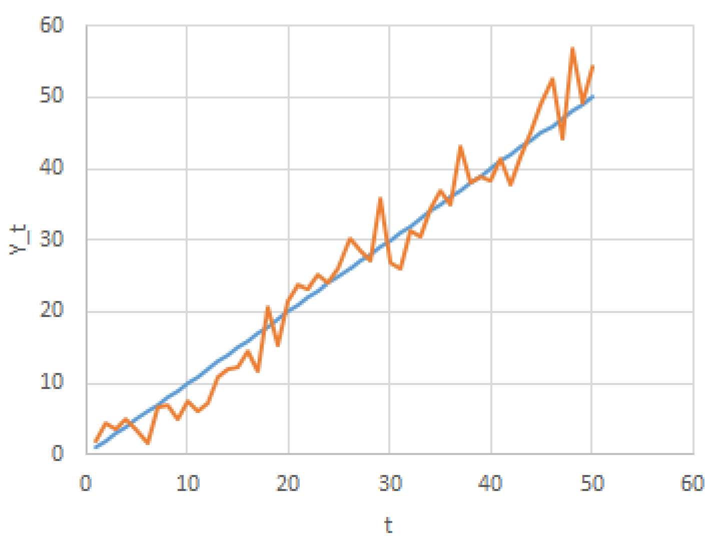


Figure 1: $Y_t = t + \varepsilon_t$, $\varepsilon_t \sim N(0, 10)$

- ε_t is detrended y_t. $\Rightarrow$ Essentially the residual from detrended regression $\varepsilon_t = Y_t - \alpha - \beta t$ is stationary. $\Rightarrow$ $\{Y_t\}$ is trend-stationary and it is a non-stationary process.

- $\mathrm{E}[Y_t] = \alpha + \beta t \Rightarrow$ Mean depends on time. Therefore, $\{Y_t\}$ is a non-stationary process. Once it is detrended, it becomes stationary.

1.4.2 Difference stationary process

- Suppose $\{Y_t\}$ is a non-stationary process. If it can be made stationary by differencing it, we call it difference stationary process.

- Example: Random Walk model $Y_t = Y_{t-1} + \varepsilon_t$. By substituting recursively,

$$\begin{aligned} Y_t &= Y_{t-1} + \varepsilon_t \\ &= Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= Y_0 + \varepsilon_1 + \cdots + \varepsilon_t \\ &= Y_0 + \sum_{k=1}^{t} \varepsilon_k \end{aligned}$$

- Variance depends on time. $\Rightarrow$ $\{Y_t\}$ is non-stationary.

$$\begin{aligned} \Rightarrow \mathrm{E}[Y_t] &= Y_0 + \mathrm{E}\left[\sum_{k=1}^{t} \varepsilon_k\right] \\ &= Y_0 \\ \Rightarrow \mathrm{Var}[Y_t] &= \mathrm{Var}\left[Y_0 + \sum_{k=1}^{t} \varepsilon_t\right] \\ &= \mathrm{Var}\left[\sum_{k=1}^{t} \varepsilon_t\right] \\ &= t\sigma^2 \end{aligned}$$

- One shock never disappears. It will last forever. So variance explodes.

- $\{Y_t\}$ is difference stationary since it can be made stationary by differencing it.

$$\begin{aligned} Y_t &= Y_{t-1} + \varepsilon_t \\ \Rightarrow Y_t - Y_{t-1} &= \Delta Y_t \\ &= \varepsilon_t, \quad \Rightarrow \text{stationary process} \end{aligned}$$

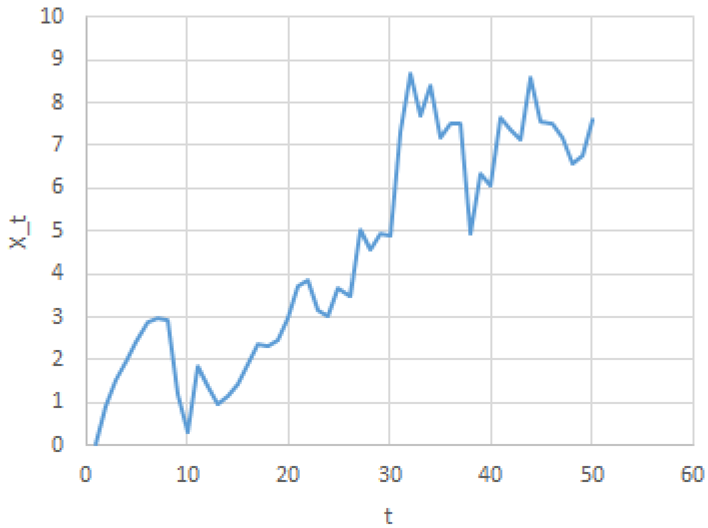


Figure 2: $X_t = X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$, $X_0 = 0$

2 ARMA modeling

- AR: Auto regression
- MA: Moving average
- ARMA(p,q) models

2.1 AR models

- Example: AR(1) model

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$$

- Example: AR(p) model

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$$

2.2 MA models

- Example: MA(1) model

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$$

- Example: MA(q) model

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$$

2.3 ARMA models

- Example: ARMA(1,1) model

$$Y_t = \phi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$$

- Example: ARMA(p,q) model

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$$

- Claim: Any stationary time-series process can be approximated by an ARMA(p,q) model. Where period p and q can be determined using different model selection criteria.

2.4 Lag operators

- Definition

$$\begin{array}{cc} LY_t = Y_{t-1} & L^{-1}Y_t = Y_{t+1} \\ \vdots & \vdots \\ L^p Y_t = Y_{t-p} & L^{-p} Y_t = Y_{t+p} \end{array}$$

- If $A(L)$ is lag polynomial of degree p

$$\begin{aligned} A(L) &= a_0 + a_1 L + \cdots + a_p L^p \\ A(L)Y_t &= \left(a_0 + a_1 L + \cdots + a_p L^p\right) Y_t \\ &= a_0 Y_t + a_1 Y_{t-1} + \cdots + a_p Y_{t-p} \end{aligned}$$

- Example: AR(1) model

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \varepsilon_t \\ &= \phi L Y_t + \varepsilon_t \\ \Rightarrow (1 - \phi L) Y_t &= \varepsilon_t \end{aligned}$$

- Example: MA(1) model

$$\begin{aligned} Y_t &= \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right) \\ &= \varepsilon_t + \theta L \varepsilon_t \\ &= (1 + \theta L)\varepsilon_t \end{aligned}$$

Announcement (01/23/17): Check following materials.

- Hamilton textbook
- Cochrane lecture notes
- Other documents posted on D2L

Econometric Methods II Lecture 02

Junyong Kim

January 25, 2017

Announcement (01/25/17)

- R & R Studio introduction session on Friday at 10 AM
- Check Eric Zivot's homepage

1 AR(1) model

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$$

AR(1) model is stationary if $|\phi| < 1$.

1.1 Properties of stationary model

$$\begin{aligned}
\mathrm{E}\left[Y_t\right] &= \mathrm{E}\left[Y_{t-1}\right] \\
\mathrm{Var}\left[Y_t\right] &= \mathrm{Var}\left[Y_{t-1}\right] \\
\Rightarrow \mathrm{E}\left[Y_t\right] &= \mathrm{E}\left[c + \phi Y_{t-1} + \varepsilon_t\right] \\
&= c + \phi \mathrm{E}\left[Y_t\right] \\
&= \frac{c}{1-\phi}, \quad \Rightarrow \text{unconditional mean} \\
\Rightarrow \mathrm{Var}\left[Y_t\right] &= \mathrm{Var}\left[c + \phi Y_{t-1} + \varepsilon_t\right] \\
&= \phi^2 \mathrm{Var}\left[Y_t\right] + \sigma^2 + \cancel{2\phi \mathrm{Cov}\left[Y_{t-1}, \varepsilon_t\right]} \\
&= \frac{\sigma^2}{1-\phi^2}, \quad \Rightarrow \text{unconditional variance} \\
&= \begin{cases} \sigma^2, & \text{as } |\phi| \to 0 \\ \infty, & \text{as } |\phi| \to 1 \end{cases}
\end{aligned}$$

Covariance in a time-series is sometimes referred to as autocovariance.

$$\mathrm{Cov}\left[Y_t, Y_{t-j}\right] = \gamma_j$$

For an AR(1) model

$$\mathrm{Cov}\left[Y_t, Y_{t-j}\right] = \mathrm{E}\left[\left(Y_t - \mu\right)\left(Y_{t-j} - \mu\right)\right]$$

Using recursive property of AR(1) model

$$
\begin{aligned}
\gamma_j &= \mathrm{Cov}\left[Y_t, Y_{t-j}\right] \\
&= \mathrm{E}\left[\underbrace{\left\{\phi\left(Y_{t-1}-\mu\right)+\varepsilon_t\right\}}_{=Y_t-\mu}\left(Y_{t-j}-\mu\right)\right]
\end{aligned}
$$

since $Y_t - \mu = \phi\left(Y_{t-1}-\mu\right)+\varepsilon_t$ (demeaned form of AR(1) model, can be shown).

$$
\begin{aligned}
\Rightarrow \gamma_j &= \mathrm{Cov}\left[Y_t, Y_{t-j}\right] \\
&= \mathrm{E}\left[\phi\left(Y_{t-1}-\mu\right)\left(Y_{t-j}-\mu\right)\right] + \underbrace{\mathrm{E}\left[\varepsilon_t\left(Y_{t-j}-\mu\right)\right]}_{=0} \\
&= \phi\gamma_{j-1} = \phi^2\gamma_{j-2} = \cdots = \phi^j\gamma_0 = \frac{\phi^j\sigma^2}{1-\phi^2}
\end{aligned}
$$

where $\gamma_0 = \mathrm{Var}\left[Y_t\right] = \frac{\sigma^2}{1-\phi^2}$. Remind the definition of correlation $\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X\sigma_Y}$.

$$
\begin{aligned}
\rho_j &= \mathrm{Corr}\left[Y_t, Y_{t-j}\right] \\
&= \frac{\mathrm{Cov}\left[Y_t, Y_{t-j}\right]}{\sqrt{\mathrm{Var}\left[Y_t\right]\mathrm{Var}\left[Y_{t-j}\right]}} \\
&= \frac{\gamma_j}{\sqrt{\gamma_0\gamma_0}} = \frac{\gamma_j}{\gamma_0} = \frac{\phi^j\cancel{\gamma_0}}{\cancel{\gamma_0}} = \phi^j
\end{aligned}
$$

$\phi^j \to 0$ as $j \to \infty$ if $|\phi| < 1$. In case of stationary time-series models, autocorrelation tends to die out as the time difference$\to \infty$.

1.2 Demeaned form of AR(1) model

$$
\begin{aligned}
Y_t &= c + \phi Y_{t-1} + \varepsilon_t \\
\Rightarrow Y_t - \frac{c}{1-\phi} &= c - \frac{c}{1-\phi} + \phi Y_{t-1} + \varepsilon_t \\
Y_t - \mu &= \frac{\cancel{c} - \phi c - \cancel{c}}{1-\phi} + \phi Y_{t-1} + \varepsilon_t \\
&= \phi\left(Y_{t-1} - \frac{c}{1-\phi}\right) + \varepsilon_t \\
&= \phi\left(Y_{t-1}-\mu\right) + \varepsilon_t \\
&= \phi^2\left(Y_{t-2}-\mu\right) + \phi\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
&= \phi^j\left(Y_{t-j}-\mu\right) + \phi^{j-1}\varepsilon_{t-j+1} + \cdots + \varepsilon_t
\end{aligned}
$$

2 Dynamic multiplier or impulse response function

$$\frac{\partial Y_t}{\partial \varepsilon_{t-j}} = \frac{\partial Y_{t+j}}{\partial \varepsilon_t} = \phi^j$$

Important: Any shock to a stationary time-series process disappears in the long-run. This property is called mean-reversion and the speed of mean-reversion depends on the magnitude of ϕ.

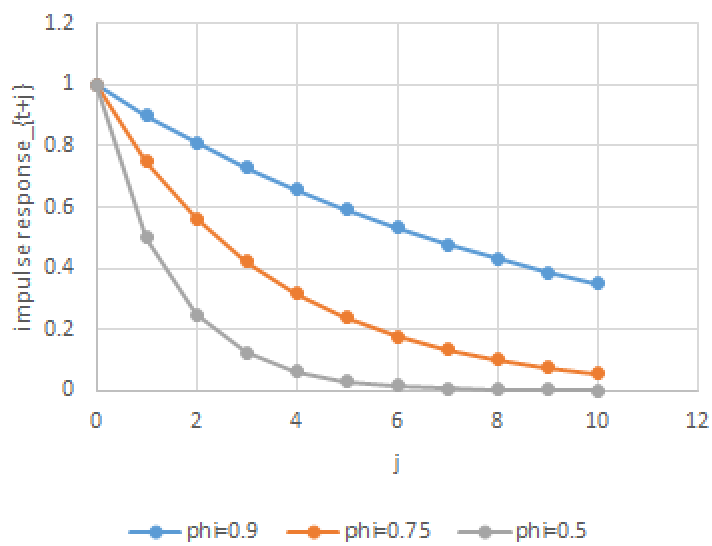


Figure 1: Impulse response functions for AR(1) models assuming $0 < \phi < 1$

2.1 Half-life

- Half-life is a measure of the speed of mean-reversion
- Half-life of AR(1): Lag at which IRF decreases to one-half
- $\rho_j = \phi^j = 0.5 \Leftrightarrow j \ln \phi = \ln 0.5 \Leftrightarrow j = \frac{\ln 0.5}{\ln \phi}$

ϕ	0.99	0.9	0.75	0.5	0.25
Half-life	68.97	6.58	2.41	1	0.5

3 Wold Decomposition theorem

Any stationary time-series process can be written down as MA(∞) process. If $\{Y_t\}$ is a stationary process, then Wold representation of $\{Y_t\}$ can be written as

$$\begin{aligned} Y_t &= \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \\ &= \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \end{aligned}$$

where $\psi_0 = 1$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

$$\begin{aligned} \mathrm{E}\left[Y_t\right] &= \mu \\ \mathrm{Var}\left[Y_t\right] &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 \\ \mathrm{Cov}\left[Y_t, Y_{t-j}\right] &= \mathrm{E}\left[\left(\mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots\right)\left(\mu + \psi_0 \varepsilon_{t-j} + \psi_1 \varepsilon_{t-j-1} + \cdots\right)\right] \\ &= \psi_j \psi_0 \mathrm{E}\left[\varepsilon_{t-j}^2\right] + \psi_{j+1} \psi_1 \mathrm{E}\left[\varepsilon_{t-j-1}^2\right] + \psi_{j+2} \psi_2 \mathrm{E}\left[\varepsilon_{t-j-2}^2\right] + \cdots \\ &= \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+j} \end{aligned}$$

4 Writing a stationary AR(1) model as MA(∞)

$$\begin{aligned} Y_t &= c + \phi Y_{t-1} + \varepsilon, \quad |\phi| < 1 \\ Y_t - \mu &= \phi\left(Y_{t-1} - \mu\right) + \varepsilon_t \\ &= \phi L\left(Y_t - \mu\right) + \varepsilon_t \\ &= \frac{\varepsilon_t}{1 - \phi L} \\ &= \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots \\ \Rightarrow Y_t &= \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \\ &= \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \end{aligned}$$

where $\psi_j = \phi^j$. Similarly, one can convert any stationary AR(p) model into a MA(∞) model. One can also show the relationship between AR & MA parameters as in the case of AR(1) model. Note that for AR(1) model

$$(1 - \phi L)\left(Y_t - \mu\right) = \varepsilon_t \Leftrightarrow Y_t - \mu = (1 - \phi L)^{-1} \varepsilon_t$$

Suppose one represent $(1 - \phi L) = \phi(L)$, which is a lag polynomial of degree 1. If we represent

$$\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots$$

For AR(1), $(Y_t - \mu) = (1 - \phi L)^{-1} \varepsilon_t = \phi(L)^{-1} \varepsilon_t = \psi(L) \varepsilon_t$, hence $Y_t = \mu + \phi(L)^{-1} \varepsilon_t = \mu + \psi(L) \varepsilon_t$ and $\phi(L)^{-1} = \psi(L)$.

4.1 AR(2) model

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \\
\Rightarrow \left(1 - \phi_1 L - \phi_2 L^2\right) Y_t &= \varepsilon_t \\
\Rightarrow \phi(L) Y_t &= \varepsilon_t \\
\Rightarrow Y_t &= \phi(L)^{-1}\varepsilon_t = \psi(L)\varepsilon_t
\end{aligned}
$$

Solve ψ_j in terms of ϕ_1 & ϕ_2. We know that $\psi(L) = \phi(L)^{-1} \Leftrightarrow \psi(L)\phi(L) = 1$.

$$
\begin{aligned}
\left(1 + \psi_1 L + \psi_2 L^2 + \cdots\right)\left(1 - \phi_1 L - \phi_2 L^2\right) &= 1 \\
\Rightarrow 1 + \left(\psi_1 - \phi_1\right) L + \left(\psi_2 - \psi_1\phi_1 - \phi_2\right) L^2 + \left(\psi_3 - \psi_2\phi_1 - \psi_1\phi_2\right) L^3 + \cdots &= 1
\end{aligned}
$$

By comparing coefficients,

$$
\begin{aligned}
\psi_1 &= \phi_1 \\
\psi_2 &= \psi_1\phi_1 + \phi_2 \\
\psi_3 &= \psi_2\phi_1 + \psi_1\phi_2 \\
\psi_4 &= \psi_3\phi_1 + \psi_2\phi_2 \\
&\vdots \\
\psi_j &= \psi_{j-1}\phi_1 + \psi_{j-2}\phi_2
\end{aligned}
$$

- Remember $\frac{\partial Y_t}{\partial \varepsilon_{t-j}} = \frac{\partial Y_{t+j}}{\partial \varepsilon_t} = \psi_j$
- One can easily obtain IRF by this recursive process

Econometric Methods II Lecture Note 03

Junyong Kim

January 30, 2017

1 Stationarity conditions for AR(p) models

- AR(1) model,
$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$$
- Stationarity: $|\phi| < 1$
- How about AR(p) models?
- Trick: Convert AR(p) model into a vector AR(1) model

1.1 State-space representation

- Write a p-th order stochastic difference equation as a first order vector stochastic difference equation
- Example: AR(2) model

$$\begin{aligned} & Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \\ \Rightarrow & \begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} \\ \Rightarrow & \mathbf{y}_t = \mathbf{F} \mathbf{y}_{t-1} + \boldsymbol{\xi}_t \end{aligned}$$

- Example: AR(p) model

$$\begin{aligned} & Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \\ \Rightarrow & \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ \Rightarrow & \mathbf{y}_t = \mathbf{F} \mathbf{y}_{t-1} + \boldsymbol{\xi}_t \end{aligned}$$

- So AR(p) model can be written as VAR(1) model

- Intuitively speaking, stationarity here is $\mathbf{F}^j \to \mathbf{0}$ as $j \to \infty$
- Stationarity condition for this model is that the eigenvalues of matrix $\mathbf{F}$ should be less than unity in absolute value
- Short note: $\mathbf{F}$ matrix can be decomposed as $\mathbf{F} = \mathbf{P\Lambda P}'$, $\mathbf{P}$ is a symmetric ($\mathbf{P} = \mathbf{P}'$) and idempotent ($\mathbf{P} = \mathbf{P}^2$) matrix
- If $\mathbf{F}$ is positive semi-definite, then $\mathbf{P}$ is lower-triangular, but here we do not impose this condition on $\mathbf{F}$
- Λ is a diagonal matrix with eigenvalues of $\mathbf{F}$ along the diagonals

$$\mathbf{F}^j = \left(\mathbf{P\Lambda P}'\right)^j = \mathbf{P}^j \mathbf{\Lambda}^j \left(\mathbf{P}'\right)^j = \mathbf{P\Lambda}^j\mathbf{P}'$$
$$\Rightarrow \mathbf{\Lambda}^j = \begin{pmatrix} \lambda_1^j & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p^j \end{pmatrix} \to \mathbf{0}, \quad \text{if} \quad |\lambda_i| < 1 \quad \forall i$$

- For stationarity, we need the eigenvalue of matrix $\mathbf{F}$ to be less than one in absolute value

1.2 Finding the eigenvalues

- λ is an eigenvalue of $\mathbf{F}$ and $\mathbf{x}$ is an eigenvector if $\mathbf{Fx} = \lambda\mathbf{x}$
- $(\mathbf{F} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \Rightarrow \det(\mathbf{F} - \lambda\mathbf{I}) = 0$, i.e. $\nexists(\mathbf{F} - \lambda\mathbf{I})^{-1}$
- AR(2) model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$
$$\Rightarrow \begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$
$$\Rightarrow \mathbf{F} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}$$
$$\Rightarrow \det \begin{pmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{pmatrix} = 0$$
$$\lambda^2 - \phi_1\lambda - \phi_2 =$$
$$\Rightarrow \lambda_i = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

- For stationarity, we need $|\lambda_i| < 1$ in case of real solution
- It is possible to get complex roots if $\sqrt{\phi_1^2 + 4\phi_2} < 0$
- In that case, λ_i can be written down as $\lambda_i = a + bi$ where $i^2 = -1$
- Complex roots induce periodic behavior in time-series

- In case of complex roots, stationarity condition is satisfied if (modulus) = $\sqrt{a^2+b^2} < 1$

- Example 1: Real roots

$$
\begin{aligned}
Y_t &= 0.6Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t, \quad |\phi_1 + \phi_2| < 1 \\
\Rightarrow \mathbf{F} &= \begin{pmatrix} 0.6 & 0.2 \\ 1 & 0 \end{pmatrix} \\
\Rightarrow \lambda^2 - 0.6\lambda - 0.2 &= 0 \\
\Rightarrow \lambda_i &= \frac{0.6 \pm \sqrt{1.16}}{2} \approx 0.8385 \quad \text{or} \quad -0.2385
\end{aligned}
$$

- Since $|0.8385|$ and $|-0.2385|$ are smaller than one, AR(2) model is stationary

- Example 2: Complex roots

$$
\begin{aligned}
Y_t &= 0.5Y_{t-1} - 0.8Y_{t-2} + \varepsilon_t, \quad |\phi_1 + \phi_2| = 0.3 < 1 \\
\Rightarrow \mathbf{F} &= \begin{pmatrix} 0.5 & -0.8 \\ 1 & 0 \end{pmatrix} \\
\Rightarrow \lambda^2 - 0.5\lambda + 0.8 &= 0 \\
\Rightarrow \lambda_i &= \frac{0.5 \pm \sqrt{-2.95}}{2} = \frac{0.5}{2} \pm \frac{\sqrt{2.95}}{2} i \\
\Rightarrow (\text{modulus}) &= \sqrt{\frac{0.25}{4} + \frac{2.95}{4}} = \sqrt{0.8} \approx 0.8944 < 1
\end{aligned}
$$

- Hence this model is stationary

1.3 Stationarity conditions on lag polynomial

- AR(2) model

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \\
\Rightarrow \left(1 - \phi_1 L - \phi_2 L^2\right) Y_t &= \varepsilon_t \\
\Rightarrow \Phi(L) &= 1 - \phi_1 L - \phi_2 L^2, \quad \text{lag polynomial of degree 2}
\end{aligned}
$$

- z is the root of this polynomial if

$$
1 - \phi_1 z - \phi_2 z^2 = 0
$$

- By the fundamental theorem of algebra,

$$
1 - \phi_1 z - \phi_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z) \Rightarrow z_1 = \frac{1}{\lambda_1}, \quad z_2 = \frac{1}{\lambda_2}
$$

  are the roots of the characteristic equation

- Result: The inverses of the roots of the characteristic equation

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0$$

are the eigenvalues of the companion matrix $\mathbf{F}$. Hence AR(p) model is stationary provided that the roots of $\Phi(z) = 0$ have modulus greater than unity

- Go back to Example 1

$$\begin{aligned} Y_t &= 0.6Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t \\ \Rightarrow \left(1 - 0.6L - 0.2L^2\right) Y_t &= \varepsilon_t \\ \Rightarrow \Phi(z) &= -0.2z^2 - 0.6z + 1 \\ \Rightarrow z_i &= -\frac{0.6 \pm \sqrt{0.36 + 0.8}}{0.4} \\ z_1 &= -\frac{0.6 + \sqrt{1.16}}{0.4}, \quad \frac{1}{z_1} \approx -0.2385 = \lambda_2 \\ z_2 &= -\frac{0.6 - \sqrt{1.16}}{0.4}, \quad \frac{1}{z_2} \approx 0.8385 = \lambda_1 \end{aligned}$$

- Hence z_1 and z_2 are the inverses of eigenvalues λ_1 and λ_2

1.4 Invertible MA model

- Digression: We know that any stationary AR model can be written down as MA(∞) process
- For example,

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \varepsilon_t \Leftrightarrow (1 - \phi L)Y_t = \varepsilon_t \Leftrightarrow Y_t = (1 - \phi L)^{-1}\varepsilon_t \\ \Rightarrow Y_t &= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots \\ &= \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots \\ &= \Psi(L)\varepsilon_t, \quad \text{where} \quad \psi_k = \phi^k \end{aligned}$$

- A corresponding theorem for MA model is that any invertible MA model can be written down as AR(∞) process
- MA(1) model: $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ is invertible if $|\theta| < 1$
- Claim: This invertible MA model can be written as AR(∞) process

$$\begin{aligned} Y_t &= (1 + \theta L)\varepsilon_t \\ \Rightarrow (1 + \theta L)^{-1}Y_t &= \varepsilon_t = Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} - \theta^3 Y_{t-3} + \cdots \\ \Rightarrow Y_t &= \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 Y_{t-3} - \cdots + \varepsilon_t \end{aligned}$$

which is AR(∞) representation of an invertible MA(1) process

2 Box–Jenkins method

- Box–Jenkins strategy for selecting ARMA models
- Question: If we observe time-series data, then how do we fit the most appropriate univariate ARMA(p,q) model to the data?
- There are few following steps involved
  1. Make the data stationary
  2. Plot autocorrelation function (ACF) & partial autocorrelation function (PACF) of the stationary time-series
  3. Choose a range of p and q for ARMA(p,q) models that can potentially fit the data
  4. Estimate all possible combinations of the models and get the AIC and BIC values for those models
  5. The model with the lowest AIC and BIC values is the best fit for the data in question

Announcement (01/30/17): Install the following R packages

- quantmod
- forecast
- tseries
- zoo

Econometric Methods II Lecture Note 04

Junyong Kim

February 1, 2017

Note: Eigendecomposition is $\mathbf{F} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$

1 ACF and PACF properties of ARMA models

- ACF: Autocorrelation function
- PACF: Partial autocorrelation function
- Autocorrelation: $\rho_j = \frac{\text{Cov}[Y_t, Y_{t-j}]}{\text{Var}[Y_t]}$
- For AR(1) model: For $Y_t = \phi Y_{t-1} + \varepsilon_t$

$$\rho_j = \frac{\phi^j \cancel{\gamma_0}}{\cancel{\gamma_0}} = \phi^j, \quad \text{where} \quad \gamma_0 = \text{Var}\left[Y_t\right] = \frac{\sigma^2}{1-\phi^2}$$

- Autocorrelation of a stationary AR(1) model declines exponentially as $j \to \infty$ since $|\phi| < 1$
- Similarly, it can be shown that for any stationary AR(p) model, autocorrelation declines exponentially

1.1 Autocorrelation property of MA models

- $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$

$$\begin{aligned}
\text{Cov}\left[Y_t, Y_{t-1}\right] &= \text{Cov}\left[\left(\varepsilon_t + \theta\varepsilon_{t-1}\right), \left(\varepsilon_{t-1} + \theta\varepsilon_{t-2}\right)\right] \\
&= \theta\text{Cov}\left[\varepsilon_{t-1}, \varepsilon_{t-1}\right] \\
&= \theta\sigma^2 \\
\text{Cov}\left[Y_t, Y_{t-2}\right] &= 0 \\
\Rightarrow \rho_1 &= \frac{\text{Cov}\left[Y_t, Y_{t-1}\right]}{\text{Var}\left[Y_t\right]} = \frac{\theta\cancel{\sigma^2}}{(1+\theta^2)\cancel{\sigma^2}} = \frac{\theta}{1+\theta^2} \\
\Rightarrow \rho_2 &= \frac{\text{Cov}\left[Y_t, Y_{t-2}\right]}{\text{Var}\left[Y_t\right]} = 0 \\
\Rightarrow \rho_j &= \text{Corr}\left[Y_t, Y_{t-j}\right] \in \begin{cases} \mathbb{R} \setminus \{0\}, & \text{for} \quad j \le p \\ \{0\}, & \text{for} \quad j = p \end{cases}
\end{aligned}$$

1.2 Partial autocorrelation function (PACF)

- Autocorrelation does not control for the effect of other lags when we calculate $\text{Corr}[Y_t, Y_{t-j}]$
- Partial autocorrelation does
- Therefore, partial autocorrelation is similar to multiple regression
- For example, AR(p) model

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

- The regression coefficients capture the partial autocorrelation at different lags

$$\begin{aligned} \text{PACF}[Y_t, Y_{t-j}] &= \phi_j, \quad \text{for AR}(p) \text{ model} \\ &\in \begin{cases} \mathbb{R} \setminus \{0\}, & \text{for} \quad j \leq p \\ \{0\}, & \text{for} \quad j = p \end{cases} \end{aligned}$$

1.3 Partial autocorrelation property of MA(p) models

- Since we know that any invertible MA process can be written down as AR(∞) process
- Partial autocorrelation (Y_t, Y_{t-j}) for invertible MA(p) models declines exponentially if $j \to \infty$

	AR(p)	MA(q)
ACF$[Y_t, Y_{t-j}]$	Declines exponentially for stationary AR(p) models	$\neq 0$ for $j \leq q$ $= 0$ for $j > q$
PACF$[Y_t, Y_{t-j}]$	$\neq 0$ for $j \leq p$ $= 0$ for $j > p$	Declines exponentially for invertible MA(q) models

1.4 Model selection criteria

- 2 Most widely used model selection criteria are
- AIC: Akaike information criterion
- BIC (or SBC): Bayesian information criterion or Schwarz Bayesian criterion
  - $\text{AIC}(p, q) = \ln\left(\hat{\sigma}^2\right) + \underbrace{\frac{2(p+q)}{T}}_{=\text{penalty factor}}$
  - $\text{BIC}(p, q) = \ln\left(\hat{\sigma}^2\right) + \underbrace{\frac{\ln T(p+q)}{T}}_{=\text{penalty factor}}$

- $\hat{\sigma}^2$ is an estimate of σ^2 from ARMA(p,q)
- Choose the model based on minimum value of AIC & BIC
  - The penalty factor for BIC is more stringent
  - BIC tends to choose models that are simpler
  - BIC: Consistent and picks the true model w.p. 1 as $T \to \infty$
  - AIC: Asymptotically efficient and chooses (asymptotically) the model with minimum MSE (mean squared error)

2 R session

- White noise simulation (Figure 1)
- AR(1) models simulation ($\phi \in \{0.995, 0.9, 0.75, 0.5\}$, Figures 2 and 3)
- Estimation of ARMA(p,q): What happens if AR(2) process is estimated with AR(1) model? (Figure 4)
- Why do we have to think about stationarity? Money making information, economic policy implications...
- Dollar–Euro exchange rate (Figure 5 for level, Figure 6 for difference)

2.1 Ljung–Box test

- Tests for serial correlation in the error term
- $\rho_1, \dots, \rho_j$ are the coefficients in a regression
- $e_t = \rho_1 e_{t-1} + \cdots + \rho_j e_{t-j} + \nu_t$, $\nu_t \sim$ iid $\left(0, \sigma^2\right)$

$$T \sum_{k=0}^{j} e_k^2 \sim \chi_j^2, \quad \text{Ljung–Box does a finite sample correction to this statistic}$$

Announcement (17/02/01): Check the assignments (deadline: next Wednesday)

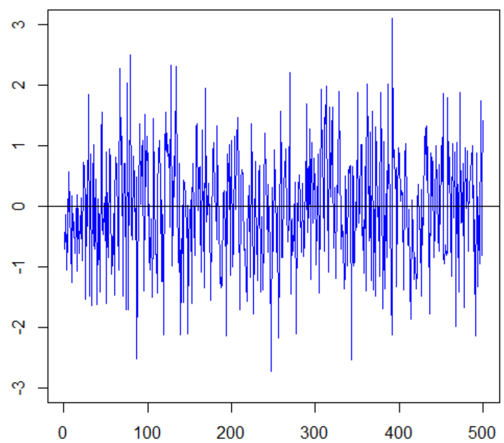


Figure 1: Simulated Standard Normal white noise process

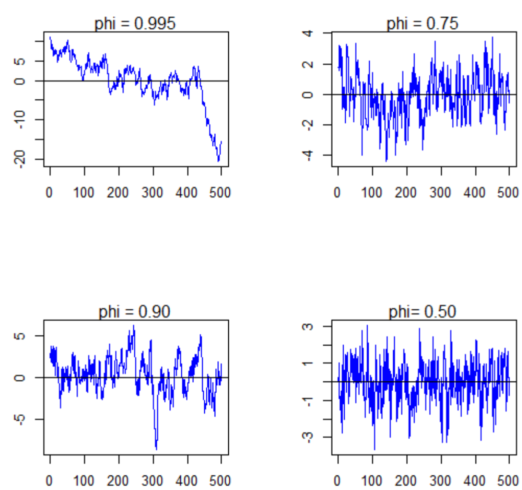


Figure 2: Simulated AR(1) processes

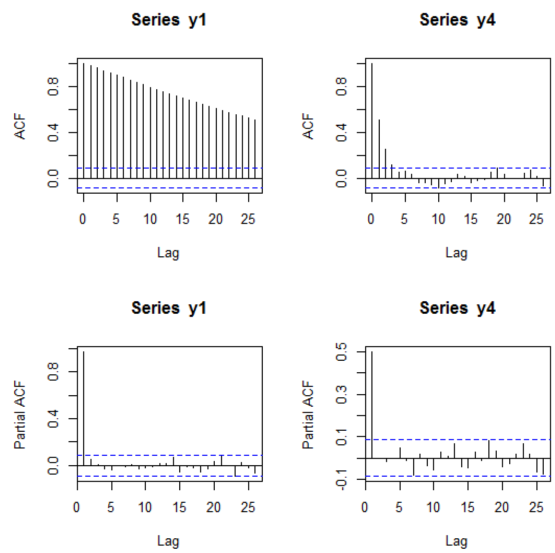


Figure 3: Estimated ACF and PACF of the processes

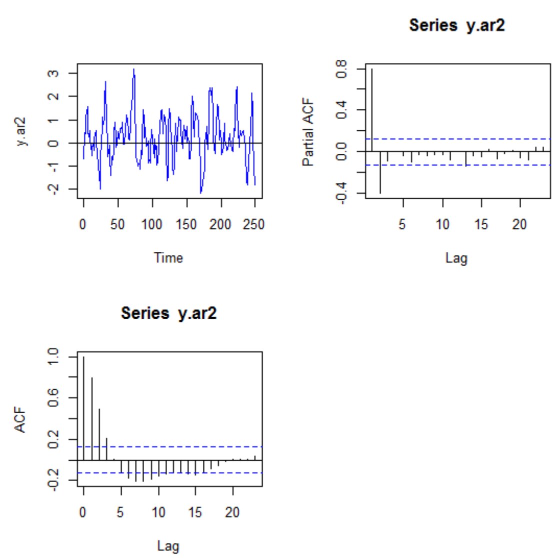


Figure 4: Simulated AR(2) process and Estimated ACF and PACF

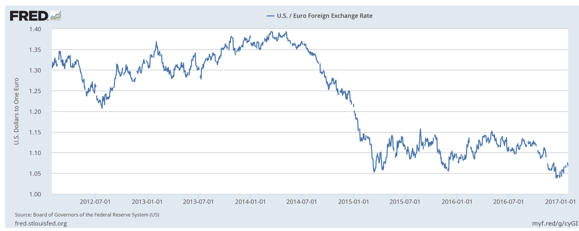


Figure 5: Dollar-to-Euro exchange rate: Is this stationary?

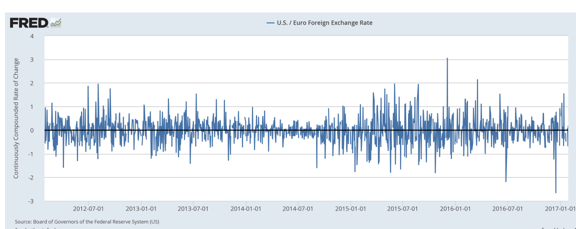


Figure 6: Log-differenced Dollar-to-Euro exchange rate: Is this stationary?

Econometric Methods II Lecture Note 05

Junyong Kim

February 6, 2017

1 MLE of ARMA models

For iid data, the joint density

$$f\left(y_1, \cdots, y_T; \theta\right) = \underbrace{\prod_{t=1}^{T} f\left(y_t; \theta\right)}_{\substack{\rightarrow\text{product} \\ \text{of marginals}}}$$

and θ is a set of parameters.

Then the log likelihood for iid data is

$$\ln L = \ln L\left(\theta | y_1, \cdots, y_T\right) = \sum_{t=1}^{T} \ln\left(y_t; \theta\right)$$

In a time-series context, this does not work because the data are not iid. One solution is to determine joint density $f\left(y_1, \cdots, y_T; \theta\right)$ directly by doing among other things, estimating $T \times T$ variance-covariance matrix.

Alternatively, use factorization of the joint density into a conditional density and marginal density.

$$\begin{aligned} f\left(y_2, y_1; \theta\right) &= f\left(y_2 | y_1\right) f\left(y_1; \theta\right) \\ f\left(y_3, y_2, y_1; \theta\right) &= f\left(y_3 | y_2\right) f\left(y_2, y_1; \theta\right) \\ &= f\left(y_3 | y_2\right) f\left(y_2 | y_1\right) f\left(y_1; \theta\right) \end{aligned}$$

In general, the factorization takes the following form.

$$f\left(y_T, \cdots, y_1; \theta\right) = \prod_{t=p+1}^{T} \left[f\left(y_t | I_{t-1}; \theta\right)\right] f\left(y_p, \cdots, y_1; \theta\right)$$

where $I_T = \left(y_1, \cdots, y_T\right)$ denotes the information at time T and $f\left(y_p, \cdots, y_1; \theta\right)$ denotes the initial values.

Log-likelihood function can be written as

$$\ln L(\theta|y) = \underbrace{\sum_{t=p+1}^{T} \ln f\left(y_t|I_{t-1};\theta\right)}_{\to\text{conditional likelihood}} + \underbrace{\ln\left(y_p,\cdots,y_1;\theta\right)}_{\substack{\to\text{marginal likelihood}\\ \text{for the initial values}}}$$

Two types of MLE estimates in this context.

$$\hat{\theta}_{\text{CMLE}} = \text{argmax} \sum_{t=p+1}^{T} \ln f\left(y_t|I_{t-1};\theta\right)$$

which maximizes the conditional likelihood (conditional MLE) and

$$\begin{aligned}\hat{\theta}_{\text{EMLE}} &= \text{Exact MLE of } \theta \\ &= \text{argmax} \sum_{t=p+1}^{T} \ln f\left(y_t|I_{t-1};\theta\right) + \ln f\left(y_p,\cdots,y_1;\theta\right)\end{aligned}$$

In stationary models, $\hat{\theta}_{\text{CMLE}}$ and $\hat{\theta}_{\text{EMLE}}$ are consistent and have the same limiting distribution.

In finite sample, however, they may differ by substantial amount and this difference gets amplified in the presence of highly persistent data.

The example here is an <u>AR(1) model</u>.

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\left(0,\sigma^2\right) \\ \text{E}\left[y_t|I_{t-1}\right] &= \text{E}\left[c + \phi y_{t-1} + \varepsilon_t|I_{t-1}\right] \\ &= \text{E}\left[c|I_{t-1}\right] + \phi\text{E}\left[y_{t-1}|I_{t-1}\right] + \text{E}\left[\varepsilon_t|I_{t-1}\right] \\ &= c + \phi y_{t-1} \\ \text{Var}\left[y_t|I_{t-1}\right] &= \text{E}\left[\left(y_t - \text{E}\left[y_t|I_{t-1}\right]\right)^2|I_{t-1}\right] \\ &= \text{E}\left[\varepsilon_t^2|I_{t-1}\right] = \sigma^2 \\ y_t|I_{t-1} &\sim \text{N}\left(c + \phi y_{t-1},\sigma^2\right)\end{aligned}$$

Therefore, we can construct the conditional log liklihood function.

$$\begin{aligned}\text{Conditional MLE} &= \sum_{t=p+1}^{T} \ln f\left(y_t|I_{t-1};\theta\right) \\ &= \sum_{t=2}^{T} \ln \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2\sigma^2}(y_t - c - \phi y_{t-1})^2} \\ &= -\frac{T-1}{2}\ln 2\pi - \frac{T-1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\left(y_t - c - \phi y_{t-1}\right)^2\end{aligned}$$

Note that $\hat{\theta}_{\text{CMLE}}$ that maximizes the conditional MLE function is the same as OLS estimate of θ from the following regression.

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right) \\ \theta &= \left(c, \phi, \sigma^2\right) \\ \hat{\theta}_{\text{OLS}} &= \hat{\theta}_{\text{CMLE}} \end{aligned}$$

For y_1, it is impossible to use the conditional log likelihood since y_0 is unavailable. Instead, one can use the unconditional mean $\frac{c}{1-\phi}$ and variance $\frac{\sigma^2}{1-\phi^2}$ for the marginal distribution.

Econometric Methods II Lecture Note 06

Junyong Kim

February 8, 2017

1 Log-likelihood for AR(1) model

For AR(1) model,

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2\right) \\ y_t | I_{t-1} &\sim \text{N}\left(c + \phi y_{t-1}, \sigma^2\right) \end{aligned}$$

We showed that the conditional log-likelihood for this AR(1) model.

$$\sum_{t-2}^{T} \ln f\left(y_t | I_{t-1}; \theta\right) = -\frac{T-1}{2} \ln 2\pi - \frac{T-1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^{T} \left(y_t - c - \phi y_{t-1}\right)^2$$

The conditional log-likelihood for this AR(1) model takes the form of log-likelihood of linear regression model with Normal errors. $\Rightarrow$ Conditional MLEs are equivalent to the OLS estimates of these parameters.

$$\text{Exact log-likelihood} = \text{Conditional log-likelihood} + \underbrace{\text{Marginal log-likelihood}}_{f(y_1;\theta)}$$

<u>Problem</u>: For y_1, we don't have any conditioning information to use. Therefore, we end up using unconditional distribution of y_1.

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{N}\left(0, \sigma^2\right) \\ \text{E}\left[y_t\right] &= \frac{c}{1-\phi} \\ \text{Var}\left[y_t\right] &= \frac{\sigma^2}{1-\phi^2} \\ \Rightarrow y_t &\sim \text{N}\left(\frac{c}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right) \end{aligned}$$

Therefore,

$$
\begin{aligned}
f(y_1;\theta) &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\phi^2}}} e^{-\frac{1}{2\frac{\sigma^2}{\phi^2}}\left(y_1-\frac{c}{1-\phi}\right)^2} \\
\underbrace{\ln f(\theta|y)}_{\substack{=\text{Marginal}\\ \text{log-likelihood}}} &= -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\frac{\sigma^2}{1-\phi^2} - \frac{1-\phi^2}{2\sigma^2}\left(y_1-\frac{c}{1-\phi}\right)^2
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\text{Exact log-likelihood} &= \text{Conditional log-likelihood} + \text{Marginal log-likelihood} \\
&= -\frac{T}{2}\ln 2\pi - \frac{1}{2}\ln\frac{\sigma^2}{1-\phi^2} - \frac{T-1}{2}\ln\sigma^2 \\
&\quad - \frac{1-\phi^2}{2\sigma^2}\left(y_1-\frac{c}{1-\phi}\right)^2 - \frac{1}{\sigma^2}\sum_{t=2}^{T}\left(y_t-c-\phi y_{t-1}\right)^2
\end{aligned}
$$

Since exact log-likelihood is a non-linear function of parameters θ, there is no closed form solution for exact MLE. Therefore, we need numerical solution for exact MLE.

If we have AR(2) model, i.e. $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$, then we have to start our log-likelihood from $t = 3$ instead of $t = 2$.

Hereafter we will study,

- Unit root
- Cointegration
- Kalman filter...

2 Unit root

Here we will talk about GDP. According to old-Kaynesian view, the output gap is determined by the difference between the actual and the potential outputs. They thought that almost all gaps are transitory and the gaps are demand shocks.

Friedman (1968, JER) thought this differently. He argued that not all shocks are demand shocks. According to his argument, in contrast, most of them are indeed permanent and supply shocks (pretty shocking at that time).

Nelson and Plosser: The gap defined above is wrong. There are permanent portions as well. If $y_t \sim I(1)$, i.e. integrated of order p, then it needs to be differenced p times to make it stationary. Their paper also states that most macroeconomic time-series also contain stochastic trend.

$$
\begin{aligned}
y_t &\sim I(1) \\
&= \tau_t + c_t, \quad \text{where } \tau \text{ is stochastic trend and } c \text{ is cycle} \\
\tau_t &= \tau_{t-1} + \nu_t, \quad \text{where } \nu_t \sim \text{iid}\left(0, \sigma_\nu^2\right) \\
c_t &= \phi_1 c_{t-1} + \phi_2 c_{t-2} + \varepsilon_t, \quad \text{where } \varepsilon_t \sim \text{iid}\left(0, \sigma_\varepsilon^2\right)
\end{aligned}
$$

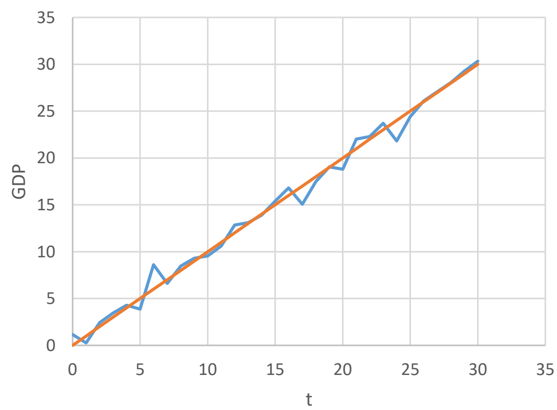


Figure 1: Simulated (old-Kaynesian style) GDP: $y_t = t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, 1)$

Roughly, The basic starting point is Dickey–Fuller test. By imposing other mean processes, one can use Augmented Dickey–Fuller test. Under the serial correlation assumption, Phillips–Perron test can be utilized. By involving other exogenous variables, KPSS test can be exploited.

If a time-series is highly persistent, then those tests will not be powerful enough. Instead, one can consider the methods proposed by Elliott, Rothensberg and Stock (1996) or Ng and Perron (2001). Under the presence of unit roots, standard asymptotic studies will not operate well, so one has to consider other asymptotics.

In R package, one can install ”urca” package to analyze the unit roots (ur.df, ur.pp, ur.ers, ur.kpss, . . .).

Econometric Methods II Lecture Note 07

Junyong Kim

February 13, 2017

1 R session

- "urca" package: Unit root and cointegration tests
- "ur.df": DF and ADF tests
- "ur.pp": Phillips–Perron test
- "ur.ers": ERS test (greater power than those of prior tests)

2 Multivariate time series models

$$Y_t = \begin{pmatrix} y_{1t} & \cdots & y_{nt} \end{pmatrix}^{\mathrm{t}}$$

For $\{Y_t\}$ to be stationary, we need to capture lead & lag cross covariances & impose restrictions

$$\begin{aligned}
\mathrm{E}\left[Y_t\right] &= \mu = \begin{pmatrix} \mu_1 & \cdots & \mu_n \end{pmatrix}^{\mathrm{t}} \\
\mathrm{Var}\left[Y_t\right] &= \mathrm{E}\left[\left(Y_t - \mu\right)\left(Y_t - \mu\right)^{\mathrm{t}}\right] \\
&= \Gamma_0 \\
\Gamma_0 &= \begin{pmatrix} \mathrm{Var}\left[y_{1t}\right] & \cdots & \mathrm{Cov}\left[y_{1t}, y_{nt}\right] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}\left[y_{nt}, y_{1t}\right] & \cdots & \mathrm{Var}\left[y_{nt}\right] \end{pmatrix} \\
\gamma_{jj}^0 &= \mathrm{Var}\left[y_{jt}\right] \\
\gamma_{ij}^0 &= \mathrm{Cov}\left[y_{it}, y_{jt}\right] \\
\mathrm{Corr}\left[Y_t\right] &= \begin{pmatrix} 1 & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & 1 \end{pmatrix} \\
&= D^{-1}\Gamma_0\left(D^{-1}\right)^{\mathrm{t}}
\end{aligned}$$

$$
\begin{aligned}
D &= \begin{pmatrix} \sqrt{\gamma_{11}^0} & & \\ & \ddots & \\ & & \sqrt{\gamma_{nn}^0} \end{pmatrix} \\
\gamma_{jj}^0 &= \mathrm{Var}\left[y_{jt}\right] \\
\gamma_{jj}^k &= \mathrm{Cov}\left[y_{jt}, y_{jt-k}\right] \\
\gamma_{ij}^k &= \gamma_{ji}^{-k} = \mathrm{Cov}\left[y_{it}, y_{jt-k}\right], \quad \text{cross-lag covariances} \\
\gamma_{ji}^k &= \gamma_{ij}^{-k} = \mathrm{Cov}\left[y_{it-k}, y_{jt}\right]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\gamma_{ij}^k &\neq 0, \quad \text{variable } j \text{ leads variable } i \\
\gamma_{ji}^k &\neq 0, \quad \text{variable } i \text{ leads variable } j
\end{aligned}
$$

so if both $\gamma_{ij}^k \neq 0$ & $\gamma_{ji}^k \neq 0$, then there exists a feedback effect between i & j.

$$
\begin{aligned}
\rho_{ij}^k &= \frac{\gamma_{ij}^k}{\sqrt{\gamma_{ii}^0 \gamma_{jj}^0}} \\
\Gamma_k &= \mathrm{E}\left[\left(Y_t - \mu\right)\left(Y_{t-k} - \mu\right)^{\mathrm{t}}\right] \\
&= \begin{pmatrix} \mathrm{Cov}\left[y_{1t}, y_{1t-k}\right] & \cdots & \mathrm{Cov}\left[y_{1t}, y_{nt-k}\right] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}\left[y_{nt}, y_{1t-k}\right] & \cdots & \mathrm{Cov}\left[y_{nt}, y_{nt-k}\right] \end{pmatrix}
\end{aligned}
$$

Γ_k is not symmetric, i.e. $\Gamma_k \neq \Gamma_k^{\mathrm{t}}$. However, $\Gamma_{-k} = \Gamma_k^{\mathrm{t}}$ since $\gamma_{ij}^k = \gamma_{ji}^{-k}$.

3 Multivariate Wold decomposition

$$Y_t = \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \cdots + \Psi_k \varepsilon_{t-k} + \cdots$$

Here, $\varepsilon_t \sim \mathrm{WN}(0, \Omega)$

Example: Bivariate model

$$
\begin{aligned}
Y_t &= \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} \\
y_{1t} &= \mu_1 + \varepsilon_{1t} + \psi_{11}^1 \varepsilon_{1t-1} + \psi_{12}^1 \varepsilon_{2t-1} + \psi_{11}^2 \varepsilon_{1t-2} + \psi_{12}^2 \varepsilon_{2t-2} + \cdots \\
y_{2t} &= \mu_2 + \varepsilon_{2t} + \psi_{21}^1 \varepsilon_{1t-1} + \psi_{22}^1 \varepsilon_{2t-1} + \psi_{21}^2 \varepsilon_{1t-2} + \psi_{22}^2 \varepsilon_{2t-2} + \cdots
\end{aligned}
$$

Is it possible to get impulse reponses from this wold representation? Or can we calculate $\frac{\partial y_{1t}}{\partial \varepsilon_{1t-k}}$ or $\frac{\partial y_{1t}}{\partial \varepsilon_{2t-k}}$? The answer is yes if $\Omega = \mathrm{E}\left[\varepsilon_t \varepsilon_t^{\mathrm{t}}\right] = \mathrm{Var}\left[\varepsilon_t\right]$ is a diagonal matrix. However, the answer is no if Ω is a non-diagonal matrix, which is usually the case.

If Ω is non-diagonal, then contemporaneous shocks ε_{1t} & ε_{2t} are correlated with each other and therefore, it is impossible to identify the effect of individual shocks.

4 Assumptions in Wold decomposition

4.1 1-summability

$$\sum_{k=0}^{\infty} \left|\psi_{ij}^{k}\right| \cdot k < \infty, \quad \text{for } i = 1, \cdots, n \text{ and } j = 1, \cdots, n$$

This $\Rightarrow$ elements of $\bar{\psi}_k \to 0$ as $k \to \infty$.

For stationarity $\{Y_t\}$ Central Limit theorem applies

$$\sqrt{T}\left(\bar{Y} - \mu\right) \to \mathrm{N}(0, \mathrm{LRV})$$

Here LRV $\Rightarrow$ long-run variance (sum of variance and covariances).

In a univariate setting,

$$\begin{aligned} y_t &= \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \\ \mathrm{Var}\left[y_t\right] &= \sigma^2 \left(1 + \psi_1^2 + \psi_2^2 + \cdots\right) \\ \mathrm{LRV} &= \sigma^2 \left(1 + \psi_1 + \psi_2 + \cdots\right)^2 \neq \mathrm{Var}\left[y_t\right] \end{aligned}$$

It can be shown that

$$\begin{aligned} \psi(L) &= \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots \\ \psi(1) &= \psi_0 + \psi_1 + \psi_2 + \cdots \\ &= 1 + \psi_1 + \psi_2 + \cdots \\ \Rightarrow \ \mathrm{LRV} &= \sigma^2 \psi(1)^2 \end{aligned}$$

Note that

$$\begin{aligned} \psi(L) &= \phi(L)^{-1} \theta(L) \\ \psi(1) &= \phi(1)^{-1} \theta(1) \\ &= \frac{1 + \theta_1 + \theta_2 + \cdots + \theta_q}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \end{aligned}$$

In a multivariate setting,

$$\begin{aligned} \mathrm{LRV} &= \sum_{k=-\infty}^{\infty} \Gamma_k \\ &= \Gamma_0 + \sum_{k=1}^{\infty} \left(\Gamma_k + \Gamma_k^{\mathrm{t}}\right) \\ &= \Psi(1) \Omega \Psi(1)^{\mathrm{t}} \end{aligned}$$

Note that for ARMA(p,q) models

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\
\left(1 - \phi_1 L - \cdots - \phi_p L^p\right) y_t &= \left(1 + \theta_1 L + \cdots + \theta_q L^q\right) \varepsilon_t \\
\phi(L) y_t &= \theta(L) \varepsilon_t \\
\text{where } \phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \\
\theta(L) &= 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q \\
\Rightarrow y_t &= \phi(L)^{-1} \theta(L) \varepsilon_t \\
&= \psi(L) \varepsilon_t, \quad \text{Wold representation} \\
\Rightarrow \psi(1) &= \phi(1)^{-1} \theta(1) = \frac{1 + \theta_1 + \theta_2 + \cdots + \theta_q}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}
\end{aligned}
$$

Econometric Methods II Lecture Note 08

Junyong Kim

February 15, 2017

1 Vector autoregression (VAR) models

$$
\begin{aligned}
\underset{n\times 1}{Y_t} &= \underset{n\times 1}{C} + \underset{n\times n}{\Phi_1}\underset{n\times 1}{Y_{t-1}} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t \\
\varepsilon_t &\sim \mathrm{N}(0, \Omega) \\
\mathrm{E}\left[\varepsilon_t \varepsilon_t^{\mathrm{t}}\right] &= \Omega
\end{aligned}
$$

Example: Bivariate VAR(2) model is

$$
\begin{aligned}
y_{1t} &= c_1 + \phi_{11}^1 y_{1t-1} + \phi_{12}^1 y_{2t-1} + \phi_{11}^2 y_{1t-2} + \phi_{12}^2 y_{2t-2} + \varepsilon_{1t} \\
y_{2t} &= c_2 + \phi_{21}^1 y_{1t-1} + \phi_{22}^1 y_{2t-1} + \phi_{21}^2 y_{1t-2} + \phi_{22}^2 y_{2t-2} + \varepsilon_{2t} \\
\phi_{ij}^p &= p\text{th lag of } j\text{th variable in } i\text{th equation.}
\end{aligned}
$$

In a n-variable VAR(p) model, the number of parameters to be estimated is $n + pn^2 + \frac{n(n+1)}{2}$. For $n = 5$ and $p = 3$, total number of parameters is $5 + 75 + 15 = 95$. This is known as *the curse of dimensionality problem* in a classical VAR model.

The first equation in a n-variable VAR(p) system can be written as

$$
\begin{aligned}
y_{1t} = c_1 &+ \phi_{11}^1 y_{1t-1} + \phi_{12}^1 y_{2t-1} + \cdots + \phi_{1n}^1 y_{nt-1} + \cdots \\
&+ \phi_{11}^p y_{1t-p} + \phi_{12}^p y_{2t-p} + \cdots + \phi_{1n}^p y_{nt-p} + \varepsilon_{1t}.
\end{aligned}
$$

If Y_t is $I(0)$ then we can write $\{Y_t\}$ as an infinite MA process that is

$$
\begin{aligned}
\Phi(L) Y_t &= \Theta(L)\varepsilon_t, \quad \text{ignoring constant} \\
Y_t &= \Phi(L)^{-1}\Theta(L)\varepsilon_t \\
&= \Psi(L)\varepsilon_t \\
\text{where } \Psi(L) &= \Phi(L)^{-1}\Theta(L) \\
\Phi(L)^{-1} &= \left(I - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p\right)^{-1}.
\end{aligned}
$$

Note that we need $[I - (\Phi_1 + \Phi_2 + \cdots + \Phi_p)]$ and

$$
\Theta(L) = \left(I + \Theta_1 L + \Theta_2 L^2 + \cdots + \Theta_q L^q\right).
$$

2 Stationarity conditions for a VAR models

Represent VAR(p) as VAR(1) using state-space representation as

$$\begin{pmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ \vdots \\ Y_{t-p+1} - \mu \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \cdots & \Phi_p \\ I & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} - \mu \\ Y_{t-2} - \mu \\ Y_{t-3} - \mu \\ \vdots \\ Y_{t-p} - \mu \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$
$$\Rightarrow Z_t = F Z_{t-1} + V_t.$$

VAR(1) representation is $Z_t = F Z_{t-1} + V_t$. This VAR(p) is stationary if the eigenvalues of matrix F have modulus less than unity. Eigenvalues of F are inverses of the roots of $I - \Phi_1 X - \Phi_2 X^2 - \cdots - \Phi_p X^p = 0$.

3 Estimation of VAR models

$$\begin{aligned} Y_t &= C + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t \\ &= \begin{pmatrix} C & \Phi_1 & \cdots & \Phi_p \end{pmatrix} \begin{pmatrix} 1 & Y_{t-1}^{\mathrm{t}} & \cdots & Y_{t-p}^{\mathrm{t}} \end{pmatrix}^{\mathrm{t}} + \varepsilon_t \\ &= \Pi^{\mathrm{t}} X_t + \varepsilon_t \\ \text{where } \Pi^{\mathrm{t}} &= \begin{pmatrix} \Pi_1 & \Pi_2 & \cdots & \Pi_n \end{pmatrix}^{\mathrm{t}} \\ \text{where } \Pi_1 &= \begin{pmatrix} c_1 & \phi_{11}^1 & \phi_{12}^1 & \cdots & \phi_{1n}^p \end{pmatrix}^{\mathrm{t}}. \end{aligned}$$

Π^{t} is $n \times k$ where $k = np + 1$.

4 SUR representation

Seemingly unrelated regression is

$$\begin{aligned} Y_1 &= X\Pi_1 + \varepsilon_1, \quad \text{where } Y_1 = \begin{pmatrix} y_{11} & \cdots & y_{1t} \end{pmatrix}^{\mathrm{t}} \\ Y_2 &= X\Pi_2 + \varepsilon_2 \\ \Rightarrow Y_n &= X\Pi_n + \varepsilon_n \\ \operatorname{Cov}\left[\varepsilon_{it}, \varepsilon_{jt}\right] &= \sigma_{ij} \end{aligned}$$

i.e. error terms across different equations are contemporaneously correlated.

Since all the equations in the VAR system have same explanatory variables on the right hand side, SUR estimation of the VAR system is equivalent of estimating these equations separately by OLS,

$$\operatorname{vec}\left(\hat{\Pi}\right) = \begin{pmatrix}\hat{\Pi}_1 & \hat{\Pi}_2 & \cdots & \hat{\Pi}_n\end{pmatrix} \to \mathrm{N}\left(\operatorname{vec}(\Pi), \Omega \otimes \left(X^{\mathrm{t}} X\right)^{-1}\right)$$

where $\Omega = \mathrm{E}\left[\varepsilon_t \varepsilon_t^{\mathrm{t}}\right]$

$$\hat{\Omega} = \frac{1}{T-k} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t^{\mathrm{t}}.$$

If you are interested in these mathematics, see Lütkepohl.

5 Testing cross-equation restriction in a VAR setting

Basically, we utilize Wald test that tests

$$\begin{aligned} \mathrm{H}_0 &: \underset{a \times nk}{R} \cdot \underset{nk \times 1}{\operatorname{vec}(\Pi)} = r \\ \mathrm{H}_1 &: R \cdot \operatorname{vec}(\Pi) \neq r \end{aligned}$$

where R is $a \times nk$, r is $a \times 1$, respectively, and

$$\text{Wald} = \left(R \cdot \operatorname{vec}(\Pi) - r\right)^{\mathrm{t}} \left[R \cdot \hat{\operatorname{Var}}\left[\operatorname{vec}\left(\hat{\Pi}\right)\right] \cdot R^{\mathrm{t}}\right]^{-1} \left(R \cdot \operatorname{vec}(\Pi) - r\right) \sim \chi_a^2.$$

Example: Bivariate VAR(3) model is

$$\begin{aligned} y_{1t} &= c_1 + \phi_{11}^1 y_{1t-1} + \phi_{12}^1 y_{2t-1} + \cdots + \phi_{11}^3 y_{1t-3} + \phi_{12}^3 y_{2t-3} + \varepsilon_{1t} \\ y_{2t} &= c_2 + \phi_{21}^1 y_{1t-1} + \phi_{22}^1 y_{2t-1} + \cdots + \phi_{21}^3 y_{1t-3} + \phi_{22}^3 y_{2t-3} + \varepsilon_{2t}. \end{aligned}$$

Examine whether the third lag is required in both equations by

$$\begin{aligned} \mathrm{H}_0 &: \phi_{11}^3 = \phi_{12}^3 = \phi_{21}^3 = \phi_{22}^3 = 0 \\ \mathrm{H}_1 &: \text{Not } \mathrm{H}_0. \end{aligned}$$

Then,

$$\begin{aligned} \Pi &= \begin{pmatrix}\Pi_1 & \Pi_2\end{pmatrix} \\ &= \begin{pmatrix} c_1 & \phi_{11}^1 & \phi_{12}^1 & \phi_{11}^2 & \phi_{12}^2 & \phi_{11}^3 & \phi_{12}^3 \\ c_2 & \phi_{21}^1 & \phi_{22}^1 & \phi_{21}^2 & \phi_{22}^2 & \phi_{21}^3 & \phi_{22}^3 \end{pmatrix}^{\mathrm{t}} \\ \operatorname{vec}(\Pi) &= \begin{pmatrix}\Pi_1^{\mathrm{t}} & \Pi_2^{\mathrm{t}}\end{pmatrix}^{\mathrm{t}} \\ &= \begin{pmatrix} c_1 & \phi_{11}^1 & \cdots & \phi_{12}^3 & c_2 & \phi_{21}^1 & \cdots & \phi_{22}^3 \end{pmatrix}^{\mathrm{t}}. \end{aligned}$$

So in this design, $r = \begin{pmatrix}0 & 0 & 0 & 0\end{pmatrix}^{\mathrm{t}}$ and

$$R = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

6 Granger causality

Granger causality examines the predictive ability of one variable, say y_{1t} for the other variable y_{2t}.
Example: Bivariate VAR(1)

$$
\begin{aligned}
y_{1t} &= c_1 + \phi_{11}^1 y_{1t-1} + \phi_{12}^1 y_{2t-1} + \varepsilon_{1t} \\
y_{2t} &= c_2 + \phi_{21}^1 y_{1t-1} + \phi_{22}^1 y_{2t-1} + \varepsilon_{2t}.
\end{aligned}
$$

We say that in this VAR(1) framework, y_{2t} Granger causes y_{1t} if $\phi_{12}^1 \neq 0$. The test is

$$
\begin{aligned}
&\mathrm{H}_0: \ y_{2t} \text{ does not Granger cause } y_{1t}, \text{ i.e. } \phi_{12}^1 = 0 \\
&\mathrm{H}_1: \ y_{2t} \text{ does Granger cause } y_{1t}, \text{ i.e. } \phi_{12}^1 \neq 0.
\end{aligned}
$$

Similarly, for VAR(2) model,

$$
\begin{aligned}
&y_{1t} = c_1 + \phi_{11}^1 y_{1t-1} + \phi_{12}^1 y_{2t-1} + \phi_{11}^2 y_{1t-2} + \phi_{12}^2 y_{2t-2} + \varepsilon_{1t} \\
&y_{2t} = c_2 + \phi_{21}^1 y_{1t-1} + \phi_{22}^1 y_{2t-1} + \phi_{21}^2 y_{1t-2} + \phi_{22}^2 y_{2t-2} + \varepsilon_{2t} \\
&\mathrm{H}_0: \ y_{2t} \text{ does not Granger cause } y_{1t}, \text{ i.e. } \phi_{12}^1 = \phi_{12}^2 = 0 \\
&\mathrm{H}_1: \ y_{2t} \text{ does Granger cause } y_{1t}, \text{ i.e. } \phi_{12}^1 \neq 0 \text{ or } \phi_{12}^2 \neq 0.
\end{aligned}
$$

Note that Granger causality does not imply causality. It only provides information about predictive power of a variable.

Econometric Methods II Lecture Note 09

Junyong Kim

February 20, 2017

Reference: Gary Koop and Dimitris Korobilis, *Bayesian multivariate time series methods for empirical macroeconomics.*
Package: Check "BMR".

1 Impulse response function

A reduced form VAR(p)

$$Y_t = c + \Pi_1 Y_{t-1} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t$$

has Wold representation of the form

$$Y_t = \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \cdots$$

where Ψ_s is $n \times n$.

Then,

$$\frac{\partial y_{it}}{\partial \varepsilon_{jt-s}} = \frac{\partial y_{it+s}}{\partial \varepsilon_{jt}} = \psi_{ij}^s,$$

i.e. the (i,j)th element of Ψ_s.

We can't say that ψ_{ij}^s represents the IRF for the impact of shock j on variable i since ε_t's may be correlated with each other, i.e. $\mathrm{E}\left[\varepsilon_t \varepsilon_t^{\mathrm{t}}\right]$ is a non-diagonal matrix.

Therefore, a shock to ε_{jt} is not independent of the shocks to other variable.

Then, how do we estimate impulse response functions if the reduced form errors are correlated with each other?

Solution: Use some restrictions to get uncorrelated shocks. These uncorrelated shocks are called structural shocks.

Structural VAR literature (to be studied later) deals with the identification of these structural shocks.

One simple way to get uncorrelated shocks across different equations is to use original Sims (1980) recursive causal ordering approach (default VAR model in most canned statistical packages).

Place the most exogenous variable as the first variable in the VAR system.

1.1 n-variable VAR(p) recursive causal ordering

Again,

$$
\begin{aligned}
y_{1t} &= c_1 + \gamma_{11}^1 y_{1t-1} + \cdots + \gamma_{1n}^1 y_{nt-1} + \cdots + \gamma_{1n}^p y_{nt-p} + \eta_{1t} \\
y_{2t} &= c_2 + \gamma_{21}^1 y_{1t-1} + \cdots + \gamma_{2n}^1 y_{nt-1} + \cdots + \gamma_{2n}^p y_{nt-p} + \eta_{2t} \\
y_{3t} &= c_3 + \gamma_{31}^1 y_{1t-1} + \cdots + \gamma_{3n}^1 y_{nt-1} + \cdots + \gamma_{3n}^p y_{nt-p} + \eta_{3t} \\
&\cdots \\
y_{nt} &= c_n + \gamma_{n1}^1 y_{1t-1} + \cdots + \gamma_{nn}^1 y_{nt-1} + \cdots + \gamma_{nn}^p y_{nt-p} + \eta_{nt}.
\end{aligned}
$$

This can be represented as

$$
BY_t = \Gamma_0 + \Gamma_1 Y_{t-1} + \Gamma_2 Y_{t-2} + \cdots + \Gamma_p Y_{t-p} + \eta_t
$$

where

$$
B = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
-\beta_{21} & 1 & 0 & \cdots & 0 \\
-\beta_{31} & -\beta_{32} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-\beta_{n1} & -\beta_{n2} & -\beta_{n3} & \cdots & 1
\end{pmatrix}
$$

is a lower triangular matrix with 1's along the diagonal. $\mathrm{E}\left[\eta_t \eta_t^{\mathrm{t}}\right] = \Omega$ is a diagonal matrix. Elements of η_t are uncorrelated with each other and therefore can be used to estimate impulse response.
$y_1 \to y_2 \to \cdots \to y_n$: Helps us in the identification of strucutral errors. The Wold representation can be written down in terms of structural errors η_t and

$$
\begin{aligned}
Y_t &= \mu + \Theta_0 \eta_t + \Theta_1 \eta_{t-1} + \Theta_2 \eta_{t-2} + \cdots \\
\frac{\partial y_{it+s}}{\partial \eta_{jt}} &= \frac{\partial y_{it}}{\partial \eta_{jt-s}} = \theta_{ij}^s,
\end{aligned}
$$

which is IRF.

In practice, these orthogonalized structural errors can be directly estimated from the reduced form VAR

$$
Y_t = c + \Pi_1 Y_{t-1} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t
$$

where $\mathrm{E}\left[\varepsilon_t \varepsilon_t^{\mathrm{t}}\right] = \Sigma$ is a non-diagonal matrix.

Decompose: Cholesky decomposition

$$
\begin{aligned}
\Sigma &= ADA^{\mathrm{t}} \\
A &= \text{Invertible lower triangular matrix with 1s along the diagonal} \\
B &= \text{Diagonal matrix with positive diagonal elements}
\end{aligned}
$$

and ε_t is a reduced form error and η_t is a structural error, respectively.

Suppose $\eta_t = A^{-1}\varepsilon_t$. Then,

$$\begin{aligned}
\mathrm{E}\left[\eta_t\eta_t^{\mathrm{t}}\right] &= A^{-1}\mathrm{E}\left[\varepsilon_t\varepsilon_t^{\mathrm{t}}\right]\left(A^{-1}\right)^{\mathrm{t}} \\
&= A^{-1}\Sigma\left(A^{-1}\right)^{\mathrm{t}} \\
&= A^{-1}ADA^{\mathrm{t}}\left(A^{-1}\right)^{\mathrm{t}} \\
&= D,
\end{aligned}$$

which is a diagonal matrix, i.e. structural errors are uncorrelated with each other. We can convert reduced form Wold representation into a structural Wold representation.

1.2 Reduced form Wold representation

$$\begin{aligned}
Y_t &= \mu + \varepsilon_t + \Psi_1\varepsilon_{t-1} + \cdots \\
&= \mu + AA^{-1}\varepsilon_t + \Psi_1 AA^{-1}\varepsilon_{t-2} + \cdots \\
&= \mu + \Theta_0\eta_t + \Theta_1\eta_{t-1} + \Theta_2\eta_{t-2} + \cdots \\
\text{where } \Theta_j &= \Psi_j A \\
\eta_t &= A^{-1}\varepsilon_t.
\end{aligned}$$

1.3 Structural Wold representation

$$Y_t = \mu + \Theta_0\eta_t + \Theta_1\eta_{t-1} + \cdots$$

from this representation, we can calculate IRFs and forecast error variance decomposition (FEVD).

2 Forecast error variance decomposition

FEVD shows the proportion of forecast error variance that can be explained by a particular shock, i.e.

$$\begin{aligned}
Y_t &= \mu + \Theta_0\eta_t + \Theta_1\eta_{t-1} + \cdots \\
Y_{t+h} &= \mu + \Theta_0\eta_{t+h} + \Theta_1\eta_{t+h-1} + \cdots \\
Y_{t+h} - Y_{t+h|t} &\to \varepsilon_{t+h|t}, \quad n\text{-period ahead forecast error} \\
Y_{t+h|t} &= \text{Forecast of } Y_{t+h} \text{ at time } t \\
&= \mathrm{E}\left[Y_{t+h|t}\right] \\
Y_{t+h|t} &= \mu + \Theta_h\eta_t + \Theta_{h+1}\eta_{t-1} + \cdots \\
Y_{t+h} - Y_{t+h|t} &= \Theta_0\eta_{t+h} + \Theta_1\eta_{t+h-1} + \cdots + \Theta_{h-1}\eta_{t+1} \\
&= \sum_{s=0}^{h-1}\Theta_s\eta_{t+h-s},
\end{aligned}$$

or for ith variable,

$$y_{it+h} - y_{it+h|t} = \sum_{s=0}^{h-1} \theta_{i1}^{s} \eta_{1t+h-s} + \cdots + \sum_{s=0}^{h-1} \theta_{in}^{s} \eta_{nt+h-s}.$$

Note that η_is are uncorrelated with others.

$$\mathrm{Var}\left[y_{it+h} - y_{it+h|t}\right] = \sigma_{\eta_1}^2 \sum_{s=0}^{h-1} (\theta_{i1}^{s})^2 + \cdots + \sigma_{\eta_n}^2 \sum_{s=0}^{h-1} (\theta_{in}^{s})^2,$$

which is n-period ahead forecast error variance of y_{it}.

The portion of Var $\left[y_{it+h} - y_{it+h|t}\right]$ due to shock η_j is

$$\mathrm{FEVD}_{ij}(h) = \frac{\sigma_{\eta_j}^2 \sum_{s=0}^{h-1} \left(\theta_{ij}^{s}\right)^2}{\sigma_{\eta_1}^2 \sum_{s=0}^{h-1} \left(\theta_{i1}^{s}\right)^2 + \cdots + \sigma_{\eta_n}^2 \sum_{s=0}^{h-1} \left(\theta_{in}^{s}\right)^2}$$

Econometric Methods II Lecture Note 10

Junyong Kim

February 22, 2017

1 Forecasting in VAR models

$$Y_t = \Pi_1 Y_{t-1} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t$$

Convert this VAR(p) model into a VAR(1) model using state-space representation, i.e.

$$\begin{aligned}
\begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix} &= \begin{pmatrix} \Pi_1 & \Pi_2 & \cdots & \Pi_p \\ I & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\
Z_t &= F Z_{t-1} + \nu_t.
\end{aligned}$$

Then, 1-steap ahead forecast is

$$Z_{t+1|t} = \mathrm{E}\left[Z_t | \Omega_t\right]$$

where Ω_t is information set at time t, hence

$$\begin{aligned}
Z_{t+1|t} &= F Z_t \\
Z_{t+2} &= F Z_{t+1} + \nu_{t+2} \\
Z_{t+2|t} &= F Z_{t+1|t} \\
&= F^2 Z_t, \quad \text{2-step ahead forecast} \\
Z_{t+k|t} &= F^k Z_t.
\end{aligned}$$

This is called chain rule of forecasting.

How do we evaluate forecasts? It depends on loss function. The conventional approach uses symmetric, quadratic loss, i.e. choose the model based on lower MSE (mean squared error) or RMSE (root mean squared error), i.e.

$$\mathrm{MSE} = \mathrm{E}\left[\varepsilon^2_{t+h|t}\right]$$

where $\varepsilon_{t+h} = Z_{t+h} - Z_{t+h|t}$.

2 Forecast evaluation tests

Idea: To test whether MSE from one forecast is significantly different than the MSE from another forecast.

2.1 2 types of forecast evaluation tests

1. Non-nested forecast comparison test
2. Nested forecast comparison test

We have several tests in both types.

2.1.1 Non-nested forecast comparison test

Diebold-Mariano-West (DMW) is the most popular non-nested forecast comparison test.
Digression: Nested model

- Model 1: $y_t = \phi y_{t-1} + \varepsilon_t$, i.e. AR(1)
- Model 2: $y_t = \phi y_{t-1} + \gamma z_{t-1} + \nu_t$

Here, Model 1 is a special case of Model 2 if $\gamma = 0$. In this case, Model 2 *nests* Model 1.
Non-nested model

- Model 1: $y_t = \phi y_{t-1} + \varepsilon_t$
- Model 2: $y_t = \gamma z_{t-1} + \nu_t$

Model 1 & Model 2 are non-nested.

Diebold-Mariano-West (DMW) Test: Suppose $y^A_{t+h|t}$ is h-period ahead forecast from Model A and $y^B_{t+h|t}$ is h-period ahead forecast from Model B and

$$\begin{aligned} \varepsilon^A_{t+h|t} &= y_{t+h} - y^A_{t+h|t} \\ \varepsilon^B_{t+h|t} &= y_{t+h} - y^B_{t+h|t}. \\ \text{Then, } \mathrm{H}_0 &: \mathrm{MSE}^A = \mathrm{MSE}^B \\ \mathrm{H}_1 &: \mathrm{MSE}^A < \mathrm{MSE}^B. \end{aligned}$$

1-tailed test since we already observed the estimated MSEs.

DMW test involves following steps, i.e.

1. Create a variable.

$$\text{DMW} = \left(\varepsilon^B_{t+h|t}\right)^2 - \left(\varepsilon^A_{t+h|t}\right)^2$$

2. Run a regression of DMW on a constant & test for the significance of constant.

To take into account serial correlation in the error term, Newey-West HAC errors are used.

Note: The forecast error from an optimal n-period forecast is MA$(n-1)$.

The standard errors need to be corrected for serial correlation for $\underline{h > 1}$, i.e. when forecast horizon is greater than one,

$$\begin{aligned}
y_t &= \phi y_{t-1} + \varepsilon_t \\
y_{t+1|t} &= \phi y_t \\
\varepsilon_{t+1|t} &= y_{t+1} - y_{t+1|t} \\
&= \varepsilon_t \\
y_{t+2|t} &= \phi^2 y_t \\
y_{t+2} &= \phi y_{t+1} + \varepsilon_{t+2} \\
&= \phi\left(\phi y_t + \varepsilon_{t+1}\right) + \varepsilon_{t+2} \\
&= \phi^2 y_t + \phi\varepsilon_{t+1} + \varepsilon_{t+2} \\
\varepsilon_{t+2|t} &= y_{t+2} - y_{t+2|t} \\
&= \phi\varepsilon_{t+1} + \varepsilon_{t+2}, \quad \text{MA}(1)
\end{aligned}$$

2.1.2 Nested forecast comparison test

This test usually examines whether inclusion of a variable in a model significantly improves the forecasting performance of the new model.

For example, if the benchmark model is AR(1), i.e. $y_t = \phi y_{t-1} + \varepsilon_t$ (Model B) & we are interested in examining whether lagged values of z_t has predictive ability for y_t beyond what is already contained in a lagged value of y_t, then the model becomes $y_t = \phi y_{t-1} + \gamma z_{t-1} + \nu_t$ (Model A).

If inclusison of z_t significantly improves forecasting performance then,

$$\text{MSE}^A < \text{MSE}^B,$$

i.e. nested forecast comparison test can be used in this case,

$$\begin{aligned}
\text{H}_0 &: \text{MSE}^A = \text{MSE}^B \\
\text{H}_1 &: \text{MSE}^A < \text{MSE}^B.
\end{aligned}$$

Clark-West test is one of the nested forecast comparison tests (just simple modification of DMW test). Define

$$\text{CW} = \left(\left(\varepsilon^B_{t+h|t}\right)^2 - \left(\varepsilon^A_{t+h|t}\right)^2\right) + \underbrace{\left(\varepsilon^B_{t+h|t} - \varepsilon^A_{t+h|t}\right)^2}_{\text{extra term}}$$

where the extra term is a compensation for adding an extra variable.

Run a regression of CW on a constant and test the significance of the constant.

3 Bayesian approach

So called Bayesian inference is

$$\begin{aligned} y &= X\beta + \varepsilon, \quad \varepsilon \sim \mathrm{N}\left(0, \sigma^2\right) \\ \hat{\beta} &= \left(X^{\mathrm{t}} X\right)^{-1} X^{\mathrm{t}} y \\ \hat{\sigma}^2 &= \frac{\hat{e}^{\mathrm{t}}\hat{e}}{T-k}, \quad \hat{e} = y - X\hat{\beta} \\ \hat{\beta} &\sim \mathrm{N}\left(\beta, \sigma^2 \left(X^{\mathrm{t}} X\right)^{-1}\right) \\ (T-k)\frac{\hat{\sigma}^2}{\sigma^2} &\sim \chi^2_{T-k}. \end{aligned}$$

These estimators are usually evaluated in terms of unbiasedness & consistency.

In Bayesian framework, $\theta = \begin{pmatrix}\beta & \sigma^2\end{pmatrix}^{\mathrm{t}}$ is a random variable having a probability distribution, and $g(\theta)$ is a prior distribution of θ without observing the data on X & y. $h(\theta, y)$ is a joint distribution of data and parameters, hence

$$\underbrace{h(\theta, y)}_{\text{joint}} = \underbrace{f(y|\theta)}_{\text{likelihood}} \underbrace{g(\theta)}_{\text{prior}} = \underbrace{p(\theta|y)}_{\text{posterior}} \underbrace{f(y)}_{\text{marginal}},$$

i.e.

$$\begin{aligned} p(\theta|y) &= \frac{f(y|\theta)g(\theta)}{f(y)} \\ p(\theta|y) &\propto f(y|\theta)g(\theta) \\ \text{where } f(y|\theta) &\rightarrow \text{likelihood function.} \end{aligned}$$

To get analytical solution of the posterior distribution, natural conjugate priors are used.

3.1 Natural conjugate prior

When natural conjugate prior is combined with the likelihood function, the posterior density has the same form as prior density.

Econometric Methods II Lecture Note 11

Junyong Kim

February 27, 2017

1 Bayesian VAR

$$\begin{aligned}\underbrace{p(\theta|y)}_{\text{posterior}} &= \frac{f(y|\theta)g(\theta)}{f(y)} \\ &\propto \underbrace{f(y|\theta)}_{\text{likelihood}}\underbrace{g(\theta)}_{\text{prior}}.\end{aligned}$$

Example 1: $y = X\beta + \varepsilon$, $\varepsilon \sim \text{iid}\left(0, \sigma^2\right)$.

Bayesian inference on β when σ^2 is known (where β is k-dimensional vector).

$$\begin{aligned}\beta|\sigma^2 &\sim \mathrm{N}\left(\beta_0, \Sigma_0\right) \\ g(\beta|\sigma^2) &= \underbrace{(2\pi)^{-k/2}\left|\Sigma_0\right|^{-1}}_{\text{constant}} e^{-\frac{1}{2}(\beta-\beta_0)^{\mathrm{t}}\Sigma_0^{-1}(\beta-\beta_0)} \\ &\propto e^{-\frac{1}{2}(\beta-\beta_0)^{\mathrm{t}}\Sigma_0^{-1}(\beta-\beta_0)} \\ f(y|\theta) &= \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2\sigma^2}(y-X\beta)^{\mathrm{t}}(y-X\beta)} \\ &\propto e^{-\frac{1}{2\sigma^2}(y-X\beta)^{\mathrm{t}}(y-X\beta)},\end{aligned}$$

and the posterior is

$$\begin{aligned}p(\beta|\sigma^2, y) &= f(y|\beta, \sigma^2)g(\beta) \\ &\propto e^{-\frac{1}{\sigma^2}(y-X\beta)^{\mathrm{t}}(y-X\beta)} e^{\frac{1}{2}(\beta-\beta_0)^{\mathrm{t}}\Sigma_0^{-1}(\beta-\beta_0)} \\ \beta|\sigma^2, y &\sim \mathrm{N}\left(\beta_1, \Sigma_1\right) \\ \beta_1 &= \left(\Sigma_0^{-1} + \sigma^2 X^{\mathrm{t}}X\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + \sigma^2 X^{\mathrm{t}}y\right) \\ \Sigma_1 &= \left(\Sigma_0^{-1} + \sigma^2 X^{\mathrm{t}}X\right)^{-1},\end{aligned}$$

i.e. β_1 is a weighted average of the prior mean & sample mean. The weights are the precision of these two measures.

Warning: In this note, there are a lot of notational distortions. So be careful in interpreting these equations.

Now consider VAR(p) model, i.e.

$$\begin{aligned} y_t &= \Pi_0 + \sum_{j=1}^{p} \Pi_j y_{t-j} + \varepsilon_t \\ \Pi &= \begin{pmatrix} \Pi_0 & \Pi_1 & \cdots & \Pi_p \end{pmatrix}^{\mathrm{t}}. \end{aligned}$$

In matrix notation, where $x_t = (1, y_{t-1}, \cdots, y_{t-p})^{\mathrm{t}}$,

$$\begin{aligned} y_t &= X\Pi + \varepsilon, \quad X = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}^{\mathrm{t}} \\ \alpha &= \mathrm{vec}(\Pi) \\ y_t &= (I \otimes X)\alpha + \varepsilon \\ \varepsilon &\sim \mathrm{N}(0, \Sigma \otimes I). \end{aligned}$$

If $\hat{\alpha}$ is the estimate of α,

$$\begin{aligned} \hat{\alpha}|\Sigma &\sim \mathrm{N}\left(\alpha, \Sigma \otimes \left(X^{\mathrm{t}} X\right)^{-1}\right) \\ \hat{\Sigma}^{-1}|y &\sim \mathrm{W}\left(S^{-1}, \nu\right) \\ S &= \left(y - X\hat{\Pi}\right)\left(y - X\hat{\Pi}\right), \end{aligned}$$

where W is Wishart distribution, which is multivariate counterpart of chi-square distribution, i.e. the sum of squares of n draws from a multivariate Normal distribution.

Suppose $y = X\pi + \varepsilon$ is the VAR model. Different approaches to estimate this VAR model are

1. Minnesota prior
2. Natural conjugate
3. Independent Normal-Wishart
4. Other methods

1 and 2 provide us analytical solutions, but for 3 and 4, analytical solutions are not feasible.

2 Minnesota prior

Replace Σ with an estimate $\hat{\Sigma}$. The original Minnesota prior assumed that Σ is a diagonal matrix. When Σ is not diagonal, an estimate of $\Sigma = \frac{S}{T}$ is used, i.e. we only need to worry about

$$\alpha|\Sigma \sim \mathrm{N}\left(\underline{\alpha}_{\mathrm{Min}}, \underline{V}_{\mathrm{Min}}\right),$$

and the corresponding posterior distribution is,

$$
\begin{aligned}
\alpha|y &\sim \mathrm{N}\left(\bar{\alpha}_{\text{Min}}, \bar{V}_{\text{Min}}\right) \\
\bar{V}_{\text{Min}} &= \left[\underline{V}_{\text{Min}}^{-1} + \left(\hat{\Sigma}^{-1} \otimes \left(X^{\text{t}} X\right)\right)\right]^{-1}.
\end{aligned}
$$

The problem with Minnesota prior: Exactly Bayesian, assumes Σ as give given.

3 Natural conjugate priors

Prior, posteriors & likelihood come from the same family of distribution, i.e.

$$
\begin{aligned}
\alpha|\Sigma &\sim \mathrm{N}\left(\underline{\alpha}, \Sigma \otimes V\right) \\
\Sigma^{-1} &\sim \mathrm{W}\left(\underline{S}^{-1}, \underline{\nu}\right).
\end{aligned}
$$

The posterior takes the form

$$
\begin{aligned}
\alpha|\Sigma, y &\sim \mathrm{N}\left(\bar{\alpha}, \Sigma \otimes \bar{V}\right) \\
\Sigma|y &\sim \mathrm{W}\left(\bar{S}^{-1}, \bar{\nu}\right).
\end{aligned}
$$

and see the formula in Koop & Korobilis for $\bar{\alpha}, \bar{V}, \bar{S}^{-1}$ & $\bar{\nu}$.

Two problems with natural conjugate prior approach are

1. Every equation must have the same set of explanatory variables.
2. Prior covariance matrix $\Sigma \otimes V$, i.e. prior covariance of the coefficients in two equations must be proportional to one another.

4 Independent Normal-Wishart prior

In natural conjugate prior, prior for α is based on Σ and hence they are not independent. But here,

$$
p\left(\beta, \Sigma^{-1}\right) = p(\beta) p\left(\Sigma^{-1}\right),
$$

for independent Normal-Wishart prior, i.e.

$$
\begin{aligned}
\beta &\sim \mathrm{N}\left(\underline{\beta}, \underline{V}_{\beta}\right) \\
\Sigma^{-1} &\sim \mathrm{W}\left(\underline{S}^{-1}, \underline{\ nu}\right).
\end{aligned}
$$

The joint posterior $p\left(\beta, \Sigma^{-1} \middle| y\right)$ does not have an analytical solution. However, the conditional posterior distribution $p\left(\beta \middle| \Sigma^{-1}, y\right)$ & $p\left(\Sigma^{-1} \middle| \beta, y\right)$ have convenient forms, i.e.

$$
\begin{aligned}
\beta|y, \Sigma^{-1} &\sim \mathrm{N}\left(\bar{\beta}, \bar{V}_{\beta}\right) \\
\Sigma^{-1}|\beta, y &\sim \mathrm{W}\left(\bar{S}^{-1}, \bar{\nu}\right).
\end{aligned}
$$

Announcement (02/27/17): Check Kishor and Koenig (2014, JMCB).

Econometric Methods II Lecture Note 12

Junyong Kim

March 1, 2017

	Pros	Cons
Minnesota prior, etc.	$\exists$Analytical solution	Unrealistic
Independent prior, etc.	More realistic	No analytical solution

1 Gibbs sampling from a Normal distribution

$$\begin{aligned}
\theta = \begin{pmatrix}\theta_1 & \theta_2\end{pmatrix}^{\mathrm{t}} &\sim \mathrm{N}_2(0, \Sigma) \\
\Sigma &= \begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix} \\
\theta_1|\theta_2 &\sim \mathrm{N}\left(\rho\theta_2, 1-\rho^2\right) \\
\theta_2|\theta_1 &\sim \mathrm{N}\left(\rho\theta_1, 1-\rho^2\right).
\end{aligned}$$

Gibbs sampling is a Markov chain Monte Carlo (MCMC) algorithm. Let θ_i^j be jth draw of θ_i. Then the iteration would be,

θ_1	θ_2
$\theta_1^1\|\theta_2^0 \sim \mathrm{N}\left(\rho\theta_2^0, 1-\rho^2\right)$	$\theta_2^1\|\theta_1^1 \sim \mathrm{N}\left(\rho\theta_1^1, 1-\rho^2\right)$
$\theta_1^2\|\theta_2^1 \sim \mathrm{N}\left(\rho\theta_2^1, 1-\rho^2\right)$	$\theta_2^2\|\theta_1^2 \sim \mathrm{N}\left(\rho\theta_1^2, 1-\rho^2\right)$
$\theta_1^3\|\theta_2^2 \sim \mathrm{N}\left(\rho\theta_2^2, 1-\rho^2\right)$	$\theta_2^3\|\theta_1^3 \sim \mathrm{N}\left(\rho\theta_1^3, 1-\rho^2\right)$
$\vdots$	$\vdots$

Why Markov chain? Because the distribution of random draw only depends on current status.
Convergence: Important issue (will not cover). Burning periods.

2 R session

We covered

- Lag selection, VAR estimation, Orthogonal impulse response
- Forecast error variance decomposition

- Granger causality test
- Forecast and plotting

```
# find optimal number of lags
info.crit=VARselect(var_1,lag.max=4,type="const")
info.crit
```

```
y=read.csv("jobs_var.csv",header=TRUE)
jobs.growth=y[,2] # Jobs growth
slo=y[,3]#LOAN OFFICER SURVEY
spread=y[,4] #Spread data
var_1=ts(cbind(jobs.growth,slo),start=1985,frequency=4)
```

```
# estimate VAR(1)
model0=VAR(var_1,p=1,type="const")
summary(model0)
```

```
###IRF
z=irf(model0, impulse="slo", response="jobs.growth",
    n.ahead = 20, ortho = TRUE, cumulative = FALSE,
    boot = TRUE, ci = 0.95, runs = 100)
plot(z)
```

Here, one utilizes Bootstrap with 100 times of resampling to find the confidence interval.

```
###Forecast error variance decomposition
fevd.model0=fevd(model0,n.ahead=20)
layout(matrix(1:2,ncol=2))
plot(fevd.model0,plot.type ="single", col=1:2)
```

At 20 quarters ahead forecast horizon 54.2% forecast error variance of jobs grows can be explained by shocks to jobs growth and 43.7% can be explained by shocks to SLO.

```
# 5 - test for Granger causality
causality(model0,cause="slo")$Granger
causality(model0,cause="jobs.growth")$Granger
```

```
# multi-step predictions
model1=VAR(var_1[1:116,],p=1,type="const")
fore0=predict(model1,n.ahead=7)
jobsgrowth.fore= ts(fore0$fcst$jobs.growth[,1:3],
    start=c(2014,1),frequency=4)
slo.fore=ts(fore0$fcst$slo[,1:3],start=c(2014,1),
    frequency=4)
```

```
# plot predictions and confidence interval
layout(1)
plot(var_1[,1],col="black",xlim=c(2014,2015),
    ylim=c(-1,2.5))
lines(jobsgrowth.fore[,1],col="blue")
lines(jobsgrowth.fore[,2],col="green")
lines(jobsgrowth.fore[,3],col="green")
```

Econometric Methods II Lecture Note 13

Junyong Kim

March 6, 2017

In the last class, we studied n-ahead forecasts. The procedure for recursive out-of-sample forecasting is

1. Estimate the VAR model from time $1, \cdots, t$
2. Using the estimated parameters make a prediction for $t+1, \cdots, t+p$
3. Add one more observation & esetimate the model for $t \in \{1, \cdots, t+1\}$ & make prediction for $t+2, \cdots, t+p+1$ and so on

Look-ahead bias: If one computes forecasts (for instance, t) using the parameters estimated from the sample that is unavailable at that time (for instance, T), then the forecasts will contain the future information (for instance, $t+1, \cdots, T$).
Pseudo out-of-sample forecasting: We can mitigate this problem by just using the past information ($1, \cdots, t$) that is available at that time to forecast the future ($t+1, \cdots, T$).

1 R session

We studied

- How to generate one-step-ahead forecasts
- How to compare RMSEs from two different models (non-nested: DMW, nested: CW)
- the result of Kishor and Koenig (2014, JMCB) paper
- the package "BMR" and how to use Bayesian VAR

Econometric Methods II Lecture Note 14

Junyong Kim

March 8, 2017

1 Structural VAR

Consider the following structural VAR model.

$$
\begin{aligned}
y_{1t} =& \gamma_{10} - b_{12} y_{2t} + \gamma_{11} y_{1t-1} + \gamma_{12} y_{2t-1} + \eta_{1t} \\
y_{2t} =& \gamma_{20} - b_{21} y_{1t} + \gamma_{21} y_{1t-1} + \gamma_{22} y_{2t-1} + \eta_{2t}
\end{aligned}
$$

and where η_{1t} and η_{2t} are structural shocks with

$$
\underbrace{\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}}_{\eta_t} \sim \text{iid} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \underbrace{\begin{pmatrix} \sigma_{\eta_1}^2 & 0 \\ 0 & \sigma_{\eta_2}^2 \end{pmatrix}}_{D} \right)
$$

where $D = \mathrm{E}[\eta_t \eta_t^{\mathrm{t}}]$ is a diagonal matrix. In general, $\mathrm{Cov}[y_{1t}, \eta_{2t}] = 0$ and $\mathrm{Cov}[y_{2t}, \eta_{1t}] = 0$. All variables are endogenous & hence OLS is not appropriate. In matrix form,

$$
\begin{aligned}
\begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \\
BY_t =& \gamma_0 + \Gamma_1 Y_{t-1} + \eta_t.
\end{aligned}
$$

Where $B(L) Y_t = \gamma_0 + \eta_t$ and where $B(L) = B - \Gamma_1 L$. This is the reduced form of the above equation.

$$
\begin{aligned}
Y_t =& B^{-1}\gamma_0 + B^{-1}\Gamma_1 Y_{t-1} + B^{-1}\eta_t \\
=& a_0 + A_1 Y_{t-1} + \varepsilon_t
\end{aligned}
$$

where $a_0 = B^{-1}\gamma_0$, $A_1 = B^{-1}\Gamma_1$ and $\varepsilon_t = B^{-1}\eta_t$.

This is the reduced form VAR ($Y_t = a_0 + A_1 Y_{t-1} + \varepsilon_t$). In practice, this reduced form can be estimated by OLS.

Problem: Shocks (ε_t) do not have structural interpretation. Instead, shocks (ε_t) are contemporaneously correlated.

$$\begin{aligned}
\varepsilon_t =& B^{-1}\eta_t \\
B =& \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix} \\
\Rightarrow B^{-1} =& \frac{1}{\Delta}\begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix}
\end{aligned}$$

where $\Delta = \det(B) = 1 - b_{12}b_{21}$.

ε_t is a linear combination of η_{1t} & η_{2t}. Reduced form errors are linear .

$$\begin{aligned}
\mathrm{E}\left[\varepsilon_t \varepsilon_t^{\mathrm{t}}\right] =& \mathrm{E}\left[B^{-1}\eta_t\eta_t^{\mathrm{t}}\left(B^{-1}\right)^{\mathrm{t}}\right] \\
=& B^{-1}\mathrm{E}\left[\eta_t\eta_t^{\mathrm{t}}\right]\left(B^{-1}\right)^{\mathrm{t}} \\
=& B^{-1}D\left(B^{-1}\right)^{\mathrm{t}} \\
=& \Omega
\end{aligned}$$

and Ω is the variance-covariance matrix of reduced form errors and this is non-diagonal.

Structural VAR: $BY_t = \gamma_0 + \Gamma_1 Y_{t-1} + \eta_t$
Reduced form VAR: $Y_t = a_0 + A_1 Y_{t-1} + \varepsilon_t$
Question: Can we recover coefficients in the structural VAR model by the estimated reduced form VAR model?
Answer: No, unless some identification restriction is imposed. In the 2 variables case for instance,

Number of reduced form VAR parameters	9
Number of structural VAR parameters	10

and hence we need one restriction to identify the structural VAR.

Under recursive identification scheme, we imposed $b_{12} = 0$. In practice there are different methods to identify structural VAR. Most of these methods involve imposing some kind of short-run or long-run restrictions and $b_{12} = 0$ is a short-run restriction.

Examples of long-run restriction

1. Money is neutral in the long-run. So shock to money supply has no impact on real economic activity

2. Demand shock does not have permanent effect on output

2 MA representation

One can start from the reduced form VAR.

$$\begin{aligned} Y_t =& a_0 + A_1 Y_{t-1} + \varepsilon_t \\ (I - A_1 L)\, Y_t =& a_0 + \varepsilon_t \\ A(L) Y_t =& \\ Y_t =& A(L)^{-1} a_0 + A(L)^{-1} \varepsilon_t \\ =& \mu + \underbrace{\Psi(L)\varepsilon_t}_{\substack{\text{reduced form moving} \\ \text{average representation}}} \end{aligned}$$

where $\mu = A(1)^{-1} a_0$.

2.1 Structural MA representation

$$\begin{aligned} Y_t =& \mu + \Psi_0 \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \cdots \\ =& \mu + \Psi_0 B^{-1} \eta_t + \Psi_1 B^{-1} \eta_{t-1} + \Psi_2 B^{-1} \eta_{t-2} + \cdots \\ =& \mu + \Theta_0 \eta_t + \Theta_1 \eta_{t-1} + \Theta_2 \eta_{t-2} + \cdots \end{aligned}$$

where $\Theta_j = \Psi_j B^{-1}$.

Note that $\Psi_0 = I$ and hence $\Theta_0 = B^{-1}$ and this captures the initial impact of structural shocks.

For bivariate case,

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \theta_{11}^0 & \theta_{12}^0 \\ \theta_{21}^0 & \theta_{22}^0 \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} + \begin{pmatrix} \theta_{11}^1 & \theta_{12}^1 \\ \theta_{21}^1 & \theta_{22}^1 \end{pmatrix} \begin{pmatrix} \eta_{1t-1} \\ \eta_{2t-1} \end{pmatrix} + \cdots \\ + \begin{pmatrix} \theta_{11}^s & \theta_{12}^s \\ \theta_{21}^s & \theta_{22}^s \end{pmatrix} \begin{pmatrix} \eta_{1t-s} \\ \eta_{2t-s} \end{pmatrix} + \cdots . \end{aligned}$$

Elements of Θ_k matrix, θ_{ij}^k represents the impulse responses (the response of ith variable on the impulse on jth variable after k periods).

$$\frac{\partial y_{1t+s}}{\partial \eta_{1t}} = \theta_{11}^s, \quad \frac{\partial y_{1t+s}}{\partial \eta_{2t}} = \theta_{12}^s, \quad \frac{\partial y_{2t+s}}{\partial \eta_{1t}} = \theta_{21}^s, \quad \frac{\partial y_{2t+s}}{\partial \eta_{2t}} = \theta_{22}^s$$

and $\lim_{s \to \infty} \theta_{ij}^s = 0$ in the case of stationary variables (the response of ith variable to jth shock after s periods).

Long-run cumulative impact matrix is

$$\begin{aligned} \sum_{s=0}^{\infty} \Theta^s = \begin{pmatrix} \sum_{s=0}^{\infty} \theta_{11}^s & \sum_{s=0}^{\infty} \theta_{12}^s \\ \sum_{s=0}^{\infty} \theta_{21}^s & \sum_{s=0}^{\infty} \theta_{22}^s \end{pmatrix} = \begin{pmatrix} \theta_{11}(1) & \theta_{12}(1) \\ \theta_{21}(1) & \theta_{22}(1) \end{pmatrix} = \Theta(1) \\ \Theta(L) = \begin{pmatrix} \theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L) \end{pmatrix} = \begin{pmatrix} \sum_{s=0}^{\infty} \theta_{11}^s L^s & \sum_{s=0}^{\infty} \theta_{12}^s L^s \\ \sum_{s=0}^{\infty} \theta_{21}^s L^s & \sum_{s=0}^{\infty} \theta_{22}^s L^s \end{pmatrix}. \end{aligned}$$

3 Identification through long-run restrictions

There are multiple ways to impose long-run restriction (Tip: the term *long-run* here implies infinity).

Example: Blanchard–Quah decomposition.

$$y_{1t} \sim I(1), \quad y_{2t} \sim I(0).$$

To impose long-run restriction, the variable has to be $I(1)$.

$$\begin{aligned} y_{1t} &= \text{log of real GDP}, \quad y_{2t} = \text{unemployment rate} \\ Y_t &= \underbrace{(\Delta y_{1t}, y_{2t})^{\mathrm{t}}}_{\substack{\text{reduced form needs to be} \\ \text{estimated in stationary variables}}} \end{aligned}$$

Structural VAR: $BY_t = \gamma_0 + \Gamma_1 Y_{t-1} + \eta_t$

Blanchard and Quah loosely interpreted η_{1t} as supply shock & η_{2t} as demand shock. η_{1t} has permanent effect on the level of y_{1t}. η_{2t} is transitory shock and it can be interpreted as demand shock.

BQ assumptions: η_{2t} (demand shock) has no long-run impact on the level of real GDP (y_1)

Econometric Methods II Lecture Note 15

Junyong Kim

March 13, 2017

1 Blanchard–Quah decomposition

$$
\begin{aligned}
Y_t &= (\Delta y_{1t}, y_{2t})^{\mathrm{t}} \\
\Delta y_{1t} &= \text{rate of growth of real GDP} \\
y_{2t} &= \text{unemployment.}
\end{aligned}
$$

Structural VAR representation: $BY_t = \gamma_0 + \Gamma_1 Y_{t-1} + \eta_t$
Reduced form VAR: $Y_t = a_0 + A_1 Y_{t-1} + \varepsilon_t$ and $\mathrm{E}[\varepsilon_t \varepsilon_t^{\mathrm{t}}] = \Omega = B^{-1} D \left(B^{-1}\right)^{\mathrm{t}}$ where $D = \mathrm{E}[\eta_t \eta_t^{\mathrm{t}}]$, which is diagonal.

Impulse responses are

$$
\begin{aligned}
\frac{\partial \Delta y_{1t+s}}{\partial \eta_{1t}} &= \theta_{11}^s, \quad \frac{\partial \Delta y_{1t+s}}{\partial \eta_{2t}} = \theta_{12}^s \\
\frac{\partial y_{2t+s}}{\partial \eta_{1t}} &= \theta_{21}^s, \quad \frac{\partial y_{2t+s}}{\partial \eta_{2t}} = \theta_{22}^s.
\end{aligned}
$$

And also,

$$
\begin{aligned}
y_{1t+s} &= y_{1t-1} + \sum_{j=0}^{s} \Delta y_{1t+j} \\
&= y_{1t-1} + \sum_{j=0}^{s} y_{1t+j} - y_{1t+j-1},
\end{aligned}
$$

and hence

$$
\begin{aligned}
\frac{\partial y_{1t+s}}{\partial \eta_{kt}} &= \frac{\partial y_{1t-1}}{\partial \eta_{kt}} + \sum_{j=0}^{s} \frac{\partial \Delta y_{1t+j}}{\partial \eta_{kt}} \\
&\approx \sum_{j=0}^{s} \frac{\partial \Delta y_{1t+j}}{\partial \eta_{kt}} \\
&= \sum_{j=0}^{s} \theta_{1k}^j,
\end{aligned}
$$

and hence

$$\lim_{s\to\infty} \frac{\partial y_{1t+s}}{\partial \eta_{kt}} = \sum_{j=0}^{\infty} \theta_{1k}^{j} \\ = \theta_{1k}(1).$$

i.e. Long-run impact of a shock

BQ assumption is that demand shock as measured by shock to unemployment does not have any long-run impact on the level of GDP, i.e.

$$\frac{\partial y_{1t+s}}{\partial \eta_{2t}} = 0, \quad \text{if } s \to \infty$$
$$\sum_{s=0}^{\infty} \theta_{12}^{s} = 0, \quad \theta_{12}(1) = 0.$$

The restriction that $\theta_{12}(1) = 0$ can be used to identify the SVAR model $BY_t = \gamma_0 + \Gamma_1 Y_{t-1} + \eta_t$ from the reduced form VAR model $Y_t = a_0 + A_1 Y_{t-1} + \varepsilon_t$.

Long-run impact matrix

$$\Theta(1) = \begin{pmatrix} \theta_{11}(1) & \theta_{12}(1) \\ \theta_{21}(1) & \theta_{22}(1) \end{pmatrix} = \begin{pmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{pmatrix}, \quad \text{where 0 is a BQ restriction}$$

Long-run covariance matrix of

$$\begin{aligned} \mathrm{LRV} = \Lambda = & \Psi(1)\Omega\Psi(1)^{\mathrm{t}} \\ = & \left(\mathbf{I} - A_1\right)^{-1} B^{-1} D \left(B^{-1}\right)^{\mathrm{t}} \left[\left(\mathbf{I} - A_1\right)^{-1}\right]^{\mathrm{t}} \\ = & \left[\left(\mathbf{I} - A_1\right) B\right]^{-1} D \left\{\left[\left(\mathbf{I} - A_1\right) B\right]^{-1}\right\}^{\mathrm{t}} \end{aligned}$$

and BQ assume that $D = \mathbf{I}$ so

$$\begin{aligned} \Lambda = & \left[\left(\mathbf{I} - A_1\right) B\right]^{-1} \left\{\left[\left(\mathbf{I} - A_1\right) B\right]^{-1}\right\}^{\mathrm{t}} \\ = & \Theta(1)\Theta(1)^{\mathrm{t}} \end{aligned}$$

because $\Theta(L) = \Psi(L)B^{-1}$.

Note that Λ can be decomposed by Cholesky decomposition; $\Lambda = \mathbf{L}\mathbf{L}^{\mathrm{t}}$ where $\mathbf{L}$ is a lower triangular matrix and $\Theta(1) = \mathbf{L}$. Once we get $\mathbf{L}$, we know $\Theta(1)$. We can recover B matrix since we know A_1, i.e.

$$\begin{aligned} \Theta(1) = & \mathbf{L} = \left(\mathbf{I} - A_1\right)^{-1} B^{-1} \\ \Rightarrow & B = \left[\left(\mathbf{I} - A_1\right) \mathbf{L}\right]^{-1} \end{aligned}$$

2 Steps in the identification of SVAR model using BQ restriction

Read: Kilian, *Structural VAR*

1. Estimate the reduced form VAR and get estimates of parameters a_0, A_1, & Ω. These estimates are $\hat{a}_0$, $\hat{A}_1$ & $\hat{\Omega}$

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t^{\mathrm{t}}$$

2. Compute a parameteric estimate of the long-run covariance matrix

$$\begin{aligned}\hat{\Lambda} =& \hat{\Psi}(1)\hat{\Omega}\hat{\Psi}(1)^{\mathrm{t}} \\ =& \left(I - \hat{A}_1\right)^{-1} \hat{\Omega} \left[\left(I - \hat{A}_1\right)^{-1}\right]^{\mathrm{t}}\end{aligned}$$

3. Cholesky factorization of $\hat{\Lambda} = \hat{P}\hat{P}^{\mathrm{t}}$ and $\hat{P} = \hat{\Theta}(1)$

4. $\hat{\Theta}(1) = \hat{P}(1) = \left(I - \hat{A}_1\right)^{-1} B^{-1}$ and hence $\hat{B} = \left[\left(I - \hat{A}_1\right)\hat{P}\right]^{-1}$

   For impulse responses,
   $$\hat{\Theta}_k = \hat{\Psi}_k \hat{B}^{-1}$$

   Once we estimate B, we can get impulse response function & forecast error variance decompositions

   Note that SMA (structural MA representation)

$$\begin{aligned}Y_t =& \mu + \Theta_0 \eta_t + \Theta_1 \eta_{t-1} + \Theta_2 \eta_{t-2} + \cdots \\ \Theta_k =& \Psi_k B^{-1} \\ \text{where } \Psi(L) =& \left(I - A_1\right)^{-1} \\ \Psi(L) =& \sum_{k=0}^{\infty} \Psi_k L^k\end{aligned}$$

3 R session

We studied

- How to deal with structural VAR model
- How to impose a structural restriction

Econometric Methods II Lecture Note 16

Junyong Kim

March 15, 2017

Announcement: April 19, 2017 (midterm)

1 R session

In the session we covered

- The way to use SVAR function
- LR test for overidentification
- The way to use BQ function

2 SVAR

We covered the powerpoint (pdf) file regarding SVARs.

- Rigobon and Sack (2003, QJE)

Econometric Methods II Lecture Note 17

Junyong Kim

March 27, 2017

1 Identification with sign restrictions

We studied...

- Agnostic identification: Repetition of random draws and disposes (computationally very intensive)
- Strategies using high frequency data (quite appealing, but the impact is seemingly too weak)
- Identification of shocks through narrative evidence: ex. the effect of the unexpected increases in Alaska fund returns on consumption behaviors (Pros: We can reflect a lot of information, Cons: Highly subjective).

2 R session

We studied how to use the SVAR method proposed by Uhlig (2005) in R package: `VARsignR`.

Econometric Methods II Lecture Note 18

Junyong Kim

March 29, 2017

Announcement: In April 10 and April 19, this class will start at 1:50 and the midterm is delayed to April 26.

1 State space models & Kalman filter

State space models typically are associated with dynamic time series models with unobserved variables.

Example 1: Time-varying parameter model, i.e.

$$\begin{aligned} y_t =& \beta_{0t} + \beta_{1t}x_{1t} + \cdots + \beta_{kt}x_{kt} + e_t \\ e_t \sim& \text{iidN}(0, R) \\ \beta_{it} =& \beta_{it-1} + \nu_{it} \\ \nu_{it} \sim& \text{iidN}\left(0, \sigma^2_{\nu_i}\right), \end{aligned}$$

and usual regression models have fixed βs. Here β is time-varying and all other assumptions of OLS hold.

GLS can be applied to this model but it is computationally intensive.

Example 2: Unobserved component model, i.e.

$$\begin{aligned} y_t =& \ln \text{GDP}_t \\ =& \underbrace{\tau_t}_{\text{trend}} + \underbrace{c_t}_{\text{cycle}}, \end{aligned}$$

and both trend & cycle are unobserved. What is observable is y_t and what is unobservable is τ_t and c_t, i.e.

$$\begin{aligned} \tau_t =& \mu + \tau_{t-1} + \nu_t, \quad \nu \sim \text{iid}\left(0, \sigma^2_\nu\right), \quad \text{i.e. a random walk with a drift} \\ c_t =& \phi_1 c_{t-1} + \phi_2 c_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma^2_\varepsilon\right), \quad \text{i.e. stationary AR(2) cycle.} \end{aligned}$$

1.1 Two equations in state space framework

- Measurement equation: Expresses observed variable as a linear function of unobserved variables

$$y_t = H_t\beta_t + Az_t + e_t, \quad e_t \sim \text{iid}(0, R).$$

- Transition equation (or state equation): Expresses unobserved variables as a linear function of its own past values

$$\beta_t = \mu + F\beta_{t-1} + \nu_t, \quad \nu_t \sim \text{iid}(0, Q),$$

and where z_t is an exogenous variable & is known. β_t is an unobservable variable.

Example 2: Unobserved component model (cont'd),

$$\begin{aligned} y_t =& \tau_t + c_t \\ \tau_t =& \mu + \tau_{t-1} + \nu_t \\ c_t =& \phi_1 c_{t-1} + \phi_2 c_{t-2} + \varepsilon_t, \end{aligned}$$

and hence, the measurement equation is

$$\begin{aligned} y_t =& \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \tau_t \\ c_t \\ c_{t-1} \end{pmatrix} \\ =& H_t\beta_t, \quad z_t = e_t = 0, \end{aligned}$$

and the transition equation is

$$\begin{aligned} \begin{pmatrix} \tau_t \\ c_t \\ c_{t-1} \end{pmatrix} =& \begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \tau_{t-1} \\ c_{t-1} \\ c_{t-2} \end{pmatrix} + \begin{pmatrix} \nu_t \\ \varepsilon_t \\ 0 \end{pmatrix} \\ \beta_t =& \mu + F\beta_{t-1} + \nu_t. \end{aligned}$$

In the time-varying parameter model,

$$\begin{aligned} y_t =& \beta_{0t} + \beta_{1t}x_{1t} + \beta_{2t}x_{2t} + e_t \\ \beta_{it} =& \beta_{it-1} + \nu_{it}, \quad i \in \{0, 1, 2\}, \quad \nu_t \sim \text{iid}\left(0, \sigma^2_{\nu_i}\right). \end{aligned}$$

Measurement equation is

$$y_t = \begin{pmatrix} 1 & x_{1t} & x_{2t} \end{pmatrix} \begin{pmatrix} \beta_{0t} \\ \beta_{1t} \\ \beta_{2t} \end{pmatrix} + e_t = H_t\beta_t + e_t.$$

Transition equation is

$$\begin{pmatrix} \beta_{0t} \\ \beta_{1t} \\ \beta 2t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_{0t-1} \\ \beta_{1t-1} \\ \beta_{2t-1} \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \\ \nu_{3t} \end{pmatrix}.$$

1.2 Kalman filter

Kalman filter: Recursive procedure for computing the optimal estimate of the unobserved state vector β_t based on the appropriate information set & the parameters of the model, $t \in \{1, \cdots, T\}$. Kalman filter provides is an estimate of $\beta_{t|t} = \mathrm{E}\left[\beta_t | \Psi_t\right]$ where Ψ_t is the information set at time t.

Sometimes, one also uses smoothed estimate of β_t, i.e. $\beta_{t|T} = \mathrm{E}\left[\beta_t | \Psi_T\right]$, where T is the full sample size.

1.2.1 Notation

$$\begin{aligned}
P_{t|t} =& \mathrm{E}\left[\left(\beta_t - \beta_{t|t}\right)\left(\beta_t - \beta_{t|t}\right)^\top\right], \quad \text{state variance} \\
y_{t|t-1} =& \mathrm{E}\left[y_t | \Psi_{t-1}\right] \\
\eta_{t|t-1} =& y_t - y_{t|t-1}, \quad \text{prediction error for measurement equation} \\
f_{t|t-1} =& \mathrm{E}\left[\eta_{t|t-1}\eta_{t|t-1}^\top\right], \quad \text{variance of prediction error.}
\end{aligned}$$

Prediction is

$$\begin{aligned}
\beta_{t|t-1} =& \mu + F\beta_{t-1|t-1} \\
P_{t|t-1} =& FP_{t-1|t-1}F^\top + Q \\
\eta_{t|t-1} =& y_t - y_{t|t-1} = y_t - H_t\beta_{t|t-1} - Az_t \\
f_{t|t-1} =& H_tP_{t|t-1}H_t^\top + R,
\end{aligned}$$

where R is the variance of measurement equation error.

Updating is

$$\begin{aligned}
\beta_{t|t} =& \beta_{t|t-1} + k_t\eta_{t|t-1} \\
P_{t|t} =& P_{t|t-1} - k_tH_tP_{t|t-1},
\end{aligned}$$

and where k_t is called as Kalman gain and $k_t = P_{t|t-1}H_t^\top f_{t|t-1}^{-1}$. k_t determines the weight assigned to the new information about β_t contained in the prediction error $\eta_{t|t-1}$.

Digression:

$$\begin{aligned}
\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} \text{ given } \Psi_{t-1} \sim& \mathrm{MVN}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}\right), \\
z_{1t}|z_{2t} \text{ given } \Psi_{t-1} \sim& \mathrm{N}\left(\mu_{1|2}, \sigma_{11|2}\right) \\
\mu_{1|2} =& \mu + \sigma_{12}\sigma_{22}^{-1}\left(z_{2t} - \mu_2\right) \\
\sigma_{11|2} =& \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{12}.
\end{aligned}$$

$z_{1t} = \beta_t$, $z_{2t} = \eta_{t|t-1}$, $\mu_1 = \beta_{t|t-1}$, $\sigma_{11} = P_{t|t-1}$, $\sigma_{22} = f_{t|t-1}$, $\sigma_{12} = P_{t|t-1}H_t^\top$.

Note that k_t is the weight assigned to $\eta_{t|t-1}$ in updating $\beta_{t|t}$. Higher the uncertainty in $\beta_{t|t-1}$, more weight (k_t) is assigned to $\eta_{t|t-1}$ ($P_{t|t-1}$: the uncertainty in $\beta_{t|t-1}$). Higher the uncertainty in $\eta_{t|t-1}$, lower the weight assigned to $\eta_{t|t-1}$ ($f_{t|t-1}$: the uncertainty in $\eta_{t|t-1}$).

What one must estimate for this model is μ, F, Q, R and A and the initial $\beta_{0|0}$ and $P_{0|0}$.

Econometric Methods II Lecture Note 19

Junyong Kim

April 3, 2017

1 Kalman filter and prediction

Prediction:

$$\begin{aligned}
\beta_{t|t-1} =& \mu + F\beta_{t-1|t-1} \\
P_{t|t-1} =& FP_{t-1|t-1}F^\top + Q \\
\eta_{t|t-1} =& y_t - y_{t|t-1} \\
=& y_t - H_t\beta_{t|t-1} - Az_t \\
f_{t|t-1} =& H_tP_{t|t-1}H_t^\top + R.
\end{aligned}$$

Updating:

$$\begin{aligned}
\beta_{t|t} =& \beta_{t|t-1} + k_t\eta_{t|t-1} \\
P_{t|t} =& P_{t|t-1} - k_tH_tP_{t|t-1},
\end{aligned}$$

where k_t is the Kalman gain $P_{t|t-1}H_t^\top f_{t|t-1}^{-1}$.

Since $\eta_{t|t-1} \sim \mathrm{N}\left(0, f_{t|t-1}\right)$, use this property of prediction to form likelihood and get the MLEs of μ, F, Q, R & H_t. And then estimate the state variable β_t and here we need initial value of $\beta_{0|0}$ & $P_{0|0}$.

For stationary transition equation, we can use unconditional mean & unconditional variance of β_t. So

$$\begin{aligned}
\beta_t =& \mu + F\beta_{t-1} + \nu_t \\
\mathrm{E}\left[\beta_t\right] =& \mu + F\mathrm{E}\left[\beta_{t-1}\right],
\end{aligned}$$

and at the steady state,

$$\begin{aligned}
\beta_{0|0} =& \mu + F\beta_{0|0} \\
=& \left(I - F\right)^{-1}\mu \\
P_{0|0} =& FP_{0|0}F^\top + Q \\
\text{note that } \mathrm{Vec}(ABC) =& \left(C^\top \otimes A\right)\mathrm{Vec}(B) \\
\mathrm{Vec}\left(P_{0|0}\right) =& \mathrm{Vec}\left(FP_{0|0}F^\top\right) + \mathrm{Vec}(Q) \\
=& (F \otimes F)\mathrm{Vec}\left(P_{0|0}\right) + \mathrm{Vec}(Q) \\
=& [I - (F \otimes F)]^{-1}\mathrm{Vec}(Q),
\end{aligned}$$

and hence we can use these unconditional moments in the case of stationary state variables.
Measurement equation:

$$y_t = H_t\beta_t + Az_t + e_t, \quad e_t \sim \text{iidN}(0, R).$$

Transition equation:

$$\beta_t = \mu + F\beta_{t-1} + \nu_t, \quad \nu_t \sim \text{iidN}(0, Q).$$

In case of non-stationary state variables, we can't use unconditional mean and variance as the initial value. Here, $\beta_{0|0}$ can be any arbitrary guess & $P_{0|0}$ can be high numbers to reflect the high uncertainty about $\beta_{0|0}$.

Calculate $\beta_{0|0}$ & $P_{0|0}$ for a stationary AR(2) model as,

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \nu_t, \quad \nu_t \sim \text{iidN}\left(0, \sigma_\nu^2\right).$$

Transition equation:

$$\begin{aligned}
\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} &= \begin{pmatrix} \mu \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \nu_t \\ 0 \end{pmatrix} \\
\beta_t &= \mu + F\beta_{t-1} + \nu_t \\
\beta_{0|0} &= (I - F)^{-1}\mu \\
&= \begin{pmatrix} 1-\phi_1 & -\phi_2 \\ -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mu \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \frac{\mu}{1-\phi_1-\phi_2} \\ \frac{\mu}{1-\phi_1-\phi_2} \end{pmatrix} \\
\text{Vec}\left(P_{0|0}\right) &= (I - F \otimes F)^{-1} \text{Vec}(Q) \\
&= \begin{pmatrix} 1-\phi_2^2 & -\phi_1\phi_2 & -\phi_1\phi_2 & -\phi_2^2 \\ -\phi_1 & 1 & -\phi_2 & 0 \\ -\phi_1 & -\phi_2 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_\nu^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}.
\end{aligned}$$

$\beta_{t|t}$ is the filtered estimate of β_t. It uses only information until time period t and does not use information from future. Thus, $\beta_{t|t}$ is one-sided-filtered estimate of β_t.

In some applications, full-sample information is needed to smooth out the noise in one-sided-filtered estimate of β_t.

If we use information from full sample, we say that we have a smoothed estimate of β_t, i.e. for $t \in \{1, \cdots, t, \cdots, T\}$, $\beta_{t|t}$ is the filtered estimate of β_t and $\beta_{t|T}$ is the smoothed estimate of β_t that uses information from the full sample.
Smoothing:

$$\begin{aligned}
\beta_{t|T} &= \beta_{t|t} + P_{t|t}F^\top P_{t+1|t}^{-1}\left(\beta_{t+1|T} - F\beta_{t|t} - \mu\right) \\
P_{t|T} &= P_{t|t} + P_{t|t}F^\top P_{t+1|t}^{-1}\left(P_{t+1|T} - P_{t+1|t}\right) P_{t+1|t}^{-1} F P_{t|t}^\top.
\end{aligned}$$

$\beta_{T|T}$ & $P_{T|T}$ are the initial values for the smoothing and are obtained from the last iteration of the basic filter. So one can go from the last observation to the first observation in order to find the smoothed estimates.

2 Identification problems in unobserved component models

$$
\begin{aligned}
y_t = & \underbrace{y_{1t}}_{\text{RW}} + \underbrace{y_{2t}}_{\text{WN}} \\
y_{1t} = & y_{1t-1} + e_{1t}, \quad e_{1t} \sim \text{iid}\left(0, \sigma_1^2\right) \\
y_{2t} = & e_{2t}, \quad e_{2t} \sim \text{iid}\left(0, \sigma_2^2\right) \\
\text{Cov}\left[e_{1t}, e_{2t}\right] = & \sigma_{12}.
\end{aligned}
$$

Reduced form version:

$$
\begin{aligned}
\Delta y_t = & \Delta y_{1t} + \Delta y_{2t} \\
= & e_{1t} + e_{2t} - e_{2t-1} \\
\approx & \varepsilon_{1t} + \theta \varepsilon_{1t-1}, \quad \varepsilon_{1t} \sim \text{iid}\left(0, \sigma_\varepsilon^2\right).
\end{aligned}
$$

Since this reduced form version has only two parameters (θ, σ_ε^2), the structural model cannot be identified because the structural model involves three parameters (σ_1^2, σ_{12}, σ_2^2). If $\sigma_{12} = 0$, then this UC model is identified.

2.1 Clark's unobserved component model

$$
\begin{aligned}
y_t = & \ln \text{GDP}_t \\
= & \tau_t + c_t \\
\tau_t = & \mu + \tau_{t-1} + \nu_t, \quad \nu_t \sim \text{iid}\left(0, \sigma_\nu^2\right) \\
c_t = & \phi_1 c_{t-1} + \phi_2 c_{t-1} + e_t, \quad e_t \sim \text{iid}\left(0, \sigma_e^2\right).
\end{aligned}
$$

Clark argued that to identify this UC model we need $\text{Cov}\left[\nu_t, e_t\right] = \sigma_{\nu e} = 0$. Shocks to permanent component & shocks to transitory component are uncorrelated with each other.

Morley, Nelson and Zivot (2003) argued that the UC model is still identified if $\sigma_{\nu e} \neq 0$ & in fact, they show that the restriction $\sigma_{\nu e} = 0$ is rejected by data in case of U.S. GDP.

In Clark's UC model, the total number of parameters to be estimated is 5 and the restriction imposed there is $\sigma_{e\nu} = 0$.
Reduced form representation:

$$
\begin{aligned}
\Delta y_t = & \Delta \tau_t + \Delta c_t \\
= & \mu + \nu_t + (1 - L) c_t,
\end{aligned}
$$

and we know that $c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + e_t = \left(1 - \phi_1 L - \phi_2 L^2\right)^{-1} e_t$. Thus,

$$\begin{aligned}
\Delta y_t =& \mu + \nu_t + (1-L)\left(1 - \phi_1 L - \phi_2 L^2\right)^{-1} e_t \\
\left(1 - \phi_1 L - \phi_2 L^2\right) \Delta y_t =& \mu \left(1 - \phi_1 L - \phi_2 L^2\right) + \nu_t \left(1 - \phi_1 L - \phi_2 L^2\right) + (1-L) e_t \\
=& \mu^* + \nu_t - \phi_1 \nu_{t-1} - \phi_2 \nu_{t-2} + e_t - e_{t-1} \\
\approx& \mu^* + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \quad \varepsilon_t \sim \text{iid}\left(0, \sigma_\varepsilon^2\right).
\end{aligned}$$

Since the process Δy_t follows ARMA(2,2), there are 6 parameters to be estimated (2 ARs, 2 MAs, intercept and sigma). So, from ARMA(2,2), one can identify Clark's UC model (5 parameters) fully and even can test whether the imposed condition $\sigma_{\nu e} = 0$ is an overidentification or not.

Announcement: Read MNZ 2003 paper. In the next class we will study `R` programmings.

Econometric Methods II Lecture Note 20

Junyong Kim

April 5, 2017

Announcement: Term paper due is May 19, 2017.

1 Unobserved component models

Clark's classical UC model assumed correlation between shock to the permanent components & shock to transitory component are uncorrelated, i.e.

$$
\begin{aligned}
y_t =& \tau_t + c_t \\
\tau_t =& \mu + \tau_{t-1} + \nu_t, \quad \nu_t \sim \text{iid}\left(0, \sigma_\nu^2\right) \\
c_t =& \phi_1 c_{t-1} + \phi_2 c_{t-2} + e_t, \quad e_t \sim \text{iid}\left(0, \sigma_e^2\right) \\
\text{Cov}\left[e_t, \nu_t\right] =& 0, \quad \text{for Clark's model} \\
\neq& 0, \quad \text{for MNZ model.}
\end{aligned}
$$

It was shown by MNZ that $\text{Cov}\left[e_t, \nu_t\right] = 0$ is not required for the identification of the UC model. In fact, this covariance can be estimated.

2 Dynamic factor model

There are 3 countries. π^f is the inflation in France, π^g is the information in Germany and π^i is the information in Italy, respectively. Then

$$
\begin{aligned}
\pi^g =& \gamma_1 c_t + \eta_{1t} \\
\pi^f =& \gamma_2 c_t + \eta_{2t} \\
\pi^i =& \gamma_3 c_t + \eta_{3t} \\
\text{where } c_t =& \text{common European factor} \\
\gamma_i =& \text{the loading on that factor} \\
\eta_i =& \text{the idiosyncratic factors.}
\end{aligned}
$$

Identification condition: Either normalize the loadings on one of the countries to be equal to one or normalize the variance of the common component to a constant.

Suppose, we normalize the variance of the common component,

$$c_t = \beta c_{t-1} + \nu_t, \quad \nu_t \sim \text{iid}(0,1) \leftarrow \text{normalized variance.}$$

Suppose idiosyncratic factors follow AR(2) process, then

$$\begin{aligned}
\eta_{1t} =& \phi_{11}\eta_{1t-1} + \phi_{12}\eta_{1t-2} + e_{1t} \\
\eta_{2t} =& \phi_{21}\eta_{2t-1} + \phi_{22}\eta_{2t-2} + e_{2t} \\
\eta_{3t} =& \phi_{31}\eta_{3t-1} + \phi_{32}\eta_{3t-2} + e_{3t} \\
e_{it} =& \text{iidN}\left(0, \sigma_{e_i}^2\right), \quad i \in \{1,2,3\}.
\end{aligned}$$

Measurement equation:

$$\begin{pmatrix} \pi^g \\ \pi^f \\ \pi^i \end{pmatrix} = \begin{pmatrix} \gamma_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \gamma_2 & 0 & 0 & 1 & 0 & 0 & 0 \\ \gamma_3 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} c_t \\ \eta_{1t} \\ \eta_{1t-1} \\ \eta_{2t} \\ \eta_{2t-1} \\ \eta_{3t} \\ \eta_{3t-1} \end{pmatrix}.$$

Transition equation:

$$\begin{pmatrix} c_t \\ \eta_{1t} \\ \eta_{1t-1} \\ \eta_{2t} \\ \eta_{2t-1} \\ \eta_{3t} \\ \eta_{3t-1} \end{pmatrix} = \begin{pmatrix} \beta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_{11} & \phi_{12} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_{21} & \phi_{22} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \phi_{31} & \phi_{32} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} c_{t-1} \\ \eta_{1t-1} \\ \eta_{1t-2} \\ \eta_{2t-1} \\ \eta_{2t-2} \\ \eta_{3t-1} \\ \eta_{3t-2} \end{pmatrix} + \begin{pmatrix} \nu_t \\ e_{1t} \\ 0 \\ e_{2t} \\ 0 \\ e_{3t} \\ 0 \end{pmatrix}.$$

Note that,

$$\text{Var}\left[\pi^g\right] = \gamma_1^2 \underbrace{\text{Var}\left[c_t\right]}_{\substack{\text{unconditional} \\ \text{variance}}} + \underbrace{\text{Var}\left[\eta_{1t}\right]}_{\substack{\text{unconditional} \\ \text{variance}}},$$

i.e. share of common component in overall variance of inflation in Germany is

$$\frac{\gamma_1^2 \text{Var}\left[c_t\right]}{\text{Var}\left[\pi^g\right]} = \frac{\gamma_1^2 \text{Var}\left[c_t\right]}{\gamma_1^2 \text{Var}\left[c_t\right] + \text{Var}\left[\eta_{1t}\right]}.$$

Econometric Methods II Lecture Note 21

Junyong Kim

April 10, 2017

1 `R` session

- In estimating the model $y_t = \phi y_{t-1} + \varepsilon_t$, we have two options to choose; conditional ML and exact ML
- OLS is equivalent to conditional ML since it excludes the very first observation
- State space model is close to exact ML
- In this session, we studied the use of `dlm` and `tis`
- Nuisance parameter problem

Econometric Methods II Lecture Note 22

Junyong Kim

April 12, 2017

1 R session

Please refer the revised UC programming written by Kundan.

1.1 No break in trend mean

$$\begin{aligned} y_t =& \tau_t + c_t \\ \tau_t =& \mu + \tau_{t-1} + \nu_t, \quad \nu_t \sim \mathrm{iidN}\left(0, \sigma_\nu^2\right) \\ c_t =& \phi_1 c_{t-1} + \phi_2 c_{t-2} + e_t, \quad e_t \sim \mathrm{iidN}\left(0, \sigma_e^2\right) \\ \mathrm{Cov}\left[\nu_t, e_t\right] =& 0. \end{aligned}$$

Measurement equation

$$y_t = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \tau_t \\ c_t \\ c_{t-1} \end{pmatrix}.$$

Transition equation

$$\begin{pmatrix} \tau_t \\ c_t \\ c_{t-1} \end{pmatrix} = \begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \tau_{t-1} \\ c_{t-1} \\ c_{t-2} \end{pmatrix} + \begin{pmatrix} \nu_t \\ e_t \\ 0 \end{pmatrix}.$$

The roundabout approach to include $\mu_t = \mu$ for all t in `R` package.

$$y_t = \begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \tau_t \\ \mu_t \\ c_t \\ c_{t-1} \end{pmatrix}$$

$$\begin{pmatrix} \tau_t \\ \mu_t \\ c_t \\ c_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \tau_{t-1} \\ \mu_{t-1} \\ c_{t-1} \\ c_{t-2} \end{pmatrix} + \begin{pmatrix} \nu_t \\ 0 \\ e_t \\ 0 \end{pmatrix}.$$

Notice: The initial value matters a lot. One has to be careful with this feature.

We studied how to use `dlm` and `tis` packages today as well.

Basically we are assuming that there is no parameter break during that time.

2 Trend–Cycle decomposition: `TrendCycle.pdf`

- We studied several filters including HP filter, BK filter, BN decomposition, etc
- In the first time, a lot of researchers just regressed the time series on t to decompose the trend from the cycle
- If we use t linearly to decompose, then this implies that the growth rate of the trend is constant over time. So the production shock plays no role in determining the trend (deterministic trend)
- This had been advocated by Keynesian economists (they emphasize the role of cyclical components) and this approach by and large ignores the production shock
- In order to apply more elaborate way to decompose these components, many filters such as HP, BK, BN have been suggested

2.1 Beveridge–Nelson decomposition

If y_t is a non-stationary series with deterministic time trend

$$y_t = \underbrace{TD_t}_{\substack{\text{deterministic}\\ \text{trend}}} + \underbrace{TS_t}_{\substack{\text{stochastic}\\ \text{trend}}} + \underbrace{c_t}_{\text{cycle}}$$

where TD is deterministic trend, TS is stochastic trend, C is cycle. If $z_t = TS_t + c_t$, Beveridge–Nelson proposed the definition of the permanent component of an $I(1)$ time series y_t with drift μ as the limiting forecast as horizon goes to infinity, adjusted for the mean rate of growth over the forecast horizon.

BN showed that if Δy_t has a Wold representation,

$$\Delta y_t = \mu + \psi(L)\varepsilon_t$$

the Beveridge–Nelson trend (BN_t) follows a pure random walk without a drift, i.e.

$$\begin{aligned} BN_t =& BN_{t-1} + \psi(1)\varepsilon_t \\ =& BN_0 + \underbrace{\psi(1)\sum_{s=1}^{t}\varepsilon_s}_{\substack{\text{BN stochastic}\\ \text{trend}}} \end{aligned}$$

2.2 Derivation of BN stochastic trend

Suppose $\psi(L)$ is polynomial of degree q.

$$\begin{aligned}
D(L) =& \psi(L) - \psi(1) \\
D(1) =& 0 \\
D(L) =& \psi^*(L)(1-L) \\
\psi^*(L)(1-L) =& \psi(L) - \psi(1) \\
\psi(L) =& \psi^*(L)(1-L) + \psi(1).
\end{aligned}$$

And

$$\begin{aligned}
\Delta y_t =& \mu + \psi(L)\varepsilon_t \\
y_t =& y_{t-1} + \mu + \psi(L)\varepsilon_t \\
=& y_0 + \mu t + \psi(L)\sum \varepsilon_t \\
=& \underbrace{y_0 + \mu t}_{TD_t} + (\psi(1) + \psi^*(L)(1-L))\sum \varepsilon_t \\
=& TD_t + \psi(1)\sum \varepsilon_t + \psi^*(L)(1-L)\sum \varepsilon_t \\
=& TD_t + \psi(1)\sum \varepsilon_t + \left(\psi^*(L)\sum \varepsilon_t - L\psi^*(L)\sum \varepsilon_t\right) \\
=& \underbrace{TD_t}_{\substack{\text{deterministic}\\ \text{trend}}} + \underbrace{\psi(1)\sum \varepsilon_t}_{\substack{\text{stochastic}\\ \text{trend}}} + \underbrace{\tilde{\varepsilon}_t - \tilde{\varepsilon}_0}_{\text{stationary component}}
\end{aligned}$$

and the BN trend is $\psi(1)\sum \varepsilon_t$.

Econometric Methods II Lecture Note 23

Junyong Kim

April 17, 2017

1 BN decomposition

Beveridge–Nelson decomposition is

$$y_t = \underbrace{y_0 + \mu t}_{TD_t} + \underbrace{\psi^*(1)\sum_{j=1}^{t}\varepsilon_j}_{TS_t} + \underbrace{\tilde{\varepsilon}_t - \tilde{\varepsilon}_0}_{\text{cycle}},$$

where TD_t is a deterministic trend, TS_t is a stochastic trend, respectively. BN_t, Beveridge–Nelson trend is the long-run forecast of y_t adjusted for mean.

$$BN_t = \lim_{h\to\infty}\big(y_{t+h|t} - \mu h\big),$$

and then,

$$\begin{aligned}
y_{t+h} =& y_0 + \mu(t+h) + \psi^*(1)\sum_{j=1}^{t+h}\varepsilon_j + \tilde{\varepsilon}_{t+h} \\
y_{t+h|t} =& y_0 + \mu(t+h) + \psi^*(1)\sum_{j=1}^{t}\varepsilon_j + \tilde{\varepsilon}_{t+h|t} \\
\lim_{h\to\infty}\big(y_{t+h|t} - \mu h\big) =& y_0 + \mu t + \psi^*(1)\sum_{j=1}^{t}\varepsilon_j + \lim_{h\to\infty}\tilde{\varepsilon}_{t+h|t} \\
\lim_{n\to\infty}\big(y_{t+h|t} - \mu h\big) =& \text{trend} = TD_t + TS_t,
\end{aligned}$$

and the stochastic trend is $\psi^*(1)\sum_{j=1}^{t}\varepsilon_j$.

Exercise 1: Calculate BN stochastic trend for the following MA(1) model.

$$\begin{aligned}\Delta y_t =& \varepsilon_t + 0.3\varepsilon_{t-1}\\ =&(1+0.3L)\varepsilon_t\\ =&\psi(L)\varepsilon_t\\ \psi(1) =&1.3\\ \text{Loading on stochastic trend} =&1.3\\ \text{BN stochastic trend} =&1.3\sum_{j=1}^{t}\varepsilon_j.\end{aligned}$$

Exercise 2: ARMA(2,2)

$$\Delta y_t = 0.82 + 1.34\Delta y_{t-1} - 0.71\Delta y_{t-2} + \varepsilon_t - 1.05\varepsilon_{t-1} + 0.52\varepsilon_{t-2},$$

then

$$\begin{aligned}\psi(L) =&\phi(L)^{-1}\theta(L)\\ \phi(L) =&1 - 1.34L + 0.71L^2\\ \phi(1) =&0.37\\ \theta(L) =&1 - 1.05L + 0.52L^2\\ \theta(1) =&0.47\\ \psi(1) =&\frac{\theta(1)}{\phi(1)}\\ =&\frac{0.47}{0.37}\\ \approx&1.27\\ \Rightarrow TS_t =&1.27\sum_{j=1}^{t}\varepsilon_j\\ TD_t =&y_0 + 0.82t.\end{aligned}$$

1.1 Steps in calculation of BN decomposition

1. Estimate ARMA(p,q) model for Δy_t
2. Calculate $\psi^*(1)$ from estimated ARMA model
3. Calculate $\sum_{j=1}^{t}\varepsilon_j$ from the residuals of estimated ARMA(p,q) model

Exercise: BN decomposition for AR(1) model; suppose

$$\Delta y_t - \mu = \phi\left(\Delta y_{t-1} - \mu\right) + \varepsilon_t,$$

and we want to calculate

$$\lim_{h\to\infty} y_{t+h|t} - \mu h,$$

hence

$$
\begin{aligned}
\Delta y_{t+1}-\mu =&\phi\left(\Delta y_t-\mu\right)+\varepsilon_{t+1}\\
\Delta y_{t+2}-\mu =&\phi^2\left(\Delta y_t-\mu\right)+\phi\varepsilon_{t+1}+\varepsilon_{t+2}\\
\Delta y_{t+h}-\mu =&\phi^h\left(\Delta y_t-\mu\right)+\sum_{j=1}^{h}\phi^{h-j}\varepsilon_{t+j}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Delta y_{t+h} =&\mu+\phi^h\left(\Delta y_t-\mu\right)+\phi^{h-1}\varepsilon_{t+1}+\cdots+\phi\varepsilon_{t+h-1}+\varepsilon_{t+h}\\
\Delta y_{t+h|t} =&\mu+\phi^h\left(\Delta y_t-\mu\right)\\
y_{t+h|t} =&y_t+\sum_{s=1}^{h}\mu+\phi^s\left(\Delta y_t-\mu\right)\\
=&y_t+h\mu+\left(\Delta y_t-\mu\right)\sum_{s=1}^{h}\phi^s\\
\lim_{h\to\infty} y_{t+h|t}-h\mu =&y_t+\frac{\phi}{1-\phi}\left(\Delta y_t-\mu\right)\\
=&TD_t+TS_t\\
\text{cycle} =&y_t-TD_t-TS_t=-\frac{\phi}{1-\phi}\left(\Delta y_t-\mu\right).
\end{aligned}
$$

2 R session

- We studied how to use `mFilter` package; HP filter, BK filter, CF filter. And also we studied how to apply BN decomposition manually
- Test if there is a break in the dynamics: Chow, CUSUMSQ style, Andrew (allow 1 change), Bai and Perron (multiple changes)
- 2-sided filter (smoothing) vs. 1-sided filter

Econometric Methods II Lecture Note 24

Junyong Kim

April 19, 2017

Today we use `cointegration.pdf` and `cointegration_slides.pdf`.

1 Cointegration

- By differencing a time series, we can apply usual statistical inferences and asymptotics
- On the other hand, by doing that one may lose a lot of information from the level of the time series
- Cointegration is developed a lot by Engle and Granger (first stage)
- Why do we have to consider the cointegration? If there is a linear relation in between two variables, then their deviation may play a role in predicting their future movements

Spourious regression: If y_{1t} & y_{2t} are independent $I(1)$ variables and are not cointegrated, i.e.

$$y_{1t} = \beta y_{2t} + u_t,$$

then this is the spurious regression in between y_1 and y_2.

1.1 Characteristics of spurious regression

1. OLS estimate of β does not converge as $t \to \infty$. In fact, it converges to a non-Normal random variable not centered at 0.
2. $R^2 \to 1$ as $n \to \infty$
3. t-statistics for $\beta \to \infty$ as $n \to \infty$

1.2 Introduction to cointegration

There are a lot of cases that involves cointegrated relations among varialbes in economics; so one should carefully apply this method (NOT ALWAYS).

- Permanent income hypothesis (income and consumption)
- Money demand models (money, income, price and interest rates)
- Growth theories (income, consumption and investment)
- Purchasing power parity (exchange rate, foreign and domestic prices)

And one should also be careful when she is using the word *long-run*; this may imply 1 day in exchange rate market, but may imply 20 years in macroeconomics, so this word can always be misleading.

Cointegration: Suppose $y_{1t} \sim I(1)$ and $y_{2t} \sim I(1)$ and,

$$y_{1t} = \beta_2 y_{2t} + u_t$$

and $u_t \sim I(0)$, then one can say that there is a cointegration and the corresponding cointegration vector is $(1, -\beta_2)^\top$.

- Usually we normalize the first element of the cointegration vector into 1
- The maximum number of the cointegration relation that n-variable system can have is $n-1$
- The number of cointegration relation can be determined by Johansen methods that involve maximum likelihood techniques

Following paragraphs are whiteboard writings.

2 Estimating cointegration relation

There are 2 cases.

1. Pre-specified cointegrating vector
2. Estimated cointegrating vector

2.1 Pre-specified cointegrating vector

If y_{1t} & y_{2t} are two $I(1)$ variables, then in many cases economic theory specifies the nature of cointegrating relation.

For example, if y_{1t} is spot exchange rate & y_{2t} is forward exchange rate, then $y_{1t} - y_{2t}$ should be stationary and this implies that the cointegrating vector is $(1, -1)^\top$.

Similarly, the statinoarity of $\frac{C_t}{Y_t}$ and $\frac{I_t}{Y_t}$ ($c_t = \ln C_t$ and $y_t = \ln Y_t$) implies the cointegration vector is $(1, -1)^\top$ (because $c_t - y_t$ is the log of the ratio, which is stationary as well).

Step 1: In case of pre-specified cointegration vector, we calculate the residuals, i.e. $u_t = y_{1t} - \beta_2 y_{2t}$ where pre-specified cointegration vector is $(1, -\beta_2)^\top$. For y_{1t} & y_{2t} to be cointegrated, the residual has to be stationary.
Step 2: Test for unit root in estimated residual. If we reject the null of unit root, then the two series on cointegrated.

If we test for stationarity of dividend-price ratio or consumption-income ratio, we are implicitly assuming a pre-specified cointegration vector of $(1, -1)^\top$ for dividend and stock price and consumption and income.

(In this framework so far, we are assuming that β_2 is known *a priori*. We just put the alleged β_2 in order to calculate the residual.)

2.2 Unknown cointegrating vector

In case of unknown cointegrating vector, we first estimate the cointegration vector β. **Engle–Granger** first proposed a simple method to estimate β.

If y_{1t} & y_{2t} are two $I(1)$ variables,

1. Run an OLS regression
$$y_{1t} = c + \beta_2 y_{2t} + u_t$$
2. Calculate an estimated residual
$$\hat{u}_t = y_{1t} - \hat{c} - \hat{\beta}_2 y_{2t}$$
& test for unit root in $\hat{u}_t$. Rejection of unit root in $\hat{u}_t$ implies the cointegration

The problem is, macroeconomic variables are by and large endogeneous in a lot of cases. If that is the case, can we use this simple regression in order to estimate the cointegration? Fortunately, the answer is yes according to Watson and it is because of *super-consistency*.

> Endogeneity problem does not lead to inconsistency in OLS estimate of β_2 because $\hat{\beta}_2$ is superconsistent since y_{1t} & y_{2t} are $I(1)$ variables.
> **Superconsistency**: $\hat{\beta}_2$ will converge to β_2 at the rate T rather than $\sqrt{T}$.

Stock and Watson also argued that this method (Engle–Granger method) works nicely only in the case that u_t does not have autocorrelation; that is, if there is a serial correlation problem, then the estimated cointegrating vector is not valid.

β_2 may be serverly biased in finite sample if u_t is serially correlated.

(Researcher usually use the lagged values in analyzing dynamic panels if this kind of autocorrelation problem is likely.) Stock–Watson dynamic OLS (DOLS) approach to estimate the cointegration vector.
Announcement: Next Monday we will finish cointegration, GARCH and then, LASSO and then.

Econometric Methods II Lecture Note 25

Junyong Kim

April 24, 2017

Announcement: Midterm on this Wednesday. Cointegration will not be included. From the beginning to Assignment 4 (e.g. state space representation, trend-cycle decomposition)
Disadvantage of EG method: The problem of Engle–Granger method is that the method does not work well under the presence of an autocorrelation (i.e. $\varepsilon_t = y_t - \mathbf{x}_t^\top \underline{\beta}$).

1 Dynamic OLS to estimate cointegration vector

Similar to Engle–Granger method.
Idea: Add leads & lags of the first difference of the right hand side variable on right hand side.

1. Estimate the following model, i.e.

$$y_{1t} = c + \beta_2 y_{2t} + \sum_{i=1}^{k} \delta_i \Delta y_{2t-i} + \sum_{j=1}^{m} \gamma_i \Delta y_{2t+j} + u_t,$$

   where k & m are chosen based on model selection criteria.

2. Calculate the cointegrating residual, $\hat{\varepsilon}_t = y_{1t} - \hat{c}^{\mathrm{DOLS}} - \hat{\beta}_2^{\mathrm{DOLS}} y_{2t}$.

Note that leads & lags are added only to estimate β_2 consistently. These leads & lags are not included in calculation of cointegrating residuals.

Cointegration involves long-run relationship as well as short-run relationship among variables. Suppose y_{1t} & y_{2t} are $I(1)$ and cointegrated. The long-run equilibrium relationship is $y_{1t} = c + \beta_2 y_{2t} + u_t$. y_{1t} & y_{2t} are cointegrated or have a long-run relationship because residual u_t is stationary.

Any deviation from this long-run relationship in the short-run will disappear in the long-run, and this exhibits error-correction property.

The short-run error correction specification is represented by vector error-correction model (VECM), i.e. VAR in first differences of variables augmented with error correction term. If our specification has VAR(1) representation in first differences, then the VECM specification is represented by,

$$
\begin{aligned}
\Delta y_{1t} =& \overbrace{c_1 + \phi_{11}^1 \Delta y_{1t-1} + \phi_{12}^1 \Delta y_{2t-1}}^{\text{standard VAR specification}} + \alpha_1 \left(y_{1t-1} - \hat{c} - \hat{\beta}_2 y_{2t-1} \right) + \varepsilon_{1t} \\
\Delta y_{2t} =& \underbrace{c_2 + \phi_{21}^1 \Delta y_{1t-1} + \phi_{22}^1 \Delta y_{2t-1}}_{\text{standard VAR specification}} + \alpha_2 \left(y_{1t-1} - \hat{c} - \hat{\beta}_2 y_{2t-1} \right) + \varepsilon_{2t},
\end{aligned}
$$

where $y_{1t-1} - \hat{c} - \hat{\beta}_2 y_{2t-1}$ is the disequilibrium in the long-run relationship last period. α_i represents the error-correction coefficient. If y_{1t} & y_{2t} are cointegrated, then at least one of these error correction coefficients has to be significant. We can also have information about the sign of these error-correction coefficient.

Suppose the long-run relationsihp is $y_{1t} = c + \beta_2 y_{2t} + u_t$. In our example, y_{1t} is LHS variable. Then the correct signs are $\alpha_1 < 0$ & $\alpha_2 > 0$.

Example: Permanent income hypothesis (Cochrane (1994, JPE)). Consumption & income are cointegrated; $c_t = \delta + \beta y_t + u_t$. Then the VECM is

$$
\begin{aligned}
\Delta c_t =& C_1 + \phi_{11}^1 \Delta c_{t-1} + \phi_{12}^1 \Delta y_{t-1} + \alpha_1 \left(c_{t-1} - \hat{\delta} - \hat{\beta} y_{t-1} \right) + \varepsilon_{1t} \\
\Delta y_t =& C_2 + \phi_{21}^1 \Delta c_{t-1} + \phi_{22}^1 \Delta y_{t-1} + \alpha_2 \left(c_{t-1} - \hat{\delta} - \hat{\beta} y_{t-1} \right) + \varepsilon_{2t}.
\end{aligned}
$$

Caveats: If we believe that the consumption walks randomly, then there should be no predictability in the model. The significance of the coefficients implies the rejection of the random-walk hypothesis.

Cochrane finds that α_1 is insignificant, and this is the validation of the random-walk hypothesis of consumption.

Note that,

- VECM can only have one error correction term
- In this case as well, IRFs should revert toward 0 because the differences are $I(0)$

So far, we have only looked at single equation cointegration model with single cointegration vector. Multivariate extension can be handled by using Johansen approach.

2 Cointegration in a multivariate setting

Suppose we have VAR(p) in levels; $Y_t = \Pi_1 Y_{t-1} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t$ and $Y_t \sim I(1)$. In first differences, we can write the above VAR as,

$$\Delta Y_t = \underbrace{\Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t}_{\text{VAR}(p\text{-}1) \text{ with one extra term } \Pi Y_{t-1}},$$

where $\Pi = \Pi_1 + \cdots + \Pi_p - I_m$ and $\Gamma_k = -\sum_{j=k+1}^{p} \Pi_j$ $(k = 1, \cdots, p-1)$.

All the terms in this equation are $I(0)$ except ΠY_{t-1} that contains terms with $I(1)$ variables.

LHS is $\Delta Y_t \sim I(0)$. It must be the case that $\Pi Y_{t-1} \sim I(0)$ and it is only if linear combination of $I(1)$ is $I(0)$.

This contains the cointegrating residuals from last period, i.e. $\Pi = \alpha \beta^\top$ where α is vector of correction coefficients & β is long-run cointegration matrix; Π is referred to as long-run impact matrix.

Example 1: $Y_t = \Pi_1 Y_{t-1} + \varepsilon_t$, i.e. VAR(1) in level, then

$$\Delta Y_t = (\Pi_1 - I) Y_{t-1} + \varepsilon_t = \Pi Y_{t-1} + \varepsilon_t.$$

Example 2: $Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \varepsilon_t$, i.e. VAR(1) in level, then

$$\begin{aligned}
\Delta Y_t =& \Pi_1 Y_{t-1} - Y_{t-1} + \Pi_2 Y_{t-1} - \Pi_2 Y_{t-1} + \Pi_2 Y_{t-2} + \varepsilon_t \\
=& (\Pi_1 + \Pi_2 - I) Y_{t-1} - \Pi_2 (Y_{t-1} - Y_{t-2}) + \varepsilon_t \\
=& \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \varepsilon_t
\end{aligned}$$

The general form for the level VAR(p) is

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t,$$

and Rank(Π) provides us information about the number of cointegrating vector.

Suppose there is no cointegration $\Pi = 0$ & Rank(Π) $= 0$. For cointegration, $0 <$ Rank(Π) $= r < n$. If Rank(Π) $= n$, all variables are stationary. Not a full-rank matrix implies Det(Π) $= 0$. And also Π is also a square matrix. Determinant is the product of eigenvalues. At least one of the eigenvalues is 0.

Johansen's method of estimating for number of cointegrating vector uses the property of Π matrix to obtain and estimate the cointegration vector; 2 methods.

1. Trace statistic

2. Maximum eigenvalue statistic

Econometric Methods II Lecture Note 26

Junyong Kim

May 1, 2017

Announcement: Exam result ($\bar{x} = 38$, $\hat{F}^{-1}(0.5) = 35$, $s = 12$)

1 Johansen cointegration test

$$\begin{aligned}\text{Rank}(\Pi) =& \gamma < n \\ =& \text{number of cointegrating vectors,}\end{aligned}$$

and since Π is not a full rank matrix, determinant of Π is 0, and this is the product of all eigenvalues, i.e. at least one of the eigenvalues is zero.

There are 2 types of test proposed by Johansen,

1. Trace test
2. Maximum eigenvalue test

and both of these tests are based on the eigenvalues of Π matrix.

1.1 Trace test

$$\begin{aligned}H_0(r) :& r = r_0, \quad \text{no. of cointegrating vectors} = r_0 \\ H_1(r) :& r > r_0,\end{aligned}$$

and the LR statistic called trace statistic,

$$LR_{\text{trace}}\left(r_0\right) = -T \sum_{i=r_0+1}^{n} \ln\left(1 - \hat{\lambda}_i\right),$$

where $\hat{\lambda}_i$s are the eigenvalues.

- If rank(Π) $= r_0$, then $\hat{\lambda}_{r_0+1}, \cdots, \hat{\lambda}_n$ are close to zero. Since there are only r_0 eigenvalues that are non-zero & we sort eigenvalues in descending order. That is, $\ln\left(1 - \hat{\lambda}_i\right) \approx 0$, so the null is not rejected.
- If rank(Π) $> r_0$, then $\ln\left(1 - \hat{\lambda}_i\right) \neq 0$, so the null will be rejected. The asymptotic distribution of trace statistic is not chi-square but a multivariate version of Dickey–Fuller distribution. This depends on $n - r_0$ and type of deterministic trend.

1.1.1 Sequential procedure for determining number of cointegrating vectors

1. First, test $H_0(r) : r = 0$ vs. $H_1(r) : r > 0$. If the null hypothesis is rejected, then go to step 2 and if not rejected, we find no evidence of cointegration

2. Test $H_0(r) : r = 1$ vs. $H_1(r) : r > 1$. Follow the same steps until $H_0(r) : r = n-1$ vs. $H_1(r) : r > n-1$

3. This sequential procedure is repeated until the null hypothesis is no longer rejected

Test statistic's asymptotic distribution is the trace of a matrix based on function of Brownian motion. It is not based on trace of Π.

So, the procedure is

1. Run the VAR using level variables and obtain the optimal p

2. Then run the VEC using differenced variables with the lag $p-1$ (refer the previous notes), and obtain $\Pi = \alpha\beta^\top$

3. By using this Π one can conduct trace or maximum eigenvalue tests

If one knows the exact structure of cointegration, then the former approaches are better than Johansen approach, because Johansen approach imposes a lot of restriction. Usually, researchers conduct such as Engle–Granger or Stock–Watson first and then Johansen in order to verify their results.

1.2 Maximum eigenvalue test

$$\begin{aligned} H_0 :& r = r_0 \\ H_1 :& r = r_0 + 1, \end{aligned}$$

and $LR_{\max} = -T \ln\left(1 - \hat{\lambda}_{r_0+1}\right)$. Asymptotic distribution is a complicated function of Brownian motion & depends on $n - r_0$ and type of deterministic trend.

Pros and cons: Trace test is more consistent, more powerful and encompasses maximum eigenvalue test. In contrast, maximum eigenvalue test exhibits proper sizes, i.e. in finite sample, trace test tends to over- or underreject the null under the null (Lütkepohl),

- Trace: Better in terms of power

- Maximum eigenvalue: Better in terms of size

2 Example

Suppose one has

$$\begin{aligned}\frac{M_t}{P_t} =& \beta_0 + \beta_1 Y_t + \beta_2 \gamma_t + \varepsilon_t \\ \frac{M_t}{P_t}, Y_t, \gamma_t \sim& I(1),\end{aligned}$$

and test whether $\hat{\varepsilon}_t$ has an unit root or not assuming one cointegration vector. To check whether there exists only one cointegration vector, one can perform Johansen test.

3 `R` session

We studied

- The libraries `urca`, `vars`, `MASS`
- Analysis using `st` and `ft`

Announcement: In the next class, we will study

- the rest of `R` example
- GARCH little bit
- LASSO little bit

Short notes after class

Trace and maximum eigenvalue tests are conducted not based on the eigenvalues of Π, but based on $S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}$, which is described more nicely in Helmut Lütkepohl's *New introduction to multiple time series analysis.* Since the latter matrix is symmetric and positive-definite, corresponding eigenvalues are always real and positive. The theoretical distributions of trace and eigenvalue test statistics are

$$
\begin{aligned}
\text{Trace} =& \lambda_{LR}\left(r_0, K\right) \\
=& 2\left[\ln \hat{L}\left(r_1\right) - \ln \hat{L}\left(K\right)\right] \\
=& -T \sum_{i=r_0+1}^{K} \ln\left(1-\lambda_i\right) \\
\xrightarrow{d}& \operatorname{tr}\left(\left(\int_0^1 \mathbf{W} d\mathbf{W}^\top\right)^\top \left(\int_0^1 \mathbf{W}\mathbf{W}^\top ds\right)^{-1} \left(\int_0^1 \mathbf{W} d\mathbf{W}^\top\right)\right) \\
\text{Max} =& \lambda_{LR}\left(r_0, r_0+1\right) \\
=& 2\left[\ln \hat{L}\left(r_1\right) - \ln \hat{L}\left(r_0+1\right)\right] \\
=& -T \ln\left(1-\lambda_{r_0+1}\right) \\
\xrightarrow{d}& \max\left(\left(\int_0^1 \mathbf{W} d\mathbf{W}^\top\right)^\top \left(\int_0^1 \mathbf{W}\mathbf{W}^\top ds\right)^{-1} \left(\int_0^1 \mathbf{W} d\mathbf{W}^\top\right)\right),
\end{aligned}
$$

where $\mathbf{W} = \mathbf{W}_{K-r_0}(s)$ is $(K-r_0)$-dimensional standard Wiener process, and $\max(\cdot)$ is the maximum eigenvalue of the matrix.

Note that, under the existence of unit root, t statistic of the autoregressive parameter has the asymptotic distribution

$$
\hat{t} \text{ of } \hat{\phi} \xrightarrow{d} \frac{\int_0^1 W dW}{\sqrt{\int_0^1 W^2 ds}}, \quad \text{if } \phi = 1
$$

where W is the standard Wiener process.

Also note that, if there is no unit root (i.e. under the stationarity assumption), the statistic is consistent and asymptotically Normal, i.e.

$$
\sqrt{T}\left(\hat{\phi} - \phi\right) \xrightarrow{d} \mathrm{N}\left(\phi, 1-\phi^2\right), \quad \text{if } -1 < \phi < 1.
$$

Econometric Methods II Lecture 27

Junyong Kim

May 3, 2017

1 VECM revisited

1. Estimate VAR in levels

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t$$

   to get the optimal lag p.

2. Estimate VECM in first differences with the lag $p-1$

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

   where $\Pi = (\Pi_1 + \Pi_2 + \cdots + \Pi_p - I)$ (ΠY_{t-1} contains $I(1)$ variables).

It can be shown that $\Pi = \alpha\beta^\top$, where α is vector of error correction coefficients & β is cointegration vector. Note that $\beta^\top Y_{t-1}$ is $I(0)$ since it is a linear combination of $I(1)$ variables with cointegration vector β.
Our example: 2 variables s_t & f_t, so

1. We found optimal lag=1 (VAR(1)), i.e.

$$\begin{aligned} s_t =& \pi_{11} s_{t-1} + \pi_{12} f_{t-1} + \varepsilon_{1t} \\ f_t =& \pi_{21} s_{t-1} + \pi_{22} f_{t-1} + \varepsilon_{2t} \end{aligned}$$

   and the VECM representation VAR(0) with ΠY_{t-1} is

$$\begin{aligned} \Delta s_t =& \alpha_1 (s_{t-1} - \beta_2 f_{t-1}) + \varepsilon_{1t} \\ \Delta f_t =& \alpha_2 (s_{t-1} - \beta_2 f_{t-1}) + \varepsilon_{2t} \end{aligned}$$

   and for estimated cointegration vector $\beta = (1, -\beta_2)^\top$ or pre-specified vector $\beta = (1, -1)^\top$.

1.1 VECM `R` session

In `urca` package, `ca.jo` function will conduct Johansen's method. The problem of this function is that VECM(1) (i.e. only ΠY_{t-1} and not $\Gamma_1 \Delta Y_{t-1}$) is unavailable. If one wants to use VECM(1), then she can consider another package called `tsdyn` (nonlinear time-series dynamics).

In `R` example (see 250th and 254th lines of `coint.R` codes), one can reject the null of $r_0 = 0$ at 5% level according to the trace test, but she can reject the null only at 10% level according to the maximum eigenvalue test. Usually these two tests reconcile with each other, but if there is a conflict, then following the result from the trace test is more usual.

One can also use `cajorls` function to exploit OLS approach. And one can consider `blrtest` to conduct likelihood ratio test.

2 GARCH family

Firstly proposed by Engle (1982) (he was in UCSD originally, but moved to NYU now). According to efficient market hypothesis, stock returns cannot be predicted, i.e.

$$\begin{aligned} P_t =& P_{t-1} + \varepsilon_t \\ r_t =& \ln P_t - \ln P_{t-1} \\ =& c + \varepsilon_t' \\ =& c + \beta r_{t-1} + \nu_t, \quad \nu_t \sim \mathrm{N}\left(0, \sigma_{\nu_t}^2\right) \end{aligned}$$

and β in the last expression should be equal to zero. Engle, in contrast, says that volatilities at least can be predicted (though returns are not predictable).

2.1 `R` session with `rugarch`

We studied how to use `rugarch` package to exploit univariate GARCH models.

Conditional mean equation: $y_t = c + \beta y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathrm{N}\left(0, \sigma_{\varepsilon_t}^2\right)$ (not iid anymore), which implies conditional heteroskedasticity.

This means, the level of ε is not predictable, but the square of ε is somewhat predictable (volatility clustering, highly meaningful results from ACF & PACF). Hence we can consider Ljung–Box test primarily to see whether GARCH is applicable.

2.2 ARCH & GARCH models

- ARCH: Autoregressive conditional heteroskedasticity
- GARCH: Generalized ARCH

Suppose

$$
\begin{aligned}
y_t =& c + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}\left(0, \sigma_t^2\right) \\
\sigma_t^2 =& a_0 + a_1 \varepsilon_{t-1}^2, \quad \text{ARCH(1) representation} \\
\sigma_t^2 =& a_0 + a_1 \varepsilon_{t-1}^2 + \cdots + a_p \varepsilon_{t-p}^2, \quad \text{ARCH}(p) \text{ representation.}
\end{aligned}
$$

We will finish this topic in the next session.

Econometric Methods II Lecture Note 28

Junyong Kim

May 8, 2017

Announcement: This is the second-last lecture.

1 ARCH models again

ARCH(1):

$$\begin{aligned}
y_t =& c + \phi(L) y_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}\left(0, \sigma_t^2\right) \\
\sigma_t^2 =& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad \text{ARCH(1)} \\
=& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2, \quad \text{ARCH}(p) \\
\mathrm{E}_{t-1}\left[y_t\right] =& c + \phi(L) y_t, \quad \text{conditional mean specification.}
\end{aligned}$$

This ARCH models are conditionally heteroskedastic but unconditionally homoskedastic.

$$\begin{aligned}
\mathrm{E}_{t-1}\left[\sigma_t^2\right] =& \mathrm{E}_{t-1}\left[\alpha_0 + \alpha_1 \varepsilon_{t-1}^2\right] \\
=& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad \text{time-varying conditional volatility.}
\end{aligned}$$

In contrast, the unconditional variance $\mathrm{E}\left[\sigma_t^2\right] = \mathrm{E}\left[\alpha_0 + \alpha_1 \varepsilon_{t-1}^2\right] = \bar{\sigma}^2$, i.e.

$$\begin{aligned}
\bar{\sigma}^2 =& \alpha_0 + \alpha_1 \mathrm{E}\left[\varepsilon_{t-1}^2\right] \\
=& \alpha_0 + \alpha_1 \bar{\sigma}^2 \\
=& \frac{\alpha_0}{1 - \alpha_1}, \quad \text{unconditional variance.}
\end{aligned}$$

Unconditional variance is time invariant, i.e. unconditional homoskedasticity.

ARCH models: AR representation in volatility. The ARCH(1) example is

$$\begin{aligned}
\sigma_t^2 =& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \\
\sigma_t^2 + \varepsilon_t^2 =& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \varepsilon_t^2 \\
\varepsilon_t^2 =& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \left(\varepsilon_t^2 - \sigma_t^2\right) \\
=& \underbrace{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \nu_t}_{\text{AR(1) representation}}, \quad \nu_t = \varepsilon_t^2 - \sigma_t^2.
\end{aligned}$$

2 GARCH models (Bollerslev)

GARCH models generalize ARCH specification (ARCH & AR, GARCH & ARMA).

$$\begin{aligned}
y_t =& c + \phi(L) y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}\left(0, \sigma_t^2\right) \\
\sigma_t^2 =& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\
=& \underbrace{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2}_{\text{GARCH}(p,q) \text{ specification}}.
\end{aligned}$$

Note that for stationarity of ARCH(1) model, we need $|\alpha_1| < 1$ and also $\alpha_0 > 0$. For GARCH(1,1) specification $|\alpha_1 + \beta_1| < 1$ guarantees stationarity. For GARCH(1,1) specification

$$\begin{aligned}
\sigma_t^2 =& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
\mathrm{E}_t\left[\sigma_t^2\right] =& \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
\mathrm{E}\left[\sigma_t^2\right] =& \bar{\sigma}^2 \\
=& \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.
\end{aligned}$$

In addition, in GARCH models, to identify GARCH parameter β, we need a significant ARCH parameter α. If ARCH parameter α is not significant, then it has been suggested to simply use ARCH(p) model specification in place of GARCH.

3 Testing for ARCH effect

1. Fit a conditional mean equation and get the residuals $\hat{\varepsilon}_t$.

2. Run a regression of squared residuals on its own lags, i.e.

$$\hat{\varepsilon}_t^2 = \beta_0 + \beta_1 \hat{\varepsilon}_{t-1}^2 + \cdots + \beta_p \hat{\varepsilon}_{t-p}^2 + \nu_t.$$

   Here the null is $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ and the alternative is $\exists \beta_i \neq 0$ for $i \in \{1, \cdots, p\}$. The test statistic is $LM_{ARCH} = T \cdot R^2 \sim \chi_p^2$.

One can also consider a multivariate GARCH model for multivariate regression models. GARCH-style modelS are widely used in finance area (option pricing, optimal hedge ratio, etc.). The textbook written by Walter Enders (*Applied Econometric Time-series*) for multivariate GARCH models.

4 Further issues

We covered `GARCH.pdf`. We are modeling

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

So $\frac{\partial \sigma_t^2}{\partial \varepsilon_{t-1}^2} = \alpha_1$ for both $\varepsilon_{t-1} \geq 0$ and $\varepsilon_{t-1} < 0$. However usually the effect is asymmetric due to the *leverage effect*.

Today, we studied

- `Box.test`
- Pure GARCH (`ugarchspec`, `ugarchfit`)
- EGARCH
- TGARCH
- Sign bias test
- News impact curve

and saw several `R` programming examples.

Econometric Methods II Lecture Note 29

Junyong Kim

May 10, 2017

> **Subset selection**: The procedure of choosing several important variables
> **Shrinkage**: The procedure of depriving the magnitude of several coefficients that are not that important

1 If there are too many variables...

We studied

- Stepwise regression (forward, backward)
- LASSO (Least absolute shrinkage and selection operator)
- Ridge regression
- Leave One Out Cross Validation (LOOCV)
- Least angle regression (LAR)
- Recommended book: *Introduction to statistical learning*, free in the internet. `R`-based. `glmnet`, `lars`...
- The use of `R` package to conduct LASSO, Ridge, etc...

2 Existence of structural break

- If we know the specific date, then we can use Chow, Wald tests
- If we don't know, then the supremum of Wald statistics (among all possible candidate dates) (Bai & Perron test?)

Slides will be delivered.

Econometric Methods II Assignment 01

Junyong Kim*

February 8, 2017

1 Analytical Exercise

1. $Y_t = \left(1 + 2.4L + 0.8L^2\right)\varepsilon_t$ and $\varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$.

$$\begin{aligned} &\Rightarrow \theta(z) = 0.8z^2 + 2.4z + 1 \\ &\Rightarrow z_i = \tfrac{-1.2 \pm \sqrt{1.44 - 0.8}}{0.8} \\ &\quad = \tfrac{-1.2 \pm 0.8}{0.8} \in \{-\tfrac{1}{2}, -\tfrac{5}{2}\} \end{aligned}$$

Since $|z_1| = |-\frac{1}{2}| < 1$, this MA(2) process is not invertible. Instead, one can find Hansen–Sargent invertible representation of this process as follows [Hamilton, 1994].

$$1 + 2.4L + 0.8L^2 = (1 + \tfrac{2}{5}L)(1 + 2L)$$

$$\Rightarrow Y_t = (1 + \tfrac{2}{5}L)(1 + 2L)\varepsilon_t, \quad \text{non-invertible} \tag{1}$$

$$\text{Cov}\left[\varepsilon_t, \varepsilon_{t-j}\right] = \begin{cases} \sigma^2, & j = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} \Rightarrow Y_t^* &= (1 + \tfrac{2}{5}L)(1 + \tfrac{1}{2}L)\varepsilon_t^*, \quad \text{invertible} \qquad (2) \\ &= \varepsilon_t^* + 0.9\varepsilon_{t-1}^* + 0.2\varepsilon_{t-2}^* \end{aligned}$$

$$\text{Cov}\left[\varepsilon_t^*, \varepsilon_{t-j}^*\right] = \begin{cases} 4\sigma^2, & j = 0 \\ 0, & \text{otherwise} \end{cases}$$

The autocovariance-generating functions of $\{Y_t\}$ in (1) and $\{Y_t^*\}$ in (2) are identical. $\square$

$$\text{Cov}\left[Y_t^*, Y_{t-j}^*\right] = \text{E}\left[\left(Y_t^* - \mu^*\right)\left(Y_{t-j}^* - \mu^*\right)\right] = \begin{cases} 7.4\sigma^2, & j = 0 \\ 4.32\sigma^2, & j = 1 \\ 0.8\sigma^2, & j = 2 \\ 0, & j \in \{3, 4, \ldots\} \end{cases}$$

2. $y_t = 2.5 + 1.1y_{t-1} - 0.18y_{t-2} + \varepsilon_t$, where $\varepsilon_t \sim \text{iid}(0, 1)$

   (a) Stability

$$\begin{aligned} \left(1 - 1.1L + 0.18L^2\right)y_t &= 2.5 + \varepsilon_t \\ \Rightarrow \phi(z) &= 0.18z^2 - 1.1z + 1 \\ \Rightarrow z_i &= \tfrac{1.1 \pm \sqrt{1.21 - 0.72}}{0.36} \in \{5, \tfrac{10}{9}\} \end{aligned}$$

   Since all roots are greater than 1, this process is stable. $\square$

*Lubar School of Business, University of Wisconsin–Milwaukee.

Moments

$$
\begin{aligned}
\mathrm{E}\left[y_t\right] &= \mathrm{E}\left[2.5+1.1y_{t-1}-0.18y_{t-2}+\varepsilon_t\right] \\
&= 2.5+1.1\mathrm{E}\left[y_{t-1}\right]-0.18\mathrm{E}\left[y_{t-2}\right]+\cancel{\mathrm{E}\left[\varepsilon_t\right]} \\
&= 2.5+1.1\mathrm{E}\left[y_t\right]-0.18\mathrm{E}\left[y_t\right] \\
&= \tfrac{2.5}{1-1.1+0.18} = \tfrac{2.5}{0.08} = 31.25 = \mu, \quad \text{mean} \\
y_t-\mu &= 1.1\left(y_{t-1}-\mu\right)-0.18\left(y_{t-2}-\mu\right)+\varepsilon_t \\
\Rightarrow \mathrm{Cov}\left[y_t, y_{t-j}\right] &= 1.1\mathrm{Cov}\left[y_{t-1}, y_{t-j}\right]-0.18\mathrm{Cov}\left[y_{t-2}, y_{t-j}\right]+\cancel{\mathrm{Cov}\left[\varepsilon_t, y_{t-j}\right]} \\
\Rightarrow \gamma_j &= 1.1\gamma_{j-1}-0.18\gamma_{j-2} \\
\Rightarrow \rho_j &= 1.1\rho_{j-1}-0.18\rho_{j-2} \\
\rho_1 &= \tfrac{1.1}{1.18} \approx 0.9322 \\
\rho_2 &= \tfrac{1.1^2}{1.18}-0.18 \approx 0.8454 \\
\therefore \quad \rho_j &= \begin{cases} 1 & j=0 \\ \tfrac{1.1}{1.18}, & j=1 \\ \tfrac{1.21}{1.18}-0.18 & j=2 \\ 1.1\rho_{j-1}-0.18\rho_{j-2} & j\in\{3,4,\dots\} \end{cases}, \quad \text{autocorrelation} \\
\Rightarrow \mathrm{Var}\left[y_t\right] &= 1.1\mathrm{Cov}\left[y_t, y_{t-1}\right]-0.18\mathrm{Cov}\left[y_t, y_{t-2}\right]+\mathrm{Cov}\left[y_t, \varepsilon_t\right] \\
\gamma_0 &= 1.1\rho_1\gamma_0-0.18\rho_2\gamma_0+\sigma^2 \\
&= \tfrac{\sigma^2}{1-1.1\rho_1+0.18\rho_2} \approx \tfrac{1}{1-1.1\times 0.9322+0.18\times 0.8454} \approx 7.8894, \quad \text{variance} \\
\therefore \quad \gamma_j &= \rho_j\gamma_0, \quad \text{autocovariance}
\end{aligned}
$$

ACF of this AR(2) process will decline exponentially. In contrast, PACF of this process will converge to zero after the second lag. One is able to compute mth partial autocorrelation $\alpha_m^{(m)}$ using the following matrices. PACF is the collection of $\left\{\alpha_i^{(i)}\right\}$.

$$
\begin{pmatrix} \alpha_1^{(m)} \\ \vdots \\ \alpha_m^{(m)} \end{pmatrix} = \begin{pmatrix} \gamma_0 & \cdots & \gamma_{m-1} \\ \vdots & \ddots & \vdots \\ \gamma_{m-1} & \cdots & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_m \end{pmatrix}
$$

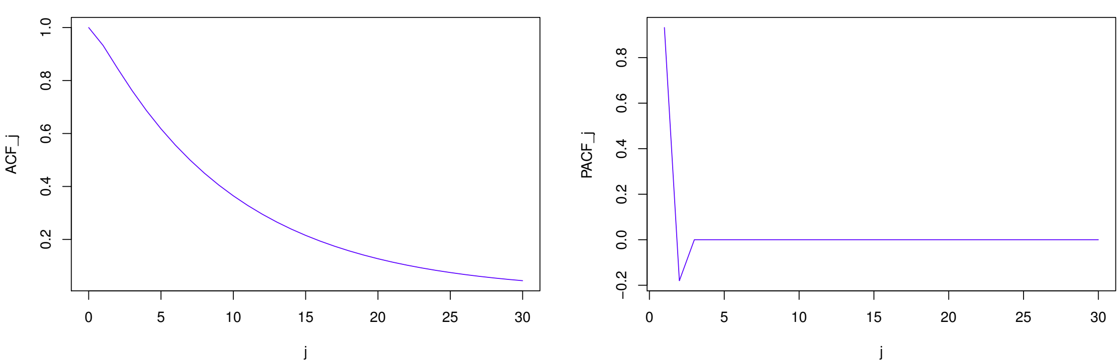


Figure 1: Correlogram (ACF) and partial correlogram (PACF)

(b) $\phi(L) = 1 - 1.1L + 0.18L^2$ and $\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \cdots$. Therefore, $\phi(L)^{-1} = \psi(L)$ and $\psi(L)\phi(L) = \left(1 + \psi_1 L + \psi_2 L^2 + \cdots\right)\left(1 - 1.1L + 0.18L^2\right) = 1$. By comparing coefficients,

$$\psi_0 = 1, \quad \psi_1 = \phi_1 = 1.1$$
$$\psi_2 = \psi_1\phi_1 + \psi_0\phi_2 = 1.1^2 - 0.18 = 1.03$$
$$\psi_3 = \psi_2\phi_1 + \psi_1\phi_2 = 1.03 \times 1.1 - 1.1 \times 0.18 = 0.935$$
$$\psi_4 = \psi_3\phi_1 + \psi_2\phi_2 = 0.935 \times 1.1 - 1.03 \times 0.18 = 0.8431$$

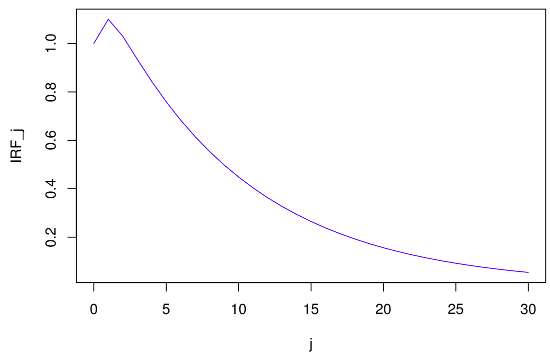


Figure 2: Impulse response function

2 Empirical Exercise

2.1 Exercise 1: Identify ARMA models for log real GDP

1. (a) *lrgdp* (Y hereafter) is defined as log level of real GDP. What is unusual here is that Y is non-stationary. According to the following correlogram, Y is very persistent.

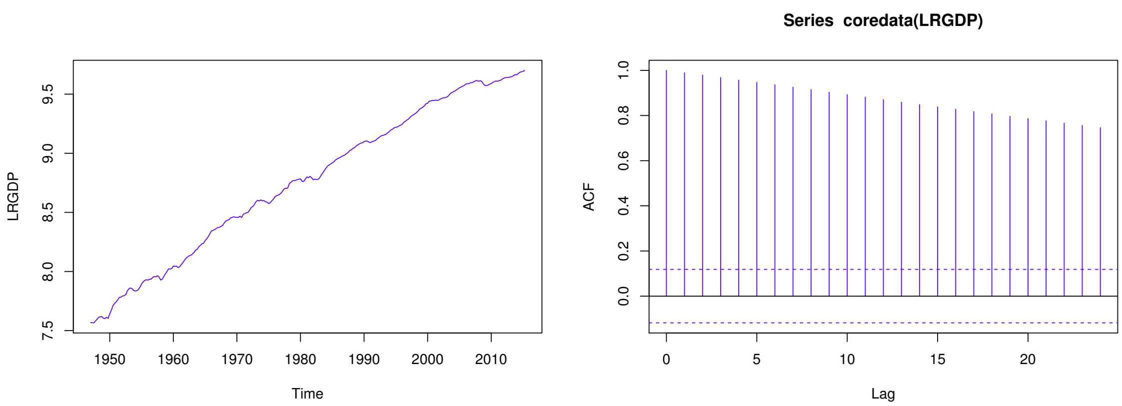


Figure 3: $\{Y_t\}$ and correlogram

Parameter	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8
Estimate	1.00	0.99	0.98	0.97	0.96	0.95	0.94	0.93	0.91

(b) $Y_t = c + \phi Y_{t-1} + \varepsilon_t$

Parameter	Estimate	St. Err.
c	0.0296	0.0079
ϕ	0.9975	0.0009

$Y_t - \mu = c + \phi\left(Y_{t-1} - \mu\right) + \varepsilon_t$

Parameter	Estimate	St. Err.
c	0.0078	0.0006
ϕ	0.9975	0.0009

- In the first regression, c is the numerator of the unconditional mean of Y_t
- Corresponding unconditional mean is $\mathrm{E}\left[Y_t\right] = \frac{c}{1-\phi}$
- In the second regression, however, c is the numerator of the unconditional mean of $Y_t - \mu$
- Corresponding unconditional mean is $\mathrm{E}\left[Y_t - \mu\right] = 0$ hence $c = 0$
- Technically, $\hat{c}$ in the second regression is exactly 0 when dependent variable $\{Y_2, Y_3, \cdots, Y_T\}$ and independent variable $\{Y_1, Y_2, \cdots, Y_{T-1}\}$ are demeaned with respective sample means
- If dependent and independent variables are demeaned with the identical sample mean $\bar{Y}$, then $\hat{c} = \frac{1}{T-1}\sum_{t=2}^{T} Y_t^{\text{demean}} - \hat{\phi}\frac{1}{T-1}\sum_{t=1}^{T-1} Y_t^{\text{demean}} \neq 0$
- This is the source of $\hat{c} \neq 0$ in the second regression
- Since this process is highly persistent with $\hat{\phi} \approx 1$, one must suspect a unit root

(c) $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$ and $Y_t^{\text{detrend}} = Y_t - \hat{\beta}_0 - \hat{\beta}_1 t = Y_t - 7.6512 - 0.0080t$

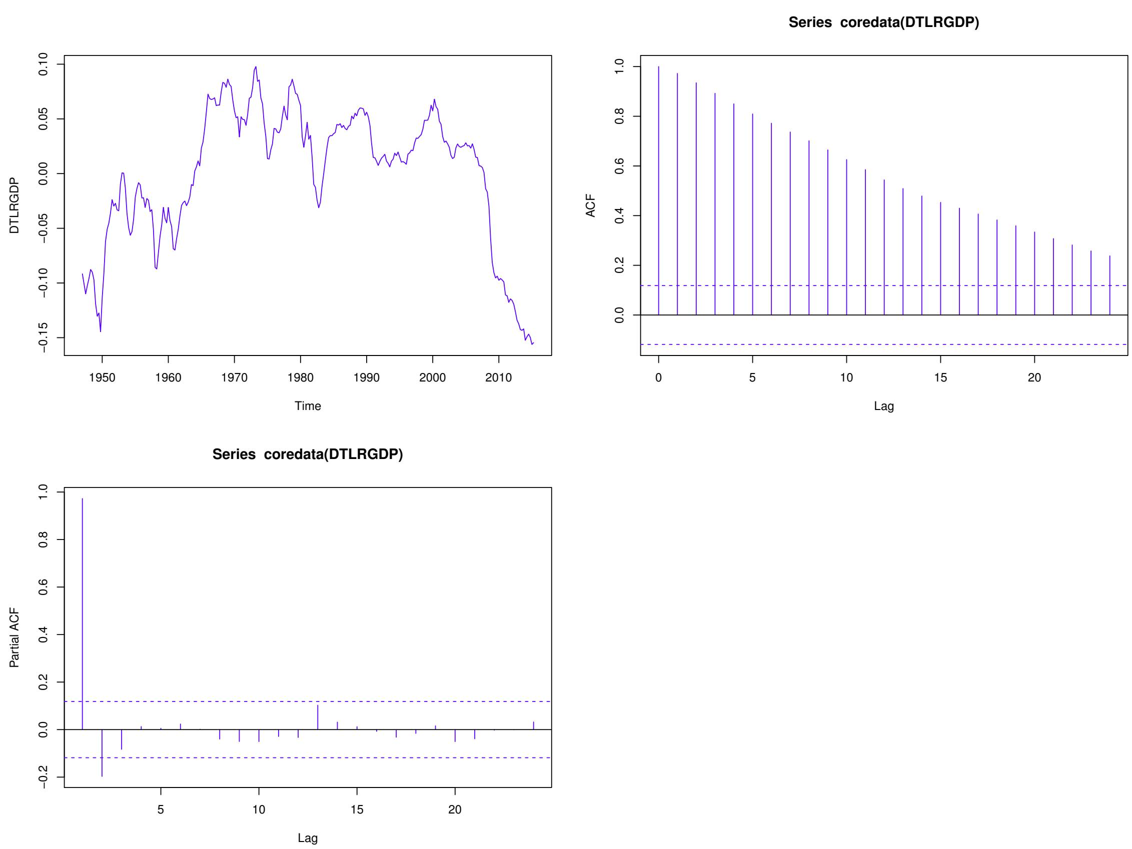


Figure 4: $\left\{Y_t^{\text{detrend}}\right\}$, ACF and PACF

Parameter	Estimate	St. Err.
β_0	7.6512	0.0075
β_1	0.0080	0.0000

$\{Y_t^{\text{detrend}}\}$ is still persistent, but less persistent than $\{Y_t\}$. While ρ_{10} for $\{Y_t\}$ is about 0.89, that of $\{Y_t^{\text{detrend}}\}$ is about 0.63. One is able to suspect that $\{Y_t\}$ is trend-stationary. Since only two partial autocorrelation coefficients are significant, one can consider AR(2) model at a glance.

Lag	0	1	2	3	4	5	6	7	8
ACF	1.00	0.97	0.93	0.89	0.85	0.81	0.77	0.74	0.70
PACF	N/A	0.97	−0.20	−0.08	0.01	0.01	0.02	0.00	−0.04

(d) $\Delta Y_t = Y_t - Y_{t-1}$

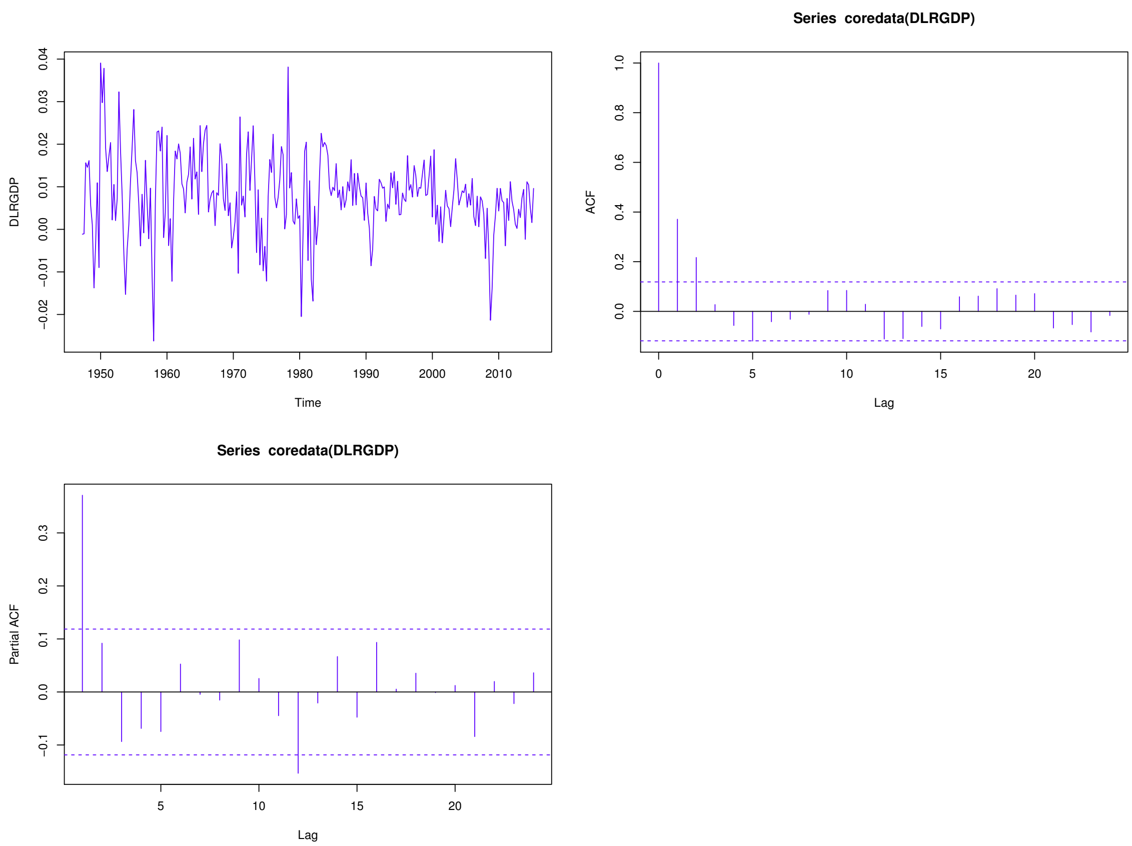


Figure 5: $\{\Delta Y_t\}$, ACF and PACF

Lag	0	1	2	3	4	5	6	7	8
ACF	1.00	0.37	0.22	0.03	−0.06	−0.12	−0.04	−0.03	−0.01
PACF	N/A	0.37	0.09	−0.09	−0.07	−0.07	0.05	0.00	−0.02

What is noticeable here is the significance of the twelfth partial correlation coefficient. ρ_{12} is -0.11 and $\rho_{12}^{\text{partial}}$ is -0.15, respectively. Hence one can consider ARMA({1,12},0) model at a glance.
In addition, one is able to find a periodic behavior in the correlogram, though it is insignificant after the third lag.
Here I use ARMA($\{p\}$,$\{q\}$) rather than ARMA(p,q) to refer the model $Y_t = c + \phi Y_{t-p} + \varepsilon_t + \theta\varepsilon_{t-q}$. In contrast, I use ARMA(p,q) rather than ARMA($\{p\}$,$\{q\}$) to refer the model $Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$.

(e) For $\{Y_t^{\text{detrend}}\}$ first,

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

p	q	AIC	Rank	BIC	Rank
0	0	−746.92	16	−739.70	16
0	1	−1,071.93	15	−1,061.09	15
0	2	−1,320.62	14	−1,306.17	14
0	3	−1,454.07	13	−1,436.00	13
1	0	−1,762.50	12	−1,751.66	12
1	1	−1,791.30	11	−1,776.85	7
1	2	−1,792.02	10	−1,773.95	9
1	3	−1,803.58	1	−1,781.90	4
2	0	−1,802.33	5	−1,787.88	1
2	1	−1,802.30	6	−1,784.24	3
2	2	−1,803.25	3	−1,781.57	5
2	3	−1,798.18	8	−1,772.89	10
3	0	−1,803.35	2	−1,785.29	2
3	1	−1,802.54	4	−1,780.86	6
3	2	−1,800.59	7	−1,775.30	8
3	3	−1,794.39	9	−1,765.48	11

According to AICs, the best three models are ARMA(1,3), ARMA(3,0) and ARMA(2,2), respectively. According to BICs, on the other hand, the best three models are ARMA(2,0), ARMA(3,0) and ARMA(2,1), respectively. By and large, AIC is more generous than BIC in selecting longer models.

For $\{\Delta Y_t\}$ second,

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

p	q	AIC	Rank	BIC	Rank
0	0	−1,761.41	16	−1,754.19	16
0	1	−1,789.70	15	−1,778.87	11
0	2	−1,800.71	4	−1,786.27	2
0	3	−1,800.80	3	−1,782.75	5
1	0	−1,799.84	9	−1,789.01	1
1	1	−1,799.31	11	−1,784.87	4
1	2	−1,800.39	6	−1,782.34	7
1	3	−1,798.83	12	−1,777.17	13
2	0	−1,800.19	8	−1,785.76	3
2	1	−1,799.59	10	−1,781.54	8
2	2	−1,801.48	2	−1,779.82	9
2	3	−1,797.83	14	−1,772.56	14
3	0	−1,800.70	5	−1,782.66	6
3	1	−1,800.33	7	−1,778.67	12
3	2	−1,804.84	1	−1,779.57	10
3	3	−1,798.01	13	−1,769.13	15

According to AICs, the best three models are ARMA(3,2), ARMA(2,2) and ARMA(0,3), respectively. According to BICs, on the other hand, the best three models are ARMA(1,0), ARMA(0,2) and ARMA(2,0), respectively. Similarly, AIC is more generous than BIC in selecting longer models by and large.

(f) $\{Y_t^{\text{detrend}}\}_{t=1947:1}^{2015:2}$ and $\{\Delta Y_t\}_{t=1947:2}^{2015:2}$ are decomposed into two different parts.

- $\left\{Y_t^{\text{detrend}}\right\}_{t=1947:1}^{1981:4}$ and $\left\{Y_t^{\text{detrend}}\right\}_{t=1982:1}^{2015:2}$
- $\{\Delta Y_t\}_{t=1947:2}^{1981:4}$ and $\{\Delta Y_t\}_{t=1982:1}^{2015:2}$
- Since $\{\Delta Y_t\}$ is the first-difference of $\{Y_t\}$, $\Delta Y_{1947:1} = Y_{1947:1} - Y_{1946:4}$ is not available

$\left\{Y_t^{\text{detrend}}\right\}$	Lag	0	1	2	3	4	5	6	7	8
1947:1	ACF	1.00	0.97	0.93	0.88	0.84	0.80	0.77	0.73	0.70
–1981:4	PACF	N/A	0.97	−0.26	−0.11	0.11	0.06	0.04	−0.06	0.00
1982:1	ACF	1.00	0.97	0.94	0.90	0.86	0.82	0.78	0.74	0.70
–2015:2	PACF	N/A	0.97	−0.06	−0.08	−0.06	−0.06	0.03	−0.04	−0.04

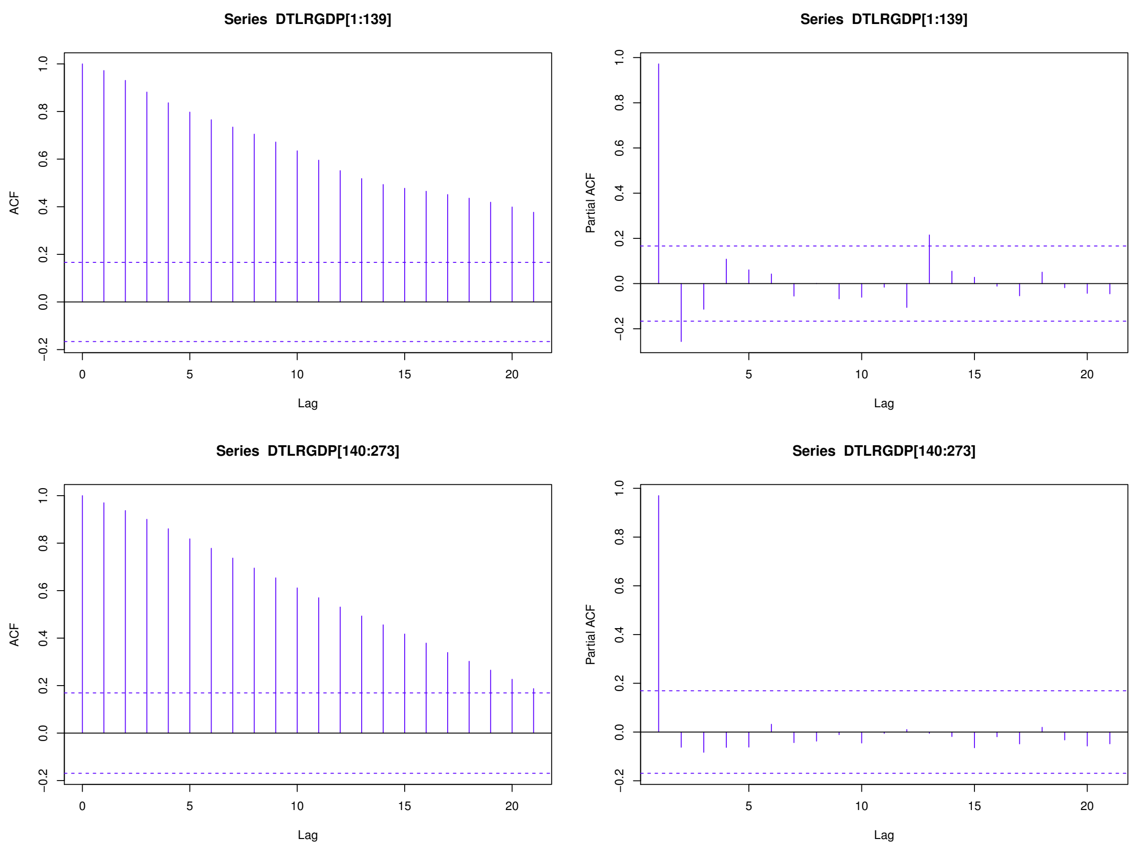


Figure 6: ACF and PACF of $\left\{Y_t^{\text{detrend}}\right\}$: 1947:1–1981:4 and 1982:1–2015:2

- What is noticeable here is that $\left\{Y_t^{\text{detrend}}\right\}$ is more persistent in the first sample rather than the second one. ρ_{20} is about 0.40 in the first sample, but it is about 0.23 in the second sample
- According to the first partial correlogram, the first (0.97), second (−0.26) and thirteenth (0.22) lags are significant. Therefore, one can consider ARMA({1,2,13},0) model at a glance
- According to the second partial correlogram, however, only the first (0.97) lag is significant. Therefore, one can consider ARMA(1,0) model at a glance
- By and large, $\left\{Y_t^{\text{detrend}}\right\}$ is less persistent in the recent sample

$\{\Delta Y_t\}$	Lag	0	1	2	3	4	5	6	7	8
1947:2	ACF	1.00	0.32	0.14	−0.06	−0.12	−0.17	−0.10	−0.09	−0.04
–1981:4	PACF	N/A	0.32	0.03	−0.13	−0.08	−0.11	−0.01	−0.05	−0.02
1982:1	ACF	1.00	0.49	0.38	0.21	0.16	−0.02	−0.03	−0.02	−0.04
–2015:2	PACF	N/A	0.49	0.18	−0.04	0.03	−0.17	0.00	0.06	−0.03

- One is able to find a strong periodic behavior in the first correlogram. The second correlogram also exhibits this periodic behavior, but it is weaker
- According to the first partial correlogram, the first (0.32) and twelfth (-0.18) lags are significant. Therefore, one can consider ARMA({1,12},0) model at a glance
- According to the second partial correlogram, however, the first (0.49) and second (0.18) lags are significant, but the twelfth (-0.01) lag is insignificant. The fifth (-0.17) and eleventh (-0.16) lags are close, but insignificant. Therefore, one can consider ARMA(2,0) model at a glance
- By and large, $\{\Delta Y_t\}$ is more persistent in the recent sample

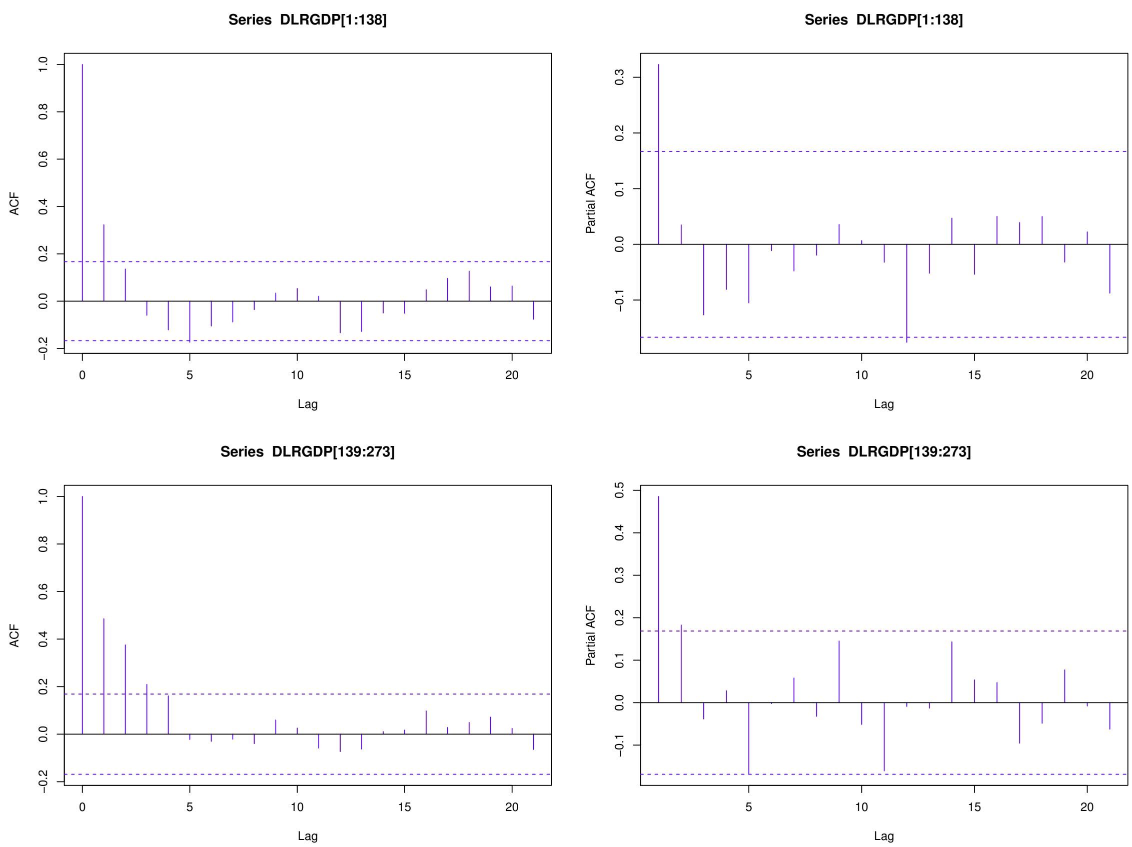


Figure 7: ACF and PACF of $\{\Delta Y_t\}$: 1947:2–1981:4 and 1982:1–2015:2

(g) Overall, $\{\Delta Y_t\}$ exhibits a strong periodic behavior and $\rho_{12}^{\text{partial}}$ plays a key role in determining this. In detail, however, $\rho_{12}^{\text{partial}}$ is significant only in the first sample and the significance is weaker in the second sample. Instead, the second sample is more persistent than the first sample with ρ_2^{partial}, which is more significant here. This implies the decay of the periodicity in the first sample and the advent of the stickiness in the second sample.

References

[Hamilton, 1994] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Econometric Methods II Assignment 02

Junyong Kim*

March 17, 2017

1 Analytical Exercise

1. By using matrix notations,

$$
\begin{aligned}
\mathbf{y}_t =& \begin{pmatrix} X_t & Y_t \end{pmatrix}^{\mathrm{t}} \\
X_t =& \varepsilon_t + \theta \varepsilon_{t-1} \\
Y_t =& h_1 X_{t-1} + u_t \\
\Rightarrow \mathbf{y}_t =& \Phi \mathbf{y}_{t-1} + \mathbf{u}_t + \Theta \mathbf{u}_{t-1} \\
\text{where } \Phi =& \begin{pmatrix} 0 & 0 \\ h_1 & 0 \end{pmatrix} \\
\text{and } \Theta =& \begin{pmatrix} \theta & 0 \\ 0 & 0 \end{pmatrix} \\
\mathbf{u}_t =& \begin{pmatrix} \varepsilon_t & u_t \end{pmatrix}^{\mathrm{t}} \\
\mathrm{E}\left[\mathbf{u}_t\right] =& \mathbf{0} \\
\mathrm{Var}\left[\mathbf{u}_t\right] =& \mathrm{E}\left[\mathbf{u}_t \mathbf{u}_t^{\mathrm{t}}\right] \\
=& \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \\
=& \Sigma,
\end{aligned}
$$

and

$$
\begin{aligned}
\Phi^2 =& \mathbf{0} \\
\text{and } \Phi\Theta =& \begin{pmatrix} 0 & 0 \\ h_1\theta & 0 \end{pmatrix}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
(\mathbf{I} - \Phi L)\mathbf{y}_t =& \mathbf{u}_t + \Theta \mathbf{u}_{t-1} \\
\Rightarrow \mathbf{y}_t =& (\mathbf{I} - \Phi L)^{-1} \mathbf{u}_t + (\mathbf{I} - \Phi L)^{-1} \Theta \mathbf{u}_{t-1} \\
=& \mathbf{u}_t + \Phi \mathbf{u}_{t-1} + \Phi^2 \mathbf{u}_{t-2} + \cdots + \Theta \mathbf{u}_{t-1} + \Phi\Theta \mathbf{u}_{t-2} + \Phi^2 \Theta \mathbf{u}_{t-3} + \cdots \\
=& \mathbf{u}_t + (\Phi + \Theta)\mathbf{u}_{t-1} + \Phi(\Phi + \Theta)\mathbf{u}_{t-2} + \cancel{\Phi^2}(\Phi + \Theta)\mathbf{u}_{t-3} + \cdots \\
=& \mathbf{u}_t + (\Phi + \Theta)\mathbf{u}_{t-1} + \Phi\Theta \mathbf{u}_{t-2}.
\end{aligned}
$$

*Lubar School of Business, University of Wisconsin–Milwaukee

Thus,

$$
\begin{aligned}
\mathrm{Var}\left[\mathbf{y}_t\right] =& \mathrm{E}\left[\mathbf{y}_t\mathbf{y}_t^{\mathrm{t}}\right] \\
=& \mathrm{E}\left[\left(\mathbf{u}_t + (\Phi+\Theta)\mathbf{u}_{t-1} + \Phi\Theta\mathbf{u}_{t-2}\right)\left(\mathbf{u}_t + (\Phi+\Theta)\mathbf{u}_{t-1} + \Phi\Theta\mathbf{u}_{t-2}\right)^{\mathrm{t}}\right] \\
=& \mathrm{E}\left[\mathbf{u}_t\mathbf{u}_t^{\mathrm{t}}\right] + \mathrm{E}\left[(\Phi+\Theta)\mathbf{u}_{t-1}\mathbf{u}_{t-1}^{\mathrm{t}}(\Phi+\Theta)^{\mathrm{t}}\right] + \mathrm{E}\left[\Phi\Theta\mathbf{u}_{t-2}\mathbf{u}_{t-2}^{\mathrm{t}}\Theta^{\mathrm{t}}\Phi^{\mathrm{t}}\right] \\
=& \Sigma + (\Phi+\Theta)\Sigma(\Phi+\Theta)^{\mathrm{t}} + \Phi\Theta\Sigma\Theta^{\mathrm{t}}\Phi^{\mathrm{t}} \\
=& \begin{pmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_u^2\end{pmatrix} + \begin{pmatrix}\theta & 0\\ h_1 & 0\end{pmatrix}\begin{pmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_u^2\end{pmatrix}\begin{pmatrix}\theta & h_1\\ 0 & 0\end{pmatrix} + \begin{pmatrix}0 & 0\\ h_1\theta & 0\end{pmatrix}\begin{pmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_u^2\end{pmatrix}\begin{pmatrix}0 & h_1\theta\\ 0 & 0\end{pmatrix} \\
=& \begin{pmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_u^2\end{pmatrix} + \begin{pmatrix}\theta^2\sigma_\varepsilon^2 & h_1\theta\sigma_\varepsilon^2\\ h_1\theta\sigma_\varepsilon^2 & h_1^2\sigma_\varepsilon^2\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & h_1^2\theta^2\sigma_\varepsilon^2\end{pmatrix} \\
=& \begin{pmatrix}\left(1+\theta^2\right)\sigma_\varepsilon^2 & h_1\theta\sigma_\varepsilon^2\\ h_1\theta\sigma_\varepsilon^2 & h_1^2\left(1+\theta^2\right)\sigma_\varepsilon^2 + \sigma_u^2\end{pmatrix} \\
\mathrm{Cov}\left[\mathbf{y}_t, \mathbf{y}_{t-1}\right] =& \mathrm{E}\left[\mathbf{y}_t\mathbf{y}_{t-1}^{\mathrm{t}}\right] \\
=& \mathrm{E}\left[\left(\mathbf{u}_t + (\Phi+\Theta)\mathbf{u}_{t-1} + \Phi\Theta\mathbf{u}_{t-2}\right)\left(\mathbf{u}_{t-1} + (\Phi+\Theta)\mathbf{u}_{t-2} + \Phi\Theta\mathbf{u}_{t-3}\right)^{\mathrm{t}}\right] \\
=& \mathrm{E}\left[(\Phi+\Theta)\mathbf{u}_{t-1}\mathbf{u}_{t-1}^{\mathrm{t}}\right] + \mathrm{E}\left[\Phi\Theta\mathbf{u}_{t-2}\mathbf{u}_{t-2}^{\mathrm{t}}(\Phi+\Theta)^{\mathrm{t}}\right] \\
=& (\Phi+\Theta)\Sigma + \Phi\Theta\Sigma(\Phi+\Theta)^{\mathrm{t}} \\
=& \begin{pmatrix}\theta & 0\\ h_1 & 0\end{pmatrix}\begin{pmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_u^2\end{pmatrix} + \begin{pmatrix}0 & 0\\ h_1\theta & 0\end{pmatrix}\begin{pmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_u^2\end{pmatrix}\begin{pmatrix}\theta & h_1\\ 0 & 0\end{pmatrix} \\
=& \begin{pmatrix}\theta\sigma_\varepsilon^2 & 0\\ h_1\sigma_\varepsilon^2 & 0\end{pmatrix} + \begin{pmatrix}0 & 0\\ h_1\theta^2\sigma_\varepsilon^2 & h_1^2\theta\sigma_\varepsilon^2\end{pmatrix} \\
=& \begin{pmatrix}\theta\sigma_\varepsilon^2 & 0\\ h_1\left(1+\theta^2\right)\sigma_\varepsilon^2 & h_1^2\theta\sigma_\varepsilon^2\end{pmatrix} \\
\mathrm{Cov}\left[\mathbf{y}_t, \mathbf{y}_{t-2}\right] =& \mathrm{E}\left[\mathbf{y}_t\mathbf{y}_{t-2}^{\mathrm{t}}\right] \\
=& \mathrm{E}\left[\left(\mathbf{u}_t + (\Phi+\Theta)\mathbf{u}_{t-1} + \Phi\Theta\mathbf{u}_{t-2}\right)\left(\mathbf{u}_{t-2} + (\Phi+\Theta)\mathbf{u}_{t-3} + \Phi\Theta\mathbf{u}_{t-4}\right)^{\mathrm{t}}\right] \\
=& \Phi\Theta\Sigma \\
=& \begin{pmatrix}0 & 0\\ h_1\theta & 0\end{pmatrix}\begin{pmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_u^2\end{pmatrix} \\
=& \begin{pmatrix}0 & 0\\ h_1\theta\sigma_\varepsilon^2 & 0\end{pmatrix} \\
\mathrm{Cov}\left[\mathbf{y}_t, \mathbf{y}_{t-j}\right] =& \mathbf{0}, \quad \forall j \in \{3, 4, \cdots\}.
\end{aligned}
$$

So,

$$
\Gamma_k = \mathrm{Cov}\left[\mathbf{y}_t, \mathbf{y}_{t-k}\right] = \begin{cases} \begin{pmatrix}\left(1+\theta^2\right)\sigma_\varepsilon^2 & h_1\theta\sigma_\varepsilon^2\\ h_1\theta\sigma_\varepsilon^2 & h_1^2\left(1+\theta^2\right)\sigma_\varepsilon^2 + \sigma_u^2\end{pmatrix} & k = 0 \\ \begin{pmatrix}\theta\sigma_\varepsilon^2 & 0\\ h_1\left(1+\theta^2\right)\sigma_\varepsilon^2 & h_1^2\theta\sigma_\varepsilon^2\end{pmatrix} & k = 1 \\ \begin{pmatrix}0 & 0\\ h_1\theta\sigma_\varepsilon^2 & 0\end{pmatrix} & k = 2 \\ \mathbf{0} & k \in \{3, 4, \cdots\} \end{cases}
$$

and $\Gamma_{-k} = \mathrm{Cov}\left[\mathbf{y}_t, \mathbf{y}_{t+k}\right] = \mathrm{Cov}\left[\mathbf{y}_{t-k}, \mathbf{y}_t\right] = \mathrm{E}\left[\mathbf{y}_{t-k}\mathbf{y}_t^{\mathrm{t}}\right] = \mathrm{E}\left[\left(\mathbf{y}_t\mathbf{y}_{t-k}^{\mathrm{t}}\right)^{\mathrm{t}}\right] = \left(\mathrm{E}\left[\mathbf{y}_t\mathbf{y}_{t-k}^{\mathrm{t}}\right]\right)^{\mathrm{t}} = \Gamma_k^{\mathrm{t}}$.

2. By using characteristic decomposition,

$$\begin{aligned}\Phi &= \begin{pmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{pmatrix} \\ \Rightarrow \det(\Phi - \lambda I) =& (0.3-\lambda)(0.4-\lambda) - 0.8 \times 0.9 = \lambda^2 - 0.7\lambda - 0.6 = 0 \\ \Rightarrow \lambda^* =& \frac{0.7 \pm \sqrt{2.89}}{2} = \frac{0.7 \pm 1.7}{2} \in \{1.2, -0.5\}.\end{aligned}$$

Since $|\lambda_1| > 1$, this process is not stable (i.e. $\lim_{j\to\infty} \Phi^j \neq \mathbf{0}$), and hence it is not covariance-stationary.

3. (a) For VAR(1) process,

$$\begin{aligned}\mathbf{y}_t =& \mathbf{A}\mathbf{y}_{t-1} + \varepsilon_t \\ =& (\mathbf{I} - \mathbf{A}L)^{-1}\varepsilon_t \\ =& \varepsilon_t + \mathbf{A}\varepsilon_{t-1} + \mathbf{A}^2\varepsilon_{t-2} + \mathbf{A}^3\varepsilon_{t-3} + \cdots\end{aligned}$$

and hence

$$\Psi_k = \mathbf{A}^k, \quad k \in \{1, 2, \cdots\}.$$

(b) For VAR(2) process,

$$\begin{aligned}\mathbf{y}_t =& \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \varepsilon_t \\ =& \left(\mathbf{I} - \mathbf{A}_1 L - \mathbf{A}_2 L^2\right)^{-1}\varepsilon_t \\ =& \left(\mathbf{I} + \Psi_1 L + \Psi_2 L^2 + \Psi_3 L^3 + \cdots\right)\varepsilon_t \\ \Rightarrow \left(\mathbf{I} - \mathbf{A}_1 L - \mathbf{A}_2 L^2\right)^{-1} =& \left(\mathbf{I} + \Psi_1 L + \Psi_2 L^2 + \Psi_3 L^3 + \cdots\right) \\ \mathbf{I} =& \left(\mathbf{I} - \mathbf{A}_1 L - \mathbf{A}_2 L^2\right)\left(\mathbf{I} + \Psi_1 L + \Psi_2 L^2 + \Psi_3 L^3 + \cdots\right) \\ =& \mathbf{I} + \Psi_1 L + \Psi_2 L^2 + \Psi_3 L^3 + \cdots \\ & - \mathbf{A}_1 L - \mathbf{A}_1\Psi_1 L^2 - \mathbf{A}_1\Psi_2 L^3 - \cdots \\ & - \mathbf{A}_2 L^2 - \mathbf{A}_2\Psi_1 L^3 - \mathbf{A}_2\Psi_2 L^4 - \cdots .\end{aligned}$$

Therefore,

$$\begin{aligned}\Psi_1 - \mathbf{A}_1 =& \mathbf{0} \\ \Psi_2 - \mathbf{A}_1\Psi_1 - \mathbf{A}_2 =& \mathbf{0} \\ \Psi_3 - \mathbf{A}_1\Psi_2 - \mathbf{A}_2\Psi_1 =& \mathbf{0} \\ \Rightarrow \Psi_k - \mathbf{A}_1\Psi_{k-1} - \mathbf{A}_2\Psi_{k-2} =& \mathbf{0}, \quad k \in \{3, 4, \cdots\},\end{aligned}$$

and hence

$$\begin{aligned}\Psi_1 =& \mathbf{A}_1 \\ \Psi_2 =& \mathbf{A}_1\Psi_1 - \mathbf{A}_2 \\ \Psi_3 =& \mathbf{A}_1\Psi_2 - \mathbf{A}_2\Psi_1 \\ \Rightarrow \Psi_k =& \mathbf{A}_1\Psi_{k-1} - \mathbf{A}_2\Psi_{k-2}, \quad k \in \{3, 4, \cdots\}.\end{aligned}$$

2 Empirical Exercise

1. Unit root test [Lothian and Taylor, 1996]

   (a) According to Figure 1, one can observe a slight downward trend at a glance.

   (b) ADF test uses the equation $\Delta y_t = \alpha + \beta y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + \varepsilon_t$ and checks if $\beta = 0$.

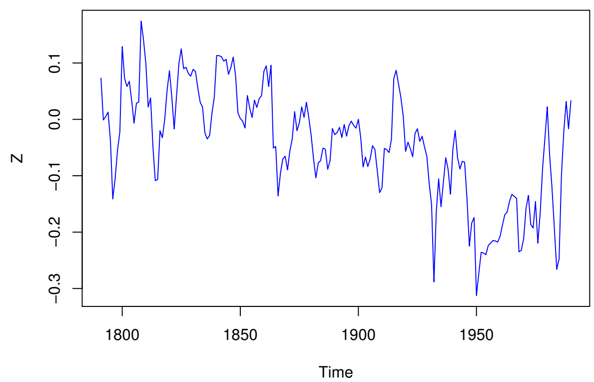


Figure 1: $z_t = \ln S_t - \ln P_t + \ln P_t^*$

			Critical value		
Case	p	τ_μ	1%	5%	10%
AIC	8	−2.25	−3.46	−2.88	−2.57
BIC	1	−3.51	−3.46	−2.88	−2.57
$\ln z_t$	0	−3.45	−3.46	−2.88	−2.57
Original paper		−3.47			
z_t	0	−3.47	−3.75	−3.00	−2.63

In the first case, p that minimizes AIC is 8 and the corresponding τ_μ is −2.25. Since the 10% critical value is −2.57, one cannot reject the null (H_0: ∃a unit root in the series).

In the second case, on the other hand, p that minimizes BIC is 1 and the corresponding τ_μ is −3.51. Since the 1% critical value is −3.46, one can reject the null at 1% significance level.

In the third case, I use $\ln z_t$ in order to replicate the original result [Lothian and Taylor, 1996]. I estimate $\Delta \ln z_t = \alpha + \beta \ln z_{t-1} + \varepsilon_t$ and the resulting τ_μ is −3.45. Since 1% and 5% critical values are −3.46 and −2.88, respectively, one can reject the null at 5% significance level. This is close to −3.47 of the original paper.

z_t without the log transformation exhibits $\tau_\mu = -3.47$ instead.

```
DFTEST1=ur.df(Z,type="drift",lags=10,selectlags="AIC")
summary(DFTEST1)
DFTEST2=ur.df(Z,type="drift",lags=10,selectlags="BIC")
summary(DFTEST2)
DFTEST3=ur.df(log(Z),type="drift",lags=0)
summary(DFTEST3)
```

(c) With 16 observations, the resulting τ_μ is −1.21. Since the 10% critical value is −2.63, one cannot reject the null even at 10% significance level.

```
DFTEST4=ur.df(Z[184:200],type="drift",lags=0)
summary(DFTEST4)
```

(d) With 43 observations, the resulting τ_μ is −2.66. Since 5% and 10% critical values are −2.93 and −2.60, respectively, one can reject the null at 10% significance level and cannot reject at 5% significance level. $\hat{\beta}_{(\mathrm{c})} = -0.23$ and $\hat{\beta}_{(\mathrm{d})} = -0.28$, but their standard errors are 0.19 and 0.10, respectively, hence the latter is closer to zero than the former.

```
DFTEST5=ur.df(Z[80:123],type="drift",lags=0)
summary(DFTEST5)
```

(e) According to Phillips–Perron test,

Sample	Z_τ	Critical value 1%	5%	10%
1791–1990	-3.62	-3.46	-2.88	-2.57
1974–1990	-1.54	-3.92	-3.07	-2.67
1870–1913	-2.84	-3.59	-2.93	-2.60

This is consistent with the result obtained from ADF tests. With the sample from 1791 to 1990, one can reject the null at 1% significance level. With the sample from 1974 to 1990, one cannot reject the null even at 10% significance level. With the sample from 1870 to 1913, one can reject the null at 10% significance level.

```
PPTEST1=ur.pp(Z,type="Z-tau",model="constant")
summary(PPTEST1)
PPTEST2=ur.pp(Z[184:200],type="Z-tau",model="constant")
summary(PPTEST2)
PPTEST3=ur.pp(Z[80:123],type="Z-tau",model="constant")
summary(PPTEST3)
```

(f) According to Elliot–Rothenberg–Stock test,

Sample	P_T^μ	Critical value 1%	5%	10%
1791–1990	2.02	1.91	3.17	4.33
1974–1990	0.09	1.87	2.97	3.91
1870–1913	2.63	1.87	2.97	3.91

This is different from the result obtained from ADF tests. With the sample from 1791 to 1990, one can reject the null at 5% significance level. With the sample from 1974 to 1990, one can reject the null at 1% significance level. With the sample from 1870 to 1913, one can reject the null at 5% significance level.

```
ERSTEST1=ur.ers(Z,type="P-test",model="constant")
summary(ERSTEST1)
ERSTEST2=ur.ers(Z[184:200],type="P-test",model="constant")
summary(ERSTEST2)
ERSTEST3=ur.ers(Z[80:123],type="P-test",model="constant")
summary(ERSTEST3)
```

2. VAR models

(a) The left panel of Figure 2 is the plot of $\Delta \mathrm{s}_t$. The right panel of Figure 2 is the plot of fp_t.

```
plot(DS,col="blue")
plot(FP,col="blue")
```

(b) The upper-left panel of Figure 3 is the ACF of $\Delta \mathrm{s}_t$ (i.e. $\mathrm{Corr}\left[\Delta \mathrm{s}_t, \Delta \mathrm{s}_{t-j}\right]$). The upper-right panel of Figure 3 is the cross-correlation function between $\Delta \mathrm{s}_t$ and fp_t (i.e. $\mathrm{Corr}\left[\Delta \mathrm{s}_t, \mathrm{fp}_{t-j}\right]$). The lower-left panel of Figure 3 is the cross-correlation function between fp_t and $\Delta \mathrm{s}_t$ (i.e. $\mathrm{Corr}\left[\mathrm{fp}_t, \Delta \mathrm{s}_{t-j}\right]$). The lower-right panel of Figure 3 is the ACF of fp_t (i.e. $\mathrm{Corr}\left[\mathrm{fp}_t, \mathrm{fp}_{t-j}\right]$).

```
acf(coredata(DS),type="correlation")
ccf(coredata(DS),coredata(FP),type="correlation")
ccf(coredata(FP),coredata(DS),type="correlation")
acf(coredata(FP),type="correlation")
```

(c) $p^* = \underset{p\in\mathbb{N}}{\operatorname{argmin}}\mathrm{AIC}(p)$. According to the result, VAR(1) provides the smallest AIC.

p	1	2	3	4
AIC	-20.28	-20.25	-20.23	-20.20

	ϕ_0 Constant	Φ_1 $\Delta \mathrm{s}_{t-1}$	fp_{t-1}	R^2	F statistic	Σ $\Delta \mathrm{s}_t$	fp_t
$\Delta \mathrm{s}_t$	-0.0046	0.0687	-1.5744	0.02	2.82	0.001158	-0.000002
	(-1.61)	(1.06)	(-1.94)			(1.0000)	(-0.0554)
fp_t	-0.0002	0.0023	0.9110	0.82	554.16	-0.000002	0.000001
	(-1.94)	(1.05)	(33.07)			(-0.0554)	(1.0000)

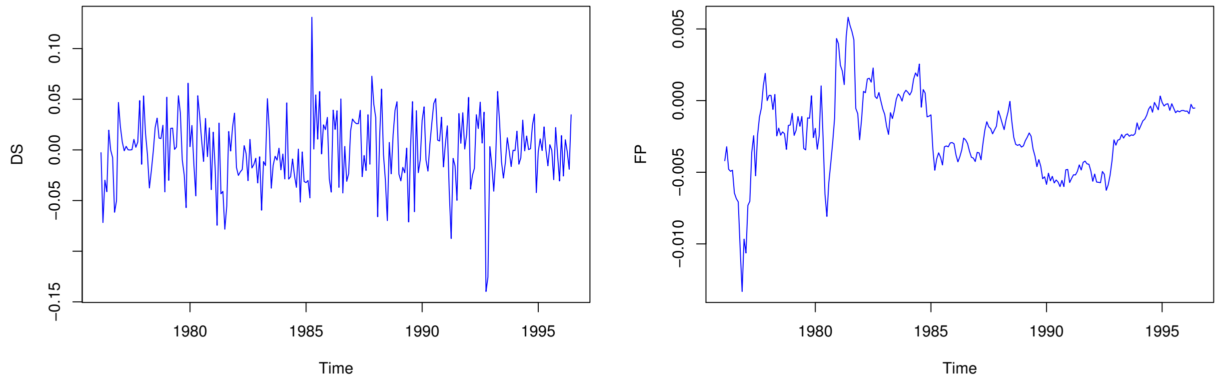


Figure 2: $\{\Delta s_t\}$ and $\{fp_t\}$

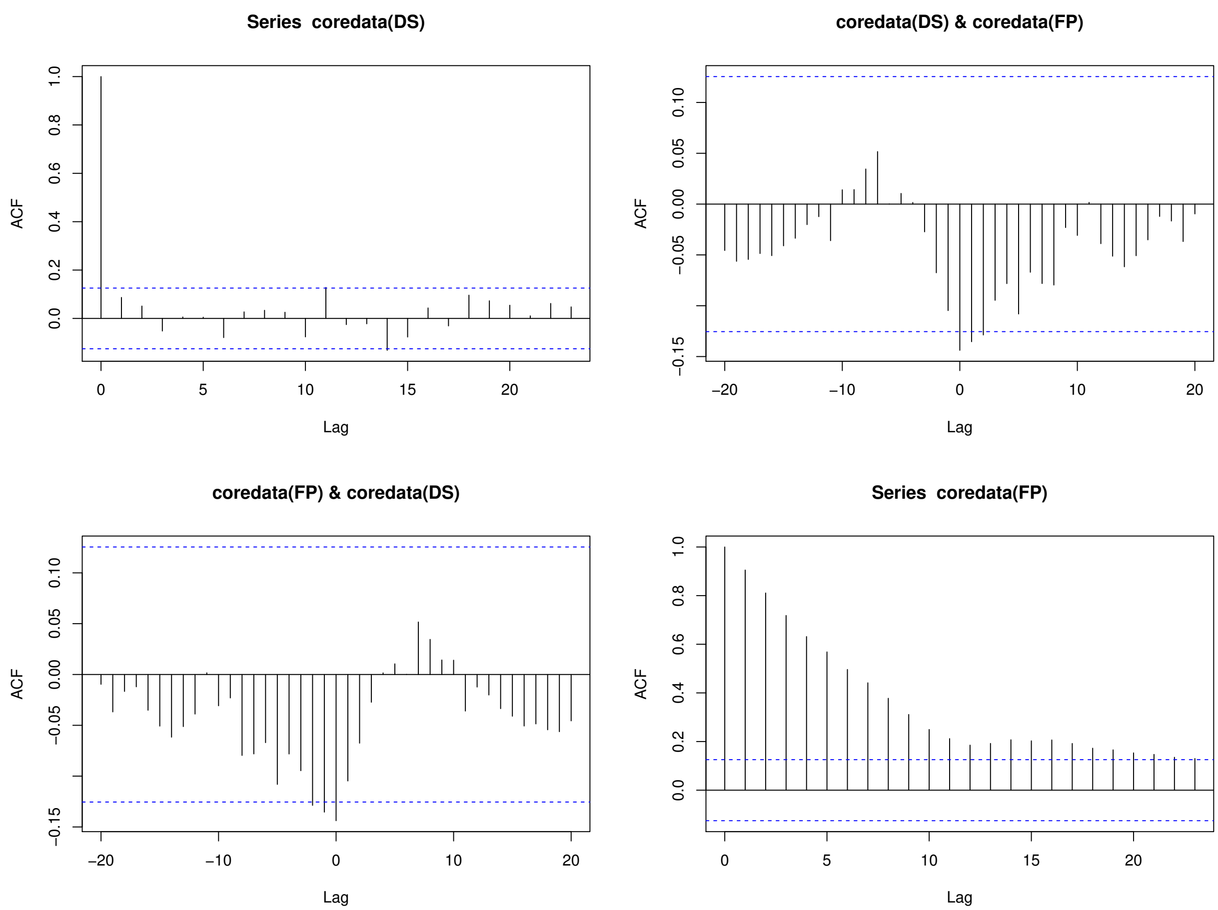


Figure 3: Auto- and Cross-correlation between Δs_t and fp_t

I report $\hat{\phi}_0$, $\hat{\Phi}_1$ and $\hat{\Sigma}$. The numbers in parentheses are t statistics for $\hat{\phi}_0$, $\hat{\Phi}_1$ and sample correlations for $\hat{\Sigma}$, respectively.

```
VARselect(Y,lag.max=4)
VAR1=VAR(Y,p=1)
summary(VAR1)
```

(d) $\Delta \mathrm{s}_t$ does not Granger-cause fp_t, but fp_t Granger-causes $\Delta \mathrm{s}_t$.

H_0	F statistic	p-value
$\Delta \mathrm{s}_t$ does not Granger-cause fp_t	1.11	0.29
fp_t does not Granger-cause $\Delta \mathrm{s}_t$	3.78	0.05

```
causality(VAR1,cause="DS")
causality(VAR1,cause="FP")
```

(e) Since $\Sigma = \mathbf{L}\mathbf{L}^{\mathrm{t}} = \begin{pmatrix} 0.001158 & -0.000002 \\ -0.000002 & 0.000001 \end{pmatrix}$, the corresponding $\mathbf{L} = \begin{pmatrix} 0.034029 & 0.000000 \\ -0.000064 & 0.001156 \end{pmatrix}$. The upper-left panel of Figure 4 shows that the orthogonalized one standard deivation (-0.034029) shock in $\Delta \mathrm{s}_t$ disappears within two periods. The lower-left panel of Figure 4 exhibits the negative effect (-0.000064) of that shock on the contemporaneous fp_t, but the response is not significantly different from 0. Since the orthogonalized shock in $\Delta \mathrm{s}_t$ is more exogenous than the shock in fp_t, the second shock cannot affect the first variable $\Delta \mathrm{s}_t$.

```
IRF11=irf(VAR1,response="DS",impulse="DS",ortho=TRUE,cumulative=FALSE,boot=TRUE,ci=0.95,runs=1000,
     seed=1)
IRF12=irf(VAR1,response="DS",impulse="FP",ortho=TRUE,cumulative=FALSE,boot=TRUE,ci=0.95,runs=1000,
     seed=1)
IRF13=irf(VAR1,response="FP",impulse="DS",ortho=TRUE,cumulative=FALSE,boot=TRUE,ci=0.95,runs=1000,
     seed=1)
IRF14=irf(VAR1,response="FP",impulse="FP",ortho=TRUE,cumulative=FALSE,boot=TRUE,ci=0.95,runs=1000,
     seed=1)
plot(IRF11)
plot(IRF12)
plot(IRF13)
plot(IRF14)
```

(f) The left panel of Figure 5 displays the forecast error variance decomposition for $\Delta \mathrm{s}_t$. The right panel of Figure 5 displays the forecast error variance decomposition for fp_t. About 1.59% of the twelve steps ahead forecast error of $\Delta \mathrm{s}_t$ is explained by the shock in fp_t, but only about 0.11% of the twelve steps ahead forecast error of fp_t is explained by the shock in $\Delta \mathrm{s}_t$.

```
FEVD1=fevd(VAR1,n.ahead=12)
plot(FEVD1,plot.type="single")
```

(g) $\Sigma = \mathbf{L}\mathbf{L}^{\mathrm{t}} = \begin{pmatrix} 0.000001 & -0.000002 \\ -0.000002 & 0.001158 \end{pmatrix}$ and hence $\mathbf{L} = \begin{pmatrix} 0.001158 & 0.000000 \\ -0.001885 & 0.033350 \end{pmatrix}$. The upper-left panel of Figure 6 implies that the orthogonalized one standard deivation (0.001158) shock in fp_t disappears exponentially. The lower-left panel of Figure 6 exhibits the negative effect (-0.001885) of that shock on the contemporaneous $\Delta \mathrm{s}_t$, but the response is not significantly different from 0. Though the 95% confidence interval includes 0, the shock in fp_t leaves a negative lasting impact on $\Delta \mathrm{s}_t$. This lower-left panel of Figure 6 is consistent with the upper-left panel of Figure 4. Since the orthogonalized shock in fp_t is more exogenous than the shock in $\Delta \mathrm{s}_t$, the second shock cannot affect the first variable fp_t.

```
VAR2=VAR(ts.union(FP,DS),p=1)
IRF21=irf(VAR2,response="FP",impulse="FP",ortho=TRUE,cumulative=FALSE,boot=TRUE,ci=0.95,runs=1000,
     seed=1)
IRF22=irf(VAR2,response="FP",impulse="DS",ortho=TRUE,cumulative=FALSE,boot=TRUE,ci=0.95,runs=1000,
     seed=1)
IRF23=irf(VAR2,response="DS",impulse="FP",ortho=TRUE,cumulative=FALSE,boot=TRUE,ci=0.95,runs=1000,
     seed=1)
IRF24=irf(VAR2,response="DS",impulse="DS",ortho=TRUE,cumulative=FALSE,boot=TRUE,ci=0.95,runs=1000,
     seed=1)
plot(IRF21)
plot(IRF22)
plot(IRF23)
plot(IRF24)
```

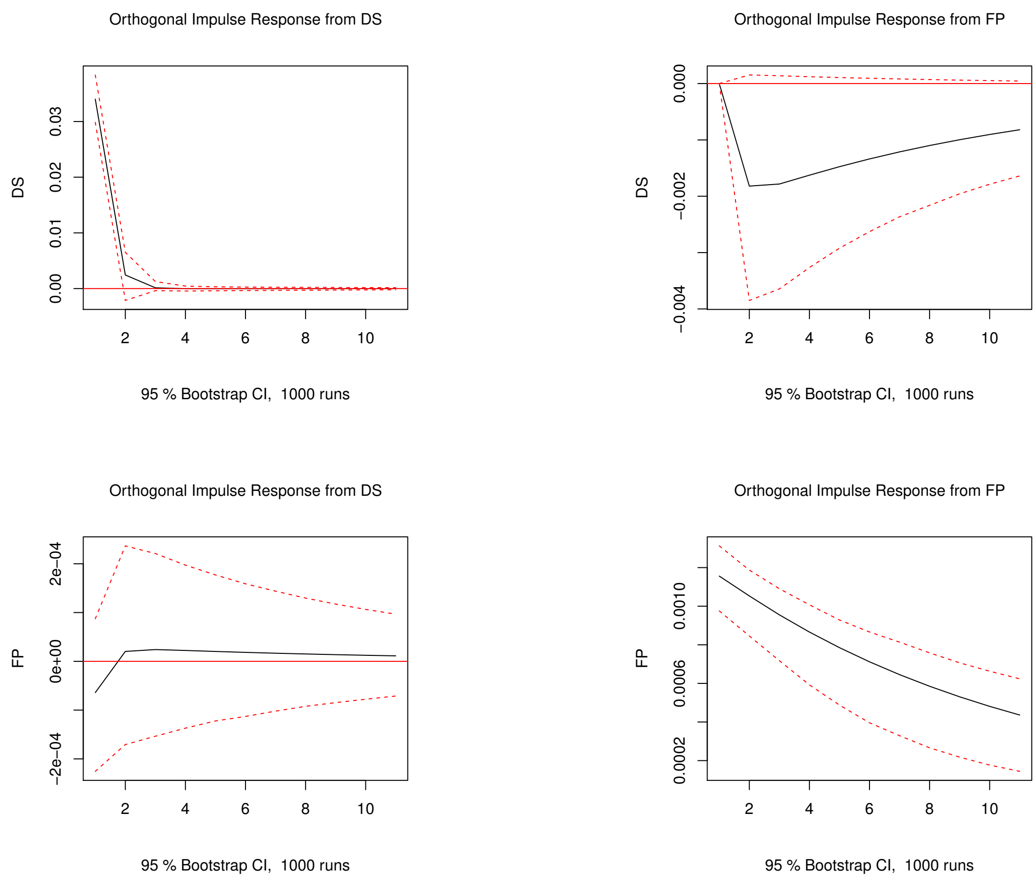


Figure 4: Impulse response functions between Δs_t and fp_t

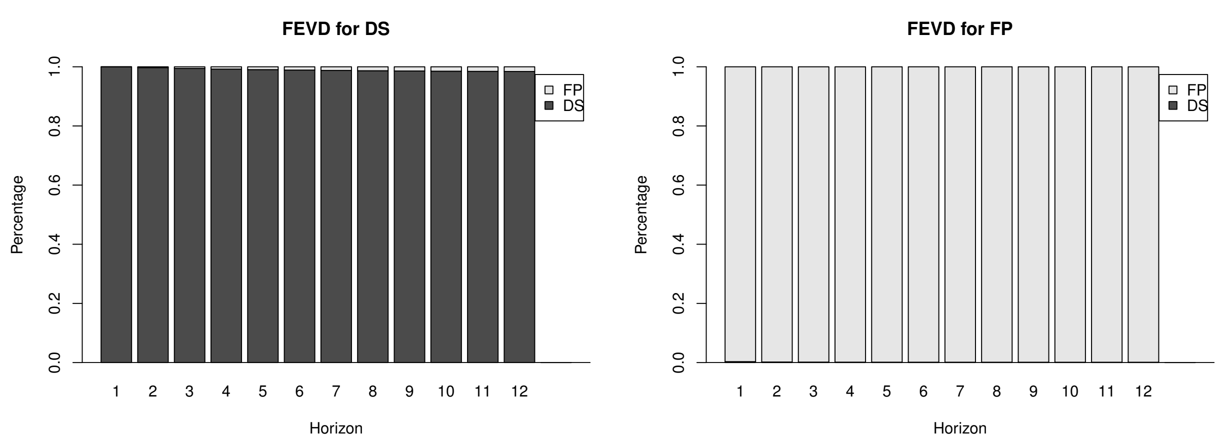


Figure 5: Forecast error variance decomposition between Δs_t and fp_t

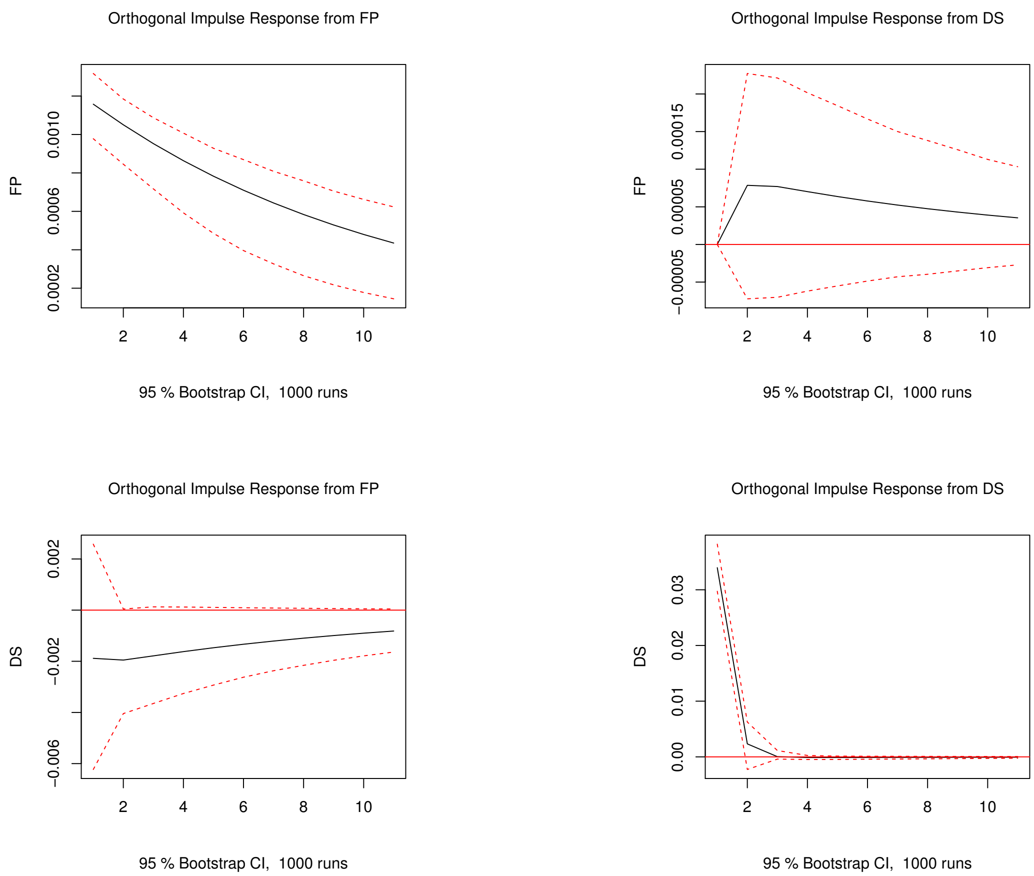


Figure 6: Impulse response functions obtained from $\mathbf{y}_t = (\mathrm{fp}_t, \Delta \mathrm{s}_t)^{\mathrm{t}}$

(h) Firstly, I estimate the model without priors. Secondly, I estimate again using priors obtained from the OLS estimation. In particular, I adopt $\phi_{11}^{(1)} = 0.0687$ and $\phi_{22}^{(1)} = 0.9110$ as the prior mean of Normal distribution for the VAR coefficient matrix and $\Sigma = \begin{pmatrix} 0.001158 & -0.000002 \\ -0.000002 & 0.000001 \end{pmatrix}$ as the prior location of inverse-Wishart distribution for the VAR covariance matrix.

Without prior

	ϕ_0	Φ_1		Σ	
	Constant	Δs_{t-1}	fp_{t-1}	Δs_t	fp_t
Δs_t	−0.0018	0.1001	−0.3844	0.005312	−0.000001
St. Dev.	(0.0051)	(0.1356)	(0.8628)	(0.000492)	(0.000302)
5%	−0.0101	−0.1226	−1.7972	0.004572	−0.000497
50%	−0.0018	0.0987	−0.3825	0.005271	−0.000001
95%	0.0065	0.3235	1.0583	0.006184	0.000497
fp_t	−0.0001	0.0035	0.9661	−0.000001	0.004153
St. Dev.	(0.0045)	(0.1201)	(0.8370)	(0.000302)	(0.000380)
5%	−0.0075	−0.1936	−0.4154	−0.000497	0.003570
50%	−0.0001	0.0038	0.9652	−0.000001	0.004125
95%	0.0073	0.2012	2.3169	0.000497	0.004818

With prior

	ϕ_0	Φ_1		Σ	
	Constant	Δs_{t-1}	fp_{t-1}	Δs_t	fp_t
Δs_t	−0.0032	0.0762	−0.9659	0.001162	−0.000002
St. Dev.	(0.0026)	(0.0645)	(0.6295)	(0.000107)	(0.000003)
5%	−0.0075	−0.0300	−1.9883	0.000996	−0.000006
50%	−0.0032	0.0769	−0.9739	0.001157	−0.000002
95%	0.0011	0.1814	0.0762	0.001343	0.000002
fp_t	−0.0002	0.0023	0.9100	−0.000002	0.000001
St. Dev.	(0.0001)	(0.0022)	(0.0274)	(0.000003)	(0.000000)
5%	−0.0004	−0.0014	0.8644	−0.000006	0.000001
50%	−0.0002	0.0023	0.9096	−0.000002	0.000001
95%	0.0000	0.0059	0.9555	0.000002	0.000002

Using simulated posteriors I report posterior mean, standard deviation and 5, 50, 95 percentiles, respectively. In both cases I generate 10,000 observations after burning 1,000 observations. Figure 7 exhibits IRFs obtained without imposing priors and Figure 8 displays IRFs obtained using priors. Red and blue lines are the median and the (5%, 95%) interval of the IRFs.

```
BVAR1=BVARW(Y,p=1,irf.periods=10,keep=10000,burn=1000)
apply(simplify2array(BVAR1$BDraws),c(1,2),mean)
apply(simplify2array(BVAR1$BDraws),c(1,2),sd)
apply(simplify2array(BVAR1$BDraws),c(1,2),quantile,probs=c(0.05,0.50,0.95))
apply(simplify2array(BVAR1$SDraws),c(1,2),mean)
apply(simplify2array(BVAR1$SDraws),c(1,2),sd)
apply(simplify2array(BVAR1$SDraws),c(1,2),quantile,probs=c(0.05,0.50,0.95))
plot(BVAR1)
IRF(BVAR1,percentiles=c(0.05,0.50,0.95),save=F)
```

```
BVAR2=BVARW(Y,coefprior=c(0.0687,0.9110),p=1,irf.periods=10,keep=10000,burn=1000,XiSigma=matrix(c
     (0.001158,-0.000002,-0.000002,0.000001),nrow=2,ncol=2))
apply(simplify2array(BVAR2$BDraws),c(1,2),mean)
apply(simplify2array(BVAR2$BDraws),c(1,2),sd)
apply(simplify2array(BVAR2$BDraws),c(1,2),quantile,probs=c(0.05,0.50,0.95))
apply(simplify2array(BVAR2$SDraws),c(1,2),mean)
apply(simplify2array(BVAR2$SDraws),c(1,2),sd)
apply(simplify2array(BVAR2$SDraws),c(1,2),quantile,probs=c(0.05,0.50,0.95))
plot(BVAR2)
IRF(BVAR2,percentiles=c(0.05,0.50,0.95),save=F)
```

(i) Without priors, estimates are insignificant and IRFs are unstable. On the other hand, the result is close to the previous outcomes when priors are employed. For instance, the patterns found in Figure 8 are similar to those found in Figure 4. Only off-diagonal plots are located oppositely.

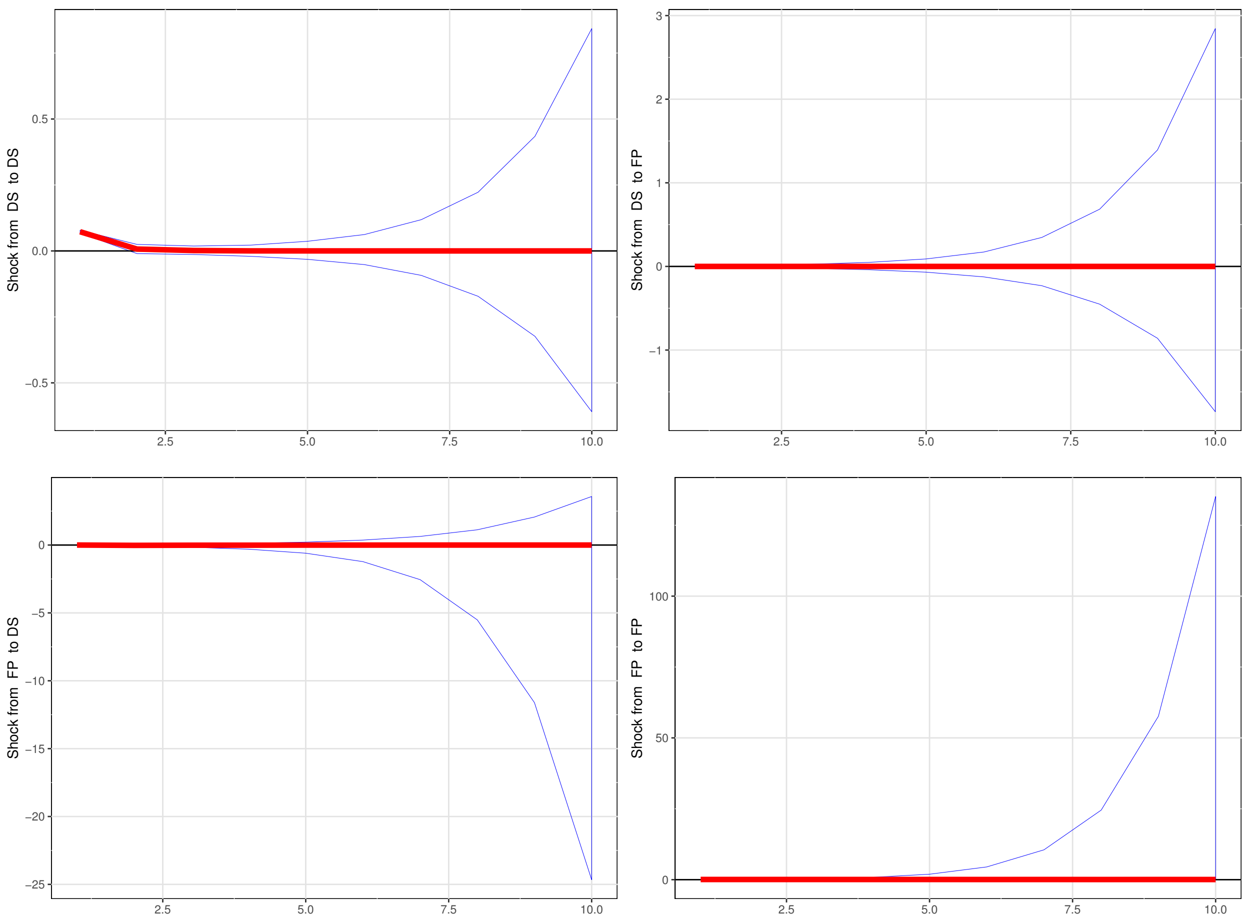


Figure 7: IRFs obtained without priors

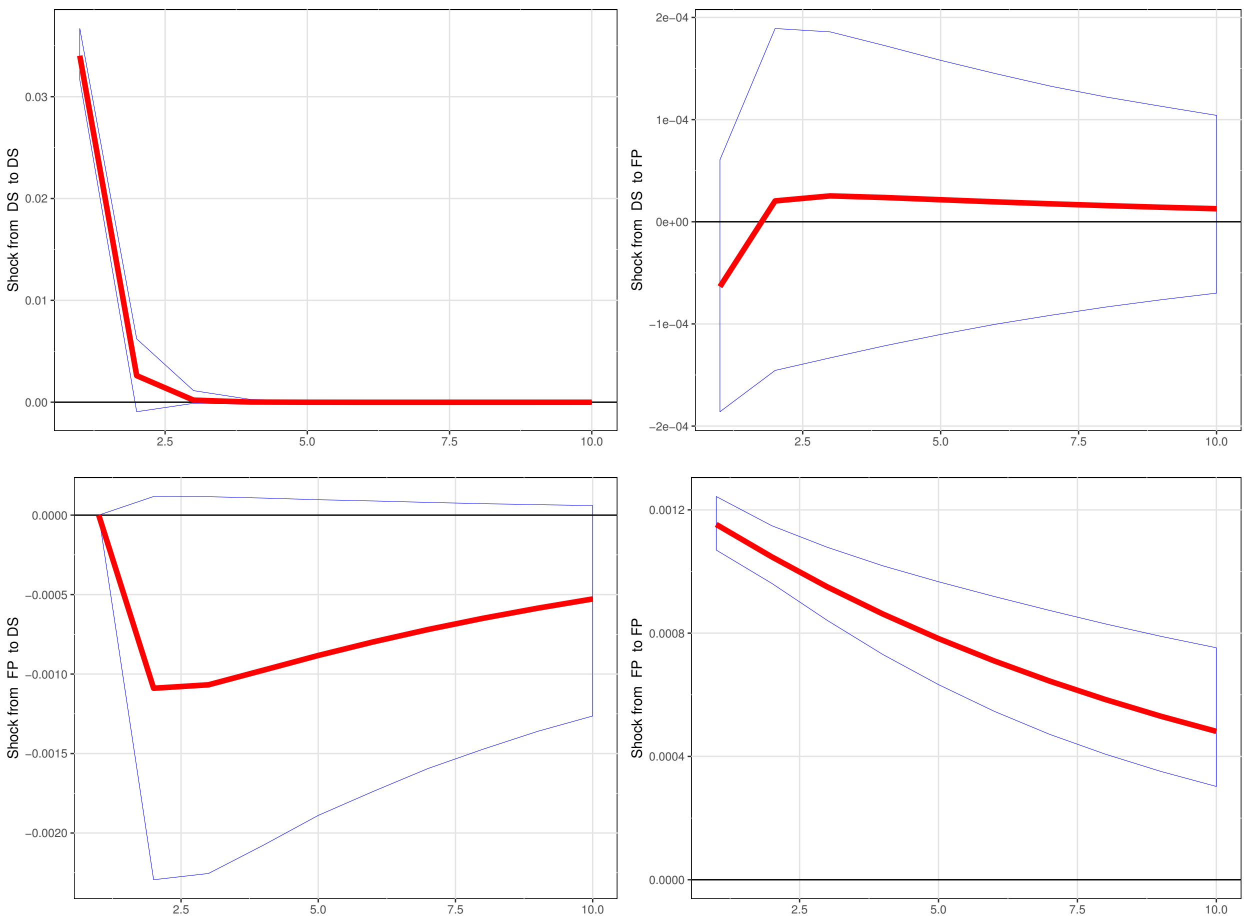


Figure 8: IRFs obtained using priors

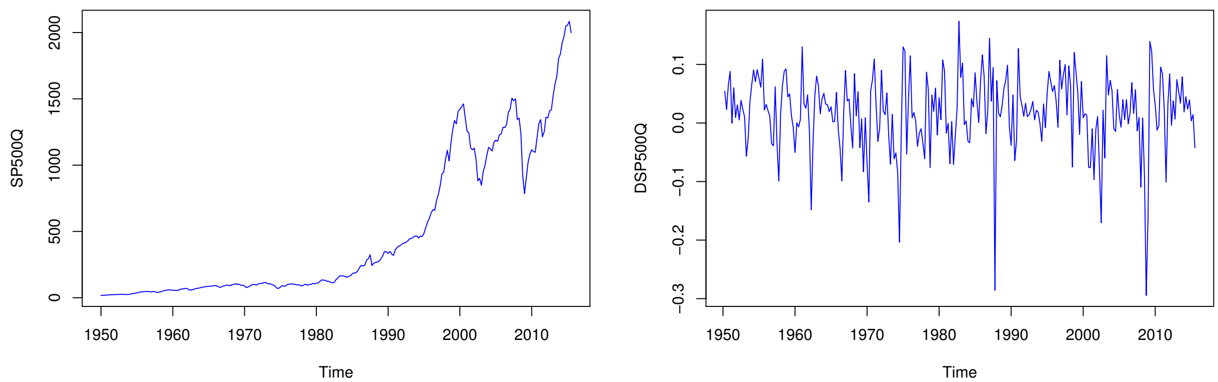


Figure 9: $\{\text{S\&P500}_t\}$ and $\{\Delta \log \text{S\&P500}_t\}$

3. Stock return predictability using consumption-wealth ratio and the spread inversion indicator

   According to empirical researches, one can consider the variable cay_t (i.e. consumption versus aggregate wealth) to predict the future stock return. [Lettau and Ludvigson, 2001].

   (a) One can use `aggregate` to transform data. The left panel of Figure 9 is the plot of the series $\{\text{S\&P500}_t\}$. The right panel of the figure is the plot after log-differencing.

```
LL1=read.csv("170303_sp500.csv")
LL2=read.csv("170303_cay.csv")
SP500M=ts(LL1$Adj.Close,start=c(1950,1),frequency=12)
SP500Q=aggregate(SP500M,nfrequency=4,FUN="mean")
plot(SP500Q,col="blue")
```

   (b) I conduct both ADF and PP tests. I compute ADF τ, τ_μ, τ_τ and PP Z_τ statistics to test the null H_0: $\beta = 0$. I choose the number of lags p using AIC for each of ADF cases. According to the result, one cannot reject the null that there exists a unit root in the series.

			Critical value		
Test	p	Statistic	1%	5%	10%
τ	0	1.78	−2.58	−1.95	−1.62
τ_μ	1	0.77	−3.44	−2.87	−2.57
τ_τ	2	−1.42	−3.98	−3.42	−3.13
Z_τ		1.09	−3.46	−2.87	−2.57

```
DFTEST6=ur.df(SP500Q,type="none",lags=10,selectlags="AIC")
summary(DFTEST6)
DFTEST7=ur.df(SP500Q,type="drift",lags=10,selectlags="AIC")
summary(DFTEST7)
DFTEST8=ur.df(SP500Q,type="trend",lags=10,selectlags="AIC")
summary(DFTEST8)
PPTEST4=ur.pp(SP500Q,type="Z-tau",model="constant")
summary(PPTEST4)
```

   (c) As in the previous case, I conduct both ADF and PP tests. According to the result, one can reject H_0: $\beta = 0$ and conclude that the series is difference-stationary. The right panel of Figure 9 is the plot of the series $\{\Delta \log \text{S\&P500}_t\}$.

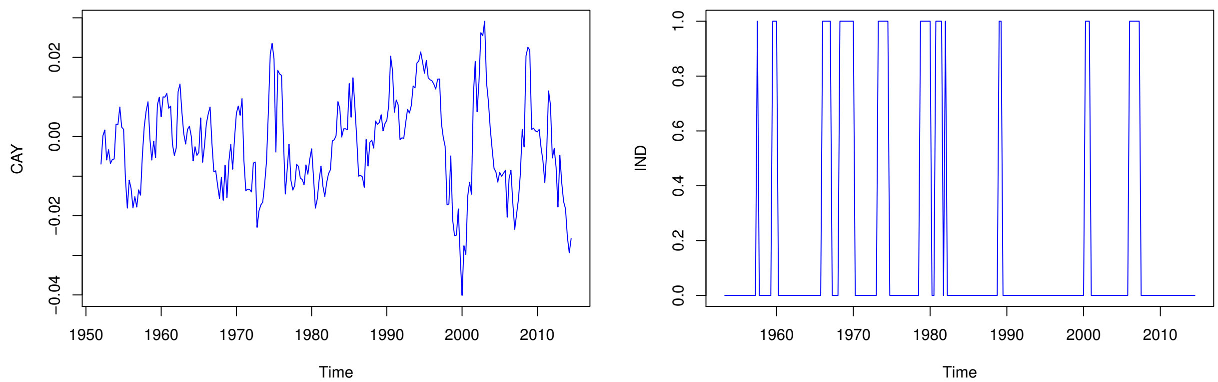


Figure 10: $\{\text{cay}_t\}$ and $\{\text{indicator}_t\}$

Test	p	Statistic	Critical value 1%	5%	10%
τ	1	-9.34	-2.58	-1.95	-1.62
τ_μ	1	-10.03	-3.44	-2.87	-2.57
τ_τ	1	-10.01	-3.98	-3.42	-3.13
Z_τ		-12.30	-3.46	-2.87	-2.57

```
DSP500Q=ts(diff(log(SP500Q)),start=c(1950,2),frequency=4)
plot(DSP500Q,col="blue")
DFTEST9=ur.df(DSP500Q,type="none",lags=10,selectlags="AIC")
summary(DFTEST9)
DFTESTa=ur.df(DSP500Q,type="drift",lags=10,selectlags="AIC")
summary(DFTESTa)
DFTESTb=ur.df(DSP500Q,type="trend",lags=10,selectlags="AIC")
summary(DFTESTb)
PPTEST5=ur.pp(DSP500Q,type="Z-tau",model="constant")
summary(PPTEST5)
```

(d) The left and right panels of Figure 10 are the plots of the series cay_t and the series indicator_t, respectively.

```
CAY=ts(LL2$cay,start=c(1952,1),frequency=4)
LL2$indicator=as.numeric(LL2$indicator)-1
IND=ts(LL2$indicator[6:251],start=c(1953,2),frequency=4)
TRIMDSP=ts(DSP500Q[13:258],start=c(1953,2),frequency=4)
TRIMCAY=ts(CAY[6:251],start=c(1953,2),frequency=4)
TRIMIND=ts(IND,start=c(1953,2),frequency=4)
plot(CAY,col="blue")
plot(IND,col="blue")
```

(e) The left, middle and right panels of Figure 11 are the plots of $\{\Delta \log \text{S\&P500}_t\}$, $\{\text{cay}_t\}$ and $\{\text{indicator}_t\}$, respectively. In terms of the speed of movement, $\Delta \log \text{S\&P500}$ is the fastest, indicator is the slowest and cay is in between these two variables.

(f) The left, middle and right panels of Figure 12 are the plots of p versus the corresponding AIC(p) from three VAR models with the vectors ($\Delta \log$ S&P500,cay), ($\Delta \log$ S&P500,indicator) and ($\Delta \log$ S&P500,cay,indicator), respectively.

```
PSELECT1=VARselect(ts.union(TRIMDSP,TRIMCAY),lag.max=20)
plot(c(1:20),PSELECT1$criteria[1,],xlab="p",ylab="AIC",type="l",col="blue")
PSELECT2=VARselect(ts.union(TRIMDSP,TRIMIND),lag.max=20)
plot(c(1:20),PSELECT2$criteria[1,],xlab="p",ylab="AIC",type="l",col="blue")
PSELECT3=VARselect(ts.union(TRIMDSP,TRIMCAY,TRIMIND),lag.max=20)
plot(c(1:20),PSELECT3$criteria[1,],xlab="p",ylab="AIC",type="l",col="blue")
```

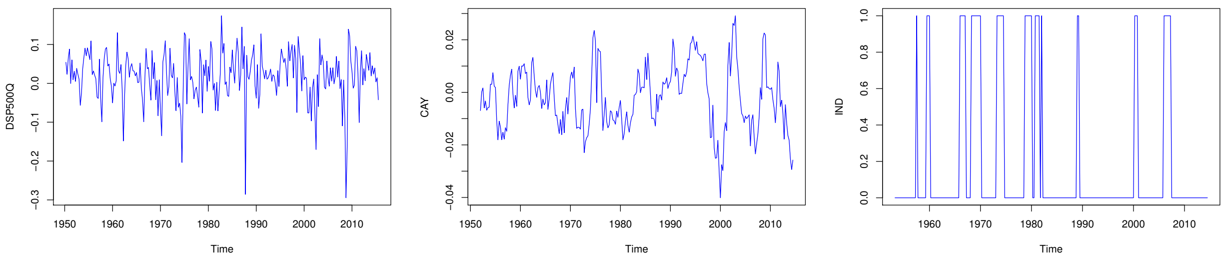


Figure 11: State variables in the VAR model: $\Delta \log$ S&P500, cay and indicator

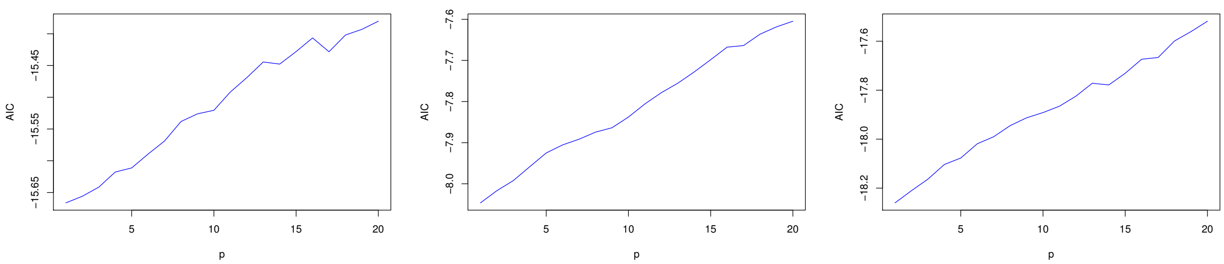


Figure 12: p versus the corresponding AIC: VAR with cay, VAR with ind, VAR with both

Since these three AIC(p)s are minimized at $p = 1$, one can conclude that the most optimal number of lags for these three cases is $p = 1$.

(g) According to the result, both cay and indicator Granger-cause $\Delta \log$ S&P500.

H_0	F statistic	p-value
cay_t does not Granger-cause $\Delta \log \text{S\&P500}_t$	5.93	0.02
indicator_t does not Granger-cause $\Delta \log \text{S\&P500}_t$	8.62	0.00

```
causality(VAR(ts.union(TRIMDSP,TRIMCAY),p=1),cause="TRIMCAY")
causality(VAR(ts.union(TRIMDSP,TRIMIND),p=1),cause="TRIMIND")
```

(h) Hereafter I substitute $\Delta \log \text{S\&P500}_t$ and indicator_t with DLSP_t and ind_t, respectively. According to AICs, the optimal number of lags is $p = 1$. $\phi_{12}^{(1)}$s for both VAR models are significant at 5% or 1% levels, respectively and the corresponding R^2s are about 10% for both models. This implies that one can exploit both cay_{t-1} and ind_{t-1} to predict DLSP_t. These findings are consistent with those from Granger causality tests; both variables Granger-cause DLSP.

$$\mathbf{y}_t = (\text{DLSP}_t, \text{cay}_t)^{\text{t}}$$

	ϕ_0	Φ_1				Σ	
	Constant	DLSP_{t-1}	cay_{t-1}	R^2	F statistic	DLSP_t	cay_t
DLSP_t	0.0146	0.2824	0.8219	0.09	11.66	0.003958	−0.000204
	(3.48)	(4.54)	(2.43)			(1.0000)	(−0.4740)
cay_t	−0.0003	−0.0077	0.8276	0.69	267.61	−0.000204	0.000047
	(−0.63)	(−1.14)	(22.54)			(−0.4740)	(1.0000)

$$\mathbf{y}_t = (\text{DLSP}_t, \text{ind}_t)^{\text{t}}$$

	ϕ_0	Φ_1				Σ	
	Constant	DLSP_{t-1}	ind_{t-1}	R^2	F statistic	DLSP_t	ind_t
DLSP_t	0.0197	0.2228	−0.0308	0.10	13.10	0.003916	−0.000947
	(4.23)	(3.60)	(−2.94)			(1.0000)	(−0.0546)
ind_t	0.0484	0.2839	0.7092	0.49	117.67	−0.000947	0.076958
	(2.35)	(1.03)	(15.24)			(−0.0546)	(1.0000)

```
VAR3=VAR(ts.union(TRIMDSP,TRIMCAY),p=1)
summary(VAR3)
VAR4=VAR(ts.union(TRIMDSP,TRIMIND),p=1)
summary(VAR4)
```

(i) One can conduct 116 iterations from 1953:2–1984:4 (i.e. the first iteration) to 1953:2–2013:3 (i.e. the last iteration). Each iteration generates 4 different out-of-sample predictions (i.e. from the one-step-ahead to the four-steps-ahead).

```
PRED=matrix(0,119,16)
for(t in 131:246){
  MODEL1=arima(TRIMDSP[1:(t-4)],order=c(1,0,0),method="ML")
  TEMP=predict(MODEL1,n.ahead=4)$pred
  PRED[(t-130),1]=TEMP[1]
  PRED[(t-129),2]=TEMP[2]
  PRED[(t-128),3]=TEMP[3]
  PRED[(t-127),4]=TEMP[4]
  MODEL2=VAR(ts.union(TRIMDSP,TRIMCAY)[1:(t-4),],p=1,type="const")
  TEMP=predict(MODEL2,n.ahead=4)$fcst$TRIMDSP
  PRED[(t-130),5]=TEMP[1,1]
  PRED[(t-129),6]=TEMP[2,1]
  PRED[(t-128),7]=TEMP[3,1]
  PRED[(t-127),8]=TEMP[4,1]
  MODEL3=VAR(ts.union(TRIMDSP,TRIMIND)[1:(t-4),],p=1,type="const")
  TEMP=predict(MODEL3,n.ahead=4)$fcst$TRIMDSP
  PRED[(t-130),9]=TEMP[1,1]
  PRED[(t-129),10]=TEMP[2,1]
  PRED[(t-128),11]=TEMP[3,1]
  PRED[(t-127),12]=TEMP[4,1]
  MODEL4=VAR(ts.union(TRIMDSP,TRIMCAY,TRIMIND)[1:(t-4),],p=1,type="const")
  TEMP=predict(MODEL4,n.ahead=4)$fcst$TRIMDSP
  PRED[(t-130),13]=TEMP[1,1]
  PRED[(t-129),14]=TEMP[2,1]
  PRED[(t-128),15]=TEMP[3,1]
  PRED[(t-127),16]=TEMP[4,1]
}
rm(TEMP)
```

(j) For AR(1) model, VAR(1) model with cay, VAR(1) model with ind and VAR(1) model with both cay and ind, I compute 4 RMSEs by using 4 different out-of-sample predictions.

	n-steps-ahead RMSE			
Model	1	2	3	4
AR(1)	0.0709	0.0714	0.0714	0.0712
VAR(1) with cay	0.0713	0.0708	0.0695	0.0687
VAR(1) with ind	0.0711	0.0715	0.0703	0.0706
VAR(1) with both	0.0718	0.0715	0.0692	0.0683

- According to four one-step-ahead RMSEs, all models are performing poorly in predicting one-step-ahead DLSP.

- RMSEs are flat by and large, but RMSE(3) and RMSE(4) for VAR(1) with cay model are lower than RMSE(1) and RMSE(2) for the model. One can also observe this pattern for two other VAR models, but the pattern for VAR(1) with ind model is less obvious than other the patterns for other VAR models.
- By adopting VAR models, one can reduce all RMSEs but one-step-ahead RMSE at a glance (but not yet significantly).

```
RMSE=matrix(0,4,4)
RMSE[1,1]=sqrt(mean((TRIMDSP[128:243]-PRED[1:116,1])^2))
RMSE[1,2]=sqrt(mean((TRIMDSP[129:244]-PRED[2:117,2])^2))
RMSE[1,3]=sqrt(mean((TRIMDSP[130:245]-PRED[3:118,3])^2))
RMSE[1,4]=sqrt(mean((TRIMDSP[131:246]-PRED[4:119,4])^2))
RMSE[2,1]=sqrt(mean((TRIMDSP[128:243]-PRED[1:116,5])^2))
RMSE[2,2]=sqrt(mean((TRIMDSP[129:244]-PRED[2:117,6])^2))
RMSE[2,3]=sqrt(mean((TRIMDSP[130:245]-PRED[3:118,7])^2))
RMSE[2,4]=sqrt(mean((TRIMDSP[131:246]-PRED[4:119,8])^2))
RMSE[3,1]=sqrt(mean((TRIMDSP[128:243]-PRED[1:116,9])^2))
RMSE[3,2]=sqrt(mean((TRIMDSP[129:244]-PRED[2:117,10])^2))
RMSE[3,3]=sqrt(mean((TRIMDSP[130:245]-PRED[3:118,11])^2))
RMSE[3,4]=sqrt(mean((TRIMDSP[131:246]-PRED[4:119,12])^2))
RMSE[4,1]=sqrt(mean((TRIMDSP[128:243]-PRED[1:116,13])^2))
RMSE[4,2]=sqrt(mean((TRIMDSP[129:244]-PRED[2:117,14])^2))
RMSE[4,3]=sqrt(mean((TRIMDSP[130:245]-PRED[3:118,15])^2))
RMSE[4,4]=sqrt(mean((TRIMDSP[131:246]-PRED[4:119,16])^2))
round(RMSE,digits=4)
```

(k) One can conduct Clark–West test since VAR(1) with cay model nests AR(1) model. Since VAR model nests AR model, one must consider only single tail for p-values [Clark and West, 2006].

$$h\text{-step ahead CW}_t = \left(\hat{\varepsilon}^{\text{AR}}_{t|t-h}\right)^2 - \left(\hat{\varepsilon}^{\text{VAR}}_{t|t-h}\right)^2 + \left(\hat{\varepsilon}^{\text{AR}}_{t|t-h} - \hat{\varepsilon}^{\text{VAR}}_{t|t-h}\right)^2, \quad t \in \{T-P+1, \cdots, T\}.$$

t statistics for all CW tests are obtained based on Newey–West standard error in order to consider heteroskedasticity and auto-correlation [Newey and West, 1987].

	n-step-ahead error			
Estimate	1	2	3	4
MSPE-adjusted	0.0002	0.0004	0.0005	0.0005
t statistic	(1.35)	(1.83)	(1.98)	(2.62)
p-value	0.0443	0.0169	0.0119	0.0022

In all four cases, MSPEs for VAR with cay model are smaller than those for AR model at 5% significance level. Therefore, one can reject H_0 and accept H_1: $\text{MSPE}_{\text{VAR with cay}} < \text{MSPE}_{\text{AR}}$.

```
TEMP=matrix(0,3,4)
CW1=(TRIMDSP[128:243]-PRED[1:116,1])^2-
  (TRIMDSP[128:243]-PRED[1:116,5])^2+
  ((TRIMDSP[128:243]-PRED[1:116,1])-(TRIMDSP[128:243]-PRED[1:116,5]))^2
TEMP[1,1]=lm(CW1~1)$coef
TEMP[2,1]=lm(CW1~1)$coef/sqrt(NeweyWest(lm(CW1~1),lag=4))
TEMP[3,1]=(1-pnorm(TEMP[2,1]))/2
CW2=(TRIMDSP[129:244]-PRED[2:117,2])^2-
  (TRIMDSP[129:244]-PRED[2:117,6])^2+
  ((TRIMDSP[129:244]-PRED[2:117,2])-(TRIMDSP[129:244]-PRED[2:117,6]))^2
TEMP[1,2]=lm(CW2~1)$coef
TEMP[2,2]=lm(CW2~1)$coef/sqrt(NeweyWest(lm(CW2~1),lag=4))
TEMP[3,2]=(1-pnorm(TEMP[2,2]))/2
CW3=(TRIMDSP[130:245]-PRED[3:118,3])^2-
  (TRIMDSP[130:245]-PRED[3:118,7])^2+
  ((TRIMDSP[130:245]-PRED[3:118,3])-(TRIMDSP[130:245]-PRED[3:118,7]))^2
TEMP[1,3]=lm(CW3~1)$coef
TEMP[2,3]=lm(CW3~1)$coef/sqrt(NeweyWest(lm(CW3~1),lag=4))
TEMP[3,3]=(1-pnorm(TEMP[2,3]))/2
CW4=(TRIMDSP[131:246]-PRED[4:119,4])^2-
  (TRIMDSP[131:246]-PRED[4:119,8])^2+
  ((TRIMDSP[131:246]-PRED[4:119,4])-(TRIMDSP[131:246]-PRED[4:119,8]))^2
TEMP[1,4]=lm(CW4~1)$coef
TEMP[2,4]=lm(CW4~1)$coef/sqrt(NeweyWest(lm(CW4~1),lag=4))
TEMP[3,4]=(1-pnorm(TEMP[2,4]))/2
round(TEMP,digits=4);rm(TEMP)
```

(l) Similarly, one can also conduct Clark–West test in this case.

Estimate	n-step-ahead error 1	2	3	4
MSPE-adjusted	0.0002	0.0001	0.0002	0.0001
t statistic	(0.81)	(0.79)	(2.21)	(1.79)
p-value	0.1046	0.1071	0.0068	0.0183

In all four cases but $n = 3$, MSPEs for VAR with ind model are not significantly smaller than those for AR model at 10% level. Hence, one cannot reject the null and can conclude that all MSPEs for VAR with ind models but MSPE(3) are not different from MSPEs for AR models.

```
TEMP=matrix(0,3,4)
CW5=(TRIMDSP[128:243]-PRED[1:116,1])^2-
  (TRIMDSP[128:243]-PRED[1:116,9])^2+
  ((TRIMDSP[128:243]-PRED[1:116,1])-(TRIMDSP[128:243]-PRED[1:116,9]))^2
TEMP[1,1]=lm(CW5~1)$coef
TEMP[2,1]=lm(CW5~1)$coef/sqrt(NeweyWest(lm(CW5~1),lag=4))
TEMP[3,1]=(1-pnorm(TEMP[2,1]))/2
CW6=(TRIMDSP[129:244]-PRED[2:117,2])^2-
  (TRIMDSP[129:244]-PRED[2:117,10])^2+
  ((TRIMDSP[129:244]-PRED[2:117,2])-(TRIMDSP[129:244]-PRED[2:117,10]))^2
TEMP[1,2]=lm(CW6~1)$coef
TEMP[2,2]=lm(CW6~1)$coef/sqrt(NeweyWest(lm(CW6~1),lag=4))
TEMP[3,2]=(1-pnorm(TEMP[2,2]))/2
CW7=(TRIMDSP[130:245]-PRED[3:118,3])^2-
  (TRIMDSP[130:245]-PRED[3:118,11])^2+
  ((TRIMDSP[130:245]-PRED[3:118,3])-(TRIMDSP[130:245]-PRED[3:118,11]))^2
TEMP[1,3]=lm(CW7~1)$coef
TEMP[2,3]=lm(CW7~1)$coef/sqrt(NeweyWest(lm(CW7~1),lag=4))
TEMP[3,3]=(1-pnorm(TEMP[2,3]))/2
CW8=(TRIMDSP[131:246]-PRED[4:119,4])^2-
  (TRIMDSP[131:246]-PRED[4:119,12])^2+
  ((TRIMDSP[131:246]-PRED[4:119,4])-(TRIMDSP[131:246]-PRED[4:119,12]))^2
TEMP[1,4]=lm(CW8~1)$coef
TEMP[2,4]=lm(CW8~1)$coef/sqrt(NeweyWest(lm(CW8~1),lag=4))
TEMP[3,4]=(1-pnorm(TEMP[1,4]))/2
round(TEMP,digits=4);rm(TEMP)
```

(m) Since VAR(1) with cay model and VAR(1) with ind model do not nest each other, one can consider Diebold–Mariano–West test [Diebold and Mariano, 1995, West, 1996] to compare two different predictive models. Here I consider all two tails to compute p-values because none of the models nests the other; i.e. $\text{MSPE}^{\text{VAR with cay}}$ can be smaller than $\text{MSPE}^{\text{VAR with ind}}$ or vice versa.

$$h\text{-step ahead DMW}_t = \left(\hat{\varepsilon}_{t|t-h}^{\text{VAR with ind}}\right)^2 - \left(\hat{\varepsilon}_{t|t-h}^{\text{VAR with cay}}\right)^2, \quad t \in \{T-P+1, \cdots, T\}.$$

Similarly, t statistics for all DMW tests are obtained based on Newey–West standard error in order to consider heteroskedasticity and auto-correlation.

Estimate	n-step-ahead error 1	2	3	4
d	0.0000	0.0001	0.0001	0.0003
t statistic	(-0.11)	(0.61)	(0.54)	(1.65)
p-value	0.5431	0.2725	0.2954	0.0492

According to the result, the null cannot be rejected in all cases but 4-steps-ahead $\bar{d}$ (i.e. MSPE) even at 10% significance level. Therefore, one can conclude that MSPEs for the two models are not significantly different from each other except in the fourth case; i.e. one can choose either cay or ind to predict DLSP and they will perform similarly. However, four-steps-ahead MSPE for VAR with cay model is smaller than that for VAR with ind model at 5% significance level. This indicates that cay performs better than ind in the long-run.

```
TEMP=matrix(0,3,4)
DMW1=(TRIMDSP[128:243]-PRED[1:116,9])^2-(TRIMDSP[128:243]-PRED[1:116,5])^2
TEMP[1,1]=lm(DMW1~1)$coef
TEMP[2,1]=lm(DMW1~1)$coef/sqrt(NeweyWest(lm(DMW1~1),lag=4))
```

```
TEMP[3,1]=1-pnorm(TEMP[2,1])
DMW2=(TRIMDSP[129:244]-PRED[2:117,10])^2-(TRIMDSP[129:244]-PRED[2:117,6])^2
TEMP[1,2]=lm(DMW2~1)$coef
TEMP[2,2]=lm(DMW2~1)$coef/sqrt(NeweyWest(lm(DMW2~1),lag=4))
TEMP[3,2]=1-pnorm(TEMP[2,2])
DMW3=(TRIMDSP[130:245]-PRED[3:118,11])^2-(TRIMDSP[130:245]-PRED[3:118,7])^2
TEMP[1,3]=lm(DMW3~1)$coef
TEMP[2,3]=lm(DMW3~1)$coef/sqrt(NeweyWest(lm(DMW3~1),lag=4))
TEMP[3,3]=1-pnorm(TEMP[2,3])
DMW4=(TRIMDSP[131:246]-PRED[4:119,12])^2-(TRIMDSP[131:246]-PRED[4:119,8])^2
TEMP[1,4]=lm(DMW4~1)$coef
TEMP[2,4]=lm(DMW4~1)$coef/sqrt(NeweyWest(lm(DMW4~1),lag=4))
TEMP[3,4]=1-pnorm(TEMP[2,4])
round(TEMP,digits=4);rm(TEMP)
```

(n) Since the trivariate VAR(1) model nests the bivariate VAR(1) with cay model in this case, one can consider CW test to compare the predictive performance.

Introduce ind to VAR with cay model

	n-step-ahead error			
Estimate	1	2	3	4
MSPE-adjusted	0.0000	0.0000	0.0001	0.0001
t statistic	(0.28)	(−0.22)	(1.16)	(1.52)
p-value	0.1942	0.2932	0.0617	0.0321

Introduce cay to VAR with ind model

	n-step-ahead errors			
Estimate	1	2	3	4
MSPE-adjusted	0.0001	0.0003	0.0004	0.0005
t statistic	(0.68)	(1.50)	(1.72)	(2.45)
p-value	0.1237	0.0336	0.0214	0.0036

According to the result, there is no meaningful improvement for 1-step-ahead and 2-steps-ahead MSPEs after introducing ind to VAR with cay model, but the improvements for 3-steps-ahead and 4-steps-ahead MSPEs are significant at 10% and 5% levels, respectively.

Similarly, the improvement for 1-step-ahead MSPE after introducing cay to VAR with ind model is insignificant, but the improvements for the other three MSPEs are significant at 5% level. In particular, the improvement for 4-steps-ahead MSPE is significant at 1% level.

These findings imply that, though the short-run predictive performance of cay and that of ind are not significantly different, both variables are jointly playing an important role in predicting the farther future. In order to avoid any myopic conclusion, as a result, one must consider both variables together to predict the future more accurately.

```
TEMP=matrix(0,3,4)
CW9=(TRIMDSP[128:243]-PRED[1:116,5])^2-
  (TRIMDSP[128:243]-PRED[1:116,13])^2+
  ((TRIMDSP[128:243]-PRED[1:116,5])-(TRIMDSP[128:243]-PRED[1:116,13]))^2
TEMP[1,1]=lm(CW9~1)$coef
TEMP[2,1]=lm(CW9~1)$coef/sqrt(NeweyWest(lm(CW9~1),lag=4))
TEMP[3,1]=(1-pnorm(TEMP[2,1]))/2
CWa=(TRIMDSP[129:244]-PRED[2:117,6])^2-
  (TRIMDSP[129:244]-PRED[2:117,14])^2+
  ((TRIMDSP[129:244]-PRED[2:117,6])-(TRIMDSP[129:244]-PRED[2:117,14]))^2
TEMP[1,2]=lm(CWa~1)$coef
TEMP[2,2]=lm(CWa~1)$coef/sqrt(NeweyWest(lm(CWa~1),lag=4))
TEMP[3,2]=(1-pnorm(TEMP[2,2]))/2
CWb=(TRIMDSP[130:245]-PRED[3:118,7])^2-
  (TRIMDSP[130:245]-PRED[3:118,15])^2+
  ((TRIMDSP[130:245]-PRED[3:118,7])-(TRIMDSP[130:245]-PRED[3:118,15]))^2
TEMP[1,3]=lm(CWb~1)$coef
TEMP[2,3]=lm(CWb~1)$coef/sqrt(NeweyWest(lm(CWb~1),lag=4))
TEMP[3,3]=(1-pnorm(TEMP[2,3]))/2
CWc=(TRIMDSP[131:246]-PRED[4:119,8])^2-
  (TRIMDSP[131:246]-PRED[4:119,16])^2+
  ((TRIMDSP[131:246]-PRED[4:119,8])-(TRIMDSP[131:246]-PRED[4:119,16]))^2
TEMP[1,4]=lm(CWc~1)$coef
TEMP[2,4]=lm(CWc~1)$coef/sqrt(NeweyWest(lm(CWc~1),lag=4))
TEMP[3,4]=(1-pnorm(TEMP[2,4]))/2
round(TEMP,digits=4);rm(TEMP)
```

```
TEMP=matrix(0,3,4)
CWd=(TRIMDSP[128:243]-PRED[1:116,9])^2-
  (TRIMDSP[128:243]-PRED[1:116,13])^2+
  ((TRIMDSP[128:243]-PRED[1:116,9])-(TRIMDSP[128:243]-PRED[1:116,13]))^2
TEMP[1,1]=lm(CWd~1)$coef
TEMP[2,1]=lm(CWd~1)$coef/sqrt(NeweyWest(lm(CWd~1),lag=4))
TEMP[3,1]=(1-pnorm(TEMP[2,1]))/2
CWe=(TRIMDSP[129:244]-PRED[2:117,10])^2-
  (TRIMDSP[129:244]-PRED[2:117,14])^2+
  ((TRIMDSP[129:244]-PRED[2:117,10])-(TRIMDSP[129:244]-PRED[2:117,14]))^2
TEMP[1,2]=lm(CWe~1)$coef
TEMP[2,2]=lm(CWe~1)$coef/sqrt(NeweyWest(lm(CWe~1),lag=4))
TEMP[3,2]=(1-pnorm(TEMP[2,2]))/2
CWf=(TRIMDSP[130:245]-PRED[3:118,11])^2-
  (TRIMDSP[130:245]-PRED[3:118,15])^2+
  ((TRIMDSP[130:245]-PRED[3:118,11])-(TRIMDSP[130:245]-PRED[3:118,15]))^2
TEMP[1,3]=lm(CWf~1)$coef
TEMP[2,3]=lm(CWf~1)$coef/sqrt(NeweyWest(lm(CWf~1),lag=4))
TEMP[3,3]=(1-pnorm(TEMP[2,3]))/2
CWg=(TRIMDSP[131:246]-PRED[4:119,12])^2-
  (TRIMDSP[131:246]-PRED[4:119,16])^2+
  ((TRIMDSP[131:246]-PRED[4:119,12])-(TRIMDSP[131:246]-PRED[4:119,16]))^2
TEMP[1,4]=lm(CWg~1)$coef
TEMP[2,4]=lm(CWg~1)$coef/sqrt(NeweyWest(lm(CWg~1),lag=4))
TEMP[3,4]=(1-pnorm(TEMP[2,4]))/2
round(TEMP,digits=4);rm(TEMP)
```

(o) To estimate Bayesian VAR, I adopt the followings, i.e.

$$\begin{aligned}\mathbf{y}_t &= \phi_0 + \Phi_1 \mathbf{y}_{t-1} + \varepsilon_t \\ \varepsilon_t | \phi_0, \Phi_0, \Sigma &\sim \mathrm{N}(0, \Sigma) \\ \phi_0, \Phi_1 &\sim \text{improper N}(?, ?) \\ \Sigma &\sim \text{improper inverse Wishart}(?, ?),\end{aligned}$$

and since the posterior of an independent Normal-inverse Wishart prior is analytically unknown, one must exploit Gibbs sampling instead using conditional posteriors. One can consider the `BMR` package to conduct Bayesian time-series analysis.

For both bivariate and trivariate VAR models from 1953:2–1984:4 (i.e. the first iteration) to 1953:2–2013:3 (i.e. the last iteration), I simulate the posterior 10,000 times after burning 1,000 observations. In order to save computing resources, I apply the first-order auto-regressive coefficients obtained using OLS as the prior mean of Φ_1 (diagonal) and the variance-cross-covariance estimates as the prior location of Σ.

Due to the lack of any special prior, the result obtained from Bayesian VAR models is similar to the result obtained from classical VAR models reported above.

	n-steps-ahead RMSE			
Model	1	2	3	4
Bayesian VAR(1) with cay	0.0709	0.0707	0.0697	0.0688
Bayesian VAR(1) with both	0.0715	0.0710	0.0692	0.0687

To compare the predictive power, I conduct DMW tests using predictions errors obtained from Bayesian and classical VAR models. According to the result, only one-step-ahead MSPE of Bayesian bivariate VAR model (RMSE=0.0709) is significantly smaller than that of classical bivariate VAR model (RMSE=0.0713) at 5% significance level. Similarly, only one-step-ahead MSPE of Bayesian trivariate VAR model (RMSE=0.0715) is significantly smaller than that of classical trivariate VAR model (RMSE=0.0718) at 10% significance level.

Since Bayesian approach exploits the prior information, the result can be more accurate. In addition, any statistical inference can be improved by Bayesian approach since the prior can complement the lack of observations as well. However, its performance depends on the quality of the prior. Here the resulting improvement is economically not that meaningful because the prior is also governed by the data.

Bayesian VAR with cay model

Estimate	n-step-ahead error 1	2	3	4
MSPE-adjusted	0.0001	0.0000	0.0000	−0.0000
t statistic	(1.65)	(0.47)	(−0.57)	(−1.09)
p-value	0.0496	0.3182	0.7157	0.8625

Bayesian VAR with both model

Estimate	n-step-ahead error 1	2	3	4
MSPE-adjusted	0.0000	0.0001	0.0000	0.0000
t statistic	(1.31)	(1.22)	(−0.03)	(−0.72)
p-value	0.0949	0.1113	0.5110	0.7634

```
BPRED=matrix(0,119,8)
for(t in 131:246){
  BMODEL1=BVARW(ts.union(TRIMDSP,TRIMCAY)[1:(t-4),],coefprior=c(0.3527,0.7523),p=1,keep=10000,burn
      =1000,XiSigma=matrix(c(0.003117,-0.000131,-0.000131,0.000039),nrow=2,ncol=2))
  TEMP=forecast(BMODEL1,periods=4,plot=FALSE)$PointForecast
  BPRED[(t-130),1]=TEMP[1,1]
  BPRED[(t-129),2]=TEMP[2,1]
  BPRED[(t-128),3]=TEMP[3,1]
  BPRED[(t-127),4]=TEMP[4,1]
  BMODEL2=BVARW(ts.union(TRIMDSP,TRIMCAY,TRIMIND)[1:(t-4),],coefprior=c(0.2623,0.7611,0.6847),p=1,
      keep=10000,burn=1000,XiSigma=matrix(c
      (0.002896,-0.000121,-0.001681,-0.000121,0.000039,0.000162,-0.001681,0.000162,0.101975),nrow
      =3,ncol=3))
  TEMP=forecast(BMODEL2,periods=4,plot=FALSE)$PointForecast
  BPRED[(t-130),5]=TEMP[1,1]
  BPRED[(t-129),6]=TEMP[2,1]
  BPRED[(t-128),7]=TEMP[3,1]
  BPRED[(t-127),8]=TEMP[4,1]
}
rm(TEMP)
```

```
BRMSE=matrix(0,2,4)
BRMSE[1,1]=sqrt(mean((TRIMDSP[128:243]-BPRED[1:116,1])^2))
BRMSE[1,2]=sqrt(mean((TRIMDSP[129:244]-BPRED[2:117,2])^2))
BRMSE[1,3]=sqrt(mean((TRIMDSP[130:245]-BPRED[3:118,3])^2))
BRMSE[1,4]=sqrt(mean((TRIMDSP[131:246]-BPRED[4:119,4])^2))
BRMSE[2,1]=sqrt(mean((TRIMDSP[128:243]-BPRED[1:116,5])^2))
BRMSE[2,2]=sqrt(mean((TRIMDSP[129:244]-BPRED[2:117,6])^2))
BRMSE[2,3]=sqrt(mean((TRIMDSP[130:245]-BPRED[3:118,7])^2))
BRMSE[2,4]=sqrt(mean((TRIMDSP[131:246]-BPRED[4:119,8])^2))
round(BRMSE,digits=4)
```

```
TEMP=matrix(0,3,4)
DMW5=(TRIMDSP[128:243]-PRED[1:116,5])^2-(TRIMDSP[128:243]-BPRED[1:116,1])^2
TEMP[1,1]=lm(DMW5~1)$coef
TEMP[2,1]=lm(DMW5~1)$coef/sqrt(NeweyWest(lm(DMW5~1),lag=4))
TEMP[3,1]=1-pnorm(TEMP[2,1])
DMW6=(TRIMDSP[129:244]-PRED[2:117,6])^2-(TRIMDSP[129:244]-BPRED[2:117,2])^2
TEMP[1,2]=lm(DMW6~1)$coef
TEMP[2,2]=lm(DMW6~1)$coef/sqrt(NeweyWest(lm(DMW6~1),lag=4))
TEMP[3,2]=1-pnorm(TEMP[2,2])
DMW7=(TRIMDSP[130:245]-PRED[3:118,7])^2-(TRIMDSP[130:245]-BPRED[3:118,3])^2
TEMP[1,3]=lm(DMW7~1)$coef
TEMP[2,3]=lm(DMW7~1)$coef/sqrt(NeweyWest(lm(DMW7~1),lag=4))
TEMP[3,3]=1-pnorm(TEMP[2,3])
DMW8=(TRIMDSP[131:246]-PRED[4:119,8])^2-(TRIMDSP[131:246]-BPRED[4:119,4])^2
TEMP[1,4]=lm(DMW8~1)$coef
TEMP[2,4]=lm(DMW8~1)$coef/sqrt(NeweyWest(lm(DMW8~1),lag=4))
TEMP[3,4]=1-pnorm(TEMP[2,4])
round(TEMP,digits=4);rm(TEMP)
```

```
TEMP=matrix(0,3,4)
DMW9=(TRIMDSP[128:243]-PRED[1:116,13])^2-(TRIMDSP[128:243]-BPRED[1:116,5])^2
TEMP[1,1]=lm(DMW9~1)$coef
TEMP[2,1]=lm(DMW9~1)$coef/sqrt(NeweyWest(lm(DMW9~1),lag=4))
TEMP[3,1]=1-pnorm(TEMP[2,1])
DMWa=(TRIMDSP[129:244]-PRED[2:117,14])^2-(TRIMDSP[129:244]-BPRED[2:117,6])^2
TEMP[1,2]=lm(DMWa~1)$coef
TEMP[2,2]=lm(DMWa~1)$coef/sqrt(NeweyWest(lm(DMWa~1),lag=4))
```

```
TEMP[3,2]=1-pnorm(TEMP[2,2])
DMWb=(TRIMDSP[130:245]-PRED[3:118,15])^2-(TRIMDSP[130:245]-BPRED[3:118,7])^2
TEMP[1,3]=lm(DMWb~1)$coef
TEMP[2,3]=lm(DMWb~1)$coef/sqrt(NeweyWest(lm(DMWb~1),lag=4))
TEMP[3,3]=1-pnorm(TEMP[2,3])
DMWc=(TRIMDSP[131:246]-PRED[4:119,16])^2-(TRIMDSP[131:246]-BPRED[4:119,8])^2
TEMP[1,4]=lm(DMWc~1)$coef
TEMP[2,4]=lm(DMWc~1)$coef/sqrt(NeweyWest(lm(DMWc~1),lag=4))
TEMP[3,4]=1-pnorm(TEMP[2,4])
round(TEMP,digits=4);rm(TEMP)
```

References

[Clark and West, 2006] Clark, T. E. and West, K. D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, 135(1-2):155–186.

[Diebold and Mariano, 1995] Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

[Lettau and Ludvigson, 2001] Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849.

[Lothian and Taylor, 1996] Lothian, J. R. and Taylor, M. P. (1996). Real exchange rate behavior: The recent float from the perspective of the past two centuries. *Journal of Political Economy*, 104(3):488–509.

[Newey and West, 1987] Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

[West, 1996] West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084.

Econometric Methods II Assignment 03

Junyong Kim*

April 7, 2017

1 Analytical Exercise

1. Here the VAR model is

$$
\begin{aligned}
\mathbf{B}\mathbf{y}_t =& \gamma_0 + \Gamma_1 \mathbf{y}_{t-1} + \varepsilon_t \\
\text{where } \mathbf{B} =& \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix} \\
\mathbf{y}_t =& \begin{pmatrix} y_{1t} & y_{2t} \end{pmatrix}^\top \\
\gamma_0 =& \begin{pmatrix} \gamma_{10} & \gamma_{20} \end{pmatrix}^\top \\
\Gamma_1 =& \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \\
\varepsilon_t =& \begin{pmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{pmatrix}^\top \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right).
\end{aligned}
$$

   (a) No. OLS estimation ignores the simultaneity between y_{1t} and y_{2t}, hence the estimator will be inconsistent due to these endogenous variables.

   (b) $\mathrm{E}\left[\mathbf{u}_t\mathbf{u}_t^\top\right] = \mathrm{E}\left[\mathbf{B}^{-1}\varepsilon_t\varepsilon_t^\top\mathbf{B}^{-1\top}\right] = \mathbf{B}^{-1}\mathrm{E}\left[\varepsilon_t\varepsilon_t^\top\right]\mathbf{B}^{-1\top} = \mathbf{B}^{-1}\Sigma\mathbf{B}^{-1\top}$. While the reduced-form VAR model has 9 parameters ($\mathbf{a}_0$, $\mathbf{A}_1$ and Ω), the structural VAR model has 10 parameters ($\mathbf{B}$, γ_0, Γ_1 and Σ). Since the latter has one more parameter than the former, it is under-identified.

   (c) One has 12 parameters for the structural VAR ($\mathbf{B}$, γ_0, Γ_1, Σ_{ε_1} and Σ_{ε_2}) and 12 parameters for the reduced-form VAR ($\mathbf{a}_0$, $\mathbf{A}_1$, $\Omega_{\mathbf{u}_1}$ and $\Omega_{\mathbf{u}_2}$), so the structural VAR model is exact-identified [Rigobon and Sack, 2003].

2 Empirical Exercise

1. $\mathbf{y}_t = \begin{pmatrix} \Delta \mathrm{s}_t & \mathrm{fp}_t \end{pmatrix}^\top$.

   (a) Figure 1 includes the IRFs computed using Cholesky decomposition and Figure 2 includes the IRFs computed using `SVAR` function, respectively. According to the result, both IRFs are identical to each other. For `SVAR`, one can adopt

$$
\mathbf{A} = \begin{pmatrix} 1 & 0 \\ . & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} . & 0 \\ 0 & . \end{pmatrix}.
$$

   (b) If one assumes that the reduced-form errors are orthogonal to each other, then one can adopt

$$
\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} . & 0 \\ 0 & . \end{pmatrix}.
$$

   Figure 3 includes the IRFs computed under this assumption.

*Lubar School of Business, University of Wisconsin–Milwaukee.

(c) One can check `LR` after conducting `SVAR` to see whether the over-identification is binding. According to the result, the $\chi^2(1)$ statistic is 0.7535 and the corresponding p-value is 0.3854. So one cannot reject the null and can conclude that the errors are orthogonal.

2. $\mathbf{y}_t = \begin{pmatrix}\Delta \log \text{output}_t & \text{unemployment}_t\end{pmatrix}^\top$.

   (a) $p^* = \underset{p\in\{1,\cdots,8\}}{\text{argmin}}\ \text{AIC}(p) = 4$ and $\text{AIC}(p^*) = 3.6647$. Table 1 includes the estimates of parameters $(\phi_0, \Phi_1, \cdots, \Phi_4, \Sigma)$. Adjusted R^2 for the two equations are 0.2261 and 0.9640, respectively. The fit is valid here because the F statistic for each equation rejects the null at 1% significance level. Figure 4 includes the IRFs computed using Cholesky decomposition.

   (b) $\hat{\mathbf{L}} = \begin{pmatrix} 19.35 & 0.00 \\ 3.00 & 6.40 \end{pmatrix}$ and $\hat{\mathbf{B}} = \begin{pmatrix} 1.29 & -22.77 \\ 0.25 & 0.19 \end{pmatrix}$.

   (c) Figures 5 and 6 include the IRFs and the FEVDs computed using Blanchard–Quah decomposition, respectively.

   If one identifies the structural shocks using Cholesky decomposition, then there will be no contemporaneous effect of endogenous variables on exogenous variables (refer the upper-right panels of Figures 1 and 2).

   On the other hand, if one identifies the shocks using Blanchard–Quah decomposition, then contemporaneous effects among the variables will not be restricted (note that the upper-right panel of Figure 5 does not start from zero). Instead several long-run effects will be constrained (note that the panel converges to zero).

   (d) To crowd out the trend, I regress $\{\text{unemployment}_t\}$ on $\{1\}$ and $\{t\}$ and exploit the residuals.

   - $p^* = \underset{p\in\{1,\cdots,8\}}{\text{argmin}}\ \text{AIC}(p) = 4$.
   - $\text{AIC}(p^*) = 3.6534$.
   - Table 2 includes the estimates of parameters $(\phi_0, \Phi_1, \cdots, \Phi_4, \Sigma)$.
   - Adjusted R^2 for the two equations are 0.2204 and 0.9504, respectively.
   - $\hat{\mathbf{L}} = \begin{pmatrix} 18.31 & 0.00 \\ -0.59 & 4.72 \end{pmatrix}$ and $\hat{\mathbf{B}} = \begin{pmatrix} 11.30 & -19.90 \\ 0.14 & 0.27 \end{pmatrix}$. [Campbell et al., 2010]

   (e) Figure 7 includes the IRFs computed using Cholesky decomposition. Figures 8 and 9 include the IRFs and the FEVDs computed using Blanchard–Quah decomposition, respectively.

   Figures 7–9 correspond to Figures 4–6; note that the effects of both shocks on unemployment_t (refer the lower panels of Figure 4–9) disappear more quickly after crowding out the trend, i.e. one is able to observe the mean-reversion more clearly because unemployment_t is somewhat trend-stationary.

3. I replicate the quadrivariate VAR(1) model with $\mathbf{y} = (r_M^e, TY, PE, VS)^\top$ [Campbell et al., 2010] using the data from Christopher Polk's homepage.[1] r_M^e is the log excess market return, TY is the term yield spread, PE is the log price-earnings ratio and VS is the small-stock value spread, respectively. The sample is from 1928 to 2001 resulting in 74 annual observations.

   In order to see the effect of the term yield spread shock on other variables, I use `SVAR` function to decompose the structural shocks. In particular, I adopt

$$\mathbf{A} = \begin{pmatrix} 1 & . & 0 & 0 \\ 0 & 1 & 0 & 0 \\ . & . & 1 & 0 \\ . & . & . & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} . & 0 & 0 & 0 \\ 0 & . & 0 & 0 \\ 0 & 0 & . & 0 \\ 0 & 0 & 0 & . \end{pmatrix}.$$

   Since both reduced-form and structural models have 30 parameters, the structural model is exact-identified. I assume that the term yield spread shock is most exogenous among others; since the original paper only uses the reduced-form model, no result regarding the structural model is reported.

[1] `http://personal.lse.ac.uk/polk`.

The SVAR estimates for both $\mathbf{A}$ and $\mathbf{B}$ are

$$\hat{\mathbf{A}} = \begin{pmatrix} 1.0000 & 0.0949 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ -0.8856 & -0.0018 & 1.0000 & 0.0000 \\ 0.5280 & 0.0147 & -0.5320 & 1.0000 \end{pmatrix}, \quad \hat{\mathbf{B}} = \begin{pmatrix} 0.2225 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.5496 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0619 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.1647 \end{pmatrix}.$$

Table 3 includes the estimates $\hat{\phi}_0$, $\hat{\Phi}_1$ and $\hat{\Sigma}$ for the reduced-form model. Figure 10 exhibits IRFs of the term yield spread shock on other variables computed using SVAR.

Since the positive term yield spread implies the positive expectation for the market condition in the future, the TY shock may be positively related to both r_M^e and PE. Figure 10, however, displays the negative contemporaneous effect of that shock on these variables.

Figure 11 includes the IRFs computed using VARsignR package. In particular, I exploit `rwz.reject` to conduct the rejection method [Rubio-Ramírez et al., 2010] with `constrained=c(+2,+1,+3)`. That is, the second shock (i.e. the TY shock) is restricted to have a positive effect on the first variable (r_M^e) and the third variable (PE). In addition, I impose more restrictions using `KMIN=1` and `KMAX=5`. The panels of Figure 11 display both median IRFs and 95% corresponding intervals.

The upper-left and upper-right panels of Figure 11 show the positive contemporaneous effect of the TY shock on both r_M^e and PE. These patterns in Figure 11 are different from the patterns in Figure 10. On the other hand, the lower-right panel of Figure 11 (i.e. the effect of the TY shock on VS) is similar to the lower-right panel of Figure 10, which is computed using SVAR function.

References

[Campbell et al., 2010] Campbell, J. Y., Polk, C., and Vuolteenaho, T. (2010). Growth or glamour? fundamentals and systematic risk in stock returns. *Review of Financial Studies*, 23(1):305–344.

[Rigobon and Sack, 2003] Rigobon, R. and Sack, B. (2003). Measuring the reaction of monetary policy to the stock market. *Quarterly Journal of Economics*, 118(2):639–669.

[Rubio-Ramírez et al., 2010] Rubio-Ramírez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, 77(2):665–696.

Parameter	Δy_{1t}	y_{2t}
ϕ_0	-9.4328	0.3363
	(-1.2720)	(3.3230)
Φ_1	0.1094	-0.0033
	(1.1120)	(-2.4950)
	-19.6694	1.4690
	(-2.7870)	(15.2500)
Φ_2	0.2165	-0.0031
	(2.1590)	(-2.2480)
	43.7807	-0.7264
	(3.6860)	(-4.4810)
Φ_3	-0.0058	-0.0023
	(-0.0570)	(-1.6530)
	-22.1867	0.0737
	(-1.8080)	(0.4400)
Φ_4	0.0622	0.0004
	(0.7270)	(0.3760)
	1.6312	0.1544
	(0.2210)	(1.5320)
$\bar{R}^2$	0.2261	0.9640
F statistic	6.6220	516.8000
(p-value)	(0.0000)	(0.0000)
Σ	520.0760	-3.9531
	[1.0000]	$[-0.5570]$
	-3.9530	0.0969
	$[-0.5570]$	[1.0000]

Table 1: VAR model estimates for $\mathbf{y}_t = (\Delta \log \text{output}_t, \text{unemployment}_t)^\top$. This table reports the estimates for VAR(4) model. t statistics (or p-values for F statistics) are reported with round brackets and correlations are reported with square brackets

Parameter	Δy_{1t}	y_{2t}
ϕ_0	8.3948	0.1729
	(2.1390)	(3.2920)
Φ_1	0.1402	−0.0035
	(1.4390)	(−2.6820)
	−17.0682	1.4290
	(−2.3750)	(14.8500)
Φ_2	0.2509	−0.0033
	(2.5240)	(−2.4710)
	42.7238	−0.7077
	(3.5800)	(−4.4290)
Φ_3	0.0300	−0.0026
	(0.2940)	(−1.9360)
	−22.0704	0.0728
	(−1.7930)	(0.4410)
Φ_4	0.0980	−0.0001
	(1.1230)	(−0.0800)
	0.6269	0.1485
	(0.0850)	(1.5060)
$\bar{R}^2$	0.2209	0.9504
F statistic	6.4580	369.5000
(p-value)	(0.0000)	(0.0000)
Σ	523.5360	−3.8331
	[1.0000]	[−0.5469]
	−3.8330	0.0938
	[−0.5469]	[1.0000]

Table 2: VAR model estimates for $\mathbf{y}_t' = \left(\Delta \log \text{output}_t, \text{unemployment}_t'\right)^\top$. This table reports the estimates for VAR(4) model. unemployment$_t'$ is the detrended unemployment$_t$. t statistics (or p-values for F statistics) are reported with round brackets and correlations are reported with square brackets

Parameter	$r^e_{M,t}$	TY_t	PE_t	VS_t
ϕ_0	0.8967	−0.0479	0.6345	0.3166
	(3.0550)	(−0.0680)	(2.3370)	(1.4630)
Φ_1	−0.0354	0.0250	0.0810	0.0291
	(−0.3090)	(0.0900)	(0.7630)	(0.3440)
	0.0643	0.3437	0.0478	−0.0429
	(1.4120)	(3.1400)	(1.1350)	(−1.2770)
	−0.2133	−0.1303	0.8354	−0.0515
	(−2.6310)	(−0.6680)	(11.1420)	(−0.8620)
	−0.1642	0.5174	−0.1149	0.9133
	(−2.1100)	(2.7650)	(−1.5970)	(15.9230)
$\bar{R}^2$	0.1080	0.3060	0.7023	0.8132
F statistic	3.2090	9.0460	44.0600	80.4300
(p-value)	(0.0178)	(0.0000)	(0.0000)	(0.0000)
Σ	0.0522	−0.0287	0.0462	−0.0026
	[1.0000]	[−0.2282]	[0.9562]	[−0.0670]
	−0.0287	0.3021	−0.0248	−0.0025
	[−0.2282]	[1.0000]	[−0.2138]	[−0.0273]
	0.0462	−0.0248	0.0447	−0.0003
	[0.9562]	[−0.2138]	[1.0000]	[−0.0071]
	−0.0026	−0.0025	−0.0003	0.0284
	[−0.0670]	[−0.0273]	[−0.0071]	[1.0000]

Table 3: VAR model estimates for $\mathbf{y}_t = \left(r^e_{M,t}, TY_t, PE_t, VS_t\right)^\top$. This table reports the estimates for VAR(1) model. r^e_M is the log excess market return, TY is the term yield spread, PE is the log price-earnings ratio and VS is the small-stock value spread, respectively. The sample is from 1928 to 2001 resulting in 74 annual observations. t statistics (or p-values for F statistics) are reported with round brackets and correlations are reported with square brackets

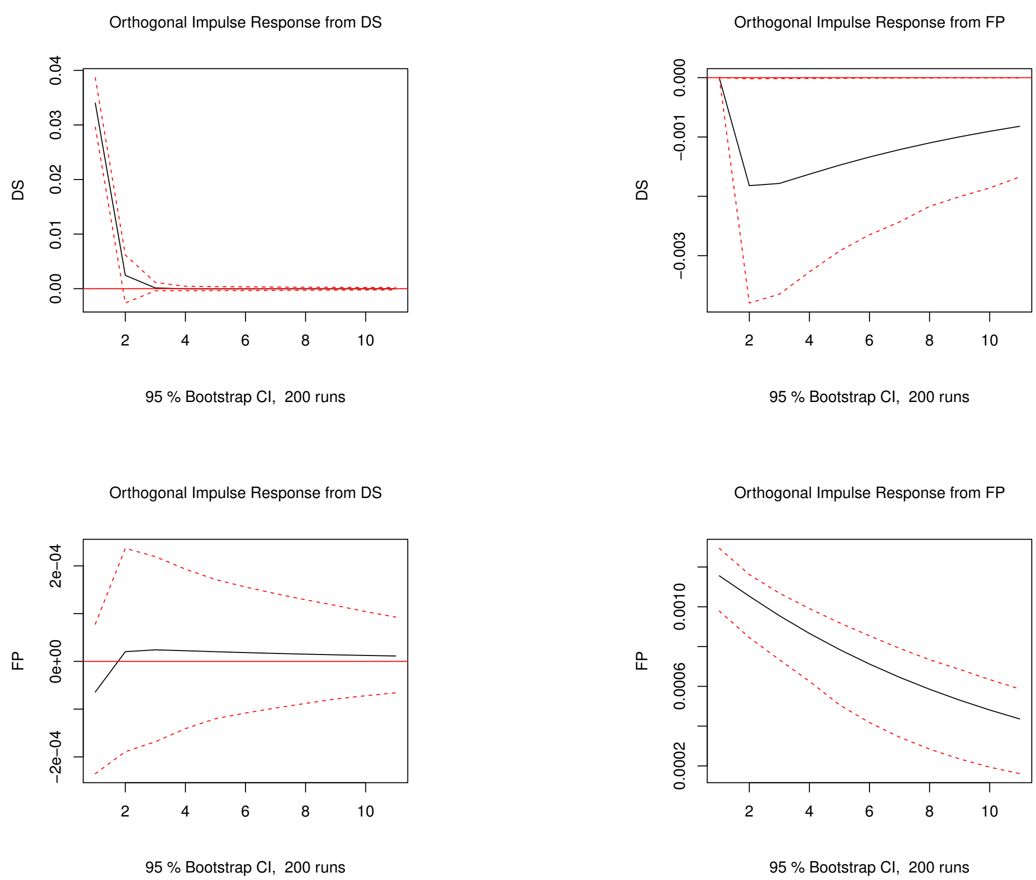


Figure 1: IRFs computed using Cholesky decomposition; $\mathbf{y}_t = (\Delta \mathrm{s}_t, \mathrm{fp}_t)^\top$

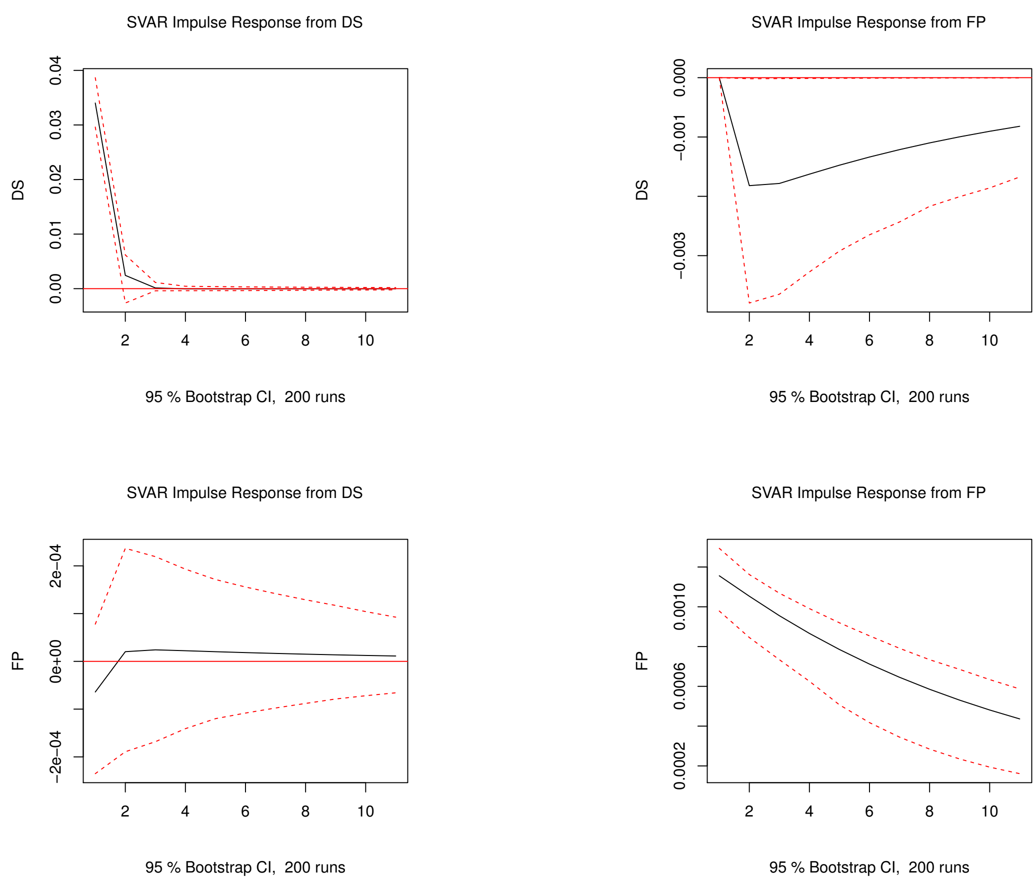


Figure 2: IRFs computed using `SVAR` function; $\mathbf{y}_t = (\Delta \mathrm{s}_t, \mathrm{fp}_t)^\top$

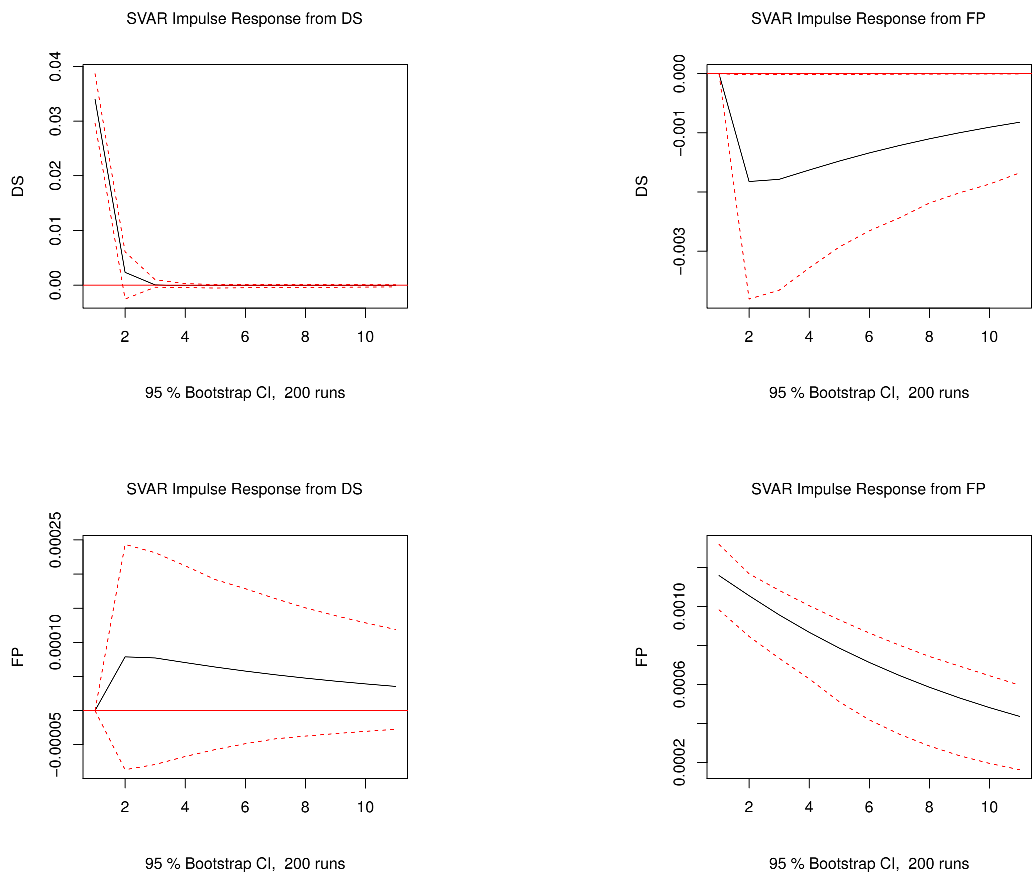


Figure 3: IRFs computed using `SVAR` function; $\mathbf{y}_t = (\Delta \mathrm{s}_t, \mathrm{fp}_t)^\top$ and reduced-form errors are assumed to be orthogonal

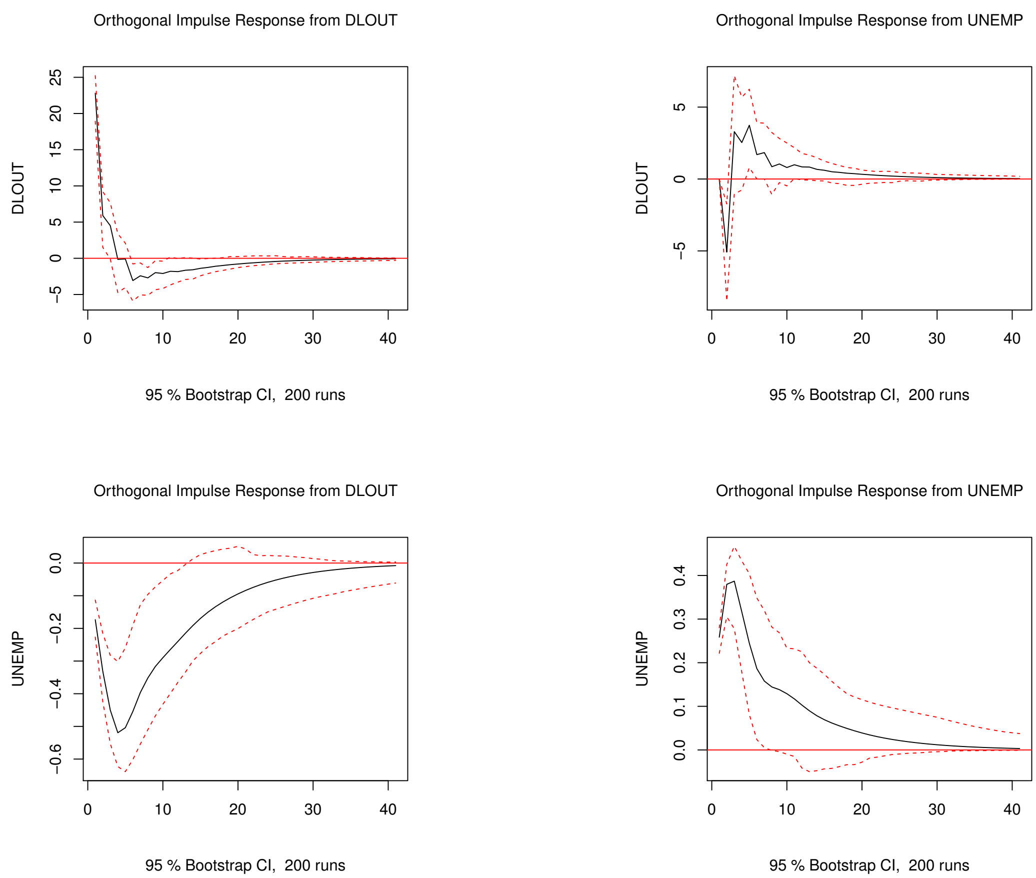


Figure 4: IRFs computed using Cholesky decomposition; $\mathbf{y}_t = (\Delta \log \text{output}_t, \text{unemployment}_t)^\top$

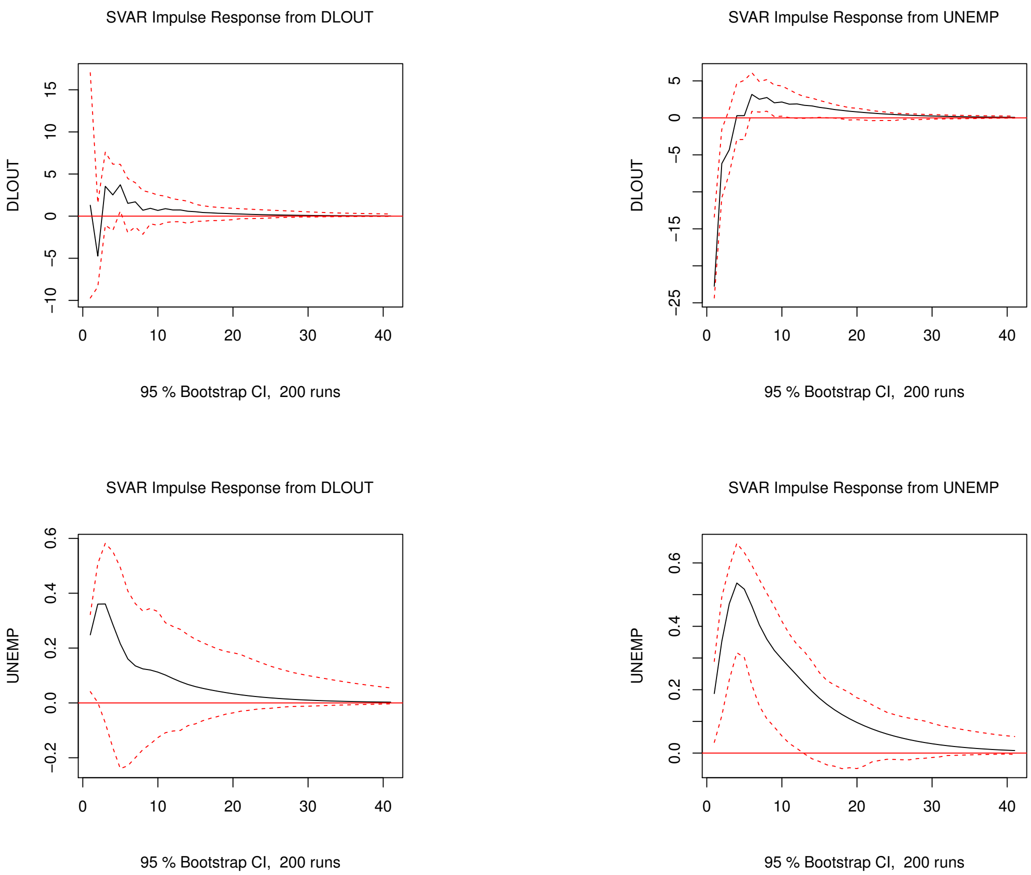


Figure 5: IRFs computed using Blanchard–Quah decomposition; $\mathbf{y}_t = (\Delta \log \text{output}_t, \text{unemployment}_t)^\top$

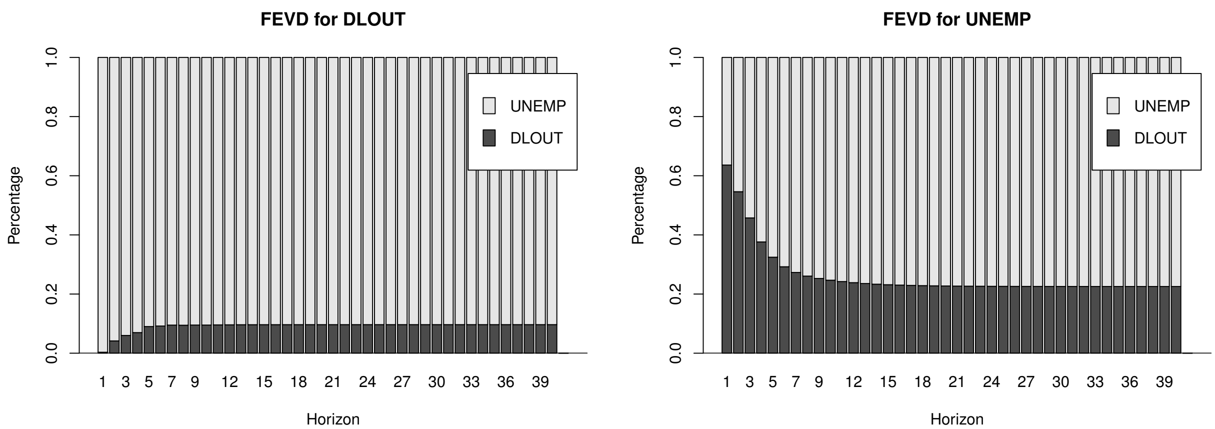


Figure 6: FEVDs computed using Blanchard–Quah decomposition; $\mathbf{y}_t = (\Delta \log \text{output}_t, \text{unemployment}_t)^\top$

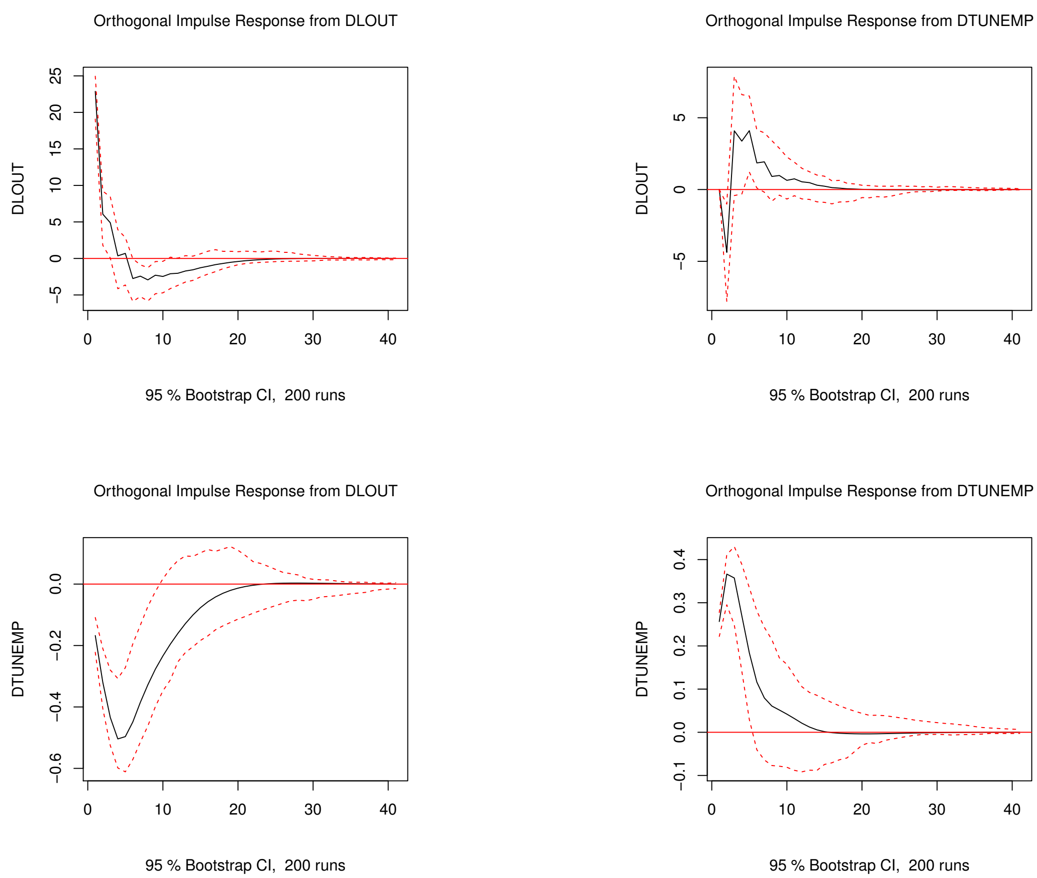


Figure 7: IRFs computed using Cholesky decomposition; $\mathbf{y}'_t = \left(\Delta \log \text{output}_t, \text{unemployment}'_t\right)^\top$, where $\text{unemployment}'_t$ is the detrended unemployment_t

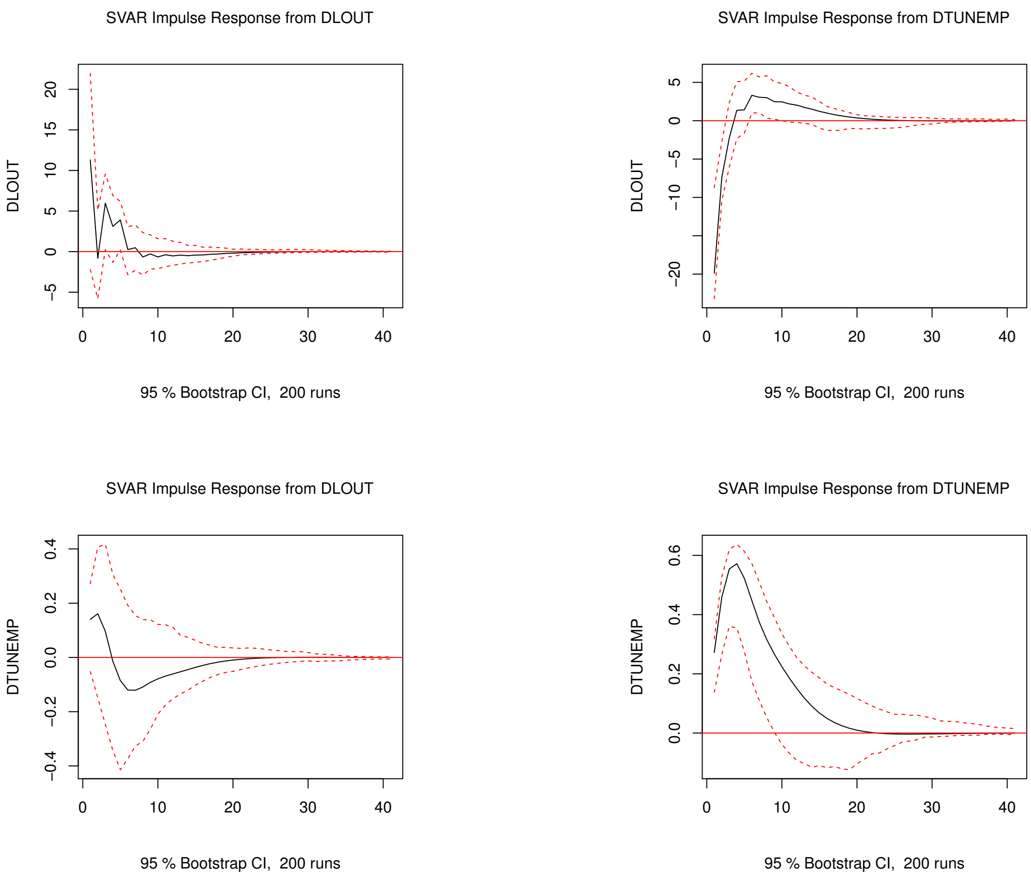


Figure 8: IRFs computed using Blanchard-Quah decomposition; $\mathbf{y}'_t = \left(\Delta \log \text{output}_t, \text{unemployment}'_t\right)^\top$, where $\text{unemployment}'_t$ is the detrended unemployment_t

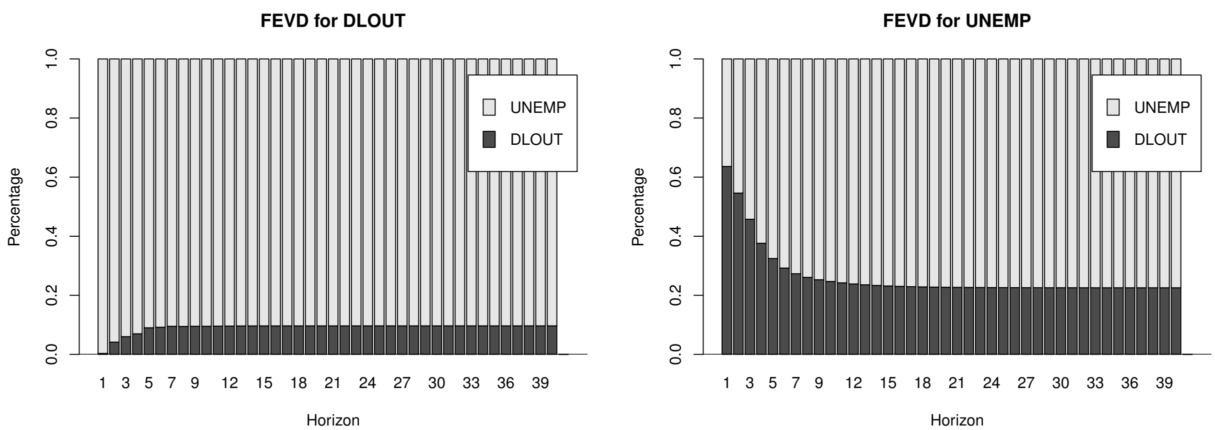


Figure 9: FEVDs computed using Blanchard–Quah decomposition; $\mathbf{y}'_t = \left(\Delta \log \mathrm{output}_t, \mathrm{unemployment}'_t\right)^\top$, where $\mathrm{unemployment}'_t$ is the detrended $\mathrm{unemployment}_t$

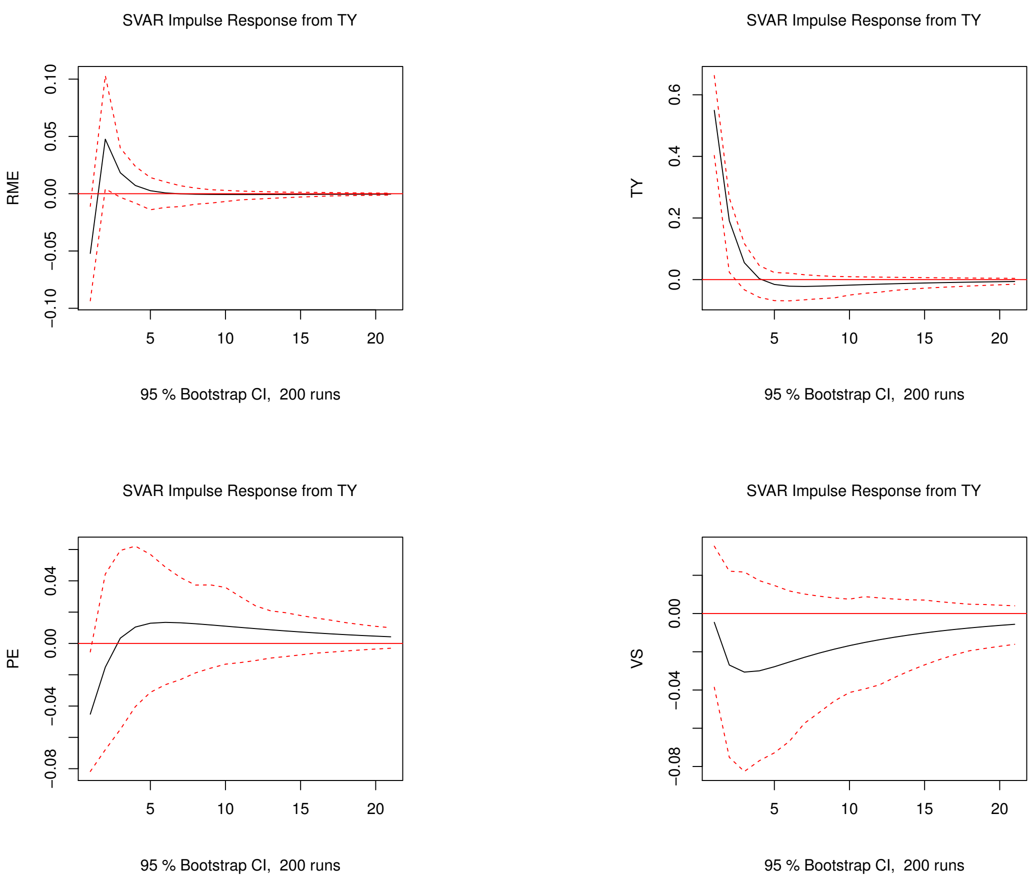


Figure 10: IRFs of the term yield spread shock on other variables computed using `SVAR` function; $\mathbf{y}_t = \left(r^e_{M,t}, TY_t, PE_t, VS_t\right)^\top$ and the TY shock is assumed to be most exogenous

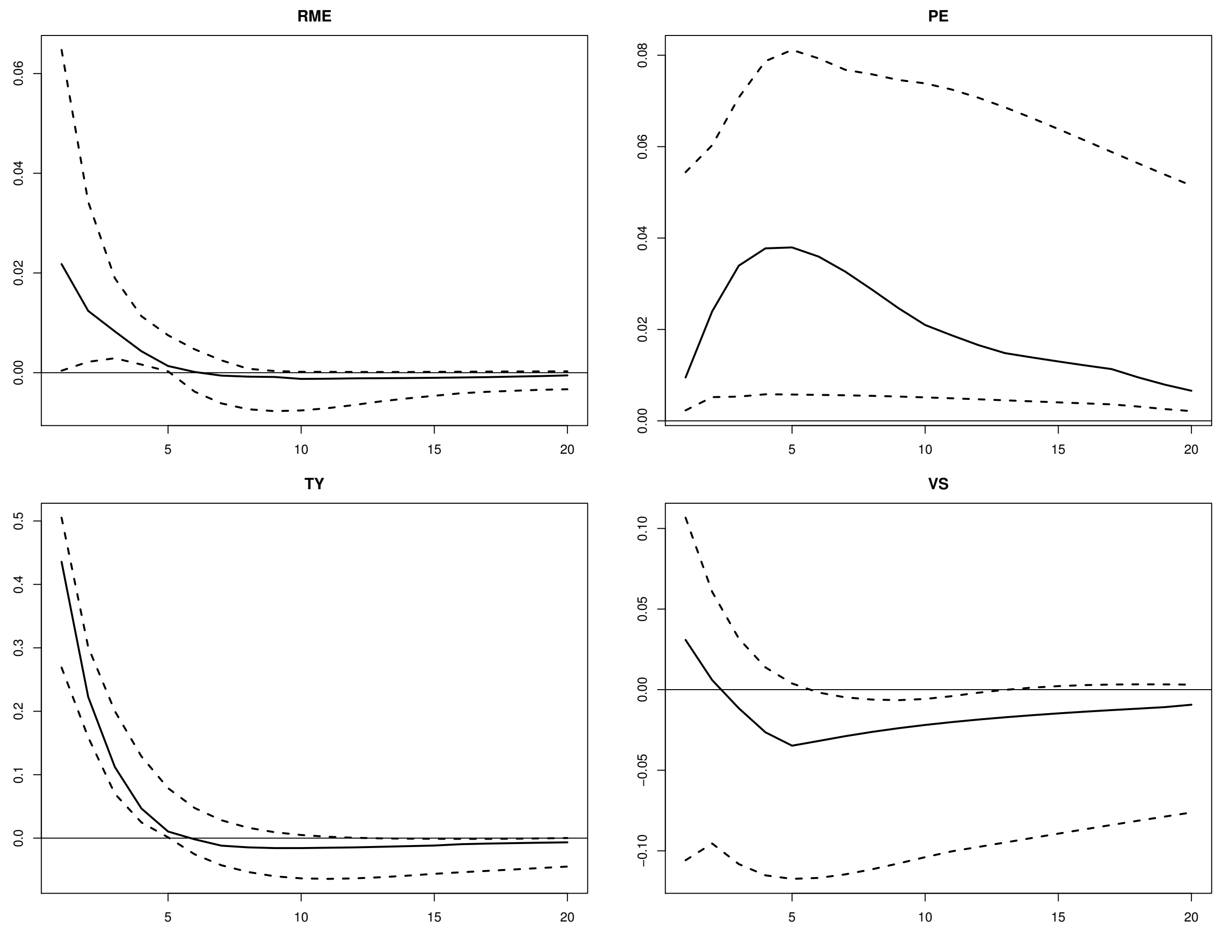


Figure 11: IRFs of the term yield spread shock on other variables computed using `rwz.reject` function; $\mathbf{y}_t = \left(r^e_{M,t}, TY_t, PE_t, VS_t\right)^\top$ and the TY shock is assumed to have a positive effect on both r^e_M and PE

Appendix: R code

```
setwd("C:/Users/Junyong/OneDrive - University of Wisconsin Milwaukee/UWM/2017
    _01_Spring/Econometric Methods II")
library(vars)
library(VARsignR)
#################################################
ZIVOT=read.table("170303_usuk.txt",header=TRUE)
DS=ts(diff(ZIVOT$USUKS),start=c(1976,3),frequency=12)
FP=ts(ZIVOT$USUKFP,start=c(1976,2),frequency=12)
FP=ts(FP[2:245],start=c(1976,3),frequency=12)
Y=ts.union(DS,FP)
VAR1=VAR(Y,p=1)
summary(VAR1)
IRF11=irf(VAR1,response="DS",impulse="DS",ortho=TRUE,cumulative=FALSE,boot=
    TRUE,ci=0.95,runs=200,seed=1)
IRF12=irf(VAR1,response="DS",impulse="FP",ortho=TRUE,cumulative=FALSE,boot=
    TRUE,ci=0.95,runs=200,seed=1)
IRF13=irf(VAR1,response="FP",impulse="DS",ortho=TRUE,cumulative=FALSE,boot=
    TRUE,ci=0.95,runs=200,seed=1)
IRF14=irf(VAR1,response="FP",impulse="FP",ortho=TRUE,cumulative=FALSE,boot=
    TRUE,ci=0.95,runs=200,seed=1)
plot(IRF11);plot(IRF12);plot(IRF13);plot(IRF14)
#################################################
AMAT=diag(2);AMAT[2,1]=NA;BMAT=diag(2);diag(BMAT)=NA
SVAR1=SVAR(VAR1,Amat=AMAT,Bmat=BMAT)
SIRF11=irf(SVAR1,response="DS",impulse="DS",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1)
SIRF12=irf(SVAR1,response="DS",impulse="FP",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1)
SIRF13=irf(SVAR1,response="FP",impulse="DS",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1)
SIRF14=irf(SVAR1,response="FP",impulse="FP",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1)
plot(SIRF11);plot(SIRF12);plot(SIRF13);plot(SIRF14)
#################################################
SVAR2=SVAR(VAR1,Bmat=BMAT)
SIRF21=irf(SVAR2,response="DS",impulse="DS",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1)
SIRF22=irf(SVAR2,response="DS",impulse="FP",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1)
SIRF23=irf(SVAR2,response="FP",impulse="DS",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1)
SIRF24=irf(SVAR2,response="FP",impulse="FP",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1)
plot(SIRF21);plot(SIRF22);plot(SIRF23);plot(SIRF24)
#################################################
SVAR2$LR
#################################################
BQDATA=read.csv("170327_BQ.csv",header=TRUE)
DLOUT=ts(diff(BQDATA$GNP82),start=c(1951,2),frequency=4)
UNEMP=ts(BQDATA$USUNRATEE[2:160],start=c(1951,2),frequency=4)
Y2=ts.union(DLOUT,UNEMP)
VARselect(Y2,lag.max=8)
```

```
VAR2=VAR(Y2,lag.max=8,ic="AIC")
summary(VAR2)
IRF21=irf(VAR2,response="DLOUT",impulse="DLOUT",ortho=TRUE,cumulative=FALSE,
   boot=TRUE,ci=0.95,runs=200,seed=1)
IRF22=irf(VAR2,response="DLOUT",impulse="UNEMP",ortho=TRUE,cumulative=FALSE,
   boot=TRUE,ci=0.95,runs=200,seed=1)
IRF23=irf(VAR2,response="UNEMP",impulse="DLOUT",ortho=TRUE,cumulative=FALSE,
   boot=TRUE,ci=0.95,runs=200,seed=1)
IRF24=irf(VAR2,response="UNEMP",impulse="UNEMP",ortho=TRUE,cumulative=FALSE,
   boot=TRUE,ci=0.95,runs=200,seed=1)
plot(IRF21);plot(IRF22);plot(IRF23);plot(IRF24)
SVAR3=BQ(VAR2)
SIRF31=irf(SVAR3,response="DLOUT",impulse="DLOUT",cumulative=FALSE,boot=TRUE,
   ci=0.95,runs=200,seed=1,n.ahead=40)
SIRF32=irf(SVAR3,response="DLOUT",impulse="UNEMP",cumulative=FALSE,boot=TRUE,
   ci=0.95,runs=200,seed=1,n.ahead=40)
SIRF33=irf(SVAR3,response="UNEMP",impulse="DLOUT",cumulative=FALSE,boot=TRUE,
   ci=0.95,runs=200,seed=1,n.ahead=40)
SIRF34=irf(SVAR3,response="UNEMP",impulse="UNEMP",cumulative=FALSE,boot=TRUE,
   ci=0.95,runs=200,seed=1,n.ahead=40)
plot(SIRF31);plot(SIRF32);plot(SIRF33);plot(SIRF34)
FEVD1=fevd(SVAR3,n.ahead=40)
plot(FEVD1,plot.type="single")
##################################################
DTUNEMP=ts(lm(UNEMP~c(1:159))$resid,start=c(1951,2),frequency=4)
Y3=ts.union(DLOUT,DTUNEMP)
VARselect(Y3,lag.max=8)
VAR3=VAR(Y3,lag.max=8,ic="AIC")
summary(VAR3)
IRF31=irf(VAR3,response="DLOUT",impulse="DLOUT",ortho=TRUE,cumulative=FALSE,
   boot=TRUE,ci=0.95,runs=200,seed=1)
IRF32=irf(VAR3,response="DLOUT",impulse="DTUNEMP",ortho=TRUE,cumulative=FALSE,
   boot=TRUE,ci=0.95,runs=200,seed=1)
IRF33=irf(VAR3,response="DTUNEMP",impulse="DLOUT",ortho=TRUE,cumulative=FALSE,
   boot=TRUE,ci=0.95,runs=200,seed=1)
IRF34=irf(VAR3,response="DTUNEMP",impulse="DTUNEMP",ortho=TRUE,cumulative=
   FALSE,boot=TRUE,ci=0.95,runs=200,seed=1)
plot(IRF31);plot(IRF32);plot(IRF33);plot(IRF34)
SVAR4=BQ(VAR3)
SIRF41=irf(SVAR4,response="DLOUT",impulse="DLOUT",cumulative=FALSE,boot=TRUE,
   ci=0.95,runs=200,seed=1,n.ahead=40)
SIRF42=irf(SVAR4,response="DLOUT",impulse="DTUNEMP",cumulative=FALSE,boot=TRUE
   ,ci=0.95,runs=200,seed=1,n.ahead=40)
SIRF43=irf(SVAR4,response="DTUNEMP",impulse="DLOUT",cumulative=FALSE,boot=TRUE
   ,ci=0.95,runs=200,seed=1,n.ahead=40)
SIRF44=irf(SVAR4,response="DTUNEMP",impulse="DTUNEMP",cumulative=FALSE,boot=
   TRUE,ci=0.95,runs=200,seed=1,n.ahead=40)
plot(SIRF41);plot(SIRF42);plot(SIRF43);plot(SIRF44)
FEVD2=fevd(SVAR4,n.ahead=40)
plot(FEVD1,plot.type="single")
##################################################
CPVDATA=read.csv("170328_CPV2010RFS.csv",header=TRUE)
RME=ts(CPVDATA$R_Me,start=1927)
TY=ts(CPVDATA$TY,start=1927)
```

```
PE=ts(CPVDATA$PE,start=1927)
VS=ts(CPVDATA$VS,start=1927)
Y4=ts.union(RME,TY,PE,VS)
VAR4=VAR(Y4,p=1)
summary(VAR4)
AMAT=rbind(c(1,NA,0,0),c(0,1,0,0),c(NA,NA,1,0),c(NA,NA,NA,1))
BMAT=diag(4);diag(BMAT)=NA
SVAR5=SVAR(VAR4,Amat=AMAT,Bmat=BMAT)
SIRF51=irf(SVAR5,response="RME",impulse="TY",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1,n.ahead=20)
SIRF52=irf(SVAR5,response="TY",impulse="TY",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1,n.ahead=20)
SIRF53=irf(SVAR5,response="PE",impulse="TY",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1,n.ahead=20)
SIRF54=irf(SVAR5,response="VS",impulse="TY",cumulative=FALSE,boot=TRUE,ci
    =0.95,runs=200,seed=1,n.ahead=20)
plot(SIRF51);plot(SIRF52);plot(SIRF53);plot(SIRF54)
SVAR6=rwz.reject(Y=Y4,nlags=1,draws=200,subdraws=200,nkeep=2000,KMIN=1,KMAX=5,
    constrained=c(+2,+1,+3),steps=20)
irfplot(irfdraws=SVAR6$IRFS,bands=c(0.025,0.975),grid=FALSE,bw=TRUE)
```

Econometric Methods II Assignment 04

Junyong Kim*

April 20, 2017

1 Analytical Exercise

1. The AR(2) model is
$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t.$$
The measurement equation for this model is
$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}.$$
The transition equation for this model is
$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}.$$
Since the AR(2) model does not have an intercept, the initial state vector equals to zero, i.e. $\beta_{0|0} = 0$. The variance of this state vector is
$$\begin{aligned}
\mathbf{y}_t =& \Phi \mathbf{y}_{t-1} + \underline{\varepsilon}_t \\
\operatorname{Var}\left[\mathbf{y}_t\right] =& \Phi \operatorname{Var}\left[\mathbf{y}_{t-1}\right] \Phi^\top + \operatorname{Var}\left[\underline{\varepsilon}_t\right] \\
\mathbf{P}_{0|0} =& \mathbf{F}\mathbf{P}_{0|0}\mathbf{F}^\top + \mathbf{Q} \\
\operatorname{vec}\left(\mathbf{P}_{0|0}\right) =& (\mathbf{I} - \mathbf{F} \otimes \mathbf{F})^{-1} \operatorname{vec}(\mathbf{Q}) \\
(\mathbf{I} - \mathbf{F} \otimes \mathbf{F})^{-1} =& \frac{1}{\left((1-\phi_2)^2 - \phi_1^2\right)(1+\phi_2)} \begin{pmatrix} 1-\phi_2 & \phi_1\phi_2 & \phi_1\phi_2 & \phi_2^2(1-\phi_2) \\ \phi_1 & 1-\phi_1^2-\phi_2 & \phi_2(1-\phi_2) & \phi_1\phi_2^2 \\ \phi_1 & \phi_2(1-\phi_2) & 1-\phi_1^2-\phi_2 & \phi_1\phi_2^2 \\ 1-\phi_2 & \phi_1\phi_2 & \phi_1\phi_2 & 1-\phi_1^2(1+\phi_2)-\phi_2 \end{pmatrix} \\
\operatorname{vec}(\mathbf{Q}) =& \begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 & 0 \end{pmatrix}^\top \\
P_{0|0} =& \frac{(1-\phi_2)\sigma_\varepsilon^2}{\left((1-\phi_2)^2 - \phi_1^2\right)(1+\phi_2)}.
\end{aligned}$$

2. The VEC model is
$$\begin{aligned}
\begin{pmatrix} \Delta y_t \\ \Delta c_t \end{pmatrix} =& \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta c_{t-1} \end{pmatrix} + \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \\
\Delta \mathbf{y}_t =& \Phi \Delta \mathbf{y}_{t-1} + \Theta \mathbf{y}_{t-1} + \varepsilon_t.
\end{aligned}$$

*Lubar School of Business, University of Wisconsin–Milwaukee

Hence,

$$\begin{aligned}\mathbf{y}_t - \mathbf{y}_{t-1} =& \Phi \mathbf{y}_{t-1} - \Phi \mathbf{y}_{t-2} + \Theta \mathbf{y}_{t-1} + \varepsilon_t \\ \mathbf{y}_t =& \left(\mathbf{I} + \Phi + \Theta\right) \mathbf{y}_{t-1} - \Phi \mathbf{y}_{t-2} + \varepsilon_t.\end{aligned}$$

Thus the measurement equation is

$$\mathbf{y}_t = \begin{pmatrix}\mathbf{I} & \mathbf{0}\end{pmatrix} \begin{pmatrix}\mathbf{y}_t \\ \mathbf{y}_{t-1}\end{pmatrix}$$
$$\begin{pmatrix} y_t \\ c_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ c_t \\ y_{t-1} \\ c_{t-1} \end{pmatrix}.$$

And the transition equation is

$$\begin{pmatrix}\mathbf{y}_t \\ \mathbf{y}_{t-1}\end{pmatrix} = \begin{pmatrix}\mathbf{I} + \Phi + \Theta & -\Phi \\ \mathbf{I} & \mathbf{0}\end{pmatrix} \begin{pmatrix}\mathbf{y}_{t-1} \\ \mathbf{y}_{t-2}\end{pmatrix} + \begin{pmatrix}\varepsilon_t \\ 0\end{pmatrix}$$
$$\begin{pmatrix} y_t \\ c_t \\ y_{t-1} \\ c_{t-1} \end{pmatrix} = \begin{pmatrix} 1 + \phi_{11} + \theta_1 & \phi_{12} - \theta_1 & -\phi_{11} & -\phi_{12} \\ \phi_{21} + \theta_2 & 1 + \phi_{22} - \theta_2 & -\phi_{21} & -\phi_{22} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ c_{t-1} \\ y_{t-2} \\ c_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ 0 \\ 0 \end{pmatrix}.$$

3. The ARMA(2,1) model is

$$\begin{aligned}\Delta y_t =& 0.3 \Delta y_{t-1} + 0.5 \Delta y_{t-2} + \varepsilon_t + 0.2 \varepsilon_{t-1} \\ =& \left(1 - 0.3L - 0.5L^2\right)^{-1} (1 + 0.2L) \varepsilon_t \\ =& \phi(L)^{-1} \theta(L) \varepsilon_t \\ =& \psi(L) \varepsilon_t.\end{aligned}$$

Then, $\psi(1) = \frac{\theta(1)}{\phi(1)} = \frac{1+0.2}{1-0.3-0.5} = 6$. Thus, the Beveridge–Nelson stochastic trend BN_t of y_t is

$$BN_t = 6 \sum_{s=1}^{t} \varepsilon_s.$$

4. This model is an unobserved component-ARMA($p = 1$,$q = 0$), so the corresponding reduced form is ARMA($p = 1$,$q^* = 1$) since $q^* = \max(p, q + 1) = 1$ [Morley et al., 2003]. This implies that one can only obtain two autocovariances (i.e. γ_0 and γ_1) from the MA side of the reduced form ARMA and this is not enough to fully identify two variances and covariance (i.e. σ_v^2, σ_e^2 and σ_{ve}) for the MA side of the UC-ARMA. In order for UC-ARMAs to be identified, the condition $p \geq q + 2$ is necessary. Hence this model cannot be identified.

2 Empirical Exercise

1. $\{LRGDP_t\}_{t=1947:1}^{2015:2}$ is the series of a log real GDP. The (1,1) panel of Figure 1 is the plot of $LRGDP$; hereafter I will use (i,j) notations in order to refer each panel.

   (a) The (1,2) panel of Figure 1 is the plot of $LRGDP^{\mathrm{LT}}$; I regress $LRGDP_t$ on $\begin{pmatrix}1 & t\end{pmatrix}^\top$ to eliminate the linear trend and take the residual as $LRGDP^{\mathrm{LT}}$. The estimated result is

$$LRGDP_t = 7.651220 + 0.008045t + \hat{\varepsilon}_t, \quad LRGDP_t^{\mathrm{LT}} \equiv \hat{\varepsilon}_t.$$

The (2,1) panel of Figure 1 is the plot of $LRGDP^{\mathrm{HP}}$, which is the cycle component obtained by Hodrick–Prescott filter.

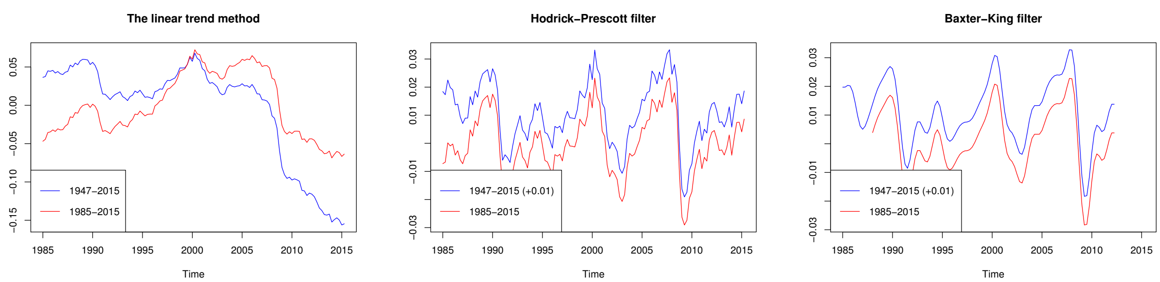
The (2,2) panel of Figure 1 is the plot of $LRGDP^{\mathrm{BK}}$, which is the cycle component obtained by Baxter–King filter.

(b) The panels of Figure 2 are obtained from the post-1985 sample. (1,1), (1,2), (2,1) and (2,2) panels of Figure 2 are the plots of $LRGDP$, $LRGDP^{\mathrm{LT}}$, $LRGDP^{\mathrm{HP}}$ and $LRGDP^{\mathrm{BK}}$, respectively. The estimated result for the linear trend method is

$$LRGDP_t = 8.958888 + 0.006605t + \hat{\varepsilon}_t, \quad LRGDP_t^{\mathrm{LT}} \equiv \hat{\varepsilon}_t.$$

For each case, I overlap the outcomes from different samples and attach the plots below. (1,1), (1,2) and (1,3) panels are the results from the linear trend method, Hodrick–Prescott filter and Baxter–King filter, respectively. While the first panel (the linear trend method) shows the obvious difference between the outcomes, two other panels (H–P and B–K filters) exhibit the results similar to each other.

For (1,2) and (1,3) panels, I intentionally add 0.01 into the first series so as to compare two results clearly.



(c) The state space form of AR(4) is

$$\begin{aligned}
\begin{pmatrix} \Delta y_t \\ \Delta y_{t-1} \\ \Delta y_{t-2} \\ \Delta y_{t-3} \end{pmatrix} = \begin{pmatrix} \mu \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \phi_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta y_{t-2} \\ \Delta y_{t-3} \\ \Delta y_{t-4} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
\Delta \mathbf{y}_t = & \underline{\mu} + \Phi \Delta \mathbf{y}_{t-1} + \underline{\varepsilon}_t,
\end{aligned}$$

and hence

$$\begin{aligned}
\mathbf{y}_{t+h|t} =& \mathbf{y}_t + h\underline{\mu} + \sum_{j=1}^{h} \Phi^j \left(\Delta \mathbf{y}_t - \underline{\mu} \right) \\
\mathbf{e1} \equiv& \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}^\top \\
y_{t+h|t} =& \mathbf{e1}^\top \mathbf{y}_{t+h|t} \\
=& y_t + h\mu + \mathbf{e1}^\top \sum_{j=1}^{h} \Phi^j \left(\Delta \mathbf{y}_t - \underline{\mu} \right) \\
\lim_{h\to\infty} y_{t+h|t} - h\mu =& y_t + \mathbf{e1}^\top \left(\mathbf{I}_4 - \Phi \right)^{-1} \Phi \left(\Delta \mathbf{y}_t - \underline{\mu} \right) \\
=& TD_t + BN_t \\
C_t =& y_t - TD_t - BN_t \\
=& - \mathbf{e1}^\top \left(\mathbf{I}_4 - \Phi \right)^{-1} \Phi \left(\Delta \mathbf{y}_t - \underline{\mu} \right).
\end{aligned}$$

In order to obtain the cycle, I estimate Φ and $\underline{\mu}$ using $\Delta LRGDP$ and the `Arima` function of `forecast` package. The estimated result is

$$\begin{aligned}\Delta y_t - 0.0078 =& 0.3390\left(\Delta y_{t-1} - 0.0078\right) + 0.1345\left(\Delta y_{t-2} - 0.0078\right)\\ &- 0.0715\left(\Delta y_{t-3} - 0.0078\right) - 0.0703\left(\Delta y_{t-4} - 0.0078\right) + \hat{\varepsilon}_t,\end{aligned}$$

and therefore,

$$\begin{aligned}\mathbf{e1}^\top\left(\mathbf{I}_4 - \hat{\Phi}\right)^{-1}\hat{\Phi} &= \begin{pmatrix}0.4963 & -0.0109 & -0.2121 & -0.1051\end{pmatrix}\\ \hat{C}_t &= -0.4963\left(\Delta y_{t-1} - 0.0078\right) + 0.0109\left(\Delta y_{t-2} - 0.0078\right)\\ &\quad + 0.2121\left(\Delta y_{t-3} - 0.0078\right) + 0.1051\left(\Delta y_{t-4} - 0.0078\right).\end{aligned}$$

The (1,1) and (1,2) panels of Figure 3 are the plots of $\Delta LRGDP$ and $\hat{C}$, respectively.

(d) Figure 4 is the plot of $LRGDP^{\text{HP}}$ versus $LRGDP^{\text{BN}}$ (the cycle component obtained from B–N decomposition). Compared to H–P cycle, B–N cycle is less volatile, less cyclical and less persistent. According to B–N decomposition, the periodicity is not that important in explaining the real output. On the other hand, H–P decomposition emphasizes the role of that periodicity.

2. The model is

$$\begin{aligned}\Delta M_t =& \beta_{0t} + \beta_{1t}\Delta i_{t-1} + \beta_{2t}INF_{t-1} + \beta_{3t}SURP_{t-1} + \beta_{4t}\Delta M_{t-1} + e_t\\ \beta_{it} =& \beta_{it-1} + \nu_{it}, \quad i \in \{0, \cdots, 4\},\end{aligned}$$

and hence the corresponding state space form is

$$\Delta M_t = \begin{pmatrix}1 & \Delta i_{t-1} & INF_{t-1} & SURP_{t-1} & \Delta M_{t-1}\end{pmatrix}\begin{pmatrix}\beta_{0t}\\ \beta_{1t}\\ \beta_{2t}\\ \beta_{3t}\\ \beta_{4t}\end{pmatrix} + e_t, \quad e_t \sim \mathrm{N}\left(0, \sigma_e^2\right)$$

$$\begin{pmatrix}\beta_{0t}\\ \beta_{1t}\\ \beta_{2t}\\ \beta_{3t}\\ \beta_{4t}\end{pmatrix} = \begin{pmatrix}\beta_{0t-1}\\ \beta_{1t-1}\\ \beta_{2t-1}\\ \beta_{3t-1}\\ \beta_{4t-1}\end{pmatrix} + \begin{pmatrix}\nu_{0t}\\ \nu_{1t}\\ \nu_{2t}\\ \nu_{3t}\\ \nu_{4t}\end{pmatrix}, \quad \begin{pmatrix}\nu_{0t}\\ \nu_{1t}\\ \nu_{2t}\\ \nu_{3t}\\ \nu_{4t}\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}0\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}\sigma_0^2 & 0 & 0 & 0 & 0\\ 0 & \sigma_1^2 & 0 & 0 & 0\\ 0 & 0 & \sigma_2^2 & 0 & 0\\ 0 & 0 & 0 & \sigma_3^2 & 0\\ 0 & 0 & 0 & 0 & \sigma_4^2\end{pmatrix}\right),$$

or

$$y_t = \mathbf{x}_t^\top \underline{\beta}_t + e_t, \quad e_t \sim \mathrm{N}\left(0, \sigma_e^2\right) \tag{1}$$

$$\underline{\beta}_t = \underline{\beta}_{t-1} + \underline{\nu}_t, \quad \underline{\nu}_t \sim \mathrm{N}(\mathbf{0}, \Sigma). \tag{2}$$

(a) Table 1 includes the estimated standard deviations and the corresponding standard errors. The panels of Figure 5 are the plots of the filtered (forward-iterating) estimates of time-varying coefficients. In particular, (1,1), (1,2), (2,1), (2,2) and (3,1) panels of Figure 5 are the plots of $\hat{\beta}_{0t|t}$, $\hat{\beta}_{1t|t}$, $\hat{\beta}_{2t|t}$, $\hat{\beta}_{3t|t}$ and $\hat{\beta}_{4t|t}$, respectively.

(b) The panels of Figure 6 are the plots of the smoothed (backward-iterating) estimates of time-varying coefficients. In particular, (1,1), (1,2), (2,1), (2,2) and (3,1) panels of Figure 5 are the plots of $\hat{\beta}_{0t|T}$, $\hat{\beta}_{1t|T}$, $\hat{\beta}_{2t|T}$, $\hat{\beta}_{3t|T}$ and $\hat{\beta}_{4t|T}$, respectively.

(c) While filtered estimates of state variables are computed using past and current observations, smoothed estimates are computed computed using all available observations.

For instance, $\hat{\beta}_{0t|t}$ relies on the information set $\Omega_t \supset \Omega_{t-1} \supset \cdots \supset \Omega_1$. In contrast, $\hat{\beta}_{0t|T}$ relies on the information set $\Omega_T \supset \Omega_{T-1} \supset \cdots \supset \Omega_t \supset \cdots \supset \Omega_1$.

3. The model is

$$\begin{pmatrix} y_t \\ u_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & 0 & 1 \end{pmatrix} \begin{pmatrix} n_t \\ x_t \\ x_{t-1} \\ x_{t-2} \\ g_t \\ l_t \end{pmatrix} + \begin{pmatrix} 0 \\ e_{ct} \end{pmatrix}, \quad e_{ct} \sim \mathrm{N}\left(0, \sigma^2_{e_c}\right)$$

$$\begin{pmatrix} n_t \\ x_t \\ x_{t-1} \\ x_{t-2} \\ g_t \\ l_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & \phi_1 & \phi_2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} n_{t-1} \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \\ g_{t-1} \\ l_{t-1} \end{pmatrix} + \begin{pmatrix} v_t \\ e_t \\ 0 \\ 0 \\ w_t \\ v_{lt} \end{pmatrix}, \quad \begin{pmatrix} v_t \\ e_t \\ w_t \\ v_{lt} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_v & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_w & 0 \\ 0 & 0 & 0 & \sigma^2_{v_l} \end{pmatrix}\right),$$

or

$$\mathbf{y}_t = \mathbf{F}\underline{\theta}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{V}) \tag{3}$$

$$\underline{\theta}_t = \mathbf{G}\underline{\theta}_{t-1} + \mathbf{w}_t, \quad \mathrm{w}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{W}). \tag{4}$$

I follow [Kim and Nelson, 1999] for other details. Table 2 includes the estimated parameters and the corresponding standard errors; standard errors are not stable due to the extreme values in the hessian matrix.

Figure 7 includes both trend and cycle components of y and u. (1,1) panel of Figure 7 is the plot of y and its trend component n. (1,2) panel of Figure 7 is the plot of y's cycle component x; shadings indicate NBER recessions. (2,1) panel of Figure 7 is the plot of u and its trend component l. (2,2) panel of Figure 7 is the plot of u's cycle component c.

According to (1,2) panel of Figure 7, NBER recessions by and large coincide with drops in x. Similarly, according to (2,2) panel, the recessions by and large coincide with jumps in c.

References

[Kim and Nelson, 1999] Kim, C.-J. and Nelson, C. R. (1999). *State-space models with regime-switching: classical and Gibbs-sampling approaches with applications*. MIT Press.

[Morley et al., 2003] Morley, J. C., Nelson, C. R., and Zivot, E. (2003). Why are the beveridge-nelson and unobserved-components decompositions of gdp so different? *Review of Economics and Statistics*, 85(2):235–243.

Parameter	Estimate	St. err.
σ_e	0.4241	(0.0626)
σ_0	0.1005	(0.0584)
σ_1	0.0195	(0.0395)
σ_2	0.2425	(0.0668)
σ_3	0.0203	(0.2558)
σ_4	0.0207	(0.0357)
Log lik.		-50.8683

Table 1: TVP regression model estimates: the model consists of the measurement equation (1) and the transition equation (2), and the model is estimated using `dlmModReg` and `dlmMLE` functions in `dlm` package. The sample is from 1959:3Q to 1985:4Q. Standard errors are estimated using the delta method.

Parameter	Estimate	St. err.
α_0	-13.7726	NaN
α_1	-2.9132	(0.0011)
α_2	-5.0756	(0.0016)
σ_{e_c}	0.0241	(0.0000)
ϕ_1	1.3099	NaN
ϕ_2	-0.3613	(0.0005)
σ_v	0.0241	(0.0000)
σ_e	0.0027	(0.0000)
σ_w	0.0091	NaN
σ_{v_l}	0.0002	(0.0002)
Log lik.		$1,206.2790$

Table 2: Bivariate unobserved component model: the model consists of the measurement equation (3) and the transition equation (4), and the model is estimated using `dlm` and `dlmMLE` functions in `dlm` package. The sample is from 1948:1Q to 2010:3Q. Standard errors are estimated using the delta method.

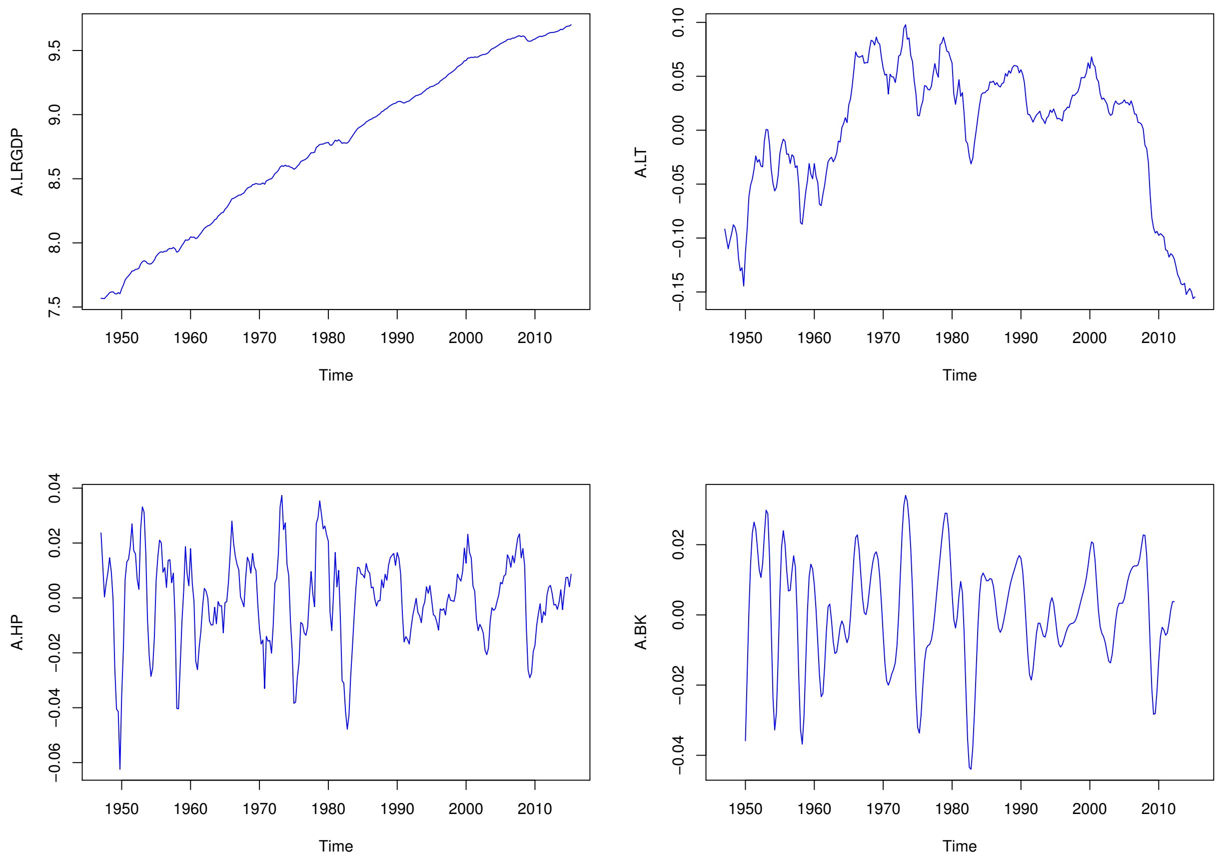


Figure 1: Various cycles: The sample is from 1947:1Q to 2015:2Q. Firstly, (1,1) panel is $LRGDP$. Secondly, (1,2) panel is the cycle component estimated using the linear trend method ($LRGDP^{\mathrm{LT}}$). Thirdly, (2,1) panel is the cycle component estimated using Hodrick–Prescott filter ($LRGDP^{\mathrm{HP}}$). Lastly, (2,2) panel is the cycle component estimated using Baxter–King filter ($LRGDP^{\mathrm{BK}}$).

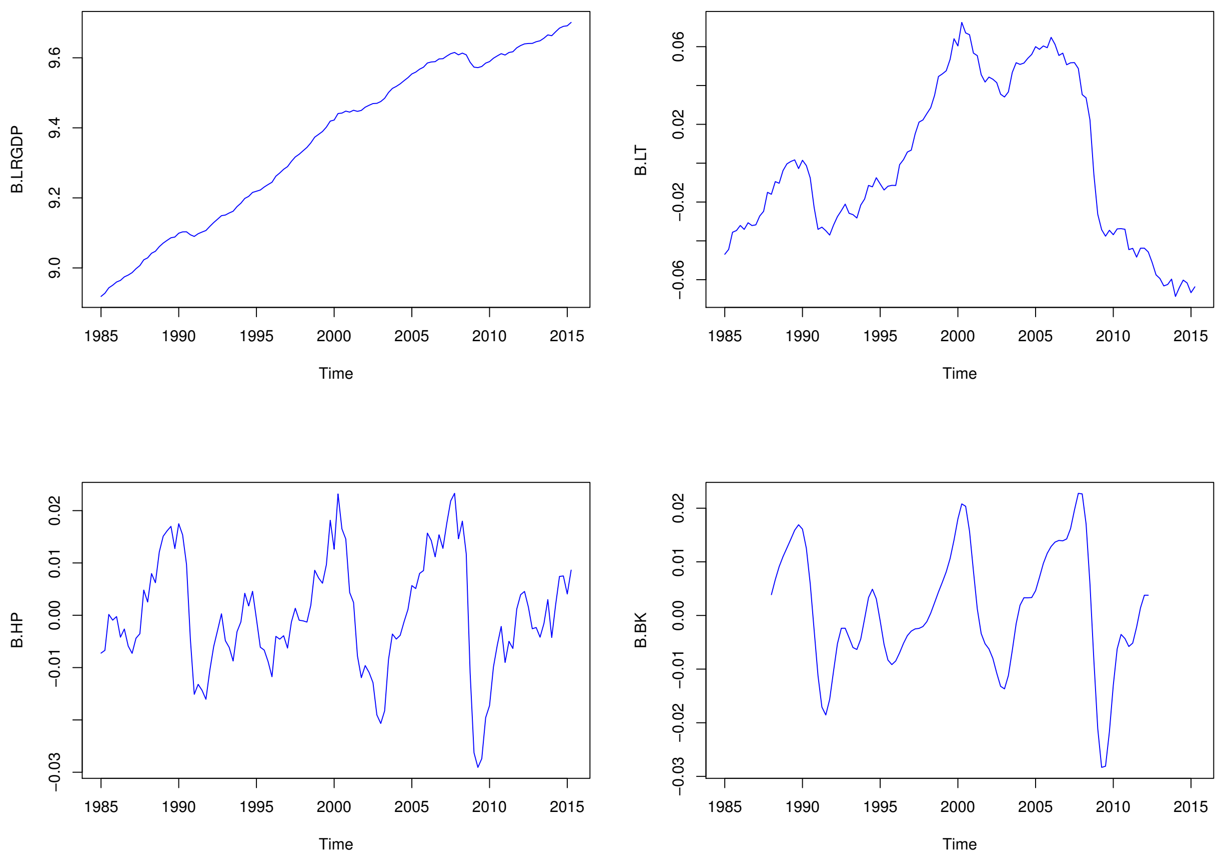


Figure 2: Various cycles: The sample is from 1985:1Q to 2015:2Q. (1,1) panel is $LRGDP$. (1,2), (2,1) and (2,2) panels are the cycles estimated using the linear trend method, Hodrick–Prescott filter and Baxter–King filter, respectively.

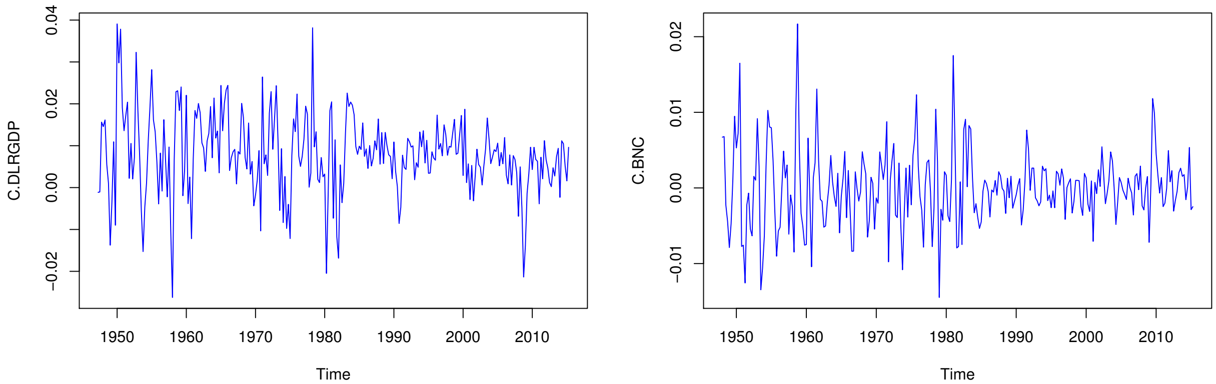


Figure 3: (1,1) panel is the plot of $\Delta LRGDP$. (1,2) panel is the cycle component BN estimated using AR(4) and `Arima` function in `forecast` package.

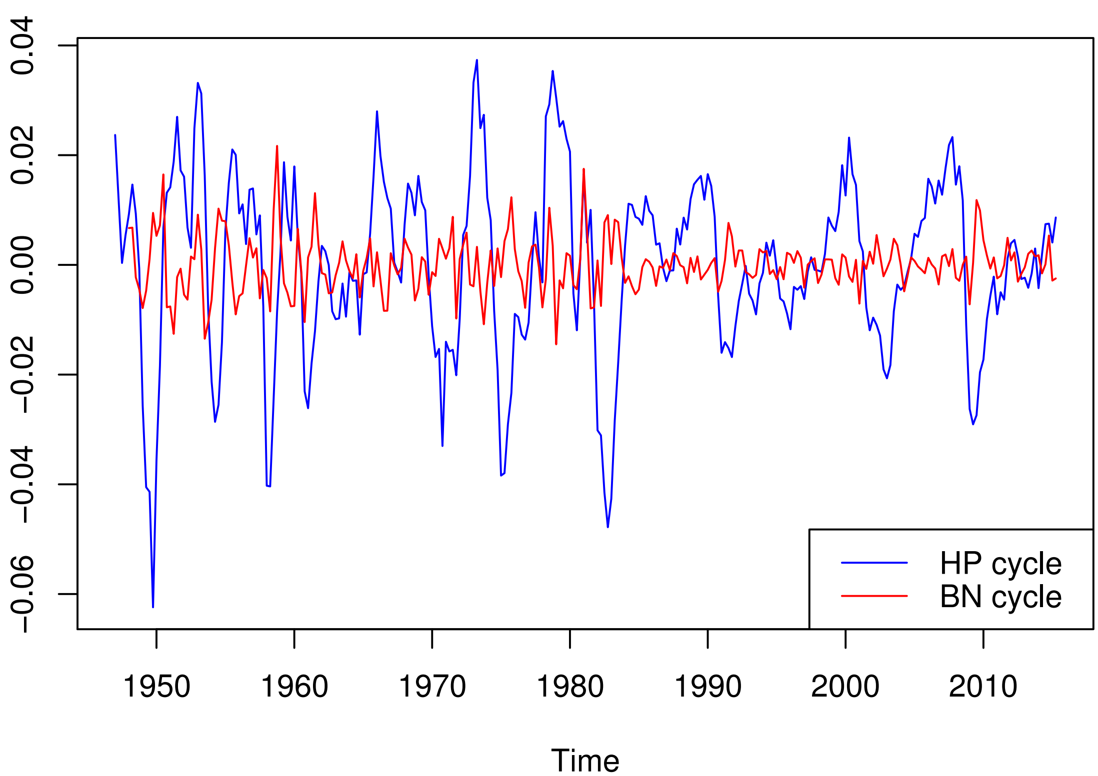


Figure 4: Hodrick–Prescott cycle versus Beveridge–Nelson cycle. The sample is from 1947:1Q to 2015:2Q.

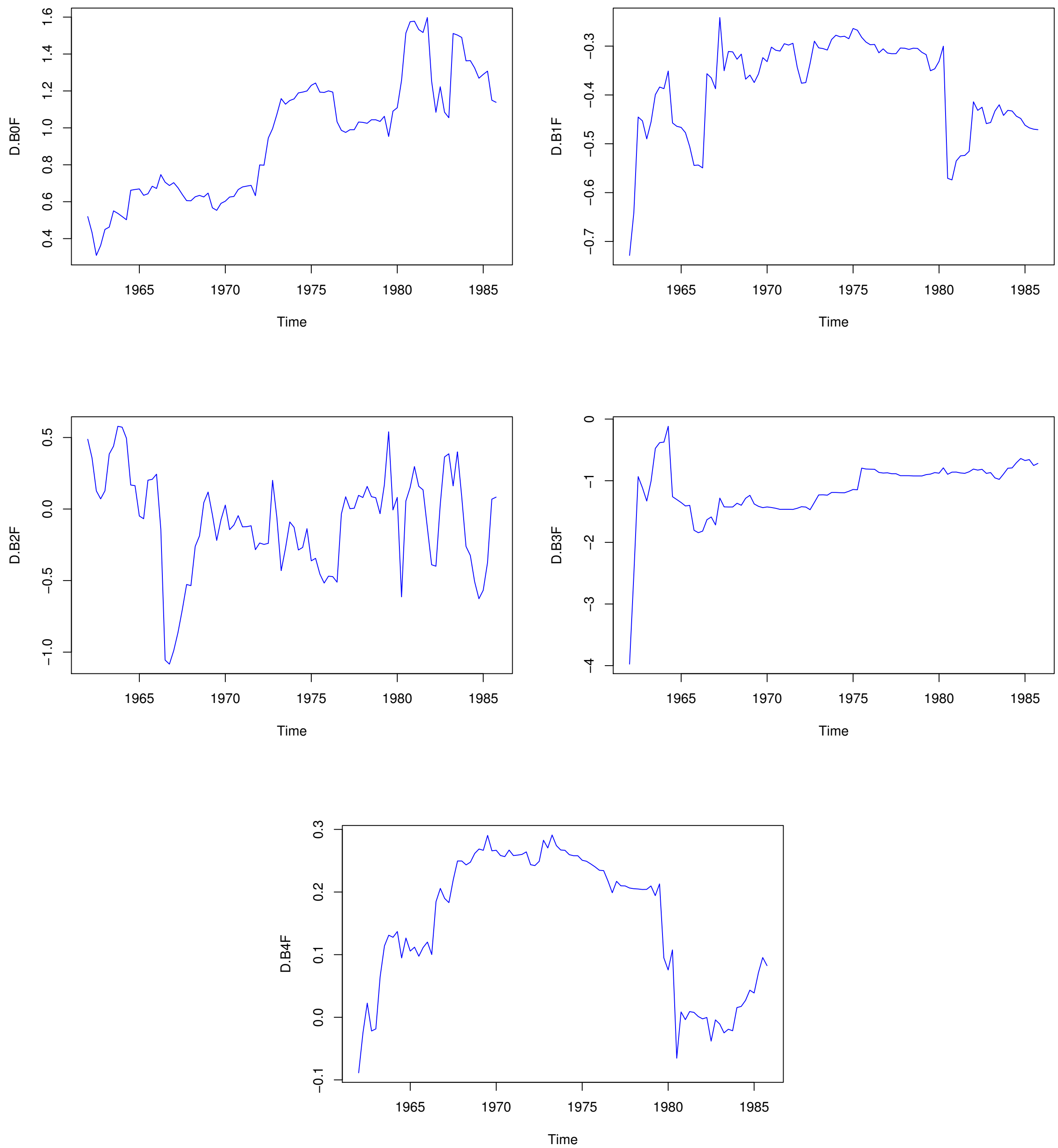


Figure 5: Filtered estimates ($\hat{\underline{\beta}}_{t|t}$) of TVP regression coefficients: the model consists of the measurement equation (1) and the transition equation (2). (1,1), (1,2), (2,1), (2,2) and (3,1) panels are $\hat{\beta}_{0t|t}$, $\hat{\beta}_{1t|t}$, $\hat{\beta}_{2t|t}$, $\hat{\beta}_{3t|t}$ and $\hat{\beta}_{4t|t}$, respectively. The result is from 1962:1Q to 1985:4Q.

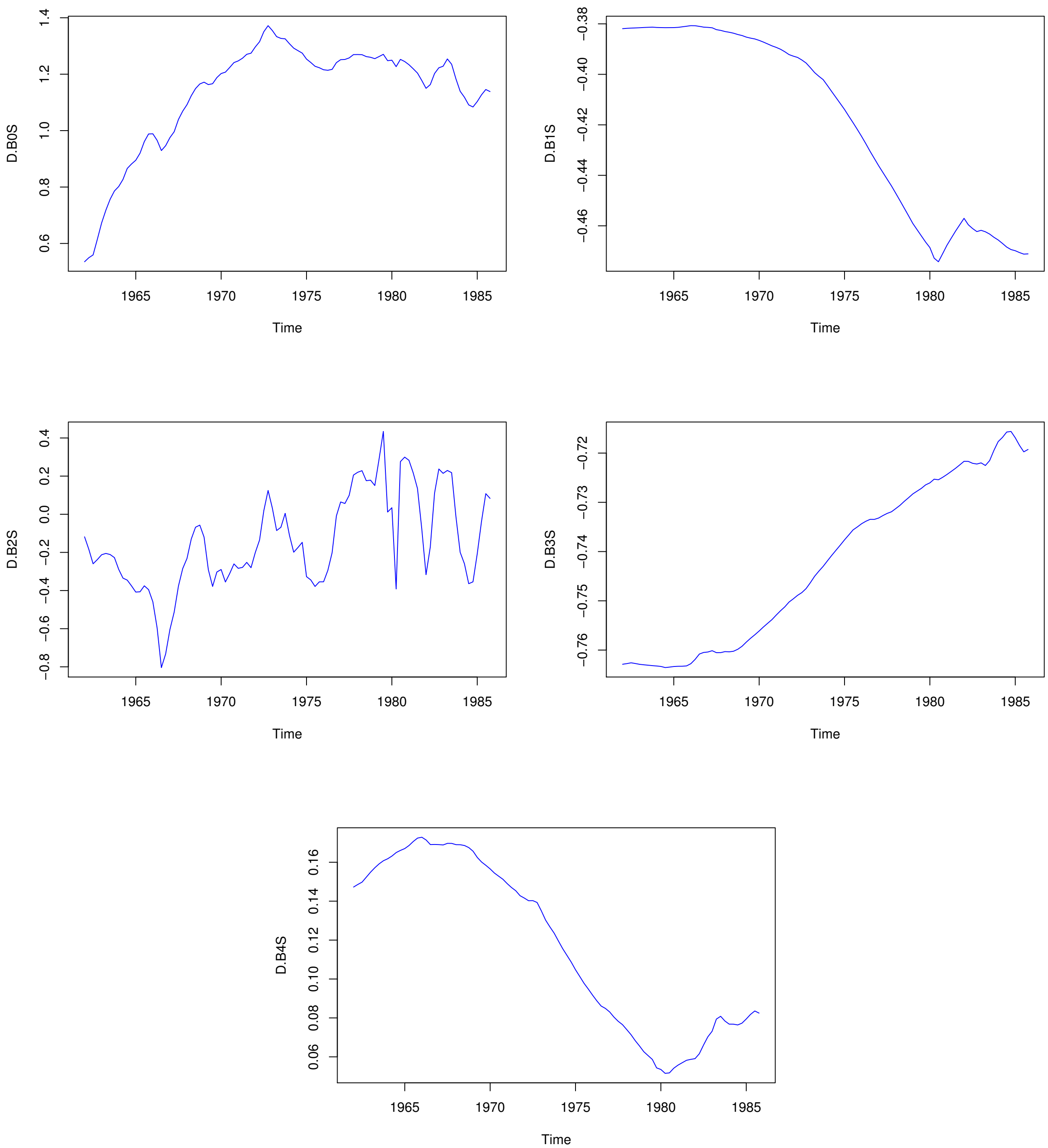


Figure 6: Smoothed estimates ($\hat{\underline{\beta}}_{t|T}$) of TVP regression coefficients: the model consists of the measurement equation (1) and the transition equation (2). (1,1), (1,2), (2,1), (2,2) and (3,1) panels are $\hat{\beta}_{0t|T}$, $\hat{\beta}_{1t|T}$, $\hat{\beta}_{2t|T}$, $\hat{\beta}_{3t|T}$ and $\hat{\beta}_{4t|T}$, respectively. The result is from 1962:1Q to 1985:4Q.

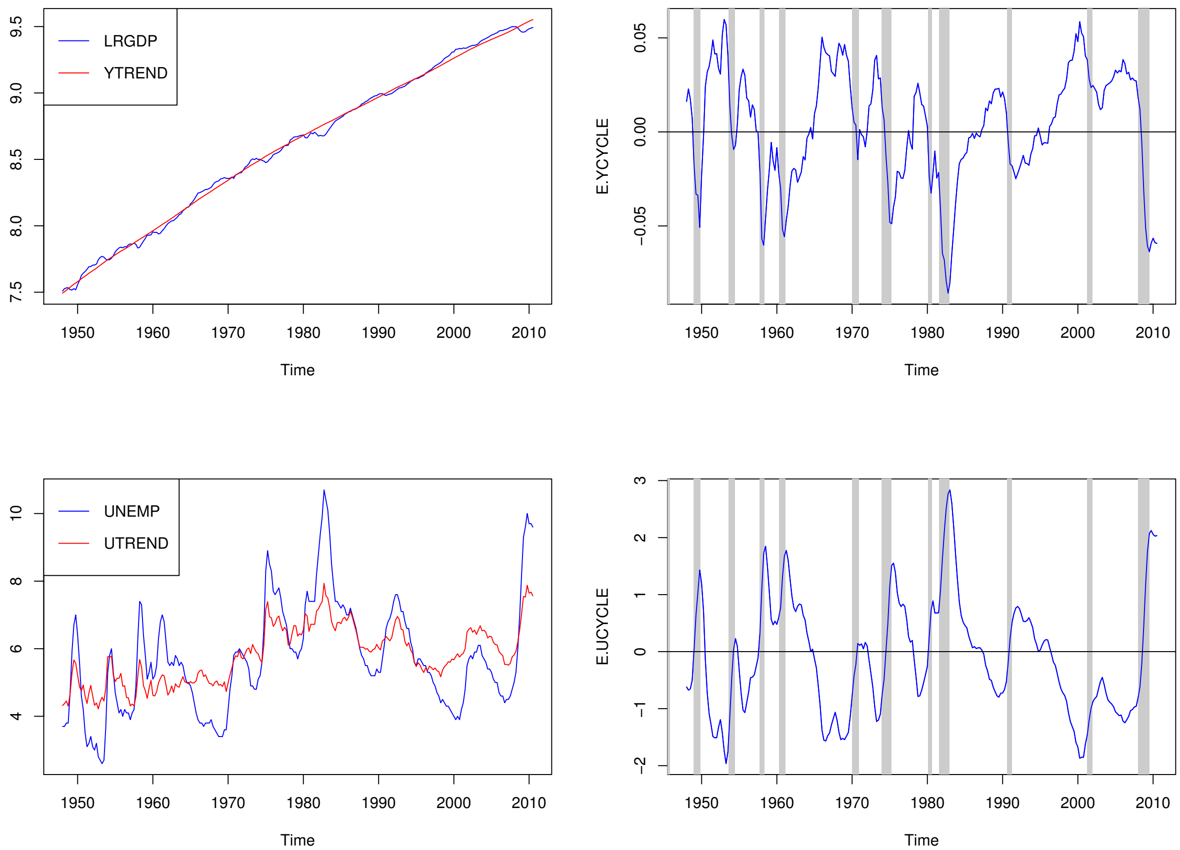


Figure 7: Unobserved component model estimates: the model consists of the measurement equation (3) and the transition equation (4). $y_t = n_t + x_t$ and $u_t = l_t + c_t$. (1,1) panel shows both y_t and n_t ($LRGDP$ trend). (1,2) panel shows x_t ($LRGDP$ cycle) and NBER recessions. Similarly, (2,1) panel shows both u_t and l_t ($UNEMP$ trend). (1,2) panel shows x_t ($UNEMP$ cycle) and NBER recessions. The sample is from 1948:1Q to 2010:3Q.

Appendix: R code

```
#########EMPIRICAL EXERCISE 1##########
library(mFilter)#install.packages("mFilter")
library(forecast)#install.packages("forecast")
library(dlm)#install.packages("dlm")
library(tis)#install.packages("tis")
setwd("C:/Users/Junyong/OneDrive - University of Wisconsin Milwaukee/UWM/2017_01_Spring/Econometric Methods II")
RAW1=read.csv("170417_lrgdp_revised.csv")
#########EMPIRICAL EXERCISE 1.A##########
A.LRGDP=ts(RAW1$lrgdp,start=c(1947,1),frequency=4);plot(A.LRGDP,col="BLUE")
A.LT=ts(lm(A.LRGDP~c(1:274))$resid,start=c(1947,1),frequency=4);plot(A.LT,col="BLUE")
A.HP=ts(hpfilter(A.LRGDP)$cycle,start=c(1947,1),frequency=4);plot(A.HP,col="BLUE")
A.BK=ts(bkfilter(A.LRGDP)$cycle,start=c(1947,1),frequency=4);plot(A.BK,col="BLUE",ylab="A.BK")
#########EMPIRICAL EXERCISE 1.B##########
B.LRGDP=ts(A.LRGDP[153:274],start=c(1985,1),frequency=4);plot(B.LRGDP,col="BLUE")
B.LT=ts(lm(B.LRGDP~c(1:122))$resid,start=c(1985,1),frequency=4);plot(B.LT,col="BLUE")
B.HP=ts(hpfilter(B.LRGDP)$cycle,start=c(1985,1),frequency=4);plot(B.HP,col="BLUE")
B.BK=ts(bkfilter(B.LRGDP)$cycle,start=c(1985,1),frequency=4);plot(B.BK,col="BLUE",ylab="B.BK")
#########COMPARISON##########
ts.plot(ts(A.LT[153:274],start=c(1985,1),frequency=4),B.LT,gpars=list(col=c("BLUE","RED")))
legend("bottomleft",legend=c("1947-2015","1985-2015"),col=c("BLUE","RED"),lty=1)
title("The linear trend method")
ts.plot(ts(A.HP[153:274]+0.01,start=c(1985,1),frequency=4),B.HP,gpars=list(col=c("BLUE","RED")))
legend("bottomleft",legend=c("1947-2015 (+0.01)","1985-2015"),col=c("BLUE","RED"),lty=1)
title("Hodrick-Prescott filter")
ts.plot(ts(A.BK[153:274]+0.01,start=c(1985,1),frequency=4),B.BK,gpars=list(col=c("BLUE","RED")))
legend("bottomleft",legend=c("1947-2015 (+0.01)","1985-2015"),col=c("BLUE","RED"),lty=1)
title("Baxter-King filter")
#########EMPIRICAL EXERCISE 1.C##########
C.DLRGDP=ts(diff(A.LRGDP),start=c(1947,2),frequency=4);plot(C.DLRGDP,col="BLUE")
C.PHI=matrix(0,nrow=4,ncol=4)
C.PHI[1,]=Arima(C.DLRGDP,order=c(4,0,0),method="ML")$coef[1:4]
C.PHI[2,1]=1;C.PHI[3,2]=1;C.PHI[4,3]=1;C.PHI
C.DMDLRGDP=C.DLRGDP-Arima(C.DLRGDP,order=c(4,0,0),method="ML")$coef[5]
C.BNC=c(1:270)
for(T in 1:270){C.BNC[T]=-c(1,0,0,0)%*%solve(diag(4)-C.PHI)%*%C.PHI%*%C.DMDLRGDP[T:(T+3)]}
C.BNC=ts(C.BNC,start=c(1948,1),frequency=4);plot(C.BNC,col="BLUE");rm(T)
```

```
#########EMPIRICAL EXERCISE 1.D##########
ts.plot(A.HP,C.BNC,gpars=list(col=c("BLUE","RED")))
legend("bottomright",legend=c("HP filter","BN decomposition"),col=c("BLUE","
    RED"),lty=1)
#########EMPIRICAL EXERCISE 2##########
RAW2=read.table("170412_tvp.txt")
D.Y=ts(RAW2$V2,start=c(1953,3),frequency=4)
D.X=ts(RAW2[,3:6],start=c(1953,3),frequency=4)
PARM_REST1=function(parm){return(exp(parm))}
SSM1=function(parm,X.MAT){parm=PARM_REST1(parm);return(dlmModReg(X=X.MAT,dV=
    parm[1],dW=c(parm[2:6])))}
D.FIT=dlmMLE(y=D.Y,parm=matrix(0,1,6),X.MAT=D.X,build=SSM1,hessian=T)
sqrt(PARM_REST1(D.FIT$par));DG=diag(6);diag(DG)=exp(D.FIT$par/2)/2
sqrt(diag(DG%*%solve(D.FIT$hessian)%*%DG))
-dlmLL(D.Y,SSM1(D.FIT$par,D.X))
D.MOD=SSM1(D.FIT$par,D.X);D.MODF=dlmFilter(D.Y,D.MOD);D.MODS=dlmSmooth(D.MODF)
#########EMPIRICAL EXERCISE 2.A##########
D.B0F=ts(D.MODF$m[-1,1][11:106],start=c(1962,1),frequency=4);plot(D.B0F,col="
    BLUE")
D.B1F=ts(D.MODF$m[-1,2][11:106],start=c(1962,1),frequency=4);plot(D.B1F,col="
    BLUE")
D.B2F=ts(D.MODF$m[-1,3][11:106],start=c(1962,1),frequency=4);plot(D.B2F,col="
    BLUE")
D.B3F=ts(D.MODF$m[-1,4][11:106],start=c(1962,1),frequency=4);plot(D.B3F,col="
    BLUE")
D.B4F=ts(D.MODF$m[-1,5][11:106],start=c(1962,1),frequency=4);plot(D.B4F,col="
    BLUE")
#########EMPIRICAL EXERCISE 2.B##########
D.B0S=ts(D.MODS$s[-1,1][11:106],start=c(1962,1),frequency=4);plot(D.B0S,col="
    BLUE")
D.B1S=ts(D.MODS$s[-1,2][11:106],start=c(1962,1),frequency=4);plot(D.B1S,col="
    BLUE")
D.B2S=ts(D.MODS$s[-1,3][11:106],start=c(1962,1),frequency=4);plot(D.B2S,col="
    BLUE")
D.B3S=ts(D.MODS$s[-1,4][11:106],start=c(1962,1),frequency=4);plot(D.B3S,col="
    BLUE")
D.B4S=ts(D.MODS$s[-1,5][11:106],start=c(1962,1),frequency=4);plot(D.B4S,col="
    BLUE")
#########EMPIRICAL EXERCISE 3##########
RAW3=read.table("170417_rgdp_us.txt")
RAW4=read.table("170417_ur_us.txt")
E.LRGDP=ts(log(RAW3$V2[5:255]),start=c(1948,1),frequency=4)
E.UNEMP=ts(RAW4$V2,start=c(1948,1),frequency=4)
E.Y=ts.union(E.LRGDP,E.UNEMP)
PARM_REST2=function(parm){
  parm[c(1:5)]=exp(parm[c(1:5)])
  return(parm)
}
SSM2=function(parm){
  parm=PARM_REST2(parm)
  F.MAT=matrix(0,2,6)
  F.MAT[1,1]=F.MAT[1,2]=F.MAT[2,6]=1
  F.MAT[2,2:4]=parm[8:10]
  V.MAT=matrix(0,2,2)
```

```
  V.MAT[2,2]=parm[1]
  G.MAT=matrix(0,6,6)
  G.MAT[1,1]=G.MAT[1,5]=G.MAT[3,2]=G.MAT[4,3]=G.MAT[5,5]=G.MAT[6,6]=1
  G.MAT[2,2:3]=parm[6:7]
  W.MAT=matrix(0,6,6)
  diag(W.MAT)[c(1,2,5,6)]=parm[2:5]
  m0.MAT=matrix(0,6,1)
  C0.MAT=diag(6)*100
  return(dlm(FF=F.MAT,V=V.MAT,GG=G.MAT,W=W.MAT,m0=m0.MAT,C0=C0.MAT))
}
E.FIT=dlmMLE(y=E.Y,parm=matrix(0,1,10),build=SSM2,hessian=T,method="SANN",
   control=list(maxit=5000))
E.COEF=PARM_REST2(E.FIT$par);E.COEF[1:5]=sqrt(E.COEF[1:5])
DG=diag(10);diag(DG)[1:5]=exp(E.FIT$par[1:5]/2)/2
E.COEF;sqrt(diag(DG%*%solve(E.FIT$hessian)%*%DG))
-dlmLL(E.Y,SSM2(E.FIT$par))
E.MOD=SSM2(E.FIT$par)
E.MODF=dlmFilter(E.Y,E.MOD)
E.MODS=dlmSmooth(E.MODF)
##########LOG REAL OUTPUT##########
E.YTREND=ts(E.MODS$s[-1,1],start=c(1948,1),frequency=4);plot(E.YTREND,col="
   BLUE")
ts.plot(E.LRGDP,E.YTREND,gpars=list(col=c("BLUE","RED")))
legend("topleft",legend=c("LRGDP","YTREND"),col=c("BLUE","RED"),lty=1)
E.YCYCLE=ts(E.MODS$s[-1,2],start=c(1948,1),frequency=4);plot(E.YCYCLE,col="
   BLUE")
nberShade();lines(E.YCYCLE,col="BLUE");abline(h=0)
##########UNEMPLOYMENT RATE##########
E.UTREND=ts(E.MODS$s[-1,6],start=c(1948,1),frequency=4);plot(E.UTREND,col="
   BLUE")
ts.plot(E.UNEMP,E.UTREND,gpars=list(col=c("BLUE","RED")))
legend("topleft",legend=c("UNEMP","UTREND"),col=c("BLUE","RED"),lty=1)
E.UCYCLE=ts(E.UNEMP-E.MODS$s[-1,6],start=c(1948,1),frequency=4);plot(E.UCYCLE,
   col="BLUE")
nberShade();lines(E.UCYCLE,col="BLUE");abline(h=0)
```

Econometric Methods II Assignment 05

Junyong Kim*

May 1, 2017

1 Analytical Exercise

1. The VECM is

$$\begin{aligned}\Delta \mathbf{y}_t =& \mathbf{c} + \boldsymbol{\alpha}\boldsymbol{\beta}^\top \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{iidN}(\mathbf{0}, \boldsymbol{\Sigma}) \\ \boldsymbol{\alpha} =& \begin{pmatrix}\alpha_1 & 0\end{pmatrix}^\top \\ \boldsymbol{\beta} =& \begin{pmatrix}1 & -\beta_2\end{pmatrix}^\top .\end{aligned}$$

   (a) Define $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^\top$, then

$$\Delta \mathbf{y}_t = \mathbf{c} + \mathbf{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t = \mathbf{c} + \begin{pmatrix}\alpha_1 & -\alpha_1\beta_2 \\ 0 & 0\end{pmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,$$

   and

$$\begin{aligned}\mathbf{y}_t - \mathbf{y}_{t-1} =& \mathbf{c} + \mathbf{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \\ \mathbf{y}_t =& \mathbf{c} + (\mathbf{I} + \mathbf{\Pi})\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t = \mathbf{c} + \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,\end{aligned}$$

   where $\mathbf{A} = \mathbf{I} + \mathbf{\Pi}$.

   (b) By premultiplying $\boldsymbol{\beta}^\top$,

$$\begin{aligned}\boldsymbol{\beta}^\top \mathbf{y}_t =& \boldsymbol{\beta}^\top \mathbf{c} + \boldsymbol{\beta}^\top \left(\mathbf{I} + \boldsymbol{\alpha}\boldsymbol{\beta}^\top\right) \mathbf{y}_{t-1} + \boldsymbol{\beta}^\top \boldsymbol{\varepsilon}_t \\ =& \boldsymbol{\beta}^\top \mathbf{c} + \left(1 + \boldsymbol{\beta}^\top \boldsymbol{\alpha}\right) \boldsymbol{\beta}^\top \mathbf{y}_{t-1} + \boldsymbol{\beta}^\top \boldsymbol{\varepsilon}_t,\end{aligned}$$

   and this process is stable when $-1 < 1 + \boldsymbol{\beta}^\top \boldsymbol{\alpha} = 1 + \alpha_1 < 1$, hence $-2 < \alpha_1 < 0$.

2 Empirical Exercise

1. (GARCH modeling)

   (a) Figure 1 is the plot of $RSP500$ from February 1960 to September 2014.

   INSERT FIGURE 1 ABOUT HERE

   (b) Figure 2 displays both ACF and PACF of $RSP500$. There is no significant coefficient in both ACF and PACF. So one can consider the random walk process with time-varying variance.

   INSERT FIGURE 2 ABOUT HERE

*Lubar School of Business, University of Wisconsin–Milwaukee

(c) I estimate the GARCH model

$$\begin{aligned} RSP500_t &= c + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}\left(0, \sigma_t^2\right) \\ \sigma_t^2 &= a_0 + a_1 \varepsilon_{t-1}^2 + b_1 \sigma_{t-1}^2, \end{aligned}$$

and the estimates are reported in Table 1.

INSERT TABLE 1 ABOUT HERE

The estimated unconditional variance is $\frac{\hat{a}_0}{1-\hat{a}_1-\hat{b}_1} = 0.002340$. Figure 3 exhibits the conditional variances.

INSERT FIGURE 3 ABOUT HERE

(d) I conduct three sign bias tests; sign bias test, positive sign bias test and negative sign bias test.

$$\begin{aligned} \hat{\varepsilon}_t^2 &= \beta_0 + \beta_1 1\left\{\hat{\varepsilon}_{t-1} < 0\right\} + \xi_t \\ &= \beta_0 + \beta_1 1\left\{\hat{\varepsilon}_{t-1} < 0\right\} \hat{\varepsilon}_{t-1} + \xi_t \\ &= \beta_0 + \beta_1 1\left\{\hat{\varepsilon}_{t-1} \geq 0\right\} \hat{\varepsilon}_{t-1} + \xi_t. \end{aligned}$$

INSERT TABLE 2 ABOUT HERE

Table 2 displays the result of these tests. Since the first test rejects the null, one can conclude that there exists an asymmetric effect.

(e) I estimate the EGARCH model

$$\begin{aligned} RSP500_t &= \mu + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}\left(0, \sigma_t^2\right) \\ \ln \sigma_t^2 &= \omega + \alpha_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \gamma_1 \left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \mathrm{E}\left[\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| \right] \right) + \beta_1 \ln \sigma_{t-1}^2, \end{aligned}$$

and the estimates are reported in Table 3.

INSERT TABLE 3 ABOUT HERE

Since $\hat{\alpha}_1$ is negative and significant, one can conclude that EGARCH better reflects the asymmetric effect.

(f) I estimate the TGARCH model

$$\begin{aligned} RSP500_t &= \mu + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}\left(0, \sigma_t^2\right) \\ \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 \varepsilon_{t-1}^2 I_{t-1} + \beta_1 \sigma_{t-1}^2, \end{aligned}$$

and the estimates are reported in Table 4.

INSERT TABLE 4 ABOUT HERE

Because $\hat{\gamma}_1$ is positive but insignificant, one cannot conclude that TGARCH better reflects the asymmetric effect. With the usual standard error, however, the t-statistic of $\hat{\gamma}_1$ is 2.2727 (the standard error is 0.0552 and the p-value is 0.0230), so one can reject the null.

(g) Figure 4 displays the news impact curve for GARCH, EGARCH and TGARCH models.

INSERT FIGURE 4 ABOUT HERE

2. (Cointegration)

(a) One can test if $SPREAD = GS - TB$ has a unit root. Figure 5 is the plot of $SPREAD$.

INSERT FIGURE 5 ABOUT HERE

I conduct Augmented Dickey–Fuller and Phillips–Perron tests. The null is that there exists a unit root.

INSERT TABLE 5 ABOUT HERE

Table 5 exhibits the result of ADF and PP tests. Since both tests reject the null, one can conclude that $SPREAD$ is stationary.

(b) One can test if $\hat{u} = GS - \hat{\beta}_0 - \hat{\beta}_1 TB$ has a unit root. Figure 6 is the plot of $\hat{u}$.

INSERT FIGURE 6 ABOUT HERE

I conduct ADF and PP tests. The null is that TB is not cointegrated with GS.

INSERT TABLE 6 ABOUT HERE

Table 6 displays the result of these tests. Since both tests reject the null, one can conclude that TB is cointegrated with GS.

(c) One can conduct Stock–Watson dynamic OLS method to address autocorrelation issues. In detail, one can estimate

$$GS_t = \alpha + \beta TB_t + \sum_{i=1}^{p^*} \delta_i \Delta TB_{t+i} + \sum_{j=1}^{q^*} \gamma_j \Delta TB_{t-j} + \varepsilon_t,$$

and use $\hat{\alpha}_{\text{DOLS}}$ and $\hat{\beta}_{\text{DOLS}}$ to obtain $\hat{\varepsilon} = GS - \hat{\alpha}_{\text{DOLS}} - \hat{\beta}_{\text{DOLS}} TB$ since Stock–Watson estimators are more consistent than OLS estimators. I check AIC for each model and choose $(p^*, q^*) = \underset{\{0,\cdots,5\}^2}{\operatorname{argmin}} AIC(p, q)$.

INSERT TABLE 7 ABOUT HERE

Table 7 displays $AIC(p, q)$ and the function is minimized at the combination $p^* = 3$ and $q^* = 5$.

INSERT TABLE 8 ABOUT HERE

Table 8 displays the estimates of Stock–Watson dynamic OLS.

INSERT FIGURE 7 ABOUT HERE

Figure 7 is the plot of $\hat{\varepsilon}$. I conduct ADF and PP tests. The null is that TB is not cointegrated with GS.

INSERT TABLE 9 ABOUT HERE

Table 9 displays the result of these tests. Since both tests reject the null, one can conclude that TB is cointegrated with GS.

(d) No, they are not that different. Figure 8 is the plot that overlaps $SPREAD$, $\hat{u}$ and $\hat{\varepsilon}$.

INSERT FIGURE 8 ABOUT HERE

Because $\hat{u}$ and $\hat{\varepsilon}$ are too close, I plot $\hat{\varepsilon} - 0.1$ instead of $\hat{\varepsilon}$ itself.

(e) I estimate VECMs with $p \in \{0, 1, 2, 3\}$ using $\hat{\varepsilon}$.

INSERT TABLE 10 ABOUT HERE

Table 10 displays the estimates of these VECMs. For the adjustment, α_1 should be negative and α_2 should be positive since $\mathbf{y} = \begin{pmatrix} GS & TB \end{pmatrix}^\top$ and $\hat{\varepsilon} = GS - \hat{\alpha} - \hat{\beta} TB$. For all models with $p \in \{0, 1, 2, 3\}$, α_1s and α_2s are all positive and significant (except α_1 in $p = 3$, which is positive but insignificant). So one can conclude that only TB shows the adjustment.

(f) For all three models, α_1s should be negative but they are positive so incorrect. On the other hand, α_2s should be positive for these models and they are positive so correct.

(g) Table 11 displays the list of $AIC(p)$s with $p \in \{1, \cdots, 10\}$ for VAR models with $\mathbf{y} = \begin{pmatrix} TB & GS \end{pmatrix}^\top$.

INSERT TABLE 11 ABOUT HERE

$AIC(p)$ is minimized at $p^* = 4$.

(h) I conduct both trace and maximum eigenvalue tests using $p^* = 4$.

INSERT TABLE 12 ABOUT HERE

Table 12 displays the results of Johansen trace and maximum eigenvalue tests. In both tests, the null hypothesis that there is no cointegrating relation (or $\text{rank}(\mathbf{\Pi}) = 0$) is rejected. However, the null hypothesis that the number of cointegrating relations is smaller than or equal to one (or $\text{rank}(\mathbf{\Pi}) \leq 0$) is not rejected. Hence one can conclude that there is only one cointegrating vector in between TB and GS.

(i) I estimate

$$\begin{aligned}\Delta\mathbf{y}_t =& \boldsymbol{\alpha}\boldsymbol{\beta}^\top\mathbf{y}_{t-1} + \mathbf{\Phi}_1\Delta\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_3\Delta\mathbf{y}_{t-3} + \boldsymbol{\varepsilon}_t \\ =& \mathbf{\Pi}\mathbf{y}_{t-1} + \mathbf{\Phi}_1\Delta\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_3\Delta\mathbf{y}_{t-3} + \boldsymbol{\varepsilon}_t,\end{aligned}$$

with both $\mathbf{y} = (TB, GS)^\top$ and $\mathbf{y} = (GS, TB)^\top$.

INSERT TABLE 13 ABOUT HERE

Table 13 reports the estimates of these Johansen VECMs. This result is similar to the result in Table 10. All significant coefficients in Table 10 are significant in Table 13 (except ϕ_{12}^1 in $p = 3$, which is significant only at 10% significance level in Table 10).

INSERT TABLE 14 ABOUT HERE

In addition, Table 14 displays the estimates of the cointegrating vectors. The cointegrating vector of Johansen VECM in Table 14 is similar to the cointegrating vector of Stock–Watson VECM in Table 8 ($\hat{\boldsymbol{\beta}}_{\text{Johansen}} = (1.0000, -0.9838, -0.3582)^\top$ in Table 14 and $\hat{\boldsymbol{\beta}}_{\text{Stock–Watson}} = (1.0000, -0.9820, -0.3678)^\top$ in Table 8).

Parameter	Estimate	t-statistic
c	0.006208^{***}	3.98
a_0	0.000085^{*}	1.75
a_1	0.130091^{***}	3.44
b_1	0.833586^{***}	21.86

Table 1: The estimates of GARCH model with $RSP500$

Test	t-statistic
Sign bias	2.4231^{**}
Negative Sign bias	0.7272
Positive Sign bias	1.5046

Table 2: The result of three sign bias tests

Parameter	Estimate	t-statistic
μ	0.005499^{***}	3.00
ω	-0.611885^{**}	-1.98
α_1	-0.119827^{*}	-1.90
β_1	0.903692^{***}	18.97
γ_1	0.224483^{***}	5.33

Table 3: The estimates of EGARCH model with $RSP500$

Parameter	Estimate	t-statistic
μ	0.005284^{***}	3.41
ω	0.000122	0.88
α_1	0.052384	0.63
β_1	0.820770^{***}	11.71
γ_1	0.125415	0.99

Table 4: The estimates of TGARCH model with $RSP500$

Test	Statistic	1% level	5% level	10% level
Augmented Dickey–Fuller	-3.54^{***}	-3.46	-2.88	-2.57
Phillips–Perron	-3.23^{**}	-3.47	-2.88	-2.58

Table 5: The result of Dickey–Fuller and Phillips–Perron tests for $SPREAD$

Test	Statistic	1% level	5% level	10% level
Augmented Dickey–Fuller	-3.57^{***}	-3.46	-2.88	-2.57
Phillips–Perron	-3.25^{**}	-3.47	-2.88	-2.58

Table 6: The result of Dickey–Fuller and Phillips–Perron tests for $\hat{u} = GS - \hat{\beta}_0 - \hat{\beta}_1 TB$

$AIC(p,q)$		p 0	1	2	3	4	5
	0	84.95	51.18	40.91	37.50	35.68	35.34
	1	34.51	17.38	16.65	16.45	12.61	11.30
	2	23.45	14.27	14.25	13.19	9.82	7.70
q	3	20.09	12.58	12.01	12.16	8.70	7.00
	4	20.62	14.17	14.28	14.40	11.04	9.55
	5	20.35	15.65	15.87	16.13	13.21	12.07

Table 7: The list of $AIC(p,q)$ for Stock–Watson dynamic OLS

Parameter	Estimate	t-statistic
α	0.3678^{***}	7.03
β	0.9820^{***}	79.00
δ_1	0.3927^{***}	3.65
δ_2	0.1915^{*}	1.78
δ_3	0.1783^{*}	1.81
γ_1	0.2528^{**}	2.23
γ_2	0.1997^{*}	1.70
γ_3	0.0615	0.56
γ_4	-0.0509	-0.44
γ_5	-0.1246	-1.08

Table 8: The estimates of Stock–Watson dynamic OLS

Test	Statistic	1% level	5% level	10% level
Augmented Dickey–Fuller	-3.57^{***}	-3.46	-2.88	-2.57
Phillips–Perron	-3.25^{**}	-3.47	-2.88	-2.58

Table 9: The result of Dickey–Fuller and Phillips–Perron tests for $\hat{\varepsilon} = GS - \hat{\alpha}_{\mathrm{DOLS}} - \hat{\beta}_{\mathrm{DOLS}} TB$

p	Parameter	Estimate	t-statistic
0	ϕ_1	-0.0064	-0.38
	α_1	0.2782^{***}	5.05
	ϕ_2	-0.0065	-0.45
	α_2	0.3749^{***}	7.79
1	ϕ_1	-0.0044	-0.29
	α_1	0.0893	1.51
	ϕ_{11}^1	0.6492^{***}	5.30
	ϕ_{12}^1	-0.2145	-1.54
	ϕ_2	-0.0053	-0.38
	α_2	0.2438^{***}	4.54
	ϕ_{21}^1	0.5067^{***}	4.56
	ϕ_{22}^1	-0.1708	-1.35
2	ϕ_1	-0.0062	-0.40
	α_1	0.1008	1.56
	ϕ_{11}^1	0.6726^{***}	5.19
	ϕ_{12}^1	-0.2393^{*}	-1.69
	ϕ_{11}^2	-0.1133	-0.83
	ϕ_{12}^2	0.1046	0.80
	ϕ_2	-0.0057	-0.41
	α_2	0.2385^{***}	4.03
	ϕ_{21}^1	0.5002^{***}	4.22
	ϕ_{22}^1	-0.1752	-1.35
	ϕ_{21}^2	0.0042	0.03
	ϕ_{22}^2	0.0179	0.15
3	ϕ_1	-0.0068	-0.44
	α_1	0.0537	0.78
	ϕ_{11}^1	0.6679^{***}	5.10
	ϕ_{12}^1	-0.2428^{*}	-1.71
	ϕ_{11}^2	-0.1278	-0.92
	ϕ_{12}^2	0.0561	0.42
	ϕ_{11}^3	0.1535	1.11
	ϕ_{12}^3	0.0053	0.04
	ϕ_2	-0.0039	-0.28
	α_2	0.1832^{***}	2.97
	ϕ_{21}^1	0.5010^{***}	4.28
	ϕ_{22}^1	-0.1811	-1.42
	ϕ_{21}^2	0.0348	0.28
	ϕ_{22}^2	-0.0879	-0.73
	ϕ_{21}^3	0.0020	0.02
	ϕ_{22}^3	0.2552^{**}	2.17

Table 10: The estimates of Stock–Watson VECMs with $p \in \{0, 1, 2, 3\}$

p	1	2	3	4	5	6	7	8	9	10
$AIC(p)$	-7.43	-7.61	-7.57	-7.66	-7.64	-7.60	-7.60	-7.56	-7.57	-7.55

Table 11: The list of $AIC(p)$s for VAR(p) models with $\mathbf{y} = (TB, GS)^\top$

Test	r	Statistic	1% level	5% level	10% level
Trace	0	20.92^{**}	24.60	19.96	17.85
	1	4.14	12.97	9.24	7.52
Maximum eigenvalue	0	16.78^{**}	20.20	15.67	13.75
	1	4.14	12.97	9.24	7.52

Table 12: The result of Johansen trace and maximum eigenvalue tests for $\mathbf{y} = (TB, GS)^\top$

Vector	Parameter	Estimate	t-statistic
	α_1	-0.1802^{***}	-2.98
	ϕ_{11}^1	-0.0011	-0.01
	ϕ_{12}^1	0.3184^{***}	2.63
	ϕ_{11}^2	-0.0868	-0.72
	ϕ_{12}^2	0.0344	0.28
	ϕ_{11}^3	0.2561^{**}	2.18
	ϕ_{12}^3	0.0020	0.02
$\mathbf{y} = (TB, GS)^\top$	α_2	-0.0529	-0.78
	ϕ_{21}^1	-0.1907	-1.41
	ϕ_{22}^1	0.6158^{***}	4.55
	ϕ_{21}^2	0.0572	0.43
	ϕ_{22}^2	-0.1281	-0.92
	ϕ_{21}^3	0.0058	0.04
	ϕ_{22}^3	0.1540	1.12
	α_1	0.0538	0.78
	ϕ_{11}^1	0.6158^{***}	4.55
	ϕ_{12}^1	-0.1907	-1.41
	ϕ_{11}^2	-0.1281	-0.92
	ϕ_{12}^2	0.0572	0.43
	ϕ_{11}^3	0.1540	1.12
	ϕ_{12}^3	0.0058	0.04
$\mathbf{y} = (GS, TB)^\top$	α_2	0.1832^{***}	2.98
	ϕ_{21}^1	0.3184^{***}	2.63
	ϕ_{22}^1	-0.0011	-0.01
	ϕ_{21}^2	0.0344	0.28
	ϕ_{22}^2	-0.0868	-0.72
	ϕ_{21}^3	0.0020	0.02
	ϕ_{22}^3	0.2561^{**}	2.18

Table 13: The estimates of Johansen VECMs with $p^* = 4$ for $\mathbf{y} = (TB, GS)^\top$ and $\mathbf{y} = (GS, TB)^\top$

Vector	Variable	Estimates
	TB	1.0000
$\mathbf{y} = (TB, GS)^\top$	GS	-1.0165
	1	0.3641
	GS	1.0000
$\mathbf{y} = (GS, TB)^\top$	TB	-0.9838
	1	-0.3582

Table 14: The cointegrating vector for $\mathbf{y} = (TB, GS)^\top$ and $\mathbf{y} = (GS, TB)^\top$

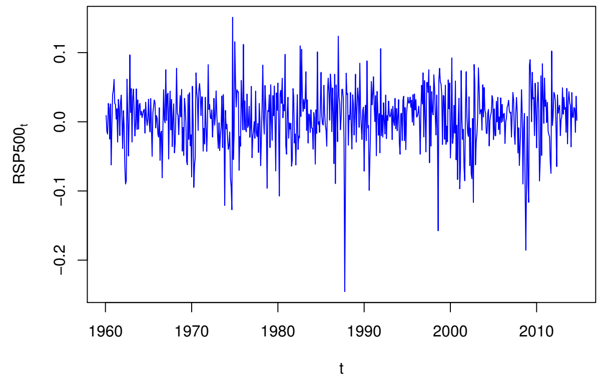


Figure 1: The plot of $RSP500$

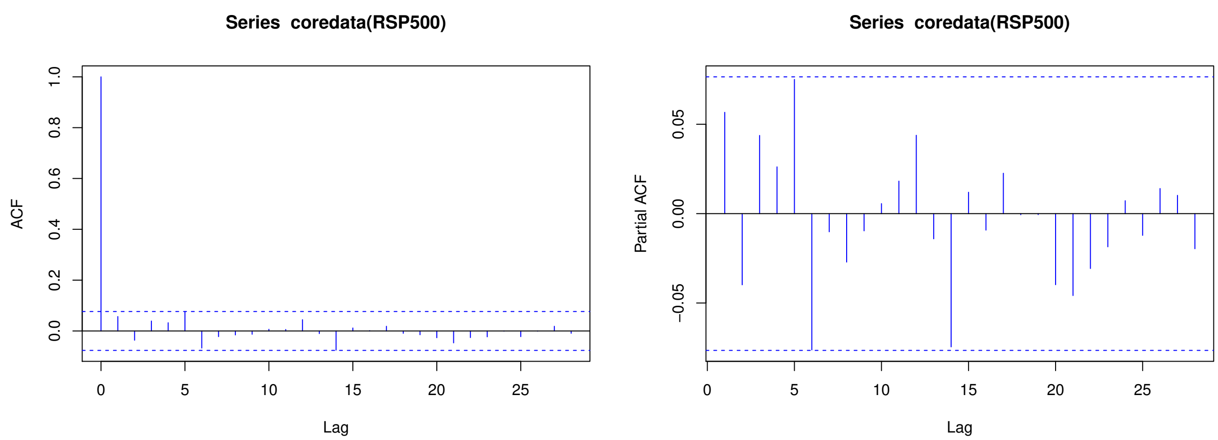


Figure 2: ACF and PACF of $RSP500$

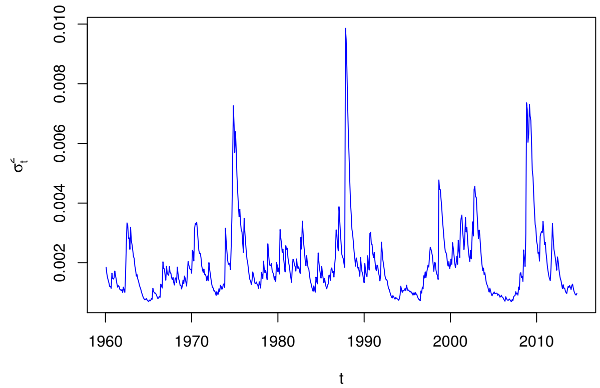


Figure 3: Conditional variances of the GARCH model

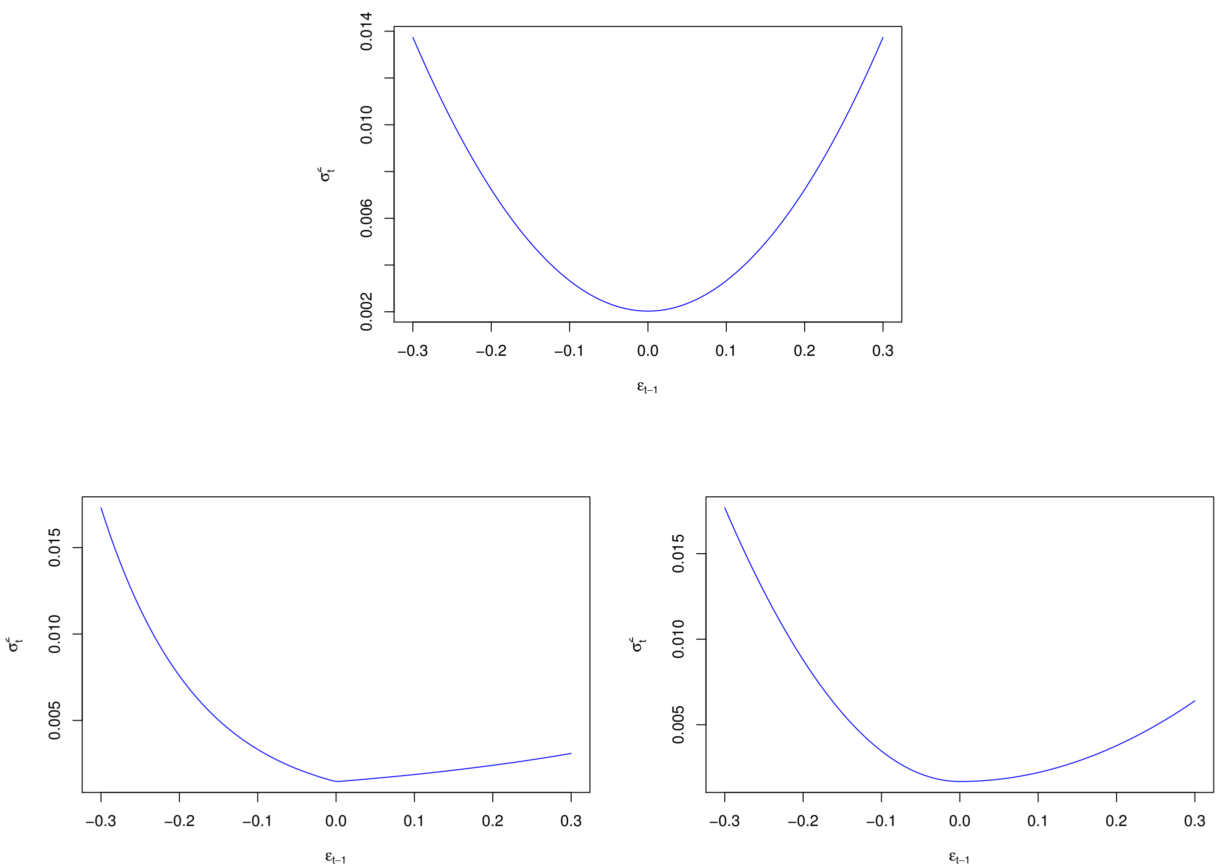


Figure 4: The estimated news impact curves for GARCH (upper), EGARCH (lower-left) and TGARCH (lower-right) models

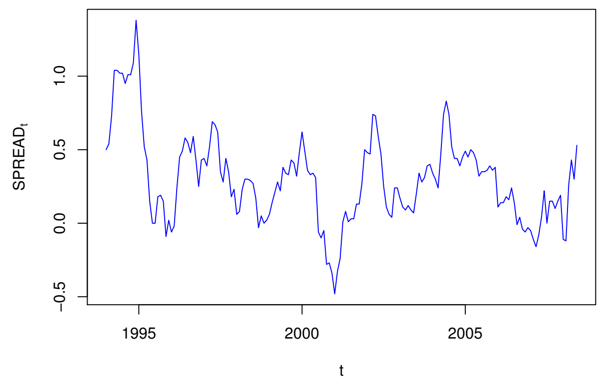


Figure 5: The plot of $SPREAD$

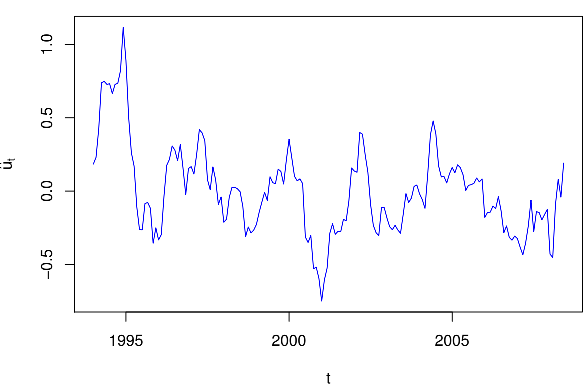


Figure 6: The plot of $\hat{u} = GS - \hat{\beta}_0 - \hat{\beta}_1 TB$

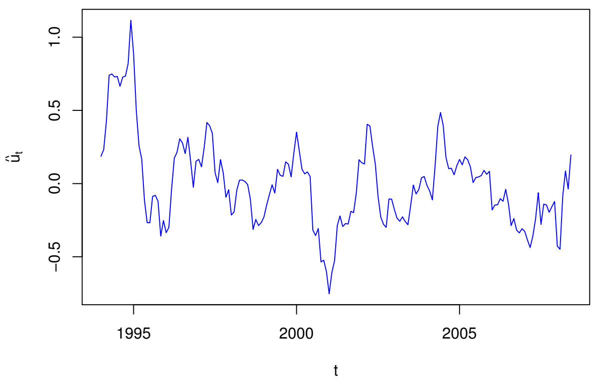


Figure 7: The plot of $\hat{\varepsilon} = GS - \hat{\alpha}_{\text{DOLS}} - \hat{\beta}_{\text{DOLS}} TB$

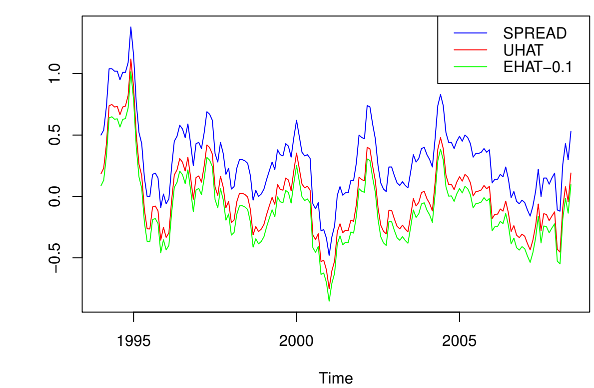


Figure 8: The plot of $SPREAD$, $\hat{u}$ and $\hat{\varepsilon}$

Appendix: R code

```
library(forecast)#install.packages("forecast")
library(urca)#install.packages("urca")
library(rugarch)#install.packages("rugarch")
library(vars)#install.packages("vars")
setwd("C:/Users/Junyong/OneDrive - University of Wisconsin Milwaukee/UWM/2017
    _01_Spring/Econometric Methods II")
#########EMPIRICAL EXERCISE 1##########
RAW1=read.table("170501_dsp.txt")
RSP500=ts(RAW1$V1,start=c(1960,2),frequency=12);plot(RSP500,xlab=expression(t)
    ,ylab=expression(RSP500[t]),col="blue")
acf(coredata(RSP500),col="blue")
pacf(coredata(RSP500),col="blue")
GARCH=ugarchfit(spec=ugarchspec(mean.model=list(armaOrder=c(0,0))),data=RSP500
    );GARCH
plot(ts(GARCH@fit$var,start=c(1960,2),frequency=12),col="blue",xlab=expression
    (t),ylab=expression(sigma[t]^2))
EGARCH=ugarchfit(spec=ugarchspec(mean.model=list(armaOrder=c(0,0)),variance.
    model=list(model="eGARCH")),data=RSP500);EGARCH
TGARCH=ugarchfit(spec=ugarchspec(mean.model=list(armaOrder=c(0,0)),variance.
    model=list(model="gjrGARCH")),data=RSP500);TGARCH
plot(newsimpact(GARCH)$zx,newsimpact(GARCH)$zy,xlab=newsimpact(GARCH)$xexpr,
    ylab=newsimpact(GARCH)$yexpr,type="l",col="blue")
plot(newsimpact(EGARCH)$zx,newsimpact(EGARCH)$zy,xlab=newsimpact(EGARCH)$xexpr
    ,ylab=newsimpact(EGARCH)$yexpr,type="l",col="blue")
plot(newsimpact(TGARCH)$zx,newsimpact(TGARCH)$zy,xlab=newsimpact(TGARCH)$xexpr
    ,ylab=newsimpact(TGARCH)$yexpr,type="l",col="blue")
#########EMPIRICAL EXERCISE 2##########
RAW2=read.table("170501_TERM.txt",header=T)
TB=ts(RAW2$TB,start=c(1994,1),frequency=12)
GS=ts(RAW2$GS,start=c(1994,1),frequency=12)
ts.plot(TB,GS,gpars=list(col=c("BLUE","RED")))
legend("topright",legend=c("TB","GS"),col=c("blue","red"),lty=1)
SPREAD=GS-TB;plot(SPREAD,col="blue",xlab=expression(t),ylab=expression(SPREAD[
    t]))
summary(ur.df(SPREAD,type="drift",selectlags="AIC"))
summary(ur.pp(SPREAD,type=c("Z-tau"),model=c("constant")))
UHAT=ts(lm(GS~1+TB)$resid,start=c(1994,1),frequency=12)
plot(UHAT,col="blue",xlab=expression(t),ylab=expression(hat(u)[t]))
summary(ur.df(UHAT,type="drift",selectlags="AIC"))
summary(ur.pp(UHAT,type=c("Z-tau"),model=c("constant")))
GSTB=ts.union(GS,TB);VARselect(GSTB,lag.max=12,type="const")
DTB=diff(TB)
DTB.1=lag(DTB,1)
DTB.2=lag(DTB,2)
DTB.3=lag(DTB,3)
DTB.4=lag(DTB,4)
DTB.5=lag(DTB,5)
DTB1=lag(DTB,-1)
DTB2=lag(DTB,-2)
DTB3=lag(DTB,-3)
DTB4=lag(DTB,-4)
DTB5=lag(DTB,-5)
```

```
PANEL=ts.union(GS,TB,DTB.1,DTB.2,DTB.3,DTB.4,DTB.5,DTB1,DTB2,DTB3,DTB4,DTB5)
rm(DTB,DTB.1,DTB.2,DTB.3,DTB.4,DTB.5,DTB1,DTB2,DTB3,DTB4,DTB5)
AIC=matrix(0,6,6)
AIC[1,1]=AIC(lm(PANEL[,1]~1+PANEL[,2]))
AIC[1,2]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,8]))
AIC[1,3]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,8:9]))
AIC[1,4]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,8:10]))
AIC[1,5]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,8:11]))
AIC[1,6]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,8:12]))
AIC[2,1]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3]))
AIC[2,2]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3]+PANEL[,8]))
AIC[2,3]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3]+PANEL[,8:9]))
AIC[2,4]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3]+PANEL[,8:10]))
AIC[2,5]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3]+PANEL[,8:11]))
AIC[2,6]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3]+PANEL[,8:12]))
AIC[3,1]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:4]))
AIC[3,2]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:4]+PANEL[,8]))
AIC[3,3]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:4]+PANEL[,8:9]))
AIC[3,4]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:4]+PANEL[,8:10]))
AIC[3,5]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:4]+PANEL[,8:11]))
AIC[3,6]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:4]+PANEL[,8:12]))
AIC[4,1]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:5]))
AIC[4,2]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:5]+PANEL[,8]))
AIC[4,3]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:5]+PANEL[,8:9]))
AIC[4,4]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:5]+PANEL[,8:10]))
AIC[4,5]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:5]+PANEL[,8:11]))
AIC[4,6]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:5]+PANEL[,8:12]))
AIC[5,1]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:6]))
AIC[5,2]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:6]+PANEL[,8]))
AIC[5,3]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:6]+PANEL[,8:9]))
AIC[5,4]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:6]+PANEL[,8:10]))
AIC[5,5]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:6]+PANEL[,8:11]))
AIC[5,6]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:6]+PANEL[,8:12]))
AIC[6,1]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:7]))
AIC[6,2]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:7]+PANEL[,8]))
AIC[6,3]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:7]+PANEL[,8:9]))
AIC[6,4]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:7]+PANEL[,8:10]))
AIC[6,5]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:7]+PANEL[,8:11]))
AIC[6,6]=AIC(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:7]+PANEL[,8:12]))
AIC;summary(lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:5]+PANEL[,8:12]))
DOLS=lm(PANEL[,1]~1+PANEL[,2]+PANEL[,3:5]+PANEL[,8:12])
UHAT2=ts(GS-DOLS$coefficients[1]-DOLS$coefficients[2]*TB,start=c(1994,1),
    frequency=12)
plot(UHAT2,col="blue",xlab=expression(t),ylab=expression(hat(u)[t]))
summary(ur.df(UHAT2,type="drift",selectlags="AIC"))
summary(ur.pp(UHAT2,type=c("Z-tau"),model=c("constant")))
ts.plot(SPREAD,UHAT,UHAT2-0.1,gpars=list(col=c("blue","red","green")))
legend("topright",legend=c("SPREAD","UHAT","EHAT-0.1"),col=c("blue","red","
    green"),lty=1)
summary(lm(diff(GS)~1+UHAT2[1:173]))
summary(lm(diff(TB)~1+UHAT2[1:173]))
summary(lm(diff(GS)[2:173]~1+UHAT2[1:172]+diff(GS)[1:172]+diff(TB)[1:172]))
summary(lm(diff(TB)[2:173]~1+UHAT2[1:172]+diff(GS)[1:172]+diff(TB)[1:172]))
```

```
summary(lm(diff(GS)[3:173]~1+UHAT2[2:172]+diff(GS)[2:172]+diff(TB)[2:172]+diff
    (GS)[1:171]+diff(TB)[1:171]))
summary(lm(diff(TB)[3:173]~1+UHAT2[2:172]+diff(GS)[2:172]+diff(TB)[2:172]+diff
    (GS)[1:171]+diff(TB)[1:171]))
summary(lm(diff(GS)[4:173]~1+UHAT2[3:172]+diff(GS)[3:172]+diff(TB)[3:172]+diff
    (GS)[2:171]+diff(TB)[2:171]+diff(GS)[1:170]+diff(TB)[1:170]))
summary(lm(diff(TB)[4:173]~1+UHAT2[3:172]+diff(GS)[3:172]+diff(TB)[3:172]+diff
    (GS)[2:171]+diff(TB)[2:171]+diff(GS)[1:170]+diff(TB)[1:170]))
VARselect(ts.union(TB,GS))
summary(ca.jo(ts.union(TB,GS),ecdet="const",type="trace",K=4,spec="transitory
    "))
summary(ca.jo(ts.union(TB,GS),ecdet="const",type="eigen",K=4,spec="transitory
    "))
summary(cajorls(ca.jo(ts.union(TB,GS),ecdet="const",type="trace",K=4,spec="
   transitory"),r=1)$rlm)
print(cajorls(ca.jo(ts.union(TB,GS),ecdet="const",type="trace",K=4,spec="
   transitory"),r=1))
summary(cajorls(ca.jo(ts.union(GS,TB),ecdet="const",type="trace",K=4,spec="
   transitory"),r=1)$rlm)
print(cajorls(ca.jo(ts.union(GS,TB),ecdet="const",type="trace",K=4,spec="
   transitory"),r=1))
```