Advanced Time Series Analysis
WS 2008/2009
Questions for review and theoretical assignments

Lecture 1

1. The stochastic process $\{\varepsilon_t\}(t = 1, 2, \cdots)$ consists of independent random variables $\varepsilon_t \sim N(0, 1)$. Compute the probability $P(\varepsilon_t \leq 0 \cap \varepsilon_{t+1} > 1.96 \cap \varepsilon_{t+2} \leq -1.96)$.

2. Write the joint density $f_{\varepsilon_t \varepsilon_{t+1}}(\varepsilon_t, \varepsilon_{t+1})$. Interpret your result.

3. Write the conditional density $f_{\varepsilon_{t+1}|\varepsilon_t}(\varepsilon_{t+1}|\varepsilon_t)$.

4. Denote a realisation of the stochastic process $\{\varepsilon_t\}$ as $\{x_1, x_2, \cdot\cdot, x_T\}$.
   Write down the joint density function of the random vector $\underline{\varepsilon} = \{\varepsilon_1, \varepsilon_2, \cdot\cdot, \varepsilon_T\}$ evaluated at $\{x_1, x_2, \cdot\cdot, x_T\}$.

   Since the random vector $\underline{\varepsilon} = \{\varepsilon_1, \varepsilon_2, \cdot\cdot, \varepsilon_T\}$ is jointly normally distributed you can use the multivariate normal density which is generally written as

   $$f_{\underline{X}} = 2\pi^{-n/2} |\Omega|^{-0.5}\, exp\left[\frac{(\underline{x} - \underline{\mu})'\Omega^{-1}(\underline{x} - \underline{\mu})}{-2}\right]$$

   What is in our example $n, \underline{x}, \underline{\mu}$ and Ω ?

5. Is the process $\{\varepsilon_t\}$ weakly stationary?

6. Is the process $\{\varepsilon_t\}$ strictly stationary?

7. A new stochastic process $\{Y_t\}$ is generated as $Y_t = a + b \cdot \varepsilon_t$
   The joint distribution of $\underline{Y} = (Y_1, Y_2, \cdot\cdot, Y_T)$ is still the multivariate normal (see 4.)
   What is $\underline{\mu}$ and Ω now?

8. $\{X_t\}$ denotes a stochastic process. We have $E(X_t) = E(X_{t+1}) = 2$
   $cov(X_t, X_{t+1}) = 2$ and $var(X_t) = var(X_{t+1}) = 1$
   using $A = \begin{bmatrix} 0.3 & 0.7 \\ 0.5 & 0.5 \end{bmatrix}$ we generate two new random variables Z_1, Z_2 by
   $\underline{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = A \cdot \begin{bmatrix} X_t \\ X_{t+1} \end{bmatrix}$
   compute $E(\underline{Z})$ and $cov(\underline{Z}) = \begin{bmatrix} var\,(Z_1)\, cov\,(Z_1, Z_2) \\ cov\,(Z_1, Z_2)\, var\,(Z_2) \end{bmatrix}$

Solutions to the assignments from Lecture 1:

1. $P(\varepsilon_t \leq 0) \cdot P(\varepsilon_{t+1} > 1.96) \cdot P(\varepsilon_t \leq -1.96) = 0.5 \cdot 0.025 \cdot 0.025 = 0.0003125$

8. $E(\underline{Z}) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

   $cov(\underline{Z}) = \begin{bmatrix} 1.42 & 1.5 \\ 1.5 & 1.5 \end{bmatrix}$

Lecture 2

1. Are the following stochastic processes $\{y_t\}$ stationary and ergodic?

$$\left[\begin{array}{ll} \{\varepsilon_t\} & \text{denotes a Gaussian white noise process} \\ \text{i.e.} & \mathbb{E}(\varepsilon_t) = 0, \quad \mathbb{E}(\varepsilon_t^2) = Var(\varepsilon_t) = \sigma^2, \quad \mathbb{E}(\varepsilon_t \cdot \varepsilon_\tau) = 0 \quad t \neq \tau \end{array}\right]$$

$$\begin{array}{llllll} \text{a)} & y_t & = & \varepsilon_t & & \\ \text{b)} & y_t & = & y_{t-1} + \varepsilon_t & \text{with} & y_1 = \varepsilon_1 \\ \text{c)} & y_t & = & y_{t-1} - y_{t-2} + \varepsilon_t & \text{with} & y_1 = \varepsilon_1 \\ \text{d)} & y_t & = & a \cdot t + \varepsilon_t & \text{with} & a \quad \text{a real number} \end{array}$$

2. Compute $\mathbb{E}(y_t - \mu)(y_{t-j} - \mu)$ [i.e. $cov(y_t, y_{t-j})$] for the stochastic processes b) and d).

3. - Check, by writing $\mathbb{E}(y_t)$, $Var(y_t)$ and $cov(y_t, y_{t-j})$ $j \geq 1$, whether a MA(2) process

   $y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$

   is stationary and ergodic.

   - Plot the autocorrelation function for a MA(2) where $\theta_1 = 0.5$ and $\theta_2 = -0.3$.

4. Write $\mathbb{E}(y_t)$ and $Var(y_t)$ for a MA(q) process.
   $y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t$

5. The sequence of autocovariances $\{\gamma_j\}_{j=0}^{\infty}$ of a Gaussian process $\{y_t\}$ evolves as
   $\gamma_j = \theta^j$ where $|\theta| < 1$.
   Is the process ergodic?

6. What do we mean by a Gaussian process?

7. Why is ergodic stationarity such an important property for the purpose of estimating the moments $\mathbb{E}(y_t)$, $Var(y_t)$, $cov(y_t, y_{t-j})$,... of a stochastic process $\{y_t\}$?

   Hint: refer to the ergodic theorem (Hayashi, *Econometrics*, p. 101) and note that if $\{y_t\}$ is stationary and ergodic, so is $\{f(y_t)\}$ where $f(\cdot)$ is a measurable function like $\ln(y_t)$, y_t^2 i.e. a function that produces a new random variable.

8. A MA(∞) is given by
   $y_t = \mu + \theta^2 \varepsilon_{t-1} + \theta^4 \varepsilon_{t-2} + \theta^6 \varepsilon_{t-3} + ...$
   where $|\theta| < 1$.
   Compute $\mathbb{E}(y_t)$ and $Var(y_t)$.

9. An AR(1) process is given by

   $Y_t = 0.5 + 0.9\, Y_{t-1} + \varepsilon_t$ where $\{\varepsilon_t\}$ is Gaussian White Noise $\varepsilon_t \sim N(0, 9)$

Compute $E(Y_t)$ and $Var(Y_t)$. Compute the first 5 auto covariances $\gamma_1, \gamma_2, \cdot\cdot, \gamma_5$ and plot the corresponding autocorrelations $\rho_1, \rho_2, \cdot\cdot, \rho_5$.

Hint $\rho_j = \frac{Cov(Y_t, Yt-j)}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t-j})}} = \frac{\gamma_j}{\gamma_0}$

10. What do the terms weak stationarity and ergodicity mean? Explain intuitively and mathematically.
11. Strict stationarity does not necessarily imply weak stationarity. When? Give an example.
12. Why are stationarity and ergodicity so important in econometrics?
13. Explain in your words the meaning of the weak law of large number (WLLN) and the central limit theorem (CLT).
14. You are generally not allowed to interchange expectation and summation operators for each and every infinite sequence of random variables. But when are you? Illustrate the general problem when computing the mean of an MA(∞).
15. Which property hast to be fulfilled that you are allowed to interchange these operators when computing the variance of an MA(∞)?
16. Explain why an AR(1) is actually also an MA(∞)?

Solutions to the assignments from Lecture 2:

9. $E(y_t) = 5$; $var(y_t) = 47.368$; $\gamma_1 = 42.632$; $\gamma_2 = 38.368$; $\gamma_3 = 34.532$; $\gamma_4 = 31.078$; $\gamma_5 = 27.971$

Lecture 3

1. Explain the general idea of ML estimation in one sentence!
2. Which problems can arise if we use the joint density $f_{Y_1,\ldots,Y_T}(y_1,\ldots,y_T)$ to conduct ML estimation? How can these problems be overcome?
3. Write down the joint density $f_{Y_1,\ldots,Y_5}(y_1,\ldots,y_5)$ as product of four conditional densities and one marginal density!
4. You want to construct the exact likelihood function of an AR(2) process

   $$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \qquad \varepsilon_t \sim N(0,\sigma^2) \text{ and i.i.d.}$$

   a) Write down the joint density of the first two observations $f_{Y_1,Y_2}(y_1,y_2)$.

   b) Using the conditional density of the third observation $f_{Y_3|Y_2,Y_1}(y_3|y_2,y_1)$ write down the joint density of the first three observations $f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3)$.

5. In what respect is the likelihood function a random variable? Explain!
6. In what respect does the conditional likelihood function differ from the exact likelihood function?
7. a) Write down the joint density of the first three observations of the $MA(3)$ process
   $Y_t = c + 0.3\varepsilon_{t-1} + 0.2\varepsilon_{t-2} - 0.1\varepsilon_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0,\sigma^2)$ and ε_t i.i.d. $N(0,\sigma^2)$.

   b) Suppose you want to set up the conditional likelihood function of this process. You condition on pre-sample values $\varepsilon_0, \varepsilon_{-1}, \varepsilon_{-2}$. Write down the first three elements of the conditional likelihood function.

   $f_{Y_1|\varepsilon_0=0,\varepsilon_{-1}=0,\varepsilon_{-2}=0} =$
   $f_{Y_2|Y_1,\varepsilon_0=0,\varepsilon_{-1}=0,\varepsilon_{-2}=0} =$
   $f_{Y_3|Y_1,Y_2,\varepsilon_0=0,\varepsilon_{-1}=0,\varepsilon_{-2}=0} =$

   c) Which condition has to hold in order to make the Conditional Maximum-Likelihood work?

8. When do we call a MA process invertible? Explain the intuition of this property!
9. What does the CAN property imply?
10. Which data requirements have to be fulfilled in order to ensure the CAN property of the ML estimates?

11. Where do bounds for parameters come from? Give examples of bounded parameter spaces.

12. What is the Cramer-Rao lower bound all about? How is it related to the Fisher information matrix?

13. You have succeeded in providing Maximum-Likelihood estimates of the parameters of an $ARMA(2,2)$ process.

$$(1 - L\phi_1 - L\phi_2)Y_t = c + (1 + \theta_1 L + \theta_2 L^2)\varepsilon_t \qquad \varepsilon_t \sim \; i.i.d.N(0, \sigma^2)$$

The (conditioned) Maximum-Likelihood estimates are

$$\begin{array}{llll} \hat{c} & = 0.2 & \hat{\theta}_1 & = 0.2 \\ \hat{\phi}_1 & = 0.6 & \hat{\theta}_2 & = -0.1 \\ \hat{\phi}_2 & = 0.1 & \hat{\sigma}^2 & = 0.8 \end{array}$$

The value of the log likelihood function evaluated at these estimates is -1432.6.

Suppose you want to test the null hypothesis

$$H_0 : \theta_1 = 0.5 \text{ against } H_A : \theta_1 \neq 0.5$$

and $H_0 : \theta_1 = 0$ against $H_A : \theta_1 \neq 0$

Perform and interpret the appropriate tests.

An estimate of the variance-covariance matrix of the estimates $\hat{\theta} = (\hat{c}, \hat{\phi}_1, \hat{\phi}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2)$ is given by

$$\widehat{Var}(\hat{\theta}) = \left[-\frac{\partial^2 lnL(\theta)}{\partial\theta\partial\theta'} \bigg|_{\hat{\theta}} \right]^{-1} = \begin{bmatrix} 0.007 & \cdots & & & & \vdots \\ 0.001 & 0.005 & & & & \\ 0.002 & 0.001 & 0.003 & & & \\ 0.003 & 0.002 & 0.001 & 0.01 & & \\ 0.001 & 0.003 & 0.004 & 0.001 & 0.002 & \\ 0.001 & 0.0001 & 0.0001 & 0.0001 & 0.00002 & 0.0001 \end{bmatrix}$$

$\theta = (c, \phi_1, \phi_2, \theta_1, \theta_2, \sigma^2)'$

You have also estimated an $ARMA(2,0)$ i.e. an $AR(2)$ model. The estimation of this restricted model yields a log likelihood value equal to -1434.3.
Compute and interpret a likelihood ratio statistic to test the hypothesis that the restrictions implied by the $ARMA(2,0)$ specification are correct. Here the $ARMA(2,2)$

specification is the unrestricted model, the $ARMA(2,0)$ is the restricted model.

As another alternative you have estimated an $MA(2)$ model. The log likelihood evaluated at the maximum likelihood estimates is -1442.2. Perform a test of the $ARMA(2,2)$ specification against the MA(2) model.

14. What are the possibilities to compute an estimate of $Var(\hat{\theta})$ in Gauss?

15. It may be the case that the Hessian based estimate of $Var(\hat{\theta})$ can not be computed due to an identification problem. Explain in which respect the curvature of the likelihood function around the true parameter value sheds light on the feasibility of identification!

16. Write a Gauss procedure to illustrate the surface of the likelihood function of an MA(1) evaluated at different parameter values θ:

    (a) Draw a sequence $\{\varepsilon_t\}$ of T i.i.d. standard normal random variables.

    (b) Compute $\{y_t\}$ from $Y_t = \theta\varepsilon_{t-1} + \varepsilon_t$ with $\theta = 0.5$.

    (c) Compute $\ln L(\theta)$ for $\theta = \{0.1 \ \ldots \ 0.9\}$ holding σ^2 fixed at 1.

    (d) Plot $\ln L(\theta)$ against θ for different T and compare your results.

Solutions to the assignments from Lecture 3:

13. test statistic: first null hypothesis $t_1 = \frac{0.2-0.5}{\sqrt{0.01}} = -3$
    second null hypothesis $t_2 = \frac{0.2}{\sqrt{0.01}} = 2$

    Likelihood ratio test statistic for ARMA(2,2) vs. AR(2): $LR_1 = 3.4$
    Likelihood ratio test statistic for ARMA(2,2) vs. MA(2): $LR_2 = 19.2$
    critical value: $\chi^2(2) = 5.99$

Lecture 4

1. What do we require to name a process a "Gaussian white noise process", an "independent white noise process", or a "white noise process"? Which process requires the strongest and which the weakest assumptions?

2. Explain the connection of ψ_j and the first row first column element of $\mathbf{F}^j$. Besides explain the eigenvalues of $\mathbf{F}^j$.

$$\mathbf{F} = \begin{pmatrix} \phi_1 & & \dots & & \phi_p \\ 1 & 0 & & & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \\ 0 & \dots & & 1 & 0 \end{pmatrix}$$

3. Given are 3 different AR processes

$$(1)\ \begin{matrix} \phi_1 = 0.6 \\ \phi_2 = -0.4 \end{matrix}\ , \quad (2)\ \begin{matrix} \phi_1 = 0.4 \\ \phi_2 = 0.8 \\ \phi_3 = -0.3 \end{matrix}\ , \quad (3)\ \begin{matrix} \phi_1 = 1.2 \\ \phi_2 = -0.1 \end{matrix}\ .$$

   Write out the $\mathbf{F}$-matrix for each process and compute the eigenvalues of $\mathbf{F}$ with Gauss, to check whether these AR processes are stationary.

4. Why do we write an $AR(p): Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t$ as a vector $AR(1): \boldsymbol{\xi} = \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$? Explain $\boldsymbol{\xi}$, $\mathbf{F}$ and $\mathbf{v}$. Write out the first row of the left and the right hand side of $\boldsymbol{\xi} = \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$.

5. Given is:

$$\begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & & \dots & & \phi_2 \\ 1 & 0 & & & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \\ 0 & \dots & & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix} = \begin{pmatrix} \omega_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

   Write out in detail the recursion and the first row of that recursion. Write a Gauss procedure to compute $\mathbf{F}^j$ and read out the first row first column element and plot these elements (x-axis: j, y-axis: $f_{11}^{(j)}$).

6. In the following, $\{\varepsilon_t\}$ denotes a Gaussian White Noise process. Which of the following processes $\{Y_t\}$ is a stationary and ergodic process? Give a brief explanatory statement and describe each process as a special case of an ARMA(p,q) process. For example 'This is a stationary AR(2) process...' et cetera.

$$
\begin{array}{ll}
(a) & (1-0.5L-0.7L^2)Y_t = \varepsilon_t \\
(b) & (1-0.9L-0.1L^2)Y_t = (1+0.3L)\varepsilon_t \\
(c) & Y_t = (1-L)\varepsilon_t \\
(d) & Y_t = (1+0.9L^2)\varepsilon_t \\
(e) & Y_t = c + 0.5Y_{t-1} + 0.3Y_{t-2} + 1.2\varepsilon_{t-1} + \varepsilon_t \\
(f) & Y_t = \dfrac{(1-1.3L^2)}{1-0.8L-0.1L^2}\varepsilon_t \\
(g) & (1-0.9L)Y_t = \varepsilon_t \\
(h) & (1-0.8L-0.1L^2)Y_t = \varepsilon_t \\
(i) & Y_t = (1+0.4L+0.3L^2)\varepsilon_t
\end{array}
$$

7. Give your opinion to the following statements. Answer "Correct, since..." or "Incorrect, rather..."

   (a) Any MA process is a stationary process .

   (b) Any finite Gaussian AR(p) process is stationary .

   (c) Whether an ARMA(p,q) is stationary is solely determined by its MA part.

   (f) A White Noise process is an ergodic process.

   (g) Any finite MA(q) is ergodic.

Solutions to the assignments from Lecture 4:

3.

(1) stationary	(2) stationary	(3) not stationary
$\lambda_1 = 0.30 + 0.55677644i$	$\lambda_1 = 0.91584462$	$\lambda_1 = 1.1099020$
$\lambda_2 = 0.30 - 0.55677644i$	$\lambda_2 = -0.88568851$	$\lambda_2 = 0.090098049$
	$\lambda_3 = 0.36984389$	

6. (a) $\lambda_1 = 1.123$ $\lambda_2 = -0.623 \rightarrow$ not stationary;

   (b) finite MA(q) stationary, Check AR part: $\lambda_1 = 1$ $\lambda_2 = -0.1 \rightarrow$ not stationary;

   (c),(d),(i) finite MA(q) stationary

   (e) $\lambda_1 = -0.352$ $\lambda_2 = 0.852 \rightarrow$ stationary;

(f),(h) $\lambda_1 = 0.910$ $\lambda_2 = -0.110 \rightarrow$ stationary

(g) stationary

Lecture 5

1. Why is the loss function important for choosing an optimal forecast?

2. How is the mean squared error (MSE) defined and what does this imply for the sensitivity towards under- and overestimation of a forecast?

3. Your client's loss function is the MSE and your task is to present a forecast of a financial time series. What is the best forecast you can deliver (a) generally, and (b) in the case of a martingale?

4. Explain the intuition of the stochastic discount factor m_{t+1}.

5. Under which circumstances does theory predict that prices (approximately or exactly) follow a random walk?

6. Why do we use log returns in modeling financial time series?

7. Illustrate a series of returns in comparison to the same series' squared returns. What does the clustering of squared returns imply for the predictability of returns?

8. Summarize what the simplest ARCH model implies for $E(r_t)$, $var(r_t)$, $cov(r_t)$, $var(r_t \mid I_t)$, and $E(r_t \mid I_t)$. In which sense are these moments compatible with financial theory?

9. For modeling high frequency returns which ARMA(p,q) would be compatible with financial theory?

Lecture 6

1. The parameters of the basic ARCH(1) model are bounded. How and why?

2. Econometricians don't like imposing constraints outside the likelihood function for computational reasons. How can you directly impose constraints into the likelihood function? Give an example and explain.

3. In which respect can the GARCH model $h_t = d + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}^2$ be understood as a kind of ARMA model? Which stylized fact of the data does such a specification take into account?

4. Empirical evidence states that the assumption of independently and identically normally distributed returns is not valid. Which special features of returns exist and how can they be captured within the ARCH framework appropriately?

5. The simple ARCH(1) model is able to generate a fat tailed distribution, but researchers argue that this is by far not enough. Propose an extension of this model that is able to generate a distribution with as much mass in the tails as required by the data.

6. Which parameter is of special interest in Nelson's E-ARCH model? Which stylized fact does this parameter allow to analyze? Are Nelson's findings in line with theory?

7. Which fundamental problems will arise if a consumption function $C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$ is estimated by OLS? How can these problems be overcome?

8. What is the fundamental methodological problem of a simultaneous equation system in macroeconomics when the equations are estimated by OLS?

9. What are the problems with the SVAR in primitive form from econometric point of view?

Lecture 7

1. There are three sensible Dickey-Fuller tests that allow to test for a unit root. As an applied guy, what is the general principle that guides your choice of the test?
2. Explain why the combination of the null hypothesis and the alternative presented in case 3 (Hamilton, table 17.1) does not make sense.
3. Which is the "correct" case (Hamilton, table 17.1) to use in order to test the null hypothesis of a unit root in the exemplary time series presented in figure 1? Defend your suggestion.
4. Explain the δ-method to a second year statistic student.
5. In a GARCH(1,1) estimation we ensure that the conditional variance h_t is always positive by imposing the following restrictions on the parameters

$$h_t = \exp(\omega^*) + \exp(\alpha^*)\varepsilon_{t-1}^2 + \exp(\beta^*)h_{t-1}.$$

Estimates of the original parameters $(\omega, \alpha, \beta)'$ can be backed out from $(\hat{\omega}^*, \hat{\alpha}^*, \hat{\beta}^*)'$ by simply taking $\hat{\omega} = \exp(\hat{\omega}^*)$, $\hat{\alpha} = \exp(\hat{\alpha}^*)$, and $\hat{\beta} = \exp(\hat{\beta}^*)$. In order to conduct proper inference we use the δ-method that delivers standard errors of $(\hat{\omega}, \hat{\alpha}, \hat{\beta})'$:

Suppose you have obtained estimates for $b = \begin{bmatrix} \omega^* \\ \alpha^* \\ \beta^* \end{bmatrix}$, $\quad \hat{b} = \begin{bmatrix} \hat{\omega}^* \\ \hat{\alpha}^* \\ \hat{\beta}^* \end{bmatrix} = \begin{bmatrix} -14.4866 \\ -2.8389 \\ -0.0661 \end{bmatrix}$.

We have

$$\sqrt{T}(\hat{b} - b) \underset{d}{\longrightarrow} N(0, \Sigma)$$

where Σ is the asymptotic variance covariance matrix.

A consistent estimate of $\frac{\Sigma}{T}$, denoted $\frac{\widehat{\Sigma}}{T}$, is given by

$$\frac{\widehat{\Sigma}}{T} = \begin{pmatrix} 0.5309 & 0.0558 & -0.0067 \\ 0.0558 & 0.0479 & -0.0031 \\ -0.0067 & -0.0031 & 0.0002 \end{pmatrix}.$$

The sample has $T = 894$ observations.

Provide estimates of $Var(\hat{\omega}^*)$, $Var(\hat{\alpha}^*)$, and $Var(\hat{\beta}^*)$ using this information.

You are interested in testing whether

$$\begin{aligned} H_0 &: \omega = \exp(\omega^*) = 0.1 & H_A &: \omega \neq 0.1 \\ H_0 &: \alpha = \exp(\alpha^*) = 0.1 & H_A &: \alpha \neq 0.1 \\ H_0 &: \beta = \exp(\beta^*) = 0.9 & H_A &: \beta \neq 0.9. \end{aligned}$$

Construct a suitable test statistic. For this purpose compute estimates of the variances of $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$, $Var(\hat{\omega})$, $Var(\hat{\alpha})$ and $Var(\hat{\beta})$, by using the δ-method.

Hints:

$$a(b) = \begin{bmatrix} \exp(\omega^*) \\ \exp(\alpha^*) \\ \exp(\beta^*) \end{bmatrix}$$

$$a(\hat{b}) \underset{p}{\longrightarrow} a(b)$$

$$\sqrt{T}(a(\hat{b}) - a(b)) \underset{d}{\longrightarrow} N(0, A(b)\Sigma A(b)')$$

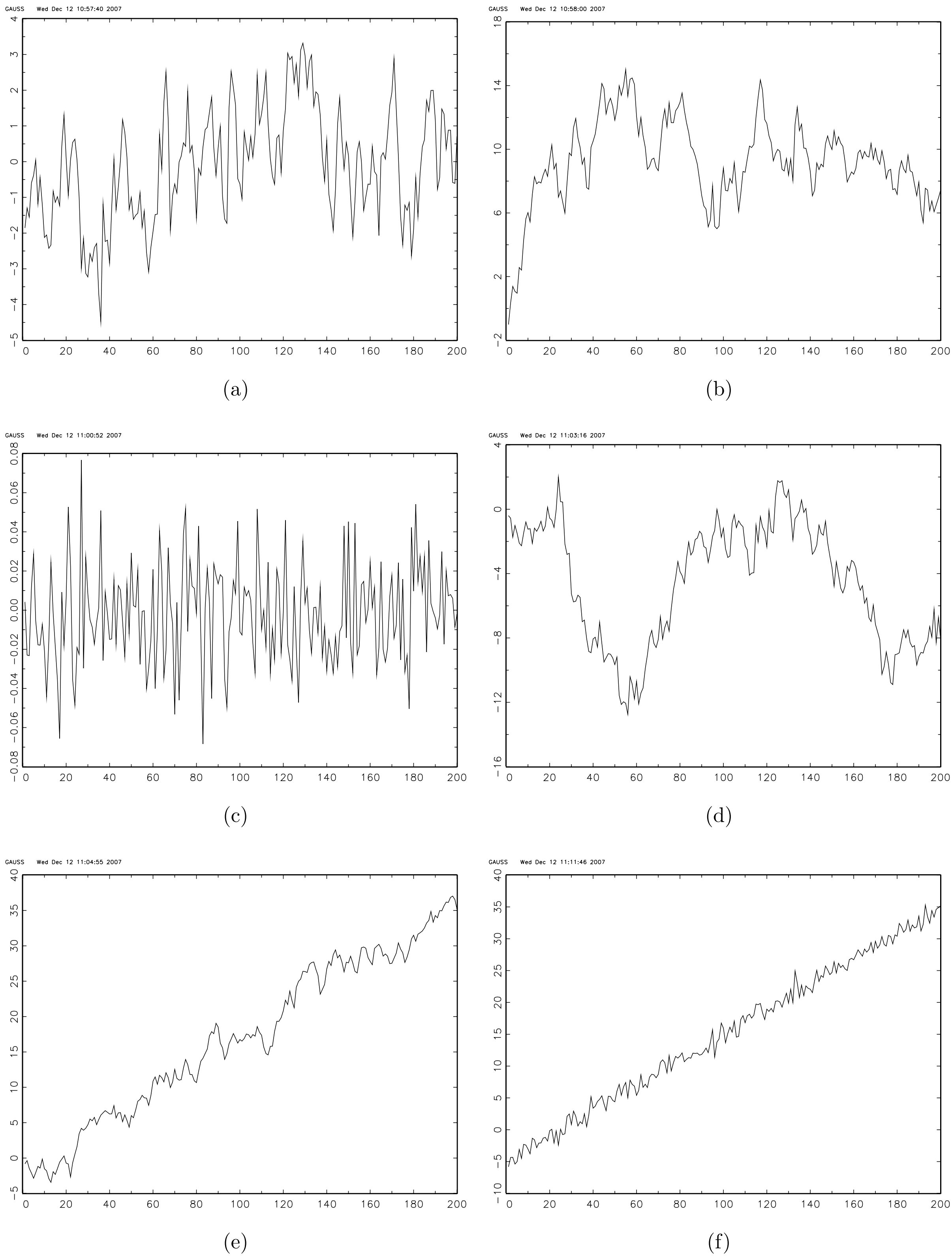
where $A(b) = \frac{\partial a(b)}{\partial b'}$ is of dimension (3×3).

The test statistics are

$$t = \frac{\hat{\omega} - 0.1}{\sqrt{\widehat{Var(\hat{\omega})}}}$$
$$t = \frac{\hat{\alpha} - 0.1}{\sqrt{\widehat{Var(\hat{\alpha})}}}$$
$$t = \frac{\hat{\beta} - 0.9}{\sqrt{\widehat{Var(\hat{\beta})}}}$$

t is approximately N(0,1) under the null hypothesis.

Figure 1: Exemplary Time Series



(a) (b)

(c) (d)

(e) (f)

Exercises: Delta Method

1. Solve the following assignemnet and check if you really understood the δ-Method:

   Suppose you have obtained estimates for $b = \begin{bmatrix} \theta \\ \phi \end{bmatrix}$ *i.e.* $\hat{b} = \begin{bmatrix} \hat{\theta} \\ \hat{\phi} \end{bmatrix}$.

   We have

   $$\sqrt{T}(\hat{b} - b) \underset{d}{\longrightarrow} N(0, \Sigma)$$

   where Σ is the asymptotic variance covariance matrix.

   A consistent estimate of Σ, denoted $\widehat{\Sigma}$, is given by

   $$\widehat{\Sigma} = \begin{pmatrix} 2 & 0.2 \\ 0.2 & 3 \end{pmatrix}$$

   The sample has $T = 100$ observations.
   Provide estimates of $Var(\hat{\theta})$ and $Var(\hat{\phi})$ using this information. The estimates are $\hat{\theta} = 0.6$ and $\hat{\phi} = 0.4$

   You are interested in testing whether

   $$r = \frac{\phi}{\phi + \theta} = 0.5$$

   Construct a suitable test statistic. For this purpose compute an estimate of the variance of $\hat{r} = \frac{\hat{\phi}}{\hat{\phi}+\hat{\theta}}$, $Var(\hat{r})$, by using the δ-method.

   Hints:

   $$a(b) = \frac{\phi}{\phi + \theta} = r$$

   $$\hat{r} = a(\hat{b}) \underset{p}{\longrightarrow} a(b)$$

   $$\sqrt{T}(a(\hat{b}) - a(b)) \underset{d}{\longrightarrow} N(0, A(b)\Sigma A(b)')$$

   where $A(b) = \frac{\partial a(b)}{\partial b'} = \left(\frac{\partial a(b)}{\partial \phi}, \frac{\partial a(b)}{\partial \theta}\right)$

   The test statistic is

$$t = \frac{\hat{r} - 0.5}{\sqrt{\widehat{Var(\hat{r})}}}$$

t is approximately N(0,1) under the Null Hypothesis that $r = 0.5$.

Lecture 8

1. The null process of an ordinary Dickey-Fuller test is a random walk which does not allow for serial correlation in first differences. Augmented Dickey-Fuller tests control for serial correlation by adding lagged first differences to the autoregressive equation.

   - When applying augmented Dickey-Fuller tests there exist a couple of information criteria that guide the lag length selection. Discuss the two most common ones.
   - How do they balance between a better fit and model parsimony?
   - What are the advantages/disadvantages of these criteria in the finite sample context?

2. A researcher wants to conduct an augmented Dickey-Fuller test to test an economic time series for a unit root. He estimates a regression model of the form:

   $$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \alpha + \rho y_{t-1} + \delta t + \varepsilon_t$$

   The researcher works under the null hypothesis that the true data generating process is given by:

   $$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \alpha + y_{t-1} + \varepsilon_t$$

   Running the regression the researcher computed the estimate $\hat{\rho} = 0.95$. The estimated OLS standard error $s.e.(\hat{\rho}) = 0.013$. The sample size is 50. Interpret the result.

Lecture 9

1. The null process of an ordinary Dickey-Fuller test is a random walk which does not allow for serial correlation in first differences. Augmented Dickey-Fuller tests control for serial correlation by adding lagged first differences to the autoregressive equation.

   - When applying augmented Dickey-Fuller tests there exist a couple of information criteria that guide the lag length selection. Discuss the two most common ones.
   - How do they balance between a better fit and model parsimony?
   - What are the advantages/disadvantages of these criteria in the finite sample context?

2. A researcher wants to conduct an augmented Dickey-Fuller test to test an economic time series for a unit root. He estimates a regression model of the form:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \alpha + \rho y_{t-1} + \delta t + \varepsilon_t$$

   The researcher works under the null hypothesis that the true data generating process is given by:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \alpha + y_{t-1} + \varepsilon_t$$

   Running the regression the researcher computed the estimate $\hat{\rho} = 0.95$. The estimated OLS standard error $s.e.(\hat{\rho}) = 0.013$. The sample size is 50. Interpret the result.

3. Consider the following SVAR in order to analyze the dependencies of the three East Asian stock markets Tokyo (T), Singapore (S) and South Korea (K)

$$\begin{aligned}
r_t^T &= k^T & &+\beta_{12}^{(0)} r_t^S + \beta_{13}^{(0)} r_t^K & &+\beta_{11}^{(1)} r_{t-1}^T + \beta_{12}^{(1)} r_{t-1}^S + \beta_{13}^{(1)} r_{t-1}^K + \beta_{11}^{(2)} r_{t-2}^T + \beta_{12}^{(2)} r_{t-2}^S + \beta_{13}^{(2)} r_{t-2}^K + u_t^T \\
r_t^S &= k^S + \beta_{21}^{(0)} r_t^T & &+ \beta_{23}^{(0)} r_t^K & &+\beta_{21}^{(1)} r_{t-1}^T + \beta_{22}^{(1)} r_{t-1}^S + \beta_{23}^{(1)} r_{t-1}^K + \beta_{21}^{(2)} r_{t-2}^T + \beta_{22}^{(2)} r_{t-2}^S + \beta_{23}^{(2)} r_{t-2}^K + u_t^S \\
r_t^K &= k^K + \beta_{31}^{(0)} r_t^T & &+\beta_{32}^{(0)} r_t^S & &+\beta_{31}^{(1)} r_{t-1}^T + \beta_{32}^{(1)} r_{t-1}^S + \beta_{33}^{(1)} r_{t-1}^K + \beta_{31}^{(2)} r_{t-2}^T + \beta_{32}^{(2)} r_{t-2}^S + \beta_{33}^{(2)} r_{t-2}^K + u_t^K
\end{aligned}$$

   (i) Write the primitive form of this SVAR in matrix notation using the notation as proposed in the script.

   (ii) Derive the SVAR in standard form.

   (iii) Write in detail the contemporaneous variance covariance matrices of both the idiosyncratic shocks ($\mathbb{E}[\mathbf{u}\mathbf{u}']$) and the composite shocks ($\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']$).

   (iv) Write the SVAR in standard form in lag operator notation.

   (v) The VAR(2) can be rewritten as VAR(1) following $\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t$. Write $\boldsymbol{\xi}_t$, $\mathbf{F}$ and $\boldsymbol{v}_t$ extensively.

4. Describe 3 options to compute the sequence of $\boldsymbol{\Psi}$ matrices in the VMA(∞) representation of an VAR(p).

5. Explain the relation of the $\mathbf{\Psi}$ sequence to the eigenvalues of $\mathbf{F}$, and argue how the latter shed light on the stationarity and ergodicity of a VAR(p). Why is it the largest eigenvalue of $\mathbf{F}$ that determines the strength of the effect of a past innovation on the present in a VAR(p)?

6. Why are we interested in tracing the effect of an orthogonal shock instead of that of a composite shock in the first place?

7. What are the problems associated with identifying idiosyncratic shocks via Cholesky decomposition of $\mathbf{\Omega}$=$\mathbf{ADA}$'?

8. How are $\mathbf{A}$ and $\mathbf{B}_0$ related? Does that relation always make sense from an economic perspective?

9. How many possibilities would you have to identify $\mathbf{B}_0$ via Cholesky decomposition with $n = 4$ variabes?

10. Think of an economic system that would justify the identification of $\mathbf{B}_0$ via Cholesky decomposition.

Lecture 10

1. Describe briefly the two most widely used methodologies to assess the dynamic relations between the variables in a VAR system.

2. Once the VMA(∞) representation of a VAR(p) is derived it is relatively easy to compute the non-orthogonalized impulse response functions (IRFs) that give the response of the system to one unit shocks in the composite innovations. Illustrate!

3. What makes it challenging to derive the orthogonalized IRFs that give the response of the system to one unit shocks in the idiosyncratic innovations?

4. The analysis of the effects of Swiss Monetary policy delivers the following orthogonalized impulse-response functions with Cholesky ordering: $\tilde{m}, \tilde{r}, \tilde{p}, \tilde{y}$. Interpret!

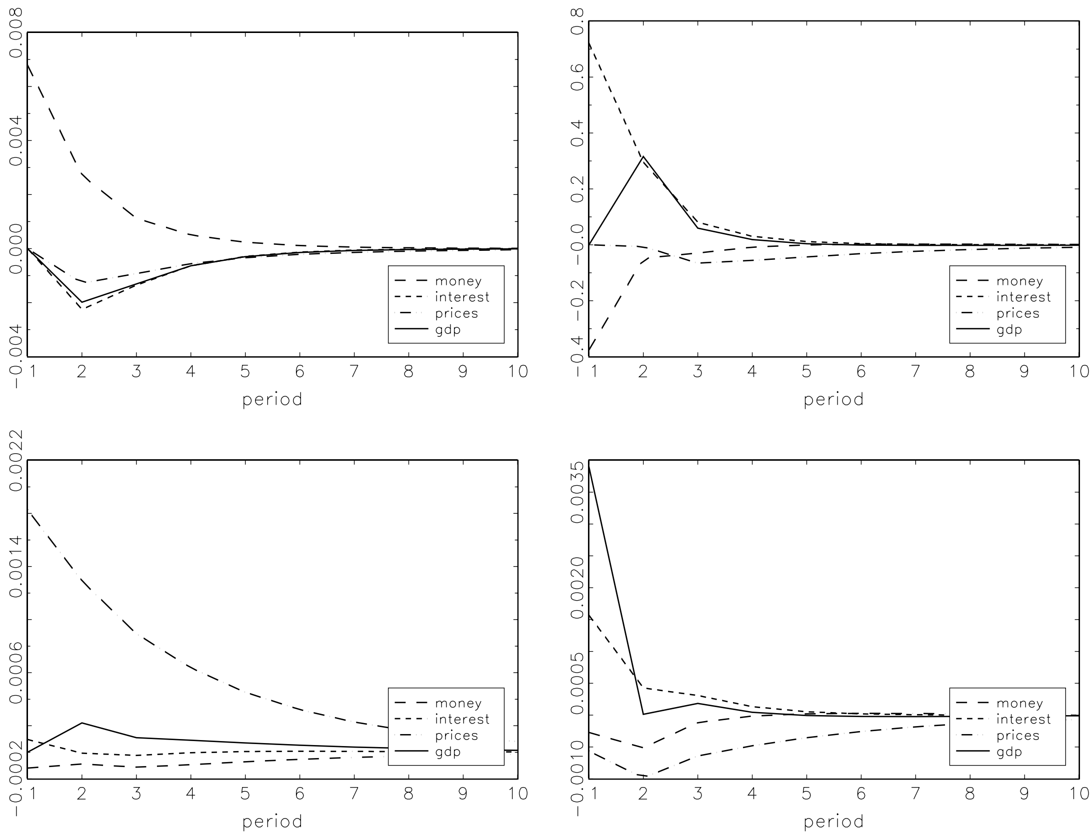


Figure 2: **Orthogonalized impulse-response functions** Upper left panel: Response of money to shocks in money, interest rate, prices, and GDP. Upper right panel: Response of interest rate to shocks in money, interest rate, prices, and GDP. Lower left panel: Response of prices to shocks in money, interest rate, prices, and GDP. Lower right panel: Response of GDP to shocks in money, interest rate, prices, and GDP.

5. In VAR modeling we consider how each of the orthogonalized disturbances $(u_{1t}, \ldots, u_{nt})$ contributes to the mean squared error of the s-period-ahead forecast of $\mathbf{y}_t$. What is written on the main diagonal of $MSE(\hat{\mathbf{y}}_{t+s|t})$?

6. Write in detail (i) $MSE(\hat{\mathbf{y}}_{t+1|t})$, and (ii) $MSE(\hat{\mathbf{y}}_{t+2|t})$ for an arbitrary three variable case $(n = 3)$. In which respect do (i) and (ii) differ, and why is it again the Cholesky decomposition that plays a crucial role?

7. Consider the following numerical examples

$$
\begin{aligned}
MSE(\hat{\mathbf{y}}_{t+1|t}) &= 1.79 \cdot \begin{bmatrix} 1.000 & 0.344 & 0.087 \\ 0.344 & 0.119 & 0.030 \\ 0.087 & 0.030 & 0.008 \end{bmatrix} + 1.78 \cdot \begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.127 \\ 0.000 & 0.127 & 0.016 \end{bmatrix} \\
&+ 2.63 \cdot \begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 \end{bmatrix} \\
&= \begin{bmatrix} 1.793 & 0.618 & 0.157 \\ 0.618 & 1.994 & 0.281 \\ 0.157 & 0.281 & 2.674 \end{bmatrix}
\end{aligned}
$$

- Where do you find the variances of the forecast errors of the first, second, and third variable?
- What is the contribution of u_{1t} to the variance of the forecast error of the first variable?
- What is the contribution of u_{1t} to the variance of the forecast error of the second variable?
- Decompose the variance of the forecast error of the third variable into the contributions of the innovations in the first, second, and third variable.

$$
\begin{aligned}
MSE(\hat{\mathbf{y}}_{t+2|t}) &= 1.79 \cdot \begin{bmatrix} 1.000 & 0.343 & 0.086 \\ 0.343 & 0.125 & 0.035 \\ 0.086 & 0.035 & 0.012 \end{bmatrix} + 1.78 \cdot \begin{bmatrix} 0.001 & 0.008 & 0.004 \\ 0.008 & 1.040 & 0.147 \\ 0.004 & 0.147 & 0.026 \end{bmatrix} \\
&+ 2.63 \cdot \begin{bmatrix} 0.001 & 0.002 & 0.002 \\ 0.002 & 0.002 & 0.003 \\ 0.002 & 0.003 & 1.004 \end{bmatrix} = \begin{bmatrix} 1.799 & 0.633 & 0.167 \\ 0.633 & 2.082 & 0.332 \\ 0.167 & 0.332 & 2.707 \end{bmatrix}
\end{aligned}
$$

- What is the contribution of u_{2t} to the variance of the forecast error of the first variable?
- What is the contribution of u_{3t} to the variance of the forecast error of the first variable?

- Decompose the variance of the forecast error of the second variable into the contributions of the innovations in the first, second, and third variable.

8. The analysis of the effects of Swiss Monetary policy delivers the following information shares with Cholesky ordering: $\tilde{m}, \tilde{r}, \tilde{p}, \tilde{y}$. Interpret!

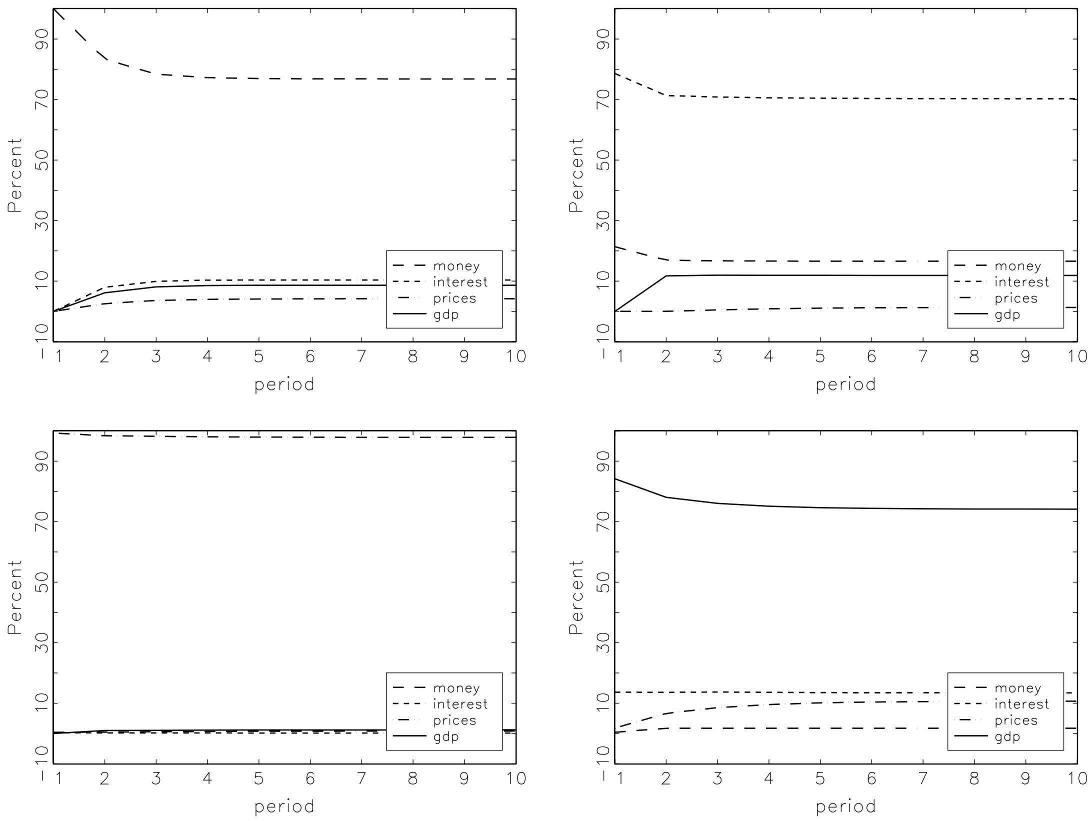


Figure 3: **Orthogonalized impulse-response functions** Upper left panel: Variance decomposition for money stock Upper right panel: Variance decomposition for interest rate. Lower left panel: Variance decomposition for consumer price index Lower right panel: Variance decomposition for GDP

Lecture 11

1. Consider the bivariate example

$$\begin{aligned} y_{1t} &= \gamma y_{2t} + u_{1t} \\ y_{2t} &= y_{2t-1} + u_{2t} \end{aligned}$$

   where u_{1t}, u_{2t} are independent Gaussian White Noise processes and y_{1t}, y_{2t} are integrated of order one ($I(1)$). Compute Δy_{1t} and Δy_{2t}.

   Given that the cointegrating relation $y_{1t} - \gamma y_{2t}$ exists why can't you simply consider a VAR in first differences to assess the dynamics of this system? Compare the VAR representation in first differences to the appropriate specification of the equilibrium correction model.

2. The equilibrium correction model cannot be estimated by OLS. Why?

3. How do you proceed to test the cointegrating relation $y_{1t} - \gamma y_{2t}$ for stationary?

4. What about the quality of the parameter estimates of the regression $y_{1t} = \beta_0 + \beta_1 y_{2t} + \varepsilon_t$ if (a) the cointegrating relation between y_{1t} and y_{2t} can be maintained, and (b) the cointegrating relation is rejected?

5. Explain the terms cointegrating rank, cointegrating vector, cointegrating relation, and base of a space of cointegrating vectors.

6. Consider the VECM

$$\Delta \mathbf{y}_t = \boldsymbol{\zeta}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\alpha} + \boldsymbol{\zeta}_0 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,$$

   where $\mathbf{y}_t$ is a (3×1) vector of the variables consumption, investment, and output, $\boldsymbol{\zeta}_1$ a (3×3) parameter matrix, $\boldsymbol{\alpha}$ a (3×1) vector of constants, and $\boldsymbol{\zeta}_0$ a (3×3) parameter matrix on which the restrictions $\boldsymbol{\zeta}_0 = -\mathbf{B}\mathbf{A}'$ are imposed, with $\mathbf{B}$ a (3×2) matrix of adjustment coefficients, and $\mathbf{A}$ a (3×2) cointegration matrix that is normalized as proposed by Phillips (see Hamilton p. 576).

   (i) Write in detail $\mathbf{A}$, and $\mathbf{y}_{t-1}$, and multiply out $\mathbf{A}'\mathbf{y}_{t-1}$. Interpret your result.

   (ii) Write in detail $\mathbf{B}$, and multiply out $\mathbf{B}\mathbf{A}'\mathbf{y}_{t-1}$. Interpret your result.

   (iii) Interpret the likelihood ratio statistic

$$LR = 2[L(\hat{\theta}) - L(\tilde{\theta})] = 1.2,$$

   where $L(\hat{\theta})$ denotes the value of the log-likelihood function at the unrestricted estimates $\boldsymbol{\zeta}_0$, and $L(\tilde{\theta})$ the value of the log-likelihood function at the restricted estimates $\boldsymbol{\zeta}_0 = -\mathbf{B}\mathbf{A}'$. The critical value associated with a significance level of 5% is $\chi^2_{0.95}(1) = 3.84$.

7. Let $\mathbf{y}_t$ be an I(1) process, and $\mathbb{E}[\Delta \mathbf{y}_t] = \boldsymbol{\delta}$, and $\mathbf{u}_t = \Delta \mathbf{y}_t - \boldsymbol{\delta} = \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \ldots$ with $\sum_{j=1}^{\infty} j \cdot |\boldsymbol{\Psi}_j| < \infty$ (which implies $\sum_{j=0}^{\infty} |\boldsymbol{\Psi}_j| < \infty$), then we have

$$\mathbf{y}_t = \mathbf{y}_0 - \boldsymbol{\eta}_0 + \boldsymbol{\Psi}(1)(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \ldots + \boldsymbol{\varepsilon}_t) + \boldsymbol{\eta}_t + \boldsymbol{\delta} \cdot t$$

with $\boldsymbol{\eta}_t$ an I(0) process.

Which relations of $\mathbf{A}$ and $\boldsymbol{\Psi}(1)$ on the one hand, and $\mathbf{A}'$ and $\boldsymbol{\delta}$ on the other hand have to hold if $\mathbf{A}'$ consists of h linearly independent cointegrating vectors, and why? Explain!

8. Let $\boldsymbol{\Psi}(L) = (\mathbf{I_n} + \boldsymbol{\Psi}_1 L + \boldsymbol{\Psi}_2 L^2 + \boldsymbol{\Psi}_3 L^3 + \ldots)$ with $\boldsymbol{\Psi}_j$ $(n \times n)$ matrices.

   (i) What is $\boldsymbol{\Psi}(L) \cdot \alpha$, with α a scalar?

   (ii) What is $\boldsymbol{\Psi}(1)$?

9. Assume $\mathbf{y}_t$ of dimension $(n \times 1)$ is an I(1) process. Why can we conclude that if the number of linearly independent cointegrating vectors equals n that $\mathbf{y}_t$ only consists of stationary variables?

10. Let $\mathbf{A}' = \begin{bmatrix} 1 & 0 & a_{13} \\ 0 & 1 & a_{23} \end{bmatrix}$. Would you conclude that there does not exist a cointegrating vector that involves all three variables, e.g. $[a_1^* \; a_2^* \; a_3^*]'$?

    If you think that such a cointegrating vector exists: How could you construct such a vector?

11. Phillip's method provides a way to construct a base of cointegrating vectors with a minimal number of free parameters. It hinges however on crucial assumptions. Which are these?

12. Suppose that the rows of the (3×4) matrix $\mathbf{A}' = \begin{bmatrix} 2 & 3 & 3 & 7 \\ 3 & 6 & 9 & 8 \\ 3 & 1 & 5 & 4 \end{bmatrix}$ form a basis for the space of cointegrating vectors. Derive Phillip's triangular representation of $\mathbf{A}'$ alongside Hamilton's recipe (pp. 576 f.) by hand, and write a Gauss procedure that does so for any $(h \times n)$ matrix $\mathbf{A}'$.

13. Let $\mathbf{A}' = \begin{bmatrix} 1 & 0 & a_{13} & 0 \\ 0 & 1 & a_{23} & 0 \end{bmatrix}$. What would you conclude regarding the role of y_{4t} in the system?