THE UNIVERSITY OF CHICAGO
Booth School of Business
Business 41914, Spring Quarter 2009, Mr. Ruey S. Tsay

Lecture 2: Vector ARMA Models

Reference: Chapter 14 of Pena, Tiao and Tsay (2001) or Chapter 8 of Tsay (2005) or Chapters 2-5 of Lutkepohl (2005).

1 Introduction

Let $\boldsymbol{z}_t = (z_{1t}, \cdots, z_{kt})'$ be a k-dimensional time series observed at equally spaced time points. The objectives of studying such a process include:

- To understand the dynamic relationships among the series z_{it}.
- To improve accuracy of forecasts.

Some basic concepts:

- Stationarity
  - Strict stationarity: distributions are time-invariant
  - Weak stionarity: the first two moments are time-invariant

  Unless stated otherwise, stationarity denotes weak stationarity in this course.

- Linearity: $\boldsymbol{z}_t$ is a linear function of independent and identically distributed random vectors, i.e.

$$\boldsymbol{z}_t = \boldsymbol{c} + \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{a}_{t-i}, \tag{1}$$

  where $\boldsymbol{c}$ is a constant, $\boldsymbol{\psi}_0 = \boldsymbol{I}$, $\{\boldsymbol{a}_j\}$ is a sequence of *iid* random vectors with mean zero and positive-definite covariance matrix $\boldsymbol{\Sigma}_a$. z_{it}'s are the components of $\boldsymbol{z}_t$.

  Remark: One can assume that $\boldsymbol{\Sigma}_a$ is a diagonal matrix provided that $\boldsymbol{\psi}_0$ would then be a lower triangular matrix with unity diagonal elements. Why?

  Remark: Wold decomposition states that any purely stochastic stationary process can be written as a linear combination as Eq.(1), but $\{\boldsymbol{a}_t\}$ is a sequence of uncorrelated series. We start with linear models for simplicity.

  For Eq. (1) to be meaningful, the coefficient matrices must satisfy

$$\sum_{i=1}^{\infty} \|\boldsymbol{\psi}_i\| < \infty,$$

  where $\|\boldsymbol{A}\|$ denotes a norm of matrix $\boldsymbol{A}$, e.g. $\|\boldsymbol{A}\| = \sqrt{tr(\boldsymbol{A}\boldsymbol{A}')}$. We shall discuss stationarity condition later. Here it suffices to say that for a linear stationary time series $\boldsymbol{z}_t$, $\boldsymbol{\psi}_i \to \boldsymbol{0}$ as $i \to \infty$.

- Invertibility: $\boldsymbol{z}_t$ is said to be invertible if

$$\boldsymbol{z}_t = \boldsymbol{a}_t + \sum_{j=1}^{\infty} \boldsymbol{\pi}_j \boldsymbol{z}_{t-j}.$$

  That is, $\boldsymbol{z}_t$ can be written as a linear combination of its past values and $\boldsymbol{a}_t$. Similarly to weak stationarity, for an invertible series $\boldsymbol{z}_t$, $\boldsymbol{\pi}_i \to \boldsymbol{0}$ as $i \to \infty$. We shall discuss invertibility condition later.

- For a stationary series $\boldsymbol{z}_t$, define

$$\begin{aligned} \boldsymbol{\mu}_z &= E(\boldsymbol{z}_t) \\ \boldsymbol{\Gamma}_\ell &= \text{Cov}(\boldsymbol{z}_t, \boldsymbol{z}_{t-\ell}) = E[(\boldsymbol{z}_t - \boldsymbol{\mu}_z)(\boldsymbol{z}_{t-\ell} - \boldsymbol{\mu}_z)']. \end{aligned}$$

  In particular, $\boldsymbol{\Gamma}_0$ is the covariance matrix of $\boldsymbol{z}_t$.

  Property: $\boldsymbol{\Gamma}_{-\ell} = \boldsymbol{\Gamma}_\ell'$ (proof?)

- For a stationary series $\boldsymbol{z}_t$, the cross-correlation matrix $\boldsymbol{\rho}_\ell$ is defined as

$$\boldsymbol{\rho}_\ell = \boldsymbol{D}^{-1} \boldsymbol{\Gamma}_\ell \boldsymbol{D}^{-1},$$

  where $\boldsymbol{D}$ = diagonal matrix consisting of standard deviations of the components of $\boldsymbol{z}_t$. Obviously, $\boldsymbol{\rho}_0$ is symmetric, but $\boldsymbol{\rho}_\ell$ is in general not symmetric for $\ell \neq 0$.

- What is the meaning of the (i, j)th element of $\boldsymbol{\rho}_\ell$ or $\boldsymbol{\Gamma}_\ell$?

  Answer: Based on the definition given in this lecture, the (i, j)th element of $\boldsymbol{\rho}_\ell$ (or $\boldsymbol{\Gamma}_\ell$) denotes the (unconditional) linear dependence of z_{it} on $z_{j,t-\ell}$.

1.1 Sample cross-correlation matrices

Given the sample $\{\boldsymbol{z}_t\}_{t=1}^n$, we construct the sample mean and covariance matrix as

$$\hat{\boldsymbol{\mu}}_z = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{z}_t, \quad \hat{\boldsymbol{\Gamma}}_0 = \frac{1}{n-1} \sum_{t=1}^{n} (\boldsymbol{z}_t - \hat{\boldsymbol{\mu}}_z)(\boldsymbol{z}_t - \hat{\boldsymbol{\mu}}_z)'. \tag{2}$$

These sample quantities are estimates of $\boldsymbol{\mu}_z$ and $\boldsymbol{\Gamma}_0$, respectively. The lag-ℓ sample cross-covariance matrix is defined as

$$\hat{\boldsymbol{\Gamma}}_\ell = \frac{1}{n-1} \sum_{t=\ell+1}^{n} (\boldsymbol{z}_t - \hat{\boldsymbol{\mu}}_z)(\boldsymbol{z}_{t-\ell} - \hat{\boldsymbol{\mu}}_z)'.$$

The lag-ℓ sample cross-correlation matrix (CCM) is then

$$\hat{\boldsymbol{\rho}}_\ell = \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{\Gamma}}_\ell \boldsymbol{D}^{-1},$$

where $\hat{\boldsymbol{D}} = \text{diag}\{\hat{\Gamma}_{0,11}^{1/2}, \cdots, \hat{\Gamma}_{0,kk}^{1/2}\}$, where $\hat{\Gamma}_{0,ii}$ is the (i,i)th element of $\hat{\boldsymbol{\Gamma}}_0$.
If $\boldsymbol{z}_t$ is a stationary linear process and $\boldsymbol{a}_t$ is normally distributed, then $\hat{\boldsymbol{\rho}}_\ell$ is a consistent estimate of $\boldsymbol{\rho}_\ell$. The asymptotic covariance matrix of elements of $\hat{\boldsymbol{\rho}}_\ell$ are complicated in general. See Chapter 10 of Box, Jenkins, and Reinsel (1994). If $\boldsymbol{z}_t$ is a white noise series, then the asymptitic variance of elements of $\hat{\boldsymbol{\rho}}_\ell$ is $1/n$, where n is the sample size.
In practice, we are often interested in testing the hypothesis that $H_o : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = \boldsymbol{0}$ vs $H_a : \boldsymbol{\rho}_i \neq \boldsymbol{0}$ for some $1 \leq i \leq m$, where m is a positive integer. Sample cross-correlation matrices can be used to generalize the univariate Ljung-Box test statistic as

$$Q_k(m) = n^2 \sum_{\ell=1}^{m} \frac{1}{n-\ell} tr(\hat{\boldsymbol{\Gamma}}_\ell' \hat{\boldsymbol{\Gamma}}_0^{-1} \hat{\boldsymbol{\Gamma}}_\ell \hat{\boldsymbol{\Gamma}}_0^{-1}), \tag{3}$$

where $tr(\boldsymbol{A})$ is the trace of the matrix $\boldsymbol{A}$. This is called the *multivariate Portmanteau test*. See Hosking (1980, JASA and 1981, JRSSB) and Li and McLeod (1981, JRSSB). This test can be re-written as

$$Q_k(m) = n^2 \sum_{\ell=1}^{m} \frac{1}{n-\ell} \boldsymbol{b}_\ell' (\hat{\boldsymbol{\rho}}_0^{-1} \otimes \hat{\boldsymbol{\rho}}_0^{-1}) \boldsymbol{b}_\ell.$$

where $\boldsymbol{b}_\ell = \text{vec}(\hat{\boldsymbol{\rho}}_\ell')$ and $\otimes$ is the Kronecker product. Here $\text{vec}(H)$ denotes the vectorization (or stacking) operation of the matrix $\boldsymbol{H}$. [The result can be obtained by using properties 7 and 8 of Lecture 1 concerning trace, vec, and Kronecker product.]
Under the null hypothesis of no serial or cross correlations and some regularity conditions, $Q_k(m)$ is asymptotically distributed as $\chi^2_{mk^2}$, i.e., chi-square with mk^2 degrees of freedom.

Remark: Strickly speaking, the test statistic of Li and McLeod (1981) is

$$Q_k^*(m) = n \sum_{\ell=1}^{m} \boldsymbol{b}_\ell' (\hat{\boldsymbol{\rho}}_0^{-1} \otimes \hat{\boldsymbol{\rho}}_0^{-1}) \boldsymbol{b}_\ell + \frac{k^2 m(m+1)}{2n},$$

which is asymptotically equivalent to $Q_k(m)$.

Remark: A simple R function to compute the $Q_k(m)$ statistic is given in the course web.

```
> x=matrix(rnorm(300),100,3) <== generate a 3-dim white noise series with 100 points.
> dim(x)
[1] 100   3
> var(x)
            [,1]        [,2]        [,3]
[1,]  0.86673283  0.05930379 -0.06466190
[2,]  0.05930379  1.25780323 -0.09803338
[3,] -0.06466190 -0.09803338  1.07372921

> source("mq.R")  <== Load the mq program.

> mq(x,10)
[1] "m,       Q(m) and  p-value:"
```

```
[1]  1.0000000 1.9463234 0.9922882
[1]  2.0000000 8.4609897 0.9709544
[1]  3.0000000 23.0220796  0.6838257
[1]  4.0000000 30.968895  0.706579
[1]  5.0000000 45.5800199  0.4478282
[1]  6.0000000 51.9960060  0.5520596
[1]  7.0000000 66.3326664  0.3627951
[1]  8.0000000 78.7491262  0.2739441
[1]  9.0000000 85.4172943  0.3471171
[1] 10.0000000 90.7172298  0.4589808
>
```

2 Vector AR Models

The process $\boldsymbol{z}_t$ follows a vector autoregressive (VAR) model of order p if

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{t-i} + \boldsymbol{a}_t, \qquad (4)$$

where $\boldsymbol{\phi}_0$ is a constant vector and $\boldsymbol{\phi}_i$ are $k \times k$ matrices for $i > 0$, and $\boldsymbol{\phi}_p \neq \mathbf{0}$. Using the backshift operator, the model becomes $\boldsymbol{\phi}(B)\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t$, where $\boldsymbol{\phi}(B) = \boldsymbol{I} - \sum_{i=1}^{p} \boldsymbol{\phi}_i B^i$ is a matrix polynomial of degree p. For simplicity, we shall start with the 2-dimensional case.

2.1 VAR(1) model

Consider the bivariate VAR(1) model

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_{t-1} + \boldsymbol{a}_t.$$

The model can be written specifically as

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}, \qquad (5)$$

or equivalently,

$$\begin{aligned} z_{1t} &= \phi_{10} + \phi_{1,11} z_{1,t-1} + \phi_{1,12} z_{2,t-1} + a_{1t} \\ z_{2t} &= \phi_{20} + \phi_{1,21} z_{1,t-1} + \phi_{1,22} z_{2,t-1} + a_{2t}. \end{aligned}$$

Thus, $\phi_{1,12}$ is the impact of $z_{2,t-1}$ on z_{1t} in the presence of $z_{1,t-1}$. Other parameters can be interpreted in a similar manner.

Some basic concepts:

- Relation to transfer function model

From the model in Eq. (5), if $\phi_{1,12} = 0$, but $\phi_{1,21} \neq 0$, then we have

$$z_{1t} = \phi_{10} + \phi_{1,11} z_{1,t-1} + a_{1t}, \quad (6)$$

$$z_{2t} = \phi_{20} + \phi_{1,21} z_{1,t-1} + \phi_{1,22} z_{2,t-1} + a_{2t}. \quad (7)$$

This structure shows that z_{1t} does not depend on the past of z_{2t}, but z_{2t} depends on the past of z_{1t}. Consequently, we have a unidirectional relationship with z_{1t} acting as the input variable and z_{2t} the output variable. The two series have a simple transfer function relationship.

However, the model representation in Eqs. (6)-(7) is in general not a transfer function model because the two innovations a_{1t} and a_{2t} might be correlated. Recall that in a transfer function model the input variables should be independent of the disturbance term N_t. To obtain a transfer function model, we need to orthogonalize the two innovations. Specifically, consider the simple linear regression

$$a_{2t} = \beta a_{1t} + \epsilon_t,$$

where $\beta = \text{cov}(a_{1t}, a_{2t})/\text{var}(a_{1t})$ and a_{1t} and ϵ_t are uncorrelated. Eq. (7) can then be written as

$$(1 - \phi_{1,22} B) z_{2t} = (\phi_{20} - \beta\phi_{10}) + [\beta + (\phi_{1,21} - \beta\phi_{1,11})B] z_{1t} + \epsilon_t,$$

which is a transfer function model.

By the same argument, if $\phi_{1,21} = 0$, but $\phi_{1,12} \neq 0$, then we also have a transfer function model with z_{2t} as the input variable and z_{1t} as the output variable. If $\phi_{1,12} = \phi_{1,21} = 0$, then the two series are not dynamically dependent.

- Stationarity condition

  The stationarity condition of a VAR(1) model is that all eigenvalues of $\boldsymbol{\phi}_1$ are less than one in modulus. To see this, we may assume that $\boldsymbol{\phi}_0 = \mathbf{0}$ and the time series starts at $t = 0$ with initial value $\boldsymbol{z}_0$. Then, it is easy to see that

$$\boldsymbol{z}_t = \sum_{i=0}^{t-1} \boldsymbol{\phi}_1^i \boldsymbol{a}_{t-i} + \boldsymbol{\phi}_1^t \boldsymbol{z}_0.$$

  If $\boldsymbol{z}_t$ is stationary, then $\boldsymbol{z}_t$ does not depend on $\boldsymbol{z}_0$ as $t \to \infty$. Consequently, $\boldsymbol{\phi}_1^i \to \mathbf{0}$ as $i \to \infty$. Thus, all eigenvalues of $\boldsymbol{\phi}_1$ should be less than one in absolute value.

- The model is invertible. This follows directly from the model and definition of invertibility.

- Moment equations under stationarity

  Let $\tilde{\boldsymbol{z}}_t = \boldsymbol{z}_t - \boldsymbol{\mu}$, where $\boldsymbol{\mu} = E(\boldsymbol{z}_t)$. The VAR(1) model can be written as

$$\tilde{\boldsymbol{z}}_t = \boldsymbol{\phi}_1 \tilde{\boldsymbol{z}}_{t-1} + \boldsymbol{a}_t.$$

  Post-multiplying the model by $\tilde{\boldsymbol{z}}'_{t-\ell}$ and taking expectation, we obtain the moment equation

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{\ell-1}, \quad \ell > 0.$$

  Consequently, for a stationary VAR(1) model, we have $\boldsymbol{\Gamma}_\ell = \boldsymbol{\phi}_1^\ell \boldsymbol{\Gamma}_0$ for $\ell > 0$. In particular, for $\ell = 1$,

$$\boldsymbol{\Gamma}_1 = \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_0.$$

Since $\mathbf{\Gamma}_0$ is nonsingular, we have $\boldsymbol{\phi}_1 = \mathbf{\Gamma}_1\mathbf{\Gamma}_0^{-1}$. Pre- and post multiplying the prior moment equation by $\boldsymbol{D}^{-1}$, we further obtain

$$\boldsymbol{\rho}_\ell = \boldsymbol{\phi}_1^*\boldsymbol{\rho}_{\ell-1},$$

for $\ell > 0$, where $\boldsymbol{\phi}_1^* = \boldsymbol{D}^{-1}\boldsymbol{\phi}_1\boldsymbol{D}$ with $\boldsymbol{D}$ being the diagonal matrix of standard deviations of the component series.

- Univariate model for the components

  Univariate ARMA(2,1) model. In general, each component follows a univariate ARMA($k, k-1$) model.

2.2 VAR(p) model

The model is

$$\boldsymbol{\phi}(B)\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t.$$

Some basic properties:

- Relation to transfer function model
- Stationarity condition
- The model is invertible
- Moment equations
- Univariate model for the components

2.3 Estimation

Two methods are commonly used. They are the least squares method and likelihood method. Under normality, the two methods are asymptotically equivalent.
Suppose that the sample $\{\boldsymbol{z}_t\}_{t=1}^n$ is available and a VAR(p) is entertained. That is,

$$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \phi_1\boldsymbol{z}_{t-1} + \cdots + +\boldsymbol{\phi}_p\boldsymbol{z}_{t-p} + \boldsymbol{a}_t, \quad t = p+1, \cdots, n.$$

Least squares (LS) method

The VAR(p) model can be written as

$$\boldsymbol{z}_t' = \boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{a}_t',$$

where $\boldsymbol{x}_t = (1, \boldsymbol{z}_{t-1}', \cdots, \boldsymbol{z}_{t-p}')'$ is a $(kp+1)$-dimensional vector and $\boldsymbol{\beta}' = [\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p]$ is a $k \times (kp+1)$ matrix. The least squares estimate of $\boldsymbol{\beta}$ is

$$\tilde{\beta} = \left[\sum_{t=p+1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t'\right]^{-1} \sum_{t=p+1}^{n} \boldsymbol{x}_t\boldsymbol{z}_t'.$$

The LS residual is

$$\tilde{a}_t = \boldsymbol{z}_t - \sum_{i=1}^{p} \tilde{\boldsymbol{\phi}}_i\boldsymbol{z}_{t-i}, \quad t = p+1, \cdots, n.$$

The LS estimate of $\mathbf{\Sigma}$ is

$$\tilde{\mathbf{\Sigma}} = \frac{1}{n-(k+1)p-1} \sum_{t=p+1}^{n} \tilde{\boldsymbol{a}}_t \tilde{\boldsymbol{a}}_t'.$$

For a stationary VAR(p) model with independent error terms $\boldsymbol{a}_t$, it can be shown that the least squares estimate $\tilde{\boldsymbol{\phi}}$ is consistent. Furthermore, let $\tilde{\boldsymbol{b}} = \text{vec}(\tilde{\boldsymbol{\beta}})$, where vec($\boldsymbol{A}$) is the column stacking operator of matrix $\boldsymbol{A}$. Then $\tilde{\boldsymbol{b}}$ is asymptotically normal with mean vec($\boldsymbol{\beta}$) and covariance matrix

$$\text{Cov}(\tilde{\boldsymbol{\beta}}) = \tilde{\mathbf{\Sigma}} \otimes \left(\sum_{t=p+1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t' \right)^{-1},$$

where $\otimes$ denotes the Kronecker product; see Lütkepohl (1991) or Result 7.10 of Johnson and Wichern (2002, *Applied Multivariate Statistical Analysis*).

<u>Maximum likelihood method</u>

The coefficient estimates are the same as those of LS estimates. However, the estimate of $\mathbf{\Sigma}$ is

$$\hat{\mathbf{\Sigma}} = \frac{1}{n-p} \sum_{t=p+1}^{n} \tilde{\boldsymbol{a}}_t \tilde{\boldsymbol{a}}_t'.$$

2.4 Order determination

Two basic approaches:

- Sequential Chi-square test

  Let P be a positive integer, denoting the maximum order entertained.

  For $\ell > 0$, consider the hypothesis H_o : VAR(ℓ) vs H_a : VAR($\ell-1$). This is to test $H_o : \boldsymbol{\phi}_\ell = \mathbf{0}$ vs $H_a : \boldsymbol{\phi}_\ell \neq \mathbf{0}$ in the autoregression

  $$\boldsymbol{z}_t = \boldsymbol{\phi}_0 + \phi_1 \boldsymbol{z}_{t-1} + \cdots + \boldsymbol{\phi}_{\ell-1} \boldsymbol{z}_{t-\ell+1} + \boldsymbol{\phi}_\ell \boldsymbol{z}_{t-\ell} + \boldsymbol{a}_t, \quad t = P+1, \cdots, n$$

  How to proceed? Likelihood ratio test in multivariate linear regression.

  If $\boldsymbol{a}_t$ is Gaussian, then we can use the LR test; see Result 7.1 on page 393 of Johnson and Wichern (2002, *Applied Multivariate Statistical Analysis*).

  Let $\hat{\mathbf{\Sigma}}_i$ be the maximum likelihood estimate of $\mathbf{\Sigma}$ of fitting a VAR(i) model to $\boldsymbol{z}_t$. Then, the LR test statistic is

  $$M(\ell) = -(n - P - \frac{3}{2} - \ell k) \ln \left[\frac{|\hat{\mathbf{\Sigma}}_\ell|}{|\hat{\mathbf{\Sigma}}_{\ell-1}|} \right], \tag{8}$$

  where $\hat{\mathbf{\Sigma}}_i$ are ML estimate of $\mathbf{\Sigma}$ using $t = P+1, \cdots, n$. For large n, $M(\ell)$ is approximately a chi-square distribution with k^2 degrees of freedom.

- Information criteria: AIC, BIC, etc.

  Three criterion functions are commonly used to determine VAR order. They are

  $$\begin{aligned} \text{AIC}(\ell) &= \ln |\hat{\mathbf{\Sigma}}_\ell| + \frac{2}{n} \ell k^2 \\ \text{BIC}(\ell) &= \ln |\hat{\mathbf{\Sigma}}_\ell| + \frac{\ln(n)}{n} \ell k^2 \\ \text{HQ}(\ell) &= \ln |\hat{\mathbf{\Sigma}}_\ell| + \frac{2 \ln[\ln(n)]}{n} \ell k^2, \end{aligned}$$

where $\hat{\mathbf{\Sigma}}_\ell$ is the MLE of $\mathbf{\Sigma}$ under normality. Suppose that the true model is VAR(p) with $p < \infty$. Asymptotically, AIC overestimates the true order with positive probability whereas BIC and HQ criteria estimate the order consistently.

2.5 Model checking

Residual plots and the $Q_k(m)$ statistics are often used as tools to check a fitted VAR model. If the residuals exhibit any inadequency, the model should be refined.

2.6 Forecasting

The minimum mean squared forecast error is the criterion commonly used to produce point forecasts of a vector time series. That is, at the forecast origin h, the forecast $z_h(\ell)$ is obtained by

$$z_h(\ell) = \text{minarg}_g E[(z_{h+\ell} - g)^2 | F_h]$$

where F_h is the information available at time h. Similarly to the univariate case, the point forecasts turn out to be the conditional expectation as

$$\boldsymbol{z}_h(\ell) = E(\boldsymbol{z}_{h+\ell} | F_h),$$

where F_h denotes the information available at the forecast origin h, i.e. F_h is the σ-field generated by $\{\boldsymbol{z}_h, \boldsymbol{z}_{h-1}, \cdots\}$ and the model. It turns out that the forecasts can be obtained recursively as ℓ increases. Specifically,

$$\begin{aligned}
\boldsymbol{z}_h(1) &= \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\phi}_i \boldsymbol{z}_{h+1-i} \\
\boldsymbol{z}_h(2) &= \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{z}_h(1) + \sum_{j=2}^{p} \boldsymbol{\phi}_j \boldsymbol{z}_{h+2-j} \\
\vdots &= \vdots
\end{aligned}$$

The associated forecast errors can also be obtained recursively. We shall return to this point later. For a stationary VAR model, the forecast $\boldsymbol{z}_h(\ell)$ converges to $\boldsymbol{\mu}_z$ as $\ell \to \infty$. This is the mean-reverting property of a stationary time series. In fact, one can show that

$$\boldsymbol{\phi}(B)\boldsymbol{z}_h(\ell) = \mathbf{0},$$

for $\ell > p$.

2.7 Some simulated examples

We consider two simulated bi-variate time series.

- Example 1: bivariate MA(1) series with 250 observations
- Example 2: bivariate AR(1) series with 150 observations

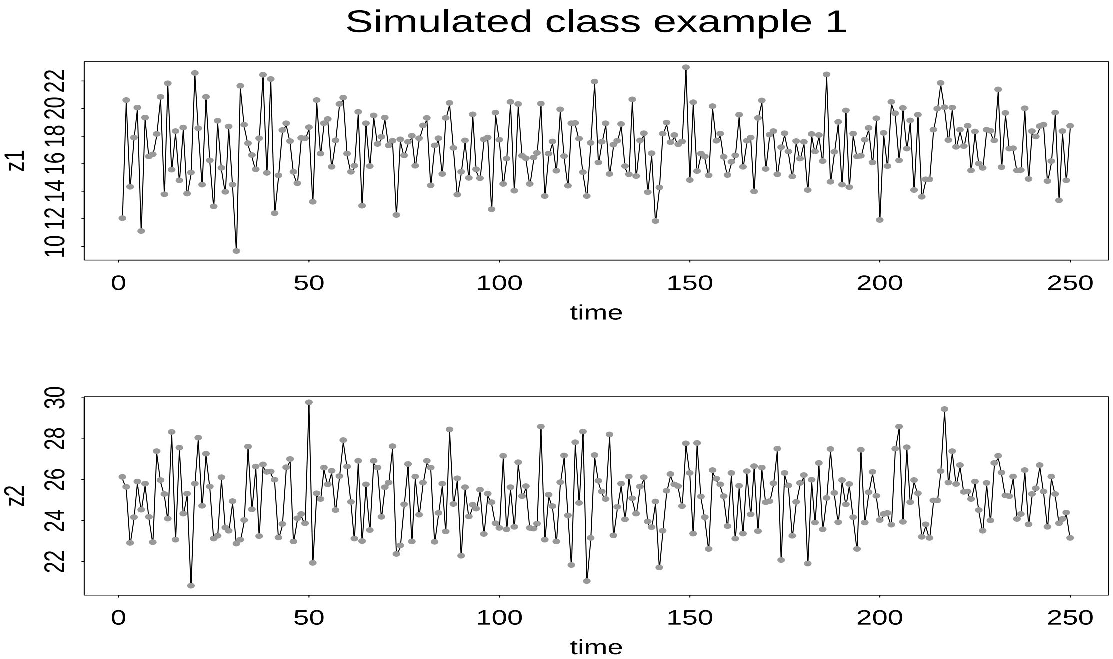

Figure 1: Time Plots of Simulated Bivariate MA(1) Series

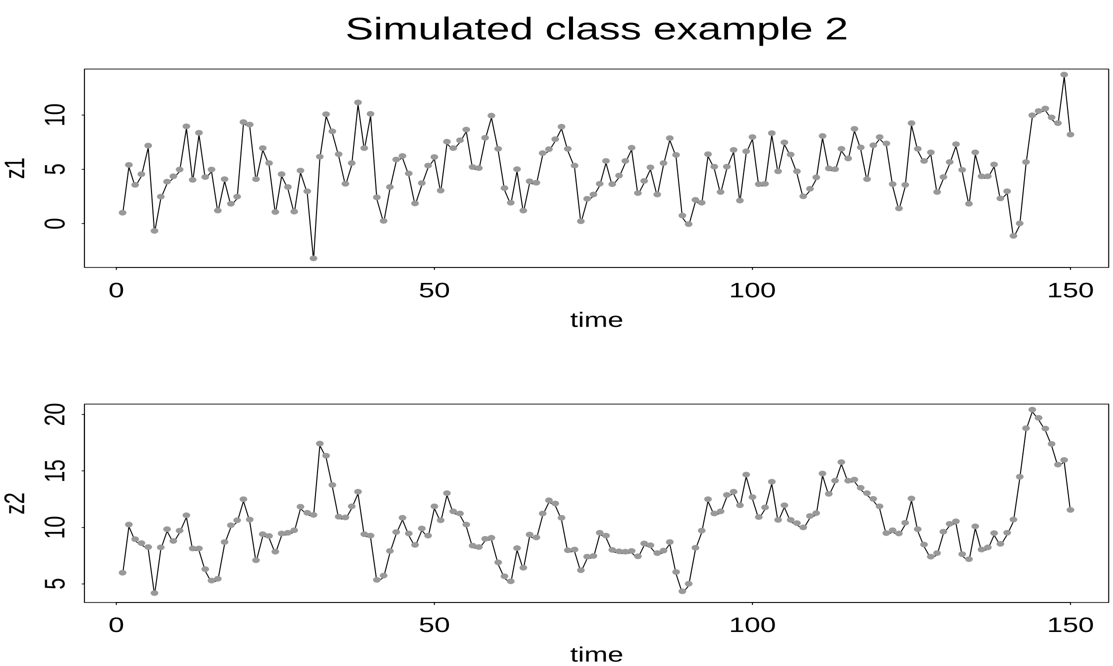

Figure 2: Time Plots of Simulated Bivariate AR(1) Series

SCA demonstration: output is edited to simplify the handout.

```
 --
input z1,z2. file 'clsma1.dat'

 Z1      , A  250  BY   1 VARIABLE, IS STORED IN THE WORKSPACE
 Z2      , A  250  BY   1 VARIABLE, IS STORED IN THE WORKSPACE
 --
miden z1,z2. maxl 12.

 TIME PERIOD ANALYZED . . . . . . . . . . . .  1  TO   250
 EFFECTIVE NUMBER OF OBSERVATIONS (NOBE). . .          250

 SERIES   NAME           MEAN       STD. ERROR

    1    Z1             17.0664      2.3334
    2    Z2             25.0393      1.5875

 NOTE: THE APPROX. STD. ERROR FOR THE ESTIMATED CORRELATIONS BELOW
       IS (1/NOBE**.5) =    0.06325

 SAMPLE CORRELATION MATRIX OF THE SERIES
   1.00
   0.30  1.00

 SUMMARIES OF CROSS CORRELATION MATRICES USING +,-,., WHERE
      + DENOTES A VALUE GREATER THAN 2/SQRT(NOBE)
      - DENOTES A VALUE LESS THAN -2/SQRT(NOBE)
      . DENOTES A NON-SIGNIFICANT VALUE BASED ON THE ABOVE CRITERION

 BEHAVIOR OF VALUES IN (I,J)TH POSITION OF CROSS CORRELATION MATRIX OVER
 ALL OUTPUTTED LAGS WHEN SERIES J LEADS SERIES I

          1              2
  1   -...........   -.....-.....

  2   +...........   -......-....

 CROSS CORRELATION MATRICES IN TERMS OF +,-,.

 LAGS  1 THROUGH  6
      - -       . .       . .       . .       . .       . .
      + -       . .       . .       . .       . .       . .

 LAGS  7 THROUGH 12
```

```
     . -      . .      . .      . .      . .      . .
     . .      . -      . .      . .      . .      . .
--
```

```
miden z1,z2. no ccm. arfits 1 to 9.

 TIME PERIOD ANALYZED . . . . . . . . . . . . .  1  TO   250
 EFFECTIVE NUMBER OF OBSERVATIONS (NOBE). . .          250

 SERIES   NAME           MEAN       STD. ERROR

    1    Z1            17.0664      2.3334
    2    Z2            25.0393      1.5875

 DETERMINANT OF S(0) = 0.121907E+02

 NOTE: S(0) IS THE SAMPLE COVARIANCE MATRIX OF W(MAXLAG+1),...,W(NOBE)

 ==========  STEPWISE AUTOREGRESSION SUMMARY  ==========
 ---------------------------------------------------------------------
     I RESIDUAL I EIGENVAL.I CHI-SQ  I          I SIGNIFICANCE
 LAG I VARIANCESI OF SIGMA I  TEST   I   AIC    I OF PARTIAL AR COEFF.
 ----+----------+----------+---------+----------+---------------------
   1 I .479E+01 I .137E+01 I  123.65 I    2.012 I - -
     I .186E+01 I .529E+01 I         I          I + -
 ----+----------+----------+---------+----------+---------------------
   2 I .476E+01 I .997E+00 I   80.66 I    1.702 I . .
     I .139E+01 I .516E+01 I         I          I + -
 ----+----------+----------+---------+----------+---------------------
   3 I .465E+01 I .908E+00 I   30.75 I    1.602 I + .
     I .122E+01 I .497E+01 I         I          I + -
 ----+----------+----------+---------+----------+---------------------
   4 I .465E+01 I .779E+00 I   36.47 I    1.476 I . .
     I .108E+01 I .494E+01 I         I          I + .
 ----+----------+----------+---------+----------+---------------------
   5 I .463E+01 I .726E+00 I   16.54 I    1.436 I . .
     I .104E+01 I .494E+01 I         I          I + .
 ----+----------+----------+---------+----------+---------------------
   6 I .455E+01 I .701E+00 I   12.76 I    1.412 I . .
     I .980E+00 I .483E+01 I         I          I + .
 ----+----------+----------+---------+----------+---------------------
   7 I .440E+01 I .670E+00 I   18.43 I    1.362 I . -
     I .931E+00 I .466E+01 I         I          I + -
 ----+----------+----------+---------+----------+---------------------
   8 I .433E+01 I .661E+00 I    7.30 I    1.362 I . .
     I .906E+00 I .457E+01 I         I          I . -
```

```
----+----------+----------+---------+----------+---------------------
   9 I .432E+01 I .654E+00 I    2.86 I    1.381 I . .
     I .895E+00 I .456E+01 I         I          I . .
----------------------------------------------------------------------
NOTE:  CHI-SQUARED CRITICAL VALUES WITH   4  DEGREES OF FREEDOM ARE

             5 PERCENT:   9.5     1 PERCENT:  13.3
--
input y1,y2. file 'clsar1.dat'
--
miden y1,y2. maxl 12. output level(deta)

TIME PERIOD ANALYZED . . . . . . . . . . . . .  1  TO   150
EFFECTIVE NUMBER OF OBSERVATIONS (NOBE). . .            150

SERIES   NAME           MEAN       STD. ERROR

   1    Y1              4.9624      2.7685
   2    Y2             10.0616      3.0082

NOTE: THE APPROX. STD. ERROR FOR THE ESTIMATED CORRELATIONS BELOW
      IS (1/NOBE**.5) =    0.08165

SAMPLE CORRELATION MATRIX OF THE SERIES
  1.00
  0.56  1.00

SAMPLE CROSS CORRELATION MATRICES FOR THE ORIGINAL SERIES.
THE (I,J) ELEMENT OF THE LAG L MATRIX IS THE ESTIMATE OF
THE LAG L CROSS CORRELATION WHEN SERIES J LEADS SERIES I

LAG = 1
  0.39  0.54
  0.06  0.81
LAG = 2
  0.08  0.49
 -0.12  0.64
LAG = 3
  0.02  0.42
 -0.16  0.47
LAG = 4
 -0.08  0.29
 -0.19  0.32
LAG = 5
 -0.04  0.25
```

```
 -0.15  0.23
LAG = 6
  0.02  0.23
 -0.11  0.16
LAG = 7
 -0.19  0.10
 -0.19  0.05
LAG = 8
 -0.22  0.02
 -0.14 -0.01
LAG = 9
 -0.10 -0.02
 -0.06 -0.04
LAG =10
 -0.03 -0.05
  0.00 -0.06
LAG =11
  0.04 -0.01
  0.05 -0.03
LAG =12
  0.07  0.03
  0.06  0.00

SUMMARIES OF CROSS CORRELATION MATRICES USING +,-,., WHERE
     + DENOTES A VALUE GREATER THAN 2/SQRT(NOBE)
     - DENOTES A VALUE LESS THAN -2/SQRT(NOBE)
     . DENOTES A NON-SIGNIFICANT VALUE BASED ON THE ABOVE CRITERION

BEHAVIOR OF VALUES IN (I,J)TH POSITION OF CROSS CORRELATION MATRIX OVER
ALL OUTPUTTED LAGS WHEN SERIES J LEADS SERIES I

         1              2
 1   +.....--....   ++++++......

 2   ...-..-.....   +++++.......

CROSS CORRELATION MATRICES IN TERMS OF +,-,.

LAGS  1 THROUGH  6
     + +      . +      . +      . +      . +      . +
     . +      . +      . +      - +      . +      . .

LAGS  7 THROUGH 12
     - .      - .      . .      . .      . .      . .
     - .      . .      . .      . .      . .      . .
```

```
--
miden y1,y2. no ccm. arfits 1 to 9.

 TIME PERIOD ANALYZED . . . . . . . . . . . . .  1  TO   150
 EFFECTIVE NUMBER OF OBSERVATIONS (NOBE). . .          150

 SERIES   NAME           MEAN       STD. ERROR

    1    Y1              4.9624      2.7685
    2    Y2             10.0616      3.0082

 DETERMINANT OF S(0) = 0.497555E+02

 ==========  STEPWISE AUTOREGRESSION SUMMARY  ==========
 ----------------------------------------------------------------------
     I RESIDUAL I EIGENVAL.I CHI-SQ  I          I SIGNIFICANCE
 LAG I VARIANCESI OF SIGMA I  TEST   I   AIC    I OF PARTIAL AR COEFF.
 ----+----------+----------+---------+----------+----------------------
   1 I .524E+01 I .683E+00 I  353.19 I    1.392 I . +
     I .103E+01 I .558E+01 I         I          I - +
 ----+----------+----------+---------+----------+----------------------
   2 I .507E+01 I .671E+00 I    7.57 I    1.389 I . +
     I .977E+00 I .538E+01 I         I          I . +
 ----+----------+----------+---------+----------+----------------------
   3 I .500E+01 I .669E+00 I    2.11 I    1.427 I . .
     I .976E+00 I .530E+01 I         I          I . .
 ----+----------+----------+---------+----------+----------------------
   4 I .495E+01 I .649E+00 I    4.82 I    1.443 I . .
     I .968E+00 I .527E+01 I         I          I . .
 ----+----------+----------+---------+----------+----------------------
   5 I .492E+01 I .646E+00 I    1.68 I    1.484 I . .
     I .958E+00 I .523E+01 I         I          I . .
 ----+----------+----------+---------+----------+----------------------
   6 I .481E+01 I .644E+00 I    3.44 I    1.510 I . .
     I .942E+00 I .510E+01 I         I          I . .
 ----+----------+----------+---------+----------+----------------------
   7 I .457E+01 I .642E+00 I    7.15 I    1.506 I . .
     I .916E+00 I .484E+01 I         I          I . .
 ----+----------+----------+---------+----------+----------------------
   8 I .445E+01 I .629E+00 I    5.33 I    1.517 I . .
     I .904E+00 I .473E+01 I         I          I . .
 ----+----------+----------+---------+----------+----------------------
   9 I .444E+01 I .625E+00 I    1.12 I    1.561 I . .
     I .902E+00 I .472E+01 I         I          I . .
 ----------------------------------------------------------------------
```

```
 NOTE:  CHI-SQUARED CRITICAL VALUES WITH   4  DEGREES OF FREEDOM ARE

             5 PERCENT:   9.5     1 PERCENT:  13.3
 --
mtsm mar. series y1,y2. model (i-p1*b)series=c1+noise.

 SUMMARY FOR MULTIVARIATE ARMA MODEL --   MAR

 VARIABLE   DIFFERENCING

  Y1
  Y2

  PARAMETER     FACTOR     ORDER    CONSTRAINT

  1       C1   CONSTANT      0        CC1
  2       P1    REG AR       1        CP1
 --
mestim mar. hold resi(r1,r2)

 SUMMARY FOR THE MULTIVARIATE ARMA MODEL

 SERIES    NAME            MEAN         STD DEV    DIFFERENCE ORDER(S)

   1     Y1               4.9624         2.7685
   2     Y2              10.0616         3.0082

 NUMBER OF OBSERVATIONS =  150  (EFFECTIVE NUMBER = NOBE =  149)

 MODEL SPECIFICATION WITH PARAMETER VALUES

 PARAMETER              PARAMETER                 PARAMETER
  NUMBER               DESCRIPTION                  VALUE
 ----------    -----------------------------     ------------
     1                CONSTANT( 1)                  3.460013
     2                CONSTANT( 2)                  8.559218
     3          AUTOREGRESSIVE ( 1, 1, 1)           0.100000
     4          AUTOREGRESSIVE ( 1, 1, 2)           0.100000
     5          AUTOREGRESSIVE ( 1, 2, 1)           0.100000
     6          AUTOREGRESSIVE ( 1, 2, 2)           0.100000
 -----------------------
 ERROR COVARIANCE MATRIX
 -----------------------
              1            2
   1     6.359580
```

```
  2     3.263123    7.681480

 ITERATIONS TERMINATED DUE TO:
   RELATIVE CHANGE IN DETERMINANT OF COVARIANCE MATRIX .LE. 0.100E-03
   TOTAL NUMBER OF ITERATIONS IS    4

 FINAL MODEL SUMMARY WITH CONDITIONAL LIKELIHOOD PARAMETER ESTIMATES

 ----- CONSTANT VECTOR (STD ERROR) -----
     0.032  (    0.655  )
     1.878  (    0.300  )

 ----- PHI MATRICES -----
 ESTIMATES OF    PHI(  1 ) MATRIX AND SIGNIFICANCE
      .133     .428      . +
     -.609    1.116      - +

 STANDARD ERRORS
      .082     .075
      .037     .034
 -----------------------
 ERROR COVARIANCE MATRIX
 -----------------------
              1           2
  1     5.271816
  2     1.331015    1.105592
 --------------------------------------------
 REDUCED CORRELATION MATRIX OF THE PARAMETERS
 --------------------------------------------
         1     2     3     4     5     6
  1   1.00
  2    .55  1.00
  3    .     .    1.00
  4   -.81  -.45  -.55  1.00
  5    .     .     .55  -.31  1.00
  6   -.45  -.81  -.31   .55  -.55  1.00
 --
miden r1,r2. maxl 12. output level(deta)

 TIME PERIOD ANALYZED . . . . . . . . . . . .  2  TO   150
 EFFECTIVE NUMBER OF OBSERVATIONS (NOBE). . .          149

 SERIES   NAME           MEAN       STD. ERROR

   1    R1              0.0000      2.2960
```

```
   2    R2              0.0000      1.0515

NOTE: THE APPROX. STD. ERROR FOR THE ESTIMATED CORRELATIONS BELOW
      IS (1/NOBE**.5) =    0.08192

SAMPLE CORRELATION MATRIX OF THE SERIES
  1.00
  0.55  1.00

SAMPLE CROSS CORRELATION MATRICES FOR THE ORIGINAL SERIES.
THE (I,J) ELEMENT OF THE LAG L MATRIX IS THE ESTIMATE OF
THE LAG L CROSS CORRELATION WHEN SERIES J LEADS SERIES I

LAG = 1
 -0.02 -0.15
 -0.05 -0.17
LAG = 2
 -0.05  0.04
  0.01  0.04
LAG = 3
  0.09  0.15
  0.02 -0.02
LAG = 4
 -0.06 -0.03
 -0.09  0.04
LAG = 5
 -0.02 -0.03
 -0.04 -0.02
LAG = 6
  0.18  0.09
  0.14  0.13
LAG = 7
 -0.18 -0.19
 -0.12 -0.10
LAG = 8
 -0.15  0.01
 -0.12 -0.05
LAG = 9
  0.00  0.08
 -0.04  0.11
LAG =10
  0.01 -0.04
  0.04 -0.07
LAG =11
  0.01 -0.05
```

```
  0.04 -0.03
 LAG =12
  0.07  0.00
  0.06 -0.04

 SUMMARIES OF CROSS CORRELATION MATRICES USING +,-,.

 BEHAVIOR OF VALUES IN (I,J)TH POSITION OF CROSS CORRELATION MATRIX OVER
 ALL OUTPUTTED LAGS WHEN SERIES J LEADS SERIES I

           1              2
  1   .....+-.....   ......-.....

  2   ............   -...........

 CROSS CORRELATION MATRICES IN TERMS OF +,-,.

 LAGS  1 THROUGH  6
      . .      . .      . .      . .      . .      + .
      . -      . .      . .      . .      . .      . .

 LAGS  7 THROUGH 12
      - -      . .      . .      . .      . .      . .
      . .      . .      . .      . .      . .      . .
 --
mfore mar. nofts 6.
 ----------------------------------------
   6 FORECASTS, BEGINNING AT ORIGIN = 150
 ----------------------------------------
 SERIES:       Y1                  Y2

 TIME    FORECAST  STD ERR   FORECAST  STD ERR
  151      5.966    2.296      9.686    1.051
  152      4.970    2.392      9.057    1.621
  153      4.568    2.450      8.961    2.182
  154      4.473    2.545      9.099    2.564
  155      4.520    2.631      9.311    2.786
  156      4.617    2.688      9.519    2.902
 --
stop
```

3 Impulse Response Functions

In some applcations, it is of interest to study the impact of an innovation on a particular component of a vector time series. For instance, if $\boldsymbol{z}_t$ consists of the monthly income and expenditure of a household, it is then of interest to study the response of the expenditure on an increase in income of a particular month. Such a study is called the *impulse response function* or *multiplier analysis*. Consider a VAR(p) model. Without loss of generality, assume that $\boldsymbol{z}_t$ has zero mean and follows the model

$$\boldsymbol{\phi}(B)\boldsymbol{z}_t = \boldsymbol{a}_t,$$

where $\boldsymbol{\phi}(B) = \boldsymbol{I} - \boldsymbol{\phi}_1 B - \cdots - \boldsymbol{\phi}_p B^p$ and $\{\boldsymbol{a}_t\}$ is a white noise series with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}_a$. To study the impulse response function of $\boldsymbol{z}_t$, we express $\boldsymbol{z}_t$ as

$$\boldsymbol{z}_t = \boldsymbol{\psi}(B)\boldsymbol{a}_t = (\boldsymbol{I} + \boldsymbol{\psi}_1 B + \boldsymbol{\psi}_2 B^2 + \boldsymbol{\psi}_3 B^3)\boldsymbol{a}_t, \tag{9}$$

where the coefficient matrices $\boldsymbol{\psi}_i$ can be obtained by equating the coefficients of

$$\boldsymbol{I} = \boldsymbol{\phi}(B)\boldsymbol{\psi}(B),$$

because $[\boldsymbol{\phi}(B)]^{-1} = \boldsymbol{\psi}(B)$. It is easy to see that

$$\begin{aligned}
\boldsymbol{\psi}_1 &= \boldsymbol{\phi}_1, \\
\boldsymbol{\psi}_2 &= \boldsymbol{\phi}_1\boldsymbol{\psi}_1 + \boldsymbol{\phi}_2, \\
\boldsymbol{\psi}_3 &= \boldsymbol{\phi}_1\boldsymbol{\psi}_2 + \boldsymbol{\phi}_2\boldsymbol{\psi}_1 + \boldsymbol{\phi}_3, \\
\vdots &= \vdots \\
\boldsymbol{\psi}_p &= \boldsymbol{\phi}_1\boldsymbol{\psi}_{p-1} + \boldsymbol{\phi}_2\boldsymbol{\psi}_{p-2} + \cdots + \boldsymbol{\phi}_{p-1}\boldsymbol{\psi}_1 + \boldsymbol{\phi}_p, \\
\boldsymbol{\psi}_\ell &= \boldsymbol{\phi}_1\boldsymbol{\psi}_{\ell-1} + \boldsymbol{\phi}_2\boldsymbol{\psi}_{\ell-2} + \cdots + \boldsymbol{\phi}_p\boldsymbol{\psi}_{\ell-p}, \quad \ell > p.
\end{aligned}$$

Since $\boldsymbol{a}_t$ are serially uncorrelated, elements of $\boldsymbol{\psi}_\ell$ denote the effects of $\boldsymbol{a}_t$ on $\boldsymbol{z}_{t+\ell}$. Thus, the coefficient matrices $\boldsymbol{\psi}_\ell$ of Eq. (9) can be regarded as the *impulse responses* of $\boldsymbol{z}_t$. Let $\psi_{ij}(\ell)$ be the (i,j)th element of $\boldsymbol{\psi}_\ell$ and define $\boldsymbol{\psi}_0 = \boldsymbol{I}$. Plots of $\{\psi_{ij}(\ell)\}$ versus ℓ (for $\ell \geq 0$) are the *impulse response functions* of $\boldsymbol{z}_t$.

In practice, elements of $\boldsymbol{a}_t$ tend to be correlated, i.e., $\boldsymbol{\Sigma}_a$ is not a diagonal matirx. In this case, change in a single element of $\boldsymbol{a}_t$ will simultaneously affect other elements of $\boldsymbol{a}_t$. Consequently, interpretations of $\{\psi_{ij}(\ell)\}$ are not as simple as they appear. To overcome this difficulty, one often takes a proper transformation of $\boldsymbol{a}_t$ such that elements of the innovation become uncorrelated, i.e. diagonize the covariance matrix. A simple way to achieve orthogonalization of innovation is to consider the Choleski decomposition of $\boldsymbol{\Sigma}_a$. Specifically, we can obtain

$$\boldsymbol{\Sigma}_a = \boldsymbol{P}\boldsymbol{P}',$$

where $\boldsymbol{P}$ is a lower diagonal matrix with positive diagonal elements. Let $\boldsymbol{\eta}_t = \boldsymbol{P}^{-1}\boldsymbol{a}_t$. Then,

$$\mathrm{cov}(\boldsymbol{\eta}_t) = \boldsymbol{P}^{-1}\mathrm{cov}(\boldsymbol{a}_t)(\boldsymbol{P}^{-1})' = \boldsymbol{P}^{-1}(\boldsymbol{P}\boldsymbol{P}')(\boldsymbol{P}')^{-1} = \boldsymbol{I}.$$

Thus, elements of $\boldsymbol{\eta}_t$ are uncorrelated.

From Eq. (9), we have

$$\begin{aligned}
\boldsymbol{z}_t &= \boldsymbol{\psi}(B)\boldsymbol{a}_t, \\
&= \boldsymbol{\psi}(B)\boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{a}_t, \\
&= [\boldsymbol{\psi}(B)\boldsymbol{P}]\boldsymbol{\eta}_t, \\
&= (\boldsymbol{\psi}_0^* + \boldsymbol{\psi}_1^* B + \boldsymbol{\psi}_2^* B^2 + \cdots)\boldsymbol{\eta}_t, \qquad (10)
\end{aligned}$$

where $\boldsymbol{\psi}_\ell^* = \boldsymbol{\psi}_\ell \boldsymbol{P}$ for $\ell \geq 0$. Let $\psi_{ij}^*(\ell)$ be the (i,j)th element of $\boldsymbol{\psi}_\ell^*$. The plot of $\psi_{ij}^*(\ell)$ against ℓ is called the *impulse response function* of $\boldsymbol{z}_t$. It denotes the impacts of an "one standard deviation" shock of the jth innovation at time t on the future values of $z_{i,t+\ell}$. In particular, the (i,j)th element of the transformation matrix $\boldsymbol{P}$ (where $i > j$) denotes the instantaneous effect of shock η_{jt} on z_{it}.
Remark: The prior definition of impulse response function depends on the *ordering* of elements in $\boldsymbol{z}_t$. The lower triangular structure of $\boldsymbol{P}$ indicates that the η_{it} is not affected by η_{jt} for $j > i$. There seems no simple way to avoid such a dependence in defining impulse response function when $\boldsymbol{\Sigma}_a$ is not a diagonal matrix.
Remark: An **R** script to calculate and plot impulse response functions of a VAR(p) is available on the course web. It is called **VARirf.R**. Demonstration of this program will be given in class.

4 Vector MA Model

A k-dimensional time series $\boldsymbol{z}_t$ follows a vector moving-average (VMA) model of order q if

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \boldsymbol{a}_t - \sum_{i=1}^{q} \boldsymbol{\theta}_i \boldsymbol{a}_{t-i}, \tag{11}$$

where $\boldsymbol{\mu}$ is a constant vector denoting the mean of $\boldsymbol{z}_t$, $\boldsymbol{\theta}_i$ are $k \times k$ matrices with $\boldsymbol{\theta}_q \neq \mathbf{0}$, and $\{\boldsymbol{a}_t\}$ is defined as before in Eq. (4). Using the backshift operator the model becomes $\boldsymbol{z}_t = \boldsymbol{\mu} + \boldsymbol{\theta}(B)\boldsymbol{a}_t$, where $\boldsymbol{\theta}(B) = \boldsymbol{I} - \sum_{i=1}^{q} \boldsymbol{\theta}_i B^i$ is a matrix polynomial of degree q.

4.1 VMA(1) model

Again, we start with the 2-dimensional vector moving-average (VAR) model,

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \boldsymbol{a}_t - \theta_1 \boldsymbol{a}_{t-1}.$$

The model can be written as

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} \theta_{1,11} & \theta_{1,12} \\ \theta_{1,21} & \theta_{1,22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix},$$

or equivalently,

$$\begin{aligned} z_{1t} &= \mu_1 + a_{1t} - \theta_{1,11} a_{1,t-1} - \theta_{1,12} a_{2,t-1} \\ z_{2t} &= \mu_2 + a_{2t} - \theta_{1,21} a_{1,t-1} - \theta_{1,22} a_{2,t-1}. \end{aligned}$$

Thus, $\theta_{1,12}$ is the impact of $a_{2,t-1}$ on z_{1t} in the presence of $a_{1,t-1}$.
Some basic properties:

- Relation to transfer function model
- The model is stationary
- The invertible condition
- Moment equations
- Univariate model for the components

4.2 VMA(q) model

The model is

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \theta(B)\boldsymbol{a}_t.$$

Some basic properties:

- Relation to transfer function model
- The model is stationary
- Invertibility condition
- Moment equations
- Univariate model for the components

4.3 Estimation

Two likelihood methods are available. The first one is the conditional likelihood method and the second exact likelihood method. For simplicity, we focus on a zero-mean VMA(1) model

$$\boldsymbol{z}_t = \boldsymbol{a}_t - \boldsymbol{\theta}\boldsymbol{a}_{t-1}.$$

Conditional likelihood method:
This method assumes that $\boldsymbol{a}_t = \mathbf{0}$ for $t \leq 0$. Consequently, using $\boldsymbol{a}_t = \boldsymbol{z}_t + \theta\boldsymbol{a}_{t-1}$, we have

$$\begin{aligned}
\boldsymbol{a}_1 &= \boldsymbol{z}_1 \\
\boldsymbol{a}_2 &= \boldsymbol{z}_2 + \boldsymbol{\theta}\boldsymbol{a}_1 = \boldsymbol{z}_2 + \boldsymbol{\theta}\boldsymbol{z}_1 \\
\boldsymbol{a}_3 &= \boldsymbol{z}_3 + \boldsymbol{\theta}\boldsymbol{a}_2 = \boldsymbol{a}_3 + \boldsymbol{\theta}\boldsymbol{z}_2 + \boldsymbol{\theta}^2\boldsymbol{z}_1 \\
\vdots &= \vdots \\
\boldsymbol{a}_n &= \boldsymbol{z}_n + \sum_{j=1}^{n-1} \boldsymbol{\theta}^{n-j}\boldsymbol{z}_{n-j}.
\end{aligned}$$

Thus, the transformation from $\{\boldsymbol{z}_t\}_{t=1}^n$ to $\{\boldsymbol{a}_t\}_{t=1}^n$ has a unity Jacobian. As such,

$$p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = p(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \prod_{t=1}^n p(\boldsymbol{a}_t|\boldsymbol{\theta}, \boldsymbol{\Sigma}).$$

The likelihood function is readily available.

Exact likelihood method:
In this method, the initial value $\boldsymbol{a}_0$ is treated as a random variable. Our goal is to obtain the joint density function of $\{\boldsymbol{z}_t\}_{t=1}^n$. To this end, consider the following relationship

$$\begin{aligned}
\boldsymbol{a}_0 &= \boldsymbol{a}_0 \\
\boldsymbol{a}_1 &= \boldsymbol{z}_1 + \boldsymbol{\theta} a_0 \\
\boldsymbol{a}_2 &= \boldsymbol{z}_2 + \boldsymbol{\theta}\boldsymbol{a}_1 = \boldsymbol{z}_2 + \boldsymbol{\theta}\boldsymbol{z}_1 + \boldsymbol{\theta}^2\boldsymbol{a}_0
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{a}_3 &= \boldsymbol{z}_3 + \boldsymbol{\theta}\boldsymbol{a}_2 = \boldsymbol{z}_3 + \boldsymbol{\theta}\boldsymbol{z}_2 + \boldsymbol{\theta}^2\boldsymbol{z}_1 + \boldsymbol{\theta}^3\boldsymbol{a}_0 \\
\vdots &= \vdots \\
\boldsymbol{a}_n &= \boldsymbol{z}_n + \sum_{j=1}^{n-1}\boldsymbol{\theta}^j\boldsymbol{z}_{n-j} + \boldsymbol{\theta}^n\boldsymbol{a}_0.
\end{aligned} \tag{12}
$$

Thus, we can transform $\{\boldsymbol{a}_0, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_n\}$ into $\{\boldsymbol{a}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_n\}$ whose joint density is known. Specifically, we have

$$
p(\boldsymbol{a}_0, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_n|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = p(\boldsymbol{a}_0, \cdots, \boldsymbol{a}_n|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \prod_{t=0}^{n} p(\boldsymbol{a}_t|\boldsymbol{\theta}, \boldsymbol{\Sigma}).
$$

Next, the joint density of the data can be obtained by integrating out $\boldsymbol{a}_0$ as

$$
p(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \int p(\boldsymbol{a}_0, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_n|\boldsymbol{\theta}, \boldsymbol{\Sigma})d\boldsymbol{a}_0.
$$

The remaining issue is how to integrate out $\boldsymbol{a}_0$.
To help better understand the concept, we shall use the special case of univariate MA(1) models in the following discussion. The same technique applies to the vector case, but the notation is much more involved.

Univariate zero-mean MA(1) model: $z_t = a_t - \theta a_{t-1}$.
Under normality, the joint density of a_0 and $\{z_t\}_{t=1}^n$ is

$$
p(a_0, z_1, \cdots, z_n|\theta, \sigma) = \prod_{t=0}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{a_t^2}{2\sigma^2}],
$$

where $a_t = z_t + \sum_{i=1}^{t-1} \theta^i z_{t-i} + \theta^t a_0 onthe$; see Eq. (12). Defining

$$
y_0 = 0, \quad y_t = z_t + \sum_{i=1}^{t-1} \theta^i z_{t-i}, \quad t = 1, \cdots, n, \tag{13}
$$

we have $a_t = y_t + \theta^t a_0$. Therefore, the exponent of the above joint density is

$$
S = \frac{-1}{2\sigma^2} \sum_{t=0}^{n} (y_t + \theta^t a_0)^2.
$$

To integrate out a_0, we must find ways to isolate a_0 in the exponent S.
From Eq. (12) and the definition of y_t in Eq. (13), we have

$$
\begin{aligned}
a_0 &= y_0 + a_0 \\
a_1 &= y_1 + \theta a_0 \\
a_2 &= y_2 + \theta^2 a_0 \\
\vdots &= \vdots \\
a_n &= y_n + \theta^n a_0.
\end{aligned}
$$

Consequently, we have a linear regression setup

$$\begin{aligned} y_0 &= -1a_0 + a_0 = -\theta^0 a_0 + a_0 \\ y_1 &= -\theta a_0 + a_1 \\ y_2 &= -\theta^2 a_0 + a_2 \\ \vdots &= \vdots \\ y_n &= -\theta^n a_0 + a_n, \end{aligned}$$

where a_0 is the unknown parameter and $x_t = -\theta^t$ for $t \geq 0$. The least squares estimate of a_0 is

$$\hat{a}_0 = \frac{\sum_{t=0}^n -\theta^t y_t}{\sum_{t=0}^n \theta^{2t}}. \tag{14}$$

Note that this is simply saying that given the data and θ we can estimate the initial innovation a_0. Next, from the linear regression, we have

- The residual is $\hat{a}_t = y_t + \theta^t \hat{a}_0$.
- $\sum_{t=0}^n (-\theta^t)(y_t + \theta^t \hat{a}_0) = \sum_{t=0}^n (-\theta^t)\hat{a}_t = 0$. (That is, regressor and residual are uncorrelated.)
- The distribution of $\hat{a}_0$ is $N(a_0, \frac{\sigma^2}{\sum_{t=0}^n \theta^{2t}})$.

Now, the exponent S can be written as

$$\begin{aligned} S &= \frac{-1}{2\sigma^2} \sum_{t=0}^n (y_t + \theta^t a_0)^2 \\ &= \frac{-1}{2\sigma^2} \left[\sum_{t=0}^n [y_t + \theta^t \hat{a}_0 + \theta^t (a_0 - \hat{a}_0)]^2 \right] \\ &= \frac{-1}{2\sigma^2} \left[\sum_{t=0}^n (y_t + \theta^t \hat{a}_0)^2 + \sum_{t=0}^n \theta^{2t} (a_0 - \hat{a}_0)^2 \right], \end{aligned}$$

where the cross-product term vanishes because of the least squares properties mentioned above. Therefore, the exponent S indeed can be decomposed into two parts with one part involving a_0 and the other without a_0. The joint density of $a_0, z_1, \cdots, z_n$ becomes

$$p(a_0, z_1, \cdots, z_n) = [\frac{1}{\sqrt{2\pi}\sigma}]^{n+1} \exp\{-\frac{1}{2\sigma^2}[\sum_{t=0}^n (y_t + \theta^t \hat{a}_0)^2 + \sum_{t=0}^n \theta^{2t}(a_0 - \hat{a}_0)^2]\}.$$

Since the term involving a_0 in the exponent S is quadratic, the conditional distribution of a_0 given the data and the parameters (θ,σ^2) is normal with mean $\hat{a}_0$ and variance $\sigma^2/(\sum_{t=0}^n \theta^{2t})$. Using properties of normal density, we obtain the joint density of the data as

$$\begin{aligned} p(z_1, \cdots, z_n|\theta, \sigma) &= \int p(a_0, z_1, \cdots, z_n|\theta, \sigma^2) da_0 \\ &= [\frac{1}{\sqrt{2\pi}\sigma}]^n \frac{1}{\sqrt{\sum_{t=0}^n \theta^{2t}}} \exp[\frac{-1}{2\sigma^2} \sum_{t=0}^n (y_t + \theta^t \hat{a}_0)^2], \end{aligned}$$

where y_t is defined in Eq. (13) and $\hat{a}_0$ in Eq. (14). The log-likelihood function is

$$L(\theta, \sigma^2) \propto -\frac{1}{2}\ln(\sum_{t=0}^{n}\theta^{2t}) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=0}^{n}(y_t + \theta^t \hat{a}_0)^2.$$

Discussion:

- The exact method requires more intensive computation. Roughly speaking, to evaluate the likelihood function, it goes through the data twice. The first round is to estimate a_0 and the second round to calculate a_t given a_0.
- The exact method is more accurate. This is particularly so when the sample size is small or when the MA part has characteristic root close to unit circle. In other words, exact likelihood method is preferred when the model is close to being non-invertible. See Hillmer and Tiao (1982, JASA) for some results of simulation study.

Example. Consider the monthly returns of the Decile 1 portfolio of NYSE/AMEX/NASDAQ from January 1960 to December 2003 with 528 observations. It is well-known that the return series exhibit significant *January* effect, which strictly speaking belongs to deterministic seasonality. In term of seasonal ARMA models, deterministic seasonality implies a near cancellation of $(1 - B^{12})$ factor in the AR and MA components. Thus, in the ARMA framework, the January effect could lead to a non-invertible ARMA model. We use this feature to demonstrate the difference between conditional and exact likelihood methods.

SCA demonstration

```
 --
input dd,x,y,z. file 'm-decile1510.txt'
 --
acf x. maxl 24.

 NAME OF THE SERIES . . . . . . . . . .            X
 TIME PERIOD ANALYZED . . . . . . . . . .  1  TO   528
 MEAN OF THE (DIFFERENCED) SERIES . . .       0.0180
 STANDARD DEVIATION OF THE SERIES . . .       0.0792
 T-VALUE OF MEAN (AGAINST ZERO) . . . .       5.2107

           -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
             +----+----+----+----+----+----+----+----+----+----+
                                     I
   1   0.23                        + IX+XXXX
   2   0.00                        + I +
   3  -0.03                        +XI +
   4   0.01                        + I +
   5   0.01                        + I +
   6  -0.03                        +XI +
```

```
   7  -0.04                        +XI +
   8  -0.09                        XXI +
   9  -0.06                        XXI +
  10   0.01                        + I +
  11   0.10                        + IX+X
  12   0.26                        + IX+XXXXX
  13  -0.02                        + I +
  14  -0.05                        +XI +
  15  -0.05                        +XI +
  16  -0.02                        +XI +
  17  -0.02                        + I +
  18  -0.07                        XXI +
  19  -0.11                       X+XI +
  20  -0.04                        +XI +
  21  -0.04                        +XI +
  22  -0.04                        +XI +
  23   0.02                        + I +
  24   0.18                        + IX+XX
 --
tsm m1. model (12)x=c+(1)(12)noise.
 --
estim m1. hold resi(r1)

 THE FOLLOWING ANALYSIS IS BASED ON TIME SPAN   1  THRU  528

 SUMMARY FOR UNIVARIATE TIME SERIES MODEL --    M1
 ------------------------------------------------------------------------
 VARIABLE   TYPE OF    ORIGINAL     DIFFERENCING
           VARIABLE   OR CENTERED

    X       RANDOM     ORIGINAL     NONE
 ------------------------------------------------------------------------
PARAMETER   VARIABLE  NUM./  FACTOR  ORDER   CONS-     VALUE    STD     T
  LABEL       NAME    DENOM.                 TRAINT             ERROR  VALUE

 1    C               CNST      1      0     NONE      .0008   .0006   1.32
 2            X        MA       1      1     NONE     -.2612   .0423  -6.17
 3            X        MA       2     12     NONE      .8751   .0305  28.68
 4            X        AR       1     12     NONE      .9863   .0161  61.25

 EFFECTIVE NUMBER OF OBSERVATIONS . .           516
 R-SQUARE . . . . . . . . . . . . . .         0.186
 RESIDUAL STANDARD ERROR. . . . . . .  0.714956E-01
 --
estim m1. method exact. hold resi(r1)
```

```
 THE FOLLOWING ANALYSIS IS BASED ON TIME SPAN   1  THRU  528

 SUMMARY FOR UNIVARIATE TIME SERIES MODEL --    M1
 -----------------------------------------------------------------------
 VARIABLE   TYPE OF    ORIGINAL     DIFFERENCING
           VARIABLE   OR CENTERED

    X       RANDOM     ORIGINAL     NONE
 -----------------------------------------------------------------------
PARAMETER   VARIABLE  NUM./  FACTOR  ORDER   CONS-     VALUE    STD     T
  LABEL       NAME    DENOM.                TRAINT             ERROR  VALUE

 1    C               CNST      1       0     NONE      .0002   .0003    .76
 2             X       MA       1       1     NONE    -.2817   .0421  -6.69
 3             X       MA       2      12     NONE      .9994   .0158  63.29
 4             X       AR       1      12     NONE      .9987   .0057 174.75

 EFFECTIVE NUMBER OF OBSERVATIONS . .           516
 R-SQUARE . . . . . . . . . . . . . . .        0.283
 RESIDUAL STANDARD ERROR. . . . . . .  0.671054E-01
 --
```

4.4 Forecasting

MA models have finite memory in the sense that the forecasts converge to the mean of the series in a finite number of steps. In other words, MA models are mean-reverting quickly. Consider the VMA(q) model

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \boldsymbol{a}_t - \theta_1 \boldsymbol{a}_{t-1} - \cdots - \boldsymbol{\theta}_q \boldsymbol{a}_{t-q}.$$

Suppose that the forecast origin is h. For 1-step ahead forecast, the model is

$$\boldsymbol{z}_{h+1} = \boldsymbol{\mu} + \boldsymbol{a}_{h+1} - \theta_1 \boldsymbol{a}_h - \cdots - \boldsymbol{\theta}_q \boldsymbol{a}_{h+1-q}.$$

Taking conditional expectation,

$$\boldsymbol{z}_h(1) = \boldsymbol{\mu} - \theta_1 \boldsymbol{a}_h - \cdots - \boldsymbol{\theta}_q \boldsymbol{a}_{h+1-q}.$$

By the same method, it is easily seen that

$$\begin{aligned}
\boldsymbol{z}_h(q) &= \boldsymbol{\mu} - \boldsymbol{\theta}_q \boldsymbol{a}_h \\
\boldsymbol{z}_h(j) &= \boldsymbol{\mu}, \quad j > q.
\end{aligned}$$