Lecture 12: Time Series Analysis III

MIT 18.S096

Dr. Kempthorne

Fall 2013

Outline

1. Time Series Analysis III
   - Cointegration: Definitions
   - Cointegrated VAR Models: VECM Models
   - Estimation of Cointegrated VAR Models
   - Linear State-Space Models
   - Kalman Filter

Cointegration

An $m-$dimensional stochastic process $\{\mathbf{X}_t\} = \{\ldots, \mathbf{X}_{t-1}, \mathbf{X}_t, \ldots\}$ is $I(d)$, **Integrated of order d** if the d-differenced process

$$\Delta^d X_t = (1-L)^d X_t \quad \text{is stationary.}$$

If $\{X_t\}$ has a $VAR(p)$ representation, i.e.,

$$\Phi(L)X_t = \epsilon_t, \text{ where } \Phi(L) = I - A_1 L - A_2 - \cdots - A_p L^p.$$

then $\Phi(L) = (1-L)^d \Phi^*(L)$

where $\Phi^*(L) = (1 - A_1^* L - A_2^* L^2 - \cdots A_m^* L^m)$ specifies the stationary $VAR(m)$ process $\{\Delta^d X_t\}$ with $m = p - d$.

Issue:

- Every component series of $\{X_t\}$ may be $I(1)$, but the process may not be jointly integrated.
- Linear combinations of the component series (without any differencing) may be stationary!

If so, the multivariate time series $\{X_t\}$ is **“Cointegrated”**

Consider $\{\mathbf{X}_t\}$ where $\mathbf{X}_t = (x_{1,t}, x_{2,t}, \dots, x_{m,t})'$ an m-vector of component time series, and each is $I(1)$, integrated of order 1.
If $\{X_t\}$ is cointegrated, then there exists an m-vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m)'$ such that

$$\boldsymbol{\beta}'\mathbf{X}_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_m x_{m,t} \sim I(0),$$

a stationary process.

- The cointegration vector $\boldsymbol{\beta}$ can be scaled arbitrarily, so assume a normalization:
$$\boldsymbol{\beta} = (1, \beta_2, \dots, \beta_m)'$$
- The expression: $\boldsymbol{\beta}'\mathbf{X}_t = u_t$, where $\{u_t\} \sim I(0)$ is equivalent to:
$$x_{1,t} = (\beta_2 x_{2,t} + \cdots \beta_m x_{m,t}) + u_t,$$
where
  - $\boldsymbol{\beta}'\mathbf{X}_t$ is the **long-run equilibrium relationship**
  i.e., $x_{1,t} = (\beta_2 x_{2,t} + \cdots \beta_m x_{m,t})$
  - u_t is the **disequilibrium error / cointegration residual**.

Examples of Cointegration

- Term structure of interest rates: expectations hypothesis.
- Purchase power parity in foreign exchange: cointegration among exchange rate, foreign and domestic prices.
- Money demand: cointegration among money, income, prices and interest rates.
- Covered interest rate parity: cointegration among forward and spot exchange rates.
- Law of one price: cointegration among identical/equivalent assets that must be valued identically to limit arbitrage.
  - Spot and futures prices.
  - Prices of same asset on different trading venues

Outline

Cointegrated VAR Models: VECM Models

The m-dimensional multivariate time series $\{\mathbf{X}_t\}$ follows the $VAR(p)$ model with auto-regressive order p if

$$\mathbf{X}_t = C + \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \mathbf{\Phi}_2 \mathbf{X}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \boldsymbol{\eta}_t$$

where

$C = (c_1, c_2, \ldots, c_m)'$ is an m-vector of constants.

$\mathbf{\Phi}_1, \mathbf{\Phi}_2, \ldots, \mathbf{\Phi}_p$ are $(m \times m)$ matrices of coefficients

$\{\boldsymbol{\eta}_t\}$ is multivariate white noise $MWN(\mathbf{0}_m, \mathbf{\Sigma})$

The $VAR(p)$ model is covariance stationary if

$$det \left[I_m - (\mathbf{\Phi}_1 z + \mathbf{\Phi}_2 z^2 + \cdots + \mathbf{\Phi}_p z^p) \right] = 0$$

has roots outside $|z| \leq 1$ for complex z.

Suppose $\{\mathbf{X}_t\}$ is $I(1)$ of order 1. We develop a **Vector Error Correction Model** representation of this model by successive modifications of the model equation:

VAR(p) Model Equation

$$\mathbf{X}_t = C + \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \mathbf{\Phi}_2 \mathbf{X}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \boldsymbol{\eta}_t$$

- Subtract $\mathbf{X}_{t-1}$ from both sides:

  $\Delta \mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-1} = C + (\mathbf{\Phi}_1 - I_m)\mathbf{X}_{t-1} + \mathbf{\Phi}_2 \mathbf{X}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \boldsymbol{\eta}_t$
- Subtract and add $(\mathbf{\Phi}_1 - I_m)\mathbf{X}_{t-2})$ from right-hand side:

  $\Delta \mathbf{X}_t = C + (\mathbf{\Phi}_1 - I_m)\Delta\mathbf{X}_{t-1} + (\mathbf{\Phi}_2 + \mathbf{\Phi}_1 - I_m)\mathbf{X}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \boldsymbol{\eta}_t$
- Subtract and add $(\mathbf{\Phi}_2 + \mathbf{\Phi}_1 - I_m)\mathbf{X}_{t-3})$ from right-hand side:

  $\Delta \mathbf{X}_t = C + (\mathbf{\Phi}_1 - I_m)\Delta\mathbf{X}_{t-1} + (\mathbf{\Phi}_2 + \mathbf{\Phi}_1 - I_m)\Delta\mathbf{X}_{t-2} + (\mathbf{\Phi}_3 + \mathbf{\Phi}_2 + \mathbf{\Phi}_1 - I_m)\mathbf{X}_{t-3} + \cdots$

$$\begin{aligned}\Longrightarrow \Delta \mathbf{X}_t &= C + (\mathbf{\Phi}_1 - I_m)\Delta\mathbf{X}_{t-1} \\ &+ (\mathbf{\Phi}_2 + \mathbf{\Phi}_1 - I_m)\Delta\mathbf{X}_{t-2} \\ &+ (\mathbf{\Phi}_3 + \mathbf{\Phi}_2 + \mathbf{\Phi}_1 - I_m)\Delta\mathbf{X}_{t-3} + \cdots \\ &+ (\mathbf{\Phi}_{p-1} + \cdots + \mathbf{\Phi}_3 + \mathbf{\Phi}_2 + \mathbf{\Phi}_1 - I_m)\Delta\mathbf{X}_{t-(p-1)} \\ &+ (\mathbf{\Phi}_p + \cdots + \mathbf{\Phi}_3 + \mathbf{\Phi}_2 + \mathbf{\Phi}_1 - I_m)\mathbf{X}_{t-p} + \eta_t\end{aligned}$$

Reversing the order of incorporating Δ-terms we can derive

$$\Delta \mathbf{X}_t = C + \mathbf{\Pi}\mathbf{X}_{t-1} + \mathbf{\Gamma}_1 \Delta\mathbf{X}_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}\Delta\mathbf{X}_{t-(p-1)} + \eta_t$$

where: $\mathbf{\Pi} = (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \cdots \mathbf{\Phi}_p - I_m)$ and $\mathbf{\Gamma}_k = (-\sum_{j=k+1}^{p} \mathbf{\Phi}_j)$,

Vector Error Correction Model (VECM)

The $VAR(p)$ model for $\{\mathbf{X}_t\}$ is a $VECM$ model for $\{\Delta\mathbf{X}_t\}$.

$$\Delta\mathbf{X}_t = C + \mathbf{\Pi}\mathbf{X}_t + \mathbf{\Gamma}_1\Delta\mathbf{X}_{t-1} + \cdots + \mathbf{\Gamma}_{p-1}\Delta\mathbf{X}_{t-(p-1)} + \eta_t$$

By assumption, the $VAR(p)$ model for $\{\mathbf{X}_t\}$ is $I(1)$, so the $VECM$ model for $\{\Delta\mathbf{X}_t\}$ is $I(0)$.

- The left-hand-side $\Delta\mathbf{X}_t$ is stationary / $I(0)$.
- The terms on the right-hand-side $\Delta\mathbf{X}_{t-j}$, $j = 1, 2, \ldots, p-1$ are stationary / $I(0)$.
- The term $\mathbf{\Pi}X_t$ must be stationary /$I(0)$.
- This term $\mathbf{\Pi}X_t$ contains any **cointegrating terms** of $\{\mathbf{X}_t\}$.
- Given that the $VAR(p)$ process had unit roots, it must be that $\mathbf{\Pi}$ is singular, i.e., the linear transformation eliminates the unit roots.

- The matrix $\mathbf{\Pi}$ is of reduced rank $r < m$ and either
  - $rank(\mathbf{\Pi}) = 0$ and $\mathbf{\Pi} = 0$ and there are no cointegrating relationships.
  - $rank(\mathbf{\Pi}) > 0$ and $\mathbf{\Pi}$ defines the cointegrating relationships.

If cointegrating relationships exist, then $rank(\mathbf{\Pi}) = r$ with $0 < r < m$, and we can write

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}',$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are each $(m \times r)$ matrices of full rank r.

- The columns of $\boldsymbol{\beta}$ define linearly independent vectors which **cointegrate $\mathbf{X}_t$**.
- The decomposition of $\mathbf{\Pi}$ is not unique. For any invertible $(r \times r)$ matrix G,

  $$\mathbf{\Pi} = \boldsymbol{\alpha}_*\boldsymbol{\beta}_*'$$

  where $\boldsymbol{\alpha}_* = \boldsymbol{\alpha}G$ and $\boldsymbol{\beta}_*' = G^{-1}\boldsymbol{\beta}'$.

Outline

1. Time Series Analysis III
   - Cointegration: Definitions
   - Cointegrated VAR Models: VECM Models
   - Estimation of Cointegrated VAR Models
   - Linear State-Space Models
   - Kalman Filter

Estimation of Cointegrated VAR Models

Unrestricted Least Squares Estimation

- Sims, Stock, and Watson (1990), and Park and Phillips (1989) prove that in estimation for cointegrated $VAR(p)$ models, the least-squares estimator of the original model yields parameter estimates which are:
  - Consistent.
  - Have asymptotic distributions identical to those of maximum-likelihood estimators.
  - Constraints on parameters due to cointegration (i.e., the reduced rank of $\mathbf{\Pi}$) hold asymptotically.

Maximum Likelihood Estimation*

- Banerjee and Hendry (1992): apply method of **concentrated likelihood** to solve for maximum likelihood estiamtes.

* Advanced topic for optional reading/study

Estimation of Cointegrated VAR Models

Maximum Likelihood Estimation (continued)

- Johansen (1991) develops a **reduced -rank regression** methodology for the maximum likelihood estimation for *VECM* models.
- This methodology provides likelihood ratio tests for the number of cointegrating vectors:
  - *Johansen's* Trace Statistic (sum of eigenvalues of $\hat{\mathbf{\Pi}}$)
  - *Johansen's* Maximum-Eigenvalue Statistic (max eigenvalue of $\hat{\mathbf{\Pi}}$).

Outline

1. Time Series Analysis III
   - Cointegration: Definitions
   - Cointegrated VAR Models: VECM Models
   - Estimation of Cointegrated VAR Models
   - Linear State-Space Models
   - Kalman Filter

Linear State-Space Model

General State-Space Formulation

y_t: $(k \times 1)$ observation vector at time t
s_t: $(m \times 1)$ state-vector at time t
ϵ_t: $(k \times 1)$ observation-error vector at time t
η_t: $(n \times 1)$ state transition innovation/error vector

State Equation / Transition Equation

$$S_{t+1} = T_t s_t + R_t \eta_t$$

where

T_t: $(m \times m)$ transition coefficients matrix
R_t: $(m \times n)$ fixed matrix; often column(s) of I_p
η_t: i.i.d. $N(0_n, Q_t)$, where Q_t $_{(n \times n)}$ is positive definite.

Linear State-Space Model Formulation

Observation Equation / Measurement Equation

$$y_t = Z_t s_t + \epsilon_t$$

where

Z_t : $(k \times m)$ observation coefficients matrix

ϵ_t: i.i.d. $N(0_k, H_t)$, where H_t $(k \times k)$ is positive definite.

Joint Equation

$$\begin{bmatrix} s_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} T_t \\ Z_t \end{bmatrix} s_t + \begin{bmatrix} R_t \eta_t \\ \epsilon_t \end{bmatrix} = \Phi_t s_t + u_t,$$

where

$$u_t = \begin{bmatrix} R_t \eta_t \\ \epsilon_t \end{bmatrix} \sim N(0, \Omega), \text{ with } \Omega = \begin{bmatrix} R_t Q_t R_t^T & 0 \\ 0 & H_t \end{bmatrix}$$

Note: Often model is time invariant (T_t, R_t, Z_t, Q_t, H_t constants)

CAPM Model with Time-Varying Betas

Consider the CAPM Model with time-varying parameters:

$$\begin{array}{lcll} r_t & = & \alpha_t + \beta_t r_{m,t} + \epsilon_t, & \epsilon_t \sim N(0, \sigma_\epsilon^2) \\ \alpha_{t+1} & = & \alpha_t + \nu_t, & \nu_t \sim N(0, \sigma_\nu^2) \\ \beta_{t+1} & = & \beta_t + \xi_t, & \xi_t \sim N(0, \sigma_\xi^2) \end{array}$$

where

r_t is the excess return of a given asset
$r_{m,t}$ is the excess return of the market portfolio
$\{\epsilon_t\}, \{\nu_t\}, \{\xi_t\}$ are mutually independent processes

Note:

$\{\alpha_t\}$ is a Random Walk with i.i.d. steps $N(0, \sigma_\nu^2)$
$\{\beta_t\}$ is a Random Walk with i.i.d. steps $N(0, \sigma_\xi^2)$
(Mutually independent processes)

Time-Varying CAPM Model: Linear State-Space Model

State Equation

$$\begin{aligned}\begin{bmatrix}\alpha_{t+1}\\ \beta_{t+1}\end{bmatrix} &= \begin{bmatrix}\alpha_t\\ \beta_t\end{bmatrix} + \begin{bmatrix}\nu_t\\ \xi_t\end{bmatrix}\\ &= \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\alpha_t\\ \beta_t\end{bmatrix} + \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\nu_t\\ \xi_t\end{bmatrix}\end{aligned}$$

Equivalently:

$$s_{t+1} = T_t s_t + R_t \eta_t$$

where:

$$s_t = \begin{bmatrix}\alpha_t\\ \beta_t\end{bmatrix},\ T_t = R_t = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},$$

$$\eta_t = \begin{bmatrix}\nu_t\\ \xi_t\end{bmatrix} \sim N_2(\mathbf{0}_2, Q_t),\quad \text{with } Q_t = \begin{bmatrix}\sigma_\nu^2 & 0\\ 0 & \sigma_\xi^2\end{bmatrix}$$

Terms:

state vector s_t, transition coefficients T_t
transition white noise η_t

Time-Varying CAPM Model: Linear State-Space Model

Observation Equation / Measurement Equation

$$r_t = \begin{bmatrix} 1 & r_{m,t} \end{bmatrix} \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + \epsilon_t$$

Equivalently

$$r_t = Z_t s_t + \epsilon_t$$

where

$Z_t = \begin{bmatrix} 1 & r_{m,t} \end{bmatrix}$ is the observation coefficients matrix

$\epsilon_t \sim N(0, H_t)$, is the observation white noise with $H_t = \sigma_\epsilon^2$.

Joint System of Equations

$$\begin{bmatrix} s_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} T_t \\ Z_t \end{bmatrix} s_t + \begin{bmatrix} R_t \eta_t \\ \epsilon_t \end{bmatrix}$$

$$= \Phi_t s_t + u_t \text{ with } Cov(u_t) = \begin{bmatrix} R_t \Omega_\eta R_t^T & 0 \\ 0 & H_t \end{bmatrix}$$

Linear Regression Model with Time-Varying β

Consider a normal linear regression model with time-varying regression coefficients:

$$y_t = \mathbf{x}_t^T \boldsymbol{\beta}_t + \epsilon_t, \text{ where } \epsilon_t \text{ are i.i.d. } N(0, \sigma_\epsilon^2).$$

where

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, \ldots, x_{p,t})^t, \; p\text{-vector of explanatory variables}$$
$$\boldsymbol{\beta}_t = (\beta_{1,t}, \beta_{2,t}, \ldots, \beta_{p,t})^t, \text{ regression parameter vector}$$

and for each parameter component j, $j = 1, \ldots, p$,

$$\beta_{j,t+1} = \beta_{j,t} + \eta_{j,t}, \text{ with } \{\eta_{j,t}, t = 1, 2, \ldots\} \text{ i.i.d. } N(0, \sigma_j^2).$$

i.e., a Random Walk with iid steps $N(0, \sigma_j^2)$.

Joint State-Space Equations

$$\begin{bmatrix} s_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} \mathbf{I}_p \\ \mathbf{x}_t^T \end{bmatrix} s_t + \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}, \text{ with state vector } \mathbf{s}_t = \boldsymbol{\beta}_t$$
$$= \begin{bmatrix} \mathbf{T}_t \\ \mathbf{Z}_t \end{bmatrix} s_t + \begin{bmatrix} R_t \eta_t \\ \epsilon_t \end{bmatrix}$$

(Time-Varying) Linear Regression as a State-Space Model

where

$$\eta_t \sim N(0, Q_t), \text{ with } Q_t = diag(\sigma_1^2, \ldots, \sigma_p^2)$$
$$\epsilon_t \sim N(0, H_t), \text{ with } H_t = \sigma_\epsilon^2.$$

Special Case: $\sigma_j^2 \equiv 0$: Normal Linear Regression Model

- Successive estimation of state-space model parameters with $t = p+1, p+2, \ldots,$ yields recursive updating algorithm for linear time-series regression model.

Autoregressive Model AR(p)

Consider the $AR(p)$ model

$$\phi(L)y_t = \epsilon_t$$

where

$$\phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j \quad \text{and } \{\epsilon_t\} \text{ i.i.d. } N(0, \sigma_\epsilon^2).$$

so

$$y_{t+1} = \sum_{j=1}^{p} \phi_j y_{t+1-j} + \epsilon_{t+1}$$

Define state vector:

$$s_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})^T$$

Then

$$s_{t+1} = \begin{bmatrix} y_{t+1} \\ y_t \\ y_{t-1} \\ \vdots \\ y_{t-(p-2)} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-(p-1)} \end{bmatrix} + \begin{bmatrix} \epsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

State-Space Model for $AR(p)$

State Equation

$$s_{t+1} = T_t s_t + R_t \eta_t,$$

where

$$T_t = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad R_t = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

and

$$\{\eta_t = \epsilon_{t+1}\}\ i.i.d. N(0, \sigma_\epsilon^2).$$

Observation Equation /Measurement Equation

$$\begin{aligned} y_t &= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix} s_t \\ &= \mathbf{Z}_t s_t \quad \text{(no measurement error)} \end{aligned}$$

Moving Average Model $MA(q)$

Consider the $MA(q)$ model

$$y_t = \theta(L)\epsilon_t$$

where

$$\theta(L) = 1 + \sum_{j=1}^{q} \theta_j L^j \quad \text{and } \{\epsilon_t\} \text{ i.i.d. } N(0, \sigma_\epsilon^2).$$

so

$$y_{t+1} = \epsilon_{t+1} + \sum_{j=1}^{q} \theta_j \epsilon_{t+1-j}$$

Define state vector:

$$s_t = (\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q})^T$$

Then

$$s_{t+1} = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-(q-1)} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{t-1} \\ \epsilon_{t-2} \\ \vdots \\ \epsilon_{t-q} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

State-Space Model for $MA(q)$

State Equation

$$s_{t+1} = T_t s_t + R_t \eta_t,$$

where

$$T_t = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad R_t = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

and

$$\{\eta_t = \epsilon_t\} \ i.i.d.N(0, \sigma_\epsilon^2).$$

Observation Equation / Measurement Equation

$$\begin{aligned} y_t &= \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_{q-1} & \theta_q \end{bmatrix} s_t + \epsilon_t \\ &= \mathbf{Z}_t s_t + \epsilon_t \end{aligned}$$

Auto-Regressive-Moving-Average Model $ARMA(p, q)$

Consider the $ARMA(p, q)$ model

$$\phi(L)y_t = \theta(L)\epsilon_t$$

where

$$\phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j \quad \text{and } \{\epsilon_t\} \text{ i.i.d. } N(0, \sigma_\epsilon^2).$$
$$\theta(L) = 1 + \sum_{j=1}^{q} \theta_j L^j \quad \text{and } \{\epsilon_t\} \text{ i.i.d. } N(0, \sigma_\epsilon^2).$$

so

$$y_{t+1} = \sum_{j=1}^{p} \phi_j y_{t+1-j} + \epsilon_{t+1} + \sum_{j=1}^{p} \theta_j \epsilon_{t+1-j}$$

Set $m = max(p, q + 1)$ and define

$$\{\phi_1, \ldots, \phi_m\} : \phi(L) = 1 - \sum_{j=1}^{m} \phi_j L^j$$
$$\{\theta_1, \ldots, \theta_m\} : \theta(L) = 1 - \sum_{j=1}^{m} \theta_j L^j$$

i.e.,

$$\phi_j = 0 \text{ if } p < j \leq m$$
$$\theta_j = 0 \text{ if } q < j \leq m$$

So: $\{y_t\} \sim ARMA(m, m-1)$

State Space Model for ARMA(p,q)

Harvey (1993) State-Space Specification

Define state vector:

$$s_t = (s_{1,t}, s_{2,t}, \dots, s_{m,t})^T, \text{ where } m = max(p, q+1).$$

recursively:

$s_{1,t} = y_t$: Use this definition and the main model equation to define $s_{2,t}$ and η_t:

$$\begin{array}{rcl} y_{t+1} & = & \sum_{j=1}^{p} \phi_j y_{t+1-j} + \epsilon_{t+1} + \sum_{j=1}^{q} \theta_j \epsilon_{t+1-j} \\ s_{1,t+1} & = & \phi_1 s_{1,t} + 1 \cdot s_{2,t} + \eta_t \end{array}$$

where

$$\begin{array}{rcl} s_{2,t} & = & \sum_{i=2}^{m} \phi_i y_{t+1-i} + \sum_{j=1}^{m-1} \theta_j \epsilon_{t+1-j} \\ \eta_t & = & \epsilon_{t+1} \end{array}$$

- Use $s_{1,t} = y_t$, $s_{2,t}$, and $\eta_t = \epsilon_{t+1}$: to define $s_{3,t}$

$$\begin{aligned} s_{2,t+1} &= \sum_{i=2}^{m} \phi_i y_{t+2-j} + \sum_{j=1}^{m-1} \theta_j \epsilon_{t+2-j} \\ &= \phi_2 y_t + 1 \cdot [\sum_{i=3}^{m} \phi_i y_{t+2-j} + \sum_{j=2}^{m-1} \theta_j \epsilon_{t+2-j}] + (\theta_1 \epsilon_{t+1}) \\ &= \phi_2 s_{1,t} + 1 \cdot [s_{3,t}] + R_{2,1} \eta_t \end{aligned}$$

where

$$\begin{aligned} s_{3,t} &= \sum_{i=3}^{m} \phi_i y_{t+2-j} + \sum_{j=2}^{m-1} \theta_j \epsilon_{t+2-j} \\ R_{2,1} &= \theta_1 \\ \eta_t &= \epsilon_{t+1} \end{aligned}$$

- Use $s_{1,t} = y_t$, $s_{3,t}$, and $\eta_t = \epsilon_{t+1}$: to define $s_{4,t}$

$$\begin{aligned} s_{3,t+1} &= \sum_{i=3}^{m} \phi_i y_{t+3-j} + \sum_{j=2}^{m-1} \theta_j \epsilon_{t+3-j} \\ &= \phi_3 y_t + 1 \cdot [\sum_{i=4}^{m} \phi_i y_{t+3-j} + \sum_{j=3}^{m-1} \theta_j \epsilon_{t+3-j}] + (\theta_2 \epsilon_{t+1}) \\ &= \phi_3 s_{1,t} + 1 \cdot [s_{4,t}] + R_{3,1} \eta_t \end{aligned}$$

where

$$\begin{aligned} s_{4,t} &= \sum_{i=4}^{m} \phi_i y_{t+3-j} + \sum_{j=3}^{m-1} \theta_j \epsilon_{t+3-j} \\ R_{3,1} &= \theta_2 \\ \eta_t &= \epsilon_{t+1} \end{aligned}$$

- Continuing until

$$\begin{aligned} s_{m,t} &= \sum_{i=m}^{m} \phi_i y_{t+(m-1)-j} + \sum_{j=m-1}^{m-1} \theta_j \epsilon_{t+m-1-j} \\ &= \phi_m y_{t-1} + \theta_{m-1}\epsilon_t \end{aligned}$$

which gives

$$s_{m,t+1} = \phi_m y_t + \theta_{m-1}\epsilon_{t+1} = \phi_m s_{1,t} + R_{m,1}\eta_t$$

where $R_{m,1} = \theta_{m-1}$ and $\eta_t = \epsilon_{t+1}$

- All the equations can be written together:

$$\begin{aligned} s_{t+1} &= T s_t + R\eta_t \\ y_t &= Z s_t \quad \text{(no measurement error term)} \end{aligned}$$

where

$$T = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{m-1} & 0 & 0 & \cdots & 1 \\ \phi_m & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-2} \\ \theta_{m-1} \end{bmatrix} \text{ and}$$

η_t i.i.d. $N(0, \sigma_\epsilon^2)$, and $Z = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ $(1 \times m)$

Outline

1. Time Series Analysis III
   - Cointegration: Definitions
   - Cointegrated VAR Models: VECM Models
   - Estimation of Cointegrated VAR Models
   - Linear State-Space Models
   - Kalman Filter

Kalman Filter

Linear State-Space Model: Joint Equation

$$\begin{bmatrix} s_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} T_t \\ Z_t \end{bmatrix} s_t + \begin{bmatrix} R_t \eta_t \\ \epsilon_t \end{bmatrix}$$
$$= \Phi_t s_t + u_t,$$

where $\{\eta_t\}$ $i.i.d. N_n(0, Q_t)$, $\{\epsilon_t\}$ $i.i.d. N_k(0, H_t)$, so

$$u_t = \begin{bmatrix} R_t \eta_t \\ \epsilon_t \end{bmatrix} \sim N_{m+k}(0_{m+k}, \Omega_t), \text{ with } \Omega_t = \begin{bmatrix} R_t Q_t R_t^T & 0 \\ 0 & H_t \end{bmatrix}$$

For $\mathcal{F}_t = \{y_1, y_2, \ldots, y_t\}$, the observations up to time t, the **Kalman Filter** is the recursive computation of the probability density functions:

$$\begin{array}{ll} p(s_{t+1} \mid \mathcal{F}_t), & t = 1, 2, \ldots \\ p(s_{t+1}, y_{t+1} \mid \mathcal{F}_t), & t = 1, 2, \ldots \\ p(y_{t+1} \mid \mathcal{F}_t), & t = 1, 2, \ldots. \end{array}$$

Define $\Theta = \{$ all parameters in $T_t, Z_t, R_t, Q_t, H_T\}$

Kalman Filter

Notation:

- **Conditional Means**

$$\begin{aligned} s_{t|t} &= E(s_t \mid \mathcal{F}_t) \\ s_{t|t-1} &= E(s_t \mid \mathcal{F}_{t-1}) \\ y_{t|t-1} &= E(y_t \mid \mathcal{F}_{t-1}) \end{aligned}$$

- **Conditional Covariances / Mean-Squared Errors**

$$\begin{aligned} \Omega_s(t \mid t) &= Cov(s_t \mid \mathcal{F}_t) &= E[(s_t - s_{t|t})(s_t - s_{t|t})^T] \\ \Omega_s(t \mid t-1) &= Cov(s_t \mid \mathcal{F}_{t-1}) &= E[(s_t - s_{t|t-1})(s_t - s_{t|t-1})^T] \\ \Omega_y(t \mid t-1) &= Cov(y_t \mid \mathcal{F}_{t-1}) &= E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})^T] \end{aligned}$$

- **Observation Innovations /Residuals**

$$\tilde{\epsilon}_t = (y_t - y_{t|t-1}) = y_t - Z_t s_{t|t-1}$$

Kalman Filter: Four Steps

(1) Prediction Step: Predict state vector and observation vector at time t given $\mathcal{F}_{t-1}$

$$s_{t|t-1} = T_{t-1} s_{t-1|t-1}$$
$$y_{t|t-1} = Z_t s_{t|t-1}$$

Predictions are conditional means with mean-squared errors (MSEs):

$$\begin{aligned}\Omega_s(t \mid t-1) &= Cov(s_t \mid \mathcal{F}_{t-1}) = T_{t-1} Cov(s_{t-1|t-1}) T_{t-1}^T + \Omega_{R_t \eta_t} \\ &= T_t \Omega_s(t-1 \mid t-1) T_t^T + R_t Q_t R_t^T \\ \Omega_y(t \mid t-1) &= Cov(y_t \mid \mathcal{F}_{t-1}) = Z_t Cov(s_{t|t-1}) Z_t^T + \Omega_{\epsilon_t} \\ &= Z_t \Omega_s(t \mid t-1) Z_t^T + H_t\end{aligned}$$

Kalman Filter: Four Steps

(2) Correction / Filtering Step: Update the prediction of the state vector and its MSE given the observation at time t:

$$s_{t|t} = s_{t|t-1} + G_t(y_t - y_{t|t-1})$$
$$\Omega_s(t \mid t) = \Omega_s(t \mid t-1) - G_t \Omega_y(t \mid t-1) G_t^T$$

where

$$G_t = \Omega_s(t-1 \mid t) Z_t^T [\Omega_s(t-1 \mid t)]^{-1}$$

is the **Filter Gain matrix**.

(3) Forecasting Step: For times $t' > t$, the present step, use the following recursion equations for $t' = t+1, t+2, \ldots$

$$\begin{aligned}
s_{t'|t} &= T_{t'-1} s_{t'-1|t} \\
\Omega_s(t' \mid t) &= T_{t'-1} \Omega_s(t'-1 \mid t) T_{t'-1}^T + \Omega_{R_{t'}\eta_{t'}} \\
y_{t'|t} &= Z_{t'} s_{t'-1|t} \\
\Omega_y(t' \mid t) &= Z_{t'} \Omega_y(t'-1 \mid t) Z_{t'}^T + \Omega_{\epsilon_{t'}}
\end{aligned}$$

Kalman Filter: Four Steps

(4) Smoothing Step: Updating the predictions and MSEs for times $t' < t$ to use all the information in $\mathcal{F}_t$ rather than just $\mathcal{F}_{t'}$. Use the following recursion equations for $t' = t-1, t-2, \ldots$

$$\begin{array}{rcl} s_{t'|t} & = & s_{t|t} + S_{t'}(s_{t'+1|t} - s_{t'+1|t'}) \\ \Omega_s(t' \mid t) & = & \Omega_s(t' \mid t') - S_{t'}[\Omega_s(t'+1 \mid t') - \Omega_s(t'+1 \mid t)\ S_{t'}^T \end{array}$$

where

$$S_{t'} = \Omega_s(t' \mid t') T_{t'}^T [\Omega_s(t'+1 \mid t')]^{-1}$$

is the **Kalman Smoothing Matrix**.

Kalman Filter: Maximum Likelihood

Likelihood Function

Given $\theta = \{$ all parameters in $T_t, Z_t, R_t, Q_t, H_T\}$, we can write the likelihood function as:

$$L(\theta) = p(y_1, \ldots, y_T; \theta) = p(y_1; \theta)p(y_2 \mid y_1; \theta) \cdots p(y_T \mid y_1, \ldots, y_{T-1}; \theta)$$

Assuming the transition errors (η_t) and observation errors (ϵ_t) are Gaussian, the observations y_t have the following conditional normal distributions:

$$[y_t \mid \mathcal{F}_{t-1}; \theta] \sim N[y_{t|t-1}, \Omega_y(t \mid t-1)]$$

The log likelihood is:

$$\begin{aligned} l(\theta) &= log\ p(y_1, \ldots, y_T; \theta) \\ &= \textstyle\sum_{i=1}^{T} log\ p(y_i; \mathcal{F}_{t-1}; \theta) \\ &= \tfrac{-kT}{2} log(2\pi) - \tfrac{1}{2} \textstyle\sum_{t=1}^{T} log|\Omega_y(t \mid t-1)| \\ &\quad -\tfrac{1}{2} \textstyle\sum_{t=1}^{T} \left[(y_t - y_{t|t-1})'[\Omega_y(t \mid t-1)]^{-1}(y_t - y_{t|t-1})\right] \end{aligned}$$

Kalman Filter: Maximum Likelihood

Computing ML Estimates of θ

- The Kalman-Filter algorithm provides all terms necessary to compute the likelihood function for any θ.
- Methods for maximizing the log likelihood as a function of θ
  - EM Algorithm; see Dempster, Laird, and Rubin (1977).
  - Nonlinear optimization methods; e.g., Newton-type methods
  - For $T \to \infty$, the MLE $\hat{\theta}_T$ is
    - Consistent: $\theta_T \longrightarrow \theta$, true parameter.
    - Asymptotically normally distributed:
      $$\hat{\theta}_T - \theta \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathcal{I}_T^{-1})$$
      where
      $$\begin{aligned} \mathcal{I}_T &= E\left[(\tfrac{\partial}{\partial \theta} log\ L(\theta))(\tfrac{\partial}{\partial \theta} log\ L(\theta))^T \right] \\ &= (-1) \times E\left[(\tfrac{\partial^2}{\partial \theta \partial \theta^T} log\ L(\theta) \right] \end{aligned}$$
      is the **Fisher Information Matrix** for θ

Kalman Filter

Note:

- Under Gaussian assumptions, all state variables and observation variables are jointly Gaussian, so the Kalman-Filter recursions provide a complete specification of the model.
- Initial state vector s_1 is modeled as $N(\mu_{S_1}, \Omega_s(1))$, where the mean and covariance parameters are pre-specified. Choices depend on the application and can reflect *diffuse* (uncertain) initial information, or ergodic information (i.e., representing the long-run stationary distribution of state variables).
- Under covariance stationary assumptions for the $\{\eta_t\}$ and $\{\epsilon_t\}$ processes, the recursion expressions are still valid for the conditional means/covariances.