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 $\{X_t\} = \{\ldots, X_{t-1}, X_t, \ldots\}$ is $I(d)$, **Integrated of order $d$** if the $d$-differenced process

$$\Delta^d X_t = (1 - L)^d X_t$$

is stationary.

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

$$\Phi(L)X_t = \epsilon_t$$

where $\Phi(L) = I - A_1L - A_2 - \cdots - A_pL^p$.

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 \( \{X_t\} \) where \( X_t = (x_{1,t}, x_{2,t}, \ldots, x_{m,t})' \) 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 \( \beta = (\beta_1, \beta_2, \ldots, \beta_m)' \) such that

\[
\beta'X_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_m x_{m,t} \sim I(0),
\]

a stationary process.

- The cointegration vector \( \beta \) can be scaled arbitrarily, so assume a normalization:
  \[ \beta = (1, \beta_2, \ldots, \beta_m)' \]
- The expression: \( \beta'X_t = u_t \), 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
- \( \beta'X_t \) 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

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

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

$$X_t = C + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \eta_t$$

where

- $C = (c_1, c_2, \ldots, c_m)'$ is an $m$-vector of constants.
- $\Phi_1, \Phi_2, \ldots, \Phi_p$ are $(m \times m)$ matrices of coefficients
- $\{\eta_t\}$ is multivariate white noise $MWN(0_m, \Sigma)$

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

$$\det [I_m - (\Phi_1 z + \Phi_2 z^2 + \cdots + \Phi_p z^p)] = 0$$

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

Suppose $\{X_t\}$ 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

\[ X_t = C + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \eta_t \]

- Subtract \( X_{t-1} \) from both sides:

\[ \Delta X_t = X_t - X_{t-1} = C + (\Phi_1 - I_m)X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \eta_t \]

- Subtract and add \( (\Phi_1 - I_m)X_{t-2} \) from right-hand side:

\[ \Delta X_t = C + (\Phi_1 - I_m)X_{t-1} + (\Phi_2 + \Phi_1 - I_m)X_{t-2} + \cdots + \Phi_p X_{t-p} + \eta_t \]

- Subtract and add \( (\Phi_2 + \Phi_1 - I_m)X_{t-3} \) from right-hand side:

\[ \Delta X_t = C + (\Phi_1 - I_m)X_{t-1} + (\Phi_2 + \Phi_1 - I_m)X_{t-2} + \cdots + (\Phi_3 + \Phi_2 + \Phi_1 - I_m)X_{t-3} + \cdots \]

\[ \implies \Delta X_t = C + (\Phi_1 - I_m)\Delta X_{t-1} + (\Phi_2 + \Phi_1 - I_m)\Delta X_{t-2} + \cdots + (\Phi_{p-1} + \cdots + \Phi_3 + \Phi_2 + \Phi_1 - I_m)\Delta X_{t-(p-1)} + (\Phi_p + \cdots + \Phi_3 + \Phi_2 + \Phi_1 - I_m)X_{t-p} + \eta_t \]

Reversing the order of incorporating \( \Delta \)-terms we can derive

\[ \Delta X_t = C + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-(p-1)} + \eta_t \]

where: \( \Pi = (\Phi_1 + \Phi_2 + \cdots + \Phi_p - I_m) \) and \( \Gamma_k = (- \sum_{j=k+1}^{p} \Phi_j) \).
Vector Error Correction Model (VECM)

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

$$\Delta X_t = C + \Pi X_t + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{p-1} \Delta X_{t-(p-1)} + \eta_t$$

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

- The left-hand-side $\Delta X_t$ is stationary / I(0).
- The terms on the right-hand-side $\Delta X_{t-j}$, $j = 1, 2, \ldots, p - 1$ are stationary / I(0).
- The term $\Pi X_t$ must be stationary / I(0).
- This term $\Pi X_t$ contains any cointegrating terms of $\{X_t\}$.
- Given that the VAR($p$) process had unit roots, it must be that $\Pi$ is singular, i.e., the linear transformation eliminates the unit roots.
The matrix \( \Pi \) is of reduced rank \( r < m \) and either

- \( \text{rank}(\Pi) = 0 \) and \( \Pi = 0 \) and there are no cointegrating relationships.
- \( \text{rank}(\Pi) > 0 \) and \( \Pi \) defines the cointegrating relationships.

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

\[
\Pi = \alpha \beta',
\]

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

- The columns of \( \beta \) define linearly independent vectors which cointegrate \( X_t \).
- The decomposition of \( \Pi \) is not unique. For any invertible \((r \times r)\) matrix \( G \),

\[
\Pi = \alpha \ast \beta_*
\]

where \( \alpha \ast = \alpha G \) and \( \beta_* = G^{-1} \beta' \).
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 $\Pi$) hold asymptotically.

Maximum Likelihood Estimation*


* Advanced topic for optional reading/study
Maximum Likelihood Estimation (continued)


- This methodology provides likelihood ratio tests for the number of cointegrating vectors:
  - *Johansen’s Trace Statistic* (sum of eigenvalues of $\hat{\Pi}$)
  - *Johansen’s Maximum-Eigenvalue Statistic* (max eigenvalue of $\hat{\Pi}$).
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**

\[ y_t: \ (k \times 1) \text{ observation vector at time } t \]

\[ s_t: \ (m \times 1) \text{ state-vector at time } t \]

\[ \epsilon_t: \ (k \times 1) \text{ observation-error vector at time } t \]

\[ \eta_t: \ (n \times 1) \text{ state transition innovation/error vector} \]

**State Equation / Transition Equation**

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

where

\[ T_t: \ (m \times m) \text{ transition coefficients matrix} \]

\[ R_t: \ (m \times n) \text{ fixed matrix; often column(s) of } I_p \]

\[ \eta_t: \text{ i.i.d. } N(0_n, Q_t), \text{ where } Q_t \ (n \times n) \text{ is positive definite.} \]
Linear State-Space Model Formulation

**Observation Equation / Measurement Equation**

\[ y_t = Z_t s_t + \epsilon_t \]

Where

\[ Z_t : (k \times m) \text{ observation coefficients matrix} \]
\[ \epsilon_t: \text{ i.i.d. } N(0, H_t), \text{ where } H_t (k \times k) \text{ is positive definite.} \]

**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 \((T_t, R_t, Z_t, Q_t, H_t \text{ constants})\)
Consider the CAPM Model with time-varying parameters:

\[
\begin{align*}
    r_t &= \alpha_t + \beta_t r_{m,t} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2_\epsilon) \\
    \alpha_{t+1} &= \alpha_t + \nu_t, \quad \nu_t \sim N(0, \sigma^2_\nu) \\
    \beta_{t+1} &= \beta_t + \xi_t, \quad \xi_t \sim N(0, \sigma^2_\xi)
\end{align*}
\]

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 \( N(0, \sigma^2_\nu) \)
- \( \{\beta_t\} \) is a Random Walk with i.i.d. steps \( N(0, \sigma^2_\xi) \)
  (Mutually independent processes)
Time-Varying CAPM Model: Linear State-Space Model

**State Equation**

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

Equivalently:

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

where:

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

\[\eta_t = \begin{bmatrix}
\nu_t \\
\xi_t
\end{bmatrix} \sim N_2(0_2, Q_t), \quad \text{with } Q_t = \begin{bmatrix}
\sigma^2 & 0 \\
0 & \sigma^2
\end{bmatrix}\]

Terms:

- state vector \(s_t\), transition coefficients \(T_t\)
- transition white noise \(\eta_t\)
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

\[ r_t = Z_ts_t + \epsilon_t \]

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 \( H_t = \sigma^2 \).

Joint System of Equations

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

\[ = \Phi_ts_t + u_t \text{ with } \text{Cov}(u_t) = \begin{bmatrix} R_t \Omega_\eta R_t^T & 0 \\ 0 & H_t \end{bmatrix} \]
Consider a normal linear regression model with time-varying regression coefficients:

\[ y_t = x_t^T \beta_t + \epsilon_t, \text{ where } \epsilon_t \text{ are i.i.d. } N(0, \sigma^2_\epsilon). \]

where

\[ x_t = (x_{1,t}, x_{2,t}, \ldots, x_{p,t})^t, \text{ } p\text{-vector of explanatory variables} \]

\[ \beta_t = (\beta_{1,t}, \beta_{2,t}, \ldots, \beta_{p,t})^t, \text{ regression parameter vector} \]

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

\[ \beta_{j,t+1} = \beta_{j,t} + \eta_{j,t}, \text{ with } \{\eta_{j,t}, t = 1, 2, \ldots\} \text{ i.i.d. } N(0, \sigma^2_j). \]

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

**Joint State-Space Equations**

\[
\begin{bmatrix}
  s_{t+1} \\
  y_t
\end{bmatrix} =
\begin{bmatrix}
  I_p \\
  x_t^T \\
  T_t \\
  Z_t
\end{bmatrix}
\begin{bmatrix}
  s_t \\
  \epsilon_t
\end{bmatrix} +
\begin{bmatrix}
  \eta_t \\
  \epsilon_t
\end{bmatrix}, \text{ with state vector } s_t = \beta_t
\]
(Time-Varying) Linear Regression as a State-Space Model

where

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

**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.
Consider the AR\((p)\) model
\[
\phi(L)y_t = \epsilon_t
\]
where
\[
\phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j
\]
and \(\{\epsilon_t\}\) i.i.d. \(N(0, \sigma^2_\epsilon)\).

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
\[
\begin{bmatrix}
y_{t+1} \\
y_t \\
y_{t-1} \\
\vdots \\
y_{t-(p-2)}
\end{bmatrix} =
\begin{bmatrix}
\phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\
1 & 0 & \ldots & 0 & 0 \\
0 & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & 1 & 0
\end{bmatrix}
\begin{bmatrix}
y_t \\
y_{t-1} \\
\vdots \\
y_{t-(p-1)}
\end{bmatrix} +
\begin{bmatrix}
\epsilon_{t+1} \\
0 \\
\vdots \\
0
\end{bmatrix}
\]
State-Space Model for $AR(p)$

**State Equation**

\[ s_{t+1} = T_s s_t + R_t \eta_t, \]

where

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

and

\[ \{ \eta_t = \epsilon_{t+1} \} \text{ i.i.d. } N(0, \sigma^2_\epsilon). \]

**Observation Equation / Measurement Equation**

\[
y_t = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0
\end{bmatrix} s_t
= Z_t s_t \text{ (no measurement error)}
\]
Moving Average Model $MA(q)$

Consider the $MA(q)$ model

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

where

$$\theta(L) = 1 + \sum_{j=1}^{q} \theta_j L^j$$

and $\{\epsilon_t\}$ 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

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

and

\[ \{ \eta_t = \epsilon_t \} \text{ i.i.d. } N(0, \sigma^2_\epsilon). \]

**Observation Equation / Measurement Equation**

\[
y_t = \begin{bmatrix}
\theta_1 & \theta_2 & \cdots & \theta_{q-1} & \theta_q \\
\end{bmatrix} s_t + \epsilon_t \\
= Z_t s_t + \epsilon_t
\]
Consider the ARMA\((p, q)\) model
\[
\phi(L)y_t = \theta(L)\epsilon_t
\]
where
\[
\phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j \quad \text{and} \quad \{\epsilon_t\} \text{ i.i.d. } N(0, \sigma^2_{\epsilon}).
\]
\[
\theta(L) = 1 + \sum_{j=1}^{q} \theta_j L^j \quad \text{and} \quad \{\epsilon_t\} \text{ i.i.d. } N(0, \sigma^2_{\epsilon}).
\]
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
\[
\{\phi_1, \ldots, \phi_m\} : \quad \phi(L) = 1 - \sum_{j=1}^{m} \phi_j L^j
\]
\[
\{\theta_1, \ldots, \theta_m\} : \quad \theta(L) = 1 - \sum_{j=1}^{m} \theta_j L^j
\]
i.e.,
\[
\phi_j = 0 \text{ if } p < j \leq m
\]
\[
\theta_j = 0 \text{ if } q < j \leq m
\]
So: \(\{y_t\} \sim ARMA(m, m - 1)\)
Harvey (1993) State-Space Specification

Define state vector:

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

recursively:

\[ s_{1,t} = y_t : \text{ Use this definition and the main model equation to define } s_{2,t} \text{ and } \eta_t: \]

\[
\begin{align*}
y_{t+1} &= \sum_{j=1}^{p} \phi_j y_{t+1-j} + \epsilon_{t+1} + \sum_{j=1}^{q} \theta_j \epsilon_{t+1-j} \\
s_{1,t+1} &= \phi_1 s_{1,t} + 1 \cdot s_{2,t} + \eta_t
\end{align*}
\]

where

\[
\begin{align*}
s_{2,t} &= \sum_{i=2}^{m} \phi_i y_{t+1-i} + \sum_{j=1}^{m-1} \theta_j \epsilon_{t+1-j} \\
\eta_t &= \epsilon_{t+1}
\end{align*}
\]
• Use \( s_{1,t} = y_t \), \( s_{2,t} \), and \( \eta_t = \epsilon_{t+1} \): to define \( s_{3,t} \)

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

where

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

• Use \( s_{1,t} = y_t \), \( s_{3,t} \), and \( \eta_t = \epsilon_{t+1} \): to define \( s_{4,t} \)

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

where

\[
\begin{align*}
    s_{4,t} &= \sum_{i=4}^{m} \phi_i y_{t+3-j} + \sum_{j=3}^{m-1} \theta_j \epsilon_{t+3-j} \\
    R_{3,1} &= \theta_2 \\
    \eta_t &= \epsilon_{t+1}
\end{align*}
\]
Continuing until

\[ s_{m,t} = \sum_{i=m}^{m} \phi_i y_{t+(m-1)-j} + \sum_{j=m-1}^{m-1} \theta_j \epsilon_{t+m-1-j} \]

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:

\[ s_{t+1} = Ts_t + R\eta_t \]
\[ y_t = Zs_t \text{ (no measurement error term)} \]

where

\[ T = \begin{bmatrix}
\phi_1 & 1 & 0 & \cdots & 0 \\
\phi_2 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\phi_{m-1} & 0 & 0 & \cdots & 1 \\
\phi_m & 0 & 0 & \cdots & 0 
\end{bmatrix}, \quad R = \begin{bmatrix} 1 \\
\theta_1 \\
\vdots \\
\theta_{m-2} \\
\theta_{m-1} \end{bmatrix} \text{ and} \]
\[ \eta_t \text{ i.i.d. } \mathcal{N}(0, \sigma^2_\epsilon), \text{ and } Z = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} (1 \times m) \]
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:

\[
p(s_{t+1} \mid \mathcal{F}_t), \quad t = 1, 2, \ldots
\]
\[
p(s_{t+1}, y_{t+1} \mid \mathcal{F}_t), \quad t = 1, 2, \ldots
\]
\[
p(y_{t+1} \mid \mathcal{F}_t), \quad t = 1, 2, \ldots
\]

Define \( \Theta = \left\{ \text{all parameters in } T_t, Z_t, R_t, Q_t, H_T \right\} \).
Kalman Filter

Notation:

- **Conditional Means**
  
  \[ s_{t|t} = E(s_t \mid \mathcal{F}_t) \]
  
  \[ s_{t|t-1} = E(s_t \mid \mathcal{F}_{t-1}) \]
  
  \[ y_{t|t-1} = E(y_t \mid \mathcal{F}_{t-1}) \]

- **Conditional Covariances / Mean-Squared Errors**
  
  \[ \Omega_s(t \mid t) = \text{Cov}(s_t \mid \mathcal{F}_t) = E[(s_t - s_{t|t})(s_t - s_{t|t})^T] \]
  
  \[ \Omega_s(t \mid t-1) = \text{Cov}(s_t \mid \mathcal{F}_{t-1}) = E[(s_t - s_{t|t-1})(s_t - s_{t|t-1})^T] \]
  
  \[ \Omega_y(t \mid t-1) = \text{Cov}(y_t \mid \mathcal{F}_{t-1}) = E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})^T] \]

- **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 $\mathcal{F}_{t-1}$

\[
\begin{align*}
    s_{t|t-1} &= T_{t-1}s_{t-1|t-1} \\
    y_{t|t-1} &= Z_{t} s_{t|t-1}
\end{align*}
\]

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

\[
\begin{align*}
    \Omega_s(t \mid t-1) &= \text{Cov}(s_t \mid \mathcal{F}_{t-1}) = T_{t-1} \text{Cov}(s_{t-1|t-1}) T_{t-1}^T + \Omega_{R_t \eta_t} \\
    &= T_{t} \Omega_s(t-1 \mid t-1) T_{t}^T + R_{t} Q_{t} R_{t}^T \\
    \Omega_y(t \mid t-1) &= \text{Cov}(y_t \mid \mathcal{F}_{t-1}) = Z_{t} \text{Cov}(s_{t|t-1}) Z_{t}^T + \Omega_{\epsilon_t} \\
    &= Z_{t} \Omega_s(t \mid t-1) Z_{t}^T + H_{t}
\end{align*}
\]
Kalman Filter: Four Steps

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

$$s_{t|t} = s_{t|t-1} + G_t (y_{t} - y_{t|t-1})$$

$$\Omega_s(t | t) = \Omega_s(t | t - 1) - G_t \Omega_y(t | t - 1)G_t^T$$

where

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

is the Filter Gain matrix.

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

$$s_{t'|t} = T_{t'-1}s_{t'-1|t}$$

$$\Omega_s(t' | t) = T_{t'-1}\Omega_s(t' - 1 | t)T_{t'-1}^T + \Omega_{R_{t'|t'\eta}}$$

$$y_{t'|t} = Z_{t'}s_{t'-1|t}$$

$$\Omega_y(t' | t) = Z_{t'}\Omega_y(t' - 1 | t)Z_{t'}^T + \Omega_{\epsilon_{t'}}$$
(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 \( t' = t - 1, t - 2, \ldots \):

\[
\begin{align*}
S_{t'}|t & = s_{t|t} + S_{t'}(s_{t'+1|t} - s_{t'+1|t'}) \\
\Omega_s(t' | t) & = \Omega_s(t' | t') - S_{t'}[\Omega_s(t' + 1 | t') - \Omega_s(t' + 1 | t)] S_{t'}^T
\end{align*}
\]

where

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

is the **Kalman Smoothing Matrix**.
Likelihood Function

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

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

Assuming the transition errors ($\eta_t$) and observation errors ($\epsilon_t$) are Gaussian, the observations $y_t$ have the following conditional normal distributions:

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

The log likelihood is:

$$l(\theta) = \log p(y_1, \ldots, y_T; \theta)$$

$$= \sum_{i=1}^{T} \log p(y_i; \mathcal{F}_{t-1}; \theta)$$

$$= -\frac{kT}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log |\Omega_y(t | t - 1)|$$

$$- \frac{1}{2} \sum_{t=1}^{T} [(y_t - y_{t|t-1})'(\Omega_y(t | t - 1))^{-1}(y_t - y_{t|t-1})]$$
Kalman Filter: Maximum Likelihood

Computing ML Estimates of $\theta$

- The Kalman-Filter algorithm provides all terms necessary to compute the likelihood function for any $\theta$.
- Methods for maximizing the log likelihood as a function of $\theta$
  - 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:
      $$\hat{\theta}_T - \theta \xrightarrow{D} \mathcal{N}(0, I^{-1}_T)$$
      where
      $$I_T = E \left[ (\frac{\partial}{\partial \theta} \log L(\theta))(\frac{\partial}{\partial \theta} \log L(\theta))^T \right]$$
      $$= (-1) \times E \left[ (\frac{\partial^2}{\partial \theta \partial \theta^T} \log L(\theta)) \right]$$
      is the Fisher Information Matrix for $\theta$
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.