

5. Multivariate Time Series and State-Space Models

Univariate methods of time series analysis can be extended to study parallel series.

Data may comprise

- stocks in a sector
- indices
- exchange rates
- macro-economic factors
- prices of bonds of different maturities

all which may exhibit evolution in time in some *dependent* fashion.

5.1 Vector Time Series

Stationarity and Correlation

Let $\mathbf{X}_t = (X_{1t}, \dots, X_{Kt})^\top$ be an observation in a K -dimensional time series: $\{\mathbf{X}_t\}$ is weakly stationary if

$$\mathbb{E}[\mathbf{X}_t] = \boldsymbol{\mu} \qquad \text{Var}[\mathbf{X}_t] = \mathbb{E}[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})^\top] = \boldsymbol{\Gamma}_0$$

and $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}_0$ are a constant $(K \times 1)$ vector and $(K \times K)$ matrix respectively.

As usual, the diagonal elements of $\boldsymbol{\Gamma}_0$ give the variances of the K elements of the time series. They will be denoted

$$\Gamma_{11}(0), \dots, \Gamma_{KK}(0)$$

Cross-Correlation

Let D be a $(K \times K)$ diagonal matrix

$$D = \text{diag}(\sqrt{\Gamma_{11}(0)}, \dots, \sqrt{\Gamma_{KK}(0)})$$

The lag-zero cross-correlation matrix $(K \times K)$ is

$$\boldsymbol{\rho}_0 = D^{-1} \boldsymbol{\Gamma}_0 D^{-1}$$

with (i, j) th element

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}}$$

This matrix gives the instantaneous correlation between the elements of the vector.

Cross-Correlation

Now consider the correlation **across** series at non-zero lags: let

$$\mathbf{\Gamma}_h = \mathbb{E}[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_{t-h} - \boldsymbol{\mu})^\top]$$

be the lag- h cross-covariance matrix, and define the lag- h cross-correlation matrix as

$$\boldsymbol{\rho}_h = \mathbf{D}^{-1}\mathbf{\Gamma}_h\mathbf{D}^{-1}$$

with (i, j) th element

$$\rho_{ij}(h) = \frac{\Gamma_{ij}(h)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}}$$

This matrix gives the instantaneous correlation between the elements of the vector.

Cross-Correlation

- ▶ $\rho_{ij}(h)$ measures the (linear) dependence of X_{it} on $X_{j,t-h}$ which occurred prior to time t .
- ▶ $\rho_{ii}(h)$ is the lag- h autocorrelation for $\{X_{it}\}$.
- ▶ Note that in general

$$\rho_{ij}(h) \neq \rho_{ji}(h)$$

so Γ_h and ρ_h are not symmetric, but

$$\Gamma_{ij}(h) = \Gamma_{ji}(-h)$$

$$\text{so } \Gamma_h = \Gamma_{-h}^\top.$$

Thus for $h > 0$, ρ_{ij} measures the linear dependence of X_{it} on $X_{j,t-h}$.

Cross-Correlation: Lead-lag relationships

In summary

- ▶ if $\rho_{ij}(h) = \rho_{ji}(h) = 0$ for all $h > 0$, then $\{X_{it}\}$ and $\{X_{jt}\}$ have **no linear relationship**.
- ▶ if $\rho_{ij}(0) \neq 0$ then $\{X_{it}\}$ and $\{X_{jt}\}$ are **concurrently correlated**.
- ▶ if $\rho_{ij}(h) = 0$ and $\rho_{ji}(h) = 0$ for all $h > 0$, then $\{X_{it}\}$ and $\{X_{jt}\}$ have no **lead-lag** relationship.
- ▶ if $\rho_{ij}(h) = 0$ for all $h > 0$ but $\rho_{ji}(h) \neq 0$ for some $h > 0$, then $\{X_{it}\}$ does **not** depend on past values of $\{X_{jt}\}$, but $\{X_{jt}\}$ **does** depend on past values of $\{X_{it}\}$.
- ▶ if $\rho_{ij}(h) \neq 0$ for some $h > 0$ and $\rho_{ji}(h) \neq 0$ for some $h > 0$, then $\{X_{it}\}$ and $\{X_{jt}\}$ exhibit a **feedback relationship**.

Sample cross-correlation

We estimate the cross-correlation matrices by the usual moment estimators: for $h > 0$,

$$\hat{\mathbf{\Gamma}}_h = \frac{1}{n} \sum_{t=h+1}^n (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_{t+h} - \bar{\mathbf{x}})^\top$$

where

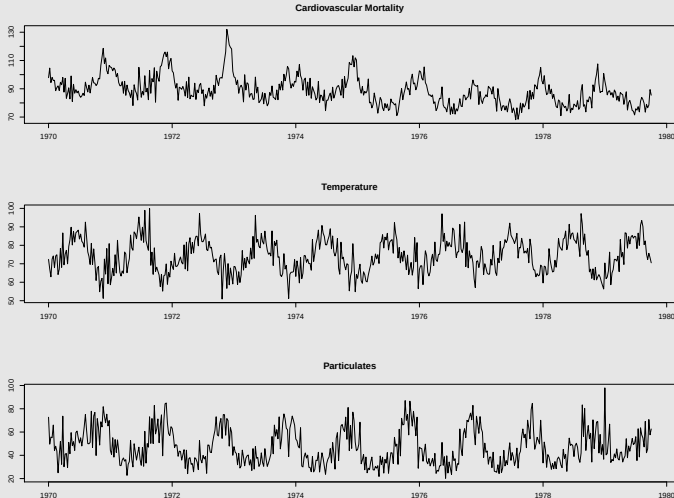
$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}_t$$

and

$$\hat{\boldsymbol{\rho}}_h = \hat{D}^{-1} \hat{\mathbf{\Gamma}}_h \hat{D}^{-1}$$

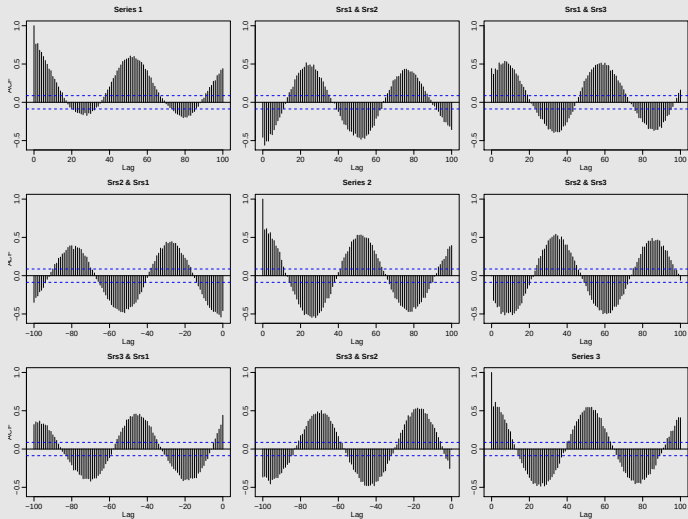
Sample cross-correlation

Example (Cardiac Mortality Data)



Sample cross-correlation

Example (Cardiac Mortality Data)



Vector Autoregression Models

A simple model for $\mathbf{X}_t$ takes the form

$$\mathbf{X}_t = \phi_0 + \sum_{j=1}^p \Phi_j \mathbf{X}_{t-j} + \mathbf{Z}_t$$

where

- ϕ_0 is a $K \times 1$ vector
- Φ_j is a $K \times K$ matrix
- $\mathbf{Z}_t$ is zero mean vector white noise process with

$$\mathbb{E}[\mathbf{Z}_t \mathbf{Z}_t^T] = \Sigma$$

$\mathbf{X}_t$ follows a **vector autoregression** (VAR) model of order p .

Vector Autoregression Models

Usually

$$\mathbf{Z}_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

is a suitable assumption. In this case, the resulting process is a multivariate Gaussian process.

Estimation of the model can proceed using the usual likelihood methods.

Restrictions on the model are required to preserve stationarity.

Example: VAR(1)

Consider the model

$$\begin{aligned}X_{1t} &= \phi_{10} + \Phi_{11}X_{1,t-1} + \Phi_{12}X_{2,t-1} + Z_{1t} \\X_{2t} &= \phi_{20} + \Phi_{21}X_{1,t-1} + \Phi_{22}X_{2,t-1} + Z_{2t}\end{aligned}$$

Then the elements of

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$$

define the dependence of current elements of X_t on past elements.

Elements of the matrix Σ define the concurrent relationship between X_{1t} and X_{2t} . This representation is termed the **reduced form** equation.

Example: VAR(1)

Let

$$\Sigma = LGL^T$$

be a Cholesky decomposition where G is a diagonal matrix, and L is a lower-triangular matrix. Define $\mathbf{U}_t = L^{-1}\mathbf{Z}_t$. Then

$$\mathbb{E}[\mathbf{U}_t] = \mathbf{0} \qquad \mathbb{V}ar[\mathbf{U}_t] = G$$

so that the elements of $\mathbf{U}_t$ are uncorrelated. Then

$$L^{-1}\mathbf{X}_t = L^{-1}\phi_0 + L^{-1}\Phi\mathbf{X}_{t-1} + L^{-1}\mathbf{Z}_t$$

so that

$$L^{-1}\mathbf{X}_t = \phi^* + \Phi^*\mathbf{X}_{t-1} + \mathbf{U}_t$$

say.

Example: VAR(1)

Thus, as L is lower-triangular, the K th row of this vector equation is

$$X_{Kt} + \sum_{k=1}^{K-1} L_{Kk} X_{kt} = \phi_{K0}^* + \sum_{k=1}^K \phi_{Kk}^* X_{k,t-1} + U_{Kt}$$

As the elements of $\mathbf{U}_t$ are uncorrelated, this equation shows explicitly the concurrent dependence of X_{Kt} on X_{kt} , for $k = 1, \dots, K - 1$.

This equation is termed the **structural equation** for X_{Kt} . A similar expression can be found for each of the elements of $\mathbf{X}_t$ by reordering.

Example: $K = 2$

Let

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix}$$

with

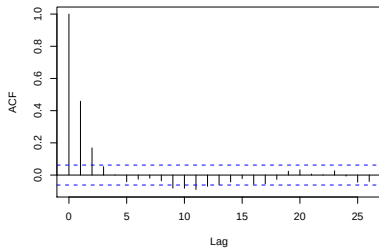
$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

Then

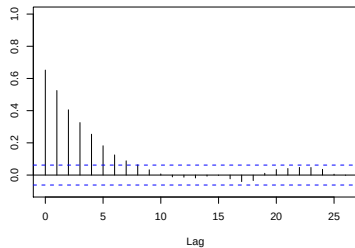
$$L^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix} \quad G = L^{-1}\Sigma(L^{\top})^{-1} = \begin{bmatrix} 2.0 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$$

Example: $K = 2$ - cross-correlations

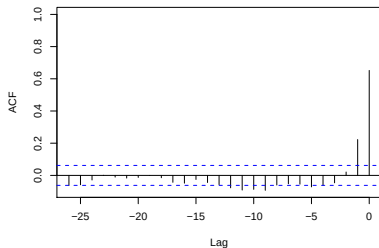
Series 1



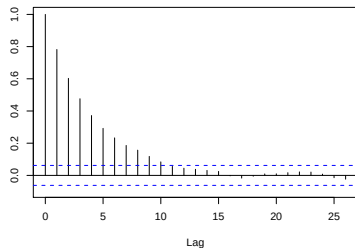
Series 1 & Series 2



Series 2 & Series 1



Series 2



Example: VAR(1)

Using the above structural equation formulation:

$$X_{2t} = 0.3 + 0.5X_{1t} - 0.7X_{1,t-1} + 0.95X_{2,t-1} + U_{2t}$$

and

$$X_{1t} = -0.2 + 1.0X_{2t} - 0.8X_{2,t-1} + 0.8X_{1,t-1} + V_{2t}$$

where $U_{2t} \sim \mathcal{N}(0, 0.5)$ and $V_{2t} \sim \mathcal{N}(0, 1)$ are independent random variables.

These equations demonstrate the concurrent linear dependence between X_{1t} and X_{2t} .

Stationarity for VAR(1)

It follows from

$$\mathbf{X}_t = \phi_0 + \Phi_1 \mathbf{X}_{t-1} + \mathbf{Z}_t$$

that, taking expectations, from stationarity

$$\mu = \phi_0 + \Phi_1 \mu \quad \therefore \quad \mu = (\mathbf{I}_K - \Phi_1)^{-1} \phi_0$$

provided $(\mathbf{I}_K - \Phi_1)$ is non-singular. Then

$$\mathbf{X}_t - \mu = \Phi_1(\mathbf{X}_{t-1} - \mu) + \mathbf{Z}_t$$

If $\mathbf{Y}_t = \mathbf{X}_t - \mu$, then by recursion

$$\mathbf{Y}_t = \mathbf{Z}_t + \Phi_1 \mathbf{Z}_{t-1} + \Phi_1^2 \mathbf{Z}_{t-2} + \cdots$$

and hence $\text{Cov}[\mathbf{Z}_t, \mathbf{Y}_{t-h}] = \text{Cov}[\mathbf{Z}_t, \mathbf{X}_{t-h}] = \mathbf{0}$, for all $h > 0$.

Stationarity for VAR(1)

In addition,

$$\mathbf{Y}_t \mathbf{Z}_t^\top = \mathbf{Z}_t \mathbf{Z}_t^\top + \Phi_1 \mathbf{Z}_{t-1} \mathbf{Z}_t^\top + \Phi_1^2 \mathbf{Z}_{t-2} \mathbf{Z}_t^\top + \cdots$$

and taking expectations yields

$$\text{Cov}[\mathbf{Y}_t, \mathbf{Z}_t] = \mathbb{E}[\mathbf{Y}_t \mathbf{Z}_t^\top] = \mathbb{E}[\mathbf{Z}_t \mathbf{Z}_t^\top] = \Sigma$$

as $\{\mathbf{Z}_t\}$ is a white-noise process.

In the representation, $\mathbf{X}_t$ depends on $\mathbf{Z}_{t-j}$ with coefficient matrix Φ_1^j ; this dependence must converge to zero. Therefore, using the eigenvalue decomposition

$$\Phi_1 = \mathbf{A}_1 \Lambda_1 \mathbf{B}_1$$

with $\mathbf{A}_1 \mathbf{B}_1 = \mathbf{I}_K$ and $\Lambda_1 = \text{diag}(\lambda_{11}, \dots, \lambda_{1K})$, we have in the usual way $\Phi_1^j = \mathbf{A}_1 \Lambda_1^j \mathbf{B}_1$.

Stationarity for VAR(1)

Thus

$$\Phi_1^k \longrightarrow \mathbf{0} \quad \Longleftrightarrow \quad |\lambda_{1k}| < 1, \quad k = 1, \dots, K$$

as in the univariate case. Note that

$$\det(\lambda \mathbf{I}_K - \Phi_1) = \lambda^K \det(\mathbf{I}_K - \lambda^{-1} \Phi_1)$$

and hence $\lambda_{11}, \dots, \lambda_{1K}$ are the solutions of

$$\det(\mathbf{I}_K - \Phi_1 z) = 0$$

and as before, $\mathbf{X}_t$ is stationary if and only if the zeroes of $\det \Phi_1(z)$ are greater than 1 in modulus.

Finally

$$\Gamma_0 = \text{Var}[\mathbf{X}_t] = \text{Var}[\mathbf{Z}_t] + \sum_{j=1}^{\infty} \text{Var}[\Phi_1^j \mathbf{Z}_{t-1}] = \Sigma + \sum_{j=1}^{\infty} \Phi_1^j \Sigma (\Phi_1^T)^j$$

Stationarity for VAR(1)

Using the equation

$$\mathbf{X}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}_1(\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \mathbf{Z}_t$$

we have for $h > 0$,

$$\mathbb{E}[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_{t-h} - \boldsymbol{\mu})^\top] = \boldsymbol{\Phi}_1 \mathbb{E}[(\mathbf{X}_{t-1} - \boldsymbol{\mu})(\mathbf{X}_{t-h} - \boldsymbol{\mu})^\top] + \mathbb{E}[\mathbf{Z}_t(\mathbf{X}_{t-h} - \boldsymbol{\mu})^\top]$$

or equivalently, as the final expectation is zero

$$\boldsymbol{\Gamma}_h = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_{h-1} = \boldsymbol{\Phi}_1^h \boldsymbol{\Gamma}_0$$

by recursion. Thus

$$\begin{aligned} \boldsymbol{\rho}_h &= D^{-1/2} \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_{h-1} D^{-1/2} = \left\{ D^{-1/2} \boldsymbol{\Phi}_1 D^{1/2} \right\} \left\{ D^{-1/2} \boldsymbol{\Gamma}_{h-1} D^{-1/2} \right\} \\ &= \boldsymbol{\Upsilon}_1 \boldsymbol{\rho}_{h-1} = \cdots = \boldsymbol{\Upsilon}_1^h \boldsymbol{\rho}_0 \end{aligned}$$

Results for VAR(p) models

We have

$$\mathbf{X}_t - \boldsymbol{\mu} = \sum_{j=1}^p \boldsymbol{\Phi}_j (\mathbf{X}_{t-j} - \boldsymbol{\mu}) + \mathbf{Z}_t$$

and

$$\text{Cov}[\mathbf{X}_t, \mathbf{Z}_t] = \Sigma \qquad \text{Cov}[\mathbf{X}_{t-h}, \mathbf{Z}_t] = \mathbf{0}, \quad h > 0$$

Also, for $h > 0$,

$$\boldsymbol{\Gamma}_h = \sum_{j=1}^p \boldsymbol{\Phi}_j \boldsymbol{\Gamma}_{h-j} \qquad \boldsymbol{\rho}_h = \sum_{j=1}^p \boldsymbol{\Upsilon}_j \boldsymbol{\rho}_{h-j}$$

where $\boldsymbol{\Upsilon}_j = D^{-1/2} \boldsymbol{\Phi}_j D^{-1/2}$

Results for VAR(p) models

As for the VAR(1),

$$\mathbf{X}_t = \boldsymbol{\mu} + \mathbf{Z}_t + \boldsymbol{\Psi}_1 \mathbf{Z}_{t-1} + \boldsymbol{\Psi}_2 \mathbf{Z}_{t-2} + \cdots$$

by the usual moving average type construction, where

$$(\mathbf{I}_K - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)(\mathbf{I}_K + \boldsymbol{\Psi}_1 B + \boldsymbol{\Psi}_2 B^2 + \cdots) = \mathbf{I}_K$$

The matrix $\boldsymbol{\Psi}_j$ is the effect of $\mathbf{Z}_t$ on $\mathbf{X}_{t+j}$; it is termed the **impulse response function** of $\mathbf{X}_t$.

The same trick to orthogonalize the system by writing

$$\Sigma = LGL^\top$$

and pre-multiplying by L^{-1} can be used.

Vector MA Models

Let

$$\mathbf{X}_t = \boldsymbol{\theta}_0 + \mathbf{Z}_t + \boldsymbol{\Theta}_1 \mathbf{Z}_{t-1} + \cdots + \boldsymbol{\Theta}_q \mathbf{Z}_{t-q} = \boldsymbol{\theta}_0 + \boldsymbol{\Theta}(B) \mathbf{Z}_t$$

where the $\boldsymbol{\Theta}_j$ are $K \times K$ matrices, and $\boldsymbol{\theta}_0$ is $K \times 1$. Then

- $\text{Cov}[\mathbf{X}_t, \mathbf{Z}_t] = \Sigma$
- $\boldsymbol{\Gamma}_0 = \Sigma + \boldsymbol{\Theta}_1 \Sigma \boldsymbol{\Theta}_1^\top + \cdots + \boldsymbol{\Theta}_q \Sigma \boldsymbol{\Theta}_q^\top$
-

$$\boldsymbol{\Gamma}_h = \begin{cases} \sum_{j=l}^q \boldsymbol{\Theta}_j \Sigma \boldsymbol{\Theta}_{j-h}^\top & 1 \leq h \leq q \\ 0 & h > q \end{cases}$$

where $\boldsymbol{\Theta}_0 = \mathbf{I}_K$.

Vector ARMA Models

Let

$$\Phi(B)\mathbf{X}_t = \phi_0 + \Theta(B)\mathbf{Z}_t$$

where

$$\Phi(B) = \mathbf{I}_K - \Phi_1 B - \dots - \Phi_p B^p$$

$$\Theta(B) = \mathbf{I}_K + \Theta_1 B + \dots + \Theta_q B^q$$

This describes a vector ARMA(p, q) (VARMA(p, q)) model.

The marginal processes arising from a K -dimensional VARMA(p, q) model are ARMA($Kp, (K-1)p + q$).

5.2 Cointegration

Consider the VARMA(1, 1) model

$$\Phi(B)X_t = \Theta(B)Z_t$$

with

$$\Phi(B) = \Phi_1(B) = \begin{bmatrix} 1 - 0.5B & B \\ 0.25B & 1 - 0.5B \end{bmatrix}$$

$$\Theta(B) = \Theta_1(B) = \begin{bmatrix} 1 - 0.2B & 0.4B \\ 0.1B & 1 - 0.5B \end{bmatrix}$$

with

$$\text{Var}[Z_t] = \Sigma$$

5.2 Cointegration

Then

$$\begin{bmatrix} 1 - 0.5B & B \\ 0.25B & 1 - 0.5B \end{bmatrix} \mathbf{X}_t = \begin{bmatrix} 1 - 0.2B & 0.4B \\ 0.1B & 1 - 0.5B \end{bmatrix} \mathbf{Z}_t$$

Premultiplying by

$$\begin{bmatrix} 1 - 0.5B & -B \\ -0.25B & 1 - 0.5B \end{bmatrix}$$

we have

$$\begin{bmatrix} 1 - B & 0 \\ 0 & 1 - B \end{bmatrix} \mathbf{X}_t = \begin{bmatrix} 1 - 0.7B & -0.6B \\ -0.15B & 1 - 0.7B \end{bmatrix} \mathbf{Z}_t$$

implying that X_{1t} and X_{2t} have unit-roots, specifically $X_{1t}, X_{2t} \sim \text{ARIMA}(0, 1, 1)$.

5.2 Cointegration

However, let

$$L = \begin{bmatrix} 1 & -2 \\ 0.5 & 1 \end{bmatrix}$$

and let

$$Y_t = LX_t \qquad U_t = LZ_t$$

Then, using Φ and Θ to denote the 2×2 AR and MA matrices,

$$\begin{aligned} Y_t = LX_t &= L\Phi X_{t-1} + LZ_t + L\Theta Z_{t-1} \\ &= L\Phi L^{-1}(LX_{t-1}) + LZ_t + L\Theta L^{-1}(LZ_{t-1}) \\ &= L\Phi L^{-1}Y_{t-1} + U_t + L\Theta L^{-1}U_{t-1} \end{aligned}$$

or equivalently

$$Y_t - \Phi^* Y_{t-1} = U_t + \Theta^* U_{t-1}$$

5.2 Cointegration

In this equation

$$\Phi^{\star} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \Theta^{\star} = \begin{bmatrix} -0.4 & 0 \\ 0 & 0 \end{bmatrix}$$

so that

$$Y_{1t} - Y_{1,t-1} = U_{1t} - 0.4U_{1,t-1}$$

$$Y_{2t} = U_{2t}$$

Hence $\{Y_{1t}\} \sim \text{ARIMA}(0, 1, 1)$ and $\{Y_{2t}\}$ is a white noise process, and there is no dependence between the two series.

Hence, the unit roots in $\{X_{1t}\}$ and $\{X_{2t}\}$ result from the unit root in $\{Y_{1t}\}$ and a subsequent linear transformation.

This phenomenon is known as *cointegration*.

Cointegration

Cointegration can be used to analyze co-movements in raw series (asset prices, exchange rates or yields).

The modelling strategy allows the detection of stable or stationary long-run relationships between non-stationary variables.

This allows the underlying relationships between the series to be discovered, and perhaps may allow forecasting.

5.2 Cointegration

The components of k -vector $\mathbf{X}_t$ are said to be *cointegrated* of order (d, b) , denoted $\mathbf{X}_t \sim CI(d, b)$ if

- (a) the components of $\mathbf{X}_t$ are $I(d)$ (stationary after d -times differencing),
- (b) there exists a linear combination of $\mathbf{X}_t$, Z_t^* say, where

$$Z_t^* = \alpha^\top \mathbf{X}_t \quad \alpha \neq \mathbf{0}$$

such that $Z_t^* \sim I(d - b)$ for some $b > 0$.

The vector α is termed the *cointegrating vector*.

Time series that cointegrate indicate that stable long-run relationships (equilibria) between the series exist irrespective of the behaviour of the individual series.

Estimation of Cointegrating Series

Consider, for illustration, $Y_t, \mathbf{X}_t = (X_{t1}, \dots, X_{tK})^\top \sim I(1)$ that are cointegrated. Suppose that

$$Y_t - (\alpha_1 X_{t1} + \dots + \alpha_K X_{tK}) \sim I(0)$$

Step 1. Form the regression model

$$Y_t = \alpha_1 X_{t1} + \dots + \alpha_K X_{tK} + Z_t^\star$$

with (cointegrating) vector

$$\boldsymbol{\alpha} = (1, -\alpha_1, \dots, -\alpha_K)^\top$$

and estimate the parameters in the model (using OLS or maximum likelihood).

Engle and Granger demonstrate that the resulting estimators are consistent with convergence rate $O_p(n^{-1})$ (that is, faster than the usual rate $O_p(n^{-1/2})$).

Estimation of Cointegrating Series

Step 2. Under cointegration, residuals $\hat{z}_t^*$ should be $I(0)$ (stationary); this can be tested using Dickey-Fuller procedures.

If $\hat{z}_t^* \sim I(0)$ is acceptable (the unit root hypothesis in the Dickey-Fuller test is rejected) then an *Error Correction Model* (ECM) can be specified.

Estimation of Cointegrating Series

For $K = 1$: Suppose $X_t, Y_t \sim I(1)$, and we conclude from the previous analysis that $\hat{z}_t^* \sim I(0)$. We have

$$\begin{aligned}(1 - B)Y_t &= \psi_0 + \gamma_1 \hat{z}_{t-1}^* + \sum_{j=1}^p \psi_{1j}(1 - B)X_{t-j} \\ &\quad + \sum_{l=1}^m \psi_{2l}(1 - B)Y_{t-l} + \epsilon_{1t} \\ (1 - B)X_t &= \xi_0 + \gamma_2 \hat{z}_{t-1}^* + \sum_{j=1}^p \xi_{1j}(1 - B)Y_{t-j} \\ &\quad + \sum_{l=1}^m \xi_{2l}(1 - B)X_{t-l} + \epsilon_{2t}\end{aligned}$$

Again, estimate using standard OLS/ML techniques.

Estimation of Cointegrating Series

The ECM in the first equation states that changes in the series Y_t are explained by

- ▶ the error in the long-run equilibrium from previous time point (coefficient γ_1),
- ▶ lagged changes in the X_t series (coefficients ψ_1),
- ▶ their own history (coefficients ψ_2).

Furthermore

- ▶ γ_1 determines the rate of re-adjustment; should have

$$\hat{\gamma}_1 < 0.$$

- ▶ Model orders p and q need to be selected; typically start with p, q large, and drop variables according to t -statistics, or using model selection criteria
- ▶ Can extend to systems of cointegrated variables

Two cointegrating series

Example (Cointegrating series)

Suppose that

$$X_t \sim I(1)$$

$$Y_t = 0.6X_t + Z_t$$

where $Z_t \sim N(0, 1)$ is an IID series. This implies that

$$Y_t - 0.6X_t = Z_t$$

is a cointegrating system with cointegrating vector $\alpha = (1, -0.6)^\top$.

To estimate the system parameters, we regress Y_t on X_t , and then form the error correction models.

Two cointegrating series

Example (Cointegrating series)

Regressions are

$$(1 - B)Y_t = \psi_0 + \gamma_1 \hat{z}_{t-1}^* + \psi_{11}(1 - B)X_{t-1} + \psi_{21}(1 - B)Y_{t-1} + \epsilon_{1t}$$

$$(1 - B)X_t = \xi_0 + \gamma_2 \hat{z}_{t-1}^* + \xi_{11}(1 - B)Y_{t-1} + \xi_{21}(1 - B)X_{t-1} + \epsilon_{2t}$$

where $\{\hat{z}_t^*\}$ are the residuals from the fit of the regression of Y_t on X_t .

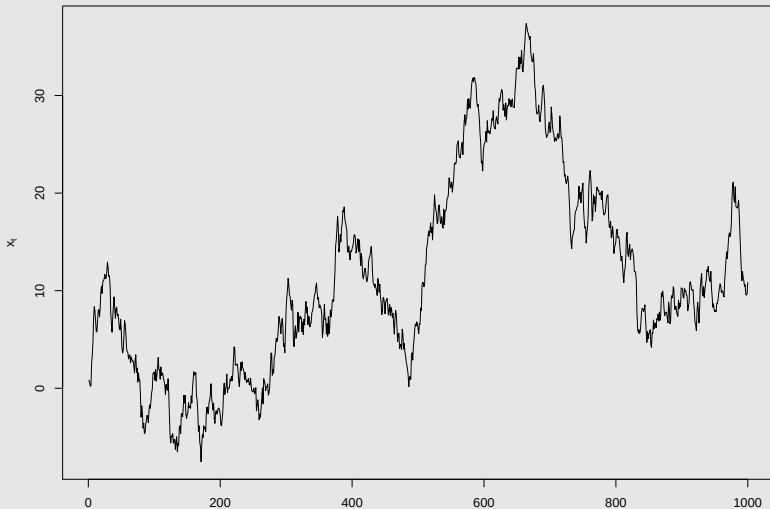
Simulated Example: R code

R code:

```
1 set.seed(123456)
2 n<-1000
3 e1<-rnorm(n);e2<-rnorm(n)
4 x<-cumsum(e1);y<-0.6*x+e2
5 lr.reg<-lm(y$\sim x)
6 error<-residuals(lr.reg)
7 error.lagged<-error[-c(n-1,n)]
8 dx<-diff(x);dy<-diff(y)
9 diff.dat<-data.frame(embed(cbind(dx,dy),2))
10 colnames(diff.dat)<-c("dx","dy","dx.1","dy.1")
11
12 ecm.reg<-lm(dy~error.lagged+dx.1+dy.1,data=diff.dat)
13 summary(ecm.reg)
14 plot(x,type="l");plot(y,type="l");plot(x,y)
```

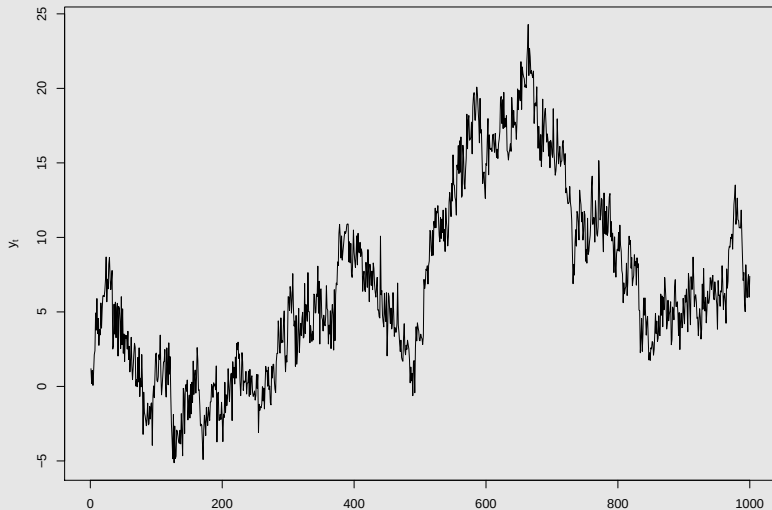
$$X_t, t = 1, \dots, 1000$$

Example (Cointegrating series)

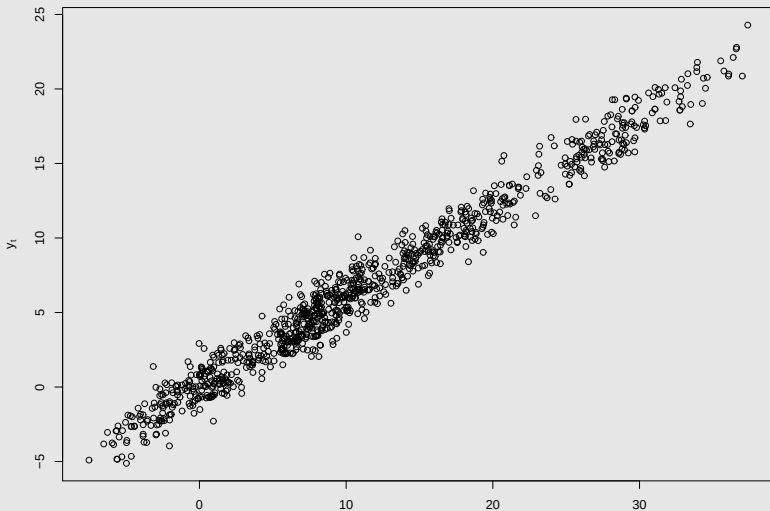


$$Y_t, t = 1, \dots, 1000$$

Example (Cointegrating series)



Example (Cointegrating series)



Results

Example (Cointegrating series)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.005528	0.037437	0.148	0.883
error.lagged	-0.936862	0.051442	-18.212	0.000
dx.1	0.602592	0.043770	13.767	0.000
dy.1	-0.972102	0.036773	-26.436	0.000

That is

$$\hat{\psi}_0 = 0.0055 \quad \hat{\gamma}_1 = -0.9369 \quad \hat{\psi}_{11} = 0.6026 \quad \hat{\psi}_{21} = -0.9721$$

Example: Consumption in the UK 1966-1991

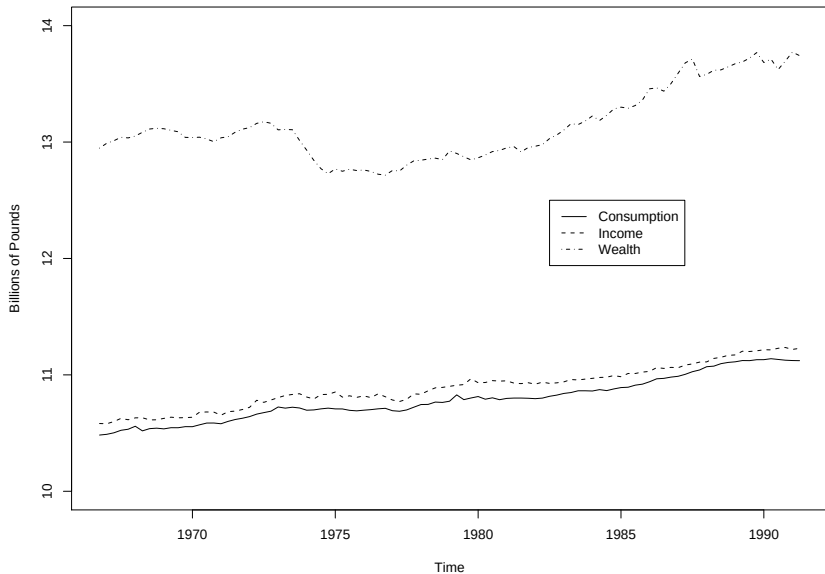
Suppose

- $\{X_{1t}\}$ is the *consumption* series (quarterly consumer expenditure seasonally adjusted)
- $\{X_{2t}\}$ is the *income* series (personal disposable income seasonally adjusted)
- $\{X_{3t}\}$ is the *wealth* series (personal financial wealth seasonally adjusted).

from 4th quarter 1966 to 2nd quarter 1991.

We wish to assess whether these three series cointegrate; this would indicate some form of long-run equilibrium exists.

UK Consumption 1966-1991



Using the Augmented Dickey-Fuller test

For some $p \geq 1$, consider the three usual models

1. No intercept, no trend

$$(1 - B)X_t = \eta X_{t-1} + \sum_{j=1}^p \gamma_j (1 - B)X_{t-j} + Z_t$$

2. Intercept, no trend

$$(1 - B)X_t = \psi_0 + \eta X_{t-1} + \sum_{j=1}^p \gamma_j (1 - B)X_{t-j} + Z_t$$

3. Intercept plus trend

$$(1 - B)X_t = \psi_0 + \psi_1 t + \eta X_{t-1} + \sum_{j=1}^p \gamma_j (1 - B)X_{t-j} + Z_t$$

Example: $p = 3$

In each case, we wish to test

$$H_0 : \eta = 0$$

$$H_1 : \eta < 0$$

R code:

```
1 library(urca)
2 data(Raotbl3)
3 attach(Raotbl3)
4 lc.mod1 <- ur.df(lc, lags=3, type="none")
5 lc.mod2 <- ur.df(lc, lags=3, type="drift")
6 lc.mod3 <- ur.df(lc, lags=3, type="trend")
7 detach(Raotbl3)
```

Cointegration

We can look for cointegration between

- ▶ $\{X_{1t}\}$ and $\{X_{2t}\}$
- ▶ $\{X_{2t}\}$ and $\{X_{1t}\}$
- ▶ $\{X_{1t}\}$ and $\{X_{3t}\}$
- ▶ $\{X_{2t}\}$ and $\{X_{3t}\}$
- ▶ ...

using the two-step regression methods described above to assess the relationships between consumption, income and wealth.

Vector Cointegration and ECM

The ECM model of the previous sections can be extended to cover the case of vectors of cointegrated variables.

Recall that $K \times 1$ vector process X_t is said to be cointegrated if at least one non-zero $K \times 1$ vector α_i exists such that

$$\alpha_i^\top X_t$$

is stationary. If r such linearly independent vectors $\alpha_1, \dots, \alpha_r$ exist, then X_t is cointegrated with rank r , with

$$A = [\alpha_1 \dots \alpha_r]$$

the $K \times r$ *cointegrating matrix*.

Vector Cointegration and ECM

The VAR(p) model

$$\mathbf{X}_t = \mu + \Lambda D_t + \sum_{j=1}^p \Phi_j \mathbf{X}_{t-j} + \epsilon_t$$

can be used as the basis of the estimation, where D_t is a vector of exogenous (external) variables. Let

$$(\mathbf{1} - \mathbf{B}) = (\mathbf{1} - \mathbf{B})\mathbf{I}_K$$

$$\mathbf{\Gamma}_i = -(\mathbf{I}_K - \mathbf{\Phi}_1 - \dots - \mathbf{\Phi}_i) \quad i = 1, \dots, p-1$$

$$\mathbf{\Gamma}_i^* = -(\mathbf{\Phi}_{i+1} + \dots + \mathbf{\Phi}_p) \quad i = 1, \dots, p-1$$

$$\mathbf{\Phi} = -(\mathbf{I}_K - \mathbf{\Phi}_1 - \dots - \mathbf{\Phi}_p)$$

with all matrices $K \times K$.

Vector Cointegration and ECM

Two models can be used to assess cointegration by inspecting estimates of Φ

- ▶ VECM I:

$$(\mathbf{1} - \mathbf{B})\mathbf{X}_t = \boldsymbol{\mu} + \Lambda D_t + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i (\mathbf{1} - \mathbf{B})\mathbf{X}_{t-i} + \boldsymbol{\Phi}\mathbf{X}_{t-p} + \boldsymbol{\epsilon}_t$$

- ▶ VECM II:

$$(\mathbf{1} - \mathbf{B})\mathbf{X}_t = \boldsymbol{\mu} + \Lambda D_t + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i^* (\mathbf{1} - \mathbf{B})\mathbf{X}_{t-i} + \boldsymbol{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t$$

5.3 State-Space Models

The *state-space* form of time series models has received considerable attention, as they can better represent the actual dynamics of a data generation process.

The principle is that an observed series is in fact generated as the **marginal** process derived from a multivariate time series model, structured so that the model is specified via two equations

- the *state* equation
- the *observation* equation

5.3 State-Space Models

The general (first-order Markov) *state equation* takes the form

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}; \boldsymbol{\theta}_{t-1}) + \boldsymbol{\eta}_t$$

and the general *observation equation* takes the form

$$\mathbf{y}_t = \mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}_t) + \boldsymbol{\epsilon}_t.$$

The process $\{\mathbf{y}_t\}$ is observed, the process $\{\mathbf{x}_t\}$ is *latent* or unobserved, and one focus of inferential interest.

Inference in General State-Space Models

Let $y_{1:t} = \{y_1, \dots, y_t\}$. Inference goals for state-space models are

- ▶ *filtering/tracking*: estimate state at time t from all observations up to time t

$$p(x_t | y_{1:t})$$

- ▶ *smoothing*: estimate state at time t from all past and possibly some future observations

$$p(x_t | y_{1:(t+s)})$$

- ▶ *prediction/forecasting*: estimate state at time t from observations up to a previous time point $t - s$

$$p(x_t | y_{1:(t-s)})$$

Inference in General State-Space Models

The fundamental inference mechanism is Bayesian; we compute the posterior quantities of interest sequentially in the following recursive calculation:

- Prediction

$$p(\mathbf{x}_t | y_{1:(t-1)}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | y_{1:(t-1)}) d\mathbf{x}_{t-1}$$

- Updating

$$p(\mathbf{x}_t | y_{1:t}) = \frac{p(y_t | \mathbf{x}_t) p(\mathbf{x}_t | y_{1:(t-1)})}{p(y_t | y_{1:(t-1)})}$$

Inference in General State-Space Models

The above equations only provide an analytical solution if all densities in the state and observation equation are Gaussian, and both the state and the observation equation are linear.

- If these conditions are met, the *Kalman filter* provides the optimal (Bayesian) solution to the tracking problem.
- If these conditions are not met, we require approximations:
 - *Extended Kalman filter* (EKF): requires Gaussian densities, approximates non-linearities (using a first-order Taylor expansion)
 - *Particle filter* (PF): approximates non-Gaussian densities and non-linear equations

Linear Gaussian State-Space Models

The general form of a linear state-space model, is given by the following two equations

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{x}_t + \mathbf{d}_t + \mathbf{G}_t \boldsymbol{\epsilon}_t, \quad t = 1, \dots, n \quad (1a)$$

$$\mathbf{x}_{t+1} = \mathbf{T}_t \mathbf{x}_t + \mathbf{c}_t + \mathbf{H}_t \boldsymbol{\epsilon}_t. \quad (1b)$$

where

- $\mathbf{y}_t$ is $(K \times 1)$
- $\mathbf{x}_t$ is $(L \times 1)$.

Equation (1a) is the *observation* or *measurement* equation, while (1b) is *transition* equation.

Linear Gaussian State-Space Models

In (1)

- $\mathbf{Z}_t$ is $(K \times L)$,
- $\mathbf{d}_t$ is $(K \times 1)$,
- $\mathbf{G}_t$ is $(n \times (K + L))$,
- $\mathbf{T}_t$ is $(L \times L)$
- $\mathbf{c}_t$ is $(L \times 1)$
- $\mathbf{H}_t$ is $(L \times (K + L))$.

All of these matrices are referred to as the *system matrices* and are assumed to be non-stochastic.

Linear Gaussian State-Space Models

Process ϵ_t is an $((K + L) \times 1)$ vector of serially independent, identically distributed disturbances with $\mathbb{E}[\epsilon_t] = \mathbf{0}$ and $\mathbb{V}ar[\epsilon_t] = \mathbf{I}_{K+L}$, the identity matrix.

The formulation is completed by the assumption that the initial state vector $\mathbf{x}_1$ is independent of ϵ_t at all times t and has unconditional mean and variance $\mathbf{a}_{1|0}$ and $\mathbf{P}_{1|0}$ respectively.

Linear Gaussian State-Space Models

The state-space model is characterized by the properties that the system matrices, the disturbance terms ϵ_t and initial state vector x_1 possess. If the system matrices do not evolve with time, the state-space model is called *time-invariant* or *time-homogeneous*.

If the disturbances, ϵ_t , and initial state vector, x_1 , are assumed to have a normal distribution, then the model is termed Gaussian.

If $G_t H_t^T = \mathbf{0}$ for all t , then the measurement and transition equations are uncorrelated.

Linear Gaussian State-Space Models

Example (AR(2) Model)

Let $\{X_t\}$ be a zero-mean, Gaussian AR(2) process given by

$$X_{t+1} = \phi_1 X_t + \phi_2 X_{t-1} + u_{2t}$$

where $u_{2t} \sim \mathcal{N}(0, \sigma_2^2)$. Suppose also that we observe the process $\{Y_t\}$ with noise, so that

$$Y_t = X_t + u_{1t}$$

with $u_{1t} \sim \mathcal{N}(0, \sigma_1^2)$ and independent of u_{2t} .

Linear Gaussian State-Space Models

Example (AR(2) Model)

A state-space representation of this model can be formed as follows

$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

$$\begin{bmatrix} X_{t+1} \\ X_t \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & \sigma_2^2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

where $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})^\top$ is $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$.

Clearly the latter state-space model is time-homogeneous and Gaussian, and the observation and state equations are uncorrelated.

5.4 Kalman Filter

The Kalman filter is used for prediction, filtering and smoothing.

- ▶ *Filtering* is concerned with computing $\mathbb{E}[\mathbf{x}_t | \mathbf{y}_{1:t}]$,
- ▶ *Smoothing* is concerned with estimating $\mathbb{E}[\mathbf{x}_t | \mathbf{y}_{1:n}]$, for all $t < n$.
- ▶ Prediction is concerned with computing $\mathbb{E}[\mathbf{x}_{t+1} | \mathbf{y}_{1:t}]$.

The Kalman filter consists of a set of simple recursions that compute the latter quantities.

5.4 Kalman Filter

For simplicity, assume that

$$\mathbf{G}_t \mathbf{H}_t^\top = \mathbf{0} \qquad \mathbf{d}_t = \mathbf{c}_t = \mathbf{0}$$

and

$$\mathbf{G}_t \mathbf{G}_t^\top = \mathbf{\Sigma}_t \qquad \mathbf{H}_t \mathbf{H}_t^\top = \mathbf{\Omega}_t.$$

The Kalman filter recursively computes the quantities

- $\mathbf{a}_{t|t} = \mathbb{E}[\mathbf{X}_t | \mathbf{y}_{1:t}]$,
- $\mathbf{a}_{t+1|t} = \mathbb{E}[\mathbf{X}_{t+1} | \mathbf{y}_{1:t}]$,
- $\mathbf{P}_{t|t} = \text{MSE}[\mathbf{X}_t | \mathbf{y}_{1:(t-1)}]$,
- $\mathbf{P}_{t+1|t} = \text{MSE}[\mathbf{X}_{t+1} | \mathbf{y}_{1:t}]$.

where MSE is the *mean-square error* or *one-step ahead prediction variance*.

5.4 Kalman Filter

Then, starting with $\mathbf{a}_{1|0}$ and $\mathbf{P}_{1|0}$, $\mathbf{a}_{t|t}$, $\mathbf{a}_{t+1|t}$ and their MSEs are obtained by running for $t = 1, \dots, n$, the recursions

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_{t|t-1} \quad , \quad \mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t^\top + \Sigma_t \quad (2a)$$

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{Z}_t^\top \mathbf{F}_t^{-1} \mathbf{v}_t, \quad (2b)$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{Z}_t^\top \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_{t|t-1}, \quad (2c)$$

$$\mathbf{a}_{t+1|t} = \mathbf{T}_t \mathbf{a}_{t|t}, \quad (2d)$$

$$\mathbf{P}_{t+1|t} = \mathbf{T}_t \mathbf{P}_{t|t} \mathbf{T}_t^\top + \Omega_t. \quad (2e)$$

5.4 Kalman Filter

Notes:

- quantities v_t and F_t in (2a) respectively denote the one-step-ahead error in forecasting y_t conditional on the information set at time $t - 1$ and its MSE. This facilitates computation of the likelihood function.
- the quantities $a_{t|t}$ and $a_{t|t-1}$ are optimal estimators of x_t conditional on the available information, in terms of minimum mean square.
- the latter property holds under Gaussianity; if this is not assumed, $a_{t|t}$ and $a_{t|t-1}$ are optimal only within the class of **linear** estimators, that is, they are the *minimum mean square linear estimators* of x_t conditional on $y_{1:t}$ and $y_{1:(t-1)}$.

5.4 Kalman Filter

- ▶ Finally, equations (2b) and (2d) can be combined together to yield recursions that only compute the one-step-ahead prediction estimates of $\mathbf{x}_{t+1}$ given $y_{1:t}$.

Similarly, by taking together (2c) and (2e), a single set of recursions for the MSE is obtained, which goes directly from $\mathbf{P}_{t|t-1}$ to $\mathbf{P}_{t+1|t}$.

- ▶ The resulting recursions, along with (2a), are

$$\mathbf{a}_{t+1|t} = \mathbf{T}_t \mathbf{a}_{t|t-1} + \mathbf{K}_t \mathbf{v}_t \quad (3a)$$

$$\mathbf{K}_t = \mathbf{T}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t^\top \mathbf{F}_t^{-1} \quad (3b)$$

$$\mathbf{P}_{t+1|t} = \mathbf{T}_t \mathbf{P}_{t+1|t} \mathbf{L}_t^\top + \mathbf{\Omega}_t \quad (3c)$$

$$\mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t. \quad (3d)$$

Parameter Estimation

Another application of the Kalman filter is the estimation of any unknown parameters θ that appear in the system matrices.

The likelihood for data $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ can be constructed as

$$p(\mathbf{y}_1, \dots, \mathbf{y}_T) = p(\mathbf{y}_n | \mathbf{y}_{1:(n-1)}) \cdots p(\mathbf{y}_2 | \mathbf{y}_1) p(\mathbf{y}_1) = \prod_{t=1}^n p(\mathbf{y}_t | \mathbf{y}_{1:(t-1)}).$$

For a Gaussian model, by taking conditional expectations on both sides of the observation equation, with $\mathbf{d}_t \equiv 0$, then

$$\mathbb{E}[\mathbf{Y}_t | \mathbf{y}_{1:(t-1)}] = \mathbf{Z}_t \mathbf{a}_{t|t-1} \qquad \mathbb{V}\text{ar}[\mathbf{Y}_t | \mathbf{y}_{1:(t-1)}] = \mathbf{F}_t.$$

Crucially, the one-step-ahead prediction density $p(\mathbf{y}_t | \mathbf{y}_{1:(t-1)})$ is the density of a multivariate normal random variable with mean $\mathbf{Z}_t \mathbf{a}_{t|t-1}$ and covariance matrix $\mathbf{F}_t$.

Thus, the log-likelihood function is given by

$$\log \ell = -\frac{n}{2} \log (2\pi) - \frac{1}{2} \sum_{t=1}^n \log (\det \mathbf{F}_t) - \frac{1}{2} \sum_{t=1}^n \mathbf{v}_t^\top \mathbf{F}_t^{-1} \mathbf{v}_t.$$

where

$$\mathbf{v}_t = (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_{t|t-1})$$

Typically, numerical procedures are used in order to maximize the log-likelihood to obtain the ML estimates of the parameters θ which are consistent and asymptotically normal.

Non-Linear Non-Gaussian State-Space Models

If the state-space model is not Gaussian, the likelihood can still be constructed in the same way using the minimum mean square linear estimators of the state vector.

However, the estimators $\hat{\theta}$ that maximize the likelihood are the *quasi-maximum likelihood* (QML) estimators of the parameters. It can be shown that the QML estimators are also consistent and asymptotically normal.

Another algorithm can be applied to a state-space model; given a fixed set of data, estimates of the state vector are computed at each time t in the sample period taking into account the full information set available.

This algorithm computes $\mathbf{a}_{t|n} = \mathbb{E}[\mathbf{X}_t | y_{1:n}]$ along with its MSE, $\mathbf{P}_{t|n}$, for all $t = 1, \dots, n - 1$.

Non-Linear Non-Gaussian State-Space Models

The latter quantities are computed via a set of backward recursions and use the quantities $\mathbf{a}_{t+1|t}$, $\mathbf{a}_{t|t}$ and the MSE matrices $\mathbf{P}_{t|t}$, $\mathbf{P}_{t+1|t}$, which are obtained from (2).

In particular, to obtain $\mathbf{a}_{t|T}$ and $\mathbf{P}_{t|T}$, we start with $\mathbf{a}_{n|n}$ and $\mathbf{P}_{n|n}$ and we run backwards for $t = n - 1, \dots, 0$

$$\mathbf{a}_{t|T} = \mathbf{a}_{t|t} + \mathbf{P}_t^* (\mathbf{a}_{t+1|T} - \mathbf{a}_{t+1|t})$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{P}_t^* (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{P}_t^* \quad , \quad \mathbf{P}_t^* = \mathbf{P}_t \mathbf{T}_t' \mathbf{P}_{t+1|t}.$$

5.5 Simulation-based approaches

During the two decades, the extensive use of the *Markov chain Monte Carlo* (MCMC), in particular the Gibbs sampler, has given rise to another smoothing algorithm called the *simulation smoother* and is also closely related to the Kalman filter.

A simulation smoother is used for drawing samples from the density $p(\mathbf{x}_0, \dots, \mathbf{x}_n | y_{1:n})$. The first simulation smoother is based on the identity

$$p(\mathbf{x}_0, \dots, \mathbf{x}_n | y_{1:n}) = p(\mathbf{x}_n | y_{1:n}) \prod_{t=0}^{n-1} p(\mathbf{x}_t | y_{1:t}, \mathbf{x}_{t+1})$$

and a draw from $p(\mathbf{x}_0, \dots, \mathbf{x}_n | y_{1:n})$ is recursively constructed in terms of $\mathbf{x}_t$.

5.5 Simulation-based approaches

Given $\hat{\mathbf{x}}_n \sim \mathcal{N}(\mathbf{x}_{n|n}, \mathbf{P}_{n|n})$, for a Gaussian state space model, $p(\mathbf{x}_t | y_{1:t}, \mathbf{x}_{t+1})$ is completely characterized by its first and second moments.

In order to compute these moments:

- ▶ the usual Kalman filter recursions (2) are run, so that $\mathbf{x}_{t|t}$ is initially obtained, then
- ▶ the draw $\hat{\mathbf{x}}_{t+1} \sim p(\mathbf{x}_{t+1} | y_{1:t}, \mathbf{x}_{t+2})$ is treated as m additional observations and a second set of m Kalman filter recursions is run for each element of the state vector $\hat{\mathbf{x}}_{t+1}$.

5.5 Simulation-based approaches

The latter procedure involves the inversion of system matrices, which are not necessarily non-singular; hence another *simulation smoother* algorithm has been devised which is computationally more efficient.

The main idea of second algorithm is to obtain a sample from the joint density of the transition equation disturbances, i.e. $p(H_0\epsilon_0, \dots, H_n\epsilon_n|y_{1:n})$, by sampling

$$\hat{\eta}_t \sim p(\mathbf{H}_t\epsilon_t|y_{1:n}, H_{t+1}\epsilon_{t+1}, \dots, H_n\epsilon_n).$$

5.5 Simulation-based approaches

This allows the state vector to be reconstructed, using the transition equation (1b) and substituting the simulated disturbances.

In particular, the new simulation smoother requires only the prediction equations (3) to be initially run and the quantities $\mathbf{v}_t$, $\mathbf{F}_t$ and $\mathbf{L}_t$ to be stored. Then, starting with $\mathbf{r}_n = \mathbf{0}$ and $\mathbf{U}_n = \mathbf{0}$, the following backward recursions are run for $t = n, n - 1, \dots, 1$.

$$\begin{aligned}\mathbf{C}_t &= \mathbf{\Omega}_t - \mathbf{\Omega}_t \mathbf{U}_t \mathbf{\Omega}_t \\ \boldsymbol{\kappa}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{C}_t), \\ \mathbf{r}_{t-1} &= \mathbf{Z}_t^\top \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}_t^\top \left(\mathbf{r}_t - \mathbf{U}_t \mathbf{\Omega}_t \mathbf{C}_t^{-1} \boldsymbol{\kappa}_t \right), \\ \mathbf{U}_{t-1} &= \mathbf{Z}_t^\top \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}_t^\top \left(\mathbf{U}_t + \mathbf{U}_t \mathbf{\Omega}_t \mathbf{C}_t^{-1} \mathbf{\Omega}_t \mathbf{U}_t \right) \mathbf{L}_t.\end{aligned}$$

5.5 Simulation-based approaches

Then, for $t = 1, \dots, n$, $\eta_t = \mathbf{\Omega}_t r_t + \kappa_t$ is a draw from

$$p(\mathbf{H}_t \epsilon_t | y_{1:n}, \mathbf{H}_{t+1} \epsilon_{t+1}, \dots, \mathbf{H}_n \epsilon_n).$$

For the initial disturbance term,

$$\eta_0 = \mathbf{P}_{1|0} r_0 + \kappa_0,$$

where $\kappa_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{1|0} - \mathbf{P}_{1|0} \mathbf{U}_0 \mathbf{P}_{1|0})$, is a draw from

$$p(H_0 \epsilon_0 | y_{1:n}, \mathbf{H}_1 \epsilon_1, \dots, \mathbf{H}_n \epsilon_n)$$

The state vector is simulated by starting with $\mathbf{x}_0 = \mathbf{0}$ and running for $t = 0, \dots, n$ the transition equation recursion with $\mathbf{H}_t \epsilon_t$ substituted by η_t

$$\mathbf{x}_{t+1} = \mathbf{T}_t \mathbf{x}_t + \eta_t.$$

Particle Filtering

The particle filter uses *Monte Carlo methods*, in particular *Importance sampling*, to construct the approximations. If

$$\mathbb{E}_f[g(X)] = \int g(x)f(x) \, dx$$

then

$$\mathbb{E}_f[g(X)] \equiv \mathbb{E}_{f_0} \left[\frac{f(X)g(X)}{f_0(X)} \right] = \int \frac{g(x)f(x)}{f_0(x)} f_0(x) \, dx$$

for some *importance sampling* density f_0 .

Particle Filtering

We then approximate the integral by sampling $x_1, \dots, x_N$ from f_0 for large N , and constructing the approximation

$$\hat{\mathbb{E}}_{f_N} = \frac{1}{N} \sum_{i=1}^N \frac{g(x_i)f(x_i)}{f_0(x_i)} = \frac{1}{N} \sum_{i=1}^N w_i g(x_i),$$

say, where, if f_0 is chosen carefully

$$\hat{\mathbb{E}}_{f_N} \longrightarrow \mathbb{E}_f[g(X)]$$

as $N \longrightarrow \infty$

Particle Filtering

If the function f is known only up to proportionality, the same procedure can work if the weights w_i are suitably normalized to sum to one, that is, if above

$$w_i = \frac{f(x_i)}{f_0(x_i)}$$

is replaced by

$$w_i^* = \frac{w_i}{W} \quad \text{with} \quad W = \sum_{i=1}^n w_i.$$

as W approximates the integral

$$\int f(x) \, dx = \int \frac{f(x)}{f_0(x)} f_0(x) \, dx$$

Particle Filtering

Given a non-linear and non-Gaussian state space model, the goal is to compute the posterior density $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ at each time t exactly.

At each time point t , perform the following *Sequential Importance Sampling* (SIS) procedure:

- ▶ Draw random samples (“particles”) from a chosen importance density that can be sampled directly.
- ▶ Compute a normalized weight for each particle
- ▶ Approximate the true posterior density at time t by the weighted sum of the particles.

Given sufficiently large numbers of particles, this IS characterization converges to the true posterior density.

Particle Filtering

In the filtering problem, at time t , we approximate $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ by

- particles $\mathbf{x}_t = (x_{t1}, \dots, x_{tN})$, sampled from density f_0 .
- normalized weights $\mathbf{w}_t^* = (w_{t1}^*, \dots, w_{tN}^*)$

via the discrete distribution $\{(w_{ti}^*, x_{ti}), i = 1 \dots, N\}$. Two major questions are

- What importance density f_0 can or should be chosen?
- How do we compute the weights efficiently (in a recursive fashion) such that for a given weight w_{ti}^* , we just need a new observation $\mathbf{y}_t$ to get an approximation to the posterior $p(\mathbf{x}_t | \mathbf{y}_{1:t})$.

A simple but effective choice for f_0 sampling density at time t is the *transition prior*: $f_0(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}_{1:t}) \equiv p(\mathbf{x}_t | \mathbf{x}_{t-1})$ as then

$$w_{ti}^* \propto w_{t-1,i}^* p(\mathbf{y}_t | \mathbf{x}_{ti}^*).$$

Particle Degeneracy

A problem with sequential importance sampling is that typically, after a few iterations, all but one particle have negligible weights. This phenomenon is termed *degeneracy*.

The degree of degeneracy is represented by the *effective sample size*

$$N_t = \left(\sum_{i=1}^N (w_{ti}^*)^2 \right)^{-1}$$

Possible solutions to overcome degeneracy effects:

- brute force: very large N
- optimized choice of importance density
- resampling

Resampling

The goal of resampling is to eliminate particles with small weights and replace by new particles drawn from the vicinity of particles with larger weights

- concentrate on particles with large weights
- model important parts of the posterior more precisely

Various efficient algorithms exist for resampling which operate in $O(N)$ time (e.g. *Sampling Importance Resampling (SIR)*)