1 Motivation

- To study a natural framework to study the dynamic structure of a collection of random variables over time \( \{ r_t \} \)

2 Linear Time Series Analysis

- Important topics:
  - Stationarity
  - Dynamic Dependence
  - Autocorrelation functions
  - Modeling
  - Forecasting

- The Models studied:
  - simple autoregressive (AR) models
  - simple moving-average (MA) models
  - mixed models (ARMA)
  - seasonal models
  - unit-root nonstationary
  - regression models with time series errors
2.1 Stationarity

- \( \{ r_t \} \) is strictly stationary if the joint distribution of \( (r_{t_1}, ..., r_{t_k}) \) is identical to that of \( (r_{t_1+\ell}, ..., r_{t_{k+\ell}}) \) for all \( \ell \).
  - joint distribution of \( (r_{t_1}, ..., r_{t_k}) \) is invariant to any shift in time

- \( \{ r_t \} \) is weakly stationary if the mean of \( r_t \) and the covariance between \( r_t \) and \( r_{t-\ell} \) are time-invariant (for all \( \ell \)).
  - \( E(r_t) = \mu, Cov(r_t, r_{t-\ell}) = \gamma_\ell \)
  - eyeballing: a time plot should show \( T \) values with constant volatility around some fixed level

2.2 Correlation and Autocorrelation Functions

- The correlation coefficient between 2 RVs \( X \) and \( Y \)
  \[
  \rho_{x,y} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E[(X - \mu_x)^2(Y - \mu_y)^2]}}
  \]

Given \( \{(x_t, y_t)\}_{t=1}^T \)
  \[
  \hat{\rho}_{x,y} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2 \sum_{t=1}^T (y_t - \bar{y})^2}}
  \]

Delivers the linear dependence between \( X \) and \( Y \)

Autocorrelation

- Usually the case where the linear dependence of \( r_t \) and \( r_{t-\ell} \) is of interest
  \[
  \rho_\ell = \frac{Cov(r_t, r_{t-\ell})}{\sqrt{Var(r_t)Var(r_{t-\ell})}} = \frac{Cov(r_t, r_{t-\ell})}{Var(r_t)} = \frac{\gamma_\ell}{\gamma_0}
  \]

Given \( \{r_t\}_{t=1}^T \)
  \[
  \hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad 0 \leq \ell < T - 1
  \]

- Is the lag-\( \ell \) sample autocorrelation of \( r_t \) significant?

- \( r_t \) is iid with \( E(r_t^2) < \infty \) \( \Rightarrow \) \( \hat{\rho}_\ell \sim N(0, 1/T) \)

- More general:
  \[
  r_t = \mu + \sum_{i=0}^q \psi_ia_{t-i} \Rightarrow \hat{\rho}_\ell \sim N\left[0, \left(1 + 2\sum_{i=1}^q \psi_i^2\right)/T\right] \quad \text{for } \ell > q
  \]

This result is used to construct a t-ratio and test the hypothesis that the
autocorrelation is equal to zero

\[ H_0 : \rho_\ell = 0 \]
\[ H_a : \rho_\ell \neq 0 \]

\[ t - ratio = \frac{\hat{\rho}_\ell}{\sqrt{1 + 2 \sum_{i=1}^{q} \rho_i^2 / T}} \]

• The t-ratio is asymptotically dist. as a standard normal (the Z statistics)

• For a given significance \( \alpha \)
  
  \(- |t - ratio| > Z_{\alpha/2} \rightarrow reject H_0\)

• We often will need to jointly test that several autocorrelations of \( r_t \) are zero

\[ H_0 : \rho_1 = \ldots = \rho_m = 0 \]
\[ H_a : \rho_i \neq 0, \text{ for some } i \in \{1, \ldots, m\} \]

\[ Q(m) = T (T + 2) \sum_{\ell=1}^{m} \frac{\hat{\rho}_\ell^2}{T - \ell} \]

• The (modified) Portmanteau statistic is asymptotically a chi-square rv with df = m.

  \(- Q(m) > \chi^2_\alpha \rightarrowreject H_0\)

  \(- In\ practice: choose m \approx \ln(T)\)

2.3 White Noise and Linear Time Series

• \( r_t \) is called a white noise series if \( \{r_t\} \) is a sequence of independent and identically distributed rvs with finite mean and variance.
  
  – all the ACFs are zero

• Significant (non-zero) ACFs of a series requires further analysis of its serial dependence.

Figure 1: Sample autocorrelation function of monthly value-weighted index of US markets from 01/26 to 12/97. Shaded region is the 95% confidence that the null hypothesis of \( \rho_\ell = 0 \) is accepted.
Linear Time Series

- A linear time series can be written as
  \[ r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \]
  - \( \mu \) is the mean of \( r_t \)
  - \( \psi_0 = 1 \)
  - \( \{a_t\} \) is a sequence of iid random variables with zero mean and a well-defined distribution (white noise)

- The lag-\( \ell \) autocovariance of \( r_t \) is
  \[ \gamma_\ell = Cov (r_t, r_{t-\ell}) = E \left[ \left( \sum_{i=0}^{\infty} \psi_i a_{t-i} \right) \left( \sum_{j=0}^{\infty} \psi_j a_{t-\ell-j} \right) \right] \]
  \[ = E \left( \sum_{i,j=0}^{\infty} \psi_i \psi_j a_{t-i} a_{t-\ell-j} \right) \]
  \[ = \sum_{j=0}^{\infty} \psi_j \psi_j E \left( a_{t-\ell-j}^2 \right) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_j \]

  The autocorrelations of \( r_t \) is given by
  \[ \rho_\ell = \frac{\gamma_\ell}{\gamma_0} = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+\ell}}{1 + \sum_{i=0}^{\infty} \psi_i^2}, \quad \ell \geq 0 \]

  Since \( \psi_i \to 0 \) as \( i \to \infty \), \( \rho_\ell \to 0 \) as \( \ell \to \infty \)

- \( a_t \) denotes new information at time \( t \) (an innovation or shock)

- The dynamic structure of \( r_t \) is governed by the coefficients \( \psi_i \) (called \( \psi \) weights)

- If \( r_t \) is weakly stationary
  \[ E (r_t) = \mu \]
  \[ Var (r_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2 \]
  - \( \{\psi_i^2\} \) must be a convergent series \( (\psi_i^2 \to 0 \text{ as } i \to \infty) \)

2.4 Simple Autoregressive Models

- \( r_t \) having a statistically significant lag-1 autocorrelation indicates that \( r_{t-1} \) may be useful in predicting \( r_t \)

- An AR(1) model is given by
  \[ r_t = \phi_0 + \phi_1 r_{t-1} + a_t \]
  - \( a_t \) is a white noise series with zero mean and variance equal to \( \sigma_a^2 \)

- This delivers
  \[ E (r_t | r_{t-1}) = \phi_0 + \phi_1 r_{t-1} \]
  \[ Var (r_t | r_{t-1}) = Var (a_t) = \sigma_a^2 \]
• AR(p) (More generally)
\[ r_t = \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_p r_{t-p} + a_t \]

\begin{align*}
\text{Properties of AR(1) models} \\
\text{• Assuming weak stationarity} \\
E(r_t) &= \mu \\
Var(r_t) &= \gamma_0 \\
Cov(r_t, r_{t-j}) &= \gamma_j
\end{align*}

we can derive
\begin{align*}
E(r_t) &= \phi_0 + \phi_1 E(r_{t-1}) \\
\mu &= \phi_0 + \phi_1 \mu \\
E(r_t) &= \mu = \frac{\phi_0}{1 - \phi_1}
\end{align*}

• Two implications
  - the mean exists iff \( \phi_1 \neq 0 \)
  - the mean is zero iff \( \phi_0 = 0 \)

• Given \( \mu \) we can find the variance (2nd central moment)
  
  using \( \phi_0 = (1 - \phi_1) \mu \)
  
  \[ r_t - \mu = \phi_1 (r_{t-1} - \mu) + a_t \]
  
  \[ = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \ldots \]
  
  \[ = \sum_{i=0}^{\infty} \phi_1^i a_{t-i} \]

  - This and independence of \( \{a_t\} \)

  \[ E[(r_t - \mu) a_{t+1}] = 0 \]
  
  \[ Cov(r_{t-1}, a_t) = E[(r_{t-1} - \mu) a_t] = 0 \]

  - \( r_{t-1} \) occurred in the past and \( a_t \) does not depend on past information
• Variance calculation:

\[
E (r_t - \mu)^2 = E (\phi_1 (r_{t-1} - \mu) + \alpha_t)^2
given -1 > \phi_1 > 1, we can show that the mean and variance of \( r_t \) are finite.
\]

\[
Var (r_t) = \phi_1^2 Var (r_{t-1}) + \sigma_\alpha^2
\]

\[
Var (r_t) = \frac{\sigma_\alpha^2}{1 - \phi_1^2}
\]

Autocorrelation Function of an AR(1) Model

• Taking our (centralized) AR(1) model

\[
\begin{align*}
\gamma_\ell &= \phi_1 \gamma_{\ell-1} + \sigma_\alpha^2, \quad \ell \geq 0 \\
\gamma_0 &= 1
\end{align*}
\]

We can take this result and look at autocorrelations

\[
\begin{align*}
\gamma_\ell &= \begin{cases} 
\phi_1 \gamma_1 + \sigma_\alpha^2 & \text{if } \ell = 0 \\
\phi_1 \gamma_{\ell-1} & \text{if } \ell > 0
\end{cases}
\end{align*}
\]

Summary Slide...

\[
\begin{align*}
\gamma_\ell &= \frac{\gamma_{\ell-1}}{1 - \phi_1^2}, \quad \ell \geq 0 \\
\gamma_0 &= 1
\end{align*}
\]

\[
\begin{align*}
\rho_\ell &= \phi_1 \rho_{\ell-1}, \quad \ell \geq 0 \\
\rho_0 &= 1
\end{align*}
\]

A weakly stationary series decays exponentially with rate \( \phi_1 \).
The model which we can estimate is...

\[ r_t = \phi_0 + \phi_1 r_{t-1} + a_t \]

\[ E (r_t) = \mu = \frac{\phi_0}{1-\phi_1} \]

\[ \text{Var} (r_t) = \gamma_0 = E (r_t - \mu)^2 = \phi_1^2 \text{Var} (r_{t-1}) + \sigma_a^2 \rightarrow \gamma_0 = \frac{\sigma_a^2}{1-\phi_1^2} \]

and the covariaviances and autocorrelations are given by:

\[ r_t = \phi_0 + \phi_1 r_{t-1} + a_t \rightarrow (r_t - \mu) = \phi_1 (r_{t-1} - \mu) + a_t \]

\[ E [(r_t - \mu) (r_{t-\ell} - \mu)] = \gamma_\ell = E [\phi_1 (r_{t-1} - \mu) (r_{t-1} - \mu)] + E [a_t (r_{t-1} - \mu)] \]

\[ \gamma_\ell = \begin{cases} \phi_1, & \text{if } \ell = 0 \\ \phi_1 \gamma_{\ell-1} + \sigma_a^2, & \text{if } \ell > 0 \end{cases} \]

*because:

\[ E [a_t (r_t - \mu)] = \phi_1 E [a_t (r_{t-1} - \mu)] + E (a_t^2) = \sigma_a^2 \]

What this all buys...

- There exists a particular equivalence between the two models:

\[ r_t - \mu = \phi_1 (r_{t-1} - \mu) + a_t \]

\[ r_t - \mu = a_t + \phi_1 a_{t-1} + \phi_2 a_{t-2} + \phi_3 a_{t-3} + ... \]

\[ r_t = \mu + \sum_{i=0}^{\infty} \phi_i a_{t-i} \]

\[ \therefore \phi_1 = \psi_i \]

- This gives us a way to think about autocorrelations:

\[ \gamma_\ell = \phi_1 \gamma_{\ell-1} \text{ if } \ell > 0 \]

\[ \rho_\ell = \frac{\gamma_\ell}{\gamma_0} = \frac{\phi_1 \gamma_{\ell-1}}{\gamma_0} = \phi_1 \rho_{\ell-1} \]

\[ \rho_0 = 1 \rightarrow \rho_\ell = \phi_1^\ell \]

- NOTE: All of these calculations were based on the ASSUMPTION that the AR(1) specification is adequate!
AR(2) Model

\[ r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \omega_t \]

- As in the AR(1) case:
  
  \[ E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2} \]

  Using \( \phi_0 = \mu (1 - \phi_1 - \phi_2) \)

  \[ r_t - \mu = \phi_1 (r_{t-1} - \mu) + \phi_2 (r_{t-2} - \mu) + \omega_t \]

- Multiplying by \((r_{t-\ell} - \mu)\) and taking expectations delivers the moment equation of a stationary AR(2)
  
  \[ \gamma_\ell = \phi_1 \gamma_{\ell-1} + \phi_2 \gamma_{\ell-2}, \ell > 0 \]

- Using the Backshift operator, this can be written as a 2nd order difference equation
  
  \[ (1 - \phi_1 B - \phi_2 B^2) \rho_\ell = 0 \]

  This equation determines the properties of a stationary AR(2) series, and is very important for forecasts

  - Consider the corresponding polynomial equation

    \[ 1 - \phi_1 x - \phi_2 x^2 = 0 \]

    \[ x = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \]

    - The (two) solutions are the characteristic roots of the AR(2) model

      \( (\omega_1, \omega_2) \)

- Divide by \( \gamma_0 \) (the variance)

  \[ \rho_\ell = \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-2}, \ell \geq 2 \]

- If the solutions are real valued:

  \[ (1 - \omega_1 B)(1 - \omega_2 B) \rho_\ell = 0 \]

  which shows that the AR(2) is just a mixture of two AR(1) models (two exponential decays)

- If \( \phi_1^2 + 4\phi_2 < 0 \), we have a complex conjugate pair of solutions

  * This delivers CYCLES with average length given by

    \[ k = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]} \]
Identifying AR models

- the order $p$ of an AR($p$) is unknown and must be empirically specified

- Partial Autocorrelation function (PACF)

  $$r_t = \phi_{0,1} + \phi_{1,1}r_{t-1} + e_{1t}$$
  $$r_t = \phi_{0,2} + \phi_{1,2}r_{t-1} + \phi_{2,2}r_{t-2} + e_{2t}$$
  $$r_t = \phi_{0,3} + \phi_{1,3}r_{t-1} + \phi_{2,3}r_{t-2} + \phi_{3,3}r_{t-3} + e_{3t}$$
  and so on...

  - The least squares estimate of $\phi_{i,j}$ is $\hat{\phi}_{i,j}$

  - Comparing models shows the contribution of adding an additional lag term to the model

- For an AR($p$) model, the lag-$p$ sample PACF ($\hat{\phi}_{p,p}$) should not be zero, but $\hat{\phi}_{j,j}$ should be close to zero $\forall j > p$

- Information Criteria (likelihood based determination)

  - Akaike Information Criterion

    $$AIC = -\frac{2}{T} \ln(\text{likelihood}) + \frac{2}{T} \text{(number of parameters)}$$

    For a Gaussian AR($\ell$) model

    $$AIC(\ell) = \ln(\hat{\sigma}^2) + \frac{2\ell}{T}$$

  - Bayesian Information Criterion

    $$BIC(\ell) = \ln(\hat{\sigma}^2) + \frac{2\ln(T)}{T}$$

    - Compute these for several values of $\ell$, and select the model which delivers the minimum value
Estimation

\[ r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \ldots + \phi_p r_{t-p} + a_t, \ t = p + 1, \ldots, T \]

OLS delivers the fitted model:

\[ \hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \hat{\phi}_2 r_{t-2} + \ldots + \hat{\phi}_p r_{t-p} \]

with associated residuals:

\[ \hat{a}_t = r_t - \hat{r}_t \]

The residual series is used to find the sample variance:

\[ \hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{T - 2p - 1} \]

Model Checking

- By Assumption, if the model is an AR(p) then the resulting residuals \((a_t)\) should be white noise

- You can run the Ljung-Box statistics to check if there are no joint autocorrelations in the residuals

\[ a_t = v_0 + v_1 a_{t-1} + v_2 a_{t-2} + \ldots + v_p a_{t-p} + \varepsilon_t, \ t = p + 1, \ldots, T \]

- If there are, then you are omitting important information in your predictions

Goodness of Fit

- The most common statistic is the \(R^2\)

\[ R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{Total sum of squares}} \]

\[ R^2 = 1 - \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{\sum_{t=p+1}^{T} (r_t - \bar{r})^2} = 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_r^2} \]

- What does this \(R^2\) tell us?

Forecasting

- Suppose we are at time \(h\) and interested in forecasting \(r_{h+\ell}\)

  - \(h\) is the forecast origin

  - \(\ell\) is the forecast horizon

- let \(\hat{r}_h(\ell)\) be the forecast of \(r_{h+\ell}\) (the \(\ell\)-step ahead forecast of \(r_t\))

\[ E \left\{ (r_{h+\ell} - \hat{r}_h(\ell))^2 \mid F_h \right\} \leq \min_y E \left\{ (r_{h+\ell} - y)^2 \mid F_h \right\} \]

- a squared error loss function
• 1-step ahead

\[ r_{h+1} = \phi_0 + \phi_1 r_h + \ldots + \phi_p r_{h+p} + a_{h+1} \]

the point forecast of \( r_{h+1} \) given \( F_h = \{ r_h, r_{h-1}, \ldots \} \) is the conditional expectation

\[ \hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i} \]

and the error is given by

\[ e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1} \]

If \( a_t \) is normally distributed white noise, then a 95% confidence around the forecast of \( r_{h+1} \) is \( \hat{r}_h(1) \pm 1.96\sigma_a \)

• Multistep ahead

\[ r_{h+c} = \phi_0 + \phi_1 r_{h+c-1} + \ldots + \phi_p r_{h+c-p} + a_{h+c} \]

Since we only have info \( F_h \) these forecasts get built up recursively

\[ \hat{r}_h(\ell) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - 1) \]

- For any AR(p) model, \( \hat{r}_h(\ell) \to E(r_t) \) as \( \ell \to \infty \)