Introduction

Kevin Kotzé

Contents

Introduction

- A time series is an ordered temporal variable
- After investigating the behaviour of these variables we may gain a better understanding of the past and in certain cases we can predict the future
- Time Domain - modelling current observations on past observations
  - Box-Jenkins methods
  - State-space methods
- Frequency domain - modelling period or systematic variations
  - Includes, Fourier methods, power spectra's & wavelet transforms
- These methodologies may be applied in the construction of different models and filters: ARMA, LLM, VAR, SVAR, CVAR, GARCH, MVGARCH, SV, DFM, MSW, . . .

Introduction

- Why do we need a separate area of study for investigations that involve time series variables?
- Consider the simple linear regression model

$$y_t = \underbrace{x_t^\top \beta}_{\text{explained}} + \underbrace{\varepsilon_t}_{\text{unexplained}}, \qquad t = 1, \ldots, T$$

- Errors should not be serially correlated for least squares estimates in such a model:
- $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots] = 0$ and
- $\mathbb{E}\left[\varepsilon_i \varepsilon_{t-j}\right] = 0$, for $j \neq 0$
- If there is serial correlation in the errors then estimates are inefficient and the standard errors that are associated with the coefficient estimates may be incorrect

Introduction

- Most economic and financial time series exhibit some form of serial correlation
  - If economic output is large during the previous quarter then there is a good chance that it is going to be large in the current quarter

- A change that arises in the current period may only affect other variables in the distant future

- Particular shock may affect variables over successive quarters
  - Hence, we need to start thinking about the dynamic structure of the system that we are investigating

- For these reasons, modelling the time dependency in the data may be challenging and if we are unable to incorporate this feature of the variables in the *explained* part of the model then the errors may display serial correlation

- Hence, most traditional time series models seek to provide a suitable description of the serial dependence in the data to provide robust results

Introduction

- The literature on various forms of time series analysis is particularly large
- A large part of the literature is concerned with forecasting, which remains an important area of study
- There is also an important section that tests various hypotheses (or theories)
- These models could also be used for policy investigations after we satisfy a number of important criteria
- Irrespective of the objective we would usually always need to identifying the *dynamic evolution* of these variables
- To accomplish this task we may need to decompose the time series into its constituent components

Figure 1: Decomposition of Time Series

Dynamic components of a time series

- Consider the following example:

$$
\begin{aligned}
Trend : T_t &= 1 + 0.05t \\
Seasonal : S_t &= 1.5\cos(t\pi \times 0.166) \\
Irregular : I_t &= 0.5 I_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i. i. d. } \mathcal{N}(0, \sigma^2_{\varepsilon_t})
\end{aligned}
$$

where ε_t is a random disturbance

- The trend component is deterministic
- The seasonal is also deterministic and uses a cosine function to impart cyclical behaviour
- The irregular component contains a stochastic term, which may be described by a statistical distribution

Figure 2: Real World Data: South African Data

Economic data

- Consider the information content that is contained in the series
- Most economic data contains trends, seasonals, and irregular components
- Usually measured in discrete time with relatively long intervals
- Could represent a stock (i.e. capital) or a flow (i.e. investment)
- This data may be expressed as rates, indices or totals
- Be cautions of using interpolation to generate higher frequency data
- Many economic variables are subject to large revisions

Economic data

- Most economic data requires a specific transformation before a subsequent analysis can be conducted
- Common transformations include the derivation of: growth rates $[\log(GDP_t/GDP_{t-1})]$, annualised rates $[(1+(i_t/100))^{(1/12)}-1]$, cyclical components, etc.
- One may need to ensure that the variables are measured at the same frequency
- Most countries follow globally accepted measurement practices

Financial data

- Data on stock prices and indices is overwhelming:
  - Does the data contain true trading prices, quotes, or proxies for trading prices?
  - Do the prices include transaction costs, commissions & effects of tax transfers?
  - Is the market sufficiently liquid?
  - Have the prices been adjusted for inflation?
  - Have they been correctly discounted?

- May only be interested in buyer initiated (ask) or seller initiated (bid) orders

Financial data

- The frequency of the data may influence the measure trading activity/returns and the associated volatility
  - Is it feasible to use extremely high frequency data and what does it represent?
  - Introduction of structural breaks could also have a dramatic effect on your results
  - Consider the implications of limiting data to a particular sub-sample
- Transformation to returns generally displays stationary behaviour
- Represents a complete scale-free summary about an investment opportunity
- Measures of risk and volatility may involve additional complexities

Processes

- Time series is a collection of observations indexed by the date of each realisation
- Using notation that starts at time, $t = 1$ and using the end point, $t = T$,

$$\{y_1, y_2, y_3, \ldots, y_T\}$$

- Time index can be of any frequency (e.g. daily, quarterly, etc.)

Deterministic & stochastic processes

- Deterministic process will always produce the same output from a given starting condition or initial state
  - No element of randomness, i.e. $Trend_t = 1 + 0.05t$

- Stochastic process has some indeterminacy that relates to the future evolution of the process
  - Usually described by some form of statistical distribution
  - Examples include: white noise processes, random walks, Brownian motions, Markov chains, martingale difference sequences, etc.

Stochastic processes: White noise

- Serially uncorrelated random variables with zero mean and finite variance
- For example, errors may be characterised by a Gaussian white noise process, where such a variable has a normal distribution
- Slightly stronger condition is that they are independent from one another

$$\varepsilon_t \sim \text{i. i. d.} \mathcal{N}(0, \sigma^2_{\varepsilon_t})$$

- Notice three implications of this assumption:
- $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-1}, \ldots] = 0$
- $\mathbb{E}[\varepsilon_t \varepsilon_{t-j}] = \textbf{cov}[\varepsilon_t \varepsilon_{t-j}] = 0$
- $\textbf{var}[\varepsilon_t] = \textbf{cov}[\varepsilon_t \varepsilon_t] = \sigma^2_{\varepsilon_t}$

Figure 3: Gaussian White Noise Process

Random walk

- Random walk would imply that the effect of a shock is permanent

$$y_t = y_{t-1} + \varepsilon_t$$

- Given the starting value y_0, and using recursive substitution, this process could be represented as

$$y_t = y_0 + \sum_{j=1}^{t} \varepsilon_j$$

- Since the effect of past shocks do not dissipate we say it has an infinite memory

- Behaves like a meandering drunken sailor, who's next movement is extremely difficult to predict

Figure 4: Random Walk - Simulated Time Series

Figure 5: Random Walk - Effect of Shock [$y_{-1} = 0, \varepsilon_0 = 1$ and $(\varepsilon_1, \ldots) = 0$]

Random walk plus drift

- Random walk plus a constant term

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

- For the starting value of zero, this could be represented as

$$y_t = \mu \cdot t + \sum_{j=1}^{t} \varepsilon_j$$

- Shocks have permanent effects and are influenced by the drift
- If you were to include an equivalent constant in the expression for the white noise process you will note that it imparts different properties on the process

Figure 6: Random Walk plus Drift - Simulated Time Series [Dotted: $\mu = 1.2$ & Solid: $\mu = 0.5$]

Figure 7: Random Walk with Drift - Effect of Shock [$y_{-1} = 0$, $\mu = 1.2$, $\varepsilon_0 = 1$ and $(\varepsilon_1, \ldots) = 0$]

Figure 8: Different Random Walks

Autoregressive process

- An $AR(1)$ process describes situations where the present value of a time series is a linear function of the previous observation

$$y_t = \phi y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \sim \text{i. i. d. } \mathcal{N}(0, \sigma^2_{\varepsilon_t})$$

- Know something about the *conditional* distribution of y_t given y_{t-1}

- After repeated substitution, and with a starting value of zero, it would take the form

$$y_t = \phi^j \sum_{j=1}^{t} \varepsilon_j$$

- Could include several lags (i.e. $AR(p)$ model) and the distribution of the error term could take many forms

- Think about the implication of future values of y_t when $\phi = 0, 0.5, 1,$ or 1.5?

Figure 9: $AR(1)$ - Simulated Time Series [$\phi = 0.9$]

Figure 10: $AR(1)$ - Effect of Shock [$\phi = 0.9$]

Moving average process

- $MA(q)$ model describes a time series by the weighted sum of the current and previous errors

- Describes examples where it takes a bit of time for the error (or "shock") to dissipate

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

- This type of expression may be used to describe a wide variety of stationary time series processes

Figure 11: $MA(1)$ - Simulated Time Series [$\theta = 0.7$]

Figure 12: $MA(1)$ - Effect of Shock [$\theta = 0.7$]

ARMA process

- A combination of these modes is termed an $ARMA(1,1)$

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

- where an $ARMA(p,q)$ takes the form

$$\begin{aligned} y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \ldots \\ \ldots + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \end{aligned}$$

- This model was popularized by Box & Jenkins, who developed a methodology that may be used to identify the terms that should be included in the model

Long memory & fractional differencing

- Most $AR(p)$, $MA(q)$ and $ARMA(p,q)$ processes are termed short-memory process because the coefficients in the representation are dominated by exponential decay

- Long-memory (or persistent) time series are considered intermediate compromises between the short-memory models and integrated nonstationary processes

- Long periods during which observations tend to be at a high level and similar long periods during which observations tend to be at a low level

Moments of distribution

- Distributions often summarised by first (mean) and second (variance) moments
- Higher order moments may be of interest (skewness & kurtosis)
- Always important to distinguish between the unconditional and conditional distributions

Moments of distribution

- The first moment of a stochastic process is the average of y_t over all possible realisations

$$\bar{y} = \mathbb{E}\left[y_t\right], \quad t = 1, \ldots, T$$

- The second moment is defined as the variance

$$\textsf{var}[y_t] = \mathbb{E}\left\{y_t \; y_t\right\} = \mathbb{E}\left\{\left(y_t - \mathbb{E}\left[y_t\right]\right)^2\right\}, \quad t = 1, \ldots, T$$

- And the covariance, for j:

$$\begin{aligned}\textsf{cov}[y_t, y_{t-j}] &= \mathbb{E}\left\{y_t \; y_{t-j}\right\} \\ &= \mathbb{E}\left\{\left(y_t - \mathbb{E}\left[y_t\right]\right)\left(y_{t-j} - \mathbb{E}\left[y_{t-j}\right]\right)\right\}, \; t = j+1, \ldots, T\end{aligned}$$

Conditional moments

- Conditional distribution is based on past realisations of a random variable
- For the $AR(1)$ model

$$y_t = \phi y_{t-1} + \varepsilon_t$$

- where $\varepsilon_t \sim \text{i. i. d.} \mathcal{N}(0, \sigma^2)$ is Gaussian white noise and $|\phi| < 1$
- Conditional moments satisfy

$$\begin{aligned}
&\mathbb{E}\left[y_t | y_{t-1}\right] = \phi y_{t-1} \\
&\text{var}[y_t | y_{t-1}] = \mathbb{E}[\phi y_{t-1} + \varepsilon_t - \phi y_{t-1}]^2 = \mathbb{E}[\varepsilon_t]^2 = \sigma^2 \\
&\text{cov}\left[\left(y_t | y_{t-1}\right), \left(y_{t-j} | y_{t-j-1}\right)\right] = 0 \;\; \text{for } j > 1
\end{aligned}$$

Conditional moments

- Conditioning on y_{t-2} for y_t

$$
\begin{aligned}
\mathbb{E}\left[y_t|y_{t-2}\right] &= \phi^2 y_{t-2} \\
\textsf{var}[y_t|y_{t-2}] &= (1+\phi^2)\sigma^2 \\
\textsf{cov}\left[(y_t|y_{t-2}),\left(y_{t-j}|y_{t-j-2}\right)\right] &= \phi\sigma^2 \;\; \text{for } j=1 \\
\textsf{cov}\left[(y_t|y_{t-2}),\left(y_{t-j}|y_{t-j-2}\right)\right] &= 0 \;\; \text{for } j>1
\end{aligned}
$$

Unconditional moments

- Unconditional distribution has slightly different moments for the $AR(1)$ model

$$y_t = \phi y_{t-1} + \varepsilon_t$$

- where $\varepsilon_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2)$ is Gaussian white noise and $|\phi|<1$

- Unconditional moments satisfy

$$\begin{aligned}
\mathbb{E}\left[y_t\right] &= 0 \\
\text{var}[y_t] &= \frac{\sigma^2}{1-\phi} \\
\text{cov}[y_t \; y_{t-j}] &= \phi^j \text{var}(y_t)
\end{aligned}$$

Stationarity: Strictly stationary

- Time series is strictly stationary if for any values

$$\{j_1, j_2, \ldots, j_n\},$$

- the joint distribution of

$$\{y_t, y_{t+j,1}, y_{t+j,2}, \ldots, y_{t+j,n}\}$$

- depends only on the intervals separating the dates

$$\{j_1, j_2, \ldots, j_n\}$$

- and not on the date itself, t

Stationarity: Covariance stationary

- If neither the mean, $\bar{y}$, nor the covariance, **cov**(y_t, y_{t-j}), depend on the date, t
- Then the process for y_t is said to be covariance (weakly) stationary, where for all t and any j

$$\mathbb{E}\left[y_t\right] = \bar{y}$$
$$\mathbb{E}\left[\left(y_t - \bar{y}\right), \left(y_{t-j} - \bar{y}\right)\right] = \textbf{cov}(y_t,\ y_{t-j})$$

- When referring to stationarity in the remainder of the course we refer to covariance stationarity
- Note that the process $y_t = \alpha t + \varepsilon_t$ would not be stationary, as the mean clearly depends on t
- In addition, we saw that the unconditional moments of the $AR(1)$ with $|\phi| < 1$ had a mean and covariance that did not depend on time

Autocorrelation function (ACF)

- For a stationary process we can plot the standardised covariance of the process over successive lags

- Makes use of the autocovariance function, which is denoted $\gamma(j) \equiv \textbf{cov}\left(y_t, y_{t-j}\right)$ for $t = 1, \ldots, T$

- Autocovariance function may be standardised by dividing each function by the variance to derive the ACF for successive values of j

$$\rho(j) \equiv \frac{\gamma(j)}{\gamma(0)}$$

- To display the results of the ACF we usually plot $\rho(j)$ against (non-negative) j to illustrate the degree of persistence in a variable

Partial autocorrelation function (PACF)

- With an $AR(1)$ process, $y_t = \phi y_{t-1} + \varepsilon_t$, the ACF would suggest y_t and y_{t-2} are correlated even though y_{t-2} does not appear in the model
- This result is due to the pass-through, where we noted that $y_t = \phi^2 y_{t-2}$ when performing the recursive substitution exercise
- PACF eliminates effects of the pass-through and focuses on the independent relationship between y_t and y_{t-2}
- Hence, the PACF (y_t, y_{t-j}) eliminates the impact of the intervening correlations between y_{t-1} and y_{t-j-1}

Partial autocorrelation function (PACF)

- To construct a PACF one would usually make use of the following steps:
  - Demean the series $(y_t^\star = y_t - \bar{y})$
  - Form the $AR(1)$ equation, $y_t^\star = \phi_{11} y_{t-1}^\star + v_t$
  - In this case, ϕ_{11} is equivalent to $\rho(1)$ in the ACF
  - Equivalent to the first coefficient in the PACF as there are no intervening values between y_t and y_{t-1}
  - Now form the second-order autoregression, $y_t^\star = \phi_{21} y_{t-1}^\star + \phi_{22} y_{t-2}^\star + v_t$
  - In this case, ϕ_{22} is the PACF between y_t and y_{t-2}, since the effects of $y_{t-1}^\star$ on $y_t^\star$ are captured by the ϕ_{21}, which isolate the effects of y_{t-1} on y_t
  - etc.

Q-statistic

- Box-Ljung Q-statistic tests whether series is white noise
- Tests whether a group of autocorrelations differ from zero
- Tests the "overall" randomness based on a number of lags

$$Q(k) = T(T+2)\sum_{j=1}^{k}\frac{\rho_j^2}{T-j}$$

- where ρ refers to the residual autocorrelation from lag j
- In this case we would express ρ as

$$\rho_k = \frac{\sum_{t=1}^{T-k}(\varepsilon_t - \bar{\varepsilon})(\varepsilon_{t+k} - \bar{\varepsilon})}{\sum_{t=1}^{T}(\varepsilon_t - \bar{\varepsilon})^2}$$

- where $\bar{\varepsilon}$ is the mean of the T residuals

Figure 13: Serial Correlation

Impact multipliers

- Used to investigate the effects of specific shocks
  - Estimate the response of GDP growth after unexpected 1% increase in demand
  - How will the exchange rate react to an unexpected increase in the interest rate

- Assuming stationarity, any $AR(p)$ process can be written as an infinite order moving average process (later in course)

- Implies that an $AR(1)$ process may be written as

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

- Suggests y_t can be described by past & present errors / shocks

Impact multipliers

- Assume the dynamic simulation started at time j, taking $y_{t-(j+1)}$ as given
- Effect of a change in the initial shock on y_t is then

$$\frac{\partial y_t}{\partial \varepsilon_{t-j}} = \phi^j$$

- Termed the dynamic multiplier that depends only on j, ε_{t-j}, and y_t
- This expression does not depend on t, (date of observation)

Impulse response functions (IRF)

- Cumulative effect of temporary shock is then used to derive the IRF:

$$\sum_{j=0}^{\infty} \frac{\partial y_t}{\partial \varepsilon_{t-j}} = 1 + \phi + \phi^2 + \cdots + \phi^j = \frac{1}{(1-\phi)}$$

- Different values of ϕ produce a variety of responses in y_t
- When $|\phi| < 1$, the process decays geometrically towards zero
- When $0 < \phi < 1$, there will be a smooth decay
- When $0 > \phi > -1$ there will be an oscillating decay
- We say that a system described in this way is stable
- Dynamic multipliers can be moved forward in time, such that $\frac{\partial y_{t+j}}{\partial \varepsilon_t} = \phi^j$

Conclusion

- Overview of fundamental concepts that we will apply throughout this course
- Provides challenge for regression models, as serial correlation in an error term provides inefficient standard errors
- Real world data may contain trends, seasonals and varying degrees of persistence
- Considered statistical properties of selected time series processes
    - Random walk: effect of errors do not disappear $\Rightarrow$ difficult to predict future behaviour
    - $AR(p)$ and $MA(q)$: effect of errors dissipate $\Rightarrow$ so we would need to model the time taken to revert to an appropriate long-term value
- Other essential tools include autocorrelation functions, hypotheses for serial correlation, impact multipliers and impulse response functions