1 Univariate and Multivariate Stochastic Processes

1.1 Univariate Processes

Macroeconomic time series can often be represented by an autoregressive moving-average process:

\[ y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}, \]  

(1)

where \( \epsilon_t \sim i.i.d.(0, \sigma_\epsilon^2) \). This type of equation is referred to as an ARMA\((p,q)\) process. An ARMA process can be used to forecast and to summarize the dynamics of a time series.

1.1.1 Stationarity

A time series is stationary if the mean, variance, and autocovariances do not change over time:

\[
\begin{align*}
E(y_t) &= \mu \\
var(y_t) &= E[(y_t - \mu)^2] = \sigma_y^2 = \gamma_0 \\
cov(y_t, y_{t-s}) &= E[(y_t - \mu)(y_{t-s} - \mu)] = \gamma_s,
\end{align*}
\]

for all \( s \). Consider two examples of stationary processes: an AR(1) and an MA(1) process.

Example #1. **AR(1) process.** An AR(1) process takes the form

\[ y_t = \phi_1 y_{t-1} + \epsilon_t, \]  

(2)

where \( |\phi_1| < 1 \). Repeated substitutions produce an MA\((\infty)\) process:

\[ y_t = \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i}. \]  

(3)

The first and second moments are

\[
\begin{align*}
E(y_t) &= 0 = \mu \\
var(y_t) &= \sigma_y^2/(1 - \phi_1^2) = \gamma_0 \\
cov(y_t, y_{t-s}) &= \phi_1^s \gamma_0.
\end{align*}
\]
Example #2. MA(1) process. An MA(1) process takes the form

\[ y_t = \epsilon_t + \theta_1 \epsilon_{t-1}. \]  

The first and second moments are

\[ E(y_t) = 0 = \mu \]
\[ \text{var}(y_t) = (1 + \theta_1^2)\sigma_\epsilon^2 = \gamma_0 \]
\[ \text{cov}(y_t, y_{t-1}) = \theta_1 \sigma_\epsilon^2 = \gamma_1 \]
\[ \text{cov}(y_t, y_{t-s}) = 0, \ s > 1. \]

1.1.2 Box-Jenkins Analysis

Box-Jenkins analysis is a methodology to fit an ARMA(p,q) process to a time series for the purpose of forecasting. The first step is to check whether the series contains a trend. If it does, the trend is removed by differencing the data to make it stationary.

- **Step 1. Identification.**
  The first step is to identify the order, p and q, of the ARMA process. This is accomplished using the autocorrelations (\( \rho_s = \gamma_s / \gamma_0 \)) and the partial autocorrelations (last coefficient of a regression on an AR(s) process, \( \rho^*_s \)). Each ARMA process has a unique combination of autocorrelations \( \rho_s \) and partial autocorrelations \( \rho^*_s \). The sample autocorrelations and partial autocorrelations are matched with their theoretical counterparts to identify the best-fitting ARMA(p,q) process. This is more of an art than a science.

- **Step 2. Estimation.**
  Once the ARMA(p,q) process is identified, the \( \phi \) and \( \theta \) coefficients are estimated using maximum likelihood methods.

- **Step 3. Diagnostic Testing.**
  The next step is to test if the residuals are random noise. This is accomplished by testing whether the autocorrelations and partial autocorrelations of the residuals are statistically significant.

- **Step 4. Forecasting.**
  The estimated ARMA(p,q) process and past observations on \( y_t \) can be used to forecast future values of \( y_t \). Most statistical packages will also present confidence intervals around the forecasts.
1.2 Multivariate Processes

One of the most common multivariate time series models is the Vector AutoRegressive (VAR) process. The standard VAR model treats all variables as endogenous and does not impose any economic restrictions. Structural VAR models (next semester) impose restrictions on the VAR consistent with economic theory.

The VAR model can handle multiple variables, but for simplicity let’s consider a bivariate system similar to the one in Sims (1972).\(^1\) The structural macro model with money \(m_t\) and income \(y_t\) is

\[
y_t - a_{12} m_t = b_{11} y_{t-1} + b_{12} m_{t-1} + \epsilon_t^y
\]

\[
m_t - a_{21} y_t = b_{21} y_{t-1} + b_{22} m_{t-1} + \epsilon_t^m.
\]

(5)

Re-written in matrix form, we have

\[
A \begin{bmatrix} y_t \\ m_t \end{bmatrix} = B \begin{bmatrix} y_{t-1} \\ m_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^m \end{bmatrix},
\]

(6)

where

\[
A = \begin{bmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.
\]

Assuming \(A\) is invertible, we can write the VAR as

\[
\begin{bmatrix} y_t \\ m_t \end{bmatrix} = A^{-1} B \begin{bmatrix} y_{t-1} \\ m_{t-1} \end{bmatrix} + A^{-1} \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^m \end{bmatrix}
\]

\[
= \Pi \begin{bmatrix} y_{t-1} \\ m_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^m \end{bmatrix}.
\]

(7)

This is a VAR model with one lag, but models with more lags are possible. The VAR is a reduced-form version of the structural model and can be used to test for Granger causality, generate impulse response functions, and calculate variance decompositions.

1.2.1 Granger Causality

Let’s continue with the money-income example. For example, money is said to Granger-cause income if past values of money \((m_{t-1})\) are able to explain current income \((y_t)\) after accounting for past values of income \((y_{t-1})\). This definition can be extended to handle more than one lag. For the case of one lag, Granger

causality from money to income can be tested using a $t$ statistic to see if $\pi_{12} = 0$. From the matrix

$$
\Pi = \begin{bmatrix}
\pi_{11} & \pi_{12} \\
\pi_{21} & \pi_{22}
\end{bmatrix}.
$$

Multiple lags would require an $F$ test. Conversely, to see if income Granger-causes money we would test if $\pi_{21} = 0$.

### 1.2.2 Impulse Response Functions

The VAR can be re-written as

$$
x_t = \Pi x_{t-1} + v_t,
$$

where $x_t = (y_t, m_t)'$ and $v_t = (v_t^y, v_t^m)'$. Impulse response functions (IRFs) are a graph of how $x_t$ responds over time to a one-time, one-unit shock in $v_t$. Of course, the vector $v_t$ includes reduced-form errors so the IRFs may be difficult to interpret. Methods for identifying the structural shocks $\epsilon_t^y$ and $\epsilon_t^m$ will be covered next semester.

### 1.2.3 Variance Decompositions

The forecast error decomposition gives the proportion of variation in a series due to its own shocks and other shocks. The variance decomposition can be calculated at various forecasting horizons.

### 2 Decomposition: Trends and Cycles

Many macroeconomic time series can be decomposed in the following way:

$$
y_t = \text{trend} + \text{stationary component} + \text{noise} = \tau_t + y_t^{sc} + \epsilon_t.
$$

Growth economists care about the "trend" ($\tau_t$). Business-cycle researchers care about the "stationary component" ($y_t^{sc}$). The "noise" ($\epsilon_t$) gives us an idea of how much is unexplained.

The stationary component $y_t^{sc}$ can be modeled as an ARMA($p,q$) process. The trend $\tau_t$ can take two forms:

1. **Deterministic trend.** The simplest deterministic trend is $\tau_t = at$, where $y_t$ increases by a fixed amount $a$ each period. This is a simple linear trend. Another possibility is $\tau_t = \exp(at)$, so that $y_t$ exhibits deterministic exponential growth. The deterministic-trend model is often referred to as trend stationary (TS).
2. Stochastic trend. If $\tau_t = y_{t-1}$, then the time series contains a stochastic trend and is difference stationary (DS). It is also said that $y_t$ is integrated of order one $I(1)$ and contains a unit root. The most famous case is when $y_t^{sc} = 0$ and $\tau_t = y_{t-1}$, the so-called random walk model. In this case,

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

or after repeated substitutions

$$y_t = \sum_{i=0}^{\infty} \epsilon_{t-i}.$$  \hspace{1cm} (10)

This shows that a random walk has a long memory, is not mean-reverting, and has infinite variance.

When $y_t^{sc} = 0$ and $\tau_t = a + y_{t-1}$, we get a random walk with drift:

$$y_t = a + y_{t-1} + \epsilon_t.$$

### 2.1 Detrending Methods

There are many different methods to remove the trend from a time series. Three common methods are linear detrending, first-differencing, the Hodrick-Prescott (HP) filter. Statistical tests (e.g., Dickey-Fuller test) and economic theory can guide us to the best detrending method.

- **Method #1. Linear Detrending.** This method involves regression of $y_t$ on time, $t$. The residuals are then the stationary component (e.g., business-cycle fluctuations). This method is appropriate if the series is TS.

- **Method #2. First Differencing.** This method involves calculating $\Delta y_t = y_t - y_{t-1}$, the first-difference of $y_t$. This will produce the stationary component and is appropriate if the series is DS. If $y_t$ is measured in natural logs, $\Delta y_t$ gives the growth rate of $y_t$.

- **Method #3. Hodrick-Prescott Filter.** This is a flexible detrending procedure which chooses $\tau_t$ to minimize

$$\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \left\{ \sum_{t=2}^{T} [\tau_{t+1} - \tau_t - (\tau_t - \tau_{t-1})]^2 \right\}$$

where $\lambda$ is the smoothing parameter. Business-cycle fluctuations are defined as $y_t - \tau_t$. Higher values of $\lambda$ result in a smoother trend. For instance,

- $\lambda = 0$ results in $\tau_t = y_t$ and there are no cycles.
- $\lambda \to \infty$ results in $\tau_t = a + bt$ or a linear trend.
- Prescott chooses $\lambda = 1600$, which causes the HP filter to focus on cycles with periodicity of 8 years or less.
3 Calibration and Estimation

There are two methods to fit macroeconomic models to the data – calibration and estimation. Calibration and simulation of macroeconomic models became popular with the work of Kydland and Prescott (1982) "Time to Build and Aggregate Fluctuations". Estimation was the preferred method for analyzing large-scale Keynesian models and has made a recent revival with advancements in computing power.

- **Calibration.** The "deep" structural parameters of the model are selected (i) based on micro evidence and (ii) to fit various long-run averages, variances and co-variances of macro time series. The model is solved simulated to produce an artificial path for the macroeconomy. Various properties of the artificial and actual data can be compared to "test" the model. This is the preferred method in the hard sciences.

- **Estimation.** The structural or reduced-form parameters are estimated using econometric methods and macro data. The estimates can then be used to develop hypotheses that allow us to reject or "accept" the model. A common problem with macro models is simultaneity bias. Many of the right-side variables are endogenous so that "instrumental variable" methods are necessary.

To highlight the differences, consider the standard Neoclassical lifecycle consumption model. The optimal consumption rule takes the form

\[
\frac{\dot{c}(t)}{c(t)} = \frac{r - \rho}{\theta}. \tag{12}
\]

The calibration method would select values for \(\rho\) and \(\theta\) consistent with the micro evidence and values for \(r\) consistent with the long-run average for interest rates. Equation (12) could then be used to simulate an artificial path for \(c(t)\). This artificial path could be compared to actual consumption profiles to test the model.

The estimation method would form an estimating equation such as

\[
\ln(c_{t+1}) - \ln(c_t) = b_0 + b_1 r_t + \epsilon_t \tag{13}
\]

and test the theoretical hypotheses \(H_0: b_0 = -\rho/\theta\) and \(H_1: b_1 = 1/\theta\).