PhD Macroeconomics 1
4. Markov Chains and Non-Stationary Processes

Karl Whelan

School of Economics, UCD

Autumn 2023

Stochastic Processes

- Having look at the dynamics of deterministic models, we will now discuss **stochastic** models, i.e. models in which variables cannot be perfectly forecasted due to the presence of random elements.
- We will start by focusing on **Markov chains**. These are a simple kind of discrete-time stochastic process that have a finite range of possible values. We will use Markov chains later in the module when we discuss stochastic dynamic programming.
- After that, we will discuss continuously-valued discrete-time stochastic processes, usually just called **time series**.
- Modelling time series is a central part of modern macroeconomics. This isn't an econometrics course but we are, by necessity, going to have to cover some econometric issues that affect empirical research in macroeconomics.
- In this set of lecture notes, we will discuss **nonstationary time series**, specifically series with upward trends, covering some of the practical issues involved in modelling these series.
- Our next few lectures will focus on modelling stationary time series.

Part I

Markov Chains

Markov Processes

- A first-order Markov process y_t is a discrete-time process with the property that conditional on the current value of the process, future realisations are independent of y_{t-1} and other past values.
- An example of a Markov process is an **autoregressive process of order 1 (i.e. an AR(1) process)**.

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

  where ϵ_t is independent identically distributed process with mean zero.
- In macroeconomics, we often use Markov processes with discrete supports i.e. the process can only ever take one of the N values $d_1, ..., d_N$. This is known as a **Markov chain**.
- Markov chains are defined by their transition probabilities.

$$p_{ij} = \text{Prob } [y_{t+1} = d_j \mid y_t = d_i]$$

- A Markov chain's transition matrix P is a matrix of probabilities in which the ith row contains the probabilities for all possible outcomes next period given $y_t = d_i$.

Simulating Markov Chains

- Markov chains are easy to simulate in Matlab using the commands `dtmc` and `simulate`.
- For instance, see below for code that simulates 100 periods of a Markov process with three possible values $(1, 2, 3)$ and a transition matrix given by

$$P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

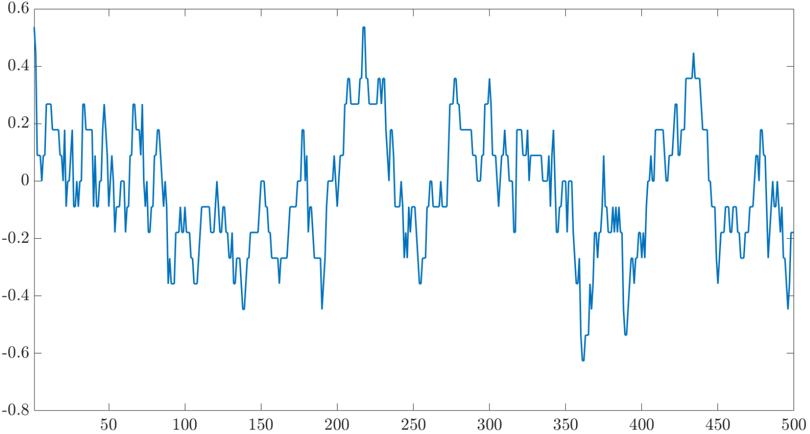
```
P = 1/3*ones(3);
mc = dtmc(P);
numsteps = 100;
State = simulate(mc,numsteps);
values = [1 2 3];
series = values(State);
plot(series);
```

- Note that Matlab uses a random number generator to pick the initial state.

Markov Chain Approximation to the AR(1) Model

- Markov chains can be used to approximate the behaviour of Markov processes with continuous supports, i.e. processes that can, in theory, take any value.
- We are going to use a Matlab programme called `rouwen.m` to approximate an AR(1) process using a Markov chain.
- This programme implements a method introduced by Geert Rouwenhorst in a 1995 paper.
- There are a number of different methods that get used to do this approximation but it appears that Rouwenhorst's method is the best. You can check out what exactly the method is if you want—I have made a paper on this available—but we don't have time to stop and cover it here.
- Among other features, Rouwenhurst's method produces series that match the mean, variance and first-order autocorrelation statistics of the AR(1) series that you are trying to mimic.
- The graph on the next page shows a sample run from approximating an AR(1) series with $\rho = 0.9$ using a 21-state Markov chain. Code for this is available on Brightspace.

Sample Simulation of a 21-State AR(1) Approximation



Changing Distributions Over Time

- Now let's consider a case where there are N possible states for a variable and we have a large population of agents, each of whom are allocated to one of these states.
- Let's denote the initial distribution of people across those states by a column vector of share values
$$\theta_0 = \begin{pmatrix} \theta_{0,1} \\ \theta_{0,2} \\ . \\ \theta_{0,N} \end{pmatrix}$$
where these shares sum to one.
- Now assume that the population is sufficiently large that the Law of Large Numbers applies to transitions. For example, if the transition matrix says there is a probability 0.5 that agents in state i remain in that state and 50 percent change that they switch to state j, then the outcome is exactly that.
- This means that distribution changes over time according to
$$\theta_1 = P'\theta_0 \Rightarrow \theta_N = (P')^N \theta_0$$

Example: Two-State Process

- Consider a two-state Markov chain with transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

- If the initial distribution across states is given by

$$\theta_0 = \begin{pmatrix} \theta_{0,1} \\ \theta_{0,2} \end{pmatrix}$$

- Then the distribution at time 1 is determined by

$$\begin{aligned} \theta_1 &= \begin{pmatrix} \theta_{1,1} \\ \theta_{1,2} \end{pmatrix} \\ &= \begin{pmatrix} (1-p)\,\theta_{0,1} + q\theta_{0,2} \\ p\theta_{0,1} + (1-q)\,\theta_{0,2} \end{pmatrix} \\ &= \begin{pmatrix} 1-p & q \\ p & 1-q \end{pmatrix} \begin{pmatrix} \theta_{0,1} \\ \theta_{0,2} \end{pmatrix} \\ &= P'\theta_0 \end{aligned}$$

- And the distribution at time T is given by $\theta_N = (P')^N \theta_0$.

Stationary Distributions

- If we let the Markov chain run for a long-time with a large population, will the distribution of outcomes settle down to specific distribution? If it does, we call this the **stationary distribution** (also sometimes called the invariant distribution).
- If it exists, the stationary distribution, θ^* satisfies

$$\theta^* = P'\theta^* \Rightarrow (I - P')\,\theta^* = 0$$

- This means that θ^* is an eigenvector of P' corresponding to an eigenvalue of one. Since transposing doesn't change eigenvalues or eigenvectors, it means θ^* is an eigenvector of P corresponding to an eigenvalue of one.
- The fact that the rows of P all sum to one guarantees that there is at least one eigenvalue. But there may be multiple unit eigenvalues and thus multiple possible stationary equilibria.
- If, however, all of the elements of the P matrix are between zero and one, that is a sufficient condition for the process to converge to a unique stationary equilibrium.

Back to the Two-State Process

- Consider again the two-state Markov chain with transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

- We can calculate the stationary distribution as the values θ_1^* and θ_2^* that solve

$$\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1-p & q \\ p & 1-q \end{pmatrix} \right] \begin{pmatrix} \theta_1^* \\ \theta_2^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- This can be solved to give

$$\begin{pmatrix} \theta_1^* \\ \theta_2^* \end{pmatrix} = \begin{pmatrix} \frac{q}{p+q} \\ \frac{p}{p+q} \end{pmatrix}$$

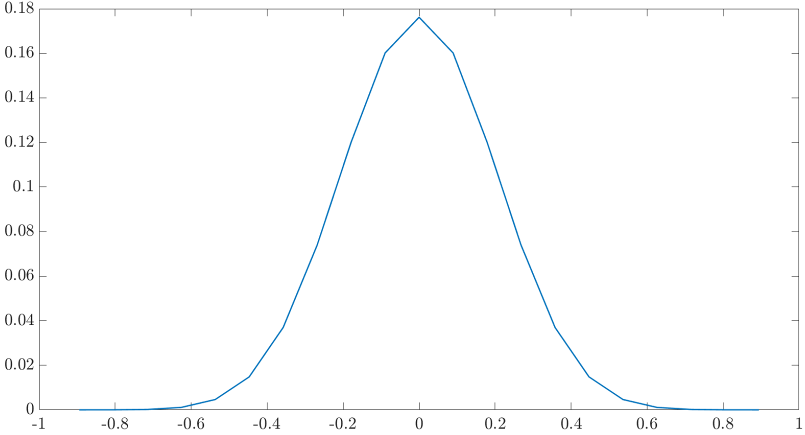
- Note that, for example, if $q = 0$, meaning that once you are in State 2 you stay there, then in the end nobody ends up in State 1. State 2 is what is known as an absorbing state.

Calculating Stationary Distributions with Matlab

- The code below uses Matlab's `eig` function to calculate the stationary distribution of a 3×3 matrix.
- Using `eig`, you can calculate the eigenvalues and also a matrix containing eigenvectors, ordered in the same order as the eigenvalues.
- You tell Matlab to find where in the ordering the unit eigenvalue is and then find the corresponding eigenvector. Once you've found it, you need to normalise its entries to sum to one.
- The graph on the next page shows the stationary distribution for the 21-state approximation to an AR(1) model disussed earlier. When using the Rouwenhurst method, I recommend using an odd number of states because this allows the mean value to be the middle point on the grid and gives a single-peaked stationary distribution.

```
% Plug in your own numbers for Markov chain probabilities
A = [0.5  0.5 0; 0 0.01 0.99; 0.1 0.1 0.8 ];
[V,D] = eig(A');                    % Find eigenvalues and eigenvectors of A'
[~,ix] = min(abs(diag(D)-1)); % Locate an eigenvalue which equals 1
v = V(:,ix)';                      % The corresponding row of V' will be a solution
v = v/sum(v);                      % Adjust it to have a sum of 1
disp(v);
```

Stationary Distribution of the 21-State AR(1) Approximation



Some Additional Points About Stationary Distributions

- A few slides back, we showed a simulation of the 21-state Markov chain and pointed out that Matlab picked a random state to start with.
- The fact that it starts out in this particular state, as opposed to one of the others, means this run of the model isn't necessarily representative.
- So if you want to get a "representative" dataset for this process it is best to "burn in" the process for a while before using it for simulation purposes, e.g. discard the first 100 periods.
- Also, while we have viewed the stationary distribution as a cross-sectional concept—this is where the cross-sectional distribution will settle down over time—it could also be viewed as the distribution of outcomes over time for a single agent who follows this Markov chain. If they lived for a very long time, this graph would be the histogram showing how much time they spent in each of the states.

Part II

Brief Introduction to Time Series

Some Types of Time Series Models

- Macroeconomic data isn't usually constrained so that it has to take up a finite set of values, so we usually model them as taking a continuous range of values.
- Because current data often depend on what has happened in the past, time series process such as the AR(1) process

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

- More generally, macroeconomists will often use AR(k) models of the form

$$y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \ldots + \rho_k y_{t-k} + \epsilon_t$$

  which depend on outcomes k periods ago.
- Next time we will look at AR models with multiple variables, where outcomes for all the variables depend on past values of all the other variables, a so-called Vector Autoregression model.
- Sometimes macroeconomists Autoregressive Moving Average processes. An ARMA(k, p) process features k lags of the dependent variable and p lags of a random shock.

$$y_t = \alpha + \epsilon_t + \sum_{i=1}^{k} \rho_i y_{t-i} + \sum_{i=1}^{p} \phi_i \epsilon_{t-i}$$

Stationarity and Random Walks

- It is easy to show that AR(1) processes such as

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

  where $-1 < \rho < 1$ will tend to move up or around a fixed average value of $\frac{\alpha}{1-\rho}$ and all samples of successive values of a fixed size will have the same expected variance. Processes of this type are termed **covariance stationary**.
- However, if $\rho = 1$, then these conditions no longer hold.
- Suppose $\alpha = 0$, then we have the classic **random walk** model.

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

  which implies the changes in y_t are completely unpredictable.
- This series will not settle down fluctuating around a specific mean value—indeed it could end up taking on any value. And variances for samples of successive values of fixed size will increase as the size of the sample increases.
- This is a **covariance non-stationary** process, more usually just termed a non-stationary process.

Generalisations

- In macroeconomics, it is more common to see a **random walk with drift**

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

  which tends to drift upwards over time but each period there will be random shocks that see the series change by either more than or less than α.

- This series is also nonstationary: It doesn't have a constant mean over time and uncertainty over future mean values grow as you look further into the future. Again, variances for samples of successive values of fixed size will increase as the size of the sample increases.

- We can generalise the point to AR(k) process

$$y_t = \alpha + \epsilon_t + \sum_{i=1}^{k} \rho_i y_{t-i}$$

  These are stationary if the sum of lag coefficients $\sum_{i=1}^{k} \rho_i$ is less than one in absolute value.

- The case in which $\sum_{i=1}^{k} \rho_i = 1$ provides a generalised form of the random walk with drift. Series of this sort are called **unit root** series because 1 is a root of the polynomial $1 - \rho_1 x - \rho_2 x^2 - .... - \rho_k x^k$.

Trends in Macroeconomic Series

- Due to technological progress, most macroeconomic variables—GDP, consumption etc—tend to grow over time.
- A simple example of a growing economy is one in which GDP experiences constant exponential growth, growing at rate β at all time.

$$Y_t = e^{\alpha + \beta t}$$

- Taking logs, we can write this as

$$y_t = \log (Y_t) = \alpha + \beta t$$

  (Note that macroeconomists tend to use lower-case letters when denoting the log of a series)

- Of course, no economy ever grows at the exact same rate each period. A more realistic model is

$$y_t = \log (Y_t) = \alpha + \beta t + u_t$$

  where u_t is a zero mean series. For example u_t could follow an AR(1) process:

$$u_t = \rho u_{t-1} + \epsilon_t$$

  where ϵ_t is a zero mean iid "white noise" series.

Deterministic and Stochastic Trends

- The model
$$y_t = \log(Y_t) = \alpha + \beta t + u_t$$
features a **deterministic trend**.
- While it may move above and below the trend value of $\alpha + \beta t$, we will always expect it to return to this trend value. It means we can predict future values with relatively high confidence.
- An alternative is the **stochastic trend** where the series is a unit root with drift, such as the random walk with drift below:
$$y_t = \alpha + y_{t-1} + u_t$$
where u_t is a zero mean series.
- On average, the series grows at rate α. However, when the series increases at a rate faster than α, this isn't offset by future negative realisations that "bring it back to trend". Future values of stochastic trend variables are harder to forecast because they of the absence of this reversion to trend feature.

Part III

Spurious Regression

A Problem When Using Non-Stationary Series

- Most of econometrics is concerned with assessing relationships between variables: Usually, we are asking the question "Does x have an effect on y?"
- But when two different unrelated nonstationary series are regressed on each other, the result is usually a so-called **spurious regression**, in which the OLS estimates and t statistics indicate that a relationship exists when, in reality, there is no such relationship.
- The modern literature on this topic dates from a famous paper by Granger and Newbold from 1974. However, the nature of the problem was known at least as far back as 1926.
- In 1926, Georges Udny Yule wrote a paper in the *Journal of the Royal Statistical Society* called "Why Do We Sometimes get Nonsense Correlations between Time-Series?"

Yule (1926) on Nonsense Correlations

SECTION I.—*The problem.*

It is fairly familiar knowledge that we sometimes obtain between quantities varying with the time (time-variables) quite high correlations to which we cannot attach any physical significance whatever, although under the ordinary test the correlation would be held to be certainly "significant." As the occurrence of such "nonsense-correlations" makes one mistrust the serious arguments that are sometimes put forward on the basis of correlations between time-series—my readers can supply their own examples—it is important to clear up the problem how they arise and in what special cases.

George Udny Yule's Chart from 1926

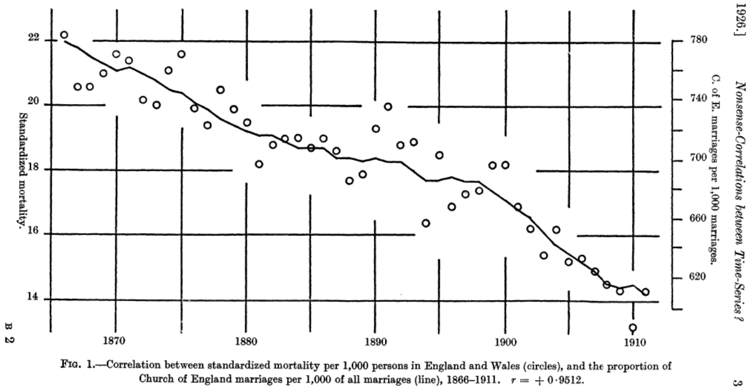


FIG. 1.—Correlation between standardized mortality per 1,000 persons in England and Wales (circles), and the proportion of Church of England marriages per 1,000 of all marriages (line), 1866–1911. $r = +0{\cdot}9512$.

Yule's Discussion of His Chart

Fig. 1 gives a very good illustration. The full line shows the proportion of Church of England marriages to all marriages for the years 1866–1911 inclusive: the small circles give the standardized mortality per 1,000 persons for the same years. Evidently there is a very high correlation between the two figures for the same year: the correlation coefficient actually works out at $+0{\cdot}9512$.

Now I suppose it is possible, given a little ingenuity and good-will, to rationalize very nearly anything. And I can imagine some enthusiast arguing that the fall in the proportion of Church of England marriages is simply due to the Spread of Scientific Thinking since 1866, and the fall in mortality is also clearly to be ascribed to the Progress of Science; hence both variables are largely or mainly influenced by a common factor and consequently ought to be highly correlated. But most people would, I think, agree with me that the correlation is simply sheer nonsense; that it has no meaning whatever; that it is absurd to suppose that the two variables in question are in any sort of way, however indirect, causally related to one another.

Spurious Regressions: Unit Roots with Drifts

- When discussing spurious regressions, econometric textbooks tend to focus on what happens when we take two random walks without drift (i.e. $y_t = y_{t-1} + \epsilon_t$ with no constant term) and regress them on each other.
- In applied econometric work, however, unit root without drift processes are not very common. Generally, we work with series that tend to be stationary or else with series that have a clear upward trend and which may be unit root processes with drift (e.g. stochastic trends of the form $y_t = \alpha + y_{t-1} + \epsilon_t$.)
- While explanations of how the spurious regression problem works for non-drifting unit root processes are quite complex, the spurious regression problem is far more relevant in the case where the processes have drift. It also turns out that the problem is easier to explain in this case.
- A property of drifting unit root processes that we will use is the following

$$\begin{aligned} y_t &= \alpha + y_{t-1} + \epsilon_t \\ &= \alpha + \alpha + y_{t-2} + \epsilon_t + \epsilon_{t-1} \\ &= \alpha t + \sum_{k=1}^{t} \epsilon_k + y_0 \end{aligned}$$

Useful Results About Infinite Sums

- Establishing properties about regressions involving drifting unit root series will require figuring out properties of sums of the form $\sum_{t=1}^{T} t$ and $\sum_{t=1}^{T} t^2$.
- Note that $1+2+3=6=\frac{(3)(4)}{2}$ and $1+2+3+4=10=\frac{(4)(5)}{2}$. The general rule is

$$\sum_{t=1}^{T} t = \frac{T(T+1)}{2} = \frac{1}{2}\left(T^2+T\right)$$

- For sums of squares, we have

$$\sum_{t=1}^{T} t^2 = \frac{T(T+1)(2T+1)}{6} = \frac{1}{6}\left(2T^3+3T^2+T\right)$$

- This means that as $T \to \infty$

$$\frac{1}{T^2}\sum_{t=1}^{T} t \rightarrow \frac{1}{2}$$

$$\frac{1}{T^3}\sum_{t=1}^{T} t^2 \rightarrow \frac{1}{3}$$

Regressions Featuring Random Walks with Drifts

- Consider regressing y_t on the completely unrelated series x_t where

$$\begin{aligned} y_t &= \alpha_y + y_{t-1} + \epsilon_t^y \\ x_t &= \alpha_x + x_{t-1} + \epsilon_t^x \end{aligned}$$

- The OLS estimator is

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2} \\ &= \frac{\sum_{t=1}^T \left(\alpha_x t + \sum_{k=1}^t \epsilon_k^x + x_0\right)\left(\alpha_y t + \sum_{k=1}^t \epsilon_k^y + y_0\right)}{\sum_{t=1}^T \left(\alpha_x t + \sum_{k=1}^t \epsilon_k^x + x_0\right)^2} \end{aligned}$$

- As T gets large, the terms in t^2 will dominate all other terms. Re-writing this as

$$\hat{\beta} = \frac{\frac{1}{T^3}\sum_{t=1}^T \left(\alpha_x t + \sum_{k=1}^t \epsilon_k^x + x_0\right)\left(\alpha_y t + \sum_{k=1}^t \epsilon_k^y + y_0\right)}{\frac{1}{T^3}\sum_{t=1}^T \left(\alpha_x t + \sum_{k=1}^t \epsilon_k^x + x_0\right)^2}$$

then all of the terms that are not of the form $\frac{1}{T^3}\sum_{t=1}^T t^2$ will go to zero.

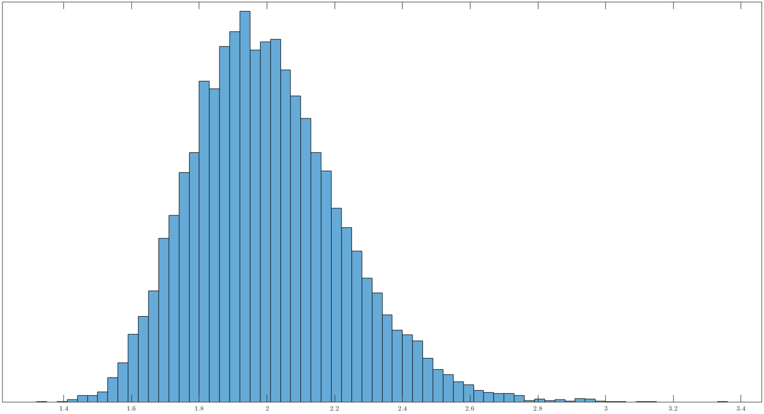
Spurious Regression Results

- This means that as T gets large, $\hat{\beta}$ tends to converge towards $\frac{\alpha_y}{\alpha_x}$
- In other words, the OLS estimator will tend towards the ratio of the two drift terms. In addition, the t statistics will generally indicate that there is a highly statistically significant relationship.
- The next pages show histograms for $\hat{\beta}$'s and t-stats from a Matlab programme with 10,000 Monte Carlo simulations regressing y_t on x_t where

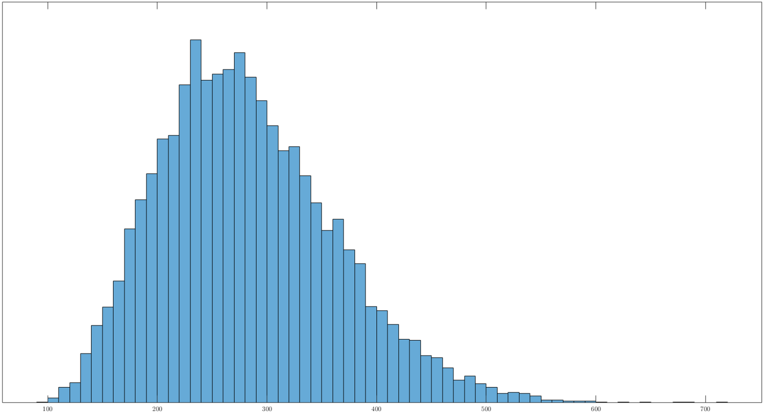
$$\begin{aligned} y_t &= 1 + y_{t-1} + \epsilon_t^y \\ x_t &= 0.5 + x_{t-1} + \epsilon_t^x \end{aligned}$$

  where the error terms are i.i.d. normally distributed errors with standard deviations equal to 1. They show OLS coefficients averaging 2 and highly significant t-stats.
- Note that the key terms driving these results were the time trends. These results also apply to "trend stationary" series like $y_t = \alpha t + \rho y_{t-1} + \epsilon_t$, so the problem is not specific to the unit root series.
- Similar results apply to regressions featuring unit roots without drifts but deriving these results analytically is beyond the scope of this module.

Histogram of β from Regressions of Unrelated Unit Roots With Drift ($T = 500$)



Histogram of t Statistics from Regressions of Unrelated Unit Roots With Drift ($T = 500$)



Matlab Programme Simulating Spurious Regressions

Spurious.m

```
clear all; clc; close all;

T      = 500;
nsims  = 10000;
mu     = 1;
alpha  = 0.5;
betas  = NaN*ones(nsims,1);
tstats = NaN*ones(nsims,1);

for i = 1:nsims

y = NaN*ones(T,1);
x = NaN*ones(T,1);

y(1) = 1;
x(1) = 1;

for t=1:T
    y(t+1) = y(t) + mu + randn(1) ;
    x(t+1) = x(t) + alpha + randn(1) ;
end

X = [ones(length(x),1) x];
mdl = fitlm(x,y);

betas(i) = mdl.Coefficients.Estimate(2);
tstats(i) = mdl.Coefficients.tStat(2);

end

figure(1)
histogram(betas);
set(gca,'ytick',[])
fig1 = figure(1);

figure(2)
histogram(tstats);
set(gca,'ytick',[])
fig2 = figure(2);
```

Part IV

Cointegration

First-Differencing as a Solution?

- So what do we do with nonstationary series when we want to understand causal relationships?
- Well suppose you have a series like $y_t = \alpha + y_{t-1} + \epsilon_t$? One option comes from the fact that if you calculate the first difference of this series $\Delta y_t = \alpha + \epsilon_t$, it becomes a stationary series.
- We say a series is **integrated of order** k (denoted $I(k)$) if it has to be differenced k times before it becomes stationary. Sometimes, one can can come across examples involving $I(2)$ series, but generally the time series in practical applications are either $I(1)$ or $I(0)$.
- So one option if we want to check if there is a relationship between two I(1) series is to use the first-differences instead. For example, if we are testing whether there is a relationship of the form

  $$y_t = \gamma + \beta x_t + \epsilon_t$$

  we could first-difference both sides and instead run the following regression

  $$\Delta y_t = \beta \Delta x_t + u_t$$

  where $u_t = \epsilon_t - \epsilon_{t-1}$.

Cointegration

- The spurious regression problem can be stated as the fact that unrelated $I(1)$ series regressed upon each other tend to appear to be related according to the usual OLS diagnostics.
- Examining whether there is a relationship in first-differences is one way to check if the relationship is spurious or real.
- However, what if there really is a relationship between the levels? For example, what if y_t and x_t are both $I(1)$ series but there existed a coefficient β such that $y_t - \beta x_t \sim I(0)$.
- In this case, there is a common trend across the series and we say that the series y_t and x_t are **cointegrated**.
- When this is the true relationship, it turns out that OLS estimates of β are consistent and using the information about the levels of the variables is very helpful for getting a good estimate of the relationship.

Consistency of OLS Under Cointegration

- Consider again the case where x_t is a unit root with drift

$$x_t = \alpha_x + x_{t-1} + \epsilon_t^x$$

  but in this case the variable y_t is cointegrated with x_t so that

$$y_t = \beta x_t + u_t$$

  where u_t is mean-zero $I(0)$ series.

- We can calculate the properties of the OLS estimator as follows:

$$\begin{aligned} \hat{\beta} &= \beta + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{t} x_t^2} \\ &= \beta + \frac{\sum_{t=1}^{T} \left(\alpha_x t + \sum_{k=1}^{t} \epsilon_k^x + x_0\right) u_t}{\sum_{t=1}^{T} \left(\alpha_x t + \sum_{k=1}^{t} \epsilon_k^x + x_0\right)^2} \end{aligned}$$

- The terms in T^2 will dominate as $T \to \infty$ so that the denominator of the last term will grow faster than the numerator. This means that $\hat{\beta}$ converges in probability to β. (In fact, it does so at a faster pace than if your regression involves I(0) variables)

The Error-Correction Representation

- Consider two $I(1)$ series, y_t and x_t. We would expect their first-differences to have stationary representations

$$\begin{aligned}\Delta y_t &= \alpha^y + \gamma_1^y \Delta y_{t-1} + \ldots + \gamma_k^y \Delta y_{t-k} + \epsilon_t^y \\ \Delta x_t &= \alpha^x + \gamma_1^x \Delta x_{t-1} + \ldots + \gamma_k^x \Delta x_{t-k} + \epsilon_t^x\end{aligned}$$

- Now suppose that y_t and x_t are cointegrated. This means there exists a value β such that $y_t - \beta x_t \sim I(0)$. But if the processes are as described above, then there is nothing about the behaviour of either series that would see the two series tending to move together. So, additional terms are required to describe these processes.

- Specifically, we need additional **error-correction** terms of the form $y_t - \beta x_t$, to get a representation of the form

$$\begin{aligned}\Delta y_t &= \alpha^y + \gamma_1^y \Delta y_{t-1} + \ldots + \gamma_k^y \Delta y_{t-k} + \theta_y \left(y_t - \beta x_t\right) + \epsilon_t^y \qquad (1) \\ \Delta x_t &= \alpha^x + \gamma_1^x \Delta x_{t-1} + \ldots + \gamma_k^x \Delta x_{t-k} + \theta_x \left(y_t - \beta x_t\right) + \epsilon_t^x \qquad (2)\end{aligned}$$

  where we expect to have $\theta_y \leq 0$ and $\theta_x \geq 0$. In other words, when y_t rises above its long-run relationship with x_t it tends to fall back and/or x_t tends to increase.

The Vector Error-Correction Representation

- When there are only two series, any potential cointegrating vector is unique up to multiplication by a scalar (e.g. we could say $y_t - \beta x_t \sim I(0)$ or that $x_t - \beta^{-1} y_t \sim I(0)$).
- However, when there are n different variables, then there may be multiple cointegrating vectors, e.g. for $Y_t = (y_{1t}, y_{2t}, y_{3t}, y_{4t})$, one could have $y_{1t} - \gamma_1 y_{3t} \sim I(0)$ and $y_{2t} - \gamma_1 y_{4t} \sim I(0)$.
- Consider the general case, in which there are r cointegrating relationships among n variables. Specifically, consider the case in which the $n \times 1$ vector of $I(1)$ series Y_t has the property that there exists an $r \times n$ matrix A such that the r series defined by $Z_t = AY_t$ are all $I(0)$. In this case, there exists an $n \times r$ matrix B such that Y_t is described by a Vector Error Correction Mechanism representation

$$\begin{aligned} \Delta Y_t &= \gamma_1^x \Delta Y_{t-1} + ... + \gamma_k^x \Delta Y_{t-k} + \alpha + BZ_{t-1} + \epsilon_t && (3) \\ &= \gamma_1^x \Delta Y_{t-1} + ... + \gamma_k^x \Delta Y_{t-k} + \alpha + BAY_{t-1} + \epsilon_t && (4) \end{aligned}$$

- This result is part of what is known as the Granger Representation Theorem.

Testing for Cointegration

- Suppose we have two $I(1)$ series, y_t and x_t. How do we test whether they are cointegrated or whether the relationship between them is spurious? Tests are based on the idea that if there is no underlying relationship than the OLS residuals, $\hat{u}_t = y_t - \hat{\beta} x_t$ will also have a unit root.
- One might be tempted to regress $\hat{u}_t$ on $\hat{u}_{t-1}$and do a t-test of whether the coefficient equals one. However, the OLS procedure produces residuals that may appear stationary, even when no cointegrating relationship exists.
- This means that special critical values must be applied when testing for cointegration. These critical values differ depending on whether the underlying y_t and x_t series have drifts and on whether the potential cointegrating regression includes a constant.
- Critical values are generally calculated via Monte Carlo simulations that regress two unrelated $I(1)$ series on its other and collect statistics on the distribution of t-statistics. Test procedures reject the null of no cointegration if the t-statistics is above specified percentiles of this distribution.
- When testing for r different cointegrating vectors among n variables, testing procedures involve estimating a VAR process and assess whether the relevant VECM is the best fit for the data.

Part V

Detrending

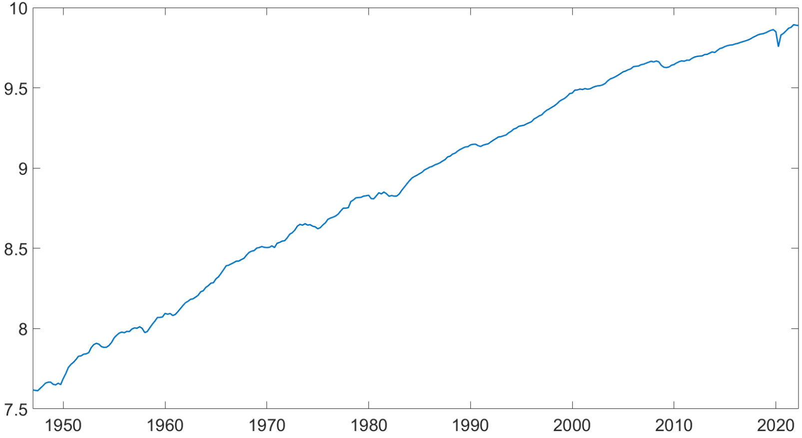
Trends and Cycles

- Macroeconomists sometimes decide to break series into a "non-stationary" long-run trend and a "stationary" cyclical component.
- "Business cycle analysis" relates to this modelling and explaining the cyclical components of the major macroeconomic variables.
- Fine in theory, but how is this done in practice?
- Simplest method: Log-linear trend
  - Estimated from regression

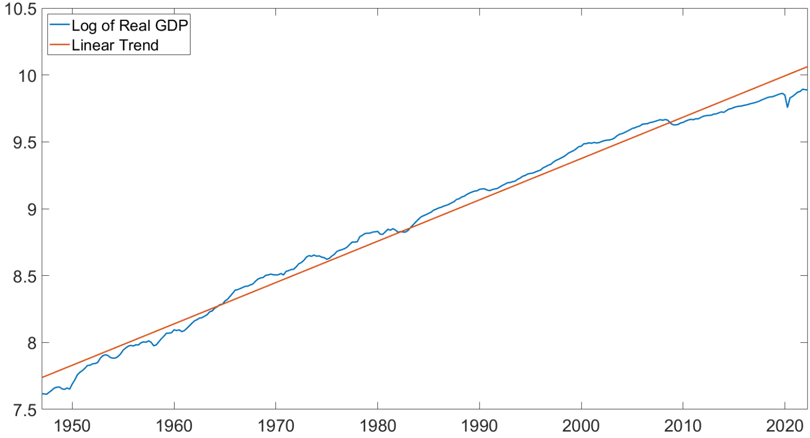
$$\log(Y_t) = y_t = \alpha + gt + \epsilon_t$$

  - Trend component $\alpha + gt$.
  - Zero-mean stationary cyclical component ϵ_t.
  - Log-difference Δy_t (equivalent to growth rate) has two components: Constant trend growth g and the change in cyclical component $\Delta\epsilon_t$.
- The next few pages show output from a Matlab programme plotting the log of real GDP and the trend and cycle based on fitting a linear trend to this series.

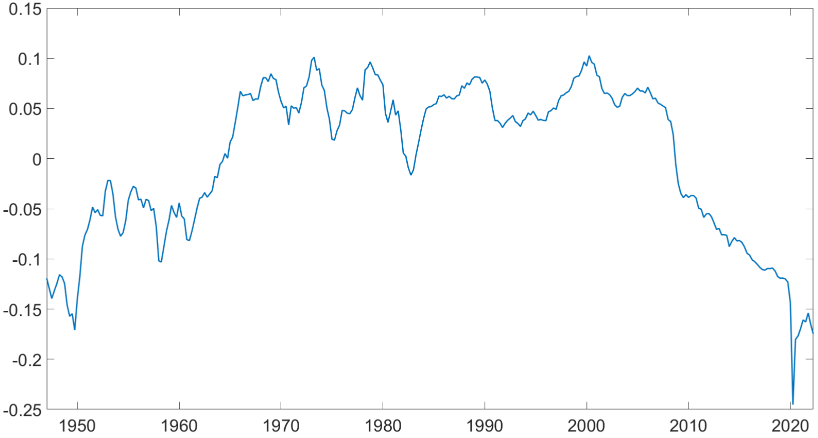
Log of US Real GDP, 1947:Q1 to 2022:Q2



Log of US Real GDP with a Linear Trend, 1947:Q1 to 2022:Q2



Detrended Log Real US GDP From a Linear Trend Model, 1947:Q1 to 2022:Q2



Variations in Trend Growth

- Two problems with a simple deterministic log-linear trend:
  - It assumes there is a constant trend growth rate for all times. But there are no fixed constant in economics and underlying growth rates for the economy may improve or decline depending on various factors.
  - It rules out stochastic trend behaviour, where the economy grows at an average rate but positive shocks are not subsequently reversed.
- A more realistic model should be one in which we accept that growth rate of the trend probably varies a bit over time leaving a cycle that moves up and down over time around this various trend.
- Macroeconomists have developed many different kinds of what are known as **filtering** methods, which take a time series and split it into its variable growth rate trend component and its cyclical component.
- This is a challenging task and ultimately relies on judgement.
  - You don't want the time series to depart by huge amounts from its so-called trend (as happened above with the log-linear trend for real GDP).
  - But if the trend series tracks too close to the actual time series, it loses its meaning as a trend.

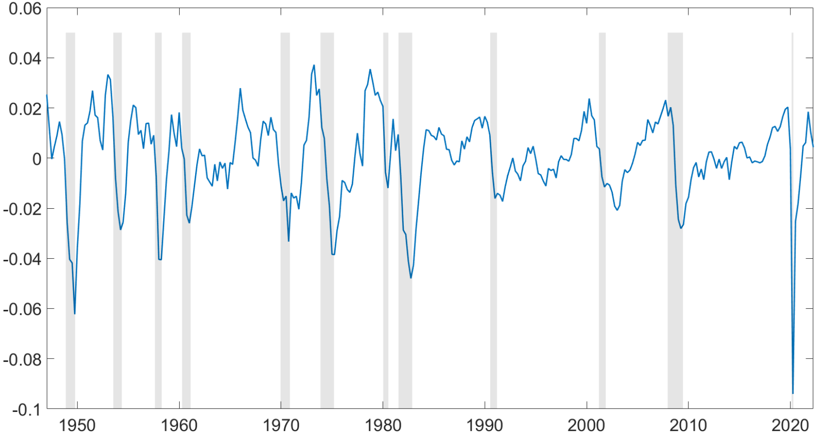
The Hodrick-Prescott Filter

- The most commonly used filtering method in macroeconomics was developed by Hodrick and Prescott (1981). It is available as a command in Matlab.
- They suggested choosing the time-varying trend Y_t^* so as to minimize

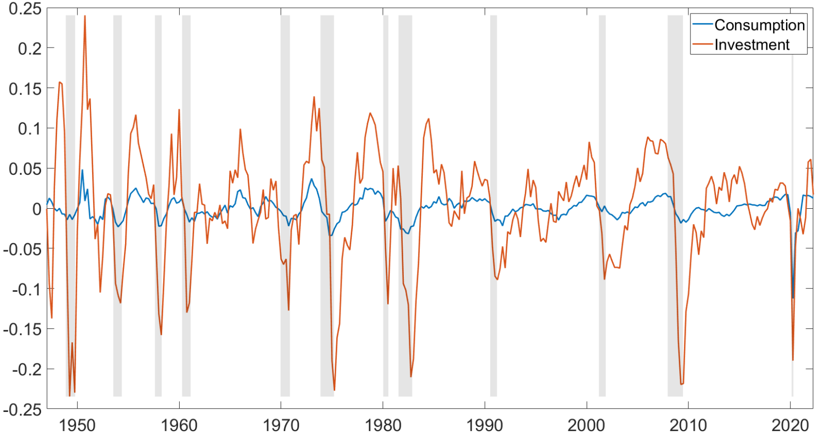
$$\sum_{t=1}^{N}\left[\left(Y_t - Y_t^*\right)^2 + \lambda\left(\Delta Y_t^* - \Delta Y_{t-1}^*\right)\right]$$

- This method tries to minimize the sum of squared deviations between output and its trend $(Y_t - Y_t^*)^2$ but also contains a term that emphasises minimizing the change in the trend growth rate $(\lambda\left(\Delta Y_t^* - \Delta Y_{t-1}^*\right))$.
- How do we choose λ and thus weight the goodness-of-fit of the trend versus smoothness of the trend?
- $\lambda = 1600$ is the standard value used in business cycle detrending with quarterly data. We will discuss this choice in more detail in a few weeks.
- Some DSGE modellers apply a HP filter to their data and then analyse only the cyclical components.

HP-Filtered Cycles Correspond Well to NBER Recessions



Investment Cycles Are Bigger than Consumption Cycles



Matlab Programme For the Previous Charts

```
clc; clear all;

NIPA = readtable('Fred-NIPA.xlsx') ;

lpce     = log(NIPA.PCECC96);
linvest  = log(NIPA.GPDIC1);
lgdp     = log(NIPA.GDPC1);
date    = NIPA.sasdate;

[~,gdpfilter] = hpfilter(lgdp,1600);
[~,pcefilter] = hpfilter(lpce,1600);
[~,investfilter] = hpfilter(linvest,1600);

figure(1)
plot(date,gdpfilter,'LineWidth',2);
set(gca,'FontSize',25);
recessionplot;

figure(2)
plot(date,pcefilter,date,investfilter,'LineWidth',2);
set(gca,'FontSize',25);
recessionplot;
legend('Consumption','Investment','Location','NorthEast');
```

Hamilton on the Hodrick-Prescott Filter

Despite its popularity, not everyone thinks the Hodrick-Prescott filter. In his 2018 *REStat* paper, "Why You Should Never Use the Hodrick-Prescott Filter" James Hamilton makes a number of specific criticisms.

- If series are generated by unit roots with trends, then the filtered cycles are spurious and spurious cyclical relationships can emerge between the filtered series.
- The HP trend and cycle have an artificial ability to "predict" the future because they are by construction a function of future as well past realizations.
- You could try to fix this by using only using a one-sided filter, so the filtered variable at time t depends only on observation seen up to that point. But end-point values for these filters tend to be very sensitive to the final observations.
- The HP filter unnecessarily imposes a value for λ, usually $\lambda = 1600$ but it is not necessary to do this. You can use "latent variable" econometric methods to estimate λ and the values it returns are far lower than $\lambda = 1600$.