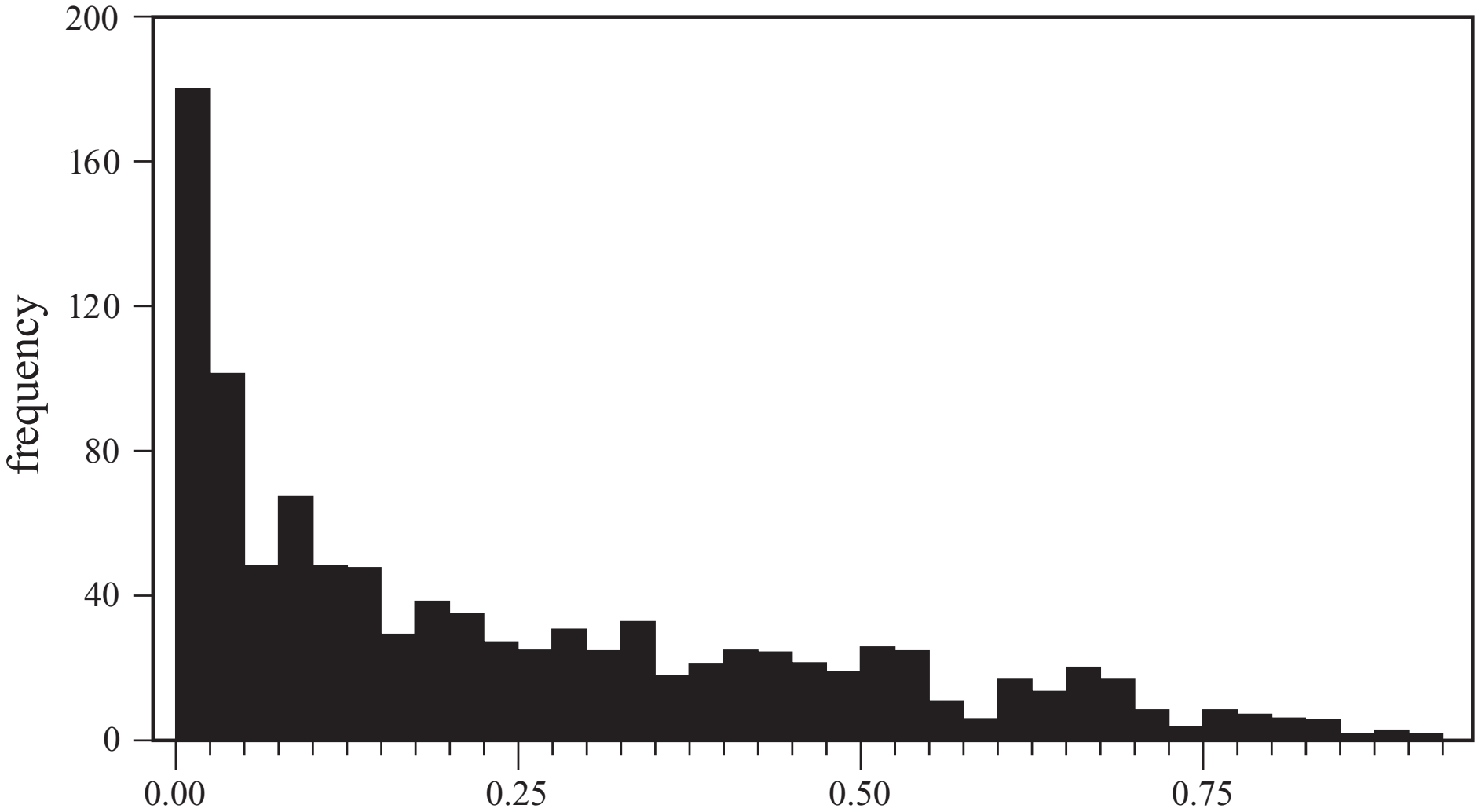
Chapter 8

Modelling long-run relationship in finance

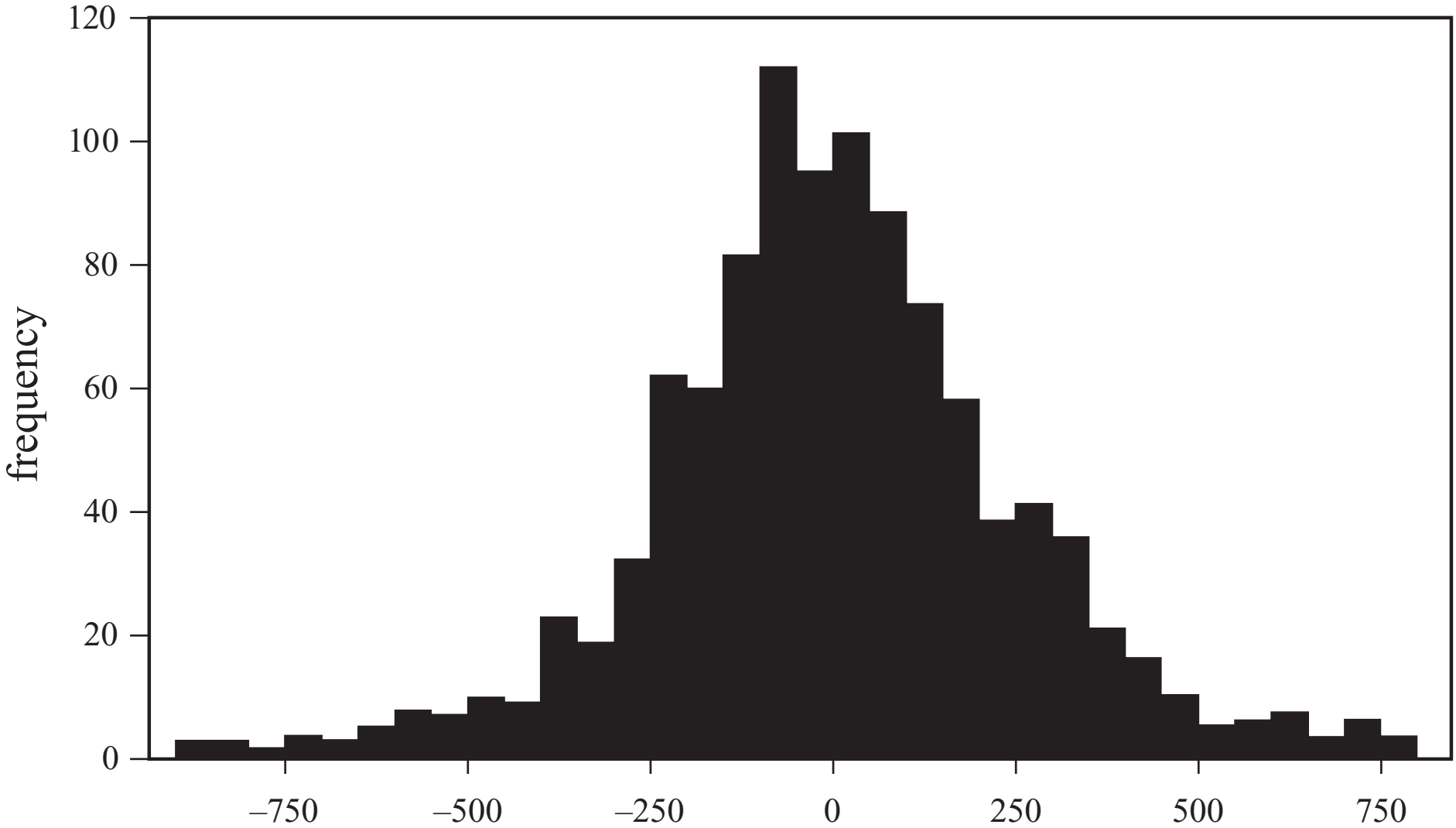
Stationarity and Unit Root Testing Why do we need to test for Non-Stationarity?

- The stationarity or otherwise of a series can strongly influence its behaviour and properties - e.g. persistence of shocks will be infinite for nonstationary series
- Spurious regressions. If two variables are trending over time, a regression of one on the other could have a high R^2 even if the two are totally unrelated
- If the variables in the regression model are not stationary, then it can be proved that the standard assumptions for asymptotic analysis will not be valid. In other words, the usual "t-ratios" will not follow a t-distribution, so we cannot validly undertake hypothesis tests about the regression parameters.

Value of R^2 for 1000 Sets of Regressions of a Non-stationary Variable on another Independent Non-stationary Variable



Value of t-ratio on Slope Coefficient for 1000 Sets of Regressions of a Non-stationary Variable on another Independent Non-stationary Variable



Two types of Non-Stationarity

- Various definitions of non-stationarity exist
- In this chapter, we are really referring to the weak form or covariance stationarity
- There are two models which have been frequently used to characterise non-stationarity: the random walk model with drift:

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

and the deterministic trend process:

$$y_t = \alpha + \beta t + u_t \tag{2}$$

where u_t is iid in both cases.

Stochastic Non-Stationarity

- Note that the model (1) could be generalised to the case where y_t is an explosive process:

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

  where $\phi > 1$.

- Typically, the explosive case is ignored and we use $\phi = 1$ to characterise the non-stationarity because
  - $\phi > 1$ does not describe many data series in economics and finance.
  - $\phi > 1$ has an intuitively unappealing property: shocks to the system are not only persistent through time, they are propagated so that a given shock will have an increasingly large influence.

Stochastic Non-stationarity: The Impact of Shocks

- To see this, consider the general case of an AR(1) with no drift:

$$y_t = \phi y_{t-1} + u_t \tag{3}$$

  Let ϕ take any value for now.

- We can write:

$$\begin{aligned} y_{t-1} &= \phi y_{t-2} + u_{t-1} \\ y_{t-2} &= \phi y_{t-3} + u_{t-2} \end{aligned}$$

- Substituting into (3) yields

$$\begin{aligned} y_t &= \phi(\phi y_{t-2} + u_{t-1}) + u_t \\ &= \phi^2 y_{t-2} + \phi u_{t-1} + u_t \end{aligned}$$

Stochastic Non-stationarity: The Impact of Shocks (Cont'd)

- Substituting again for y_{t-2}

$$\begin{aligned} y_t &= \phi^2(\phi y_{t-3} + u_{t-2}) + \phi u_{t-1} + u_t \\ &= \phi^3 y_{t-3} + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \end{aligned}$$

- Successive substitutions of this type lead to:

$$\begin{aligned} y_t &= \phi^T y_0 + \phi u_{t-1} + \phi^2 u_{t-2} + \phi^3 u_{t-3} + \cdots \\ &\quad + \phi^T u_0 + u_t \end{aligned}$$

The Impact of Shocks for Stationary and Non-stationary Series

- We have 3 cases:

(1) $\phi < 1 \Rightarrow \phi^T \rightarrow 0$ as $T \rightarrow \infty$
So the shocks to the system gradually die away.

(2) $\phi = 1 \Rightarrow \phi^T = 1 \ \forall \ T$
So shocks persist in the system and never die away. We obtain

$$y_t = y_0 + \sum_{t=0}^{\infty} u_t \text{ as } T \rightarrow \infty$$

So the current value of y is just an infinite sum of past shocks plus some starting value of y_0.

(3) $\phi > 1$. Now given shocks become more influential as time goes on, since if $\phi > 1$, $\phi^3 > \phi^2 > \phi$, etc.

Detrending a Stochastically Non-stationary Series

- Going back to our 2 characterisations of non-stationarity, the r.w. with drift:

$$y_t = \mu + y_{t-1} + u_t \tag{4}$$

and the trend-stationary process

$$y_t = \alpha + \beta t + u_t \tag{5}$$

- The two will require different treatments to induce stationarity. The second case is known as deterministic non-stationarity and what is required is detrending.

Detrending a Stochastically Non-stationary Series (Cont'd)

- The first case is known as stochastic non-stationarity, where there is a stochastic trend in the data. Letting $\Delta y_t = y_t - y_{t-1}$ and $Ly_t = y_{t-1}$ so that $(1 - L)\, y_t = y_t - Ly_t = y_t - y_{t-1}$. If (4) is taken and y_{t-1} subtracted from both sides

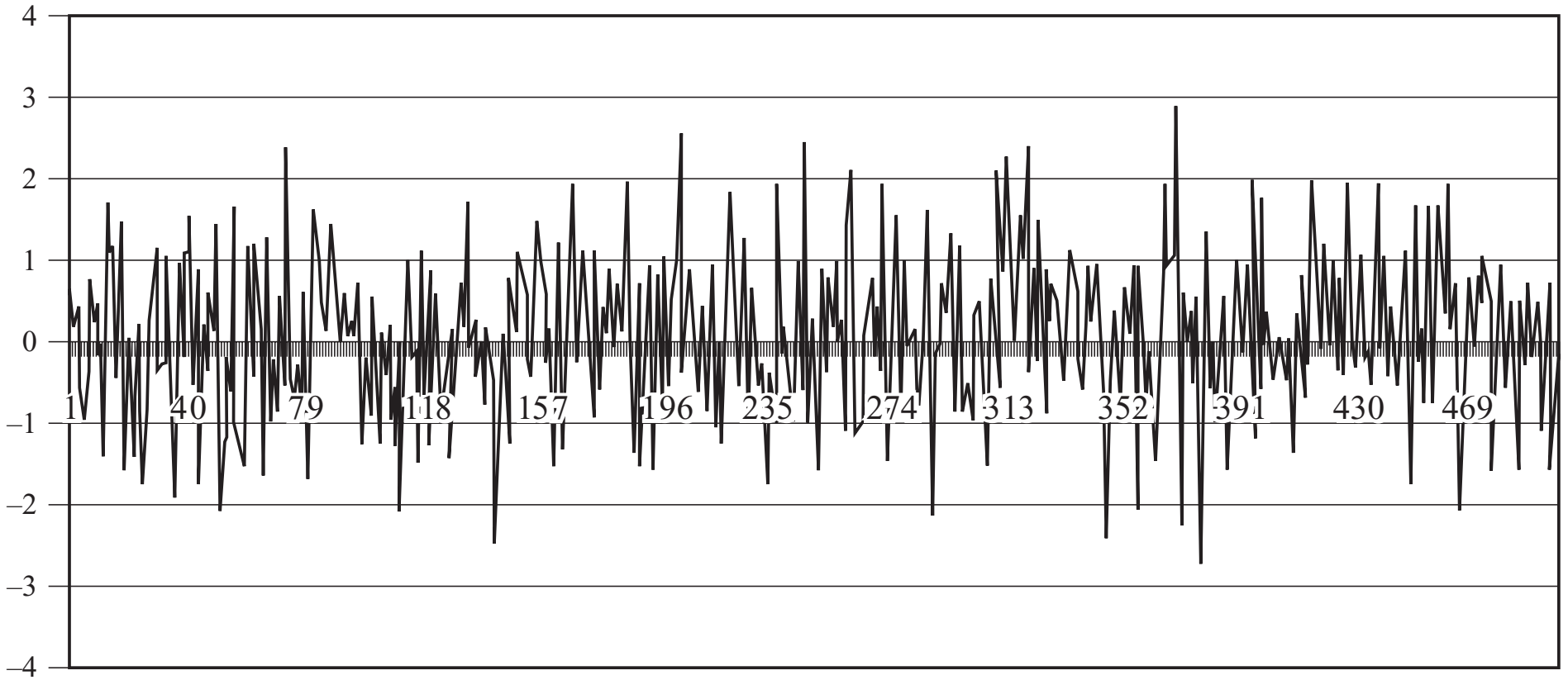
$$y_t - y_{t-1} = \mu + u_t$$
$$\Delta\, y_t = \mu + u_t$$

We say that we have induced stationarity "by differencing one".

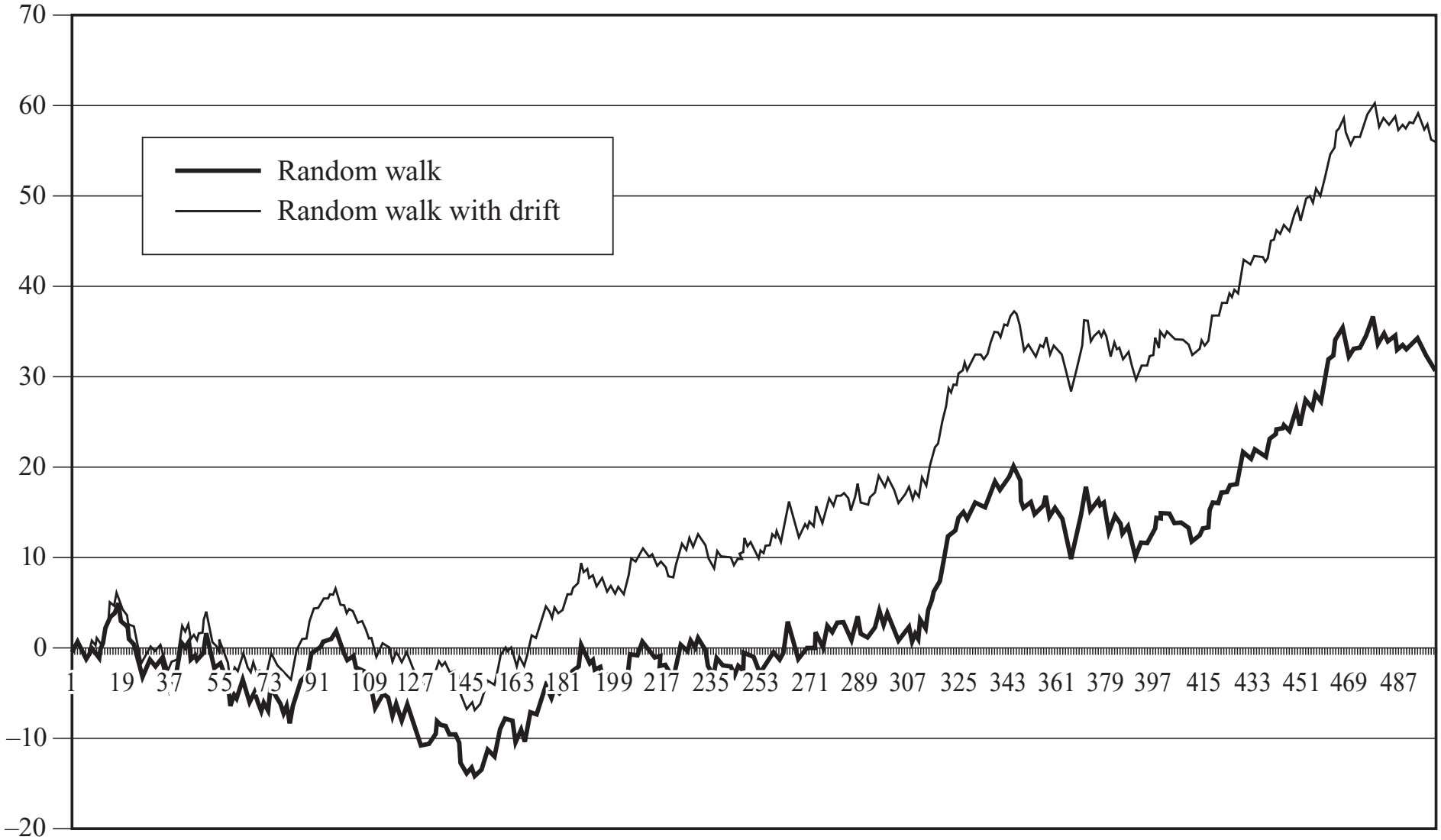
Detrending a Series: Using the Right Method

- Although trend-stationary and difference-stationary series are both "trending" over time, the correct approach needs to be used in each case.
- If we first difference the trend-stationary series, it would "remove" the non-stationarity, but at the expense on introducing an MA(1) structure into the errors.
- Conversely if we try to detrend a series which has stochastic trend, then we will not remove the non-stationarity.
- We will now concentrate on the stochastic non-stationarity model since deterministic non-stationarity does not adequately describe most series in economics or finance.

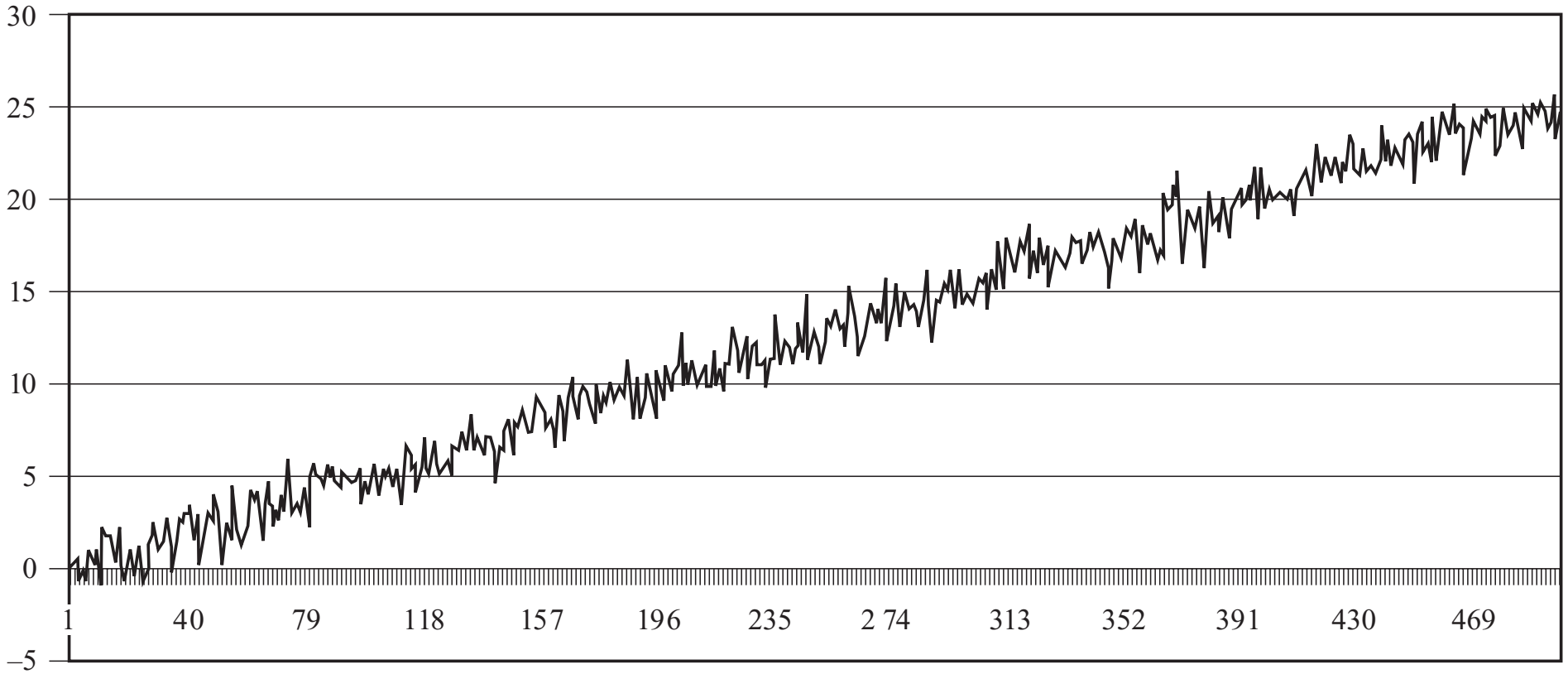
Sample Plots for various Stochastic Processes: A White Noise Process



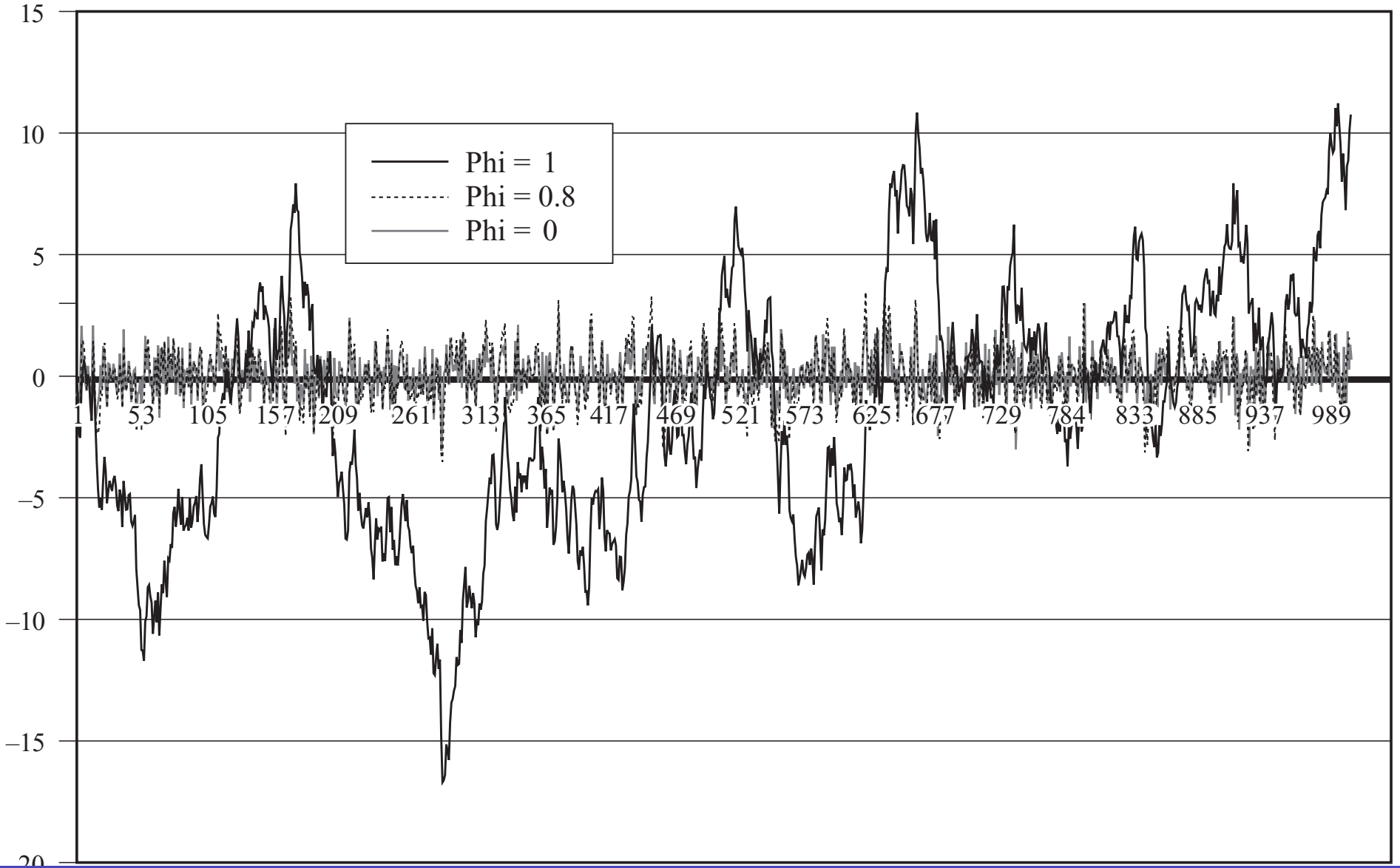
Sample Plots for various Stochastic Processes: A Random Walk and a Random Walk with Drift



Sample Plots for various Stochastic Processes: A Deterministic Trend Process



Autoregressive Processes with differing values of ϕ (0, 0.8, 1)



Definition of Non-Stationarity

- Consider again the simplest stochastic trend model:

$$y_t = y_{t-1} + u_t$$

or

$$\Delta y_t = u_t$$

- We can generalise this concept to consider the case where the series contains more than one "unit root". That is, we would need to apply the first difference operator, Δ, more than once to induce stationarity.

Definition

If a non-stationary series, y_t must be differenced d times before it becomes stationary, then it is said to be integrated of order d. We write $y_t \sim \mathrm{I}(d)$. So if $y_t \sim \mathrm{I}(d)$ then $\Delta^d y_t \sim \mathrm{I}(0)$.

An $\mathrm{I}(0)$ series is stationary

Definition of Non-Stationarity (Cont'd)

An I(1) series contains one unit root

$$y_t = y_{t-1} + u_t$$

Characteristics of I(0), I(1) and I(2) Series

- An I(2) series contains two unit roots and so would require differencing twice to induce stationarity.
- I(1) and I(2) series can wander a long way from their mean value and cross this mean value rarely.
- I(0) series should cross the mean frequently.
- The majority of economic and financial series contain a single unit root, although some are stationary and consumer prices have been argued to have 2 unit roots.

How do we test for a unit root?

- The early and pioneering work on testing for a unit root in time series was done by Dickey and Fuller (Dickey and Fuller 1979, Fuller 1976). The basic objective of the test is to test the null hypothesis that ϕ =1 in:

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

against the one-sided alternative $\phi < 1$. So we have

H_0: series contains a unit root
versus H_1: series is stationary.

- We usually use the regression:

$$\Delta y_t = \psi y_{t-1} + u_t$$

so that a test of $\phi = 1$ is equivalent to a test of $\psi = 0$ (since $\phi - 1 = \psi$).

Different forms for the DF Test Regressions

- Dickey Fuller tests are also known as τ tests: τ, τ_μ, τ_τ.
- The null (H_0) and alternative (H_1) models in each case are
  i. H_0: $y_t = y_{t-1} + u_t$
  H_1: $y_t = \phi y_{t-1} + u_t, \quad \phi < 1$

  This is a test for a random walk against a stationary autoregressive process of order one (AR(1))

  ii. H_0: $y_t = y_{t-1} + u_t$
  H_1: $y_t = \phi y_{t-1} + \mu + u_t, \quad \phi < 1$

  This is a test for a random walk against a stationary AR(1) with drift.

  iii. H_0: $y_t = y_{t-1} + u_t$
  H_1: $y_t = \phi y_{t-1} + \mu + \lambda t + u_t, \quad \phi < 1$

  This is a test for a random walk against a stationary AR(1) with drift and a time trend.

Computing the DF Test Statistic

- We can write

$$\Delta y_t = u_t$$

where $\Delta y_t = y_t - y_{t-1}$, and the alternatives may be expressed as

$$\Delta y_t = \psi y_{t-1} + \mu + \lambda t + u_t$$

with $\mu = \lambda = 0$ in case i), and $\lambda = 0$ in case ii) and $\psi = \phi - 1$. In each case, the tests are based on the t-ratio on the y_{t-1} term in the estimated regression of Δy_t on y_{t-1}, plus a constant in case ii) and a constant and trend in case iii). The test statistics are defined as

$$test\ statistic = \frac{\hat{\psi}}{\hat{SE}(\hat{\psi})}$$

Computing the DF Test Statistic (Cont'd)

- The test statistic does not follow the usual t-distribution under the null, since the null is one of non-stationarity, but rather follows a non-standard distribution. Critical values are derived from Monte Carlo experiments in, for example, Fuller (1976). Relevant examples of the distribution are shown in table 4.1 below

Critical Values for the DF Test

Significance level	10%	5%	1%
CV for constant but no trend	−2.57	−2.86	−3.43
CV for constant and trend	−3.12	−3.41	−3.96

The null hypothesis of a unit root is rejected in favour of the stationary alternative in each case if the test statistic is more negative than the critical value.

The Augmented Dickey Fuller (ADF) Test

- The tests above are only valid if u_t is white noise. In particular, u_t will be autocorrelated if there was autocorrelation in the dependent variable of the regression (Δy_t) which we have not modelled. The solution is to "augment" the test using p lags of the dependent variable. The alternative model in case (i) is now written:

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + u_t$$

- The same critical values from the DF tables are used as before. A problem now arises in determining the optimal number of lags of the dependent variable

- There are 2 ways
  - use the frequency of the data to decide
  - use information criteria

Testing for Higher Orders of Integration

- Consider the simple regression:

$$\Delta y_t = \psi y_{t-1} + u_t$$

We test that H_0: $\psi = 0$ vs. H_1: $\psi < 0$.

- If H_0 is rejected, we simply conclude that y_t does not contain a unit root.

- But what do we conclude if H_0 is not rejected? The series contains a unit root, but is that it? No! What if $y_t \sim I(2)$? We would still not have rejected. So we now need to test

$$H_0 : y_t \sim I(2) \text{ vs. } H_1 : y_t \sim I(1)$$

We would continue to test for a further unit root until we rejected H_0.

Testing for Higher Orders of Integration (Cont'd)

- We now regress $\Delta^2 y_t$ on Δy_{t-1} (plus lags of $\Delta^2 y_t$ if necessary).
- Now we test H_0: $\Delta y_t \sim I(1)$ which is equivalent to H_0: $y_t \sim I(2)$.
- So in this case, if we do not reject (unlikely), we conclude that y_t is at least I(2).

The Phillips-Perron Test

- Phillips and Perron have developed a more comprehensive theory of unit root nonstationarity. The tests are similar to ADF tests, but they incorporate an automatic correction to the DF procedure to allow for autocorrelated residuals.
- The tests usually give the same conclusions as the ADF tests, and the calculation of the test statistics is complex.

Criticism of Dickey-Fuller and Phillips-Perron-type tests

- Main criticism is that the power of the tests is low if the process is stationary but with a root close to the non-stationary boundary.

  e.g. the tests are poor at deciding if ϕ=1 or ϕ=0.95,

  especially with small sample sizes.

- If the true data generating process (dgp) is

$$y_t = 0.95y_{t-1} + u_t$$

  then the null hypothesis of a unit root should be rejected.

- One way to get around this is to use a stationarity test as well as the unit root tests we have looked at.

Stationarity tests

- Stationarity tests have

  H_0: y_t is stationary
  versus H_1: y_t is non-stationary

  So that by default under the null the data will appear stationary.

- One such stationarity test is the KPSS test (Kwaitowski, Phillips, Schmidt and Shin, 1992).

- Thus we can compare the results of these tests with the ADF/PP procedure to see if we obtain the same conclusion.

- A Comparison

	ADF / PP	KPSS
	H_0: $y_t \sim I(1)$	H_0: $y_t \sim I(0)$
	H_1: $y_t \sim I(0)$	H_0: $y_t \sim I(1)$

Stationarity tests (Cont'd)

- 4 possible outcomes

Reject H_0	and	Do not reject H_0
Do not Reject H_0	and	Reject H_0
Reject H_0	and	Reject H_0
Do not reject H_0	and	Do not reject H_0

Cointegration: An Introduction

- In most cases, if we combine two variables which are I(1), then the combination will also be I(1)
- More generally, if we combine variables with differing orders of integration, the combination will have an order of integration equal to the largest. i.e.,

  if $X_{i,t} \sim \mathrm{I}(d_i)$ for $i = 1, 2, 3, \ldots, k$

  so we have k variables each integrated of order d_i.

  Let

$$z_t = \sum_{i=1}^{k} \alpha_i X_{i,t} \tag{6}$$

  Then $z_t \sim \mathrm{I}(\max d_i)$.

Linear Combinations of Non-stationary Variables

- Rearranging (6), we can write

$$X_{1,t} = \sum_{i=2}^{k} \beta_i X_{i,t} + z_t'$$

where $\beta_i = -\frac{\alpha_i}{\alpha_1}, z_t' = \frac{z_t}{\alpha_1}, i = 2, \ldots, k$.

- This is just a regression equation.
- But the disturbances would have some very undesirable properties: z_t' is not stationary and is autocorrelated if all of the X_i are I(1).
- We want to ensure that the disturbances are I(0). Under what circumstances will this be the case?

Definition of Cointegration (Engle & Granger, 1987)

- Let z_t be a $k \times 1$ vector of variables, then the components of z_t are cointegrated of order (d,b) if
  i. All components of z_t are I(d)
  ii. There is at least one vector of coefficients α such that

$$\alpha' z_t \sim \mathrm{I}(d - b)$$

- Many time series are non-stationary but "move together" over time.
- If variables are cointegrated, it means that a linear combination of them will be stationary.
- There may be up to r linearly independent cointegrating relationships (where $r \leq k - 1$), also known as cointegrating vectors. r is also known as the cointegrating rank of z_t.
- A cointegrating relationship may also be seen as a long term relationship.

Cointegration and Equilibrium

- Examples of possible Cointegrating Relationships in finance:
  - spot and futures prices
  - ratio of relative prices and an exchange rate
  - equity prices and dividends
- Market forces arising from no arbitrage conditions should ensure an equilibrium relationship.
- No cointegration implies that series could wander apart without bound in the long run.

Equilibrium Correction or Error Correction Models

- When the concept of non-stationarity was first considered, a usual response was to independently take the first differences of a series of I(1) variables.
- The problem with this approach is that pure first difference models have no long run solution.
  e.g. Consider y_t and x_t both I(1).
  The model we may want to estimate is

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

  But this collapses to nothing in the long run.
- The definition of the long run that we use is where

$$y_t = y_{t-1} = y; x_t = x_{t-1} = x.$$

- Hence all the difference terms will be zero, i.e.

$$\Delta y_t = 0; \Delta x_t = 0.$$

Specifying an ECM

- One way to get around this problem is to use both first difference and levels terms, e.g.

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \gamma x_{t-1}) + u_t \tag{7}$$

- $y_{t-1} - \gamma x_{t-1}$ is known as the error correction term.
- Provided that y_t and x_t are cointegrated with cointegrating coefficient γ, then $(y_{t-1} - \gamma x_{t-1})$ will be I(0) even though the constituents are I(1).
- We can thus validly use OLS on (7)
- The Granger representation theorem shows that any cointegrating relationship can be expressed as an equilibrium correction model.

Testing for Cointegration in Regression

- The model for the equilibrium correction term can be generalised to include more than two variables:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t \tag{8}$$

- u_t should be I(0) if the variables y_t, $x_{2t}, \ldots x_{kt}$ are cointegrated
- So what we want to test is the residuals of equation (8) to see if they are non-stationary or stationary. We can use the DF/ADF test on u_t.

  So we have the regression

$$\Delta \hat{u}_t = \psi \hat{u}_{t-1} + v_t \quad \text{with} \quad v_t \sim iid.$$

- However, since this is a test on the residuals of an actual model, $\hat{u}_t$, then the critical values are changed.

Testing for Cointegration in Regression: Conclusions

- Engle and Granger (1987) have tabulated a new set of critical values and hence the test is known as the Engle Granger (E.G.) test.
- We can also use the Durbin Watson test statistic or the Phillips Perron approach to test for non-stationarity of $\hat{u}_t$.
- What are the null and alternative hypotheses for a test on the residuals of a potentially cointegrating regression?

  H_0 : unit root in cointegrating regression's residuals

  H_1 : residuals from cointegrating regression are stationary

Methods of Parameter Estimation in Cointegrated Systems: The Engle-Granger Approach

- There are (at least) 3 methods we could use: Engle Granger, Engle and Yoo, and Johansen.
- The Engle Granger 2 Step Method

  This is a single equation technique which is conducted as follows:

  Step 1:
  - Make sure that all the individual variables are I(1).
  - Then estimate the cointegrating regression using OLS.
  - Save the residuals of the cointegrating regression, .
  - Test these residuals to ensure that they are I(0).

  Step 2:

Methods of Parameter Estimation in Cointegrated Systems: The Engle-Granger Approach (Cont'd)

– Use the step 1 residuals as one variable in the error correction model e.g.

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2(\hat{u}_{t-1}) + u_t$$

where $\hat{u}_{t-1} = y_{t-1} - \hat{\tau} x_{t-1}$.

An Example of a Model for Non-stationary Variables: Lead-Lag Relationships between Spot and Futures Prices

Background

- We expect changes in the spot price of a financial asset and its corresponding futures price to be perfectly contemporaneously correlated and not to be cross-autocorrelated.

$$\begin{aligned}
&\text{corr}(\Delta\log(f_t), \Delta\ln(s_t)) \approx 1 \\
&\text{corr}(\Delta\log(f_t), \Delta\ln(s_{t-k})) \approx 0 \quad \forall\ k > 0 \\
&\text{corr}(\Delta\log(f_{t-j}), \Delta\ln(s_t)) \approx 0 \quad \forall\ j > 0
\end{aligned}$$

- We can test this idea by modelling the lead-lag relationship between the two.
- We will consider two papers Tse(1995) and Brooks et al (2001).

Futures & Spot Data

- Tse (1995): 1055 daily observations on NSA stock index and stock index futures values from December 1988 - April 1993.
- Brooks *et al* (2001): 13,035 10-minutely observations on the FTSE 100 stock index and stock index futures prices for all trading days in the period June 1996 – 1997.

Methodology

- The fair futures price is given by

$$F_t^* = S_t \, e^{(r-d)(T-t)}$$

where F_t^* is the fair futures price, S_t is the spot price, r is a continuously compounded risk-free rate of interest, d is the continuously compounded yield in terms of dividends derived from the stock index until the futures contract matures, and $(T-t)$ is the time to maturity of the futures contract. Taking logarithms of both sides of equation above gives

$$f_t^* = s_t + (r-d)(T-t)$$

- First, test f_t and s_t for nonstationarity.

Dickey-Fuller Tests on Log-Prices and Returns for High Frequency FTSE Data

	Futures	Spot
Dickey–Fuller statistics for log-price data	−0.1329	−0.7335
Dickey–Fuller statistics for returns data	−84.9968	−114.1803

Cointegration Test Regression and Test on Residuals

- Conclusion: $\log F_t$ and $\log S_t$ are not stationary, but $\Delta \log F_t$ and $\Delta \log S_t$ are stationary.
- But a model containing only first differences has no long run relationship.
- Solution is to see if there exists a cointegrating relationship between f_t and s_t which would mean that we can validly include levels terms in this framework.
- Potential cointegrating regression:

$$s_t = \gamma_0 + \gamma_1 f_t + z_t$$

  where z_t is a disturbance term.
- Estimate the regression, collect the residuals, $\hat{Z}_t$, and test whether they are stationary.

Estimated Equation and Test for Cointegration for High Frequency FTSE Data

Coefficient	*Estimated value*
$\hat{\gamma}_0$	0.1345
$\hat{\gamma}_1$	0.9834
DF test on residuals	*Test statistic*
$\hat{z}_t$	-14.7303

Source: Brooks, Rew and Ritson (2001).

Conclusions from Unit Root and Cointegration Tests

- Conclusion: $\hat{Z}_t$ are stationary and therefore we have a cointegrating relationship between log F_t and log S_t.
- Final stage in Engle-Granger 2-step method is to use the first stage residuals, $\hat{Z}_t$ as the equilibrium correction term in the general equation.
- The overall model is

$$\Delta \log s_t = \beta_0 + \delta \hat{z}_{t-1} + \beta_1 \Delta \ln s_{t-1} + \alpha_1 \Delta \ln f_{t-1} + v_t$$

Estimated Error Correction Model for High Frequency FTSE Data

Coefficient	Estimated value	*t*-ratio
$\hat{\beta}_0$	9.6713E−06	1.6083
$\hat{\delta}$	−0.8388	−5.1298
$\hat{\beta}_1$	0.1799	19.2886
$\hat{\alpha}_1$	0.1312	20.4946

Source: Brooks, Rew and Ritson (2001).

Look at the signs and significances of the coefficients:

- $\hat{\alpha}_1$ is positive and highly significant
- $\hat{\beta}_1$ is positive and highly significant
- $\hat{\delta}$ is negative and highly significant

Forecasting High Frequency FTSE Returns

- Is it possible to use the error correction model to produce superior forecasts to other models?

Comparison of Out of Sample Forecasting Accuracy

	ECM	ECM-COC	ARIMA	VAR
RMSE	0.0004382	0.0004350	0.0004531	0.0004510
MAE	0.4259	0.4255	0.4382	0.4378
% Correct direction	67.69%	68.75%	64.36%	66.80%

Source: Brooks, Rew and Ritson (2001).

Can Profitable Trading Rules be Derived from the ECM-COC Forecasts?

- The trading strategy involves analysing the forecast for the spot return, and incorporating the decision dictated by the trading rules described below. It is assumed that the original investment is £1000, and if the holding in the stock index is zero, the investment earns the risk free rate.
  - Liquid Trading Strategy - making a round trip trade (i.e. a purchase and sale of the FTSE100 stocks) every ten minutes that the return is predicted to be positive by the model.
  - Buy-&-Hold while Forecast Positive Strategy - allows the trader to continue holding the index if the return at the next predicted investment period is positive.
  - Filter Strategy: Better Predicted Return Than Average - involves purchasing the index only if the predicted returns are greater than the average positive return.

Can Profitable Trading Rules be Derived from the ECM-COC Forecasts? (Cont'd)

- Filter Strategy: Better Predicted Return Than First Decile - only the returns predicted to be in the top 10% of all returns are traded on
- Filter Strategy: High Arbitrary Cut Off - An arbitrary filter of 0.0075% is imposed,

Spot Trading Strategy Results for Error Correction Model Incorporating the Cost of Carry

Trading strategy	Terminal Wealth (£)	Return(%) annualised	Terminal Wealth (£) with slippage	Return(%) Annualised with slippage	Number of trades
Passive investment	1040.92	4.09 {49.08}	1040.92	4.09 {49.08}	1
Liquid trading	1156.21	15.62 {187.44}	1056.38	5.64 {67.68}	583
Buy-and-Hold while forecast positive	1156.21	15.62 {187.44}	1055.77	5.58 {66.96}	383
Filter I	1144.51	14.45 {173.40}	1123.57	12.36 {148.32}	135
Filter II	1100.01	10.00 {120.00}	1046.17	4.62 {55.44}	65
Filter III	1019.82	1.98 {23.76}	1003.23	0.32 {3.84}	8

Source: Brooks, Rew and Ritson (2001).

Conclusions

- The futures market "leads" the spot market because:
  - the stock index is not a single entity, so
    - some components of the index are infrequently traded
    - it is more expensive to transact in the spot market
    - stock market indices are only recalculated every minute
- Spot & futures markets do indeed have a long run relationship.
- Since it appears impossible to profit from lead/lag relationships, their existence is entirely consistent with the absence of arbitrage opportunities and in accordance with modern definitions of the efficient markets hypothesis.

The Engle-Granger Approach: Some Drawbacks

This method suffers from a number of problems:

1. Unit root and cointegration tests have low power in finite samples
2. We are forced to treat the variables asymmetrically and to specify one as the dependent and the other as independent variables.
3. Cannot perform any hypothesis tests about the actual cointegrating relationship estimated at stage 1.

- Problem 1 is a small sample problem that should disappear asymptotically.
- Problem 2 is addressed by the Johansen approach.
- Problem 3 is addressed by the Engle and Yoo approach or the Johansen approach.

The Engle & Yoo 3-Step Method

- One of the problems with the EG 2-step method is that we cannot make any inferences about the actual cointegrating regression.
- The Engle & Yoo (EY) 3-step procedure takes its first two steps from EG.
- EY add a third step giving updated estimates of the cointegrating vector and its standard errors.
- The most important problem with both these techniques is that in the general case above, where we have more than two variables which may be cointegrated, there could be more than one cointegrating relationship.

  In fact there can be up to r linearly independent cointegrating vectors (where $r \leq g - 1$), where g is the number of variables in total.

The Engle & Yoo 3-Step Method (Cont'd)

- So, in the case where we just had y and x, then r can only be one or zero.
- But in the general case there could be more cointegrating relationships.
- And if there are others, how do we know how many there are or whether we have found the "best"?
- The answer to this is to use a systems approach to cointegration which will allow determination of all r cointegrating relationships - Johansen's method.

Testing for and Estimating Cointegrating Systems Using the Johansen Technique Based on VARs

- To use Johansen's method, we need to turn the VAR of the form

$$\underset{g\times 1}{y_t} = \underset{g\times g\ g\times 1}{\beta_1 y_{t-1}} + \underset{g\times g\ g\times 1}{\beta_2 y_{t-2}} + \cdots + \underset{g\times g\ g\times 1}{\beta_k y_{t-k}} + \underset{g\times}{u}$$

into a VECM, which can be written as

$$\Delta y_t = \Pi y_{t-k} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{k-1} \Delta y_{t-(k-1)} + u_t$$

where $\Pi = (\sum_{i=1}^{k} \beta_i) - I_g$ and $\Gamma_i = (\sum_{j=1}^{i} \beta_j) - I_g$

Π is a long run coefficient matrix since all the $\Delta y_{t-i} = 0$.

Review of Matrix Algebra necessary for the Johansen Test

- Let Π denote a $g \times g$ square matrix and let c denote a $g \times 1$ non-zero vector, and let λ denote a set of scalars.
- λ is called a characteristic root or set of roots of Π if we can write

$$\Pi c = \lambda c$$
$$g \times g \quad g \times 1 \quad g \times 1$$

- We can also write

$$\Pi c = \lambda I_g c$$

and hence

$$(\Pi - \lambda I_g)c = 0$$

where I_g is an identity matrix, and hence

Review of Matrix Algebra necessary for the Johansen Test (Cont'd)

- Since $c \neq 0$ by definition, then for this system to have zero solution, we require the matrix $(\Pi - \lambda I_g)$ to be singular (i.e. to have zero determinant).

$$|\Pi - \lambda I_g| = 0$$

- For example, let Π be the 2×2 matrix

$$\Pi = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix}$$

- Then the characteristic equation is

$$|\Pi - \lambda I_p| = \left| \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$= \begin{vmatrix} 5-\lambda & 1 \\ 2 & 4-\lambda \end{vmatrix} = (5-\lambda)(4-\lambda) - 2 = \lambda^2 - 9\lambda + 18$$

Review of Matrix Algebra necessary for the Johansen Test (Cont'd)

- This gives the solutions $\lambda = 6$ and $\lambda = 3$.
- The characteristic roots are also known as Eigenvalues.
- The rank of a matrix is equal to the number of linearly independent rows or columns in the matrix.
- We write Rank(Π) = r
- The rank of a matrix is equal to the order of the largest square matrix we can obtain from Π which has a non-zero determinant.
- For example, the determinant of Π above $\neq 0$, therefore it has rank 2.

The Johansen Test and Eigenvalues

- Some properties of the eigenvalues of any square matrix A:
  1. the sum of the eigenvalues is the trace
  2. the product of the eigenvalues is the determinant
  3. the number of non-zero eigenvalues is the rank

- Returning to Johansen's test, the VECM representation of the VAR was

$$\begin{aligned}\Delta y_t &= \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{k-1} \Delta y_{t-(k-1)} \\ &\quad + u_t\end{aligned}$$

- The test for cointegration between the y's is calculated by looking at the rank of the Π matrix via its eigenvalues (to prove this requires some technical intermediate steps).

The Johansen Test and Eigenvalues (Cont'd)

- The rank of a matrix is equal to the number of its characteristic roots (eigenvalues) that are different from zero.
- The eigenvalues denoted λ_i are put in order:

$$\lambda_1 \geq \lambda_2 \geq \lambda_g$$

- If the variables are not cointegrated, the rank of Π will not be significantly different from zero, so $\lambda_i = 0 \quad \forall \quad i$.

  Then if $\lambda_i = 0$, $\ln(1 - \lambda_i) = 0$

  if the λ 's are roots, they must be less than 1 in absolute value.
- Say rank $(\Pi) = 1$, then $\ln(1 - \lambda_1)$ will be negative and $\ln(1 - \lambda_i) = 0$
- If the eigenvalue λ_i is non-zero, then $\ln(1 - \lambda_i) < 0 \quad \forall \quad i > 1$.

The Johansen Test Statistics

- The test statistics for cointegration are formulated as

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{g} \ln(1 - \hat{\lambda}_i)$$

and

$$\lambda_{max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1})$$

where $\hat{\lambda}$ is the estimated value for the ith ordered eigenvalue from the Π matrix.

λ_{trace} tests the null that the number of cointegrating vectors is less than equal to r against an unspecified alternative.

$\lambda_{trace} = 0$ when all the $\lambda_i = 0$, so it is a joint test.

λ_{max} tests the null that the number of cointegrating vectors is r against an alternative of $r + 1$.

Decomposition of the Matrix

- For any $1 < r < g$, Π is defined as the product of two matrices:

$$\begin{array}{ccc} \Pi & = & \alpha\beta' \\ g \times g & & g \times r \quad r \times g \end{array}$$

- β contains the cointegrating vectors while α gives the "loadings" of each cointegrating vector in each equation.

Decomposition of the Matrix (Cont'd)

- For example, if $g = 4$ and $r = 1$, α and β will be 4×1, and Πy_{t-k} will be given by:

$$\Pi = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}_{t-k}$$

or

$$\Pi = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} \begin{pmatrix} \beta_{11} y_1 + \beta_{12} y_2 + \beta_{13} y_3 + \beta_{14} y_4 \end{pmatrix}_{t-k}$$

Johansen Critical Values

- Johansen & Juselius (1990) provide critical values for the 2 statistics. The distribution of the test statistics is non-standard. The critical values depend on:
  1. the value of $g - r$, the number of non-stationary components
  2. whether a constant and/or trend are included in the regressions.
- If the test statistic is greater than the critical value from Johansen's tables, reject the null hypothesis that there are r cointegrating vectors in favour of the alternative that there are more than r.

The Johansen Testing Sequence

- The testing sequence under the null is $r = 0, 1, \ldots, g-1$
- so that the hypotheses for λ_{trace} are

$$\begin{array}{lll} \mathrm{H}_0 : r = 0 & \text{versus} & \mathrm{H}_1 : 0 < r \leq g \\ \mathrm{H}_0 : r = 1 & \text{versus} & \mathrm{H}_1 : 1 < r \leq g \\ \mathrm{H}_0 : r = 2 & \text{versus} & \mathrm{H}_1 : 2 < r \leq g \\ \vdots & \vdots & \vdots \\ \mathrm{H}_0 : r = g-1 & \text{versus} & \mathrm{H}_1 : r = g \end{array}$$

- We keep increasing the value of r until we no longer reject the null.

Interpretation of Johansen Test Results

- But how does this correspond to a test of the rank of the Π matrix?
- r is the rank of Π.
- Π cannot be of full rank (g) since this would correspond to the original y_t being stationary.
- If Π has zero rank, then by analogy to the univariate case, Δy_t depends only on Δy_{t-j} and not on y_{t-1}, so that there is no long run relationship between the elements of y_{t-1}. Hence there is no cointegration.
- For $1 < \text{rank}(\Pi) < g$, there are multiple cointegrating vectors.

Hypothesis Testing Using Johansen

- EG did not allow us to do hypothesis tests on the cointegrating relationship itself, but the Johansen approach does.
- If there exist r cointegrating vectors, only these linear combinations will be stationary.
- You can test a hypothesis about one or more coefficients in the cointegrating relationship by viewing the hypothesis as a restriction on the Π matrix.
- All linear combinations of the cointegrating vectors are also cointegrating vectors.
- If the number of cointegrating vectors is large, and the hypothesis under consideration is simple, it may be possible to recombine the cointegrating vectors to satisfy the restrictions exactly.

Hypothesis Testing Using Johansen (Cont'd)

- As the restrictions become more complex or more numerous, it will eventually become impossible to satisfy them by renormalisation.
- After this point, if the restriction is not severe, then the cointegrating vectors will not change much upon imposing the restriction.
- A test statistic to test this hypothesis is given by

$$test\ statistic = -T \sum_{i=1}^{r} [\ln(1-\lambda_i) - \ln(1-\lambda_i^*)] \sim \chi^2(m)$$

where

λ_i^* are the characteristic roots of the restricted model,

λ_i are the characteristic roots of the unrestricted model,

r is the number of non-zero characteristic roots in the unrestricted model and m is the number of restrictions.

Cointegration Tests using Johansen: Three Examples

Example 1: Hamilton(1994, pp.647)

- Does the PPP relationship hold for the US / Italian exchange rate - price system?
- A VAR was estimated with 12 lags on 189 observations. The Johansen test statistics were

r	λ_{max}	critical value
0	22.12	20.8
1	10.19	14.0

- Conclusion: there is one cointegrating relationship.

Example 2: Purchasing Power Parity (PPP)

- PPP states that the equilibrium exchange rate between 2 countries is equal to the ratio of relative prices
- A necessary and sufficient condition for PPP is that the log of the exchange rate between countries A and B, and the logs of the price levels in countries A and B be cointegrated with cointegrating vector [1 –1 1].
- Chen (1995) uses monthly data for April 1973-December 1990 to test the PPP hypothesis using the Johansen approach.

Cointegration Tests of PPP with European Data

Tests for cointegration between	$r = 0$	$r \leq 1$	$r \leq 2$	α_1	α_2
FRF–DEM	34.63*	17.10	6.26	1.33	−2.50
FRF–ITL	52.69*	15.81	5.43	2.65	−2.52
FRF–NLG	68.10*	16.37	6.42	0.58	−0.80
FRF–BEF	52.54*	26.09*	3.63	0.78	−1.15
DEM–ITL	42.59*	20.76*	4.79	5.80	−2.25
DEM–NLG	50.25*	17.79	3.28	0.12	−0.25
DEM–BEF	69.13*	27.13*	4.52	0.87	−0.52
ITL–NLG	37.51*	14.22	5.05	0.55	−0.71
ITL–BEF	69.24*	32.16*	7.15	0.73	−1.28
NLG–BEF	64.52*	21.97*	3.88	1.69	−2.17
Critical values	31.52	17.95	8.18	–	–

Notes: FRF – French franc; DEM – German mark; NLG – Dutch guilder; ITL – Italian lira; BEF – Belgian franc.
Source: Chen (1995). Reprinted with the permission of Taylor & Francis Ltd <www.tandf.co.uk>.

Example 3: Are International Bond Markets Cointegrated?

- Mills & Mills (1991)
- If financial markets are cointegrated, this implies that they have a "common stochastic trend".
- Data:

  Daily closing observations on redemption yields on government bonds for 4 bond markets: US, UK, West Germany, Japan.
- For cointegration, a necessary but not sufficient condition is that the yields are nonstationary. All 4 yields series are I(1).

Testing for Cointegration Between the Yields

- The Johansen procedure is used. There can be at most 3 linearly independent cointegrating vectors.
- Mills & Mills use the trace test statistic:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{g} \ln(1 - \hat{\lambda}_i)$$

where λ_i are the ordered eigenvalues.

Johansen Tests for Cointegration between International Bond Yields

r (number of cointegrating vectors under the null hypothesis)	Test statistic	Critical values 10%	Critical values 5%
0	22.06	35.6	38.6
1	10.58	21.2	23.8
2	2.52	10.3	12.0
3	0.12	2.9	4.2

Source: Mills and Mills (1991). Reprinted with the permission of Blackwell Publishers.

Testing for Cointegration Between the Yields (Cont'd)

- Conclusion: No cointegrating vectors.
- The paper then goes on to estimate a VAR for the first differences of the yields, which is of the form

$$\Delta X_t = \sum_{i=1}^{k} \Gamma_i \Delta X_{t-i} + v_t$$

where:

$$X_t = \begin{bmatrix} X(US)_t \\ X(UK)_t \\ X(WG)_t \\ X(JAP)_t \end{bmatrix}, \Gamma_i = \begin{bmatrix} \Gamma_{11i} & \Gamma_{12i} & \Gamma_{13i} & \Gamma_{14i} \\ \Gamma_{21i} & \Gamma_{22i} & \Gamma_{23i} & \Gamma_{24i} \\ \Gamma_{31i} & \Gamma_{32i} & \Gamma_{33i} & \Gamma_{34i} \\ \Gamma_{41i} & \Gamma_{42i} & \Gamma_{43i} & \Gamma_{44i} \end{bmatrix}, v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \end{bmatrix}$$

They set $k = 8$

Variance Decompositions for VAR of International Bond Yields

Explaining movements in	Days ahead	Explained by movements in			
		US	UK	Germany	Japan
US	1	95.6	2.4	1.7	0.3
	5	94.2	2.8	2.3	0.7
	10	92.9	3.1	2.9	1.1
	20	92.8	3.2	2.9	1.1
UK	1	0.0	98.3	0.0	1.7
	5	1.7	96.2	0.2	1.9
	10	2.2	94.6	0.9	2.3
	20	2.2	94.6	0.9	2.3
Germany	1	0.0	3.4	94.6	2.0
	5	6.6	6.6	84.8	3.0
	10	8.3	6.5	82.9	3.6
	20	8.4	6.5	82.7	3.7
Japan	1	0.0	0.0	1.4	100.0
	5	1.3	1.4	1.1	96.2
	10	1.5	2.1	1.8	94.6
	20	1.6	2.2	1.9	94.2

Source: Mills and Mills (1991). Reprinted with the permission of Blackwell Publishers.

Impulse Responses for VAR of International Bond Yields

	Response of US to innovations in			
Days after shock	US	UK	Germany	Japan
0	0.98	0.00	0.00	0.00
1	0.06	0.01	−0.10	0.05
2	−0.02	0.02	−0.14	0.07
3	0.09	−0.04	0.09	0.08
4	−0.02	−0.03	0.02	0.09
10	−0.03	−0.01	−0.02	−0.01
20	0.00	0.00	−0.10	−0.01
	Response of UK to innovations in			
Days after shock	US	UK	Germany	Japan
0	0.19	0.97	0.00	0.00
1	0.16	0.07	0.01	−0.06
2	−0.01	−0.01	−0.05	0.09
3	0.06	0.04	0.06	0.05
4	0.05	−0.01	0.02	0.07
10	0.01	0.01	−0.04	−0.01
20	0.00	0.00	−0.01	0.00
	Response of Germany to innovations in			
Days after shock	US	UK	Germany	Japan
0	0.07	0.06	0.95	0.00
1	0.13	0.05	0.11	0.02
2	0.04	0.03	0.00	0.00
3	0.02	0.00	0.00	0.01
4	0.01	0.00	0.00	0.09
10	0.01	0.01	−0.01	0.02
20	0.00	0.00	0.00	0.00
	Response of Japan to innovations in			
Days after shock	US	UK	Germany	Japan
0	0.03	0.05	0.12	0.97
1	0.06	0.02	0.07	0.04
2	0.02	0.02	0.00	0.21
3	0.01	0.02	0.06	0.07
4	0.02	0.03	0.07	0.06
10	0.01	0.01	0.01	0.04
20	0.00	0.00	0.00	0.01

Source: Mills and Mills (1991). Reprinted with the permission of Blackwell Publishers.