Unit Root Tests in Time Series Volume 1

Palgrave Texts in Econometrics
General Editor: **Kerry Patterson**

Titles include:
Simon P. Burke and John Hunter
MODELLING NON-STATIONARY TIME SERIES

Michael P. Clements
EVALUATING ECONOMETRIC FORECASTS OF ECONOMIC AND FINANCIAL VARIABLES

Lesley Godfrey
BOOTSTRAP TESTS FOR REGRESSION MODELS

Terence C. Mills
MODELLING TRENDS AND CYCLES IN ECONOMIC TIME SERIES

Kerry Patterson
A PRIMER FOR UNIT ROOT TESTING

Kerry Patterson
UNIT ROOT TESTS IN TIME SERIES VOLUME 1
Key Concepts and Problems

Palgrave Texts in Econometrics
Series Standing Order ISBN 978–1–4039–0172–9 (hardback) 978–1–4039-0173–6 (paperback)
(outside North America only)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and the ISBNs quoted above.

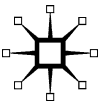
Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England

Unit Root Tests in Time Series Volume 1

Key Concepts and Problems

Kerry Patterson

palgrave
macmillan



Softcover reprint of the hardcover 1st edition 2011 978-0-230-25024-6

First published 2011 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN 978-0-230-25025-3 ISBN 978-0-230-29930-6 (eBook)
DOI 10.1057/9780230299306

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Patterson, K. D.
Unit root tests in time series volume 1 : key concepts and problems / Kerry Patterson.
p. cm.
Includes bibliographical references and index.

1. Econometrics. 2. Time-series analysis. I. Title.

HB139.P374 2010
519.5′5—dc22 2010027477

10 9 8 7 6 5 4 3 2 1
20 19 18 17 16 15 14 13 12 11

To Bella and a continuing treasure at Auton Farm

Contents

Detailed Contents

List of Tables

List of Figures

Symbols and Abbreviations

$\Rightarrow_D$	convergence in distribution (weak convergence)
$\rightarrow_p$	convergence in probability
$\rightarrow$	tends to; for example, ε tends to zero, $\varepsilon \rightarrow 0$
$\mapsto$	mapping
$\Rightarrow$	implies
$\sim$	is distributed as
$\equiv$	definitional equality
$\neq$	not equals
$\Phi(z)$	the cumulative distribution function of the standard normal distribution
$\Re$	the set of real numbers; the real line ($-\infty$ to ∞)
$\Re^+$	the positive half of the real line
N^+	the set of non-negative integers
ε_t	white noise, unless explicitly excepted
$\prod_{j=1}^{n} x_j$	the product of x_j, $j = 1, \ldots, n$
$\sum_{j=1}^{n} x_j$	the sum of x_j, $j = 1, \ldots, n$
0^+	approach zero from above
$\_0$	approach zero from below
L	the lag operator, $L^j y_t \equiv y_{t-j}$
Δ	the first-difference operator, $\Delta \equiv (1 - L)$
Δ_s	the s-th difference operator, $\Delta_s \equiv (1 - L^j)$
Δ^s	the s-th multiple of the first difference operator, $\Delta^s \equiv (1 - L)^s$
$\subset$	a proper subset of
$\subseteq$	a subset of
$\cap$	intersection of sets
$\cup$	union of sets
$\in$	an element of
$\|a\|$	the absolute value (modulus) of a
iid	independent and identically distributed
m.d.s	martingale difference sequence
$N(0, 1)$	the standard normal distribution, with zero mean and unit variance
niid	independent and identically normally distributed
B(t)	standard Brownian motion; that is, with unit variance
W(t)	nonstandard Brownian motion

Preface

The purpose of this book is to provide a review and critical assessment of some key procedures related in one way or another to the problem of testing for a unit root in a stochastic process. As is now well known, the presence of a unit root implies a form of nonstationarity that is considered to be relevant for economic time series, so that a non-standard inferential and distributional framework is required.

The research and literature on this topic has grown almost exponentially since Nelson and Plosser's seminal article published in 1982. Therein they applied the framework due to Dickey (1976) and Fuller (1976) to testing for a unit root in a number of macroeconomic time series. Subsequent key articles by Dickey and Fuller (DF) (1979, 1981) developed some aspects of the initial testing framework. The basic set-up for a DF unit root test is now familiar enough, being taught in most intermediate, if not introductory, courses in econometrics; however, the underlying distribution theory is somewhat more advanced, and the many complications that have arisen in practice has meant the development of a voluminous literature that, because of its extent, is difficult to comprehend, especially for the non-specialist. Indeed, it is probably the case that a simple survey of the field of methods and applications is virtually infeasible; indeed, the topic is so extensive that even some 20 years ago Diebold and Nerlove (1990) noted the scale of the literature on this topic.

The articles on unit root tests are amongst the most cited in economics and econometrics and have clearly influenced the direction of economic research at a much wider level than simply testing for a unit root. A citation summary for articles based on univariate processes is presented in Table P.1. The numbers shown here clearly indicate that there has been a sustained interest in the topic over the last 30 years or so and, looking at the wider influence, the unit root literature led to the concept of cointegration and to some of the most cited of econometric articles, including Engle and Granger (1987), Johansen (1988, 1991) and Johansen and Juselius (1990) (see the note to Table P.1).

The appropriate prerequisites for this book include some knowledge of econometric theory at an intermediate or graduate level as, for example, in Davidson (2000), Davidson and Mackinnon (2004) or Mittlehammer et al. (2000), and, possibly, with some additional directed study, as in good introductory books such as Gujarati (2006), Dougherty (2007), Ramanathan (2002) and Stock and Watson (2007). It would also be helpful to have had an introduction to the

Table P.1 Number of citations of key articles on unit roots.

Author(s)		Number of citations
Dickey and Fuller (1979)	1	7,601
Phillips and Perron (1988)	2	4,785
Dickey and Fuller (1981)	3	4,676
Perron (1989)	4	3,371
Kwiatkowski, Phillips, Schmidt, Shin (1992)	5	3,280
Nelson and Plosser (1982)	6	3,035
Phillips (1987a)	7	1,881
Zivot and Andrews	8	1,694
Elliott, Rothenberg and Stock (1996)	9	1,556
Said and Dickey (1984)	10	1,342

Notes: Articles relate to univariate unit root tests. Prominent articles, on a citation basis, involving largely multivariate methods are: Engle and Granger (1987): 12,366; Johansen (1988): 8,236; Johansen and Juselius (1990): 4,886; Johansen (1991): 4,150; and in econometric methods more generally, White (1980), 12,359. (The last article is the most highly cited on the basis of several citation methods.) On the index of economics articles since 1970, compiled by Han Kim et al. (2006), Dickey and Fuller (1979) ranks 7, whilst Engle and Granger (1987) and Johansen (1988) rank 4 and 8, respectively.
Source: Google Scholar, accessed 22 February 2010.

methods of maximum likelihood and generalised least squares (GLS), for example as provided in Greene (2006). An introduction to time series analysis and mathematical statistics would also be useful; for example, for the former at the level of Chatfield (2004) and the latter along the lines covered by Mittelhammer (1996). A book designed especially as a primer for this one is Patterson (2010). Some familiarity with the application of unit root tests would also be helpful to set the context.

I have taken the brief of this book to include issues that are related to but theoretically separate from the central concern of testing for a unit root. For example, one of these is the problem of the bias in estimating the coefficients in an autoregressive model; whilst this is strictly a finite sample effect, it is of practical interest and serves as a 'lead in' to the problems associated with unit testing. This book is, therefore, not about listing tests for a unit root. Not only would there not be enough space for such an enterprise, it is not the best way to indicate which test statistics, and methods more generally, have been taken up by practitioners. The main tests are of course presented and their rationale explained, together with examples to illustrate how they are used. However, the problem suggested by the presence of a unit root or near-unit root is more than just the design of a test statistic. There are two other important practical issues that a researcher has to face. The first is to consider what the appropriate alternative hypothesis is. In the Dickey-Fuller paradigm, followed by many in practice, the alternative to a stochastic trend is a deterministic trend, usually

characterised as a low order polynomial in time. However, the choice of the order of the polynomial, typically representing reversion to a constant or a linear trend, has a critical effect on the power of the test if over-specified; on the other hand, if under-specified, a test will have no power to detect a true alternative. The representation of the 'attractor' as being generated by a low order polynomial trend is likely to be a shorthand, or reduced form, for a far more complex process. In part, the pre-eminent role of a simple deterministic trend in providing the mean or trend reverting alternative to a non-reverting process (nonstationary by way of a unit root or roots), is historical and, of itself, deserves further study and evaluation.

A second key, practical aspect arises from the usual need to choose some form of a truncation parameter. In the context of the familiar augmented Dickey-Fuller tests, this is a lag truncation parameter and in the context of the semi-parametric tests, which require an estimate of the long-run variance, a parameter limiting the bandwidth in the formation of a sum of autocovariances. Whilst familiar criteria, such as the AIC, BIC and general-to-specific (g-t-s), rules are in frequent use, the combination of each one of these with a test statistic defines test procedures with potentially differing characteristics; their use in combination then leads to the accumulation of type I error.

Some of the developments covered in this book are as follows.

- The distinction between difference stationary and trend stationary processes and the implications of this distinction for the permanence or otherwise of shocks.
- An outline of the autoregressive moving average (ARMA) modelling framework and its role in testing for a unit root.
- The finite sample bias in estimating models and its implications for inference even when seemingly well into the region of stationarity.
- Forming confidence intervals that are robust to the problem of quantiles that are not constant.
- The DF unit root tests and developments of them to account for weak serially correlated processes.
- Bootstrapping confidence intervals and unit root tests.
- Tests that:
  - are based on a direct maximum likelihood approach;
  - are based on a GLS, or quasi-GLS, approach, including the influential tests by Elliott, Rothenberg and Stock (1996);
  - combine the backward and forward recursions of a random walk;
  - are based on recursive estimation of the mean or trend;
  - are robust to the initial condition;
  - allow for more than one unit root;

- are based on stationarity as the null hypothesis;
- allow for unit roots in seasonal data.

The results of a number of Monte Carlo studies are reported in various chapters. Indeed, simulation is a key tool that is used throughout to provide some guidance on finite sample issues. Consider, for example, the problems caused by the presence of weakly dependent errors when testing for a unit root. Then under fairly weak assumptions, the asymptotic properties of several frequently used test statistics are unaffected by such errors, but typically, the finite sample properties do not reflect the asymptotic properties, an example being the difficulty caused by the near cancellation of a root, especially a near-unit root, in the AR and MA components of an ARMA model. To understand the finite sample nature of such problems, many more simulations were run than are reported in the various chapters; the results are then typically illustrated for one or two sample sizes where they are representative of a wider range of sample sizes.

There are a number of developments and problems not covered in this volume, but which are included in Volume 2. These include the following.

Nonparameteric tests: the tests that have been considered in this volume, such as the family of Dickey-Fuller tests, are parametric tests in the sense that they are concerned with direct estimation in the context of the parametric structure of an AR or ARMA model. Nonparametric tests use less structure in that no such explicit parametric framework is required and inference is based on other information in the data, such as ranks, signs and runs. Semi-parametric tests use some structure, but it falls short of a complete parametric setting; an example here is the rank score based test, which is based on ranks.

Fractional integration: this considers the case of fractional values of the integration parameter. That is, suppose that a stochastic process generates a time series that is integrated of order d, where d is a fractional number. What meaning can we attribute to such an operation and how can the parameter d be estimated? There are two general approaches to the analysis and estimation of fractional I(d) process, as they may be either analysed in the time domain or the frequency domain.

Bounded random walks: the application of random walk models to some economic time series can be inappropriate, as where there are natural bounds or limits to the values that the series can take, such as in the case of unemployment rates and nominal interest rates. One way of modelling this is to allow unit root behavior; for example, persistence and the absence of mean reversion over a range of possible values but reversion at other values. These models have

in common that they involve some form of nonlinearity. Perhaps the simplest from of nonlinearity actually arises from piecewise linearity; that is, an overall model comprises two or more linear models for subperiods where the component models differ not in their form, for example all are AR(p), but in their parameters. A popular class of such models is the smooth transition autoregressive – or STAR – class, of which the exponential and logistic members are the most frequent in application, giving rise to the acronyms STAR and LSTAR.

Structural breaks: Perron's (1989) seminal article began another thread of the unit root literature. What if, instead of a unit root process generating the data, there was a trend subject to a break due to 'exceptional' events? How would standard unit root tests perform? For example, what would be their power characteristics, if the break was ignored in the alternative hypothesis? The idea of regime change that could affect led to a fundamental re-evaluation of the simplicity of the simple 'opposing' mechanisms of a unit root process, on the one hand, and a trend stationary process, on the other. In practice, although there are likely to be some contemporaneous and, later, historical indications of regime changes, there is almost inevitably likely to be uncertainty not only about the dating of such changes but also the nature of the changes. This poses another set of problems for econometric applications. If a break is presumed, when did it occur? Which model captures the nature of the break? If multiple breaks occurred, when did they occur?

My sincere thanks go to Lorna Eames, my secretary at the University of Reading, for her unfailing assistance in the many tasks needed to bring the manuscript into shape.

The graphs in this book were prepared with MATLAB www.mathworks.co.uk, which was also used, together with TSP (www.tspintl.com) and RATS (www.estima.com), for the numerical examples. Martinez and Martinez (2002) provide an invaluable guide to statistics with many MATLAB examples; guides to MATLAB include Hanselman and Littlefield (2004), Moler (2004) and Hahn and Valentine (2007).

If you have comments on any aspects of the book, please contact me at my email address given below.

Author's email address: k.d.patterson@reading.ac.uk
Palgrave Macmillan Online: http://www.palgrave.com/economics/
Palgrave Macmillan email address: orders@palgrave.com

1
Introduction to Random Walks and Brownian Motion

Introduction

The first part of this chapter introduces the random walk initially in a form with stochastic shocks generated by a random variable with a binomial distribution. In the simplest version of this process the random variable has two equally likely outcomes resulting in a symmetric binomial random walk. The idea is simple enough and the terminology is due to a problem posed by Pearson (see Hughes, 1995, p.53), although the concept dates from much earlier, originating in games of chance. Starting from the origin, at regular intervals a walker takes equally spaced steps either to the left (north) or to the right (south), with equal probability. The walker's progress can be plotted in two dimensions by recording the distance from the origin on the vertical axis and the elapsed time on the horizontal axis; for example, one step to the north followed by one step to the south returns the walker to the origin. Such a graph will look like a series of equally sized steps, see Figure 1.1; perhaps surprisingly, the resulting path does not generally oscillate around zero, the theoretical mean of the process. This lack of 'mean reversion' is one of the key characteristics of a random walk. In this form the process has its origins in gambling, where the gambler gains or loses an equal amount at each gamble, with equal probability, and intuition might suggest that the gambler is not systematically losing or winning. The 'distance from the origin' corresponds to the gambler's win/lose tally, which is one-dimensional and can be represented on the vertical axis.

There are several ways to generalise the random walk described above. Indeed, Pearson's problem was originally posed in a more general form. In particular, whilst the walker takes equally spaced steps of length x, he is allowed to pivot through any angle before taking the next step, not simply going north or south, which is an angle of $\pm 90^\circ$ to his orientation. For example, taking an equally spaced step at 45° would place the walker in a north-east direction. The

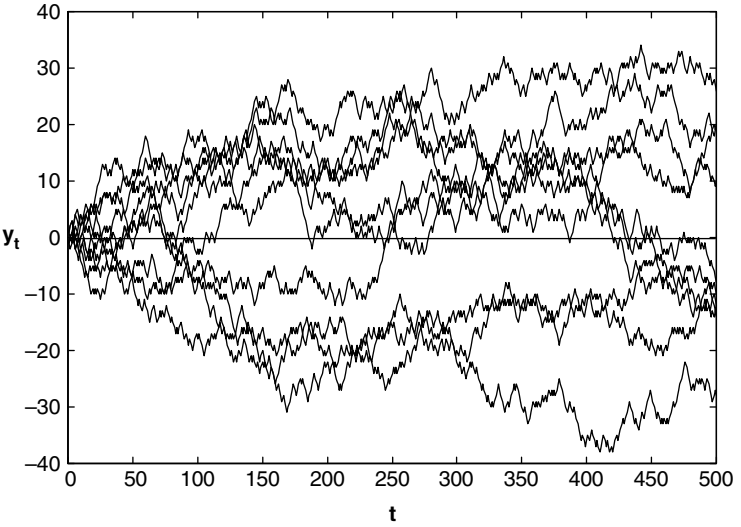


Figure 1.1 Sample paths of a binomial random walk.

walk could therefore be represented as a bird's-eye view of his progress in two-dimensional geographic space (with time an implicit third dimension). In this case starting at the origin, one step in a northward direction followed by one step in a southward direction will not necessarily return the walker to the origin, as the step angle may differ from $\pm 90^\circ$. By focusing on, say, the north-south direction, this random walk can be represented in two dimensions with time as one of the dimensions. The step size in the single dimension will depend on the pivoting angle and so will no longer be a constant; moreover, the resulting random variable is continuous as the step size defined in this dimension can vary continuously between 0 and x.

The random walk is of interest in its own right in economics as it provides a statistical model that is paired with rational expectations and some versions of the efficient markets hypothesis. For example, suitably defined, a random walk has the martingale property that the expected value of the random variable y_t at time t – 1, conditional on the information set Ψ, comprising lagged values of y_t, is y_{t-1}; that is, $E_{t-1}(y_t|\Psi) = y_{t-1}$, where E_{t-1} is the expectation formed at time t – 1. An implication of this property is that $E_t[y_t - E_{t-1}(y_t|\Psi)] = 0$, so that the difference between outturn and expectation, $y_t - E_{t-1}(y_t|\Psi)$, is 'news' relative to the information contained in Ψ, a property associated with the rational expectations hypothesis.

An important property of a random walk process arises in the limit as the time divisions, the time steps, are 'shrunk' toward zero. The limiting process is

Brownian motion, knowledge of which is essential to an understanding of the distribution of several important unit root test statistics. Hence an introduction to the random walk also serves as an introduction to Brownian motion.

It is a matter of choice as to the order in which the random walk, and developments thereof, and the concept of a 'unit root' can be introduced. The preference expressed here is motivated by the fascination often expressed in the counter-intuitive properties of a simple random walk and how quickly one can link such processes to economic time series. Given this background, the more formal testing framework then follows quite naturally, informed by some useful features of random walk type behaviour that can be illustrated graphically.

Section **1.1** outlines the basic random walk with a number of generalisations and key properties including, for example, its lack of mean reversion, sometimes referred to as mean aversion. Section **1.2** defines Brownian motion (BM) and sections **1.3** and **1.4** state two key theorems, based on BM, required in unit root statistics, that is the functional central limit theorem (FCLT) and the continuous mapping theorem (CMT); these are often used together to obtain the limiting distribution of a unit root test statistic. Section **1.5** provides a brief background to the development of unit root and related tests and section **1.6** provides some selective economic examples in which unit root processes are of interest.

This chapter assumes a basic familiarity with probability and some time series concepts, which are the more detailed subjects of later chapters. It is, for example, almost impossible to talk of a unit root without the context of an autoregressive model, which is itself a special case of the ARMA (autoregressive moving average) class of models, considered in greater detail in Chapter 3. Similarly, the idea of a unit root is greatly aided by first studying the lag operator and lag polynomial. The reader may find it useful to review the material in Appendices 1 and 2: Appendix 1 is a brief introduction to random variables and Appendix 2 provides some background material on the lag operator. Where necessary, the ARMA model of Chapter 3 is anticipated in this chapter in just sufficient detail to make the context self-sufficient.

1.1 Random walks

The concept of a random walk has two roles of interest. First, it is prototypical model for the representation of the time series of many economic variables, including real variables such as output and employment and nominal or financial variables such as price levels, inflation rates and exchange rates. It is often taken as a 'default' or 'baseline' model from which other models are evaluated. Second, random walks, or their limiting forms, appear as partial sum processes, psp, in econometric estimators, especially in distribution theory for estimators and test statistics. In the limit, here interpreted as taking smaller and smaller

steps in a given time interval, the random walk leads to Brownian motion, which is considered in section **1.7**.

1.1.1 The random walk as a partial sum process (psp)

We start with the random variable y_t, with a sequence of length T of such variables written as $\{y_t\}_{t=1}^{T}$. Apart from the starting value y_0, values of y_t are determined by the following one period recursion:

$$y_t = y_{t-1} + \varepsilon_t \qquad t = 1, \ldots, T \tag{1.1}$$

Thus, y_t is determined as its previous value, y_{t-1}, plus the intervention of a random variable denoted ε_t, usually referred to as a 'shock' in the sense that y_t would otherwise just be y_{t-1}. The sequence $\{\varepsilon_t\}_{t=1}^{T}$ is assumed to be independent white noise. (White noise, WN, requires $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ for all t, and $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$; if all members of the sequence $\{\varepsilon_t\}_{t=1}^{T}$ are independently and identically distributed, then $\{\varepsilon_t\}_{t=1}^{T}$ is referred to as independent or strong white noise sequence.)

This random walk is a partial sum process (psp) as, by successive back substitution, y_t can be expressed as y_0 plus the cumulated sum of the sequence, to that point, of stochastic inputs:

$$y_t = y_0 + \sum_{j=1}^{t} \varepsilon_j \tag{1.2}$$

One of the insights in distinguishing the behaviour of a time series generated by a random walk $y_t = y_{t-1} + \varepsilon_t$ (and generalisations of this process) is the nature of its evolution over time. Viewing $\{\varepsilon_t\}_1^t$ as a sequence of 'shocks', a particular sample path of y_t is determined by the starting position y_0 and the cumulated shocks, each of which receives an equal weight – in particular, there is no sense in which the past is forgotten, a feature that is sometimes referred to as infinite 'memory'.

Such processes occur quite readily in gambling and have been the subject of considerable study and extension (see, for example, the classic text by Feller, 1968). A simple case that illustrates much of interest is the symmetric, binomial random walk generated by a sequence of gambles on the outcome of the toss of a fair coin: a coin is tossed, with the gambler winning one (+1) 'chip' if it lands heads and losing one (–1) 'chip' if it lands tails; the game continues sequentially with further tosses of the coin, indexed by the counter $t = 1, \ldots, T$. For simplicity, the games are assumed to be played at the rate of one per period t, so t increments in units of 1 from 1 to T. Each toss of the coin is referred to as a 'trial', a term that originates from Bernoulli trials, resulting in the binary outcomes 'success' or 'failure', with probabilities p and q, respectively. Successive trials are independent in the sense that the outcome on any one is unaffected by any of the others.

To put the process more formally, let x_t be the random variable representing the coin toss on the t-th toss of the coin; the sample space associated with each single toss of the coin is $\Omega_j = (H, T)$, and the associated probability measure is $P_j^x = (p_1^x, p_2^x)$; note that an italicised T, *T*, refers to the outcome that the coin toss is a tail. Next let ε_j be the random variable that is derived from x_j, such that the original sample space Ω_j is mapped to $\Omega_j^\varepsilon = (+1, -1)$. The probability measures of x_j and ε_j are the same; for example, for a symmetric random walk $p = q = 1/2$, so that $P_j^x = P_j^\varepsilon = (1/2, 1/2)$; if $p \neq q$ then the resulting process is an asymmetric random walk.

The random variable y_t is the partial sum of the derived random variables ε_j, $j = 1, \ldots, t$, and is usually referred to as the 'tally'. We assume that there is no 'charge' to enter the game, so that $y_0 = 0$. The sample space of y_t is the t-dimensional product space of Ω_1, that is $\Omega_t^y = (\Omega_1^\varepsilon)^t = \Omega_1^\varepsilon \times \Omega_1^\varepsilon \times \ldots \times \Omega_1^\varepsilon$ and, by independence, the probability measure associated with y_t is the product measure $P_t^y = (P_1^\varepsilon)^t = P_1^\varepsilon \times P_1^\varepsilon \times \ldots \times P_1^\varepsilon$. Note that $E(\varepsilon_t) = 0$, $t = 1, \ldots, T$ and $E(y_t) = E(y_0) + \sum_{i=1}^t E(\varepsilon_i) = 0$, so that the theoretical mean of the tally is zero.

The counterpart of the random walk sequence generated by the gambler is the random walk of the 'banker':

$$\begin{aligned} y_{B,t} &= y_{B,t-1} - \varepsilon_t \\ &= y_{B,0} - \sum_{j=1}^t \varepsilon_j \end{aligned} \tag{1.3}$$

Note that this is an example of a 'zero sum' game since $y_{B,t} + y_t = y_{B,0} - \sum_{j=1}^t \varepsilon_j + y_0 + \sum_{j=1}^t \varepsilon_j = y_{B,0} + y_0$, where the latter equals zero if both parties start with zero capital. It will occasionally be useful to look at the random walk from the banker's perspective.

The random walk of Equation (1.1) was simulated with $T = 500$ and binomial inputs; ten sample paths were shown in Figure 1.1. These paths tend to confound intuition. A line of reasoning that seems attractive is that as the expected value of each component random variable ε_j is zero, the expectation of the tally, $E(y_t)$, is zero, hence the sample paths can be expected to fluctuate reasonably evenly about zero. The figure shows that this line of reasoning is false. There is very little reversion to the expected value of zero; indeed, once started in a particular direction, whether that is in the positive or negative half of the diagram, a sample path only rarely crosses (traverses), or is reflected from, the zero axis; these are collectively referred to as 'visits to the origin' or mean reversions, the number of which is a key characteristic that is considered further in section **1.1.3**.

1.1.2 Random walks: visits to the origin (sign changes and reflections)

As noted in the previous section, one of the characteristics of a random walk is that there is very little mean reversion; that is, although $E(y_t) = 0$, which is the

'origin' for this process, this is not an 'attractor' for the sample path. To formalise this idea, let V_T be the number of visits to the origin (including reflections) of a symmetric binomial random walk of length T, then $\zeta_T = V_T/\sqrt{T}$ is the normalised number of such visits with distribution function $F(\zeta_T)$ for finite T. It turns out that $F(\zeta_T)$ has a simple limiting distribution as follows:

$$F(\zeta_T) \Rightarrow_D F(\zeta) = 2\Phi(\zeta) - 1 \tag{1.4}$$

where $\Phi(\zeta)$ is the (cumulative) standard normal distribution function. Equally, one can write in terms of the random variables that $\zeta_T \Rightarrow_D \zeta$, where ζ is a random variable with the half-normal distribution. Thus, the distribution function $F(\zeta)$ is that of the absolute value of a normally distributed random variable with mean $\mu = 0$ and variance σ^2 (see Feller, 1968; Burridge and Guerre, 1996; Garciá and Sansó, 2006). If $\sigma^2 = 1$, as in the symmetric binomial random walk, then $E(\zeta) = \sqrt{1/2\pi} = 0.7979$, so that $E(V_T) = 0.7979\sqrt{T}$, which is to the right of the median; if $T = 500$, then the integer part of $E(V_{500})$ is 17. The median of $F(\zeta)$ is 0.674, so that the median number of mean reversions for (large) T trials is about $0.674\sqrt{T}$ (about because of the need to be an integer); for example, if $T = 500$, then median is about 15. The distribution function, $F(\zeta)$, and its mean and median are illustrated in Figure 1.2.

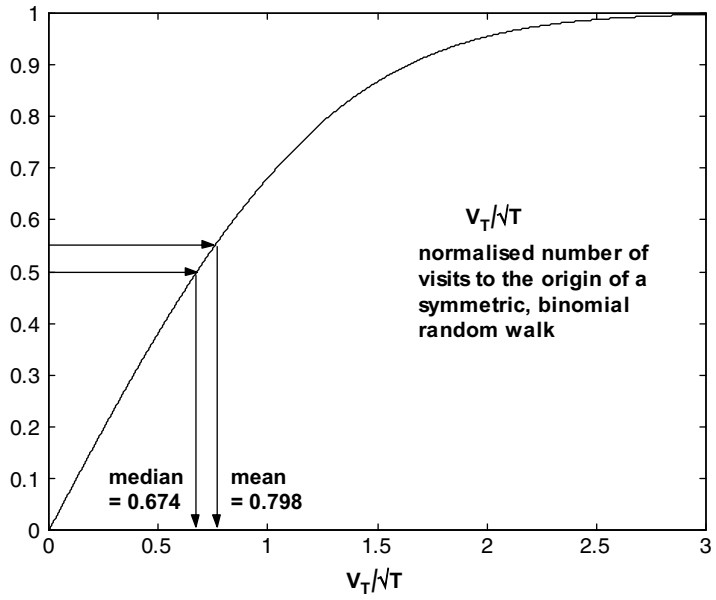


Figure 1.2 Distribution function of visits to the origin.

1.1.3 Random walk: an example of a stochastic process

Note that a random walk is an example of a stochastic process and, in the language of such processes, the realisations in Figure 1.1 are trajectories or sample paths. Our interest generally lies not in the outcome of a random variable at a single point in time, but in a sample path, or the distribution of sample paths, of a sequence of random variables over an interval of time.

To conceptualise how such sample paths arise the idea of a stochastic process involves a sample space Ω, a probability space and time. In the case of the stochastic process defined by the symmetric, binomial random walk then $\Omega = \Omega_t^y$, with probability measure P_t^y and $t = 1, \ldots, T$; note that if $T \to \infty$, then the sample space and the probability space associated with the product measure become of infinite dimension. This can also occur if T is fixed and then partitioned into a grid with 'mesh' size Δt and $\Delta t \to 0$.

Let Θ be the set of possible values taken by the time index, then in the random walk of Equation (1.1), time is discrete, so that Θ has a finite or countably infinite number of elements. Indeed, in this case Θ is the set of positive integers or, if the process is viewed as starting at $t = 0$, the set of nonnegative integers, $t \in \Theta = N^+ = (0, 1, 2, \ldots)$; equally the process might be viewed as starting in the infinite past and, hence, $t \in \Theta = N = (0, \pm 1, \pm 2, \ldots)$. In the continuous time case, Θ is an interval, for example $\Theta = \Re$, or the positive half line $\Theta = \Re^+$ or an interval on $\Re$, for example $\Theta = [0, 1]$ or $\Theta = [0, T]$.

Stochastic processes may be viewed as taking place in discrete time or in continuous time, which are represented, respectively, as follows:

$$Y = (y_t(\omega) : t \in \Theta, \omega \in \Omega) \qquad \text{discrete time} \tag{1.5a}$$

$$Y = (y(t, \omega) : t \in \Theta \subseteq \Re, \omega \in \Omega) \qquad \text{continuous time} \tag{1.5b}$$

The continuous-time stochastic process represented at a discrete or countably infinite number of points is then written either as $y(s)$ or $y(t)$ if only two or three points in time are being referenced or, more generally, as $y(t_1), y(t_2), \ldots, y(t_n)$. Note that reference to ω may be suppressed if it is not material to the presentation.

For given $t \in \Theta$, $y_t(\omega)$, or $y(t, \omega)$, is a function of $\omega \in \Omega$ and is, therefore, a random variable. A realisation is a single number – the point on the sample path relating to, say, $t = s$. By varying the element of Ω, whilst keeping $t = s$, there is a distribution of outcomes at that point. For given $\omega \in \Omega$, $y_t(\omega)$ is a function of time, $t \in \Theta$. In this case an 'outcome' is a complete sample path; that is, a function of $t \in \Theta$, rather than a single number. A description of the sample path would require a functional relationship rather than a single number. By varying ω a different complete sample path is obtained; that is, (potentially) different realisations for all $t \in \Theta$.

The component random variables in the binomial random walk are discrete, with a simple countable number of outcomes; indeed, just two in this case. Hence this specification of the random walk is an example of a discrete time, discrete variable stochastic process. Later we will consider a random walk where the stochastic inputs, ε_t, are distributed as $N(0, \sigma_\varepsilon^2)$, in which case the stochastic process so generated is an example of a discrete time, continuous variable stochastic process. A case of particular interest that arises in the context of Brownian motion is a continuous time stochastic process. One can view this as the limit of a discrete time process in which a given interval of time is divided into a finer and finer grid, so that $\Delta t \to 0$ and $\Theta = \Re$, or the positive half-line $\Theta = \Re^+$ or an interval on $\Re$, for example, $\Theta = [0, 1]$.

As noted, often the reference to $\omega \in \Omega$ is suppressed and a single random variable in the stochastic process is written y_t, but the underlying dependence on the sample space should be recognised and means that different $\omega \in \Omega$ give rise to potentially different sample paths.

1.1.4 Random walk: an example of a nonstationary process

A key distinction in econometrics and statistics is between processes that are stationary and those that are nonstationary. In a time series context, these are said to generate time series that are, respectively, stationary or nonstationary, it being understood that it is the underlying generating process that is stationary or nonstationary. Intuitively, stochastic processes that are stationary are unchanging in some key aspects, which give rise to several definitions of stationarity, differences between them depending on what is assumed to be unchanging. A strong form of stationarity requires that the joint probability distribution of the random variables that comprise the stochastic process is unchanging; however, the most often used definition in econometrics relates to a weakly (or second-order) stationary process. These definitions are considered in the next two subsections.

1.1.4.i A strictly stationary process

Let $\tau \neq s$ and T be arbitrary, if Y is a strictly stationary, discrete time process for a discrete random variable, y_t, then:

$$P(y_{\tau+1}, y_{\tau+2}, \ldots, y_{\tau+T}) = P(y_{s+1}, y_{s+2}, \ldots, y_{s+T}) \tag{1.6}$$

That is, the joint probability mass function (pmf) for the sequence of length T starting at time $\tau+1$ is the same for any shift in the time index from τ to s and for any choice of T. This means that it does not matter which T-length portion of the sequence we observe. Since a special case of this result in the discrete case is for $T = 1$; that is, $P(y_\tau) = P(y_s)$, the marginal pmfs must also be the same for $\tau \neq s$, implying that $E(y_\tau) = E(y_s)$. These results imply that other moments, including joint moments, such as the covariances, are invariant to arbitrary time shifts.

If the random variables are continuous and also defined in continuous time, a strictly stationary random process must satisfy the following:

$$F[y(\tau+t_1), y(\tau+t_2), \ldots, y(\tau+t_N)] = F[y(s+t_1), y(s+t_2), \ldots, y(s+t_N)] \tag{1.7}$$

where $t_1 < t_2 \ldots < t_N$, $\tau \neq s$ and F(.) is the joint distribution function, (see Appendix 1). If the probability density functions (pdfs) exist, then an analogous condition holds where F(.) is replaced by the joint pdf, denoted f(.):

$$f[y(\tau+t_1), y(\tau+t_2), \ldots, y(\tau+t_N)] = f[y(s+t_1), y(s+t_2), \ldots, y(s+t_N)] \tag{1.8}$$

1.1.4.ii Weak or second-order stationarity (covariance stationarity)

A less demanding form of stationarity is weak or second-order stationarity, which requires that the following three conditions are satisfied for arbitrary τ and s, $\tau \neq s$:

$$SS1: E(y_\tau) = E(y_s) = \mu \tag{1.9a}$$

$$SS2: var(y_\tau) = var(y_s) = \sigma_y^2 \tag{1.9b}$$

$$SS3: cov(y_\tau, y_{\tau+k}) = cov(y_s, y_{s+k}) \tag{1.9c}$$

The moments in SS1–SS3 are assumed to exist. The first condition states that the mean is constant, the second that the variance is constant and the third that the k-th order autocovariance is invariant to an arbitrary shift in the time origin. The extension to continuous time is straightforward, replacing y_τ by $y(\tau)$, and so on. From these three conditions, it is evident that a stochastic process could fail to be weakly stationary because its mean is changing; and/or its variance is changing; and/or the k-th order autocovariances depend on time for some k.

A stochastic process that is not stationary is said to be nonstationary. A nonstationary process could be: nonstationary in the mean; nonstationary in the variance; and/or nonstatonary in the autocovariances. Usually it is apparent from the context whether the stationarity being referred to is strict or weak. When the word 'stationary' is used without qualification, it is taken to refer to weak stationarity, shortened to WS, but, perhaps, most frequently referred to as covariance stationarity. (Weak or covariance stationarity is also referred to as wide-sense stationary, leading to the initials WSS.)

Two particular cases of interest relate to difference stationarity and trend stationarity, generally referred to as DS and TS, respectively. A process that is DS is nonstationary in the levels of its component random variables, but stationary in their first differences. Thus, the stochastic process Y is DS if:

$$Y = (y_t(\omega) : t \in \Theta, \omega \in \Omega) \qquad \text{nonstationary process}$$

$$DY = (\Delta y_t(\omega) : t \in \Theta^* \subseteq \Theta, \omega \in \Omega) \quad \text{stationary process}$$

The random walk $y_t = y_{t-1} + \varepsilon_t$ is an example of a DS process for, as elaborated below, neither its variance nor its autocovariances satisfy conditions SS2 and SS3; however, $\Delta y_t = \varepsilon_t$ satisfies these conditions for $t \in \Theta^* \subseteq \Theta$. The nonstationarity in Y is due to the implied accumulation of shocks, which is evident from the representation in Equation (1.2).

A TS process is one that is stationary after the removal of a deterministic trend. Typically, a linear trend is assumed for the generating process, thus observations that are generated from such a process will tend to have the direction given by the sign of the trend implying $E(y_t) \neq$ constant, and will not, therefore, satisfy SS1, even though they may satisfy SS2 and SS3. However, removal of the trend gives the detrended series, which will be stationary.

$$
\begin{aligned}
Y &= (y_t(\omega) : t \in \Theta, \omega \in \Omega) && \text{nonstationary process} \\
\tilde{Y} &= (\tilde{y}_t(\omega) : t \in \Theta, \omega \in \Omega) && \text{stationary process} \\
\tilde{y}_t(\omega) &\equiv y_t(\omega) - \mu_t && \text{detrended observation}
\end{aligned}
$$

The deterministic components are captured by the term μ_t; for example, $\mu_t = \beta_0 + \beta_1 t$, so that $\tilde{y}_t(\omega)$ is the detrended data for period t. For practical applications, an estimate of μ_t will be required.

The next two subsections show that the random walk is not a stationary process by virtue of an increasing variance and autocovariances that are not invariant to a translation of the time index.

1.1.4.iii The variance of a random walk increases over time

One of the problems for an intuitive understanding of the behaviour of a random walk sample path, is that the variance of y_t is not constant; indeed, it increases linearly with t – this means that the range of y_t increases with t. This characteristic reflects the lack of stationarity of the distribution of y_t as t varies. The variance of y_t is as follows:

$$
\begin{aligned}
\operatorname{var}(y_t) &= \sum_{j=1}^{t} \operatorname{var}(\varepsilon_j) + 2\sum_{i=1}^{t}\sum_{j>i}^{t} \operatorname{cov}(\varepsilon_i, \varepsilon_j) && \text{(1.10a)} \\
&= \sum_{j=1}^{t} \operatorname{var}(\varepsilon_j) \quad \text{as } \operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j && \text{(1.10b)} \\
&= t\sigma_\varepsilon^2 \qquad\qquad \text{as } \operatorname{var}(\varepsilon_j) = \sigma_\varepsilon^2 \text{ for all } j && \text{(1.10c)}
\end{aligned}
$$

In the case of the symmetric, binomial random walk $\sigma_\varepsilon^2 = 1$ and $E(y_t) = 0$ for all t; but the variance increases with t, such that the variance of y_1 is 1, the variance of y_{100} is 100 and the variance of y_{500} is 500. Note that provided $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$ and $\operatorname{var}(\varepsilon_j) = \sigma_\varepsilon^2$ for all j, then, holds, so the result would also hold for $\varepsilon_t \sim \text{niid}(0, \sigma_\varepsilon^2)$ or, weaker still, that ε_t is white noise.

1.1.4.iv The autocovariances of a random walk are not constant

The k-th order autocovariance, $\gamma(k)$, is a measure of the (linear) dependence between y_t and its k-th lag, y_{t-k}, equivalently, the k-th lead, if the process generating the data is covariance stationary. $\gamma(k)$ is defined as follows:

$$\begin{aligned}\gamma(k) &= E[y_t - E(y_t)][(y_{t-k} - E(y_{t-k})] \quad k = \pm 1, 2, 3, \ldots \\ &= \text{cov}(y_t, y_{t-k})\end{aligned} \tag{1.11}$$

Clearly, the expectations in (1.11) must exist for $\gamma(k)$ to be defined. The variance $\gamma(0)$ is given by setting $k = 0$:

$$\gamma(0) = E[y_t - E(y_t)]^2 \tag{1.12}$$

For a stationary process, the k-th order autocorrelation coefficient $\rho(k)$ is $\gamma(k)$ scaled by the variance, $\gamma(0)$ (which is constant on this assumption), so that:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \tag{1.13}$$

The scaling ensures that $0 \leq |\rho_k| \leq 1$. Considered as a function of k, $\gamma(k)$ and $\rho(k)$ give rise to the autocovariance and autocorrelation functions; the latter portrayed graphically, with k on the horizontal axis and ρ_k on the vertical axis, is referred to as the correlogram. (See equation (1.16) for an adjustment to $\rho(k)$ for a nonstationary process.)

Covariance (or second-order) stationarity requires that the $\gamma(k)$ should be invariant to a translation of the time index, provided that a distance of k periods is maintained between the random variables. This is not the case for a random walk. To illustrate, assume for simplicity that y_0 is a fixed constant, so that $\text{var}(y_0) = 0$, then $\gamma(1)$ for $t = 2$ and $t = 3$, are, respectively, as follows:

$$\begin{aligned}\text{cov}(y_1, y_2) &= \text{cov}(\varepsilon_1, \varepsilon_1 + \varepsilon_2) \\ &= \text{var}(\varepsilon_1) + \text{cov}(\varepsilon_1, \varepsilon_2) \\ &= \sigma_\varepsilon^2 \\ \text{cov}(y_2, y_3) &= \text{cov}(\varepsilon_1 + \varepsilon_2, \varepsilon_1 + \varepsilon_2 + \varepsilon_3) \\ &= \text{var}(\varepsilon_1) + \text{var}(\varepsilon_2) + 2\text{cov}(\varepsilon_1, \varepsilon_2) + \text{cov}(\varepsilon_1, \varepsilon_3) + \text{cov}(\varepsilon_2, \varepsilon_3) \\ &= 2\sigma_\varepsilon^2\end{aligned}$$

Hence $\text{cov}(y_1, y_2) \neq \text{cov}(y_2, y_3)$ although both relate to an index value $k = 1$. These derivations exploit the properties of white noise $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ for all t, and

$E(\varepsilon_t\varepsilon_s)=0$ for $t \neq s$. In general, $\gamma(1)$ for arbitrary t is given by:

$$\begin{aligned} \text{cov}(y_{t-1}, y_t) &= \text{cov}(\textstyle\sum_{j=1}^{t-1}\varepsilon_j, \sum_{j=1}^{t}\varepsilon_j) \\ &= (t-1)\sigma_\varepsilon^2 \end{aligned} \tag{1.14}$$

Hence, $\gamma(1)$ varies as the time index varies, increasing linearly with t. This result generalises to $\gamma(k)$, so that

$$\begin{aligned} \text{cov}(y_{t-k}, y_t) &= \text{cov}(\textstyle\sum_{j=1}^{t-k}\varepsilon_j, \sum_{j=1}^{t}\varepsilon_j) \\ &= (t-k)\sigma_\varepsilon^2 \end{aligned} \tag{1.15}$$

As noted above, the k-th order autocorrelation coefficient, $\rho(k)$, is the standardised, or scaled, k-th order autocovariance. If the sequence $\{y_t\}_{t=1}^{T}$ is stationary, such that (inter alia) $\text{var}(y_{t-k})=\text{var}(y_t)=\gamma(0)$, for all t given k, then the appropriate scaling is $\gamma(0)$. However, in the nonstationary case, such as the random walk of Equation (1.1), $\text{var}(y_{t-k}) \neq \text{var}(y_t)$, leading to the following variation:

$$\begin{aligned} \rho(k) &= \frac{\text{cov}(y_{t-k}, y_t)}{\sqrt{\text{var}(y_{t-k})\text{var}(y_t)}} \\ &= \frac{(t-k)\sigma_\varepsilon^2}{\sqrt{((t-k)\sigma_\varepsilon^2)(t\sigma_\varepsilon^2)}} \\ &= \sqrt{(1-k/t)} \end{aligned} \tag{1.16}$$

For finite t, $\rho(k)$ depends on t for a given k and is not, therefore, invariant to the time index t. Note that $\rho(k) \to 1$ as $t \to \infty$.

1.1.5 A simple random walk with Gaussian inputs

An obvious extension of the symmetric random walk is to generate the stochastic inputs as draws from a normal distribution or some other symmetric continuous distribution. This gives a smother pattern to the sample paths, but otherwise replicates the pattern of long sojourns of the paths in one half or the other. This is illustrated in Figure 1.3 where $\varepsilon_t \sim$ niid(0, 1), but otherwise the details are as for Figure 1.1. In the case of Gaussian inputs, $E(V_T)=0.6363\sqrt{T}$ compared to $E(V_T)=0.7979\sqrt{T}$ for the binomial inputs. For example, if $T=500$ then the integer part of $E(V_T)$ is 14, compared to 17 for binomial inputs.

1.1.6 Variations on the simple random walk

There are several interesting variations on the basic or 'pure' random walk of $y_t=y_{t-1}+\varepsilon_t$. One of the most useful imparts a direction to the random walk, which can be done in one of two ways. First, the random walk can be made asymmetric. This is very simple to do in the case of binomial inputs, and corresponds to $p \neq q$; for example, to continue the gambling example, suppose

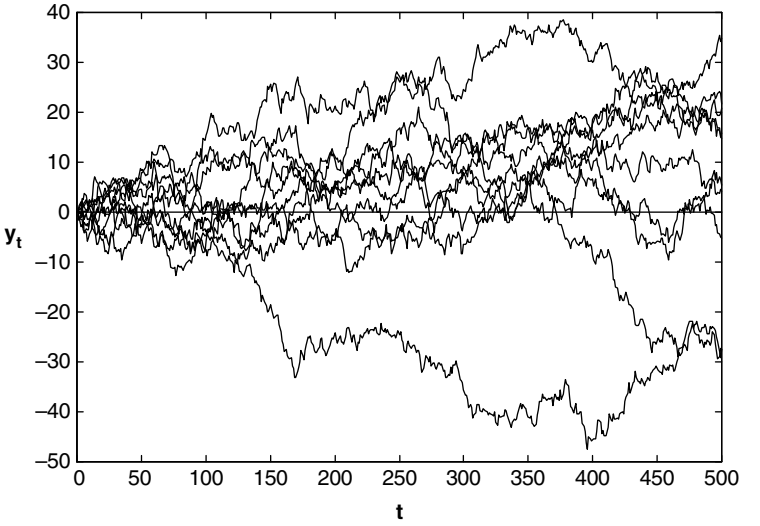


Figure 1.3 Sample paths of a Gaussian random walk.

that $p > q$, then this will impart a positive direction to the walk. To illustrate, the simulations underlying Figure 1.1 were repeated but with $p=0.55$, 0.6, 0.65 and 0.7, with the results shown in Figure 1.4. Even in the case of $p=0.55$, this change is sufficient to make the walk almost entirely positive, and as p increases further the walk has a clear positive direction.

The second and perhaps more familiar way to impart a direction to the random walk is to introduce 'drift', so that the random walk becomes:

$$y_t = \mu + y_{t-1} + \varepsilon_t \tag{1.17}$$

Thus, ceteris paribus, the increment/decrement to the random walk each period is μ, and the sign of μ will determine the direction of the drift subject to the realisation of ε_t. As in (1.2), by repeated back substitution y_t can be expressed as y_0 plus the cumulated stochastic inputs, but in this case there is an additional deterministic time trend due to the accumulation of drift:

$$y_t = y_0 + \mu t + \sum_{j=1}^{t} \varepsilon_j \tag{1.18}$$

The direction to the sample path of y_t is imparted by the term μt, and the random walk generated by the cumulated sum of shocks will, depending on the sign and magnitude of μ, tend to be observed mostly on either the positive or the negative side of the zero axis. In the context of a gambling game $\mu > 0$ could be the cost per play if the random walk is viewed from the banker's perspective, whereas it is the negative of the cost per play if the random walk is from the gambler's perspective.

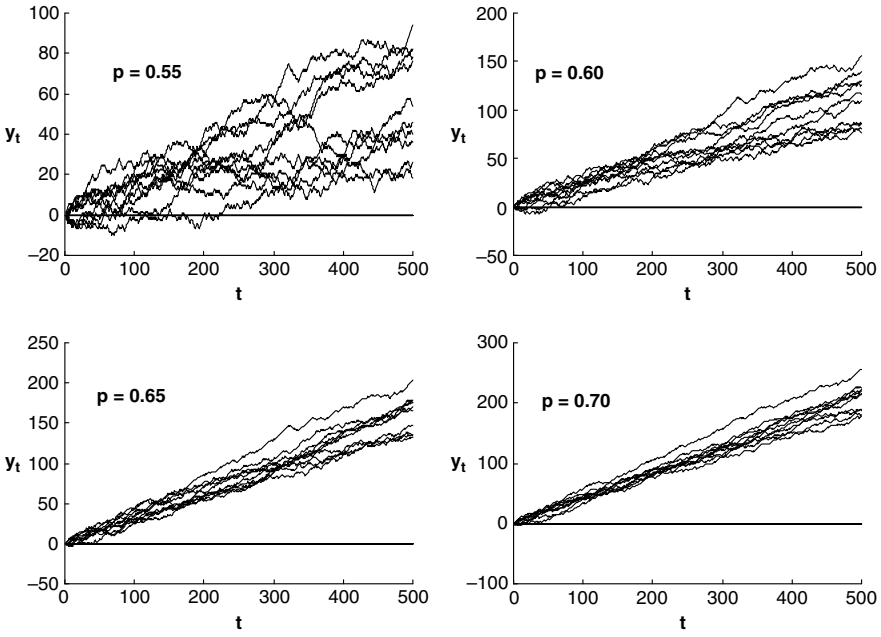


Figure 1.4 Sample paths of asymmetric binomial random walks.

The drifted random walk is important because it is a possible characterisation of economic time series that inherently have a direction, as is usually the case with macroeconomic aggregates such as the expenditure components of GNP, employment and price indices. It offers an explanation that is alternative to serially correlated deviations about a deterministic time trend. To illustrate, ten sample paths are shown in Figure 1.5 for a symmetric binomial random walk with $\mu = 0.05$, 0.1, 0.15, 0.2. As the standard deviation of ε_t is unity, the drift coefficient is in units of σ_ε. The positive drift to the random walk becomes clearer as μ increases.

1.1.7 An empirical illustration

To illustrate random walk-like behaviour in a real time series, we consider the exchange rate of the Swiss Franc (SWFR) against the UK £, with $T = 7{,}347$ daily observations from 2 January 1980. The data are graphed in Figure 1.6, with the mean of 2.59 superimposed on the figure. Note that the time axis has been scaled so that its range is from 0 to 1; in effect, each time division is represented as $1/T$ units of time. There are just 39 crossings of the sample mean during the observation period compared to an expected number of 68 for a sample of this size generated by a random walk with Gaussian inputs. The last 1,000 observations are plotted in Figure 1.7 as a scatter graph of y_t and y_{t-1}; this figure

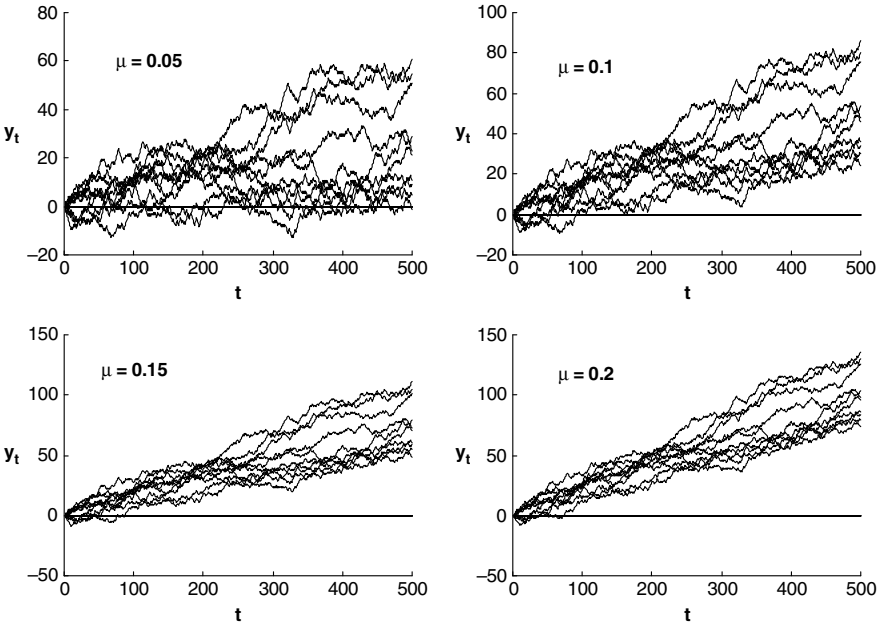


Figure 1.5 Sample paths of drifted symmetric binomial random walks.

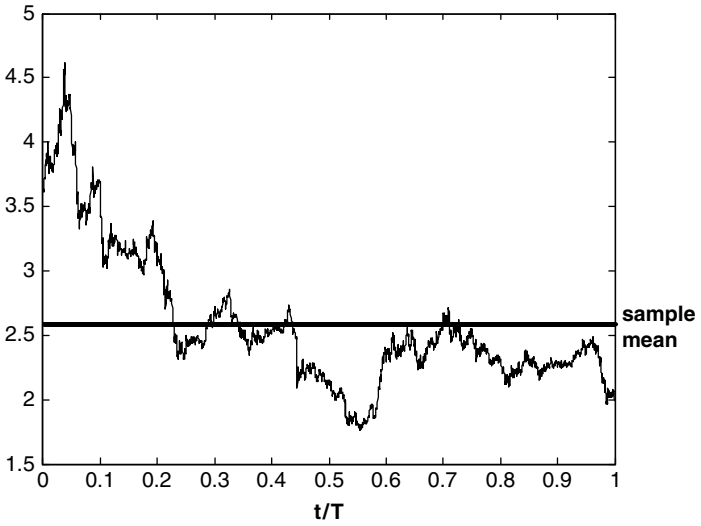


Figure 1.6 Exchange rate (daily), SWFR:£.

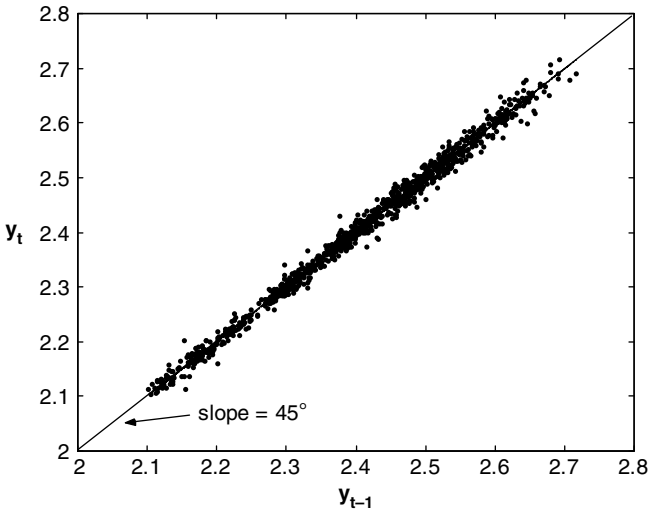


Figure 1.7 Scatter graph of daily, SWFR:£.

show the observations clustered around a line with a slope of 45°, which is suggestive of a random walk, although more formally testing would be required to assess this hypothesis.

1.1.8 Finer divisions within a fixed interval: towards Brownian motion

The next step in terms of obtaining the limiting process is to fix the length of the time interval, and then divide it into smaller and smaller parts, so that in the limit as the size of the divisions tends to zero, the random walk becomes a continuous process. The random walk is then defined on an interval on the real line with range zero to unity.

The length of the walk T is fixed and then divided into N small time steps, so that $\Delta t = T/N$; N is then allowed to increase, so that the time divisions approach 0. There is no loss in setting $T = 1$ and, therefore, $\Delta t = 1/N$. Within the unit interval an individual instant of time is denoted t_j, which satisfies $t_j = t_{j-1} + \Delta t$, so that $\Theta = [t_0 = 0, t_1, \ldots, t_{N-1}, t_N = 1]$, where $t_j = j/N$. The other parameter in the random walk is the size of each step, or win/loss amount in a gamble, which is taken to be $\Delta y_t = (\sqrt{\Delta t})\varepsilon_t$. The variance of Δy_t is therefore $(\Delta t)\sigma_\varepsilon^2$, a choice which ensures that if $\Delta t = 1$ then the step size is ε_t, as in the standard random walk.

The random walk is now:

$$y_{t_j} = y_{t_{j-1}} + (\sqrt{\Delta t})\varepsilon_t \tag{1.19}$$

The conditional variance of y_{t_j} is $\text{var}(y_{t_j} | y_{t_{j-1}}) = \Delta t \sigma_\varepsilon^2$, whereas the unconditional variance of y_{t_j} is $\text{var}(y_{t_j}) = t_j \sigma_\varepsilon^2$; and if $\sigma_\varepsilon^2 = 1$, then $t_j \sigma_\varepsilon^2 = t_j$.

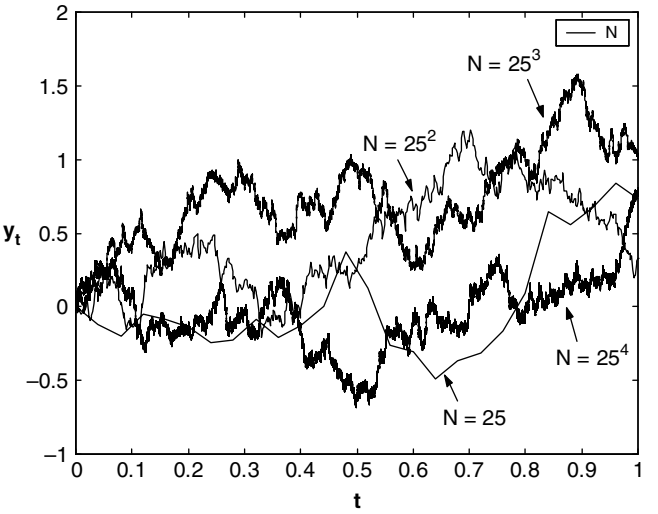


Figure 1.8 Gaussian random walk as time divisions approach zero.

To illustrate the sample paths as the time division tend to zero, the Gaussian random walk was simulated over the unit interval, so that $t \in \Theta = [0, 1]$, $\varepsilon_t \sim N(0, 1)$, $T = 1$ and $N = 25^h$, for $h = 1, 2, 3, 4$; with these values the unit interval is first divided into 25 equal parts and finally into 390,625 equal parts, so that the grid of time divisions is at first very coarse, but becomes finer and finer as N increases. The resulting sample paths are shown in Figure 1.8.

An interesting question is whether the sample paths generated as $\Delta t \to 0$ have any characteristics of interest. The answer is yes, but we first need a limiting result. Define a scaled version of y_{t_j} as follows:

$$Y_{t_j} \equiv \frac{y_{t_j}}{\sigma_\varepsilon \sqrt{N}} \tag{1.20}$$

If $\varepsilon_t \sim$ iid$(0, \sigma_\varepsilon^2)$ and Y_{t_j} is generated as in (1.14) then as $N \to \infty$, with T fixed, so that $\Delta t \to 0$, it follows that:

$$\begin{aligned} Y_{t_j} &\Rightarrow_D N(0, t_j) \\ &= \sqrt{t_j} N(0, 1) \end{aligned} \tag{1.21}$$

This result follows by application of the standard central limit theorem (CLT) and is an example of an invariance principle in the sense that although ε_t is not necessarily normally distributed, in the limit, as $N \to \infty$, a suitably scaled version of y_{t_j} is normally distributed. In fact, the assumption that $\varepsilon_t \sim$ iid$(0, \sigma_\varepsilon^2)$ is sufficient rather than necessary for (1.21) to hold. The CLT still goes through

if $\{\varepsilon_t\}_{t=1}^{T}$ is a martingale difference sequence (see Billingsley, 1995, p.475; and for generalisations and references, the interested reader is referred to Merlevède, Peligrad and Utev, 2006; Ibragimov and Linnik, 1971).

Equation (1.21) states that y_{t_j} scaled by $\sigma_\varepsilon\sqrt{N}$ has a limiting normal distribution with variance $\text{var}(y_{t_j}) = t_j$; thus, dividing the scaled partial sum by $\sqrt{t_j}$ results in a random variable, denoted Z_{t_j}, which is distributed as $N(0, 1)$. In summary:

$$Z_{t_j} \equiv \frac{y_{t_j}}{\sigma_\varepsilon \sqrt{t_j}\sqrt{N}} \Rightarrow_D N(0,1) \tag{1.22}$$

These results, especially (1.22), and the division of a fixed time interval into smaller and smaller parts, lead naturally to the concept of Brownian motion, which is considered next.

1.2 Definition of Brownian motion

The stochastic process W(t) defined in continuous time is said to be a Brownian motion (BM) process if the following three conditions are met:

BM1: $W(0) = 0$.
BM2: the increments are independent and stationary over time.
BM3: $W(t) \sim N(0, t\sigma^2)$; that is W(t) is normally distributed with mean zero and variance $t\sigma^2$.

W(t) is a standard Brownian motion process if $\sigma^2 = 1$, when it will be denoted B(t). If $\sigma^2 \neq 1$ and $W(0) = 0$, then $B(t) = W(t)/\sigma$ converts the process to have a unit variance and become standard BM. If $W(0) = \mu \neq 0$, and $\sigma^2 \neq 1$, then $B(t) = [W(t) - \mu]/\sigma$ is standard BM, which implies $W(t) = \mu + \sigma B(t)$. A trended BM is obtained if $W(t) = \beta t + \sigma B(t)$, so that $B(t) = [W(t) - \beta t]/\sigma$ is standard BM. In the case of standard BM, BM3 above is replaced by $B(t) \sim N(0, t)$. Given BM2 and BM3, and assume that we are dealing with standard BM, then two related results are of interest.

First, the difference between BM at times t and s is normally distributed, thus:

$$\begin{aligned} B(t) - B(s) &\sim N(0, t-s) \\ &= \sqrt{(t-s)}N(0,1) \end{aligned} \tag{1.23}$$

where $0 \leq s < t$; this says that the increment of (standard) BM over the interval $t - s$ is normally distributed with zero mean and variance $t - s$. A consequence of this result is that letting Δt be an increment of time, then:

$$B(t + \Delta t) - B(t) \sim \sqrt{\Delta t}N(0,1) \tag{1.24}$$

The connection between the scaled random walk of equation (1.19) and BM should now be evident: the random walk is specified in discrete time and if

$\varepsilon_t \sim$ iid$(0, \sigma_\varepsilon^2)$, but is not niid, then Y_{t_j} is approximately normally distributed for finite N; BM is specified in continuous time and W(t) and B(t) are exactly normally distributed. Both of these differences disappear in the limit as $\Delta t \to 0$ a result formalised below as the functional central limit theorem (FCLT). Referring back to Figure 1.8, which graphs some sample paths of a random walk with increasingly fine time divisions, $\Delta t \to 0$, the last of the sub-figures has $\Delta t = 1/25^4 = 0.00000256$, and thus this could equally be taken to illustrate some sample paths of BM.

BM provides a mathematical model of the diffusion, or motion over time, of erratic particles; for example, the biologist Robert Brown's original observation in 1827 that pollen grains suspended in water exhibited a ceaseless erratic motion; being bombarded by water molecules, the pollen seemed to be the subject of myriad chance movements. A similar phenomenon can be observed with smoke particles colliding with air molecules. In both examples, the trajectory of the particle over any small period is spiky and seemingly chaotic, but observed over a longer period the particle traces out a smoother path that has local trends. In an economic context, it is evident that the behaviour of stock prices over time, particularly very short periods of time, can be quite erratic – or noisy; however, over a longer period, a direction is imparted to the level of the series.

BM is used to model these phenomena: at any one point, or over a small interval, the movement, as captured by the 'increments', is erratic and seemingly without structure, whereas over a longer period, the individual erratic movements are slight relative to the whole path. Hence a key element of BM is the way that the erratic increments are built up into the level of the series. Whilst BM specifies normal increments, it can be generalised to increments from other distributions, as might be appropriate for some financial asset prices, whose distributions exhibit much greater kurtosis than is found in a normal distribution.

1.3 Functional central limit theorem (FCLT)

A result of particular use in establishing the distribution of many unit root test statistics is the functional central limit theorem. Whereas the standard CLT applies to a suitably scaled random variable, the FCLT applies to a stochastic process, which defines a function rather than a single random variable. Below, for example, the simple random walk of length T, which is an example of a partial sum process, is written as a function of a variable r, such that $0 \leq r \leq 1$. Allowing r to vary over this range emphasises that the random walk is a random function of r; this is evident from, for example, Figure 1.8, which plots some sample paths, or trajectories, from a random walk – the whole paths generally

differ. As noted above, the device of letting $\Delta t \to 0$, used in plotting Figure 1.8, leads, in the limit, to Brownian motion.

Assume, for simplicity that $y_0 = 0$, then the simple random walk $y_t = \sum_{t=1}^{T} \varepsilon_t$, $t = 1, \ldots, T$, can be written equivalently as follows:

$$y_T(r) = \sum_{t=1}^{[rT]} \varepsilon_t \tag{1.25}$$

The notation $y_T(r)$ emphasises the fixed length T of the sequence and the functional dependence on r. The notation [rT] indicates the integer part of rT; thus rT is exactly an integer for $r = j/T$, $j = 1, \ldots, T$. (Note that $j = 0$ would follow if the lower limit of the summation in (1.25) was 0.) The virtue of (1.25) is that $y_T(r)$ can be considered as continuous function of r, albeit it will be a step function; however, the 'steps' become increasingly smaller as $T \to \infty$, so that, in the limit, $y_T(r)$ is a continuous function of r. To consider this limit, $y_T(r)$ is first normalised as follows:

$$Z_T(r) \equiv \frac{y_T(r)}{\sigma_\varepsilon \sqrt{T}} \tag{1.26}$$

Let $\varepsilon_t \sim$ iid$(0, \sigma_\varepsilon^2)$, with $\sigma_\varepsilon^2 < \infty$, then the FCLT states that:

$$Z_T(r) \Rightarrow_D B(r) \tag{1.27}$$

This is sometimes stated in slightly abbreviated form as $Z_T \Rightarrow_D B$ (or with a variant of the $\Rightarrow_D$ notation). Equation (1.27) states that a suitably normalised version of $y_T(r)$ converges to standard Brownian motion. If $y_T(r)$ is not normalised by σ_ε^2, that is, define, say, $v_T(r) \equiv y_T(r)/\sqrt{T}$, then:

$$v_T(r) \Rightarrow_D W(r) = \sigma_\varepsilon B(r) \tag{1.28}$$

The FCLT is another example of an invariance principle in that the convergence result is invariant to the distribution of the stochastic inputs that drive $y_T(r)$ and so $Z_T(r)$, in particular they do not have to be Gaussian. Of course, some assumptions have to be made about these inputs, but these assumptions, discussed below, are relatively weak, and the FCLT is simply extended to cover such cases. The nature of Brownian motion B(r) means that it is normally distributed for all r in its domain, its increments are normally distributed and it is jointly normally distributed for different values of r. The CLT is in this sense a by-product of the FCLT.

The notation $\Rightarrow_D$ is used here as it would be in the case of conventional asymptotic results, where it indicates convergence in distribution; however, here it refers here to the weak convergence of the probability measures (see Billingsley, 1995; Davidson, 1994, for more detail). The latter is more encompassing than convergence in distribution, which just compares the distribution of one random variable with another. In effect, the convergence relates to the convergence of one stochastic process to another, rather than of a single random variable to another.

1.4 Continuous mapping theorem (CMT)

The FCLT is often used in combination with the CMT applied to function spaces to establish distributional results for unit root test statistics. We first state the CMT and then its extension to function spaces.

Consider the random variable x_T and the continuous function $f(.)$, then the CMT states that:

$$\text{if } x_T \Rightarrow_D x \quad \text{and} \quad P(x \in D_g) = 0, \quad \text{then} \quad f(x_T) \Rightarrow_D f(x) \tag{1.29}$$

where D_g is the set of discontinuity points of $f(x)$ and $P(.)$ indicates probability, (for an elaboration, see Billingsley, 1995, Theorem 25.7; Davidson, 1994, Theorem 22.11, 2000, Theorem 3.1.3). A familiar case from elementary texts is when $x_T \Rightarrow_D x \sim N(0, 1)$ and $f(x) = x^2$, then $f(x) \Rightarrow_D \chi^2(1)$; thus, if x_T is asymptotically distributed as standard normal, then x_T^2 is asymptotically $\chi^2(1)$. An example is provided by the standard regression t test, which has a small sample 't' distribution but converges in distribution to $N(0, 1)$, thus its square converges to $\chi^2(1)$.

The next step is to extend the CMT to functionals, that is functions of stochastic processes which are themselves functions; in this case, interest is in a function of a stochastic process, $Z_T(r)$, such as $g(Z_T(r)) = Z_T(r)^2$, where $g(.)$ is a continuous mapping, apart from a set of measure zero. The (extended) CMT for functionals of the stochastic process $Z_T(r)$ is as follows (where D is the domain of the argument of $g(.)$).

Let $g(.)$ be a functional that maps D to the real line, $g: D \mapsto \Re$, and which is continuous apart from a set of measure zero, if $Z_T(r) \Rightarrow_D B(r)$, then:

$$g(Z_T(r)) \Rightarrow_D g(B(r)) \tag{1.30}$$

An application of the extended CMT for $g(Z_T(r)) = Z_T(r)^2$ yields the following: if $Z_T(r) \Rightarrow_D B(r)$, then $Z_T(r)^2 \Rightarrow_D B(r)^2$.

An application of the extended CMT, of particular importance in unit root tests, relates to the least squares (LS) estimator in the AR(1) model that nests the simple random walk. Consider estimating the following:

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{1.31}$$

where $\{\varepsilon_t\}_{t=1}^T$ is assumed to be a sequence of iid random variables, with zero mean and constant variance, written $\varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2)$. Clearly, $\rho = 1$ corresponds to the simple random walk of Equation (1.1), so that a natural hypothesis testing approach is to set H_0: $\rho = 1$ against H_A: $|\rho| < 1$. In the context of (1.31) this is *the* unit root hypothesis, of which more in the next section and in particular a rationalisation of the word 'root' in this context. In this framework, equation (1.31) is the hypothesised data-generating process and the maintained regression, but more generally these may differ. Implicit in this set-up is that y_0 is a

starting value and is either a constant – for example, $y_0 = 0$ – or a draw from a distribution with a finite variance.

One possible test statistic, suggested by Dickey and Fuller (see Fuller, 1976), is $\hat{\delta} \equiv T(\hat{\rho} - 1)$, where $\hat{\rho}$ is the LS estimator of ρ, large negative values of which will be inconsistent with H_0. The quantities needed to construct $\hat{\delta}$ are given as follows:

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \tag{1.32}$$

$$\hat{\rho} - 1 = \frac{\sum_{t=1}^{T} y_{t-1}(y_t - y_{t-1})}{\sum_{t=1}^{T} y_{t-1}^2} \tag{1.33}$$

$$= \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2} \qquad \text{using } y_t - y_{t-1} = \varepsilon_t$$

$$\hat{\delta} \equiv T(\hat{\rho} - 1)$$

$$= T\frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$= \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t / T}{\sum_{t=1}^{T} y_{t-1}^2 / T^2} \tag{1.34}$$

Hypothesis testing requires the limiting distribution of $\hat{\delta}$, which is the ratio of two quantities whose limiting distributions are known (see, for example, Phillips, 1987; Banerjee et al., 1993; Patterson, 2010). In particular:

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \Rightarrow_D \int_0^1 W(r)dW(r) = \frac{1}{2}\sigma_\varepsilon^2[B(1)^2 - 1] \tag{1.35}$$

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \Rightarrow_D \int_0^1 W(r)^2 dr = \sigma_\varepsilon^2 \int_0^1 B(r)^2 dr \tag{1.36}$$

The limiting distribution, $F(\hat{\delta})$, then follows from the extended CMT so that:

$$\hat{\delta} \Rightarrow_D \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr} \equiv F(\hat{\delta}) \tag{1.37}$$

$$= \frac{1}{2}\frac{[B(1)^2 - 1]}{\int_0^1 B(r)^2 dr}$$

The second line uses (1.35) and (1.36), so that σ_ε^2 cancels from the numerator and the denominator.

Note that equations (1.35), (1.36) and (1.37) involve integrals of Brownian motion; however, these are not integrals in the standard sense of Reimann or Reimann-Stieltjes integrals. Indeed, whilst BM is continuous it is nowhere differentiable and so these integrals do not exist. Rather, the integrals are defined

according to the Itô calculus. To cover this topic here would require quite a lengthy digression; for an excellent introduction to the topic, the interested reader is referred to Mikosch (1998); and for an introduction tailored to unit root tests, see Patterson (2010).

1.5 Background to unit root and related tests

The previous section introduced one of the 'family' of Dickey-Fuller (DF) test statistics for a unit root, which is just one of many tests for a unit root. Much has been written about the 'unit root' hypothesis, with a multiplicity of tests and a wide range of applications to be found in journal articles, textbooks and theses. In a selective survey published in 1990, Diebold and Nerlove (1990) noted then that 'The unit root literature is vast ... .' It is two decades since that survey, with no abatement in the interest in unit roots, and the topic in some form is still one of the key areas of interest in journal articles. Additionally, many econometric software packages, those available both commercially and academically, include at least one and usually more such test statistics in their programmed options, and the results of such tests are routinely computed for inclusion in undergraduate and graduate project work, including doctoral theses, and in journal articles. A section on 'unit root testing' is now close to compulsory in all but the most elementary of econometric courses and textbooks.

1.5.1 What is a unit root?

To gain some understanding of what is meant by a unit root, first consider the simplest case where a sequence of random variables $\{y_t\}_{t=1}^{T}$ is generated by an AR(1) model so that, as in Equation (1.31), $y_t = \rho y_{t-1} + \varepsilon_t$, with ε_t, $t = 1, \ldots, T$. If $\rho = 1$, then $y_t = y_{t-1} + \varepsilon_t$; that is, $\Delta y_t = \varepsilon_t$, where $\Delta y_t \equiv y_t - y_{t-1}$, and there is said to be a unit root, strictly in the generating process, but often loosely referred to as in y_t or in the time series of observations or realisations of y_t. The next section considers how to generalise this idea.

1.5.1.i Generalising the random walk

There is more than one way of representing the generalisation. In the first representation the AR(1) model is extended directly with further lags on y_t; for example, the AR(2) model is written as $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$. For consistency, the AR(1) model would then be written with the coefficient denoted ϕ_1 rather than ρ. The AR(2) model could, potentially, have a single unit root, which corresponds to $H_0: \phi_1 + \phi_2 = 1$, or two unit roots, which corresponds to $H_0: \phi_1 = 2$ and $\phi_2 = -1$. In the latter case the model can be written as $\Delta^2 y_t = \varepsilon_t$, so that $\Delta y_t = \Delta y_{t-1} + \varepsilon_t$, where $\Delta^2 y_t \equiv \Delta y_t - \Delta y_{t-1}$. In this specification, there is a unit root in the first differences, which necessarily already have a unit root.

An alternative way of writing an AR model, which at the same time allows a simple but important generalisation, is to adopt a common factor interpretation (see Chapter 3, section **3.8.1.ii**), in which the possible unit root is isolated, with any other dynamic effects originating from the error term. This model is, as is the simple AR model, more easily written with the use of the lag operator and lag polynomial, considered in detail in Appendix 2, introduced briefly here with what is sufficient for present purposes.

First define the lag operator L, such that when applied to the variable y_t, it induces the j-th order lag, that is $L^j y_t \equiv y_{t-j}$; if $j < 0$, then $L^j y_t$ is a lead. The lag operator can be used to define a model with lags, such as an AR model or, when combined with a moving average (MA) error, an ARMA (autoregressive moving average) model. For example, the AR(2) model can be written as $(1 - \phi_1 L - \phi_2 L^2)y_t = \varepsilon_t$; and the ARMA(2, 1) model, which is of order 2 in its AR component and 1 in its MA component, is written as $(1 - \phi_1 L - \phi_2 L^2)y_t = (1 + \theta_1 L)\varepsilon_t$. In general the ARMA(p, q) model is represented compactly as follows:

$$\phi(L)(y_t - \mu_t) = \theta(L)\varepsilon_t \tag{1.38a}$$

$$\phi(L) = 1 - \sum_{j=1}^{p} \phi_i L^i \tag{1.38b}$$

$$\theta(L) = 1 + \sum_{j=1}^{q} \theta_j L^j \tag{1.38c}$$

where μ_t comprises deterministic terms; for example, a constant or a constant and a linear trend, so that y_t is adjusted for a nonzero long-run (deterministic) component by subtracting μ_t.

With the benefit of the lag operator and lag polynomial, the common factor form of the model is written as follows:

$$y_t = \mu_t + z_t \tag{1.39a}$$

$$\phi(L)z_t = \theta(L)\varepsilon_t \tag{1.39b}$$

$$\phi(L) = (1 - \rho L)\varphi(L) \tag{1.39c}$$

From the perspective of this model, the AR polynomial $\phi(L)$ has been factored as the product of two polynomials $(1 - \rho L)$ and $\varphi(L)$, one of which is a first-order polynomial, with a unit root if $\rho = 1$, and the other is a polynomial of one lower order than $\phi(L)$. The unit root null hypothesis can then always be expressed as $H_0: \rho = 1$, whatever the order of $\phi(L)$. Contrast this with H_0 in the direct ARMA(p, q) model, which is $\sum_{j=1}^{p} \phi_j = 1$.

At this point note a convention that should be borne in mind (it is elaborated on in Appendix 2), which can be briefly illustrated with the model $(1 - \rho L)y_t = \varepsilon_t$. The definition adopted here is that the root of the lag polynomial $(1 - \rho L)$ is $\delta_1 = \rho^{-1}$, with reciprocal $\delta_1^{-1} = \rho$; hence, strictly, it is δ_1 not ρ that is the root. When $\rho = 1$ there is no contradiction in referring to ρ as *the* root, since $\delta_1 = 1$

also. When $\rho < 1$ but close to 1, this situation is usually referred to as a 'near'-unit root; this is correct terminology and a usage we follow, but to be precise (on the definition adopted here), the root in such a case is $\delta_1 = \rho^{-1} > 1$.

The existence of a unit root or roots generates a nonstationary process; that is, the probability structure is not constant over time. For example, in the first example above, y_t has a variance that increases linearly with time and autocovariances and autocorrelations that depend on t (that is they are not invariant to a translation of the time axis); nevertheless, taking the first difference results in a stationary process. These properties were demonstrated in sections **1.1.4.iii** and **1.1.4.iv**.

1.5.1.ii Integrated of order d: the I(d) notation

The idea that there are some nonstationary stochastic processes that can be made stationary by applying the differencing operator $\Delta \equiv (1 - L)$ to the component random variables a sufficient number of times leads to a commonly adopted definition of an I(d) series. The following definition was suggested by Engle and Granger (EG) (1987): 'A series with no deterministic component which has a stationary, invertible, ARMA representation after differencing d times, is said to be integrated of order d, denoted $y_t \sim$ I(d).' (EG used the notation x_t, whereas here y_t is used throughout.) The reader is very likely to have encountered expressions such as $y_t \sim$ I(1) or $y_t \sim$ I(0). (For a detailed discussion of what constitutes an I(0) series, see Davidson, 2009.) Although EG focus on the integer cases $d = 0$ and $d = 1$, they note that their definition applies to fractional d. One could add, as a clarification, that d is the minimum number of differences needed to ensure stationarity.

Of particular importance in empirical work is being able to distinguish between I(0) and I(1) series, and five properties of interest, due to Engle and Granger (1987), are summarised in Table 1.1.

Given the critical nature of the differences between I(0) and I(1) series, and more generally, I(d) series with $d \geq 1$, it is perhaps not a surprise that a number of tests have been suggested with this aim in mind. Most of the tests take the null hypothesis as $y_t \sim$ I(d) with the alternative hypothesis as $y_t \sim$ I(d – 1), the most frequent case being $d = 1$. However, it is also possible to reverse these roles so that the null hypothesis is $y_t \sim$ I(d – 1) and the alternative is $y_t \sim$ I(d). (More precisely, one should refer to the data-generating process as generating series, or observations, that are I(d).) A brief development of 'unit root' tests is considered in the next section.

1.5.2 The development of unit root tests

In the applications including and immediately following the seminal contribution by Nelson and Plosser (1982), the unit root test statistics were usually those due to Dickey and Fuller (see Fuller, 1976), which have come to be known

Table 1.1 Contrasting properties of I(0) and I(1) series.

	I(0)	I(1)
variance, var(y_t)	var(y_t) is finite	var(y_t) $\to \infty$ as $t \to \infty$
autocorrelations, $\rho(k)$	invariant to translation of time axis $\rho(k) \to 0$ for k large enough; finite sum	not invariant to translation of time axis $\rho(k) \to 1$ for all k as $t \to \infty$; not summable
innovation effect	transient	permanent
spectrum, $f(\lambda)$, at zero, f(0)	$0 < f(0) < \infty$	$f(0) \to \infty$ as $\lambda \to 0$
expected time between crossings of E(y_t)	finite	infinite

Note: See Engle and Granger (1987); y_t is either I(0) with zero mean or I(1) with $y_0 = 0$.

as DF or, in their augmented form, ADF statistics. Subsequently, the Phillips and Perron (1988) developments of these statistics in a semi-parametric form, known collectively as PP tests, became popular and it was not unusual to see joint reporting of the ADF and PP tests. However, several other test statistics were suggested and the battery of such tests started to grow after Nelson and Plosser's article.

The development of further unit root test statistics continued in the 1990s, with a significant contribution by Elliott, Rothenberg and Stock (ERS, 1996) and an allied paper by Elliott (1999); both papers were available in discussion paper form several years before their publication dates. These articles noted that whilst it was not possible to obtain a single test statistic that was uniformly most powerful across the entire parameter space of interest, it was possible to develop a test statistic that was most powerful against a particular point in the parameter space, hence the terminology of a 'point-optimal' test. The problem with this approach was that it seemed to require an infinity of test statistics, one for each point in the parameter space under the alternative hypothesis. However, ERS were able to show that very little, if any, power was lost by, in effect, choosing just one value of the root in the stationary region as representative of all of those near to the unit root and then computing the test statistic using that value. Moreover, a variation of the approach led to the use of quasi-differenced data in standard tests such as the DF/ADF tests, so that ERS-type tests became easy to apply and popular – and familiar from the quasi-differencing approach to deal with weakly dependent errors in a regression model, as in the Cochrane-Orcutt procedure.

ERS-type tests joined the set of DF tests as those to which practitioners would most readily turn, partly because they were quickly incorporated into commercially available software. Indeed, they were often preferred because of their superior power under the assumptions of their derivation.

At the risk of simplification, the next important development came in exploiting the difference between the assumptions in ERS (1996) and Elliott (1999). In essence, the difference was quite simple: what was the nature of the starting point or initial condition in a time series process? For example, consider a trended series that is adequately modelled under the alternative hypothesis as stationary around a linear trend. Does it make much difference to the test results if the starting point is close to or far away from the trend (in some well-defined measure)? For example, consider one of the data sets used in this book comprising US industrial production with 1,044 observations for 1919m1 to 2005m12: would it matter if the observations were taken as starting in 1925, which was relatively close to trend, or in 1935, which was relatively a long way from a linear trend? With about 1,000 observations one might be tempted into thinking that this would not make a difference. However, it transpires that it does make a difference (both in this case and in general) and the test results are markedly affected by the relative scale of the initial observation. What this means is that it is possible for contradictory test results to be obtained on the same data set either by using the same test statistic but with starting points with different characteristics, or by using different test statistics with different characteristics but at the same starting point.

In a revival of use, it turns out that the DF tests, which were dominated by the ERS-type tests, in terms of power, for an initial observation that was close to trend (or a constant mean in the non-trended case), are more robust as the initial observation departs from trend.

Given the importance of calibrating the trend to this result, the question of the appropriate specification of the trend, an issue that had been largely sidelined, returns to have importance. However, to continue the précis of developments, the next task was to seek a unit root test statistic that was robust to the initial condition. One way of doing this was to combine test statistics with different characteristics. Simple linear combinations seem to work well, offering protection against an unknown initial condition at not too much cost in terms of power. Alternatively, as demonstrated by Elliott and Müller (2006), it is possible to construct a unit root test that is robust to the scale of the initial condition. For a discussion of current issues in unit root testing, including the specification of the trend, the role of initial conditions, see the special issue of *Econometric Theory* (2009), starting with Harvey et. al. (2009), and followed by a number of commentaries.

Of course, there are many variations and complications that occur in practice and which have attracted attention. Perhaps the simplest practical consideration arises where a time series has a seasonal pattern, which characterises many

production, employment and consumption series. It is not surprising, therefore, that this area attracted attention not long after the Nelson and Plosser article. In some ways it is more natural to use the seasonally unadjusted rather than the seasonally adjusted data, as the latter necessarily involve some assumptions about the form of the seasonality. There is the risk in the latter of inducing patterns that are not present in the unadjusted series. However, it soon became clear that the presence of a seasonal period allowed the possibility of unit roots at frequencies other than at the zero frequency associated with the conventional unit root tests, and thus the tests became apparently quite complicated due to the need to consider the possibility of unit roots occurring at different seasonal frequencies.

Another significant development that affected the course of unit root testing was to swap the null and alternative hypotheses, so that the null hypotheses became that of stationarity whilst the alternative became that of nonstationarity. This development was not entirely straightforward because the null hypothesis is not now that of a point but of a region in the parameter space. One of the key contributions was a test due to Kwiatkowski, Phillips, Schmidt and Shin (1992), referred to as the KPSS test. This test exploited the duality between a structural model of a time series and its reduced form to solve the testing problem. In the former, the time series is viewed as being built up from components; for example, an unobservable level plus an irregular component. In turn, a number of other stationarity tests were suggested, including those for seasonally unadjusted time series.

Notwithstanding the duality between confidence interval construction and hypothesis testing, the emphasis in much of the early empirical literature was on hypothesis testing, that is, coming to a decision with two possible outcomes: either to reject or not to reject the null hypothesis of a unit root. Of course, it was well known that this dichotomy was often too simple: surely a confidence interval would be more informative? To some extent reporting a p-value, as in the elementary textbook case, would help, but in fact it was rarely done, partly because the quantiles generally had to be obtained by simulation rather than reference to a standard set of tables.

Two developments eased this case. The hypothesis testing/confidence interval approaches were 're-connected'; for example, by Stock (1991), Hansen (1999) and Elliott and Stock (2001). Constructing a confidence interval by inverting a test statistic is familiar from the standard 't' statistic and the approach can be carried across to unit root tests. Indeed, as Elliott and Stock (2001) demonstrated, advantage can be taken of the more powerful unit root test statistics to invert one of these to get a shorter interval for a given confidence level.

Indeed, the circle was in a sense completed as a point-optimal test of the unit root hypothesis, along the lines of ERS (1996) and Elliott (1999), was equally a point-optimal test if the null and alternative hypotheses were swapped, so that

a point-optimal stationary test, but of a null close to the unit root, could be obtained by using the other tail of the corresponding unit root test.

The second development was encouraged by the increasing capacity of personal computers, so that large-scale simulations could be undertaken at little cost. This enabled bootstrapping to be applied to unit root tests and confidence interval construction, and a straightforward outcome of the former was the p-value associated with a particular sample value of a test statistic.

1.6 Are unit root processes of interest?

There have been a very large number of studies addressed to the issue of whether a particular series has been generated by a stochastic process with a unit root, and the question arises as to why there is such an (enduring) interest. This question is answered more fully in the next chapter, the present intention being to give an idea of some of the topics that have been studied. Nelson and Plosser (1982) considered 14 macroeconomic time series, such as GNP, industrial production, some price indices, and employment and bond yields. Subsequent research included a more detailed analysis of a number of these series, with particular interest focusing on aggregate measures of output, such as GDP or GNP, especially for industrialised countries (see, for example, Campbell and Mankiw, 1987a, 1987b; Cochrane, 1988; Rudebusch, 1992, 1993; Harvey, 1985). However, interest widened and many articles that involved the use of economic time series included a test of some form on the unit root hypothesis, in part because there was an underlying theoretical base for the distinction between unit root and non-unit root processes from an economic perspective (that is, it was not just a matter of the econometric aspects of the application). To give an idea of the underlying motivation, three areas of application are considered below.

1.6.1 Are there constant 'great ratios'?

An area of interest for the importance of unit roots relates to the implications of some growth models for the ratios of economic variables, sometimes referred to as the 'great ratios'. In a seminal article, Klein and Kosobud (KK) (1961), suggested five celebrated ratios of economics, namely the savings-income (or consumption-income) ratio, the capital-income ratio, labour's share of income, the income velocity of circulation and the capital-labour ratio (see also Kaldor, 1957), to which other, monetary ratios, such as the real money supply and the real interest rate (not strictly a ratio) have been added. KK constructed a small macroeconomic model which showed the connections between their five ratios. Later research developed the balanced growth implications of neoclassical growth models (see, for example, Brock and Mirman, 1972; King, Plosser and Rebelo, 1988a, 1988b; and King and Rebelo, 2000).

Of course, whether such constancy of the great ratios, which would anyway only be approximate, held empirically is another matter, and on examining these ratios for the US economy, by regressing them on a constant and a time trend, KK concluded that only one of the ratios, that for labour's share of income, could be considered approximately constant. On the other hand, at a descriptive level and with a more recent data set, King and Rebolo (2000) suggested that for US data the ratios of investment and labour income to output fluctuated about a constant mean, and whereas there was an upward trend for the ratio of consumption to income (since 1952), the trend was relatively slight.

We can interpret KK's interest in the possible existence of approximately constant ratios, where the individual components are themselves trended, in the following way. Consider two time series, each with a stochastic trend of the form generated by the accumulation of shocks, as in Equation (1.2); then, in general, the stochastic trends will be unrelated, so that the ratio of the two series, or the logarithmic difference, will also have a stochastic trend. The exception to this rule is when the stochastic trends are annihilated, resulting in a trendless ratio; in such a case the time series are said to have a common trend and are cointegrated. For example, consider consumption, c, and output, q, on a per capita basis, and suppose each of these to have a stochastic trend, but this trend is common to each variable, such that the log difference c – q; that is, the log of the consumption-output ratio, is trendless. Similarly, extending the analysis to include per capita investment, i, so that each of c, i and q, have a stochastic trend, but balanced growth implies there is a single stochastic trend, such that the log ratios c – q and i – q are trendless.

There are (at least) two ways to assess whether there is evidence to support the stability of the great ratios. The first is to consider each (log) ratio separately and carry out a unit root test (or swap the null hypothesis and carry out a stationarity test); non-rejection of the null hypothesis would then be evidence against the stability of the ratio. Tests of this kind have been reported by Harvey et al. (2003) for the G7 countries; earlier work includes Kunst and Neusser (1990). An alternative is to consider a system approach in which several series are jointly modelled, and then tested for the extent of cointegration in the system. Both the references cited above also use this approach; additionally, Mills (2001) extends the analysis to consider whether there are not only common trends but also common cycles.

1.6.2 Purchasing power parity

The theory of purchasing power parity (PPP) is fundamental to the theory of the real exchange rate and is a cornerstone of international economics. It is the macroeconomic analogue of the law of one price (LOOP). At the microeconomic level, the idea is that the price of a homogeneous good should be the same when converted to units of a common currency; in this case the nominal exchange

rate is regarded as a variable exogenous to the firm's decision, whereas at the macroeconomic level it is an endogenous variable determined by the ratio of (aggregate) price levels for the domestic and foreign economies.

To consider this aspect, the following notation is adopted. The nominal exchange rate E_t is defined as the domestic currency price of a unit of foreign currency; and the prices of a homogeneous good in domestic and foreign currencies are denoted $P^d_{g,t}$ and $P^f_{g,t}$, respectively. Then LOOP implies $P^d_{g,t} = E_t \times P^f_{g,t}$; that is, the domestic price equals the foreign price expressed in units of the domestic currency; expressing the price in the units of the foreign currency would give the same result. From the perspective of an individual producer, who has no market power, the right-hand-side variables are exogenous. In a perfect, frictionless market (without tariffs or transaction costs), setting $P^d_{g,t}$ to be greater than $E_t \times P^f_{g,t}$ means that the domestically produced good is not competitive with its foreign counterpart and will not attract any market share.

At the macroeconomic level, E_t is the endogenous variable, determined by the aggregation of individual market decisions, rather than exogenous as at the microeconomic level. Let P^d_t and P^f_t be suitably defined aggregate price indices for the domestic and foreign countries, respectively, then PPP states that the following relationship should hold:

$$E_t = A\left(\frac{P^d_t}{P^f_t}\right) \tag{1.40}$$

where A is a constant. In the simplest version of PPP, $A = 1$, but A differing from unity is permissible within the general theory; for example, $A \neq 1$ could arise from the use of different base years in the construction of the indices P^d_t and P^f_t. If $A = 1$, then PPP implies $P^d_t = E_t P^f_t$, so that the price levels are equalised in units of a common currency (here the domestic currency), which is the direct analogue of LOOP.

In a weaker version of PPP, the elasticity of the nominal exchange rate with respect to relative prices is allowed to differ from unity. That is:

$$E_t = A\left(\frac{P^d_t}{P^f_t}\right)^{\delta} \tag{1.41}$$

Allowing $\delta \neq 1$ is weaker in the sense that a 1% change in relative prices results in a δ% change in the nominal exchange rate, so that $P^d_t \neq E_t P^f_t$ even if $A = 1$. The weak form of PPP results in:

$$\begin{aligned} \ln E_t &= \ln A + \delta \ln\left(\frac{P^d_t}{P^f_t}\right) \\ &= \ln A + \delta(\ln P^d_t - \ln P^f_t) \end{aligned} \tag{1.42}$$

$$e_t = a + \delta(p^d_t - p^f_t) \tag{1.43}$$

where a lower-case letter denotes the logarithm of the corresponding upper-case variable. If $\delta = 1$, so that the elasticity of the nominal exchange rate with respect to relative prices is unity, then taking the term in relative prices to the left-hand side gives the log of the real exchange rate:

$$e_t - (p_t^d - p_t^f) = a \tag{1.44}$$

Thus, this equation states that the real exchange rate is constant. In terms of the level of the real exchange rate:

$$RE_t \equiv E_t \frac{P_t^f}{P_t^d} \tag{1.45a}$$

$$= A \tag{1.45b}$$

In practice, the real exchange rate is not expected to be constant, but rather stochastic and mean-reverting. For convenience of notation, define $re_t \equiv e_t - (p_t^f - p_t^d)$, then the stochastic version of this equation is:

$$re_t \equiv e_t - (p_t^f - p_t^d) \tag{1.46a}$$

$$= a + u_t \tag{1.46b}$$

where u_t is I(0), hence some dependency is allowed in the structure of u_t, but it must be weak in the sense of allowing the log of the real exchange rate to return to its mean given a shock; the lack of an immediate return to the PPP rate, following a shock, has been variously attributed to sticky prices, incomplete information and incomplete arbitrage. However, notwithstanding these 'short run' impediments, the argument goes, in the long run the real exchange rate reverts to the rate implied by PPP, although that reversion may be quite slow. One often-cited measure of the speed of return is the half-life of a shock; that is, when 50% of the overall adjustment to as shock has been reached.

One way of testing this property is by way of a test for a unit root on re_t, the presence of which is not compatible with mean reversion. Early studies used data from the post-Bretton Woods period of floating exchange rates for industrialised countries and one or more of the Dickey-Fuller tests or the Phillips and Perron (1988) semi-parametric versions of these tests, largely finding that the null hypothesis of a unit root in the generating process for re_t could not be rejected. One of the difficulties in establishing the robustness of this finding was that measures of the persistence of shocks suggested that they had very long life, with estimated half-lives of five years or more (see, for example, Rogoff, 1996). Thus a key problem was seen as distinguishing a unit root from a very near-unit root, highlighting the problem of the power of unit root tests, which is the ability of a test to find in favour of the alternative when the alternative is true.