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



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



All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6-10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The author has asserted his right to be identified as the author of this work in accordance with the Copyright, Designs and Patents Act 1988.

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 $^{\mathbb{R}}$  and Macmillan $^{\mathbb{R}}$  are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN 978-0-230-29930-6 (eBook)

ISBN 978-0-230-25025-3 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

 To Bella and a continuing treasure at Auton Farm

# Contents

Det	ailed Contents	ix			
List	t of Tables	xxii			
List	of Figures	xxvii			
Syn	nbols and Abbreviations	xxxii			
Preface					
1	Introduction to Random Walks and Brownian Motion	1			
2	Why Distinguish Between Trend Stationary and				
	Difference Stationary Processes?	38			
3	An Introduction to ARMA models	68			
4	Bias and Bias Reduction in AR Models	123			
5	Confidence Intervals in AR Models	158			
6	Dickey-Fuller and Related Tests	189			
7	Improving the Power of Unit Root Tests	260			
8	Bootstrap Unit Root Tests	319			
9	Lag Selection and Multiple Tests	348			
10	Testing for Two (or More) Unit Roots	385			
11	Tests with Stationarity as the Null Hypothesis	434			
12	Combining Tests and Constructing Confidence Intervals	497			
13	Unit Root Tests for Seasonal Data	519			
App	pendix 1: Random Variables; Order Notation	597			
App	pendix 2: The Lag Operator and Lag Polynomials	603			
Ref	erences	619			
Aut	hor Index	633			
Sub	oject Index	637			

# **Detailed** Contents

Li	st of T	ables			xxii
Li	st of F	igures			xxvii
Sy	ymbols and Abbreviations reface				
Pr	eface				xxxiii
1	Intr	oductio	n to Rand	om Walks and Brownian Motion	1
	1.1	Rando	m walks		3
		1.1.1	The rand	om walk as a partial sum process (psp)	4
		1.1.2	Random	walks: visits to the origin (sign changes and	
			reflection	ns)	5
		1.1.3	Random	walk: an example of a stochastic process	7
		1.1.4	Random	walk: an example of a nonstationary process	8
			1.1.4.i	A strictly stationary process	8
			1.1.4.ii	Weak or second-order stationarity	
				(covariance stationarity)	9
			1.1.4.iii	The variance of a random walk increases	
				over time	10
			1.1.4.iv	The autocovariances of a random walk are	
				not constant	11
		1.1.5	A simple	random walk with Gaussian inputs	12
		1.1.6	Variation	s on the simple random walk	12
		1.1.7	An empir	rical illustration	14
		1.1.8	Finer div	isions within a fixed interval: towards	
			Browniar	n motion	16
	1.2	Definit	ion of Bro	wnian motion	18
	1.3	Function	onal centra	al limit theorem (FCLT)	19
	1.4	Contin	uous map	ping theorem (CMT)	21
	1.5	Backgr	ound to ui	nit root and related tests	23
		1.5.1	What is a	a unit root?	23
			1.5.1.i	Generalising the random walk	23
			1.5.1.ii	Integrated of order d: the I(d) notation	25
		1.5.2	The deve	lopment of unit root tests	25
	1.6	Are un	it root pro	cesses of interest?	29
		1.6.1	Are there	constant 'great ratios'?	29
		1.6.2	Purchasiı	ng power parity	30
		1.6.3	Asset retu	irns	33
	1.7	Conclu	ıding rema	urks	35
		Questi	ons		36

2	Why	y Distin	guish Bety	ween Trend Stationary and Difference	
	Stat	ionary	Processes?		38
	2.1	What a	accounts fo	or the trend?	39
	2.2	Inappr	opriate det	rending and inappropriate first differencing	40
		2.2.1	Spurious	detrending	41
			2.2.1.i	Limit distributions and test statistics	41
			2.2.1.ii	Spurious periodicity	42
		2.2.2	Spurious	first differencing	46
			2.2.2.i	The over-differencing effect	46
			2.2.2.ii	Implications of over-differencing for the	
				spectral density function	46
	2.3	Persist	ence and th	he impact of shocks	47
		2.3.1	The role of	of shocks	48
			2.3.1.i	A simple model	48
			2.3.1.ii	More complex models	49
	2.4	Measu	res of the i	mportance of the unit root	51
		2.4.1	The varia	nce ratio for a simple random walk	51
		2.4.2	Interpreta	ation in terms of autocorrelations and	
			spectral d	lensity	53
		2.4.3	Interpreta	ation in terms of a permanent-temporary	
			(P-T) deco	omposition	54
		2.4.4	Interpreta	ation in terms of unconditional and	
			conditior	nal variances	55
		2.4.5	The relati	ionship between the persistence measure and	
			the varia	nce ratio	56
	2.5	Illustra	tion: US G	NP	57
	2.6	Unit ro	oot tests as	part of a pre-test routine	61
		2.6.1	Spurious	regression	61
	2.7	Conclu	iding Rema	arks	64
		Questi	ons		64
3	An Introduction to ARMA Models				
	3.1	ARMA	(p, q) mode	els	69
		3.1.1	Determin	listic terms	70
		3.1.2	The long	run and persistence	71
	3.2	Roots a	and stabilit	у	72
		3.2.1	The mod	ulus of the roots of the AR polynomial must	
			lie outsid	e the unit circle	72
		3.2.2	The speci	al case of unit roots	73
			3.2.2.i	Unit roots in the AR polynomial	73
			3.2.2.ii	A notational convention	74
			3.2.2.iii	Unit roots in the MA polynomial	75

3.3	Infinite	moving a	verage representation of the ARMA model	76
	3.3.1	Examples	5	76
	3.3.2	The gene	ral MA representation	77
	3.3.3	Inversion	of the MA polynomial	78
3.4	Approx	imation o	f MA and ARMA models by a 'pure' AR model	79
3.5	Near-ca	ncellation	of roots	80
3.6	The Bev	veridge-Ne	elson (BN) decomposition of a lag polynomial	81
	3.6.1	The BN d	ecomposition of a lag polynomial	81
	3.6.2	The BN d	ecomposition for integrated processes	
		(optional	on first reading)	83
		3.6.2.i	The BN decomposition in terms of the	
			ARMA lag polynomials	83
		3.6.2.ii	Precise and efficient computation of the BN	
			decomposition	86
		3.6.2.iii	Illustrations of the BN decomposition	87
3.7	The Die	ckey-Fuller	r (DF) decomposition of the lag polynomial	
	and the	e ADF mod	lel	89
	3.7.1	The DF d	ecomposition	90
3.8	Differen	nt ways of	writing the ARMA model	91
	3.8.1	Alternativ	ve ways of representing the dynamics	92
		3.8.1.i	An error dynamics model	92
		3.8.1.ii	Common factor model	94
		3.8.1.iii	Direct specification	96
	3.8.2	A pseudo	t test of the null hypothesis, $\gamma = 0$ : $\hat{\tau}$	98
	3.8.3	DF n-bias	s tests: δ	98
3.9	An out	line of ma	ximum likelihood (ML) estimation and	
	testing	for a unit	root in an ARMA model	100
	3.9.1	ML estim	ation	100
		3.9.1.i	The ARMA model (Shin and Fuller, 1998)	100
		3.9.1.ii	The log-likelihood functions	101
		3.9.1.iii	AR(1), stationary case with niid errors	102
	3.9.2	Likelihoo	d-based unit root test statistics	104
		3.9.2.i	ML: conditional and unconditional	
			approaches	105
	3.9.3	Estimatio	on, critical values and finite sample performance	107
		3.9.3.i	Estimation	107
		3.9.3.ii	Asymptotic distributions	108
		3.9.3.iii	Critical values and estimation	108
		3.9.3.iv	Finite sample performance	109
3.10	Illustra	tion: UK c	ar production	110
	3.10.1	Estimated	l model and test statistics	111
	3.10.2	The BN d	ecomposition of the car production data	112

	3.11	Conclu	ding remark	KS	115
		Questio	ons		116
4	Bias	and Bia	as Reduction	n in AR Models	123
	4.1	Finite s	ample bias o	of the LS estimator	124
		4.1.1	The bias to	order T	124
		4.1.2	First-order	bias and the long-run multiplier	127
	4.2	Bias ree	luction		128
		4.2.1	Total bias a	nd first-order bias	129
			4.2.1.i T	Total bias	129
			4.2.1.ii F	First-order bias	129
		4.2.2	First-order	unbiased estimators in the AR model	129
		4.2.3	Simulating	the bias	131
			4.2.3.i I	llustration of linear bias correction using	
			t	he AR(1) model	132
		4.2.4	Constant b	ias correction	133
		4.2.5	Obtaining	a linear bias corrected estimator	133
		4.2.6	Linear bias	correction in AR(p) models, $p \ge 2$	134
		4.2.7	The connee	ction	135
		4.2.8	The varian	ce and mean squared error of the FOBC	
			estimators		137
			4.2.8.i N	Mean squared error comparison	138
	4.3	Recursi	ve mean adj	ustment	138
		4.3.1	AR(1) mod	el	139
			4.3.1.i C	Constant mean	139
			4.3.1.ii 7	Frend in the mean	140
		4.3.2	Extension t	to AR(p) models	141
			4.3.2.i (	Constant mean, AR(p) models	141
			4.3.2.ii 1	Frend in the mean, AR(p) models	142
	4.4	Bootsti	apping		142
		4.4.1	Bootstrapp	ing to reduce bias	142
	4.5	Results	of a simulat	tion: estimation by least squares, first-order	
		bias co	rrection and	recursive mean adjustment	145
	4.6	Illustra	tions		147
		4.6.1	An AR(1) n	nodel for US five-year T-bond rate	147
		4.6.2	An AR(2) n	nodel for US GNP	149
	4.7	Conclu	ding remark	KS	150
		Questio	ons		151
	4.8	Appen	dix: First-ord	ler bias in the constant and linear trend cases	155
5	Con	fidence	Intervals in	AR models	158
	5.1	Confid	ence interva	ls and hypothesis testing	160
		5.1.1	Confidence	e intervals	160

	5.1.2	The link with hypothesis testing	162		
5.2	The quantile function				
	5.2.1	Constant quantiles	164		
	5.2.2	Nonconstancy of the quantiles	165		
	5.2.3	Complications in the AR case	165		
		5.2.3.i Simulation of the LS quantiles	166		
		5.2.3.ii Empirical example: an AR(2) model for US			
		GNP	167		
5.3	Constr	ucting confidence intervals from median unbiased			
	estima	tion	169		
	5.3.1	Inverse mapping and median unbiased estimation	169		
	5.3.2	Confidence intervals based on median unbiased			
		estimation	173		
	5.3.3	Extension of the median unbiased method to general			
		AR(p) models	175		
5.4	Quanti	iles using bias adjusted estimators	175		
	5.4.1	Quantiles	175		
	5.4.2	Coverage probabilities	177		
	5.4.3	Section summary	178		
5.5	Bootst	rap confidence intervals	179		
	5.5.1	Bootstrap confidence intervals	179		
		5.5.1.i The bootstrap percentile-t confidence interval	179		
		5.5.1.ii The grid-bootstrap percentile-t confidence			
		interval	182		
5.6	Conclu	iding remarks	184		
	Questi	ons	186		
Dicl	kev-Full	er and Related Tests	189		
6.1	The ba	sic set-up	190		
	6.1.1	An AR(1) process	190		
		6.1.1.i Back-substitution	190		
		6.1.1.ii Variances	192		
		6.1.1.iii Weakly dependent errors	193		
	6.1.2	The LS estimator	195		
	6.1.3	Higher-order AR models	197		
	6.1.4	Properties of the LS estimator, stationary AR(p) case	199		
6.2	Near-u	nit root case; simulated distribution of test statistics	200		
6.3	DF test	s for a unit root	203		
	6.3.1	First steps: formulation and interpretation of the			
		maintained regression	203		
	6.3.2	Three models	205		
	6.3.3	Limiting distributions of $\hat{\delta}$ and $\hat{\tau}$ test statistics	207		

6

7

	6.3.4	Obtainin	g similar tests	208
		6.3.4.i	Similarity and invariance in error	
			dynamics/common factor approach	209
		6.3.4.ii	Similarity and invariance in the direct (DF)	
			specification	209
	6.3.5	Continuo	ous record asymptotics	211
	6.3.6	The long	-run variance as a nuisance parameter	212
	6.3.7	Joint (F-t	ype) tests in the DF representation	214
	6.3.8	Critical v	alues: response surface approach	215
6.4	Size an	d power		217
6.5	Nonlin	lear trends		220
6.6	Exploi	ting the for	rward and reverse realisations	221
	6.6.1	DF-max t	ests (Leybourne, 1995)	221
	6.6.2	Weighted	l symmetric tests	223
6.7	Non-ii	d errors an	d the distribution of DF test statistics	226
	6.7.1	Departur	es from iid errors	227
		6.7.1.i	Distinguishing variances	227
		6.7.1.ii	Effects on the limit distributions	227
		6.7.1.iii	Expectations of the limit distributions	229
		6.7.1.iv	Effect of non-iid errors on the size of the tests	229
	6.7.2	Illustratio	ons	231
		6.7.2.i	MA(1) errors	231
		6.7.2.ii	AR(1) errors	232
		6.7.2.iii	Effect of non-iid errors on the size of DF tests	233
6.8	Phillip	s-Perron (P	PP) semi-parametric unit root test statistics	236
	6.8.1	Adjusting	g the standard test statistics to obtain the	
		same lim	it distribution	238
	6.8.2	Estimatic	on of the variances	239
		6.8.2.i	The unconditional variance, $\sigma_z^2$	239
		6.8.2.ii	Estimating the long-run variance	240
	6.8.3	Modified	Z tests	243
	6.8.4	Simulatio	on results	244
		6.8.4.i	PP tests, $Z\hat{\rho}_{\mu}$ and $Z\hat{\tau}_{\mu}$ , and modified PP tests,	
			$MZ\hat{\rho}_{\mu}$ and $MZ\hat{\tau}_{\mu}$	244
6.9	Power:	a compari	son of the ADF, WS and MZ tests	247
6.10	) Conclu	ıding rema	ırks	249
	Questi	ons		250
6.11	Appen	dix: Respo	nse surface function	255
Imp	proving	the Power	of Unit Root Tests	260
7.1	A GLS	approach		262
	7.1.1	GLS estin	nation	263

	7.1.2	Feasible (	GLS	263
		7.1.2.i	Simple examples	265
	7.1.3	Approxin	nating an MA process by an AR model	267
		7.1.3.i	Using the AR version of FGLS	267
		7.1.3.ii	Using the MA version of FGLS via an AR	
			approximation	268
		7.1.3.iii	ARMA errors	270
7.2	FGLS u	nit root te	sts	272
	7.2.1	DF-type t	ests based on FGLS	272
	7.2.2	GZW uni	t root test (ARMA variation)	272
7.3	Illustra	tion		276
	7.3.1	US Unem	ployment rate	276
		7.3.1.i	Standard ADF estimation	276
		7.3.1.ii	PP tests	277
		7.3.1.iii	FGLS (GZW version)	278
7.4	Tests th	at are opti	imal for particular alternatives	
	(point-	optimal te	sts)	279
	7.4.1	Likelihoo	d ratio test based on nearly-integrated	
		alternativ	7es	280
		7.4.1.i	Known deterministic components	280
		7.4.1.ii	Unknown deterministic components	282
		7.4.1.iii	A likelihood-based test statistic, with	
			estimated trend coefficients	283
		7.4.1.iv	The power envelope	285
	7.4.2	A family	of simple-to-compute test statistics	286
		7.4.2.i	The PT tests	286
		7.4.2.ii	Estimating $\sigma_{z,lr}^2$	287
	7.4.3	GLS detre	ending and conditional and unconditional tests	288
	7.4.4	Uncondit	tional case: the QT family of test statistics	290
	7.4.5	Exploitin	g the quasi-differencing interpretation	291
	7.4.6	ADF tests	using QD data	293
	7.4.7	Critical v	alues	294
	7.4.8	Illustratio	on of a power envelope and tangential power	
		functions	5	294
		7.4.8.i	Graphical illustrations	294
		7.4.8.ii	Power gains?	297
	7.4.9	Sensitivit	y to the initial condition	298
	7.4.10	A test stat	tistic almost invariant to the initial condition	300
	7.4.11	Weighted	l test statistics to achieve robustness to the	
		initial co	ndition	303
		7.4.11.i	Combining test statistics	304
		7.4.11.ii	Power and the initial condition	304

		7.4.12	Illustrati	on: US industrial production	306	
	7.5	Detren	ding (or d	emeaning) procedure by recursive mean		
		adjustr	nent		309	
	7.6	Conclu	iding rema	arks	312	
		Questions				
8	Boo	tstrap U	nit Root '	Tests	319	
	8.1	Bootst	rap schem	es with an exact unit root	321	
		8.1.1	The rand	lom inputs to the bootstrap replications	322	
		8.1.2	A bootst	rap approach to unit root tests	323	
		8.1.3	The AR(1	1) case	323	
			8.1.3.i	The AR(1) case: no constant	324	
			8.1.3.ii	Variations due to different specifications of		
				the deterministic terms	325	
		8.1.4	Extensio	n to higher-order ADF/AR models	326	
		8.1.5	Sieve boo	otstrap unit root test	327	
			8.1.5.i	Chang and Park sieve bootstrap	328	
			8.1.5.ii	Psaradakis bootstrap scheme	333	
	8.2	Simula	tion studi	es of the bootstrap unit root tests	335	
		8.2.1	Chang a	nd Park	336	
		8.2.2	Psaradak	is	336	
			8.2.2.i	Summary of results	337	
			8.2.2.ii	Lag truncation criteria	338	
		8.2.3	A brief c	omparison of the methods of Chang and Park		
			and Psar	adikis	338	
			8.2.3.i	AIC versus marginal-t as lag selection criteria	339	
			8.2.3.ii	Summary	341	
		8.2.4	'Asympto	otic' simulations	342	
	8.3	Illustra	tion: US u	inemployment rate	342	
	8.4	Conclu	iding rema	arks	345	
		Questi	ons		346	
9	Lag	Selectio	on and Mu	ıltiple Tests	348	
	9.1	Selection	on criteria		349	
		9.1.1	ADF set-	up: reminder	350	
		9.1.2	Illustrati	on of the sensitivity of size and power to lag		
			length se	election	352	
		9.1.3	Consiste	ncy and simulation lag selection criteria	356	
		9.1.4	Lag selec	tion criteria	358	
			9.1.4.i	Fixed lag	358	
			9.1.4.ii	Choose k* as a function of T	359	
			9.1.4.iii	Data-dependent rules	359	
		9.1.5	Simulati	on results	362	

			9.1.5.i	MA(1) errors	363	
			9.1.5.ii	AR(1) errors	364	
			9.1.5.iii	Lag lengths	365	
	9.2	Multip	le tests		367	
		9.2.1	Test chara	acteristics	367	
		9.2.2	Defining	the indicator sets	368	
		9.2.3	Overall ty	vpe I error	370	
		9.2.4	Test powe	er and test dominance	372	
			9.2.4.i	Test dependency function	373	
			9.2.4.ii	Test conflict function	373	
		9.2.5	Illustratio	ons: different test statistics and different lag		
			selection	criteria	374	
			9.2.5.i	Illustration 1: a comparison of $\hat{\tau}_{\mu}$ and $\hat{\tau}_{\mu}^{ws}$	374	
			9.2.5.ii	Illustration 2: using different lag selection		
				methods	376	
	9.3	Empirie	cal illustra	tion	381	
	9.4	Conclu	ding rema	rks	382	
		Questio	ons		383	
10	Testi	ng for ]	Гwo (or M	ore) Unit Roots	385	
	10.1	Prelimi	inaries			
		10.1.1	I(2) characteristics			
		10.1.2	Illustrative series			
		10.1.3	Alternativ	ves to I(2)	391	
		10.1.4	Some fact	torisations and restrictions	393	
	10.2	The Ha	sza-Fuller	(HF) decomposition of a lag polynomial	395	
		10.2.1	Derivatio	n of the HF decomposition	395	
			10.2.1.i	Example 1	397	
			10.2.1.ii	Numerical example (continuation)	397	
			10.2.1.iii	Example 2	398	
	10.3	DF test	as a secon	d unit root is approached	398	
	10.4	Testing	the null h	ypothesis of two unit roots	401	
		10.4.1	The HF te	est	402	
			10.4.1.i	Direct demeaning	402	
			10.4.1.ii	Initial conditions	404	
		10.4.2	Critical v	alues for HF F-type tests	405	
		10.4.3	Sen-Dicke	ey version of the HF test	405	
			10.4.3.i	Double-length regression for the SD		
				symmetric test statistic	407	
			10.4.3.ii	SD F-type test, $p=2$	409	
			10.4.3.iii	SD F-type test, $p > 2$	410	
			10.4.3.iv	Deterministic components	410	

10.5 The Dickey-Pantula (DP) approach	412
10.5.1 An extension of the DF procedure	412
10.5.2 Numerical illustration	413
10.5.3 Start from the highest order of integration	414
10.6 Power	415
10.6.1 Asymptotic local power	415
10.6.2 Small sample simulation results	416
10.6.2.i SD comparison of the HF and SD versions of	
the F test	416
10.6.2.ii Illustrative simulations: F test and DP tests	417
10.7 Illustrations of testing for two unit roots	418
10.7.1 Illustration 1: the stock of US consumer credit	418
10.7.2 Illustration 2: CPI Denmark	424
10.8 Concluding remarks	428
Questions	429
11 Tests with Stationarity as the Null Hypothesis	434
11.1 A structural time series model interpretation	436
11.1.1 The local level model	437
11.1.2 The importance of the limits	440
11.1.3 The 'fluctuations' testing framework	441
11.2 Tests for stationarity	442
11.2.1 The KPSS test	442
11.2.2 Modified rescaled range test statistic (MRS) (Lo, 1991)	443
11.2.3 Komogoroff-Smirnoff-type test (KS) (Xiao, 2001)	445
11.2.4 A Durbin-Watson-type test (SBDH) (Choi and Ahn,	
1999)	446
11.3 A test with parametric adjustment	448
11.3.1 The Leybourne and McCabe (1994) test	448
11.3.2 The modified Leybourne and McCabe test	450
11.3.3 Data-dependent lag selection for the LBM test(s)	451
11.4 The long-run variance	452
11.4.1 Estimation of the long-run variance	452
11.4.2 Illustrative simulation results	454
11.5 An evaluation of stationarity tests	456
11.5.1 Selection of the bandwidth parameter	456
11.5.2 Summary tables of limiting distributions and critical	
values	458
11.5.3 Empirical size	460
11.5.3.1 Illustrative simulations	461
11.5.3.11 Looking at the quantiles: comparing	
empirical distribution functions	467

		11.5.4	The power of tests for stationarity against the unit	
			root alternative	471
			11.5.4.i Power against a fixed alternative	472
			11.5.4.ii Power against local alternatives	478
	11.6	Illustra	tions: applications to some US consumer price indices	482
		11.6.1	US CPI (aggregate series)	482
		11.6.2	US regional consumer prices: using tests of	
			nonstationarity and stationarity in testing for	
			convergence	486
			11.6.2.i Framework for tests	486
			11.6.2.ii Consumer prices for two regions of the US	487
	11.7	Conclu	iding remarks	491
		Questic	ons	494
12	Com	bining	Tests and Constructing Confidence Intervals	497
	12.1	The im	portance of the initial condition	498
	12.2	Probler	ns with stationarity tests for highly correlated series	500
	12.3	Confid	ence intervals revisited	503
		12.3.1	Inverting the PT unit root test statistic to obtain a	
			confidence interval	505
		12.3.2	Constructing a confidence interval using DF tests with	
			GLS detrended data	508
			12.3.2.i Test inversion	508
			12.3.2.ii Illustration using time series on US GNP	511
	12.4	Conclu	iding remarks	515
		Questic	ons	516
13	Unit	Root To	ests for Seasonal Data	519
	13.1	Seasona	al effects illustrated	521
	13.2	Seasona	al 'split' growth rates	524
	13.3	The spe	ectral density function and the seasonal frequencies	525
		13.3.1	Frequencies, periods and cycles	526
		13.3.2	Power spectrum and periodogram	527
		13.3.3	Illustrations (randomising the input)	529
		13.3.4	Aliasing (artificial cycles)	532
		13.3.5	Seasonal integration from a frequency domain	
			perspective	534
	13.4	Lag ope	erator algebra and seasonal lag polynomials	535
		13.4.1	The seasonal operator in terms of L	535
	13.5	An intr	oduction to seasonal unit root tests: the DHF	
		(1984)	test	539
		13.5.1	The seasonal DGP	539
			13.5.1.i No deterministic components	540

13.5.1.ii Seasonal intercepts	540
13.5.1.iii Seasonal deterministic trends	542
13.5.2 Limiting distributions	543
13.6 HEGY tests	543
13.6.1 The HEGY regression	545
13.6.2 The constructed variables and the seasonal frequencies	s 546
13.6.3 The structure of hypothesis tests	547
13.6.3.i Tests for the roots $+1, -1$	547
13.6.3.ii Tests for unit roots at the harmonic seasona	1
frequencies, $\lambda_s, s \neq 0, s \neq S/2$	547
13.6.3.iii Overall tests	548
13.6.4 The HEGY tests: important special cases	548
13.6.4.i Quarterly data	548
13.6.4.ii Monthly data	550
13.6.5 Limiting distributions and critical values	552
13.6.5.i Monthly case	553
13.6.5.ii Quarterly case	555
13.6.5.iii Nuisance parameters	555
13.6.5.iv Empirical quantiles	556
13.6.6 Multiple testing	558
13.6.7 Lag augmentation	561
13.7 Can the (A)DF test statistics still be used	
for seasonal data?	561
13.8 Improving the power of DHF and HEGY tests	562
13.8.1 Recursive mean adjustment for seasonal unit root test	s 563
13.8.2 Improving power with monthly data	564
13.8.2.i Systematic sampling	564
13.8.2.ii QM-HEGY	564
13.8.2.iii Choice of test statistic for quarterly data	566
13.8.2.iv Critical values	567
13.8.2.v Illustration: US industrial production	567
13.9 Finite sample results, DHF and HEGY	568
13.9.1 Power, initial assessment	569
13.9.2 Not all roots present under the null	570
13.9.2.i AR(1) DGP	572
13.9.2.ii Other DGPs: one cycle and two cycles per y	ear 574
13.9.3 Extension to include RMA versions of	
the HEGY and DHF tests	577
13.9.4 Size retention	578

13.10	Empirical illustrations	582
	13.10.1 Illustration 1, quarterly data: employment in US	
	agriculture	582
	13.10.1.i A linear trend alternative	582
	13.10.1.iiA nonlinear trend alternative	586
	13.10.2 Illustration 2, monthly data: UK industrial	
	production	587
13.11	Some other developments	590
	13.11.1 Periodic models	590
	13.11.2 Stationarity as the null hypothesis	592
13.12	Concluding remarks	593
	Questions	594
Appendix	x 1: Random Variables; Order Notation	597
A1.1	Discrete and continuous random variables	597
A1.2	The (cumulative) distribution function, cdf	598
	A1.2.1 CDF for discrete random variables	598
	A1.2.2 CDF for continuous random variables	599
A1.3	The order notation: 'big' O, 'small' o	602
Appendix	c 2: The Lag Operator and Lag Polynomials	603
A2.1	The lag operator, L	603
A2.2	The lag polynomial	603
A2.3	Differencing operators	604
A2.4	Roots	605
	A2.4.1 Solving for the zeros	605
	A2.4.2 Graphical representation of the roots	605
	A2.4.3 Roots associated with seasonal frequencies	608
A2.5	The inverse of a lag polynomial and the summation operator	609
	A2.5.1 The inverse lag polynomial	609
	A2.5.2 The summation operator $\Delta^{-1}$ as an inverse	
	polynomial	610
A2.6	Stability and the roots of the lag polynomial	610
A2.7	Some uses of the lag operator	612
	A2.7.1 Removing unit roots	612
	A2.7.2 Data filter	613
	A2.7.3 Summing the lag coefficients	614
	A2.7.4 Application to a (deterministic) time trend	615
	Questions	615
Reference	S	619
Author In	dex	633
Subject In	ldex	637

# List of Tables

P.1	Number of citations of key articles on unit roots	xxxiv
1.1	Contrasting properties of I(0) and I(1) series	26
2.1	Errors of specification	41
2.2	Estimated ARMA models for US GNP (real)	58
2.2a	ARMA (3, 2) model for y <sub>t</sub>	58
2.2b	ARMA (2, 2) model for $\Delta y_t$	58
2.3	Roots of the polynomials of the ARMA models	58
2.3a	ARMA (3, 2) model for y <sub>t</sub>	58
2.3b	ARMA (2, 2) model for $\Delta y_t$	58
3.1	Finite sample critical values for conditional and	
	unconditional LR-type test statistics	109
3.2	LS, CML and ML: illustrative simulation results, MA(1)	
	errors, $T = 100$	110
3.3	ARMA models for UK car production, coefficient estimates	112
3.4	Test statistics for a unit root, car production data	112
4.1	First-order least squares bias: $\lim_{T\to\infty} T[E(\hat{\varphi}_i - \varphi_i)]$	126
4.1a	Constant included in regression	126
4.1b	Constant and trend included in regression	126
4.2	Total bias of LS estimators and fixed points	126
4.2a	Constant included in regression	126
4.2b	Constant and trend included in regression	126
4.3	A summary of the abbreviations used in the following	
	sections	128
4.4	Bias correction in an AR(1) model	146
4.5	LS and bias adjusted estimates of an AR(2) model for US	
	five-year T-bond	148
4.6	LS and bias adjusted estimates of an AR(2) model for US GNP	150
4.7	Roots for the zero bias case: constant included in regression	153
5.1	Quantiles to obtain a median unbiased estimator	173
5.2	Simulated coverage probabilities	178

6.1	Percentage of times that $t_{\rho} < -1.645$	203
6.2	Summary of simple maintained regressions: error	
	dynamics/common factor approach	206
6.3	Summary of maintained regressions: DF representations	207
6.4	Null hypothesis for DF joint F-type tests	215
6.5	Illustration of response surface coefficients	216
6.6	Illustrative size and power comparisons of DF-type tests	218
6.6a	MA(1) error structure: $z_t = (1 + \theta_1 L)\varepsilon_t$	218
6.6b	AR(1) error structure: $(1 - \varphi_1 L)z_t = \varepsilon_t$	219
6.7	Critical values of the DF-max tests, $\hat{\tau}_{\mu}^{max}$ and $\hat{\tau}_{\beta}^{max}$	223
6.8	Effect of non-iid errors on means of limit distribution of	
	test statistics	230
6.9	Effect on (asymptotic) size of DF tests with non-iid errors	234
6.10a	Effect of MA(1) errors on the size of PP-type tests, 5%	
	nominal size	245
6.10b	Effect on AR(1) errors on the size of PP-type tests, 5%	
	nominal size	246
6.11	Illustrative size and power comparisons of DF-type tests	248
6.11a	MA(1) error structure	248
6.11b	AR(1) error structure	248
6.12	Response surface coefficients	257
7.1	Summary of GLS matrices, general and special cases	264
7.2	Size and power comparisons: CGLS and ML test statistics,	
	AR(1) error structure	275
7.3	Estimation results for US unemployment, Nelson and	
	Plosser data	277
7.4	Summary of unit root test values for US unemployment	
	data	278
7.5	Critical values (fixed initial condition)	295
7.6	Critical values (initial condition: drawn from	
	unconditional distribution)	295
7.7	Critical values for $\hat{Q}^{(i)}(\mathbf{g}, \lambda)$	303
7.8	Critical values for $\hat{\tau}_i^{DFG}$ , $\hat{\tau}_i^{DFQ}$	305
7.9	Critical values for $\hat{\tau}_{\mu}^{\text{rma}}$ and $\hat{\tau}_{\beta}^{\text{rma}}$	311
7.10	Estimation results for US unemployment	314
7.11	Summary of unit root test values for US unemployment	
	data	314
8.1	Simulated size and power of the Chang and Park ADF tests	337

8.2	Simulated size ( $\rho = 1$ ) and power ( $\rho = 0.9$ ) of Psaradakis	
	bootstrap tests for 5% nominal size, T = 100; $\theta_1 = -0.8$	
	throughout	338
8.3	Comparison of bootstrapping procedures	340
8.3a	Simulated size: $\theta_1 = 0$ and $\rho = 1$ ; simulated power: $\theta_1 = 0$	
	and $\rho = 0.95$ , $\theta_1 = 0$ and $\rho = 0.9$	340
8.3b	Simulated size: $\theta_1 = -0.4$ and $\rho = 1$ ; simulated power:	
	$\theta_1 = -0.4$ and $\rho = 0.95$ , $\theta_1 = -0.4$ and $\rho = 0.9$	340
8.3c	Simulated size: $\theta = -0.8$ and $\rho = 1$ ; simulated power:	
	$\theta_1 = -0.8$ and $\rho = 0.95$ , $\theta_1 = -0.8$ and $\rho = 0.9$	341
8.4	'Aysmptotic' size: $\theta_1 = -0.8$ and $\rho = 1$	342
8.5	ML estimation of ARIMA(2, 0, 2) for US unemployment rate	343
8.6	Test statistics for US unemployment rate	344
9.1	$MA(1)$ errors: empirical size and power of $\hat{\tau}_{\mu}$ using different	
	lag selection criteria	363
9.2	AR(1) errors: size and power of $\hat{\tau}_{\mu}$ , different lag selection	
	criteria	364
9.3	Intersection of sets using different lag selection criteria, $\%$	
	same choice, MA(1)	366
9.4	Intersection of sets using different lag selection criteria, $\%$	
	same choice, AR(1)	366
9.5	Joint probabilities for two test statistics	369
9.6	Probabilities under independence of tests given $H_0$	370
9.7	Illustrative probabilities if tests are distributed as joint	
	normal, $\alpha = 0.05$ , with dependency indicated by $\eta$	372
10.1	A taxonomy of possibilities in the two root case	392
10.2	Critical values for HF F-type tests	406
10.3	Empirical power of tests for two unit roots; data directly	
	demeaned assuming a constant mean	418
10.4	Empirical power of tests for two unit roots; data directly	
	demeaned by a constant and linear trend	419
10.5	Empirical power of tests for two unit roots; data directly	
	demeaned by a constant, linear trend and a quadratic trend	420
10.6	Number of unit roots; estimation details for log consumer	
	credit, US	423
10.7	Number of unit roots; estimation details for log of CPI,	
	Denmark	427
10.8	Dependent and explanatory variables in the simple	
	symmetric test of Sen and Dickey	432

11.1	Limiting distributions of test statistics for stationarity	459
11.2	Quantiles of null distributions for stationary tests (critical	
	values)	460
11.3	Effect of lag selection on quantiles of $LBM^{(\mu)}$	461
11.4	Stationarity tests: empirical size for 5% nominal size AR(1)	
	errors	462
11.5	Stationarity tests: empirical size for 5% nominal size, MA(1)	
	errors	463
11.6a	LBM tests, US CPI (logs)	485
11.6b	Stationarity tests, US CPI (logs)	485
11.7	Long-run variance estimates and bandwidth parameters	485
11.8	Unit roots tests on US regional CPIs; (log) levels, N-E and W	490
11.9	US regional inflation contrast: LBM <sup><math>(\mu)</math></sup> tests for stationarity	491
11.10	US regional inflation contrast: tests for stationarity	492
12.1	Sensitivity of empirical size to the initial condition	500
12.2	Upper quantiles of unit root test statistics when used for	
	staionarity	504
12.3	The roles of different values of c and $\rho$	505
12.4	$H_0: \rho = \rho^*$ against $H_{\Delta}: \rho = \bar{\rho}$ or, equivalently, $c = c^*$	
	against $c = \bar{c}$	508
12.5	Stationarity tests for US GNP	512
13.1	Calculation of seasonal 'split' growth rates:	
	$\Delta_1 y_{n,s} = y_{n,s} - y_{n,s-1}$	525
13.2	Aliased frequencies: monthly data sampled at quarterly	
	intervals	533
13.3	Seasonal test statistics distinguished by mean and trend	
	components	542
13.4	Quantiles for DFH tests	544
13.5	Seasonal frequencies, periods and roots: the quarterly case	549
13.6	Constructed variables and seasonal frequencies: the	
	quarterly case	549
13.7	Hypothesis tests: the quarterly case	549
13.8	Seasonal frequencies, periods and roots: the monthly case	552
13.9	Constructed variables and seasonal frequencies: the	
	monthly case	553
13.10	Hypothesis tests: the monthly case	553
13.11	Limiting null distributions of HEGY test statistics	554
13.12	HEGY critical values: quarterly data	557
13.13	HEGY critical values: monthly data	559
13.14	Matrix of estimated pairwise probabilities (monthly tests)	561

13.15	Some critical values for QM-HEGY, 5% tests (4% quarterly,	
	1% monthly)	567
13.16	Summary of results for US industrial production	568
13.17	Size and power of tests	578
13.18	Size of tests: AR(1) errors	580
13.19	Size of tests: MA(4) errors	581
13.20	Average lags for AR(1) errors and MA(4) errors	582
13.21	Test statistics for unit roots for US agricultural employment	585
13.22	Test statistics for unit roots for UK industrial production	
	(monthly)	589
13.23	Constructed variables and seasonal frequencies for the	
	weekly case	596
13.24	The structure of hypotheses: the	
	weekly case	596

# List of Figures

1.1	Sample paths of a binomial random walk	2
1.2	Distribution function of visits to the origin	6
1.3	Sample paths of a Gaussian random walk	13
1.4	Sample paths of asymmetric binomial random walks	14
1.5	Sample paths of drifted symmetric binomial random walks	15
1.6	Exchange rate (daily), SWFR:£	15
1.7	Scatter graph of daily, SWFR:£	16
1.8	Gaussian random walk as time divisions approach zero	17
2.1	AC functions for residuals of detrended series	44
2.2a	Simulated spectral density function, T = 100	45
2.2b	Simulated spectral density function, T = 200	45
2.3	Power spectrum of over-differenced series	47
2.4	US GNP (quarterly, s.a.)	57
2.5	Impulse responses based on ARMA (2, 2) model	59
2.6	Estimates of G(1) and variance ratio from AR(k) models	60
2.7	Simulated density function of R <sup>2</sup>	63
2.8	Simulated density function of $t_{\beta_2}$	63
2.9	Measures of persistence for an MA(1) process	67
3.1	UK car production (logs, p.a., s.a.)	111
3.2	Permanent and transitory car production (logs)	115
4.1	Bias functions of estimators	131
4.2	RMSE functions of estimators	147
4.3	US five-year T-bond rate	148
4.4	US GNP (in logs, constant prices)	149
5.1	5% and 10% quantiles and t-function	166
5.2	Quantiles for LS estimates	167
5.3	Quantiles for LS estimates: US GNP	168
5.4	Quantile functions	172
5.5	Inverse quantile functions	172
5.6	90% confidence interval	174
5.7	5% quantiles for different estimators	176
5.8	95% quantiles for different estimators	177

5.9a	Simulated cumulative distribution function	181
5.9b	Simulated inverse cumulative distribution function	181
5.10	Bootstrap quantiles: US GNP	185
5.11	Obtaining a confidence interval	187
6.1a	Distribution of $t_{\rho}$ (no constant in estimated model)	201
6.1b	Distribution of $t_{\rho}$ (constant in estimated model)	202
6.1c	Distribution of $t_{\rho}$ (constant and trend in estimated model)	202
6.2	Distribution of $t_{\rho}$ with drift in DGP	211
6.3	Ratio of variances: long-run/unconditional (MA(1) errors)	232
6.4	Ratio of variances: long-run/unconditional (AR(1) errors)	233
6.5a	Distribution of $\hat{\tau}$ with MA(1) errors	235
6.5b	Distribution of $\hat{\tau}_{\mu}$ with MA(1) errors	235
6.5c	Distribution of $\hat{\tau}_{\beta}$ with MA(1) errors	236
6.6a	Distribution of $\hat{\tau}$ with AR(1) errors	236
6.6b	Distribution of $\hat{\tau}_{\mu}$ with AR(1) errors	237
6.6c	Distribution of $\hat{\tau}_\beta$ with AR(1) errors	237
7.1	US unemployment rate (Nelson and Plosser series)	277
7.2	Variation in 5% critical values for test statistics	296
7.3	Illustration of power envelope for $\rho_{T,\mu}$ (c)	296
7.4	Comparison of power envelopes	297
7.5	Power as scale of initial observation varies, $\pi(\kappa   \rho)$ , constant	
	in estimated model	299
7.6	Power as initial observation varies, $\pi(\kappa   \rho)$ , trend in	
	estimated model	300
7.7	Power comparison, $\rho = 0.95$	305
7.8	Power comparison: the weighted tests	306
7.9	US Industrial production (logs, p.a., s.a.)	307
7.10a	Estimated scaled residuals $\hat{\xi}_{ ext{c}}$	308
7.10b	Estimated weights â	308
7.11a	DF and GLSC unit root test	308
7.11b	Weighted unit root test	308
8.1	US unemployment rate (monthly data)	343
9.1	MA (1) errors: size and power, $\hat{\delta}_{\mu}$	352
9.2	MA(1) errors: size and power, $\hat{\tau}_{\mu}$	353
9.3	MA(1) errors: relative frequency of lags chosen by AIC	353
9.4	AR(1) errors: size and power, $\hat{\delta}_{\mu}$	355
9.5	AR(1) errors: size and power, $\hat{\tau}_{\mu}$	355

9.6	AR(1) errors: relative frequency of lag chosen by AIC	356
9.7a	Bivariate normal with $\eta = 0.75$	371
9.7b	Probability of conflict, two tests, bivariate normal	372
9.8	Power functions for $\hat{\tau}_{\mu}$ and $\hat{\tau}_{\mu}^{ws}$	374
9.9	Conditional test dependency functions	375
9.10	Unconditional test conflict	376
9.11	Power of $\hat{\tau}_{\mu}$ using G-t-S and MAIC	377
9.12	Conditional test dependencies, G-t-S and MAIC	378
9.13	Unconditional test conflict $\hat{\tau}_{\mu}$ with G-t-S and MAIC	379
9.14	Power of $\hat{\tau}_{\mu}$ using G-t-S and MAIC	379
9.15	Conditional test dependencies $\hat{\tau}_{\mu}$ with G-t-S and MAIC	380
9.16	Unconditional test conflict $\hat{\tau}_{\mu}$ with G-t-S and MAIC	380
9.17	US (log) wheat production	381
9.18	The impact of different lag lengths	382
10.1	Simulated I(d) series, $d = 0, 1, 2, 3$	389
10.2a	US average hourly earnings (logs)	390
10.2b	Denmark CPI (logs)	390
10.3a	US M1 (logs)	391
10.3b	US consumer credit (logs)	391
10.4a	Empirical size of DF $\hat{\tau}_{\mu}$ lower tail	399
10.4b	Empirical size of DF $\hat{\tau}_{\mu}$ upper tail	400
10.5a	Empirical size of DF $\hat{\tau}_{\beta}$ lower tail	400
10.5b	Empirical size of DF $\hat{\tau}_{\beta}$ upper tail	401
10.6a	US Consumer credit, actual and trends (logs)	421
10.6b	Residuals from trends, US credit	421
10.6c	US credit, monthly growth rate and trend	422
10.6d	Residuals for linear trend for growth rate	422
10.7a	Denmark CPI (logs), actual and trends	425
10.7b	Residuals from trends, Denmark CPI	426
10.7c	Denmark CPI, annual inflation rate and trend	426
10.7d	Residuals from linear trend for growth rate	427
11.1	Bartlett and quadratic spectral kernals	453
11.2	Estimators of long-run variance: rmse, AR(1)	455
11.3	QS, BW and QS (CAB) estimators: rmse, AR(1)	456
11.4	Estimators of long-run variance: rmse, MA(1)	457
11.5	Comparison by bandwidth criteria: rmse, MA(1)	457
11.6a	$KS^{\mu}$ , empirical size for AR(1) errors: 4 kernels	464
11.6b	LBM <sup><math>\mu</math></sup> , empirical size for AR(1) errors	464
11.7	Best of each test, empirical size for AR(1) errors	465
11.8a	KPSS <sup><math>\mu</math></sup> , empirical size for MA(1) errors: 4 kernels	466

11.8b	LBM $^{\mu}$ , empirical size for MA(1) errors	467
11.9	Best of each test, empirical size for MA(1) errors	467
11.10	EDFs for KPSS <sup>(<math>\mu</math>)</sup> and LBM <sup>(<math>\mu</math>)</sup>	468
11.11	EDFs for KPSS <sup>(<math>\mu</math>)</sup> , AR(1) errors	469
11.12	EDFs for LBM <sup>(<math>\mu</math>)</sup> , AR(1) errors	469
11.13	EDFs for KPSS <sup>(<math>\mu</math>)</sup> , MA(1) errors	470
11.14	EDFs for LBM <sup>(<math>\mu</math>)</sup> , MA(1) errors	471
11.15a	Power of test statistics in standard versions (T = 100)	473
11.15b	Size-adjusted power of test statistics (T = 100,	
	semi-parametric tests use $m^{(qs)}(4)$ )	473
11.15c	Size-adjusted power of test statistics (T = 100,	
	semi-parametric tests use m <sup>(qs)</sup>	474
11.15d	Size-adjusted power of test statistics ( $T = 100$ ,	
	semi-parametric tests use $m^{(bw)}(4)$ )	474
11.15e	Size-adjusted power of test statistics ( $T = 100$ ,	
	semi-parametric tests use $m_{(bw)}^{(bw)}$	475
11.15f	Size-adjusted power of LBM test statistics ( $T = 100$ )	475
11.16a	Best power combinations ( $T = 100$ )	476
11.16b	Power comparison with best-size combinations ( $T = 100$ )	476
11.17a	Size-adjusted power. $AR(1)$ DGP (T = 500. semi-parametric	
	tests use $m^{(qs)}(4)$ )	477
11.17b	Best power and size, $AR(1)$ DGP (T = 100)	478
11.18a	Best power combinations ( $T = 500$ )	479
11.18b	Power comparison with best size combinations ( $T = 500$ )	479
11.19a	Size adjusted power, $AR(1)$ DGP (T = 500, semi-parametric	
	tests use $m^{(qs)}(4)$ )	480
11.19b	Best power and size, $AR(1)$ DGP (T = 500)	480
11.19c	Best power and size, $AR(1)$ DGP (T = 500, semi-parametric	
	tests use $m_{ab}^{(bw)}(4)$	481
11.20	US CPI and linear trend (logs)	483
11.21	US CPI inflation rate	483
11.22	US regional CPIs (logs)	488
11.23	CPI contrast (log)	488
11.24	Regional CPI inflation rates	489
11.25	Contrast of inflation rates	489
12.1	90% confidence interval using QD data: US GNP	514
12.2	90% confidence interval, QD data, with c(L): US GNP	515
13.1a	Consumers' expenditure (UK)	522
13.1b	Consumers' expenditure: seasonal component	522
13.2a	Consumers' expenditure (logs)	523

13.2b	Consumers' expenditure (logs): seasonal component	523
13.3	Consumers' expenditure: DV seasonal component	524
13.4	Consumers' expenditure	525
13.5	Expenditure on restaurants and hotels (UK)	526
13.6	The periodogram for a single frequency	529
13.7a	Periodogram for periods of 12 and 120	530
13.7b	Periodogram for periods of 4 and 120	530
13.8a	Detrended expenditure on restaurants and hotels	531
13.8b	Periodogram for levels	531
13.9a	Quarterly growth rate	532
13.9b	Periodogram for growth rate	532
13.10	Aliasing of monthly cycles when sampled quarterly	533
13.11a	Periodogram for $\Delta_4 y_t = \varepsilon_t$	537
13.11b	Seasonal splits for $\Delta_4 y_t = \varepsilon_t$	537
13.12a	Periodograms for $(1 + L)y_t = \varepsilon_t$ and $(1 + L^2)y_t = \varepsilon_t$	538
13.12b	Seasonal splits for $y_t = -y_{t-1} + \varepsilon_t$	538
13.12c	Seasonal splits for $y_t = -y_{t-2} + \varepsilon_t$	539
13.13	The monthly, seasonal frequencies $\lambda_i$	551
13.14a	Simulated data for $\Delta_{12}y_t = \varepsilon_t$	551
13.14b	Periodogram for $\Delta_{12}y_t = \varepsilon_t$	552
13.15a	Power of DFH $\hat{\tau}_{S,S\mu}$ and HEGY $F_{All}$ tests, no lags	571
13.15b	Power of DFH $\hat{\tau}_{S,S\mu}$ and HEGY $F_{All}$ tests, 4 lags	571
13.16a	Power of tests with RW(1) DGP	573
13.16b	EDFs for $\hat{\tau}_{S,S\mu}$ with RW(1) DGP	573
13.17a	Power of tests with Nyquist DGP	575
13.17b	EDFs for $\hat{\tau}_{S,S\mu}$ with Nyquist DGP	575
13.18a	Power of tests with seasonal harmonic DGP	576
13.18b	EDFs for $\hat{\tau}_{S,S\mu}$ with seasonal harmonic DGP	576
13.19	Power of DFH $\hat{\tau}_{S,S\mu}$ and HEGY $F_{All,S\mu}$	579
13.20	Power of DFH $\hat{\tau}_{S,S\mu,S\beta}$ and HEGY $F_{All,S\mu,S\beta}$	579
13.21	US agricultural employment	583
13.22	Seasonal split growth rates	584
13.23	Periodogram for US agricultural employment	584
13.24	UK Index of Industrial Production (logs) n.s.a.	587
13.25	Periodogram for UK IIP (log) n.s.a.	588
A1.1	Binomial distribution, pmf and cdf	600
A1.2	Standard normal distribution, pdf and cdf	602
A2.1a	Finding the roots of $\varphi(z)$	606
A2.1b	$\phi(z)$ with complex roots	607

# Symbols and Abbreviations

$\Rightarrow_{\mathrm{D}}$	convergence in distribution (weak convergence)
$\rightarrow_{\mathrm{p}}$	convergence in probability
$\rightarrow$	tends to; for example, $\varepsilon$ tends to zero, $\varepsilon \rightarrow 0$
$\mapsto$	mapping
$\Rightarrow$	implies
$\sim$	is distributed as
≡	definitional equality
≠	not equals
$\Phi(z)$	the cumulative distribution function of the standard normal distribution
R	the set of real numbers; the real line $(-\infty \text{ to } \infty)$
$\Re^+$	the positive half of the real line
$N^+$	the set of non-negative integers
ε <sub>t</sub>	white noise, unless explicitly excepted
$\prod_{j=1}^{n} x_j$	the product of $x_j$ , $j = 1,, n$
$\sum_{j=1}^{n} x_j$	the sum of $x_j$ , $j = 1,, n$
$0^+$	approach zero from above
_0	approach zero from below
L	the lag operator, $L^j y_t \equiv y_{t-j}$
$\Delta$	the first-difference operator, $\Delta \equiv (1 - L)$
$\Delta_{s}$	the s-th difference operator, $\Delta_s \equiv (1 - L^j)$
$\Delta^{s}$	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
U	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

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  $45^{\circ}$  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 twodimensional 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  $\Psi$ , 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  $\Psi$ , 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:

$$y_t = y_{t-1} + \epsilon_t$$
  $t = 1, ..., T$  (1.1)

Thus,  $y_t$  is determined as its previous value,  $y_{t-1}$ , plus the intervention of a random variable denoted  $\varepsilon_t$ , 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  $\varepsilon_j$  be the random variable that is derived from  $x_j$ , such that the original sample space  $\Omega_j$  is mapped to  $\Omega_j^\varepsilon = (+1, -1)$ . The probability measures of  $x_j$  and  $\varepsilon_j$  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  $\epsilon_j, \ j=1, \ \ldots, \ t$ , 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  $\Omega_1$ , that is  $\Omega_t^y \!=\! (\Omega_1^\epsilon)^t \!=\! \Omega_1^\epsilon \times \Omega_1^\epsilon \times \ldots \times \Omega_1^\epsilon$  and, by independence, the probability measure associated with  $y_t$  is the product measure  $P_t^y \!=\! (P_1^\epsilon)^t \!=\! P_1^\epsilon \times P_1^\epsilon \times \ldots \times P_1^\epsilon$ . Note that  $E(\epsilon_t)\!=\!0, \ t=1, \ \ldots, \ T$  and  $E(y_t)\!=\! E(y_0)\!+\! \Sigma_{i=1}^t E(\epsilon_i)\!=\!0$ , 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':

$$y_{B,t} = y_{B,t-1} - \varepsilon_t$$

$$= y_{B,0} - \sum_{j=1}^t \varepsilon_j$$
(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  $\varepsilon_j$  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  $\zeta$  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  $\sigma^2$  (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  $\Omega$ , 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, ..., 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  $\Delta t$  and  $\Delta t \to 0$ .

Let  $\Theta$  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  $\Theta$  has a finite or countably infinite number of elements. Indeed, in this case  $\Theta$  is the set of positive integers or, if the process is viewed as starting at t=0, the set of nonnegative integers, t  $\in \Theta = N^+ = (0, 1, 2, ...)$ ; equally the process might be viewed as starting in the infinite past and, hence, t  $\in \Theta = N = (0, \pm 1, \pm 2, ...)$ . In the continuous time case,  $\Theta$  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:

$$Y = (y_t(\omega) : t \in \Theta, \omega \in \Omega)$$
 discrete time (1.5a)

 $Y = (y(t, \omega) : t \in \Theta \subseteq \Re, \omega \in \Omega) \quad \text{continuous time}$ (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  $\omega$  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  $\Omega$ , 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  $\omega$  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,  $\varepsilon_t$ , 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  $\Delta t \rightarrow 0$  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:

$$P(y_{\tau+1}, y_{\tau+2}, \dots, y_{\tau+T}) = P(y_{s+1}, y_{s+2}, \dots, y_{s+T})$$
(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  $\tau$  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:

$$F[y(\tau+t_1), y(\tau+t_2), \dots, y(\tau+t_N)] = F[y(s+t_1), y(s+t_2), \dots, y(s+t_N)]$$
(1.7)

where  $t_1 < t_2 ... < t_N$ ,  $\tau \neq s$  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(.):

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

#### 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  $\tau$  and s,  $\tau \neq$  s:

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

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

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

$$\begin{split} &Y\!=\!(y_t(\omega)\!:\!t\in\Theta,\omega\in\Omega) \qquad \text{nonstationary process} \\ &DY\!=\!(\Delta y_t(\omega)\!:\!t\in\Theta^*\subseteq\Theta,\omega\in\Omega) \quad \text{stationary process} \end{split}$$

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  $E(y_t) \neq \text{constant}$ , 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.

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

The deterministic components are captured by the term  $\mu_t$ ; 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  $\mu_t$  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:

$$var(y_t) = \sum_{j=1}^{t} var(\varepsilon_j) + 2\sum_{i=1}^{t} \sum_{j>i}^{t} cov(\varepsilon_i, \varepsilon_j)$$
(1.10a)

$$= \sum_{j=1}^{t} var(\epsilon_j) \quad \text{as } cov(\epsilon_i, \epsilon_j) = 0, i \neq j \tag{1.10b}$$

$$=\!t\sigma_{\epsilon}^2 \qquad \quad \text{as } var(\epsilon_j)\!=\!\sigma_{\epsilon}^2 \ \text{for all } j \qquad \qquad (1.10c)$$

In the case of the symmetric, binomial random walk  $\sigma_{\epsilon}^2 = 1$  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  $cov(\epsilon_i, \epsilon_j) = 0$ ,  $i \neq j$  and  $var(\epsilon_j) = \sigma_{\epsilon}^2$  for all j, then, holds, so the result would also hold for  $\epsilon_t \sim niid(0, \sigma_{\epsilon}^2)$  or, weaker still, that  $\epsilon_t$  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:

$$\gamma(k) = E[y_t - E(y_t)][(y_{t-k} - E(y_{t-k})] \quad k = \pm 1, 2, 3, \dots$$

$$= cov(y_t, y_{t-k})$$
(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:

$$\gamma(0) = E[y_t - E(y_t)]^2$$
(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:

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

The scaling ensures that  $0 \le |\rho_k| \le 1$ . 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  $\rho_k$  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  $var(y_0)=0$ , then  $\gamma(1)$  for t=2 and t=3, are, respectively, as follows:

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

Hence  $cov(y_1, y_2) \neq cov(y_2, y_3)$  although both relate to an index value k=1. These derivations exploit the properties of white noise  $E(\epsilon_t^2) = \sigma_{\epsilon}^2$  for all t, and  $E(\epsilon_t \epsilon_s)\!=\!0$  for  $t\neq\,s.$  In general,  $\gamma(1)$  for arbitrary t is given by:

$$\begin{aligned} & \operatorname{cov}(y_{t-1}, y_t) \!=\! \operatorname{cov}(\sum_{j=1}^{t-1} \epsilon_j, \sum_{j=1}^{t} \epsilon_j) \\ & = \! (t-1) \sigma_{\epsilon}^2 \end{aligned} \tag{1.14}$$

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

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

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)  $var(y_{t-k}) = var(y_t) = \gamma(0)$ , 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),  $var(y_{t-k}) \neq var(y_t)$ , leading to the following variation:

$$\rho(\mathbf{k}) = \frac{\operatorname{cov}(\mathbf{y}_{t-\mathbf{k}}, \mathbf{y}_t)}{\sqrt{\operatorname{var}(\mathbf{y}_{t-\mathbf{k}})\operatorname{var}(\mathbf{y}_t)}}$$

$$= \frac{(t-\mathbf{k})\sigma_{\epsilon}^2}{\sqrt{((t-\mathbf{k})\sigma_{\epsilon}^2)(t\sigma_{\epsilon}^2)}}$$

$$= \sqrt{(1-\mathbf{k}/t)}$$
(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  $\rho(k) \rightarrow 1$  as  $t \rightarrow \infty$ .

### 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  $\varepsilon_t \sim \operatorname{niid}(0, 1)$ , 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:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{1.17}$$

Thus, ceteris paribus, the increment/decrement to the random walk each period is  $\mu$ , and the sign of  $\mu$  will determine the direction of the drift subject to the realisation of  $\epsilon_t$ . 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:

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

The direction to the sample path of  $y_t$  is imparted by the term  $\mu t$ , and the random walk generated by the cumulated sum of shocks will, depending on the sign and magnitude of  $\mu$ , 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  $\varepsilon_t$  is unity, the drift coefficient is in units of  $\sigma_{\varepsilon}$ . The positive drift to the random walk becomes clearer as  $\mu$  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<sub>t</sub> and y<sub>t-1</sub>; 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  $\Theta = [t_0 = 0, t_1, ..., t_{N-1}, t_N = 1]$ , 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  $\Delta y_t$  is therefore  $(\Delta t)\sigma_{\varepsilon}^2$ , a choice which ensures that if  $\Delta t = 1$  then the step size is  $\varepsilon_t$ , as in the standard random walk.

The random walk is now:

$$\mathbf{y}_{t_i} = \mathbf{y}_{t_{i-1}} + (\sqrt{\Delta t})\boldsymbol{\varepsilon}_t \tag{1.19}$$

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



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  $\Delta t \rightarrow 0$  have any characteristics of interest. The answer is yes, but we first need a limiting result. Define a scaled version of  $y_{t_i}$  as follows:

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

If  $\epsilon_t \sim iid(0, \sigma_{\epsilon}^2)$  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:

$$Y_{t_j} \Rightarrow_D N(0, t_j)$$

$$= \sqrt{t_j} N(0, 1)$$
(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  $\epsilon_t$  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  $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$  is sufficient rather than necessary for (1.21) to hold. The CLT still goes through

if  $\{\epsilon_t\}_{t=1}^T$  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  $\sigma_{\epsilon}\sqrt{N}$  has a limiting normal distribution with variance  $var(y_{t_j})=t_j$ ; 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:

$$Z_{t_j} \equiv \frac{y_{t_j}}{\sigma_{\epsilon} \sqrt{t_j} \sqrt{N}} \Rightarrow_D N(0, 1)$$
(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) ~ 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:

$$B(t) - B(s) \sim N(0, t - s)$$
  
=  $\sqrt{(t - s)}N(0, 1)$  (1.23)

where  $0 \le s < t$ ; 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  $\Delta t$  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

 $\epsilon_t \sim \, iid(0,\,\sigma_\epsilon^2)$ , 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  $\Delta t \rightarrow 0$  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,  $\Delta t \rightarrow 0$ , 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  $0 \le r \le 1$ . 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  $\Delta t \rightarrow 0$ , 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, ..., T, can be written equivalently as follows:

$$y_{\rm T}(r) = \sum_{t=1}^{[rT]} \varepsilon_t$$
 (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, ..., 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:

$$Z_{\rm T}(\mathbf{r}) \equiv \frac{\mathbf{y}_{\rm T}(\mathbf{r})}{\sigma_{\rm \epsilon}\sqrt{T}} \tag{1.26}$$

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

$$Z_{\rm T}(\mathbf{r}) \Rightarrow_{\rm D} \mathbf{B}(\mathbf{r}) \tag{1.27}$$

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  $\sigma_{\epsilon}^2$ , that is, define, say,  $v_T(r) \equiv y_T(r)/\sqrt{T}$ , then:

$$v_{\rm T}(r) \Rightarrow_{\rm D} W(r) = \sigma_{\epsilon} B(r)$$
 (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  $\mathbf{x}_{T}$  and the continuous function f(.), then the CMT states that:

if 
$$x_T \Rightarrow_D x$$
 and  $P(x \in D_g) = 0$ , then  $f(x_T) \Rightarrow_D f(x)$  (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:  $D \mapsto \mathfrak{R}$ , and which is continuous apart from a set of measure zero, if  $Z_T(r) \Rightarrow_D B(r)$ , then:

$$g(Z_{\rm T}(\mathbf{r})) \Rightarrow_{\rm D} g(\mathbf{B}(\mathbf{r})) \tag{1.30}$$

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:

$$\mathbf{y}_t = \rho \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{1.31}$$

where  $\{\varepsilon_t\}_{t=1}^T$  is assumed to be a sequence of iid random variables, with zero mean and constant variance, written  $\varepsilon_t \sim iid(0, \sigma_{\varepsilon}^2)$ . 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  $\rho$ , large negative values of which will be inconsistent with H<sub>0</sub>. The quantities needed to construct  $\hat{\delta}$  are given as follows:

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

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

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

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

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

$$= \frac{\sum_{t=1}^{T} y_{t-1} \varepsilon_t / T}{\sum_{t=1}^{T} y_{t-1}^2 / T^2}$$
(1.33)

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:

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

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

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

$$\hat{\delta} \Rightarrow_{D} \frac{\int_{0}^{1} W(r) dW(r)}{\int_{0}^{1} W(r)^{2} dr} \equiv F(\hat{\delta})$$

$$= \frac{1}{2} \frac{[B(1)^{2} - 1]}{\int_{0}^{1} B(r)^{2} dr}$$
(1.37)

The second line uses (1.35) and (1.36), so that  $\sigma_\epsilon^2$  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  $\varepsilon_t$ ,  $t=1, \ldots, T$ . 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  $\phi_1$  rather than  $\rho$ . 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  $(1 - \phi_1 L - \phi_2 L^2)y_t = \epsilon_t$ ; and the ARMA(2, 1) model, which is of order 2 in its AR component and 1 in its MA component, is written as  $(1 - \phi_1 L - \phi_2 L^2)y_t = (1 + \theta_1 L)\epsilon_t$ . In general the ARMA(p, q) model is represented compactly as follows:

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

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

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

where  $\mu_t$  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  $\mu_t$ .

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

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

$$\phi(\mathbf{L})\mathbf{z}_t = \theta(\mathbf{L})\boldsymbol{\varepsilon}_t \tag{1.39b}$$

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

From the perspective of this model, the AR polynomial  $\phi(L)$  has been factored as the product of two polynomials  $(1 - \rho L)$  and  $\phi(L)$ , one of which is a firstorder 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  $\sum_{i=1}^{p} \phi_i = 1$ .

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  $\delta_1$  not  $\rho$  that is the root. When  $\rho = 1$  there is no contradiction in referring to  $\rho$  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  $d \ge 1$ , 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

	I(0)	I(1)
variance, $var(y_t)$	$var(\boldsymbol{y}_t) \text{ is finite}$	$var(y_t) \to \ \infty \ as \ t \to \ \infty$
autocorrelations, $\rho(k)$	invariant to translation of time axis $\rho(k) \rightarrow 0$ for k large enough; finite sum	not invariant to translation of time axis $\rho(k) \rightarrow \ 1 \ for \ all \ k \ as \ t \rightarrow \infty;$ 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 $E(\boldsymbol{y}_t)$	finite	infinite

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

*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 station-arity 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  $P_{g,t}^d$  and  $P_{g,t}^f$ , respectively. Then LOOP implies  $P_{g,t}^d = E_t \times P_{g,t}^f$ ; 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  $P_{g,t}^d$  to be greater than  $E_t \times P_{g,t}^f$  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  $P_t^d$  and  $P_t^f$  be suitably defined aggregate price indices for the domestic and foreign countries, respectively, then PPP states that the following relationship should hold:

$$E_t = A \left( \frac{P_t^d}{P_t^f} \right)$$
(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  $P_t^d$  and  $P_t^f$ . If A=1, then PPP implies  $P_t^d = E_t P_t^f$ , 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:

$$E_t = A \left(\frac{P_t^d}{P_t^f}\right)^{\delta}$$
(1.41)

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

$$ln E_{t} = ln A + \delta ln \left(\frac{P_{t}^{d}}{P_{t}^{f}}\right)$$
$$= ln A + \delta (ln P_{t}^{d} - ln P_{t}^{f})$$
(1.42)

$$\mathbf{e}_t = \mathbf{a} + \delta(\mathbf{p}_t^d - \mathbf{p}_t^t) \tag{1.43}$$

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:

$$\mathbf{e}_t - (\mathbf{p}_t^d - \mathbf{p}_t^f) = \mathbf{a} \tag{1.44}$$

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

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

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 ret, 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 ret 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 tests, which is the ability of a test to find in favour of the alternative when the alternative is true.