

# Fraktionale Integration und Kointegration in Theorie und Praxis

Inauguraldissertation

zur Erlangung des akademischen Grades  
eines Doktors der Wirtschaftswissenschaften  
an der Wirtschaftswissenschaftlichen Fakultät  
der Julius-Maximilians-Universität Würzburg

Vorgelegt von  
Diplom-Kaufmann Andreas Dechert

Würzburg, im Juli 2014

Erstgutachter:

Prof. Dr. Martin Kukuk

Zweitgutachter:

Prof. Toker Doganoglu, Ph.D.

## Zusammenfassung

Das Ziel der Arbeit ist eine Zusammenfassung über den Stand der Forschung über das Thema der fraktionalen Integration und Kointegration sowie Weiterentwicklungen der aktuellen Methoden im Hinblick darauf, dass sie robuster auf eine Reihe von empirischen Gegebenheiten anwendbar sind. Hierzu wurden insbesondere die Möglichkeiten von Strukturbrüchen in deterministischen Prozessanteilen vorgeschlagen sowie deren Auswirkungen auf Schätzeigenschaften analysiert. Mit diesem Wissen können Schätzstrategien entwickelt werden, die auch im empirischen Teil der Arbeit angewandt wurden. Der Aufbau der Arbeit gestaltet sich so, dass nach der Einleitung und Problemstellung im zweiten Kapitel der Arbeit zunächst in die Zeitreihenanalyse eingeführt wird. Hierbei wird auch eine intuitive Motivation für die Betrachtung von Long-Memory-Prozessen gegeben. Diese gestaltet sich so, dass der klassischerweise als ganzzahlig angenommene Integrationsgrad eines Prozesses nun jede beliebige Zahl, also auch Brüche, annehmen kann. Diese Annahme führt wiederum dazu, dass hiermit sehr langfristige Abhängigkeiten von Zeitreihen effizient beschrieben werden können, da diese lediglich von einem einzigen Parameter abhängen. Die Schätzung dieses nunmehr fraktionalen Integrationsgrads wird im dritten Kapitel ausführlich beschrieben und in mehreren Simulationsstudien ausgiebig analysiert. Hierzu werden neben parametrischen Schätzmethoden, die einer genauen Spezifizierung der Korrelationsstruktur von Zeitreihen bedürfen, auch semi-parametrische Methoden angeführt, die in der Praxis robuster einsetzbar sind, da ihre Schätzgenauigkeit und Effizienz nicht von einer korrekten Klassifizierung von sog. Short-Memory-Komponenten beeinflusst werden. Die Analyse dieser Methode erfolgt in erster Linie im Hinblick auf eine empirische Anwendbarkeit und bietet auch als Ergebnis Empfehlungen für eine optimale Schätzstrategie.

Das vierte Kapitel beschäftigt sich in erster Linie mit Integrationstests wie z.B. Einheitswurzeltests und deren Anwendbarkeit bei Existenz von Long-Memory-Prozessbestandteilen. Darüber hinaus werden auch Schätz- und Testmethoden für das Vorliegen von deterministischen Trends thematisiert, die wiederum auch die Möglichkeit von Strukturbrüchen zulassen. Eine multivariate Betrachtungsweise ermöglicht das fünfte Kapitel mit der Einführung der fraktionalen Kointegration. Auch hier liegt der Fokus der Arbeit darin, die empirische Anwendbarkeit zu verbessern, indem in Simulationsstudien Effekte von empirischen Gegebenheiten – wie Strukturbrüche – analysiert und optimale Schätzstrategien entwickelt werden. Im sechsten Kapitel der Arbeit wird im Rahmen der ökonomischen Theorie der Markterwartungshypothese die Verzinsung deutscher im Zeitraum Oktober 1998 bis No-

vember 2011 untersucht. Diese Hypothese impliziert, dass zwischen den einzelnen Zinssätzen eine multivariate Beziehung in Form von Kointegrationsbeziehungen bestehen sollte, da die Zinssatzdifferenzen einer Liquiditätsprämie entsprechen. Von dieser wurde in bisherigen Studien angenommen, dass sie stationär ist, d.h. dass sie allenfalls eine Short-Memory-Eigenschaft aufweist, welche nur relativ kurzfristige Abhängigkeiten impliziert. Von dieser Sichtweise löst sich die Arbeit, indem sie die Möglichkeit von fraktionalen Kointegrationsbeziehungen ermöglicht, die eine Aussage über die Persistenz der Liquiditätsprämie ermöglicht. Im Rahmen dieser Analyse konnten eine Reihe interessanter Erkenntnisse gewonnen werden, wie z.B. dass das Ausmaß der Persistenz (d.h. die Trägheit der Anpassung auf ökonomische Schocks) mit ansteigender Laufzeitdifferenz sukzessive größer wird und auch nicht mehr durch klassisch angenommene Prozesstrukturen erklärt werden kann. Nichtsdestotrotz können die Ergebnisse der empirischen Analyse die Annahmen der Markterwartungshypothese nicht bestätigen, da insbesondere der Integrationsgrad für sehr lange Laufzeitdifferenzen so großausfällt, dass selbst eine relativ schwache fraktionale Kointegrationsbeziehung abgelehnt werden muss.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>ix</b>
<b>Tabellenverzeichnis</b>	<b>xii</b>
<b>Abkürzungsverzeichnis</b>	<b>xiii</b>
<b>1 Einleitung</b>	<b>1</b>
<b>2 Zeitreihenmodelle</b>	<b>4</b>
2.1 Grundlegende Definitionen . . . . .	4
2.1.1 Stochastische Prozesse . . . . .	4
2.1.2 Der Stationaritätsbegriff . . . . .	5
2.1.3 Integrierte Prozesse . . . . .	5
2.2 Lineare stochastische Prozesse . . . . .	6
2.2.1 Moving-Average-Prozesse . . . . .	7
2.2.2 Autoregressive Prozesse . . . . .	9
2.2.3 Gemischte ARMA-Prozesse . . . . .	13
2.3 Frequenzanalyse von Zeitreihen . . . . .	15
2.3.1 Das Spektrum . . . . .	16
2.3.2 Das Periodogramm . . . . .	20
2.3.3 Das Periodogramm als Schätzung für die Spektraldichte . . . . .	22
2.4 Gedächtnis von Zeitreihen . . . . .	24
2.5 Fraktional integrierte Prozesse . . . . .	24
2.5.1 Fraktional differenziertes Rauschen . . . . .	25
2.5.2 ARFIMA( $p, d, q$ )-Prozesse . . . . .	32
2.5.3 Alternative Long-Memory-Modelle . . . . .	34
2.5.4 Die empirische Relevanz von Long-Memory-Modellen . . . . .	35
<b>3 Schätzverfahren zur Bestimmung des Integrationsgrads</b>	<b>39</b>
3.1 Parametrische Maximum-Likelihood-Schätzung . . . . .	39
3.1.1 Die Cholesky-Zerlegung . . . . .	40
3.1.2 Die Berechnung der Kovarianzmatrix . . . . .	40
3.1.3 Autoregressive Approximation nach Hasslett und Raftery . . . . .	41
3.1.4 Tanaka-Approximation . . . . .	43
3.1.5 Der Whittle-Schätzer . . . . .	48
3.1.6 Vergleich der parametrischen Verfahren . . . . .	50

3.1.7	Parameterwahl und Informationskriterien . . . . .	55
3.2	Semiparametrische Schätzverfahren . . . . .	65
3.2.1	Das Geweke/Porter-Hudak-Verfahren . . . . .	67
3.2.2	Lokale Whittle-Schätzung . . . . .	70
3.2.3	Exakte lokale Whittle-Schätzung . . . . .	72
3.2.4	Vergleich der semiparametrischen Verfahren . . . . .	74
3.2.5	Wahl der Bandbreite $m$ . . . . .	81
<b>4</b>	<b>Spezifikationstests für fraktional integrierte Prozesse</b>	<b>93</b>
4.1	Tests für deterministische Trends . . . . .	93
4.1.1	Der Fall für lineare Trends . . . . .	96
4.1.2	Strukturbruch eines linearen Trends . . . . .	102
4.1.3	Vergleich des Tests für Trendbruch mit dem Meanchange-Test nach Shao (2011) . . . . .	107
4.2	Herkömmliche Einheitswurzeltests . . . . .	112
4.2.1	Der Test nach Dickey-Fuller . . . . .	113
4.2.2	Der KPSS-Test . . . . .	116
4.3	Lagrange-Multiplikatoren Tests . . . . .	119
4.4	Wald-Tests . . . . .	125
4.4.1	Der fraktionale Dickey-Fuller-Test . . . . .	125
4.4.2	Der effiziente Wald-Test nach Lobato und Velasco (2007) . . .	131
4.5	Varianzverhältnistest-Ansatz . . . . .	139
4.6	Vergleichende Monte-Carlo-Simulation . . . . .	150
<b>5</b>	<b>Fraktionale Kointegration</b>	<b>155</b>
5.1	Begriff der fraktionalen Kointegration . . . . .	155
5.2	Schätzung der Kointegrationsbeziehung . . . . .	157
5.3	Bestimmung des Integrationsgrads von Kointegrationsfehlern . . . . .	162
5.3.1	Der ELW-Schätzer im Kointegrationsfall . . . . .	163
5.3.2	Der effiziente Wald-Test im Kointegrationsfall . . . . .	166
5.3.3	Vergleichende Monte-Carlo-Studie . . . . .	169
5.4	Multivariate Schätzverfahren von fraktional kointegrierten Systemen .	172
5.4.1	Exakte lokale Whittle-Schätzung im Falle der fraktionalen Ko- integration . . . . .	172
5.4.2	Das allgemeine fraktionale Fehlerkorrekturmodell und dessen Schätzung . . . . .	177

<b>6</b>	<b>Fraktionale Kointegration und die Markterwartungshypothese</b>	<b>189</b>
6.1	Die Markterwartungshypothese und Kointegration . . . . .	190
6.2	Überprüfung anhand täglicher Bundesanleihendaten . . . . .	193
6.2.1	Univariate Analyse . . . . .	194
6.2.2	Multivariate Modellierung der Zinssätze . . . . .	203
<b>7</b>	<b>Schlussfolgerungen und Ausblick</b>	<b>217</b>
	<b>Literatur</b>	<b>224</b>

## Abbildungsverzeichnis

2.1	Beispiele für ein Weißes Rauschen und einen Random Walk . . . . .	5
2.2	Beispiele für MA(1)-Prozesse . . . . .	8
2.3	(Partielle) Autokorrelationsfunktion von MA(1)-Prozessen . . . . .	9
2.4	Beispiele für AR(1)-Prozesse . . . . .	10
2.5	(Partielle) Autokorrelationsfunktion von AR(1)-Prozessen . . . . .	11
2.6	(Partielle) Autokorrelationsfunktion von ARMA(1,1)-Prozessen . . . . .	14
2.7	Kosinusschwingung in Abhängigkeit der jeweiligen Fourierfrequenzen . . . . .	16
2.8	Logarithmierte Spektren von ausgewählten AR(1)-Prozessen . . . . .	17
2.9	Logarithmierte Spektren von ausgewählten MA(1)-Prozessen . . . . .	19
2.10	Logarithmierte Spektren von ausgewählten ARMA(1,1)-Prozessen . . . . .	20
2.11	Schätzung und Anpassungen des Periodogramms für einen saisonalen AR-Prozess . . . . .	23
2.12	Gewichte der MA-Darstellung $\pi_j$ im Vergleich . . . . .	26
2.13	Beispiele für fraktionale Rauschen . . . . .	28
2.14	Logarithmierte Spektren von fraktionalen Rauschen . . . . .	29
2.15	(Partielle) Autokorrelationsfunktion von fraktionalen Rauschen . . . . .	30
2.16	Gewichte der MA-Darstellung von nicht-stationären fraktionalen Rau- schen . . . . .	31
2.17	Klassifizierung der Prozesseigenschaften in Abhängigkeit des Integra- tionsgrads . . . . .	32
2.18	Zusammensetzung des ARFIMA(1, $d$ , 1)-Spektrums . . . . .	33
2.19	Tägliche Handelsmenge der E.on-Aktie im Zeitraum November 2005 bis März 2014 . . . . .	36
2.20	ACF der täglichen Handelsmenge der E.on-Aktie . . . . .	37
2.21	Periodogramm der täglichen Handelsmenge der E.on-Aktie . . . . .	38
3.1	Geschätzte Dichtefunktionen für fraktionale Rauschen mit $d = 0,4$ und $d = 1,4$ . . . . .	44
3.2	Geschätzte Dichtefunktionen der Schätzergebnisse für ARFIMA(1, $d$ , 0)- Prozesse mit $\theta = 0,5$ und $d = 0,4$ bzw. $d = 1,4$ . . . . .	45
3.3	Streudiagramme der Schätzergebnisse $\hat{d}$ und $\hat{\theta}$ . . . . .	47
3.4	Verzerrung und RMSE von $\hat{d}$ für ARFIMA(0,0,0)-Prozesse in Abhän- gigkeit von $n$ . . . . .	51
3.5	Verzerrung und RMSE von $\hat{d}$ für ARFIMA(0, $d$ , 0)-Prozesse in Abhän- gigkeit von $d$ . . . . .	52
3.6	Verzerrung und RMSE von $\hat{d}$ für ARFIMA(1,0,0)-Prozesse in Abhän- gigkeit von $\theta$ . . . . .	53
3.7	GPH-Schätzung für $d = -0,4$ . . . . .	70
3.8	GPH-Schätzung für $d = 0,4$ . . . . .	70
3.9	Gewichtungsfunktion $w(d)$ für $d \in [1/2; 3/4]$ . . . . .	73
3.10	Verzerrung und RMSE in Abhängigkeit von $n$ mit $m = n^{0,8}$ und $d = 0$ . . . . .	75
3.11	Geschätzte Dichtefunktionen für $\hat{d}_{GPH}$ , $\hat{d}_{LW}$ und $\hat{d}_{ELW}$ . . . . .	76
3.12	Optimale Bandbreiten in Abhängigkeit des AR(1)-Koeffizienten $\theta$ und Stichprobenumfang $n$ . . . . .	77

3.13	Optimale Bandbreiten in Abhängigkeit des AR(1)-Koeffizienten $\theta$ für das LW- und ELW-Verfahren mit Stichprobenumfang $n = 500$ . . . . .	78
3.14	Vergleich der Schätzverfahren für ARFIMA(1, $d$ , 0)-Prozesse . . . . .	79
3.15	Geschätzte Bandbreiten für ARFIMA(1, $d$ , 0)-Prozesse . . . . .	84
3.16	Schätzeigenschaften von GPH, LW und ELW bei geschätzten Bandbreiten für ARFIMA(1, $d$ , 0)-Prozesse . . . . .	85
3.17	Geschätzte Bandbreiten für ARFIMA(1, 0, 1)-Prozesse bei GPH und LW . . . . .	87
3.18	Geschätzte Bandbreiten für ARFIMA(1, 0, 1)-Prozesse bei ELW . . . . .	88
4.1	Durchschnittliche Verzerrung bei der Schätzung von $d$ bei Vorliegen eines linearen Trends . . . . .	94
4.2	Durchschnittliche Verzerrung bei der Schätzung von $d$ bei Vorliegen eines gebrochenen Trends (ELW) bzw. gebrochenen Mittelwertes (GPH) . . . . .	96
4.3	Geschätzte Dichte- und Verteilungsfunktionen für die lineare Teststatistik $R_{LT}^2$ . . . . .	97
4.4	Abhängigkeit der 5%-kritischen-Werte für $R_{LT}^2$ und $n$ . . . . .	98
4.5	Asymptotische kritische Werte für $R_{LT}^2$ in Abhängigkeit von $d$ . . . . .	99
4.6	Beispiel der Anwendung der linearen Trendstatistik $R_{LT}^2$ . . . . .	100
4.7	Vergleich der Annahmewahrscheinlichkeiten für einen linearen Trend für $\mu_1 = 0$ und $n = 250$ . . . . .	101
4.8	Asymptotisches Verhalten der 5%-kritischen-Werte der t-Statistik für $\mu_2$ . . . . .	103
4.9	Struktur der 5%-kritischen-Werte für $R_{TB}^2$ und t-Stat( $\hat{\mu}_2$ ) <sub>norm</sub> . . . . .	104
4.10	Beispiel des Tests für einen gebrochenen Trend . . . . .	106
4.11	Vergleich der kritischen Werte nach Shao (2011) mit den geschätzten Response-Surface-Funktionen . . . . .	109
4.12	Macht des TB-Tests gegen MC . . . . .	111
4.13	Relative Ablehnungshäufigkeiten für $\beta = 1$ in Abhängigkeit der Wahl von $p$ . . . . .	114
4.14	Dichtefunktion des KPSS-Tests für $n = 2000$ . . . . .	117
4.15	Relative Ablehnungshäufigkeiten für $d = 0$ in Abhängigkeit von $l$ . . . . .	119
4.16	Geschätzte Dichtefunktion der LM-Statistiken . . . . .	123
4.17	Vergleich der simulierten kritischen Werte für $\alpha = 5\%$ mit den jeweiligen Response-Surface-Funktionen . . . . .	124
4.18	Geschätzte Dichtefunktion des fraktionalen Dickey-Fuller-Test mit bekanntem und unbekanntem $d$ . . . . .	127
4.19	Kritische Werte für die Situation mit Trendbruch in Abhängigkeit von $\tau$ . . . . .	128
4.20	5% Kritische Werte des fraktionalen Dickey-Fuller-Tests in Abhängigkeit von $n$ . . . . .	130
4.21	Geschätzte Dichtefunktion der effizienten Wald-Statistik mit unbekanntem $d$ . . . . .	135
4.22	5%-kritische-Werte des effizienten Wald-Tests in Abhängigkeit von $n$ . . . . .	137
4.23	Struktur der 5%-kritische-Werte für VR-Statistik in Abhängigkeit von $d$ und des Tuning-Parameters $d_1$ für $n = 1000$ . . . . .	141
4.24	Anpassung der Gammaverteilung für die VR-Statistik mit $d_1 = 0,1$ . . . . .	145

4.25	Anpassung der Gammaverteilung für die VR-Statistik im relevanten kritischen Bereich (für $d_1 = 1$ auf logarithmierter Basis) . . . . .	146
4.26	Struktur der 5%-kritische-Werte für die VR-Statistik ohne Trendbeseitigung in Abhängigkeit von $d$ und des Beobachtungsumfangs $n$ . . .	147
4.27	Struktur der 5%-kritischen-Werte der VR-Statistik in Abhängigkeit von $\tau$ für $d_1 = 0,1$ und $d_1 = 1$ . . . . .	149
5.1	Beispiele für bivariate fraktionale Kointegrationsbeziehungen . . . . .	157
5.2	Geschätzte Dichtefunktionen des ELW-Schätzergebnisse der Kointegrationsresiduen für $\beta = 0$ . . . . .	164
5.3	Geschätzte Dichtefunktionen für den effizienten Wald-Test im Kointegrationsfall . . . . .	166
5.4	Kritische Werte für den effizienten Wald-Test im Kointegrationsfall im Vergleich zum univariaten Integrationsfall . . . . .	167
5.5	Optimale Bandbreite für den ELW-Schätzer für Kointegration in Abhängigkeit von $\theta$ . . . . .	174
5.6	Vergleich der Abbildungen des Einheitskreises nach $\mathbb{C}_d$ . . . . .	179
5.7	Geschätzte Dichtefunktionen für $\beta$ und $d$ . . . . .	183
6.1	Verlauf der Renditen deutscher Bundesanleihen im Zeitraum August 1997 bis Mai 2011 . . . . .	193
6.2	(Partielle) Autokorrelationsfunktion der Zinssätze . . . . .	197
6.3	Anpassung der theoretischen Spektren an die geschätzten Periodogramme auf log-Basis . . . . .	200
6.4	Zinsdifferenzen deutscher Bundesanleihen verschiedener Laufzeiten . . .	203
6.5	(Partielle) Autokorrelationsfunktion der Zinsdifferenzen . . . . .	205
6.6	ELW-Schätzergebnisse für den Integrationsgrad der Zinsdifferenzen in Abhängigkeit der Bandbreite $m$ . . . . .	207

## Tabellenverzeichnis

3.1	Verteilungsparameter von $\hat{d}$ bei fraktionalen Rauschen . . . . .	44
3.2	Verteilungsparameter der geschätzten Parameter bei ARFIMA(1, $d$ , 0)- Prozessen mit $d = 0,4$ . . . . .	46
3.3	Verteilungsparameter der geschätzten Parameter bei ARFIMA(1, $d$ , 0)- Prozessen mit $d = 1,4$ . . . . .	46
3.4	Schätzeigenschaften von $\hat{d}$ bei ARFIMA(1,0,1)-Prozessen mit $n = 250$	54
3.5	Schätzeigenschaften von $\hat{d}$ bei ARFIMA(1,0,1)-Prozessen mit $n = 500$	54
3.6	Schätzeigenschaften in Abhängigkeit der Wahl $p$ und $q$ für ARFIMA(0,0,0)- Prozesse . . . . .	56
3.7	Schätzeigenschaften in Abhängigkeit der Wahl $p$ und $q$ für ARFIMA(1,0,0)- Prozesse mit $\theta = 0,5$ . . . . .	56
3.8	Gewählte Parameterordnungen $p$ und $q$ für $n = 250$ unter Verwen- dung der Whittle-Approximation . . . . .	58
3.9	Gewählte Parameterordnungen $p$ und $q$ für $n = 500$ unter Verwen- dung der Whittle-Approximation . . . . .	59
3.10	Gewählte Parameterordnungen $p$ und $q$ für $n = 250$ unter Verwen- dung der HR-Approximation . . . . .	60
3.11	Gewählte Parameterordnungen $p$ und $q$ für $n = 500$ unter Verwen- dung der HR-Approximation . . . . .	61
3.12	Relative Selektionshäufigkeiten des wahren Modells bei $n = 250$ . . .	63
3.13	Relative Selektionshäufigkeiten des wahren Modells bei $n = 500$ in Abhängigkeit von $d$ . . . . .	63
3.14	Durchschnittliche Verzerrung von $\hat{d}$ bei Verwendung von Modellselek- tionskriterien $n = 250$ in Abhängigkeit von $d$ . . . . .	64
3.15	Durchschnittliche Verzerrung von $\hat{d}$ bei Verwendung von Modellselek- tionskriterien $n = 500$ in Abhängigkeit von $d$ . . . . .	65
3.16	RMSE von $\hat{d}$ bei Verwendung von Modellselektionskriterien $n = 250$ in Abhängigkeit von $d$ . . . . .	66
3.17	RMSE von $\hat{d}$ bei Verwendung von Modellselektionskriterien $n = 500$ in Abhängigkeit von $d$ . . . . .	66
3.18	Verteilungsparameter für das GPH-, LW- und ELW-Verfahren für $n =$ 1000 . . . . .	76
3.19	Beispiel der parametrischen Bandbreitenschätzung für das ELW-Verfahren bei einem ARFIMA(1, 0, 0)-Prozess mit $\theta = 0,5$ und $n = 500$ . . . . .	83
3.20	Schätzeigenschaften für das GPH-Verfahren bei ARFIMA(1,0,1)-Prozessen mit $n = 500$ . . . . .	89
3.21	Schätzeigenschaften für das LW-Verfahren bei ARFIMA(1,0,1)-Prozessen mit $n = 500$ . . . . .	90
3.22	Schätzeigenschaften für das ELW-Verfahren bei ARFIMA(1,0,1)-Prozessen mit $n = 500$ . . . . .	91
3.23	Schätzeigenschaften für das ELW-Verfahren bei ARFIMA(1,1,1)-Prozessen mit $n = 500$ . . . . .	92
4.1	Response-Surface-Parameter für $R_{LT,(1-\alpha)}^2$ . . . . .	99

4.2	Kritische Werte für $R_{LT}^2$ . . . . .	100
4.3	Response-Surface-Parameter für $R_{TB,(1-\alpha)}^2$ . . . . .	105
4.4	Response-Surface-Parameter für t-Stat $(\hat{\mu}_2)_{\text{norm},(1-\alpha)}$ . . . . .	105
4.5	Kritische Werte für $R_{TB}^2$ und t-Stat $(\hat{\mu}_2)_{\text{norm}}$ . . . . .	107
4.6	Response-Surface-Parameter für $MC(d, \epsilon = 0,15)_{(1-\alpha)}$ . . . . .	108
4.7	Durchschnittliche mit TB und MC geschätzte relative Bruchpunkte $\hat{\tau}$ für $n = 100$ mit $\mu_1 = 0$ . . . . .	110
4.8	Durchschnittliche mit TB und MC geschätzte relative Bruchpunkte $\hat{\tau}$ für $n = 250$ mit $\mu_1 = 0$ . . . . .	110
4.9	Relative Ablehnungshäufigkeiten für $\beta = 1$ in Abhängigkeit von $n$ . . . . .	114
4.10	Asymptotische kritische Quantile des KPSS-Tests . . . . .	118
4.11	Relative Ablehnungshäufigkeiten für $H_0 : d = 0$ bei Verwendung des KPSS-Tests . . . . .	119
4.12	Übersicht der Verteilungscharakteristiken und der kritischen Werte für $\alpha = 5\%$ bei unbekanntem und bekannten Integrationsparametern . . . . .	123
4.13	Response-Surface-Parameter für LM . . . . .	125
4.14	Behandlung von deterministischen Bestandteilen für $y_t$ für den Test von $H_0 : d = 1$ gegen $H_1 : d < 1$ . . . . .	126
4.15	Übersicht der kritischen Werte für $\alpha = 5\%$ bei unbekanntem und bekanntem Integrationsparametern für die FDF-Statistik . . . . .	129
4.16	Wahl des Polynomgrades von $n^{-1}$ für die Response-Surface-Funktion der FDF-Statistik ohne Trendbereinigung für $\alpha = 5\%$ . . . . .	130
4.17	Geschätzte Response-Surface-Funktionen für die kritischen Werte des fraktionalen Dickey-Fuller-Tests . . . . .	132
4.18	Übersicht der kritischen Werte für $\alpha = 5\%$ bei unbekanntem Integra- tionsparameter für die effiziente Wald-Statistik und $n = 1000$ . . . . .	135
4.19	Wahl des Polynomgrades von $n^{-1}$ für die Response-Surface-Funktion der NEW-Statistik ohne Trendbereinigung . . . . .	136
4.20	Vergleich der asymptotischen kritischen Werte mit Sephton (2009) . . . . .	136
4.21	Geschätzte Response-Surface-Funktionen für die kritischen Werte des effizienten Wald-Tests . . . . .	138
4.22	Ablehnungsraten für $H_0 : d_0 = d$ bei Verwendung von $\rho(d_1)$ mit $\alpha = 5\%$ . . . . .	143
4.23	Ablehnungsraten für $H_0 : d_0 = d$ bei Verwendung von $\rho(d_1)$ mit AR(1)-Störtermen ( $\theta = 0,5$ ) und $\alpha = 5\%$ . . . . .	144
4.24	Geschätzte Response-Surface-Funktionen für die kritischen Werte des univariaten VR-Tests mit $d_1 = 0,1$ und $d_1 = 1$ . . . . .	148
4.25	Ablehnungsraten für $H_0 : d = 1$ bei Verwendung von fraktionalem Integrationstests für $\theta = 0$ und $\theta = 0,3$ . . . . .	151
4.26	Ablehnungsraten für $H_0 : d = 1$ bei Verwendung von fraktionalem Integrationstests für $\theta = 0,5$ und $\theta = 0,8$ . . . . .	152
5.1	Durchschnittliche Verzerrung und RMSE für $\hat{\beta}$ bei Anwendung des NBL- und OLS-Schätzers für $n = 500$ und $d = 1$ . . . . .	160
5.2	Simulierte Quantile für den geschätzten Integrationsgrad der Kointe- grationsresiduen . . . . .	164

5.3	Durchschnittliche Verzerrung und RMSE für $\hat{b}$ bei Anwendung des ELW-Schätzers für die Kointegrationsresiduen . . . . .	165
5.4	Geschätzte Response-Surface-Funktionen für die Kritische Werte des effizienten Wald-Tests für den Kointegrationsfall . . . . .	168
5.5	Vergleich der asymptotischen kritischen Werte für den effizienten Wald-Test im Integrations- und Kointegrationsfall . . . . .	169
5.6	Ablehnungsraten für keine Kointegration ( $b = 0$ ) bei Anwendung verschiedener Integrationstests und fixiertem Irrtumsniveau $\alpha = 5\%$ . . .	170
5.7	Vergleich der Schätzeigenschaften für den ELW-Schätzer in der multivariaten und univariaten Ausführung für $n = 500$ . . . . .	176
5.8	Vergleich der Schätzeigenschaften für den ELW-Schätzer bzgl. $\beta$ im Vergleich zu OLS für $n = 500$ . . . . .	177
5.9	Durchschnittliche Verzerrung und RMSE für $\hat{\beta}$ bei Schätzung des FCVAR und des klassischen CVAR . . . . .	183
5.10	Durchschnittliche Verzerrung und RMSE für $\hat{d}$ und $\hat{b}$ bei Schätzung des FCVAR-Modells . . . . .	185
5.11	Durchschnittliche Ablehnungsraten für $r = 0$ bei Anwendung des LR-Tests für (F)CVAR-Modelle . . . . .	187
6.1	Deskriptive Statistiken der Anleiherenditen . . . . .	194
6.2	Geschätzte Integrationsgrade auf Basis der ersten Differenzen . . . . .	195
6.3	Spezifikationstests für deterministische Bestandteile . . . . .	196
6.4	Übersicht der nach dem BIC-Kriterium selektierten Modelle bei Verwendung des parametrischen Whittle-Schätzers . . . . .	198
6.5	Übersicht der nach dem BIC-Kriterium selektierten Modelle für M6 bei Verwendung des parametrischen Whittle-Schätzers . . . . .	199
6.6	Geschätzte Integrationsgrade für M6 und Y10 in verschiedenen Stichproben . . . . .	202
6.7	Deskriptive Statistiken der Zinsdifferenzen . . . . .	205
6.8	Geschätzte Integrationsgrade der Zinsdifferenzen . . . . .	206
6.9	Integrationstests für die Zinsdifferenzen . . . . .	207
6.10	ELW-Schätzergebnisse für bivariate Kointegrationsbeziehungen . . . . .	208
6.11	(F)CVAR-Schätzergebnisse für bivariate Kointegrationsbeziehungen . . . . .	210
6.12	LR-Teststatistiken für den Kointegrationsparameter $\beta = 1$ für bivariate Kointegrationsbeziehungen . . . . .	212
6.13	LR-Teststatistiken für den Kointegrationsrang $r$ für das alle Renditenreihen umfassende Zeitreihensystem $(Y1, Y2, Y5, Y10, M6)$ im (F)CVAR-Kontext . . . . .	213
6.14	Geschätzte Kointegrations- und Fehlerkorrekturmatrizen für Kointegrationsrang $r = p - 1$ mit $k = 1$ für den Vektor $(Y1, Y2, Y5, Y10, M6)$ im (F)CVAR-Kontext . . . . .	214

## Abkürzungsverzeichnis

ACF	Autokorrelationsfunktion (engl. Autocorrelation Function)
ADF	Erweiterter Dickey-Fuller-Test (engl. Augmented Dickey-Fuller Test)
AIC	Akaike-Informationskriterium
AR	Autoregressiver Prozess
ARFIMA	Fraktional integrierter autoregressiver Moving-Average-Prozess
ARIMA	Integrierter autoregressiver Moving-Average-Prozess
ARMA	Autoregressiver Moving-Average-Prozess
BIC	Bayesianisches(bzw. Schwarz-)Informationskriterium
ca.	circa
Cov	Kovarianz (engl. Covariance)
CVAR	Kointegriertes vektorautoregressives Modell (engl. Cointegrated Vector Autoregressive Model)
E	Erwartungswert
ELW	Exakter lokaler Whittle-Schätzer
evtl.	eventuell
ExK	Exzess-Kurtosis, definiert als viertes standardisiertes Moment abzüglich drei
FELW	Durchführbarer exakter lokaler Whittle-Schätzer (engl. Feasible Exact Local Whittle Estimator)
FCVAR	Fraktional kointegriertes vektorautoregressives Modell (engl. Fractional cointegrated vector autoregressive model)
FDF	Fraktionaler Dickey-Fuller-Test
FR	Fraktional integriertes Rauschen
GARCH	Generalisierte autoregressive bedingte Heteroskedastizität (engl. Generalized Autoregressive Conditional Heteroscedasticity)
GPH	Geweke/Porter-Hudak-Schätzer
HR	Hasslett/Raftery-Approximation
HQC	Hannan/Quinn-Informationskriterium

$I(d)$	Integriert vom Grad $d$
$IID$	Unabhängig und identisch verteilt (engl. Independent and Identically Distributed)
KPSS	Test nach Kwiatkowski, Phillips, Schmidt und Shin
KQ	Kleinst-Quadrate-Methode
LM	Lagrange-Multiplikator
LR	Likelihood-Verhältnis (engl. Likelihood Ratio)
LT	Linearer Trend
LTB	Linearer Trend mit Bruch
LW	Lokaler Whittler-Schätzer
OLS	Ordinary Least Squares als engl. Synonym für Kleinst-Quadrate-Methode
PACF	Partielle Autokorrelationsfunktion (engl. Partial Autocorrelation Function)
QT	Quadratischer Trend
MA	Moving-Average-Prozess
MC	Mittelwertbruch (engl. Mean Change)
ML	Maximum-Likelihood
MW	Mittelwert, definiert als arithmetisches Mittel
NBLS	Narrow Band Least Squares
NEW	Neuer effizienter Wald-Test
RMSE	Wurzel des mittleren quadratischen Fehlers (engl. Root Mean Squared Error)
S	Schiefe, definiert als das dritte standardisierte Moment
SD	Standardabweichung, definiert als die Wurzel des zweiten zentrierten Moments
TB	Trendbruch-Modell, synonym verwendet für LTB
Var	Varianz, definiert als das zweite zentrierte Moment
VR	Varianzverhältnis (engl. Variance Ratio)
WKN	Wertpapierkennnummer

u.a.	unter anderen
u.U.	unter Umständen
usw.	und so weiter
z.B.	zum Beispiel
z.T.	zum Teil

# 1 Einleitung

Die hier vorliegende Arbeit lässt sich im Bereich der Zeitreihenanalyse ansiedeln. Das Ziel der Zeitreihenanalyse ist die Beschreibung einer zeitlich geordneten Folge von Beobachtungen, wie wir sie in sämtlichen wissenschaftlichen Disziplinen erhalten können. Dies kann z.B. dadurch geschehen, dass wir einen Prozess in mehrere Komponenten zerlegen können wie etwa einen deterministischen, einen zyklischen und einen stochastischen Bestandteil<sup>1</sup>. Mit zyklischen Komponenten, worunter man sich regelmäßig wiederholende Schwankungen verstehen kann, werden wir uns in dieser Arbeit nicht beschäftigen. Das Hauptaugenmerk werden wir auf die Modellierung der stochastischen Komponente legen, für die wir zum einen klassische Konzepte anführen werden als auch einen weiteren verallgemeinerten Ansatz. Dieser wiederum lässt sich dadurch motivieren, dass wir eine weiterreichende Abhängigkeitsstruktur im stochastischen Teil effizienter modellieren können als mit den klassischen Modellen.

Diese Modelle finden vor allem ihren Zweck in der Überprüfung ökonomischer Theorien, die wir auch im empirischen Teil dieser Arbeit vornehmen werden. Daraus lassen sich Aussagen und Diagnosen z.B. vom Zustand der Wirtschaft treffen, was wiederum eine wichtige Entscheidungsgrundlage für die Politik sein kann. Aus diesem Grund sollte der Beschreibung von Zeitreihen bzw. von wirtschaftlichen Prozessen ein hohes Gewicht beigemessen werden, da womöglich Fehlspezifizierungen zu falschen Rückschlüssen führen können. Letztendlich lassen sich auch Zeitreihenmodelle zur Prognose von zukünftigen Werten einsetzen und ermöglichen uns auch eine genauere Beschreibung von Eigenschaften von Zeitreihen wie etwa deren Schwankungsbreite in Form der Volatilität, was ein fundamentaler Bestandteil des Risikomanagements darstellt.

Diese Anwendungsgebiete wiederum setzen ein tieferes Verständnis des Verhaltens von Zeitreihen voraus, zu dem wir auch mit dieser Arbeit beitragen wollen. Insbesondere die Unterscheidung der Stationarität und Nicht-Stationarität von Zeitreihen blickt auf eine seit langem geführten Diskussion zurück und konnte in vielen Bereichen nicht abschließend geklärt werden. Hier vermag eine flexiblere Modellierung die Grenzen zwischen diesen beiden Polen aufzuweichen und zeigt weitere interessante Erklärungsansätze, die mit dieser Arbeit diskutiert werden sollen. Der Hauptbestandteil dieses Beitrages ist aber wohl das Aufzeigen von Schätzern und Testme-

---

<sup>1</sup>Vgl. hierzu das Vorwort zu Schlittgen und Streitberg (1994)

thoden, die eine empirische Anwendung der Modellierungsansätze ermöglichen sollen. Hierzu werden unterschiedliche Schätzstrategien bewertet und diskutiert, was einen großen Teil dieser Arbeit einnimmt. Auch sollen Schätzverfahren robust gemacht werden gegen möglicherweises Auftreten von Nicht-Linearitäten, wie sie etwa z.B. Strukturbrüche von Trendkomponenten darstellen. Alles in allem ist die Arbeit darauf fokussiert, bereits vorhandene Schätz- und Testansätze für empirische Anwendungen in einer möglichst robusten Weise anwendbar zu machen.

Den Aufbau der Arbeit wollen wir wie folgt gestalten: Im zweiten Kapitel werden grundlegende Definitionen und die am weitesten verbreiteten Zeitreihenmodelle wiederholt. Dazu werden wir kurz in die Frequenzanalyse einführen und die Eigenschaften von ARMA-Modellen im Frequenzbereich darlegen. Daraufhin werden wir das Gedächtnis von Zeitreihen definieren und mit ARFIMA-Prozessen, welche eine Verallgemeinerung der ARIMA-Modelle darstellen und als Long-Memory-Prozesse bezeichnet werden, diskutieren. Das zweite Kapitel werden wir mit Gedanken zur empirischen Relevanz von Long-Memory-Modellen beschließen.

Das dritte Kapitel widmen wir komplett der Schätzproblematik des Integrationsparameters von ARFIMA-Prozessen. Dazu werden wir sowohl parametrische als auch semiparametrische Verfahren motivieren, die wir gesondert in Monte-Carlo-Studien vergleichen werden. Auch werden wir die Problematik der Anwendung der Schätzverfahren diskutieren, welche bei den parametrischen Verfahren in der Wahl der Modellordnung liegt und für die semiparametrischen die Wahl einer geeigneten Bandbreite darstellt. Die hierzu vorgestellten Modelle werden wir ebenso mit Monte-Carlo-Simulationen analysieren, welche damit den Abschluss des dritten Kapitels bilden.

Das vierte Kapitel ist den Spezifikationstests für fraktional integrierten Prozessen gewidmet, wo wir zunächst einen Test für lineare Trends als eine Verallgemeinerung eines Testansatzes von Gómez und Ventosa-Santaulària (2010) vorschlagen werden, mit dem wir auch Strukturbrüche von linearen Trends bei gleichzeitiger nicht-stationärer fraktionaler Integration berücksichtigen können. Einen großen Teil des vierten Kapitels werden die Tests bzgl. des Integrationsgrads einnehmen, für die wir auch kritische Werte in Form von Response-Surface-Funktionen angeben. Auch hier setzt eine vergleichende Monte-Carlo-Studie, die einen Aufschluss über die Macht der einzelnen Tests geben soll, den Schlusspunkt des Kapitels.

Das fünfte Kapitel stellt das letzte der theoretischen Kapitel der Arbeit dar und übersetzt den univariaten Modellierungsansatz der fraktionalen Integration auf den

multivariaten Bereich der Kointegration. Dazu werden wir das Verhalten der univariaten Schätzer im Kointegrationsfall in Simulationsstudien untersuchen und zwar besonders im Hinblick darauf, dass die Parameter der Kointegrationsbeziehung in der Regel unbekannt sind. Auch werden wir die mögliche multivariate Erweiterung der Schätzverfahren behandeln und deren Vor- und Nachteile diskutieren. Das Kapitel wird dann mit einer Betrachtung des fraktional kointegrierten Fehlerkorrekturmodells beschlossen, das eine Verallgemeinerung des klassischen Johansen-Kointegrationsmodells darstellt.

Im sechsten und vorletzten Abschnitt wollen wir die in der Arbeit vorgestellten Methoden empirisch anwenden und betrachten hierzu die Verzinsung deutscher Bundesanleihen mit verschiedenen Laufzeiten über einen relativ langen Zeitraum. Wir werden untersuchen, inwieweit Anhaltspunkte für eine fraktionale Modellierung der univariaten Zinssätze vorliegen und auch eine multivariate Betrachtung anstrengen, da sich aufgrund der ökonomischen Theorie der Markterwartungshypothese eine Kointegration der verschieden langlaufenden Verzinsungen ergeben sollte. Diese Implikationen werden wir versuchen zu testen und eine mögliche Evidenz für fraktionale Kointegration diskutieren.

Zum Schluss werden wir noch die Erkenntnisse der Arbeit zusammenfassen und einen möglichen Ausblick für eine weitere Forschungsrichtung im Bereich der fraktionalen Integration versuchen.

## 2 Zeitreihenmodelle

In diesem Kapitel werden wir zunächst die grundlegenden Begriffe der Zeitreihenanalyse wie Stationarität und Integration definieren, die für den Aufbau dieser Arbeit notwendig sind und stellen kurz klassische lineare stochastische Prozesse vor. Darauf aufbauend werden wir die Frequenzanalyse von Zeitreihen motivieren und dieses Kapitel damit beschließen, dass wir die Eigenschaft des langen Gedächtnis von Zeitreihen, was wir in dieser Arbeit als Long Memory bezeichnen werden, einführen. Diese werden wir dann insbesondere im Rahmen der fraktionalen Integration diskutieren. Die folgenden Darstellungen basieren in größten Teilen auf den Ausführungen von Schlittgen und Streitberg (1994) und Hamilton (1993), wobei das Themengebiet der fraktionalen Integration dem Lehrbuch von Palma (2007) und der Dissertation von Tschernig (1994) entstammt.

### 2.1 Grundlegende Definitionen

#### 2.1.1 Stochastische Prozesse

Unter einem stochastischen Prozess  $y_t$  mit  $t = 1, 2, \dots, n$  verstehen wir eine Folge  $\{y_1, y_2, \dots, y_n\}$  von Zufallsvariablen, die definiert ist als eine messbare Abbildung, die einem Ereignis eines Zufallsvorgangs einen reellen Wert zuordnet<sup>2</sup>. Als Beispiel hierfür lässt sich ein Weißes Rauschen vorstellen mit

$$y_t = \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim IID(0; \sigma_\varepsilon^2) \quad (2.1)$$

wobei wir mit einer *IID*-Verteilung eine identisch und unabhängig verteilte Zufallsvariable (von engl. *i*ndependent and *i*dentically *d*istributed) meinen. Die Forderung der identischen Verteilung zu jedem Zeitpunkt  $t$  gewährleistet, dass wir eine feste Varianz für  $y_t$  von  $\sigma_\varepsilon^2$  haben. Zusätzlich wird noch der Erwartungswert des Weißen Rauschen bei null fixiert, um es von möglicherweise vorhandenen deterministischen Prozessbestandteilen wie etwa einem Mittelwert bzw. Startwert zu unterscheiden. Die Bezeichnung „weiß“ rührt aus der Spektralanalyse im Bereich der Optik, da ein Weißes Rauschen eine konstante Spektraldichte aufweist und sich somit sämtliche Frequenzanteile zu der Farbe weiß aufheben. Ein Beispiel für ein Weißes Rauschen stellen wir in Abbildung (2.1) mit  $\varepsilon_t \sim N(0; 1)$  dar. Hieraus lässt sich erkennen, dass die einzelnen Realisierungen des Zufallsprozess keine Struktur aufweisen und nicht voneinander abhängen.

---

<sup>2</sup>Vgl. hierzu die Ausführungen von Schlittgen und Streitberg (1994) S. 92

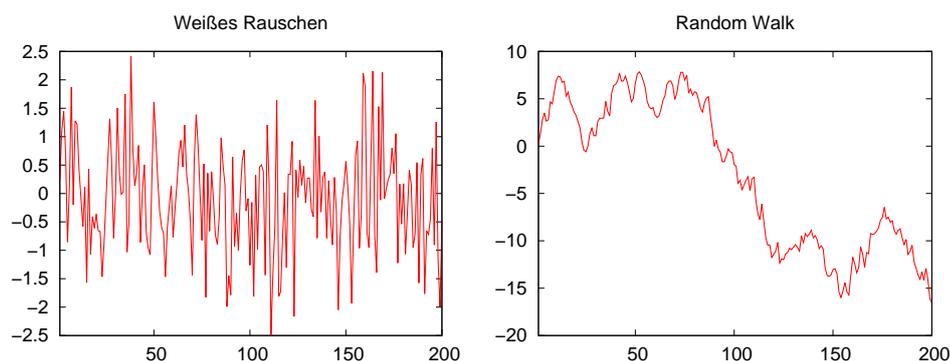


Abb. 2.1: Beispiele für ein Weißes Rauschen und einen Random Walk

### 2.1.2 Der Stationaritätsbegriff

Die Stationarität ist eine der wichtigsten Eigenschaften einer Zeitreihe, denn Schätzungen auf Basis von nicht-stationären Prozessen können zu Fehlentscheidungen aufgrund ihres zum Teil explosiven Verhaltens führen, was Granger und Newbold (1974) in ihrer Arbeit festgestellt hatten. Ausgehend von der Forderung nach schwacher Stationarität, lassen sich folgende Bedingungen der Stationarität einer Zeitreihe  $y_t$  für  $t = 1, \dots, n$  ableiten:

- $E(y_t) = \text{konstant} \quad \forall \quad t$
- $\text{Var}(y_t) = \gamma(0) < \infty \quad \forall \quad t$
- $\text{Cov}(y_t, y_{t+j}) = \gamma(j) = \text{konstant} \quad \forall \quad t, j$

Mit  $\gamma(j)$  bezeichnen wir im Folgenden die Autokovarianzfunktion zur Ordnung  $j$ . Im Falle der starken oder strengen Stationarität wird noch zusätzlich gefordert, dass die Varianz von  $y_t$  einen endlichen und konstanten Wert aufweist und somit für alle Zeitpunkte gleich ist. Dies wiederum wäre z.B. für das Weiße Rauschen in Gleichung (2.1) gegeben, da neben der identischen Verteilung auch die Bedingung der Unabhängigkeit gilt. Den Begriff der starken Stationarität wollen wir aber in dieser Arbeit nicht weiter verfolgen und verwenden den Begriff Stationarität als Synonym für den schwachen Stationaritätsbegriff.

### 2.1.3 Integrierte Prozesse

In empirischen Anwendungen der Zeitreihenanalyse erweisen sich Prozesse allerdings recht häufig als nicht-stationär. Um solche Prozesse dennoch untersuchen zu können, wird häufig der Begriff der Integration angeführt. Dieser besagt, dass ein Prozess  $y_t$

integriert vom Grad  $d$  ist, falls er nach  $d$ -maliger Anwendung der Differenzenbildung wiederum stationär ist. Klassisch nimmt man an, dass der Integrationsparameter  $d$  lediglich ganzzahlige Werte annehmen kann, wovon wir uns aber im Verlauf dieser Arbeit trennen möchten. Weiterhin verwenden wir auch die vereinfachende Notation  $y_t \sim I(d)$ , um auszudrücken, dass ein Prozess integriert vom Grad  $d$  ist.

Um uns mit dem Begriff der Integration vertraut zu machen, betrachten wir nun als ein Beispiel einen sog. Random-Walk-Prozess, der in der Form

$$y_t = y_{t-1} + \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim IID(0, \sigma_\varepsilon^2) \quad (2.2)$$

nicht-stationär ist, da die Varianz von  $y_t$  linear über alle Grenzen anwächst. Dies kann man daran erkennen, dass wir durch rekursives Einsetzen die Darstellung  $y_t = \sum_{i=1}^t \varepsilon_i$  erhalten können. Bringen wir nun  $y_{t-1}$  auf die linke Seite der Gleichung (4.15) und führen die Beschreibung mit Hilfe des Lag-Operators  $L^j y_t = y_{t-j}$  ein, so erhalten wir mit

$$y_t - y_{t-1} = (1 - L)y_t = \varepsilon_t \quad (2.3)$$

ein Weißes Rauschen, welches aufgrund des Erwartungswerts von null und der endlichen Varianz  $\sigma_\varepsilon^2$  ein stationärer Prozess darstellt. Somit sehen wir, dass ein einmaliges Differenzieren ausreicht, um einen Random Walk stationär zu machen, da es sich hierbei um ein integriertes Weißes Rauschen handelt. Der Integrationsgrad  $d$  beträgt damit genau eins bzw.  $y_t \sim I(1)$ . Als Beispiel für einen Random Walk haben wir die rechte Graphik der Abbildung 2.1, welches sich als Integration des Weißen Rauschen in der linken Graphik ergibt. Dies zeigt uns die Implikationen des nicht-stationären Verhaltens, welches sich in der starken Persistenz des Prozessverlaufs widerspiegelt.

## 2.2 Lineare stochastische Prozesse

In dieser Arbeit wollen wir uns vorrangig mit linearen stochastischen Prozessen beschäftigen. Die Eigenschaft der Linearität rührt daher, dass wir einen Prozess  $y_t$  durch eine Lineartransformation der eigenen Vergangenheitswerte (sog. autoregressive Prozess) bzw. von Störtermen (Moving-Average-Prozesse) modellieren<sup>3</sup>. Trotz dieser einfachen Darstellungsweise vermögen lineare stochastische Prozesse eine weite Bandbreite an empirischen Phänomenen zu erklären und modelltheoretisch zu er-

---

<sup>3</sup>Vgl. hierzu Schlittgen und Streitberg (1994) S. 105

klären.

### 2.2.1 Moving-Average-Prozesse

Ein einfaches Beispiel für einen stochastischen Prozess stellt ein Moving-Average-Prozess der ersten Ordnung (weiterhin abgekürzt als MA(1)) der Form

$$y_t = \mu + \varepsilon_t - \phi\varepsilon_{t-1} \quad \text{mit} \quad \varepsilon_t \sim IID(0, \sigma_\varepsilon^2) \quad (2.4)$$

dar. Dieser wiederum weist als erstes unzentriertes Moment den Erwartungswert

$$E(y_t) = E(\mu + \varepsilon_t - \phi\varepsilon_{t-1}) = \mu \quad (2.5)$$

auf und hat somit einen konstanten Mittelwert, was eine der Stationaritätsmerkmale erfordert. Betrachten wir nun weiter die Varianz von MA(1)-Prozessen, so erhalten wir

$$\begin{aligned} \gamma(0) &= E(y_t - \mu)^2 = E(\varepsilon_t - \phi\varepsilon_{t-1})^2 = E(\varepsilon_t^2 - 2\phi\varepsilon_t\varepsilon_{t-1} + \phi^2\varepsilon_{t-1}^2) \\ &= \sigma_\varepsilon^2 - 0 + \phi^2\sigma_\varepsilon^2 = (1 + \phi^2)\sigma_\varepsilon^2 \end{aligned} \quad (2.6)$$

und für die erste Autokovarianz  $\gamma(1)$

$$\begin{aligned} E(y_t - \mu)(y_{t-1} - \mu) &= E(\varepsilon_t - \phi\varepsilon_{t-1})(\varepsilon_{t-1} - \phi\varepsilon_{t-2}) \\ &= E(\varepsilon_t\varepsilon_{t-1} - \phi\varepsilon_{t-1}^2 - \phi\varepsilon_t\varepsilon_{t-2} + \phi^2\varepsilon_{t-1}\varepsilon_{t-2}) \\ &= -\phi\sigma_\varepsilon^2. \end{aligned} \quad (2.7)$$

Für  $j > 1$  ergibt sich für die Autokovarianzfunktion  $\gamma(j)$  stets 0 aufgrund der zeitlichen Unabhängigkeit der Störterme  $\varepsilon_t$ . Allgemein lässt sich festhalten, dass die Varianz und Autokovarianz keine Funktion der Zeit darstellt, sondern lediglich konstante Werte annehmen. Daraus lässt sich der Schluss ziehen, dass ein MA(1)-Prozess unabhängig von den Werten  $\mu$  und  $\phi$  zumindest schwach stationäre Prozesse umschreibt. Zum Zweck der Modelldiagnose lässt sich die Autokorrelationsfunktion (abgekürzt als ACF von autocorrelation function) heranziehen, die sich auch empirisch als  $\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0)$  schätzen lässt. Diese wiederum ist lediglich eine lineare Transformation der Autokovarianzfunktion und es ergeben sich somit ähnliche Eigenschaften wie für diese. Die ACF weist lediglich für  $j = 1$  einen Wert ungleich null auf, der sich durch Einsetzen der Autokovarianzfunktionswerte als  $\rho(1) = -\phi/(1 + \phi)^2$  ergibt. Zusätzlich steht uns noch die partielle Autokorrelations-

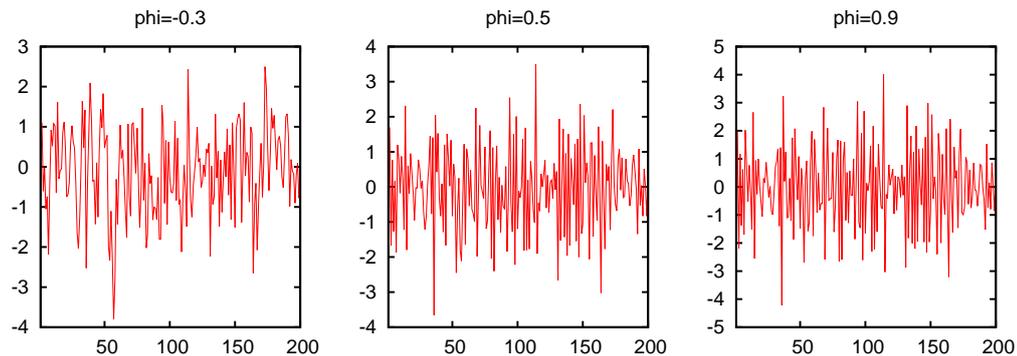


Abb. 2.2: Beispiele für MA(1)-Prozesse

funktion (PACF) zur Verfügung, die sich als geschätzter Koeffizient des jeweiligen Lag  $j$  der Zeitreihe auf sich selbst ergibt. Hierfür wiederum ergibt sich ein anderer Funktionsverlauf als für die ACF, die wir in Abbildung 2.3 gegenüber stellen für MA(1)-Prozesse mit  $\phi \in \{-0,3; 0,5; 0,9\}$ . Im Gegensatz zur ACF bricht die PACF nicht bei  $j = 1$  für MA(1)-Prozesse ab und weist einen Anstieg auf, der steiler ausfällt, je größer der Betrag von  $\phi$  ist. Für negative Werte von  $\phi$  erhalten wir hingegen einen vom Vorzeichen alternierenden Verlauf der mit steigendem Lag  $j$  betragsmäßig abnimmt. Dies kann man dann z.B. in der Zeitreihengraphik 2.2, welche mit den jeweils gleichen Störtermen  $\varepsilon_t$  erzeugt worden sind, erkennen, dass wir für  $\phi > 0$  einen spitzen Verlauf haben aufgrund der negativen Autokorrelation und für  $\phi < 0$  einen etwas glatteren Verlauf, da wir hier eher positive Autokorrelationen beobachten. Es sei aber auch angemerkt, dass die Interpretation im Wesentlichen von der Annahme des datengenerierenden Prozesses abhängt. So stellt u.a. Hamilton (1994) einen MA(1)-Prozess als

$$y_t = \mu + \varepsilon_t + \phi\varepsilon_{t-1} \quad \text{mit} \quad \varepsilon_t \sim IID(0; \sigma_\varepsilon^2) \quad (2.8)$$

dar, was dann genau eine entgegengesetzte Interpretation impliziert. Betrachten wir nun einen allgemeinen MA( $q$ )-Prozess der Form

$$y_t = \mu + \varepsilon_t - \phi_1\varepsilon_{t-1} - \dots - \phi_q\varepsilon_{t-q} = \mu + \underbrace{(1 - \phi_1L^1 - \dots - \phi_qL^q)}_{\phi(L)}\varepsilon_t \quad (2.9)$$

mit  $\varepsilon_t \sim IID(0; \sigma_\varepsilon^2)$ . So erkennen wir hier deutlich die Natur von MA-Prozessen, dass der Prozess  $y_t$  eine gewichtete Summe bzw. einen Mittelwert der vergangenen Werte von  $\varepsilon$  darstellt. Im Vergleich zu der Situation mit dem MA(1)-Prozess ändert sich der Erwartungswert  $E(y_t) = \mu$  nicht, da immer noch  $E(\varepsilon_t) = 0$  gilt. Die Varianz

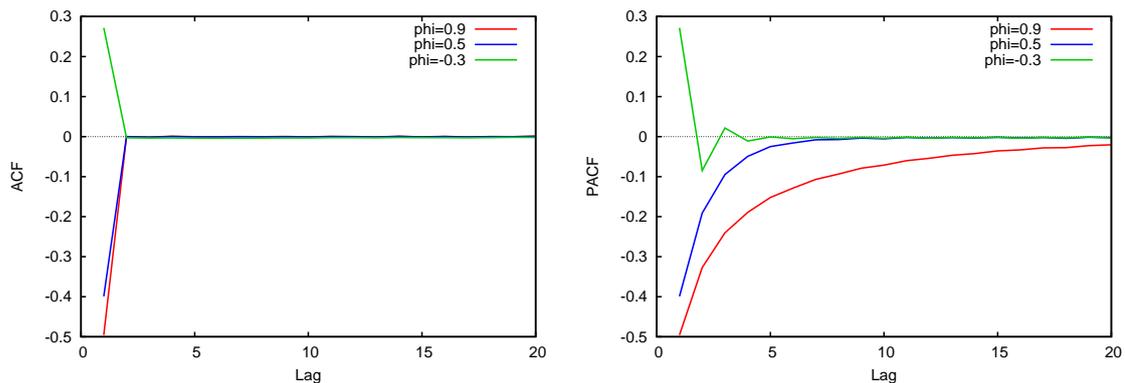


Abb. 2.3: (Partielle) Autokorrelationsfunktion von MA(1)-Prozessen

von  $y_t$  ergibt sich als

$$\gamma(0) = E(y_t - \mu)^2 = E(\varepsilon_t - \phi_1\varepsilon_{t-1} - \dots - \phi_q\varepsilon_{t-q})^2. \quad (2.10)$$

und lässt sich aufgrund der Unkorreliertheit der  $\varepsilon_t$  vereinfachen zu

$$\gamma(0) = (1 + \phi_1^2 + \phi_2^2 + \dots + \phi_q^2)\sigma_\varepsilon^2 \quad (2.11)$$

Die Autokovarianz der Ordnung  $j$  von  $y_t$  lässt sich allgemein schreiben als

$$\gamma(j) = \begin{cases} -\sigma_\varepsilon^2 \sum_{i=0}^{q-j} \phi_i \phi_{i+j}, & 0 \leq j \leq q \quad \text{mit } \phi_0 = 1 \\ 0, & \text{sonst.} \end{cases} \quad (2.12)$$

Sowohl Erwartungswert als auch die Kovarianz hängen damit nur von den konstanten Größen  $\mu$ ,  $\phi_i$  und  $\sigma_\varepsilon^2$  ab, die in endlicher Anzahl in den Ausdrücken vorkommen. Damit lässt sich aussagen, dass jeder MA( $q$ )-Prozess mit endlichem  $q$  zumindest schwach stationär ist.

## 2.2.2 Autoregressive Prozesse

Neben Moving-Average-Prozessen wollen wir uns nun mit autoregressiven Prozessen beschäftigen. Ein autoregressiver Prozess erster Ordnung (weiterhin als AR(1) abgekürzt) ist definiert als

$$y_t = c + \theta y_{t-1} + \varepsilon_t \quad \text{mit} \quad \varepsilon \sim IID(0; \sigma_\varepsilon^2) \quad (2.13)$$

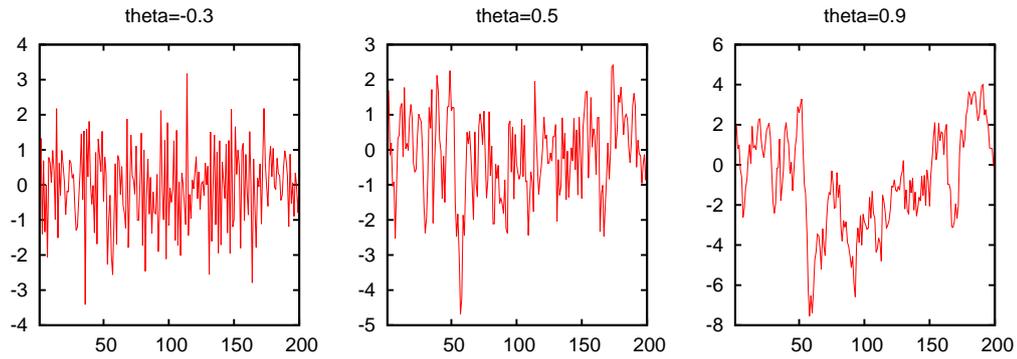


Abb. 2.4: Beispiele für AR(1)-Prozesse

und stellt somit einen Prozess dar, der sich aus den eigenen Vergangenheitswerten erklären lässt. Der Parameter  $c$  lässt sich als ein Drift-Parameter interpretieren, der einen linearen Zuwachs (falls  $c > 0$ ) für jeden Zeitpunkt  $t$  von  $y_t$  bedeutet. Die Diskussion der Stationarität von AR(1)-Prozessen ist im Vergleich zu Moving-Average-Prozessen schwieriger, da die Bedingung hier nicht nur auf die Endlichkeit der Ordnung  $p$  abstellt. Schreiben wir den datengenerierenden Prozess aus (2.13) mit dem Lag-Operator um, dann erhalten wir

$$y_t - \underbrace{\theta y_{t-1}}_{\theta(L)} = c + \varepsilon_t. \quad (2.14)$$

Für diesen Prozess lässt sich prinzipiell zeigen, dass die Stationaritätsbedingung erfüllt ist, falls eine Lösung  $|z| > 1$  existiert für die charakteristische Gleichung

$$|\theta(z)| = |1 - \theta z| = 0. \quad (2.15)$$

Diese gilt in diesem Fall bei einem Autoregressionsparameter von  $|\theta| < 1$ . Bei einem Parameter  $|\theta| > 1$  weist der Prozess  $y_t$  ein explosives Verhalten auf und entwickelt sich über alle Grenzen. Der Grenzfall  $\theta = 1$  entspricht dem Random-Walk-Prozess aus Abschnitt 2.1.3 und ist gerade nicht-stationär aufgrund der tendenziell unendlich großen Varianz. Im Falle der Stationarität lässt sich dieser AR(1) als einen Moving-Average-Prozess darstellen, indem wir schreiben

$$\begin{aligned} y_t &= (1 - \theta L)^{-1}(c + \varepsilon_t) = \frac{c}{1 - \theta} + (1 - \theta L)^{-1}\varepsilon_t \\ &= \frac{c}{1 - \theta} + \sum_{i=0}^{\infty} \theta^i L^i \varepsilon_t = \frac{c}{1 - \theta} + \sum_{i=0}^{\infty} \phi_i L^i \varepsilon_t \quad \text{mit} \quad \phi_i = \theta^i. \end{aligned} \quad (2.16)$$

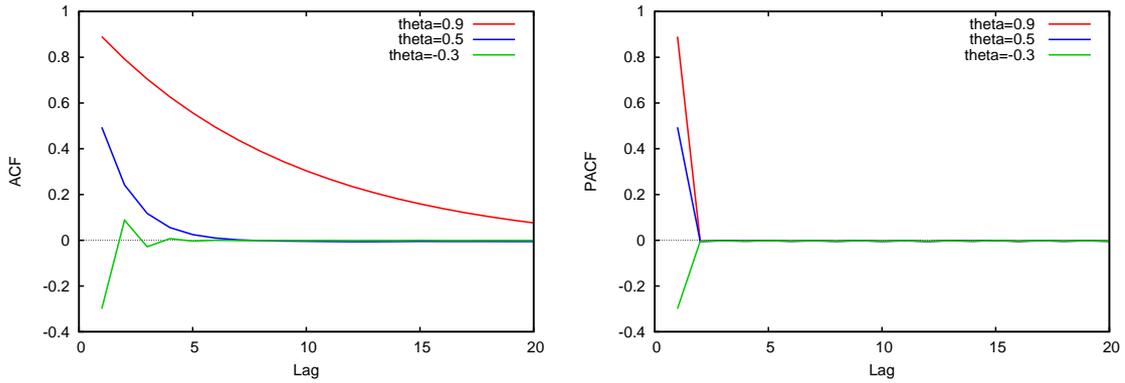


Abb. 2.5: (Partielle) Autokorrelationsfunktion von AR(1)-Prozessen

Dieser MA-Prozess hat die Ordnung  $\infty$  und dessen Gewichte  $\phi_i$  konvergieren für den Fall  $|\theta| < 1$  gegen 0. Folglich lässt sich der Erwartungswert für  $y_t$  wie für die bereits angeführten MA-Prozesse bestimmen mit

$$E(y_t) = \frac{c}{1 - \theta} \quad (2.17)$$

$$\gamma(0) = \sum_{i=0}^{\infty} (\theta^i)^2 \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \theta^2} \quad (2.18)$$

$$\gamma(j) = \sum_{i=0}^{\infty} (\theta^{j+i})^2 \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2 \theta^{2j}}{1 - \theta^2}. \quad (2.19)$$

Die Struktur von  $\gamma(j)$  gibt wiederum schon Anhaltspunkte, dass die empirisch ermittelbare ACF einen exponentiell abklingenden Verlauf annimmt und für negative Werte von  $\theta$  einen alternierenden. Für  $\theta \in \{-0,3; 0,5; 0,9\}$  haben wir die ACF und PACF in Abbildung 2.5 auch graphisch dargestellt und können erkennen, dass die PACF im Gegensatz zu MA(1)-Prozessen beim ersten Lag abbricht und somit ein Unterscheidungsmerkmal zu endlichen MA-Prozessen bei empirischen Analysen bildet.

Versuchen wir nun die vorangegangenen Ergebnisse für einen autoregressiven Prozess der Ordnung  $p$  der Form

$$y_t = c + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t. \quad (2.20)$$

zu verallgemeinern, der sich auch kompakt als

$$(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p) y_t = \theta(L) y_t = c + \varepsilon_t \quad (2.21)$$

umschreiben lässt. Dieser Prozess genügt nun wiederum der Stationaritätsbedingung, wenn die Lösungen  $z_1, \dots, z_p$  der charakteristischen Gleichung

$$|\theta(L)| = |1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p| = 0 \quad (2.22)$$

außerhalb des Einheitskreises liegen, da die Lösungen durchaus auch komplexe Zahlen annehmen können. Im stationären Fall bilden wir den Erwartungswert von  $y_t$  über die Prozessgleichung

$$E(y_t) = c + \theta_1 E(y_{t-1}) + \theta_2 E(y_{t-2}) + \dots + \theta_p E(y_{t-p}) + E(\varepsilon_t) \quad (2.23)$$

und nutzen dabei die Annahme eines konstanten Erwartungswertes  $\mu = E(y_t) = E(y_{t-1}) = E(y_{t-i})$  aus. Dann stellt sich das Problem wie folgt dar:

$$(1 - \theta_1 - \dots - \theta_p)\mu = c \quad \Leftrightarrow \quad \mu = \frac{c}{1 - \theta_1 - \dots - \theta_p}. \quad (2.24)$$

Zentrieren wir die Prozessgleichung unter der vereinfachenden Annahme  $c = 0$  um den Erwartungswert der einzelnen Prozesselemente, erhalten wir

$$y_t - \mu = \theta_1(y_{t-1} - \mu) + \theta_2(y_{t-2} - \mu) + \dots + \theta_p(y_{t-p} - \mu) + \varepsilon_t. \quad (2.25)$$

Diese Gleichung, multipliziert mit  $(y_{t-i} - \mu)$ , ergibt die Autokovarianzfunktion für  $y_t$  mit

$$\gamma(j) = \begin{cases} \theta_1 \gamma(j-1) + \theta_2 \gamma(j-2) + \dots + \theta_p \gamma(j-p) & \text{für } j \neq 0 \\ \theta_1 \gamma(1) + \theta_2 \gamma(2) + \dots + \theta_p \gamma(p) + \sigma_\varepsilon^2 & \text{für } j = 0 \end{cases} \quad (2.26)$$

Dividieren wir die Autokovarianzfunktion  $\gamma(j)$  durch die Varianz  $\gamma(0)$ , so ergibt sich daraus die Autokorrelationsfunktion  $\rho(j)$  und resultiert in den Yule-Walker-Gleichungen:

$$\rho(j) = \theta_1 \rho(j-1) + \theta_2 \rho(j-2) + \dots + \theta_p \rho(j-p) \quad \text{für } j \neq 0. \quad (2.27)$$

Diese Gleichungen zeigen den Zusammenhang der empirisch ermittelbaren Autokorrelationskoeffizienten zu den einzelnen Parametern des datengenerierenden AR( $p$ )-Prozesses aus Gleichung (2.20) und legen nahe, dass die Parameter  $\theta_1, \dots, \theta_p$  per Momentenmethode geschätzt werden können.

### 2.2.3 Gemischte ARMA-Prozesse

Wie wir bereits gesehen haben, lässt sich ein  $AR(p)$ -Prozess in einen  $MA(\infty)$ -Prozess überführen, falls dieser stationär ist. Im Gegenzug kann auch ein  $MA(q)$ -Prozess in einen  $AR(\infty)$ -Prozess invertiert werden, falls wiederum die  $q$  Anzahl Lösungen der charakteristischen Gleichung des  $MA(q)$ -Prozesses außerhalb des Einheitskreises liegen. Diese gegenseitige Invertierbarkeit legt die Möglichkeit nahe, dass wenn wir z.B.  $AR$ -Prozesse mit einer relativ hohen Ordnung  $p$  beschreiben möchten, dass man durch Aufnahme von  $MA$ -Prozesselementen eine Einsparung von Parametern vornehmen und somit die einzelnen Parameter einfacher und effizienter schätzen kann. Da dies natürlich auch umgekehrt gilt, werden wir uns in diesem Abschnitt mit gemischten autoregressiven Moving-Average-Prozessen (auch abgekürzt als  $ARMA$ -Prozesse) beschäftigen. Ein allgemeiner  $ARMA(p, q)$ -Prozess ist definiert als

$$(1 - \theta_1 L - \dots - \theta_p L^p)y_t = c + (1 - \phi_1 L - \dots - \phi_q L^q)\varepsilon_t \quad (2.28)$$

mit  $\varepsilon_t \sim IID(0; \sigma_\varepsilon^2)$ . Die Diskussion der Stationarität von  $ARMA(p, q)$ -Prozessen gestaltet sich dahingehend recht einfach, da wir von einem  $MA(q)$ -Prozess wissen, dass er mit endlichem  $q$  stationär ist und somit nur  $AR(p)$ -Teil des gemischten Prozesses nicht-stationär sein kann. Die Kriterien für die Stationarität sind wiederum die gleichen, wie sie im vorangegangenen Abschnitt beschrieben wurden. Die Darstellung der Autokovarianzfunktion von  $ARMA$ -Prozessen lässt sich durch die Betrachtung des Prozesses  $y_t$  als Abweichung von dessen Erwartungswert  $\mu$  motivieren und ergibt unter der Vereinfachung  $c = 0$

$$\begin{aligned} y_t - \mu &= \theta_1(y_{t-1} - \mu) + \theta_2(y_{t-2} - \mu) + \dots + \theta_p(y_{t-p} - \mu) \\ &+ \varepsilon_t - \phi_1\varepsilon_{t-1} - \phi_2\varepsilon_{t-2} - \dots - \phi_q\varepsilon_{t-q}. \end{aligned} \quad (2.29)$$

Hierüber lässt sich der Erwartungswert bilden und ergibt für  $j > q$  die Autokovarianzfunktion

$$\gamma(j) = \theta_1\gamma(j-1) + \theta_2\gamma(j-2) + \dots + \theta_p\gamma(j-p). \quad (2.30)$$

Das heißt wiederum, dass jede Autokovarianz und damit auch Autokorrelation der Ordnung  $j$ , die über die  $MA$ -Ordnung  $q$  eines  $ARMA(p, q)$ -Prozesses hinausgeht, alleine von den  $AR$ -Bestandteilen bestimmt wird. Daran lässt sich schon erkennen, dass die ACF von  $ARMA(p, q)$ -Prozessen je nach Parameterkonstellation nicht abbrechen wird wie für  $MA(q)$ -Prozesse. Betrachten wir nun den speziellen Fall des

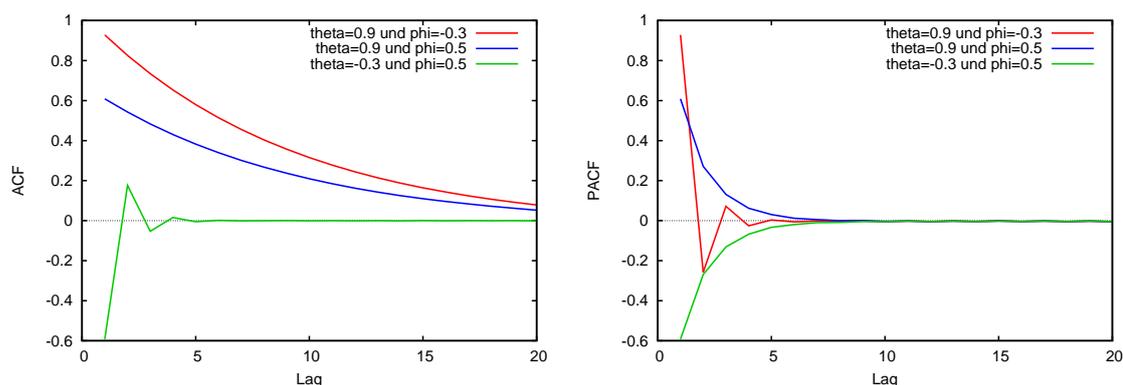


Abb. 2.6: (Partielle) Autokorrelationsfunktion von ARMA(1,1)-Prozessen

ARMA(1,1)-Prozesses der Form

$$y_t = \theta y_{t-1} + \varepsilon_t - \phi \varepsilon_{t-1}, \quad (2.31)$$

wofür sich somit die Autokovarianzfunktion mit

$$\gamma(j) = \theta \gamma(j-1) = \theta^{j-1} \gamma(1) \quad \text{für } j > 1 \quad (2.32)$$

aufstellen lässt. Die Varianz und die erste Autokovarianz lassen sich, ohne darauf näher einzugehen, angeben mit

$$\gamma(0) = \frac{1 - 2\theta\phi + \phi^2}{1 - \theta^2} \sigma_\varepsilon^2 \quad (2.33)$$

$$\gamma(1) = \frac{(1 - \theta\phi)(\theta - \phi)}{1 - \theta^2} \sigma_\varepsilon^2. \quad (2.34)$$

Um dies zu veranschaulichen stellen wir in Abbildung 2.6 die ACF und PACF für insgesamt drei Parameterkonstellationen von ARMA(1,1)-Prozessen dar, die durch Simulation der jeweiligen Prozesse gewonnen wurden. Hier kann man u.a. erkennen, dass die PACF einen zu null konvergierenden Verlauf hat, aber auch nicht bei der AR-Ordnung  $p = 1$  abbricht. Der Verlauf der Kurven selbst ergibt sich aber aus den Parametern  $\theta$  und  $\phi$ , die für die ACF besonders durch  $\theta$  bestimmt wird, da sich nur für  $\theta < 0$  ein vom Vorzeichen der Korrelationswerte alternierender Verlauf ergibt und für  $\theta > 0$  ein exponentiell abklingender. Im Gegensatz dazu wird die PACF vor allem durch den MA-Parameter  $\phi$  getrieben, der die Gestalt der PACF zu definieren scheint. Aufgrund dieser ambivalenten Modellierungsmöglichkeit wird auch rasch deutlich, dass ARMA-Prozesse in der empirischen Anwendung einen hohen Stellenwert genießen, da sie besonders dann in Frage kommen, wenn Anwender

sich aufgrund der Gestalt von ACF und PACF nicht für die Modellierung mit AR- und MA-Prozessen entscheiden können.

### 2.3 Frequenzanalyse von Zeitreihen

Die bisherigen Analysen beschäftigten sich mit den Eigenschaften von stochastischen Prozessen im Zeitbereich. Dazu wurde versucht, einen Prozess  $y_t$  durch seine vorgegangenen Werte zu erklären bzw. durch eine Aufsummierung von Störgrößen. Eine weitere Möglichkeit lautet, einen Prozess durch die Überlagerung von Zyklen darzustellen. So lässt sich ein stationärer Prozess  $y_t$  durch eine gewichtete Summe von Sinus- und Kosinustermen ausdrücken als

$$y_t = a_0 + \sum_{j=1}^M [a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)] \quad \text{mit} \quad \lambda_j = \frac{2\pi j}{n} \quad (2.35)$$

für  $t = 1, \dots, n$ . Diese Darstellungsweise wird auch als Fourierreihe und  $\lambda_j$  als Fourierfrequenz bezeichnet.  $M$  umschreibt die Anzahl der verwendeten Frequenz zur Beschreibung des Prozesses  $y_t$ . Dieser Parameter lässt sich über die Betrachtung einzelner Fourierfrequenzen von

$$\begin{aligned} \lambda_1 &= 2\pi/n \\ \lambda_2 &= 4\pi/n \\ &\vdots \\ \lambda_M &= 2M\pi/n \end{aligned} \quad (2.36)$$

spezifizieren als  $M = (n - 1)/2$ , da die höchste in Gleichung (2.35) aufgenommene Frequenz  $\lambda_M = \frac{2(n-1)\pi}{2n} < \pi$  sein muss aufgrund der Periodizität der Sinus- und Kosinusfunktion. In Abbildung 2.7 haben wir noch die Kosinusschwingungen für die Fourierfrequenzen  $\lambda_j$  mit  $j \in \{1; 2; 5\}$  unter  $n = 500$  graphisch dargestellt und können daran erkennen, dass, je kleiner die Frequenz ist, desto länger die jeweilige Welle wird, die zur Beschreibung der Zeitreihe  $y_t$  dient. Zudem ist von cosinoiden Prozessen bekannt, dass sie stationär sind, wenn die einzelnen Gewichtspaare  $(a_j, b_j)$  unabhängig sind von  $(a_k, b_k)$  für  $j \neq k$ . Daraus und aus weiteren Überlegungen, die wir hier nicht formal darstellen wollen, folgt, dass  $a_j$  und  $b_j$  sich dabei als Zufallsvariablen mit Erwartungswert 0 und Varianzen  $E(a_j^2) = E(b_j^2) = \sigma_j^2$  interpretieren lassen.

Hiermit lässt sich nun Varianz von  $y_t$  unter der vereinfachenden Annahme

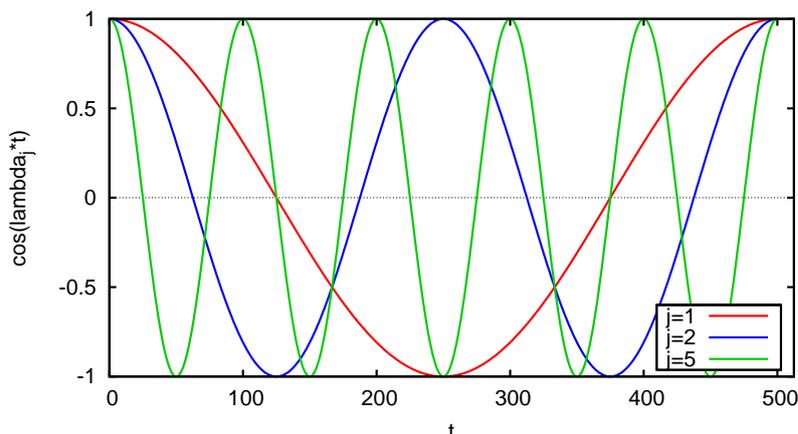


Abb. 2.7: Kosinusschwingung in Abhängigkeit der jeweiligen Fourierfrequenzen

$a_0 = 0$  analysieren:

$$\text{Var}(y_t) = \sum_{j=1}^M [\text{Var}(a_j) + \text{Var}(b_j)] = 2 \sum_{j=1}^M \sigma_j^2. \quad (2.37)$$

Somit zeigt die Varianz der Koeffizienten  $a_j$  bzw.  $b_j$  den jeweiligen Beitrag der Fourierfrequenz  $\lambda_j$  zur Gesamtvarianz von  $y_t$  an und lässt sich damit als „Stärke“ der jeweiligen Schwingungskomponente in  $y_t$  interpretieren.<sup>4</sup>

### 2.3.1 Das Spektrum

Ein wichtiges Instrument der Frequenzanalyse ist die Spektraldichte oder auch Spektrum genannt, das sich für einen stationären Prozess  $y_t$  als die Fouriertransformierte der theoretischen Kovarianzfunktion  $\gamma(j)$  ergibt als

$$f(\lambda) := \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) \exp(-i\lambda j) = \frac{1}{2\pi} \left[ \gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j) \cos(\lambda j) \right] \quad (2.38)$$

mit  $i = \sqrt{-1}$ . Der Begriff Spektraldichtefunktion lässt sich dadurch begründen, dass viele Eigenschaften des Spektrums formal an eine Dichtefunktion erinnern<sup>5</sup>. Diese lauten nach Schlittgen und Streitberg (1994):

- (1)  $f(\lambda = 0)$  liefert die Summe aller Kovarianzen, da  $f(0) = \sum_{j=-\infty}^{\infty} \gamma(j)$
- (2)  $f(\lambda)$  ist symmetrisch, da  $f(-\lambda) = f(\lambda)$

<sup>4</sup>Vgl. hierzu Schlittgen und Streitberg (1994) S. 155

<sup>5</sup>Vgl. hierzu Schlittgen und Streitberg (1994) S. 156ff.

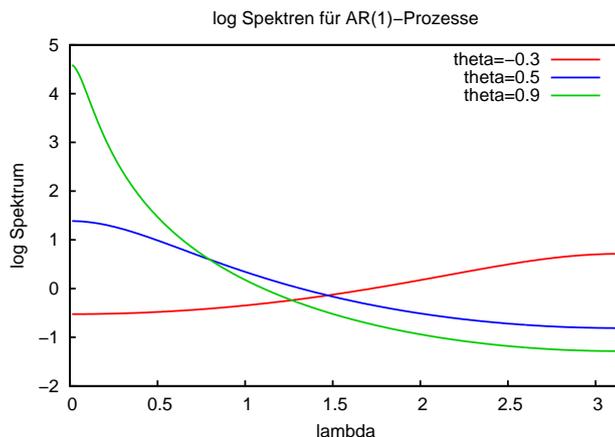


Abb. 2.8: Logarithmierte Spektren von ausgewählten AR(1)-Prozessen

- (3)  $f(\lambda)$  ist periodisch mit der Periode  $\pi$ :  $f(\lambda + \pi) = f(\lambda)$
- (4)  $f(\lambda)$  ist nicht-negativ:  $f(\lambda) \geq 0$
- (5)  $f(\lambda)$  ist stetig.

Aufgrund der Eigenschaft der Periodizität reicht es aus, das Spektrum eines Prozesses auf einen Bereich für  $\lambda$  von  $[0; \pi]$  zu untersuchen. Ein weiterer wichtiger Punkt zum Verständnis ist das Parsevalsche Theorem, das wiederum besagt, dass mit

$$\gamma(j) = \int_{-\pi}^{\pi} f(\lambda) \exp(-i\lambda j) d\lambda = \int_{-\pi}^{\pi} f(\lambda) \cos(\lambda j) d\lambda \quad (2.39)$$

sämtliche Werte der Autokovarianzfunktion  $\gamma(j)$  aus dem Spektrum über die inverse Fouriertransformation berechenbar ist. Besonders gilt für die Varianz  $\gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda$ , dass sich diese als Fläche unter dem Spektrum ergibt. Hieraus motivieren Schlittgen und Streitberg (1994) die alternative Bezeichnung des Varianzspektrums für die Spektraldichte.

Betrachten wir nun die Spektraldichten der zuvor vorgestellten stochastischen Prozesse und beginnen zunächst mit einem AR(1)-Prozess der Form (2.13). Für

diesen lässt sich anhand seiner Autokovarianzfunktion (2.19) das Spektrum als

$$\begin{aligned}
f(\lambda) &= \frac{1}{2\pi} \frac{\sigma_\varepsilon^2}{1 - \theta^2} \sum_{j=-\infty}^{\infty} \theta^{|j|} \exp(i\lambda j) \\
&= \frac{1}{2\pi} \frac{\sigma_\varepsilon^2}{1 - \theta^2} \left( 1 + \sum_{j=1}^{\infty} \theta^j \exp(i\lambda j) + \sum_{j=1}^{\infty} \theta^j \exp(-i\lambda j) \right) \\
&= \frac{1}{2\pi} \frac{\sigma_\varepsilon^2}{1 - \theta^2} \left( \frac{1}{1 - \theta \exp(i\lambda)} + \frac{1}{1 - \theta \exp(-i\lambda)} - 1 \right) \\
&= \frac{1}{2\pi} \frac{\sigma_\varepsilon^2}{1 - \theta^2} \left( \frac{1 - \theta^2}{(1 - \theta \exp(i\lambda))(1 - \theta \exp(-i\lambda))} \right) \\
&= \frac{1}{2\pi} \frac{\sigma_\varepsilon^2}{1 + \theta^2 - 2\theta \cos(\lambda)} \tag{2.40}
\end{aligned}$$

angeben. Die Implikationen des Parameters  $\theta$  auf das Frequenzverhalten lässt sich an der Darstellung in Abbildung 2.8 erkennen, wo wir die gleichen Parameterkonstellationen wie für die Abbildung 2.5 verwendet haben. Für positive Werte von  $\theta$  ist die Varianz bei niedrigen Frequenzen konzentriert und deswegen bildet die Spektraldichte bei  $\lambda = 0$  ein Maximum. Das bedeutet, dass sich der Prozess in langen Wellen entwickelt, da wir implizit ein hohes Gewicht auf langen Kosinus- und Sinuswellen haben. Diese Eigenschaft wird umso stärker ausgeprägt, je größer der Parameter  $\theta$  ausfällt. Bei negativen  $\theta$  liegt hingegen der Hauptteil der Varianz bei hohen Frequenzen und diese können für solche Prozesse einen kurzwelligen Verlauf feststellen. Es sei noch angemerkt, dass Darstellungen in Abbildung 2.8 sich wie die folgenden jeweils auf die logarithmierte Spektraldichte beziehen, was auch häufig als Logspektrum bezeichnet wird, da die jeweiligen Funktionswerte zum Teil sehr stark auseinander liegen und so eine bessere Vergleichbarkeit gewährleistet ist. Zusätzlich normieren wir die verwendete Störtermvarianz auf  $\sigma_\varepsilon^2 = 1$  für alle Betrachtungen der Spektraldichte.

Auf ähnliche Weise lässt sich auch die Spektraldichte für Moving-Average-Prozesse ermitteln. Das Spektrum eines MA(1)-Prozesses der Form  $y_t = \varepsilon_t - \phi \varepsilon_{t-1}$  aus Gleichung (2.4) lautet dann unter Verzicht auf eine genauere Herleitung

$$f(\lambda) = \frac{1}{2\pi} [1 + \phi^2 - 2\phi \cos(\lambda)] \sigma_\varepsilon^2. \tag{2.41}$$

Auch für MA(1)-Prozesse wollen wir die logarithmierten Spektren für die gleichen Parameterkonstellationen wie im Abschnitt zuvor betrachten. Die jeweiligen Graphen befinden sich in Abbildung 2.9 und lassen wiederum ähnliche Rückschlüsse

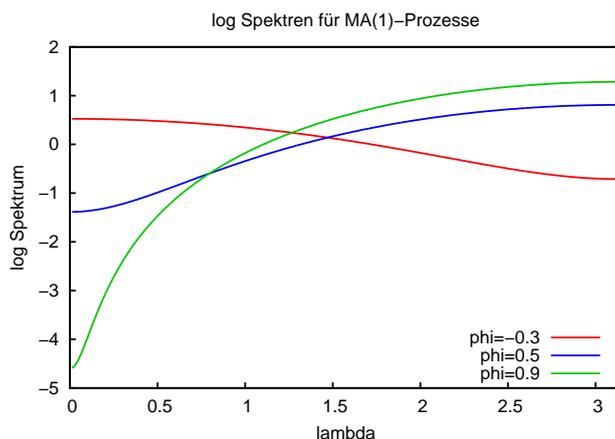


Abb. 2.9: Logarithmierte Spektren von ausgewählten MA(1)-Prozessen

wie bei der Betrachtung der ACF und PACF zu. Für  $\phi < 0$  haben wir aufgrund der implizierten positiven Autokorrelation einen längerwelligen Verlauf und für  $\phi > 0$  einen kürzerwelligen, da negative Autokorrelationen tendenziell ein antipersistentes Verhalten der Zeitreihe hervorrufen.

Diesen Abschnitt wollen wir nun mit der Betrachtung des Spektrums von ARMA(1,1)-Prozessen abschließen, das sich wiederum recht einfach motivieren lässt. Deren Prozessgleichung lässt sich durch Umstellen in eine MA( $\infty$ )-Darstellung überführen als

$$\begin{aligned} (1 - \theta L)y_t &= (1 - \phi L)\varepsilon_t \\ y_t &= (1 - \phi L)(1 - \theta L)^{-1}\varepsilon_t. \end{aligned} \quad (2.42)$$

Hierfür bestimmt sich die Spektraldichte mit

$$f(\lambda) = \frac{1}{2\pi} \frac{1 + \phi^2 - 2\phi \cos(\lambda)}{1 + \theta^2 - 2\theta \cos(\lambda)} \sigma_\varepsilon^2, \quad (2.43)$$

was wiederum einfach der Multiplikation der Spektren der einzelnen Prozessbestandteile entspricht. Es gilt nämlich für das Spektrum von beliebigen ARMA( $p, q$ )-Prozessen, dass sich als

$$f_{\text{ARMA}}(\lambda) = f_{\text{AR}}(\lambda) \cdot f_{\text{MA}}(\lambda) \quad (2.44)$$

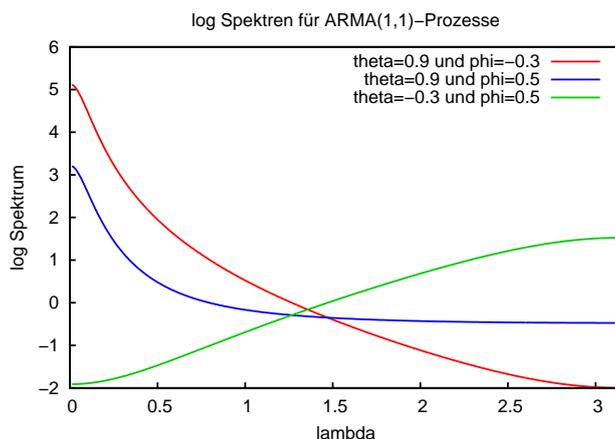


Abb. 2.10: Logarithmierte Spektren von ausgewählten ARMA(1,1)-Prozessen

ergibt, was sich formal auch als<sup>6</sup>

$$f(\lambda) = \frac{1}{2\pi} \left| \frac{\phi(\exp(-i\lambda))}{\theta(\exp(-i\lambda))} \right|^2 \cdot \sigma_\varepsilon^2 = \frac{1}{2\pi} \frac{|1 - \sum_{j=1}^q \phi_j \exp(i\lambda j)|^2}{|1 - \sum_{k=1}^p \theta_k \exp(i\lambda k)|^2} \cdot \sigma_\varepsilon^2 \quad (2.45)$$

darstellen lässt.

In Abbildung 2.10 haben wir nun die Darstellung der Spektren der ARMA(1,1)-Prozesse, die wir auch zur Darstellung der ACF und PACF im vorangegangenen Abschnitt verwendet haben, graphisch aufgearbeitet. Auch hier zeigt sich, dass die Spektraldichten Extremstellen bei  $\lambda = 0$  und  $\lambda = \pi$  aufweisen, was sich im Endeffekt wieder durch die Struktur der Autokovarianzen erklärt. Für komplexere ARMA( $p, q$ )-Strukturen mit  $p, q > 1$  ist es aber durchaus möglich, dass wir andere Extremstellen als die Randlösungen der zu beobachtenden Fourierfrequenzen haben. Ein solches Beispiel werden wir in einem der folgenden Abschnitte mit einem saisonalen AR-Prozess betrachten.

### 2.3.2 Das Periodogramm

Das Periodogramm für einen Zeitreihe  $y_t$  stellt mit

$$I(\lambda_j) = \frac{1}{2\pi} \left[ \hat{\gamma}(0) + 2 \sum_{k=1}^{n-1} \hat{\gamma}(k) \cos(\lambda_j k) \right] \quad (2.46)$$

<sup>6</sup>Vgl. hierzu Schlittgen und Streitberg (1994) S. 184

das empirische Analogon zur Spektraldichte dar und basiert auf den empirisch geschätzten Autokovarianzen

$$\hat{\gamma}(j) = \frac{1}{n} \sum_{t=j+1}^n (y_t - \bar{y})(y_{t-j} - \bar{y}) \quad \text{mit} \quad \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t. \quad (2.47)$$

Eine weitere Motivation für das Periodogramm liefert die Betrachtung der Modellgleichung (2.35). Für ungerade  $n$  müssen zur Schätzung dieser Gleichung insgesamt  $(2M + 1) = n$  Parameter spezifiziert werden und damit haben wir genau so viele zu schätzende Parameter wie Beobachtungspunkte. Da aber die einzelnen Regressoren linear unabhängig sind, bedeutet eine einfache KQ-Regression mit den Regressanden  $\cos(\lambda_j t), \sin(\lambda_j t)$  für  $j = 1, \dots, M$  und einem Absolutglied auf den Prozess  $y_t$  eine perfekte Anpassung.<sup>7</sup> Prinzipiell lässt sich die gleiche Aussage auch für gerade  $n$  machen, was aber nicht weiter thematisiert werden soll. Die Schätzformeln für die Koeffizienten  $a_j$  und  $b_j$  gibt Hamilton (1994) an mit

$$\hat{a}_j = \frac{2}{n} \sum_{t=1}^n y_t \cos(\lambda_j(t-1)) \quad (2.48)$$

$$\hat{b}_j = \frac{2}{n} \sum_{t=1}^n y_t \sin(\lambda_j(t-1)) \quad \text{für} \quad j = 1, \dots, M \quad (2.49)$$

und  $\hat{a}_0 = n^{-1} \sum_{i=1}^n y_i$ . Weiterhin zeigt Hamilton (1994), dass durch die geschätzten Koeffizienten der Modellgleichung diese sich in Kombination mit dem Periodogramm ergeben als

$$\frac{1}{2}(\hat{a}_j^2 + \hat{b}_j^2) = \frac{4\pi}{n} I(\lambda_j). \quad (2.50)$$

Das Periodogramm liefert damit an der Stelle  $\lambda_j$  die Summe der quadrierten Gewichte der harmonischen Darstellungsweise des Prozesses und lässt sich in der Weise interpretieren, dass die einzelnen Periodogrammkoordinaten die Varianzanteile von  $y_t$  zu bestimmten Schwingungen entsprechen.

Brockwell und Davis (1991) geben eine alternative Berechnungsmöglichkeit zu (2.46) auf Basis der Fouriertransformation des zu untersuchenden Prozesses  $y_t$  für das Periodogramm mit

$$I(\lambda_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^n y_t \exp(-i\lambda_j t) \right|^2 \quad (2.51)$$

---

<sup>7</sup>Vgl. hierzu Hamilton (1994) S. 159

an. Hierbei gilt es zu beachten, dass die Fouriertransformation eines Prozesses komplexwertig ist und diese Berechnungsweise im Endeffekt der Summe des quadrierten Real- und Imaginärteiles dem jeweiligen Transformationswert entspricht. Diese Rechenroutine entspricht dem Vorgehen der meisten Statistikprogramme und lässt sich einfacher handhaben als die Berechnung auf Basis der Autokovarianzfunktion  $\gamma(j)$ . Brockwell und Davis (1991) können zudem auch zeigen, dass die Definition der Berechnung in (2.46) sich eigentlich aus der Formel (2.51) ergibt.

### 2.3.3 Das Periodogramm als Schätzung für die Spektraldichte

Zunächst scheint es intuitiv, das Periodogramm als Schätzung für die Spektraldichte anzusehen und auch so in empirischen Untersuchungen zu verwenden. Dies ist aber durchaus mit Problemen behaftet, auf die wir nun eingehen werden<sup>8</sup>. Ansatzpunkt der Betrachtung ist die Erkenntnis von Fuller (1976), der für  $\lambda \neq 0$  und unter Verwendung eines genügend großen Stichprobenumfanges  $n$  zeigt, dass

$$\frac{2 \cdot I(\lambda)}{f(\lambda)} \stackrel{a}{\sim} \chi_2^2 \quad (2.52)$$

verteilt ist. Hieraus und aus dem Wissen, dass  $\chi_2^2$ -verteilte Zufallsvariablen einen Erwartungswert von zwei aufweisen, folgt mit

$$E[I(\lambda)] = f(\lambda), \quad (2.53)$$

dass das Periodogramm ein asymptotisch erwartungstreuer Schätzer für das Spektrum darstellt. Nichtsdestotrotz muss man aber festhalten, dass die Verteilung (2.52) impliziert, dass das 95%-Konfidenzintervall der Periodogramms mit  $[0,05f(\lambda); 7,38f(\lambda)]$  recht weit ausfällt und dass die Breite des Konfidenzbandes nicht kleiner wird, wenn der Beobachtungsumfang  $n$  von  $y_t$  steigt. Dies liegt daran, dass mit steigendem  $n$  auch gleichzeitig die Anzahl der Koordinaten des Periodogramms anwächst.

Um eine empirische Analyse zu erleichtern, existieren noch eine Reihe von Möglichkeiten, die Spektraldichte genauer zu schätzen. Neben parametrischen Ansätzen gibt es auch mit der Bartlett-Kerndichtenschätzung einen nicht-parametrischen Ansatz, der sich als

$$\hat{f}(\lambda) = \frac{1}{2\pi} \left\{ \hat{\gamma}_0 + 2 \sum_{j=1}^m \left[ 1 - \frac{j}{m+1} \right] \hat{\gamma}(j) \cos(\lambda j) \right\} \quad (2.54)$$

---

<sup>8</sup>Vgl. hierzu die Ausführungen von Hamilton (1994) S. 163 - 167

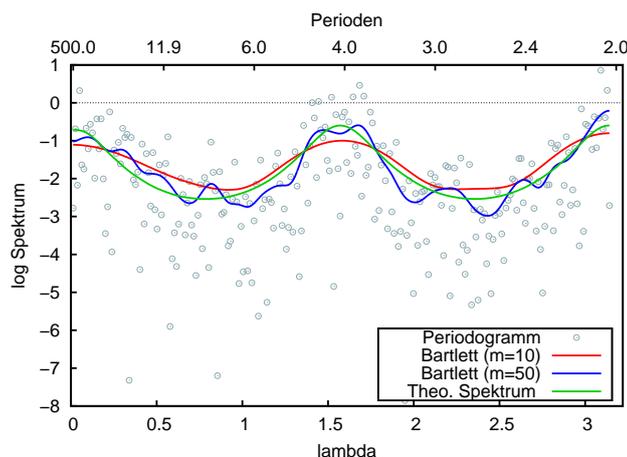


Abb. 2.11: Schätzung und Anpassungen des Periodogramms für einen saisonalen AR-Prozess

errechnet.  $m$  ist dabei ein vorher festzulegender Bandbreitenparameter und bewirkt, dass die Autokovarianzfunktion für  $j > m$  gesetzt wird zu  $\gamma(j) = 0$ . Dies wollen wir nun an einem Beispiel kurz demonstrieren, und zwar wollen wir hierfür einen saisonalen AR-Prozess mit

$$y_t = 0,5y_{t-4} + \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim N(0; 1) \quad (2.55)$$

und einen Beobachtungsumfang  $n = 500$  simulieren und schätzen hierfür die jeweiligen Periodogrammkoordinaten sowie die Bartlett-Kerndichtschätzung mit  $m = 10$  und  $m = 50$ . Das Ergebnis stellen wir wiederum mit dem theoretischen Spektrum des Prozesses graphisch in Abbildung 2.11 dar. Zusätzlich geben wir an einer zweiten Abszisse mit  $1/j$  den jeweiligen Kehrwert der jeweiligen Ordnung der Fourierfrequenzzahl  $\lambda_j$  an, was wiederum als Periodizität des Prozesses interpretiert werden kann. Hierbei können wir erkennen, dass zwar das Periodogramm den Ausschlag des Spektrums bei der Periode 4 andeutet, aber doch eine recht starke Streuung aufweist. Die Bartlett-Schätzungen sind ungleich glatter und können dem Betrachter wesentlich einfacher einen Aufschluss über eine vermutliche Periodizität des Prozesses geben. Je kleiner auch die Bandbreite  $m$  gewählt wird, desto glatter fällt die jeweilige Bartlett-Schätzung aus, da hier dann mehr Autokovarianzen auf null gesetzt werden. Die Wahl der Bandbreite ist hiermit auch von einer recht hohen Bedeutung, da ein zu hoch gewähltes  $m$  die Aussage über den vermuteten Verlauf des Spektrums schwieriger macht, aber ein zu niedriges  $m$  womöglich dazu führt, dass Periodizitäten des Prozesses unentdeckt bleiben könnten.

## 2.4 Gedächtnis von Zeitreihen

Die Definition des Gedächtnisses einer Zeitreihe geht auf die Arbeit von McLeod und Hipel (1978) zurück. Ausgangspunkt ist das Verhalten der Autokovarianzfunktion für unendlich ansteigende Ordnungen. Ein stochastischer Prozess hat ein langes Gedächtnis, falls dessen Autokovarianzfunktion nicht betragsmäßig aufsummierbar ist:

$$\sum_{j=-\infty}^{\infty} |\gamma(j)| \rightarrow \infty. \quad (2.56)$$

Synonym werden wir hierzu auch den Begriff der Long-Memory-Prozesse verwenden. Die Bedingung (2.56) impliziert, dass die Autokovarianzfunktion für  $j \rightarrow \infty$  nicht schnell genug oder gar nicht gegen null konvergiert, was eine unendlich lange Abhängigkeit des Prozesses bedeutet. Damit geht der Einfluss von Ereignissen, die sehr lange zurückliegen, nicht gegen null und haben immer noch eine Auswirkung auf gegenwärtige Realisierungen von  $y_t$ <sup>9</sup>. Diese Definition beruht ursprünglich auf dem Schätzproblem des Erwartungswerts von stationären Prozessen. Falls nämlich die Summe der Autokovarianzen absolut addierbar ist, d.h.  $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$ , dann ist das arithmetische Mittel mit der Konvergenzrate  $\sqrt{n}$  ein konsistenter Schätzer. Solche Prozesse werden dann häufig auch als Prozesse mit kurzem Gedächtnis oder Short-Memory-Prozesse bezeichnet<sup>10</sup>. Zu diesen Prozessen gehören alle stationären AR-Prozesse sowie sämtliche MA-Prozesse, die wir bisher beschrieben haben. In den folgenden Abschnitten wollen wir dann mögliche Long-Memory-Prozesse beschreiben und wiederum deren wichtigsten Eigenschaften diskutieren.

## 2.5 Fraktional integrierte Prozesse

Im Folgenden führen wir die Gruppe der fraktional integrierten Prozesse ein, die unter bestimmten Voraussetzungen ein langes Gedächtnis aufweisen. Die Basis hierfür bildet die Einführung der integrierten ARMA-Prozesse (abgekürzt als ARIMA( $p, d, q$ )-Prozesse) von Box und Jenkins (1976):

$$\theta(L)\Delta^d y_t = \phi(L)\varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim IID(0; \sigma_\varepsilon^2) \quad \text{und} \quad \Delta = (1 - L). \quad (2.57)$$

Die Idee dahinter ist, dass selbst ein nicht-stationärer Prozess nach  $d$ -maliger Differenzenbildung einen Integrationsgrad von null aufweist und somit als ein (statio-

---

<sup>9</sup>Vgl. Tschernig (1994) S. 17

<sup>10</sup>Vgl. hierzu Palma (2007) S. 39ff

närer) ARMA( $p, q$ )-Prozess beschrieben werden kann. Dabei wurde klassischerweise angenommen, dass  $d$  nur ganze positive Zahlen umfassen kann. Diese Restriktion werden wir mit Einführung der fraktional integrierten ARMA-Prozesse (ARFIMA-Prozesse) aufheben und nun annehmen, dass  $d$  jede beliebige reelle Zahl sein kann.

### 2.5.1 Fraktional differenziertes Rauschen

Um die Eigenschaften und Implikationen des fraktionalen Integrationsgrads  $d$  für ARFIMA-Zeitreihen zu untersuchen, bietet es sich an, zunächst deren einfachste Ausprägung  $p = q = 0$  mit

$$\Delta^d y_t = \varepsilon_t \quad \text{und} \quad \varepsilon_t \sim IID(0; \sigma_\varepsilon^2) \quad (2.58)$$

zu betrachten. Dieser Prozess wiederum lässt sich in eine MA-Darstellung umstellen zu

$$y_t = \Delta^{-d} \varepsilon_t = (1 - L)^{-d} \varepsilon_t = \pi(L) \varepsilon_t = \sum_{j=0}^{\infty} \pi_j L^j \varepsilon_t. \quad (2.59)$$

Diesen Prozess bezeichnen wir als fraktionales (differenziertes) Rauschen (auch abgekürzt als FR). Er lässt sich wiederum als einen unendlich langen MA-Prozess darstellen, dessen einzelne Terme wir mit  $\pi_j$  notieren werden. Diesem konnte Arendt (1974) mit Hilfe einer Potenzreihenentwicklung von  $(1 - L)^d$  eine Struktur verleihen:

$$\begin{aligned} (1 - L)^{-d} &= 1 + dL + \frac{d(d+1)}{2!} L^2 + \frac{d(d+1)(d+2)}{3!} L^3 + \dots \\ &\quad + \frac{d(d+1) \cdots (d+j-1)}{j!} L^j + \dots = \pi(L). \end{aligned} \quad (2.60)$$

Alternativ lassen sich die MA-Gewichte auch kompakt mit Hilfe der Gammafunktion  $\Gamma(z)$  darstellen<sup>11</sup>:

$$\begin{aligned} \pi_j &= \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \prod_{k=1}^j \frac{k-1+d}{k} \\ \text{mit } \Gamma(z) &= \begin{cases} \int_0^\infty \exp(-t)t^{z-1} dt, & \text{für } z > 0 \\ \infty, & \text{für } z = 0 \\ z^{-1}\Gamma(1+z), & \text{für } z < 0 \end{cases} \end{aligned} \quad (2.61)$$

<sup>11</sup>Vgl. hierzu Brockwell und Davis (1994) S. 522

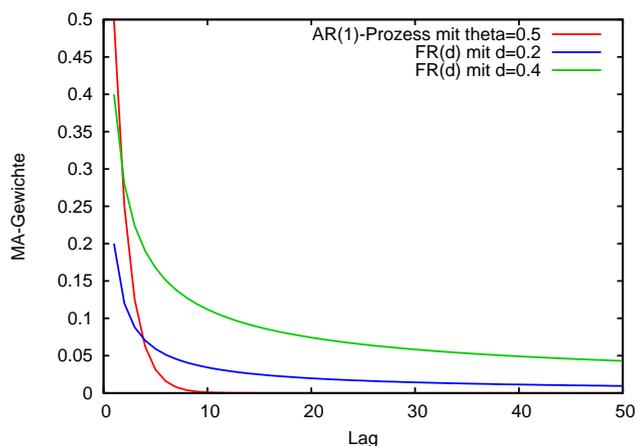


Abb. 2.12: Gewichte der MA-Darstellung  $\pi_j$  im Vergleich

In der Abbildung 2.12 wollen wir nun die Gewichte der Moving-Average-Darstellung eines fraktional integrierten Rauschens mit  $d = 0,2$  und  $d = 0,4$  mit der eines AR(1)-Prozesses mit  $\theta = 0,5$  vergleichen. Hier fällt auf, dass der AR(1)-Prozess zwar ein höheres Gewicht für weniger lang zurückliegende Störterme aufweist als die fraktionalen Rauschen, aber dass bereits ab einem Lag von zehn die Gewichte kaum noch von null verschieden sind. Die fraktionalen Rauschen hingegen messen auch sehr lange zurückliegenden Störtermen ein Gewicht  $\pi_j > 0$  zu, das umso stärker ausfällt, je größer der Integrationsgrad  $d$  ist.

Alternativ lassen sich die MA-Gewichte für große Lags mit Hilfe der Stirlingschen Formel approximieren mit

$$\pi_j \approx \frac{j^{d-1}}{\Gamma(d)} \quad \text{für } j \rightarrow \infty. \quad (2.62)$$

Mit dieser Darstellung wiederum lässt sich die Stationarität von fraktionalen Rauschen leichter definieren. Diese leitet sich in diesem Fall nicht notwendigerweise aus der MA-Struktur, da lediglich endlich lange MA-Prozesse stationär sind. Als Kriterium der Stationarität lässt sich auf die Summierbarkeit der quadrierten MA-Gewichte abstellen und somit muss<sup>12</sup>

$$\sum_{j=0}^{\infty} \pi_j^2 < \infty \quad (2.63)$$

gelten. Diese Bedingung gewährleistet im Endeffekt eine endliche Prozessvarianz von

<sup>12</sup>Vgl. hierzu Hamilton (1994) S. 52

$y_t$ . Setzen wir nun die Approximation der MA-Gewichte auf Basis der Stirlingschen Formel aus Gleichung (2.62) in diese Bedingung ein, erhalten wir die vereinfachende Stationaritätsbedingung

$$\sum_{j=0}^{\infty} j^{2(d-1)} < \infty. \quad (2.64)$$

Diese Reihe kann nur gegen einen endlichen Wert konvergieren, falls gilt:

$$2(d-1) < -1 \quad \Leftrightarrow \quad d < 0,5. \quad (2.65)$$

Damit wäre ein fraktional integriertes Rauschen mit  $d < 0,5$  immer stationär.

In der Abbildung 2.13 haben wir verschiedene Rauschprozesse mit dem Integrationsgrad  $d \in \{-0,4; 0; 0,4; 0,6; 0,8; 1,2; 1,4\}$  simuliert. Jede der einzelnen Prozesse wurde mit dem gleichen zugrunde liegenden Störterm  $\varepsilon_t \sim N(0; 1)$  simuliert, damit wir die Implikationen des Integrationsgrads vergleichen können. Es zeigt sich nämlich, dass mit steigendem  $d$  die Persistenz der Prozesse sukzessive zunimmt. Für  $d = -0,4$  und  $d = 0$  können wir prinzipiell kein träges Verhalten feststellen und tatsächlich besitzen diese Prozesse auch kein langes Gedächtnis im Gegensatz zu den Prozessen mit  $d > 0$ , was wir später sehen werden. Interessant ist auch der Vergleich der beiden Simulationen mit  $d = 0,4$  und  $d = 0,6$ , weil ersterer gerade noch stationär und zweiter schon zu den nicht-stationären Prozessen gehört. Bei graphischer Betrachtung ergibt sich zwar schon ein Unterschied in der Persistenz, aber dieser fällt nicht stark auf. Dies ist in der empirischen Anwendung in der Tat aber ein Problem, da durch die fraktionale Modellierung die Unterschiede zwischen Stationarität und Nicht-Stationarität sich fließender gestalten und manche Methoden dadurch schwieriger anzuwenden sind. Gleichzeitig ergibt sich aber dadurch natürlich ein weites Spektrum an weiteren Modellierungsmöglichkeiten, die besonders eine empirische Anwendung des fraktionalen Ansatzes interessant machen.

Die Spektraldichte von fraktional integrierten Rauschen lässt sich über die Betrachtung des Prozesses als MA-Prozess aus Gleichung (2.59) motivieren und ergibt sich als

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} |1 - \exp(-i\lambda)|^{-2d} = \frac{\sigma_\varepsilon^2}{2\pi} |2 \sin(\lambda/2)|^{-2d}. \quad (2.66)$$

Betrachten wir das Spektrum rechtsseitig in der Nähe seines Ursprungs  $\lambda \rightarrow 0+$  so

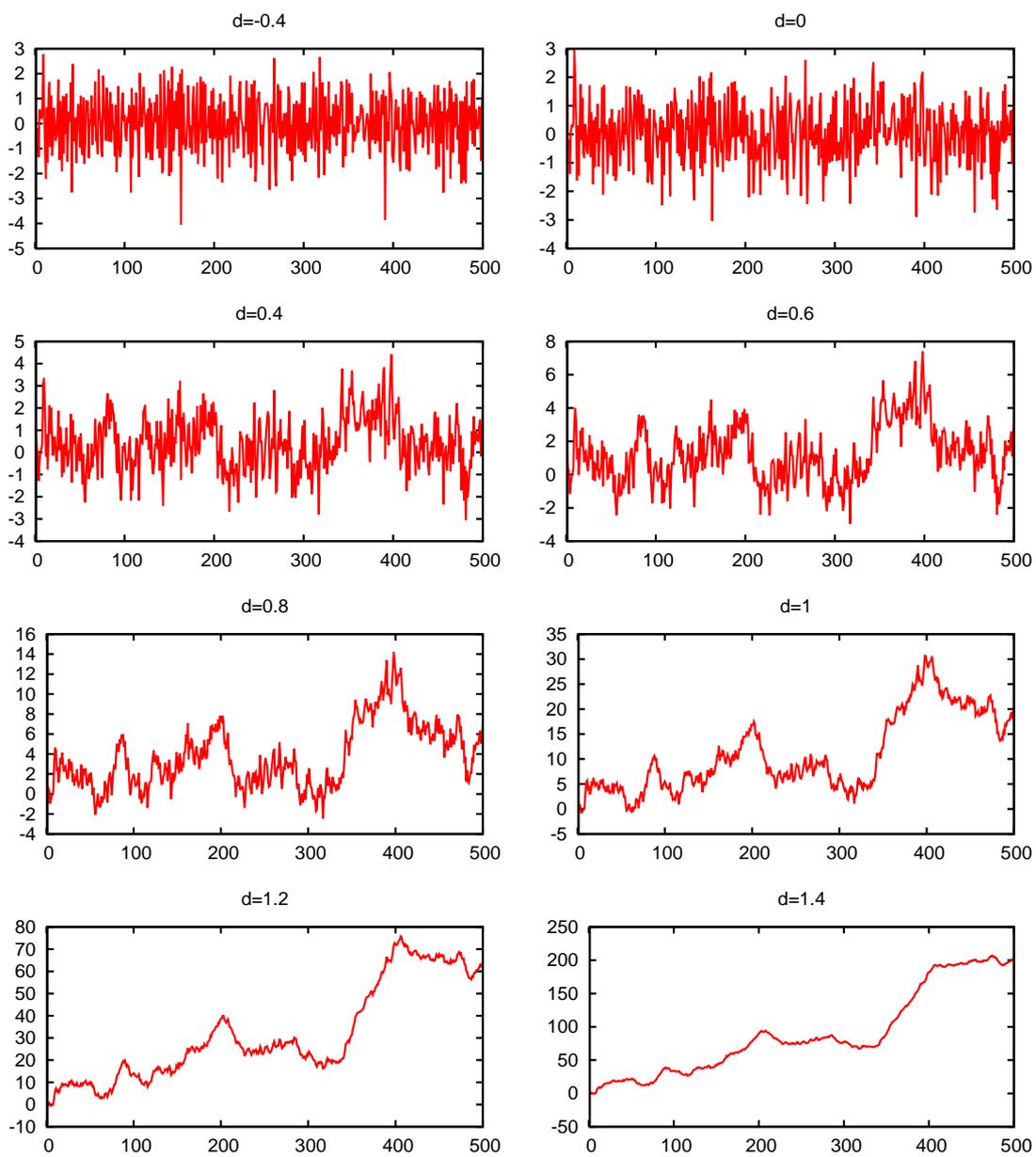


Abb. 2.13: Beispiele für fraktionale Rauschen

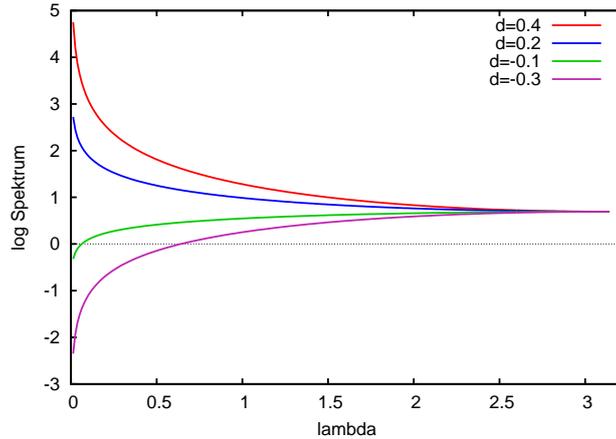


Abb. 2.14: Logarithmierte Spektren von fraktionalen Rauschen

ergibt sich mit  $\sin(\lambda) \approx \lambda$  die Approximation des Spektrums mit

$$f(\lambda) \approx \frac{\sigma_\varepsilon^2}{2\pi} \lambda^{-2d}, \quad (2.67)$$

was eine erhebliche Vereinfachung darstellt. Diese hat sich bereits in ähnlicher Weise bei Anwendung der Stirlingschen Formel für die MA-Darstellung gezeigt, da die Fourierfrequenzen nahe bei null sehr langen Schwingungen entsprechen und damit wiederum für die implizite Autokorrelation zu sehr hohen Lags. Die Spektraldichte für eine Reihe von fraktionalen Rauschen mit  $d \in \{-0,3; -0,1; 0,2; 0,4\}$  stellen wir in der Abbildung 2.14 dar, wo wir wiederum erkennen können, dass für  $d > 0$  die Steigung des Spektrums nahe des Ursprungs mit  $d$  größer wird, was sich wiederum durch die höheren Gewichte der hohen Lags der MA-Darstellung erklären lässt. Für negative Integrationsgrade haben wir hingegen interessanterweise ein Maximum bei  $\lambda = \pi$ , was nahe legt, dass solche Prozesse insgesamt stärker durch kürzere Schwingungen erklärt werden können. Dies verstärkt weiter die Vermutung, dass fraktionale Rauschen mit  $d < 0$  tendenziell eher nicht zu Long-Memory-Prozessen gehören. Die Autokovarianzfunktion motivieren Brockwell und Davis (1994) dann auf Basis des Spektrums mit dem Parsevalschen Theorem

$$\begin{aligned} \gamma(j) &= \int_{-\pi}^{\pi} f(\lambda) \exp(ij\lambda) d\lambda = \frac{2\sigma_\varepsilon^2}{2\pi} \int_0^{\pi} \cos(j\lambda) [2 \sin(\lambda/2)]^{-2d} d\lambda \\ &= \sigma_\varepsilon^2 \frac{(-1)^j \Gamma(1-2d)}{\Gamma(j-d+1)\Gamma(1-j-d)} \quad \text{für } j = 0, 1, 2, \dots \end{aligned} \quad (2.68)$$

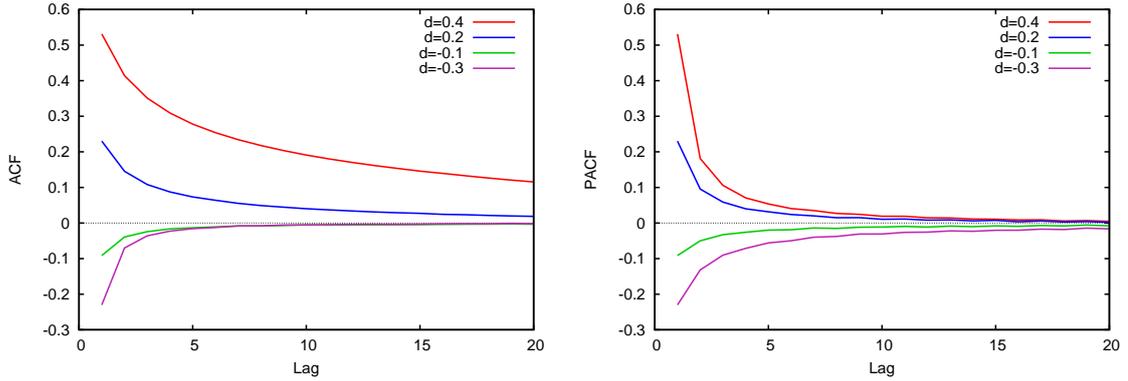


Abb. 2.15: (Partielle) Autokorrelationsfunktion von fraktionalem Rauschen

und speziell für die Varianz von  $y_t$  ergibt sich

$$\gamma(0) = \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)} \sigma_\varepsilon^2. \quad (2.69)$$

Hieraus lässt sich wieder die Autokorrelationsfunktion bestimmen, die wir wiederum in Abbildung 2.15 graphisch darstellen. Hierzu stellen wir auch die jeweilige PACF dar, die wir in diesem Fall durch eine Simulation bestimmt haben. Die Erkenntnisse hieraus bleiben prinzipiell die gleichen wie bei der Betrachtung der Spektraldichte in Abbildung 2.14. Interessant ist in diesem Zusammenhang höchstens, dass eine Modellierung mit fraktionalem Integrationsparameter sowohl eine lange ACF- und eine lange PACF-Struktur empirisch erklären kann, was sie wiederum mit ARMA( $p, q$ )-Prozessen gemeinsam haben. Ein reines fraktionales Rauschen hat aber den Vorteil, dass es lediglich von einem einzigen Parameter abhängt, was somit eine noch sparsamere Parametrisierung als für ARMA-Prozesse bedeutet und somit unter Umständen effizienter geschätzt werden kann. Die in der Gleichung (2.68) dargestellte Autokovarianzfunktion  $\gamma(j)$  lässt sich in veränderter Schreibweise für  $j \rightarrow \infty$  vereinfachen zu

$$\gamma(j) = \sigma_\varepsilon^2 \frac{\Gamma(1 - 2d)}{\Gamma(1 - d)\Gamma(d)} \cdot \frac{\Gamma(j + d)}{\Gamma(1 + j - d)} \approx \sigma_\varepsilon^2 \frac{\Gamma(1 - 2d)}{\Gamma(1 - d)\Gamma(d)} \cdot j^{2d-1}. \quad (2.70)$$

Diese Darstellung erlaubt uns nun eine Aussage über das Verhalten von  $\rho(j)$  für große Lags  $j$  zu treffen. Für  $d > 0$  konvergiert  $\gamma(j)$  für  $j \rightarrow \infty$  und wir haben somit dann den Fall, dass die Summe aller absoluten Autokovarianzen unendlich groß wird. Für diesen Fall beobachten wir damit eine Long-Memory-Eigenschaft für solche Prozesse. Der Fall  $d = 0$  entspricht einem einfachen Weißen Rauschen, wo wir keine einzige Autokovarianz ungleich null haben. In diesem Fall sprechen wir

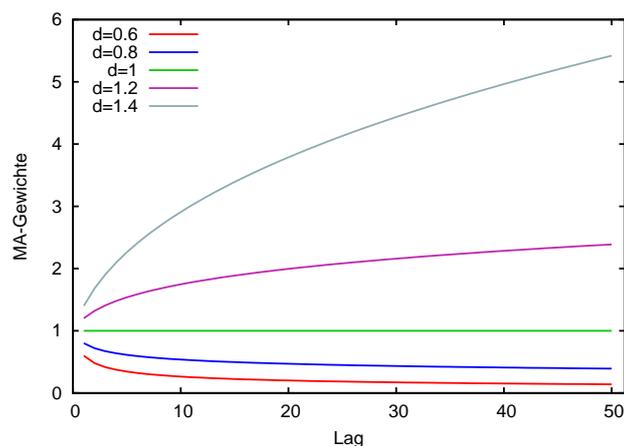


Abb. 2.16: Gewichte der MA-Darstellung von nicht-stationären fraktionalen Rauschen

davon, dass der Prozess überhaupt kein Gedächtnis hat. Letztendlich können wir noch festhalten, dass der Fall negativer Integrationsparameter  $d < 0$  eine gegen null konvergierende Autokovarianz impliziert. Dies wird von vielen Autoren auch häufig als Intermediate Memory bezeichnet, um sie von Long-Memory-Prozessen abzugrenzen.<sup>13</sup>

Wir wollen diesen Abschnitt mit einer näheren Betrachtung der nicht-stationären fraktionalen Rauschen beschließen. Somit betrachten wir nun Integrationsparameter  $d > 0,5$  und selbst hier lassen sich noch Unterschiede bezüglich des Prozessverhaltens feststellen. Wenn wir uns nochmal die Abbildung 2.13 in Erinnerung rufen, so fällt auf, dass die Prozesse mit  $d = 0,6$  und  $d = 0,8$  trotz ihrer Nicht-Stationarität eine Tendenz zum Prozesswert null und somit zum Startwert des Prozesses haben. Die anderen nicht-stationären Prozesse mit  $d \geq 1$  hingegen weisen nicht eine solche Tendenz auf und scheinen sich über alle Grenzen zu entwickeln. Dieses Verhalten kann nicht überraschen, wenn wir wiederum die Gewichte der einzelnen MA-Darstellungen betrachten, die für nicht-stationäre fraktionale Rauschen in Abbildung 2.16 dargestellt sind. Für  $d < 1$  haben wir ein abklingendes Verhalten der Gewichte auf länger zurückliegende Störterme, was wiederum impliziert, dass der Prozess  $y_t$  sich wieder zurück zu seinem Startwert entwickelt. Aus diesem Grund bezeichnet man solche Prozesse auch häufig als Mean-Reverting-Prozesse. Für  $d = 1$  hingegen beobachten wir konstant bei eins liegende und für  $d > 1$  sogar zunehmende Gewichte. Dieser Umstand verhindert aber, dass der Prozesswert  $y_t$  sich wieder zurück zu seinem Startwert entwickelt und sich im Zeitablauf immer weiter in positiver oder negativer

<sup>13</sup>Vgl. hierzu Brockwell und Davis (1994) S. 520

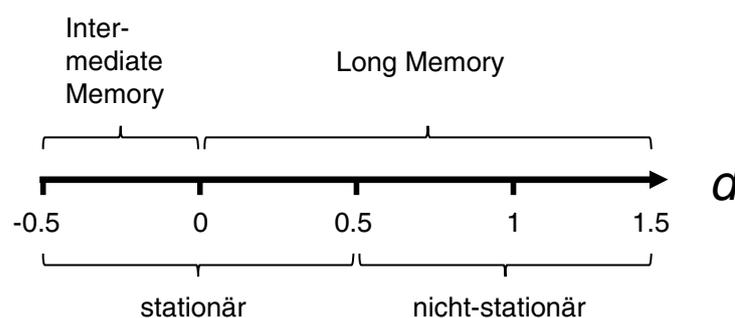


Abb. 2.17: Klassifizierung der Prozesseigenschaften in Abhängigkeit des Integrationsgrads

Richtung von ihm entfernt, was wir auch in der Abbildung 2.13 sehen können. Zusammenfassend aus diesem Abschnitt können wir festhalten, dass Prozesse mit  $d > 0$  eine Long-Memory-Eigenschaft aufweisen und mit  $d > 0,5$  nicht-stationär sind. Bis zu einem Integrationsgrad von  $d < 1$  haben Prozesse eine Mean-Reverting-Eigenschaft und kehren wieder zu ihrem Startwert bzw. unbedingten Erwartungswert zurück.

### 2.5.2 ARFIMA( $p, d, q$ )-Prozesse

Nachdem wir die Long-Memory-Eigenschaften von fraktionalen Rauschen untersucht haben, liegt es nahe, dass man zusätzlich das Auftreten von Short-Memory-Bestandteilen in empirisch zu untersuchenden Prozessen modelliert. Dies wiederum erreichen wir durch Hinzunahme weiterer ARMA-Bestandteile zu fraktionalen Rauschen, womit wir dann zu ARFIMA( $p, d, q$ )-Prozessen kommen. Die Prozessgleichung von ARFIMA( $p, d, q$ )-Prozessen haben wir bereits in Gleichung (2.57) aufgestellt und wollen sie nun in diesem Abschnitt nicht noch einmal wiederholen. Die Stationarität von ARFIMA-Prozessen hängt zum einen vom Integrationsgrad des Prozesses  $d$  und sowohl von Stationarität des AR-Teils ab, welche gegeben ist, wenn die Lösungen  $z_i$  der charakteristischen Gleichung  $|\theta(z)| = 0$  außerhalb des Einheitskreises liegen. Die Autokovarianzfunktion von ARFIMA-Prozessen gestaltet sich als recht kompliziert und soll in diesem Abschnitt nicht diskutiert werden, da wir verschiedene Approximationsweisen im folgenden Kapitel thematisieren werden. Eine Grundlage hierfür ist die Betrachtung der Spektraldichte, die sich für einen statio-

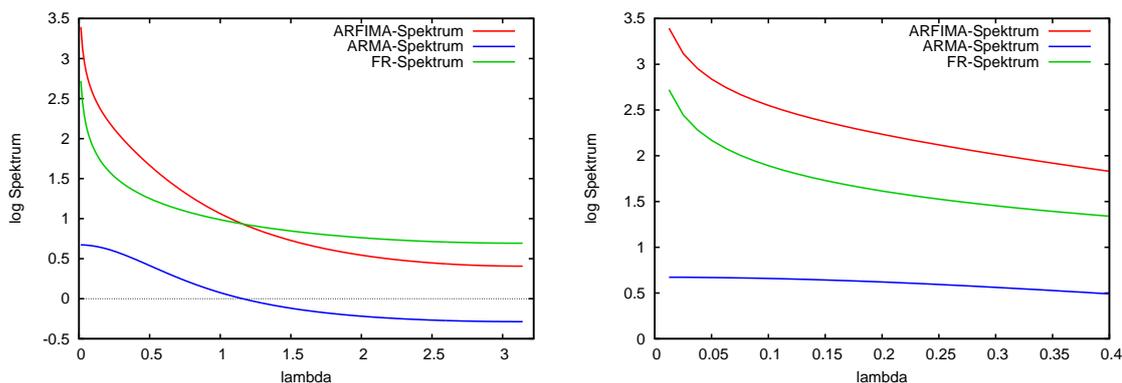


Abb. 2.18: Zusammensetzung des ARFIMA(1,  $d$ , 1)-Spektrums

nären ARFIMA( $p, d, q$ )-Prozess als<sup>14</sup>

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \underbrace{\left| \frac{1 - \sum_{j=1}^q \phi_j \exp(i\lambda j)}{1 - \sum_{k=1}^p \theta_k \exp(i\lambda k)} \right|^2}_{\text{ARMA-Spektrum}} \cdot \underbrace{[2 \sin(\lambda/2)]^{-2d}}_{\text{FR-Spektrum}} \quad (2.71)$$

ergibt und damit die Multiplikation des ARMA-Spektrums mit dem des fraktionalen Rauschen darstellt. Ähnlich wie für das fraktionale Rauschen lässt sich die Spektraldichte für den Fall  $\lambda \rightarrow 0+$  annähern mit

$$\begin{aligned} f(\lambda) &\approx \sigma_\varepsilon^2 \underbrace{\left( \frac{1 - \phi_1 - \dots - \phi_q}{1 - \theta_1 - \dots - \theta_p} \right)^2}_{\rightarrow \text{konstant}} [2 \sin(\lambda/2)]^{-2d} \\ &\approx G \lambda^{-2d} \quad \text{für } \lambda \rightarrow 0+. \end{aligned} \quad (2.72)$$

Hierbei zeigt sich, dass die AR- und MA-Teile des Prozesses zu einem von  $\lambda$  unabhängigen Bestandteil des Spektrums werden. Dies zusammengefasst mit der Varianz des Störterms  $\sigma_\varepsilon^2$  ergibt eine Konstante  $G > 0$ . Das Spektrum des fraktionalen Bestandteils dominiert die Spektraldichte des gesamten Prozesses in der näheren Umgebung  $\lambda = 0$ . Diese Eigenschaft des Spektrums wird noch für eine Reihe von Schätzmethoden im Verlauf dieser Arbeit von Bedeutung sein.

In der Abbildung 2.18 können wir die Motivation zur Approximation auch noch graphisch betrachten. Hier haben wir das Spektrum eines ARFIMA(1; 0,2; 1) mit  $\theta = 0,5$  und  $\phi = 0,3$  und dessen Prozessbestandteile auf logarithmierter Basis dargestellt, wobei sich in diesem Fall das Gesamtspektrum als Summe der Teilspektren ergibt. Hierbei können wir erkennen, dass der fraktionale Teil das ARFIMA-

<sup>14</sup>Vgl. hierzu Schlittgen und Streitberg (1994) S. 184 und S. 186

Spektrum nahe des Ursprungs dominiert, was in der zweiten Graphik von 2.18 durch die Beschränkung auf eine kleinere Anzahl von Fourierfrequenzen noch deutlicher wird. Die Steigung des Gesamtspektrums entspricht dort nahezu der Steigung des fraktionalen Spektrums, da der ARMA-Teil nahezu konstant verbleibt.

### 2.5.3 Alternative Long-Memory-Modelle

Neben dem ARFIMA-Modell existieren noch eine Reihe von weiteren Zeitreihenmodellen, die ebenfalls Long-Memory-Eigenschaften aufweisen. Z.B. haben Mandelbrot und Wallis (1969) fraktionale Gaußprozesse vorgeschlagen, die durch ihre Autokorrelationsfunktion

$$\rho(j) = \frac{1}{2} [|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}] \quad \text{mit} \quad 0 < H < 1 \quad (2.73)$$

charakterisiert werden können<sup>15</sup>. Die Prozesseigenschaften ergeben sich dann auf Basis des Parameters  $H$ , den Mandelbrot im Angedenken an den britischen Hydrologen Harold Edwin Hurst, dessen Studien der Speicherkapazität von Stauseen die Modellierung von langfristigen Abhängigkeiten mitbegründet haben, und an den deutschen Mathematiker Otto Hölder gewählt hat. Für  $H = 0,5$  beobachten wir hierbei einen Prozess ohne Korrelation entsprechend einem Weißen Rauschen und für  $0,5 < H < 1$  einen stationären Prozess mit Long Memory. Somit scheint es wiederum plausibel, dass wir für  $H = d + 0,5$  die gleichen asymptotischen Eigenschaften für fraktionale Gaußprozesse haben wie für fraktionale Rauschen mit einem Integrationsgrad von  $d$ . Der große Nachteil der fraktionalen Gaußprozesse ist aber, dass Short-Memory-Effekte im Gegensatz zu ARFIMA-Modellen nicht berücksichtigt werden können<sup>16</sup>. Eine weitere Möglichkeit zur Modellierung von Long-Memory-Effekten stellt das Gegenbauer-ARMA-Modell (häufig auch abgekürzt als GARMA-Modell) der Form

$$\theta(L)(1 - 2\xi + L^2)^\lambda y_t = \phi(L)\varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim IID(0; \sigma_\varepsilon^2) \quad (2.74)$$

dar, welches wiederum selbst eine Verallgemeinerung des ARFIMA( $p, d, q$ )-Modells darstellt. Das ARFIMA-Modell mit  $d = 2\lambda$  resultiert nämlich als spezieller GARMA-Prozess unter  $\xi = 1$ . Das GARMA-Modell ermöglicht aber keine genauere Modellierungsmöglichkeit der Long-Memory-Komponente, sondern bietet lediglich eine tiefer gehende Spezifizierung der Short-Memory-Eigenschaften, die über das eines ge-

<sup>15</sup>Vgl. hierzu Tschernig (1994) S. 47ff.

<sup>16</sup>Vgl. hierzu Tschernig (1994) S. 48

wöhnlichen ARMA-Modells hinausgehen<sup>17</sup>. Aus diesem Grund und da das GARMA-Modell eine relativ komplexe und unhandliche Spektraldichte aufweist, werden wir uns aber mit diesem in der folgenden Arbeit nicht weiter beschäftigen.

#### 2.5.4 Die empirische Relevanz von Long-Memory-Modellen

Wie wir in den vorangegangenen Abschnitten gesehen haben, eignen sich Long-Memory-Prozesse wie das ARFIMA-Modell, um persistente Zeitreihen zu beschreiben. Die theoretische Fundierung von ARFIMA scheint auch soweit plausibel, da es sich als Verallgemeinerung der relativ weit verbreiteten AR(I)MA-Modelle ergibt. Wir wollen uns aber nun in diesem Abschnitt die Frage stellen, inwieweit Long-Memory-Modelle in der Empirie relevant erscheinen und auf ein einfaches Beispiel anwenden.

Die ersten empirischen Anwendungen von Long-Memory-Modellen erfolgten im Bereich der Hydrologie, da insbesondere Wasserstände von Stauseen und Flüssen eine sehr langfristige zeitliche Abhängigkeit aufweisen und sich damit scheinbar in sehr langen Wellen entwickeln, was z.B. Hurst (1951) festhielt. Hosking (1980) motiviert mit seiner Arbeit ebenfalls im Bereich der Hydrologie das Modellieren mit fraktional integrierten Rauschprozessen und diskutierte die Eigenschaften derer, die wir auch teilweise in Abschnitt 2.5.1 beschrieben haben. In Lehrbüchern zu Long-Memory-Prozessen wie z.B. Palma (2007) wird aus diesem Grund häufig die Modellierung des Wasserstands des Nils als Beispiel angeführt, da für diesen über eine sehr lange Zeitspanne Daten verfügbar sind und sich deswegen ein sehr langes Gedächtnis in diesen Daten zeigt. Ein weiteres häufig verwendetes Beispiel für langfristige Abhängigkeiten findet sich im Bereich der Biologie bei der Betrachtung von Wachstumsprozessen von Bäumen. Auch hier kann über die Betrachtung von Baumringdicken das Wachstum auch von sehr lang zurückliegenden Perioden betrachtet und tendenziell eine Long-Memory-Eigenschaft dieser bestätigt werden.

Im Bereich der Wirtschaftswissenschaften war die Arbeit von Granger (1980) eine der ersten, die mögliche Long-Memory-Eigenschaften in ökonomischen Zeitreihen thematisierte. Granger (1980) argumentiert, dass eine Aggregation von dynamischen Zeitreihen, wie z.B. bei der Betrachtung von volkswirtschaftlichem Angebot und Nachfrage, die sich aus den individuellen Gegebenheiten der einzelnen Wirtschaftssubjekte ergeben, ein langes Gedächtnis ergeben kann. Dies zeigt Granger (1980) an dem Beispiel, dass, wenn individuelle Reihen für  $i = 1, 2, \dots, N$  erzeugt werden durch  $N$ -verschiedene AR(1)-Prozesse mit betaverteilten Autokorrelations-

---

<sup>17</sup>Vgl. hierzu Lildholdt (2000) S. 19

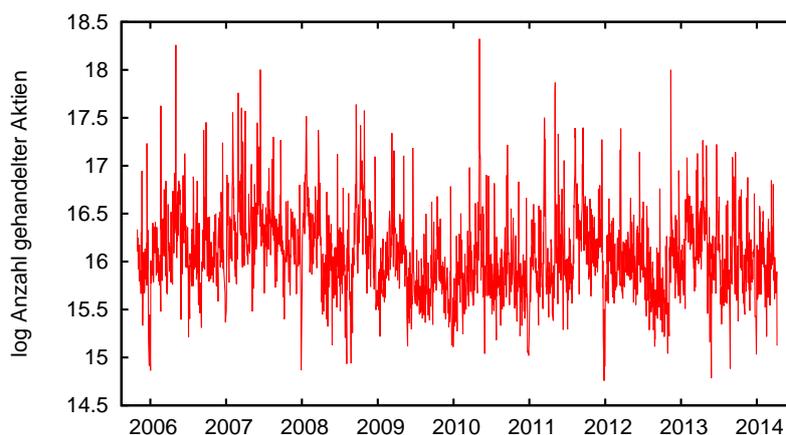


Abb. 2.19: Tägliche Handelsmenge der E.on-Aktie im Zeitraum November 2005 bis März 2014

koeffizienten  $\alpha_i$  der Form

$$y_{it} = \alpha_i y_{i,t-1} + \varepsilon_{it} \quad \text{mit} \quad \varepsilon_{it} \sim IID(0, \sigma_i^2) \quad \text{und} \quad \alpha_i \sim \text{Beta}(p, q). \quad (2.75)$$

In diesem Fall ergibt sich mit

$$\bar{y}_t = \frac{1}{N} \sum_{i=1}^N y_{it} \sim I(1 - q/2) \quad (2.76)$$

eine Long-Memory-Eigenschaft für das arithmetische Mittel<sup>18</sup>. Dieses Beispiel wiederum erscheint auf den ersten Blick nicht realitätsnah, zeigt aber, dass eine Zusammenfassung mehrerer dynamischer Zeitreihen eine Long-Memory-Eigenschaft bewirken kann<sup>19</sup>. Somit scheint eine Anwendung der ARFIMA-Modellierung auf ökonomische Zeitreihen wie Inflationsraten und BIP-Wachstumsdaten sinnvoll.

Wir wollen nun als Beispiel die tägliche Anzahl der gehandelten E.on-Aktien (WKN: ENAG999) betrachten, welche z.B. auf der Internetseite des Finanzportals Google Finance oder Ariva frei zugänglich sind. In der Abbildung 2.19 haben wir den Verlauf der logarithmierten Zeitreihe ab dem 31. Oktober 2005 bis 31. März 2014 dargestellt, was abzüglich aller Nicht-Handelstage in  $n = 2142$  Beobachtungspunkten resultiert. Bei einer rein graphischen Beurteilung des zeitlichen Ablaufs kann man zu der Beurteilung kommen, dass die tägliche Handelsmenge eher stationäre Merkmale aufweist, was auch durch die Betrachtung der empirisch ermittelten Autokorrelationsfunktion

<sup>18</sup>Siehe hierzu Lildholdt (2000) S. 20

<sup>19</sup>Vgl. hierzu Lildholdt (2000) S. 21

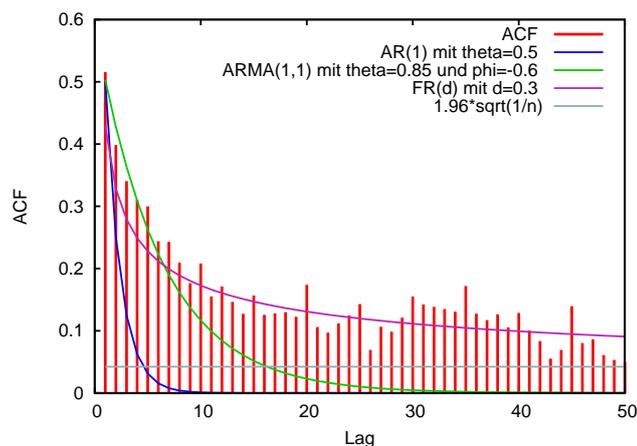


Abb. 2.20: ACF der täglichen Handelsmenge der E.on-Aktie

in Abbildung 2.20 verstärkt wird, da deren Maximum kaum über dem Wert 0,5 liegt. Die Anwendung von gebräuchlichen ARMA-Schätzverfahren legt für eine gewählte Spezifikation eines AR(1)-Modells einen AR-Parameter von  $\theta = 0,5$  nahe und für ein ARMA(1,1)-Modell die Parameterkonstellation  $\theta = 0,85$  und  $\phi = -0,6$ . Auf Basis dieser Werte haben wir in der Abbildung 2.20 die jeweiligen theoretischen Autokorrelationsfunktionen hinzugefügt und können hierbei erkennen, dass besonders das AR(1)-Modell kaum den Verlauf der ACF erklären kann. Das ARMA(1,1)-Modell mit der erwähnten Parametrisierung passt zwar hingegen besser, kann aber besonders die Autokorrelationen ab einem Lag von ungefähr zehn nicht mehr nachvollziehen. Diese sind aber immer noch bis zu einem Lag von 50 signifikant von null verschieden bei einem Irrtumsniveau von 5 %<sup>20</sup>, was wiederum aber mit Hilfe eines einfachen fraktionalen Rauschen mit Integrationsgrad  $d = 0,3$  relativ gut beschrieben werden kann.

Dieser Eindruck, dass ein fraktionales Rauschen die statistischen Eigenschaften der täglichen Handelsmenge der E.on-Aktie relativ gut beschreiben kann, verfestigt sich auch durch die Betrachtung des Periodogramms in Abbildung 2.21. Hier vergleichen wir auch die jeweiligen modelltheoretischen Spektraldichten mit den geschätzten Periodogrammkoordinaten, welches wir auch in einer nach Bartlett mit einer Bandbreite von  $m = 92$  geglätteten Version darstellen. Insbesondere das Spektrum des fraktionalen Rauschen trifft dabei den Verlauf des geglätteten Periodogramms gut, während die reinen Short-Memory-Modelle die Struktur weniger nachvollziehen können.

<sup>20</sup>Der Annahmebereich für eine hypothetische Autokorrelation von null ergibt sich bei einem beliebigen Lag im vorliegenden Fall als  $\pm 1,96/\sqrt{2142} \approx \pm 0,0423$ , was auch in der Abbildung 2.20 dargestellt ist.

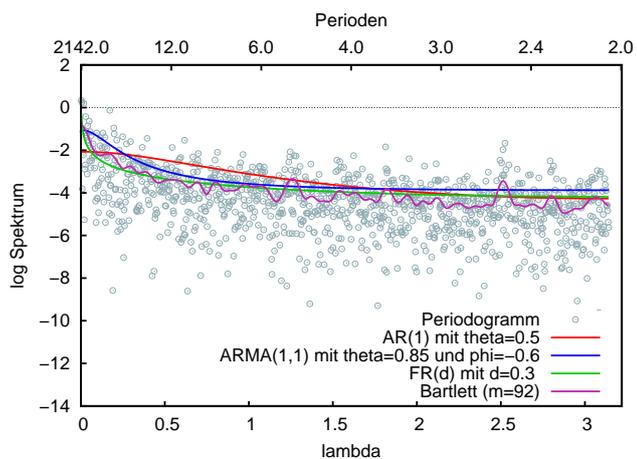


Abb. 2.21: Periodogramm der täglichen Handelsmenge der E.on-Aktie

Wir sehen also, dass Long-Memory-Modelle auch in der empirischen Modellierung der Wirtschaftswissenschaften ein Anwendungsgebiet finden und wollen hiermit dieses grundlegende Kapitel beschließen, während wir im folgenden die Schätzung des Integrationsgrads diskutieren werden.

### 3 Schätzverfahren zur Bestimmung des Integrationsgrads

Im Allgemeinen unterscheiden wir zwei verschiedene Gruppen von Schätzern für den Integrationsgrad, nämlich parametrische und semiparametrische Schätzverfahren. Parametrische Schätzer sind geeignet, die Parameter eines ARFIMA( $p, d, q$ )-Prozesses zu schätzen, soweit man die Ordnung kennt bzw. man sie mit einem geeigneten Verfahren bestimmen kann. Dabei gilt es zu bedenken, dass bei einer Fehlspezifizierung der Ordnungen  $p$  und  $q$  die Konsistenz der Schätzergebnisse gefährdet ist und die Anwendung der Verfahren somit behutsam erfolgen sollte. Bei den semiparametrischen Verfahren hingegen ist die Kenntnis der Ordnung eines ARFIMA-Prozesses nicht unbedingt nötig, dafür sind diese in der Regel ineffizienter. Ein weiteres Unterscheidungsmerkmal für die verschiedenen Schätzer sind die Schätzmethoden. Diesbezüglich existieren regressionsbasierende Methoden als auch Maximum-Likelihood-Schätzer. Im Folgenden sollen nun die verschiedenen Verfahren vorgestellt und am Ende auch miteinander verglichen werden.

#### 3.1 Parametrische Maximum-Likelihood-Schätzung

Für einen stationären gaussianischen Prozess  $y_t$  mit Mittelwert von 0 und Beobachtungen  $t = 1, \dots, n$  können wir die logarithmierte Likelihood-Funktion angeben mit

$$L(\boldsymbol{\psi}) = -\frac{1}{2} \ln \det \boldsymbol{\Gamma}_{\boldsymbol{\psi}} - \frac{1}{2} \mathbf{y}' \boldsymbol{\Gamma}_{\boldsymbol{\psi}}^{-1} \mathbf{y}. \quad (3.1)$$

Die Elemente des Vektors  $\mathbf{y} = (y_1, \dots, y_n)'$  sind dabei die Ausprägungen des Prozesses  $y_t$ , die Matrix  $\boldsymbol{\Gamma}_{\boldsymbol{\psi}} = \text{Var}(\mathbf{y})$  ist die Kovarianzmatrix von  $\mathbf{y}$ , die sich aus dem Parametervektor des Prozesses  $\boldsymbol{\psi} = (d, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_q)'$  ergibt. Die Maximum-Likelihood-Schätzung  $\hat{\boldsymbol{\psi}}$  wird dann durch die Maximierung von  $L(\boldsymbol{\psi})$  bestimmt.<sup>21</sup> Die Probleme dieser Methode in der Praxis sind zum einen, dass die Struktur des Vektors  $\boldsymbol{\psi}$  bekannt sein muss, d.h. falls  $y_t$  durch einen ARFIMA( $p, d, q$ )-Prozess charakterisiert wird, müssen die Ordnungen  $p$  und  $q$  bekannt sein. Des Weiteren muss die Determinante und die Inverse der Kovarianzmatrix bestimmt werden, was relativ schwierig in der Praxis ist.

Im Folgenden wollen wir eine mögliche Umsetzung diskutieren.

<sup>21</sup>Vgl. hierzu die Ausführungen von Palma (2007) S. 66 ff

### 3.1.1 Die Cholesky-Zerlegung

Da es sich bei  $\mathbf{\Gamma}_\psi$  um eine Kovarianzmatrix handelt, wissen wir, dass sie symmetrisch und positiv definit ist. Damit lässt sich folgende Faktorisierung mit Hilfe der Cholesky-Zerlegung durchführen:

$$\mathbf{\Gamma}_\psi = \mathbf{U}\mathbf{U}', \quad (3.2)$$

wobei es sich bei  $\mathbf{U}$  um eine obere Dreiecksmatrix handelt. Auf Grund dieses Umstandes ist die Determinante von  $\mathbf{\Gamma}_\theta$  bestimmt durch

$$\det \mathbf{\Gamma}_\psi = (\det \mathbf{U})^2 = \prod_{j=1}^n u_{jj}^2, \quad (3.3)$$

wobei  $u_{jj}$  das  $j$ -te Hauptdiagonalelement von  $\mathbf{U}$  darstellt. Die Inverse von  $\mathbf{\Gamma}_\psi$  lässt sich dann durch

$$\mathbf{\Gamma}_\psi^{-1} = (\mathbf{U}^{-1})'\mathbf{U}^{-1} \quad (3.4)$$

unter der Annahme, dass  $\mathbf{\Gamma}$  bekannt ist, bestimmen. Das Problem des Verfahren ist jedoch, dass es gerade bei längeren Zeitreihen ineffizient wird.

### 3.1.2 Die Berechnung der Kovarianzmatrix

In diesem Abschnitt werden wir versuchen, die Kovarianz eines ARFIMA( $p, d, q$ )-Prozesses zum einen exakt zu bestimmen und zum anderen eine für die Praxis eher gebräuchliche approximative Variante herzuleiten.<sup>22</sup> Im vorherigen Kapitel haben wir die Spektraldichte eines ARFIMA( $p, d, q$ )-Prozesses bestimmt. Aus dieser Formel lässt sich gemäß des Parsevalschen Theorems wieder die Kovarianz gewinnen, indem folgende Umkehrformel angewandt wird:

$$\begin{aligned} \gamma_t &= \int_{-\pi}^{\pi} f(\lambda) \exp(-i\lambda t) d\lambda \\ &= \frac{1}{2\pi} \sigma_\varepsilon^2 \int_{-\pi}^{\pi} \frac{|\phi[\exp(-i\lambda)]|^2}{|\theta[\exp(-i\lambda)]|^2} |1 - \exp(-i\lambda)|^{-2d} \exp(i\lambda t) d\lambda \end{aligned} \quad (3.5)$$

---

<sup>22</sup>Vgl. hierzu Palma (2007) S. 47 ff

Für den AR-Teil  $\theta(L)$ -des ARFIMA-Prozesses verwenden wir folgende Faktorisierung

$$\theta(L) = \prod_{i=1}^p (1 - \rho_i L). \quad (3.6)$$

Unter der Annahme, dass das Produkt dieser Lösung genau eins ergibt, konnte Sowell (1992) folgende Berechnungsweise der Kovarianzfunktion angeben:

$$\gamma_h = \sigma_\varepsilon^2 \sum_{i=-q}^q \sum_{j=1}^p \varphi(i) \xi_j C(d, p + i - h, \rho_j), \quad (3.7)$$

$$\text{mit } \varphi(i) = \sum_{k=\max(0,i)}^{\min(q,q+i)} \beta_k \beta_{k-i}$$

$$\text{und } \xi_j = \left[ \rho_j \prod_{i=1}^p (1 - \rho_i \rho_j) \prod_{m \neq j} (\rho_j - \rho_m) \right]^{-1}$$

$$C(d, h, \rho_j) = \frac{\gamma_0(h)}{\sigma_\varepsilon^2} [\rho_j^{2p} F(d + h, 1, 1 - d + h, \rho_j) + F(d - h, 1, 1 - d - h, \rho_j) - 1].$$

Dabei umschreibt  $F(a, b, c, x)$  die gaussianische hypergeometrische Funktion, die sich wie folgt berechnet:

$$F(a, b, c, x) = 1 + \frac{a \cdot b}{c \cdot 1} x + \frac{a \cdot (a + 1) \cdot b \cdot (b + 1)}{c \cdot (c + 1) \cdot 1 \cdot 2} x^2 + \dots \rightarrow \infty. \quad (3.8)$$

Diese Berechnungsweise der einzelnen Kovarianzelemente lässt bereits erkennen, dass diese sich in der Praxis als äußerst rechenintensiv darstellt und somit eine empirische Anwendung erschwert. Implementiert ist dieses Verfahren mit exakter Berechnungsweise der Autokovarianzfunktion aktuell nur in dem Softwarepaket OxMetrics, das in der Programmiersprache Ox geschrieben ist. Da wir später die Eigenschaften des Schätzverfahrens in Simulationen bewerten wollen, werden wir im Folgenden einige Approximationsvorschläge von verschiedenen Forschern kennenlernen.

### 3.1.3 Autoregressive Approximation nach Hasslett und Raftery

Eine autoregressive Approximationsweise schlagen Hasslett und Raftery (1989) vor. Sie prognostizieren die stationäre Sequenz  $\{y_t\}$ , die einer ARFIMA( $p, d, q$ )-Ordnung

folgt, jeweils um einen Schritt im voraus mit

$$\hat{y}_t = \theta(L)\phi(L)^{-1} \sum_{j=1}^{t-1} \tilde{\theta}_{tj} y_{t-j}, \quad (3.9)$$

wobei  $\tilde{\theta}_{tj} = - \binom{t}{j} \frac{\Gamma(j-d)\Gamma(t-d-j+1)}{\Gamma(-d)\Gamma(t-d+1)}$ .

Diese Vorhersage wiederum weist die folgende Varianz auf:

$$v_t = \text{Var}(y_t - \hat{y}_t) = \sigma_y^2 \kappa \prod_{j=1}^{t-1} (1 - \tilde{\theta}_{jj}^2) \quad (3.10)$$

mit  $\sigma_y^2 = \text{Var}(y_t)$  und  $\kappa$  als das Verhältnis der Störtermvarianz zur Varianz der ARMA( $p, q$ )-Bestandteile. Problematisch wird nun die Berechnung der Koeffizienten  $\tilde{\theta}_{tj}$ , wenn der Beobachtungsumfang  $n \rightarrow \infty$  geht, da deren Anzahl dann stark steigt. Deswegen machen Hasslett und Raftery (1989) den Vorschlag, die Anzahl der Koeffizienten auf  $M$  zu beschränken mit

$$\sum_{j=1}^{t-1} \tilde{\theta}_{tj} y_{t-j} \approx \sum_{j=1}^M \tilde{\theta}_{tj} y_{t-j} - \sum_{j=M+1}^{t-1} \pi_j y_{t-j}, \quad (3.11)$$

da sich  $\tilde{\theta}_{tj} \propto -\pi_j$  für große  $j$  verhalten. Der zweite Term der rechten Seite der Gleichung wird dann wiederum approximiert mit

$$\sum_{j=M+1}^{t-1} \pi_j y_{t-j} \approx M \tilde{\theta}_{t-1, M} d^{-1} [1 - (M/t)^d] \bar{y}_{M+1, t-1-M}, \quad (3.12)$$

wobei  $\bar{y}_{M+1, t-1-M} = (t-1-2M)^{-1} \sum_{j=M+1}^{t-1-M} y_j$ . Somit kann der Parametervektor  $\boldsymbol{\psi} = \{d, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_q\}$  durch die Maximierung von

$$l(\boldsymbol{\psi}) = -\frac{n}{2} \ln(\hat{\sigma}^2(\boldsymbol{\psi})) \quad (3.13)$$

mit  $\hat{\sigma}^2(\boldsymbol{\psi}) = \frac{1}{n} \sum_{t=1}^n \frac{(y_t - \hat{y}_t)^2}{v_t}$

geschätzt werden. Programme zur Durchführung der Approximation sind derzeit für Fortran und R öffentlich zugänglich. Der Wertebereich für  $d$  ist auf positive, stationäre Werte, also  $d \in [0; 0,5]$  beschränkt. Crato und Rao (1996) modifizieren den Algorithmus dahingehend, dass sie den Wertebereich auf  $[-0,5; 0,5]$  erweitern

können.

### 3.1.4 Tanaka-Approximation

Eine rechentechnische effiziente Approximationsweise hat Tanaka (1999) vorgeschlagen. Für das einfache Modell

$$\Delta^d y_t = \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim N(0; \sigma^2) \quad \text{für} \quad t = 1, \dots, n \quad (3.14)$$

kann er aufgrund der Annahme der Normalverteilung der Störterme die logarithmierte Likelihood angeben mit

$$L(d, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^n (\Delta^d y_t)^2, \quad (3.15)$$

die sich wiederum konzentrieren lässt zu

$$l(d) = -\frac{n}{2} \ln \left[ \sum_{t=1}^n (\Delta^d y_t)^2 \right]. \quad (3.16)$$

Für diese Situation kann er mit Hilfe von Standard-ML-Methoden die asymptotische Verteilung von  $\hat{d}$  angeben mit

$$\sqrt{n}(\hat{d} - d) \xrightarrow{a} N(0; 6/\pi^2). \quad (3.17)$$

Die Annahme einer Normalverteilung der Störterme ist zwar eine sehr restriktive Annahme, aber trotzdem ist es bemerkenswert, dass für den Integrationsparameter  $d$  keinerlei Restriktionen gelten, so dass auch nicht-stationäre Prozesse untersucht werden können.

Hierzu wollen wir nun einige Simulationsergebnisse vorstellen. Es werden 100.000 fraktional integrierte Prozesse der Form (3.14) mit  $n = 1000$  und  $d = 0,4$  bzw.  $d = 1,4$  erzeugt. Als Störterm erzeugen wir zum einen standardnormalverteilte Zufallsvariablen als auch t-verteilte mit zwei Freiheitsgraden, um zu sehen, wie sich die Approximationsmethode bei einer Abkehr von der Normalverteilungsannahme von  $\varepsilon_t$  verhält. Des Weiteren erzeugen wir Störterme, die eine GARCH(1,1)-Ordnung in der Form

$$\varepsilon_t \sim N(0; \sigma_t^2) \quad \text{mit} \quad \sigma_t^2 = 0,1 + 0,1\varepsilon_{t-1}^2 + 0,8\sigma_{t-1}^2 \quad (3.18)$$

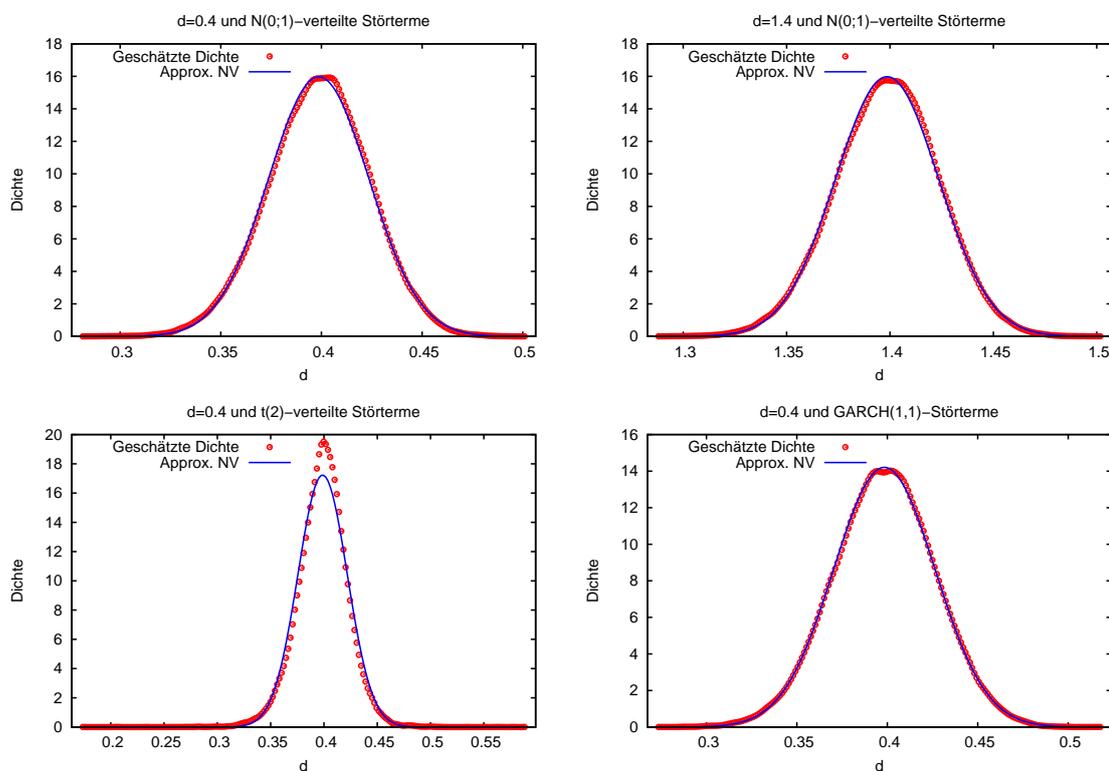


Abb. 3.1: Geschätzte Dichtefunktionen für fraktionales Rauschen mit  $d = 0,4$  und  $d = 1,4$

Tab. 3.1: Verteilungsparameter von  $\hat{d}$  bei fraktionalem Rauschen

$d$	MW	SD	S	ExK
$N(0;1)$ -verteilte Störterme				
0,4	0,3987	0,0250	-0,1089	0,0475
1,4	1,3986	0,0249	-0,0941	0,0219
$t(2)$ -verteilte Störterme				
0,4	0,3988	0,0231	-0,1731	2,1367
1,4	1,3988	0,0232	-0,1856	2,0990
GARCH(1,1)-Störterme				
0,4	0,3983	0,0281	-0,0607	0,0240
1,4	1,3984	0,0282	-0,0509	0,0791

MW ist hier die Abkürzung für Mittelwert, SD für Standardabweichung, S für Schiefe und ExK für den Exzess (=Kurtosis-3)

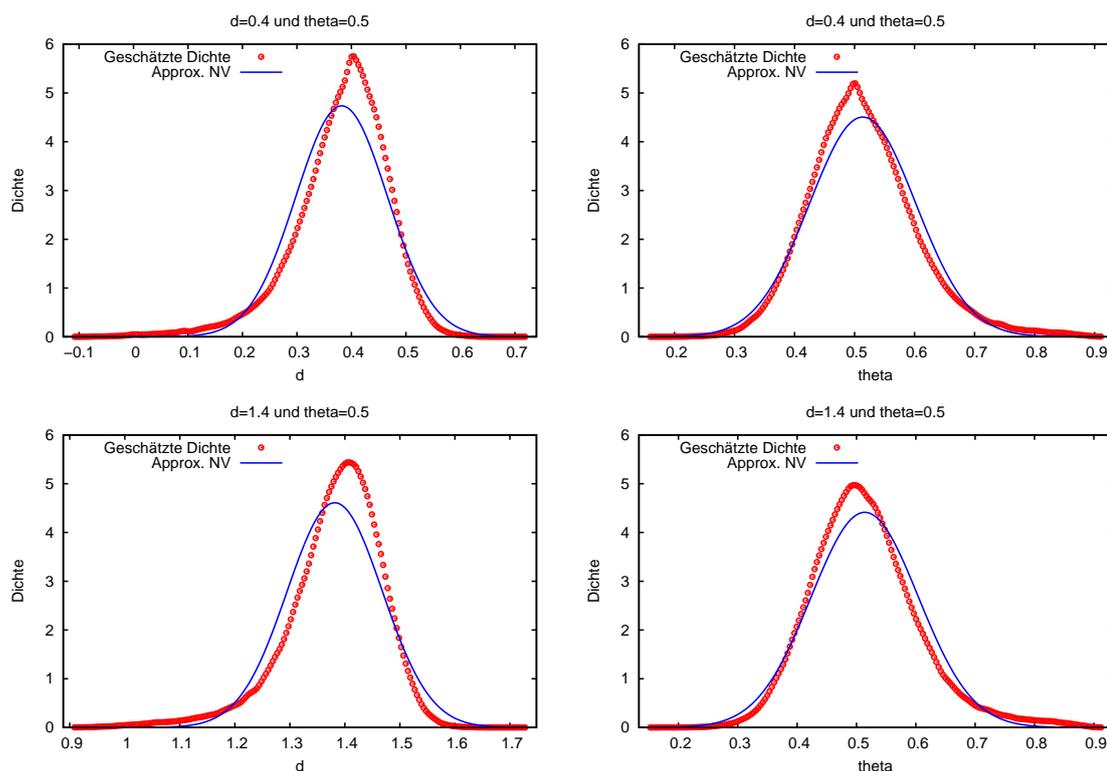


Abb. 3.2: Geschätzte Dichtefunktionen der Schätzergebnisse für ARFIMA(1, $d$ ,0)-Prozesse mit  $\theta = 0,5$  und  $d = 0,4$  bzw.  $d = 1,4$

haben, was eine relativ stark persistente Störtermvarianz bedeutet.

Die Abbildung 3.1 zeigt jeweils die mit einem Kerndichtenschätzer ermittelten Dichtefunktionen für  $\hat{d}$  mit einer approximierten Normalverteilung für ausgewählte Fälle. Die Tabelle 3.1 enthält die jeweiligen Verteilungsparameter für alle Situationen.

Für den Fall standardnormalverteilter Störterme ergibt sich optisch eine gute Anpassung an eine Normalverteilung. Führt man aber übliche Normalitätstests für die geschätzten Werte  $\hat{d}$  durch, so ergibt sich eine Ablehnung der Normalverteilung auf allen üblichen Signifikanzniveaus. Im Mittel werden die wahren Parameter gut getroffen und auch die Standardabweichung, für die wir  $\sqrt{6/(\pi^2 \cdot 1000)} \approx 0,0247$  erwarten, wird nahezu erreicht.

Für den Fall der  $t$ -verteilen Störterme hingegen ergibt sich eine stärkere Abweichung von der Normalverteilung der geschätzten Integrationsparameter, während bei den GARCH-Störtermen diese wiederum gut getroffen wird. Etwas überraschend ist, dass die Standardabweichung bei  $t$ -verteilten Störtermen sogar geringer ist als beim standardnormalverteilten Fall.

Tanaka (1999) erörtert auch den Fall, dass  $y_t$  einer ARFIMA( $p, d, q$ )-Ordnung

Tab. 3.2: Verteilungsparameter der geschätzten Parameter bei ARFIMA(1,  $d$ , 0)-Prozessen mit  $d = 0,4$

	$\theta = -0,5$		$\theta = 0$		$\theta = 0,5$	
	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$
MW	0,3974	-0,4975	0,3936	0,0039	0,3821	0,5136
SD	0,0299	0,0330	0,0394	0,0505	0,0842	0,0885
S	-0,1429	0,1596	-0,3482	0,2143	-0,9448	0,5769
ExK	0,0458	0,0591	0,6967	0,4338	1,9486	0,9829

MW ist hier die Abkürzung für Mittelwert, SD für Standardabweichung, S für Schiefe und ExK für den Exzess (=Kurtosis-3)

Tab. 3.3: Verteilungsparameter der geschätzten Parameter bei ARFIMA(1,  $d$ , 0)-Prozessen mit  $d = 1,4$

	$\theta = -0,5$		$\theta = 0$		$\theta = 0,5$	
	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$
MW	1,3975	-0,4978	1,3935	0,0074	1,3816	0,5140
SD	0,0300	0,0332	0,0385	0,0500	0,0865	0,0904
S	-0,1293	0,1462	-0,4301	0,2739	-1,0218	0,6488
ExK	0,0635	0,0279	0,9473	0,6019	2,1159	1,0977

MW ist hier die Abkürzung für Mittelwert, SD für Standardabweichung, S für Schiefe und ExK für den Exzess (=Kurtosis-3)

$$\Delta^d \theta(L) y_t = \phi(L) \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim N(0; \sigma^2) \quad \text{für} \quad t = 1, \dots, n \quad (3.19)$$

folgt, wobei die Lösungen der charakteristischen Gleichung  $z$  der AR-Bestandteile  $\theta(z) = 0$  und für die MA-Bestandteile  $\phi(z) = 0$  allesamt außerhalb des Einheitskreises liegen. Die konzentrierte logarithmierte Likelihood ergibt sich dann als

$$l(\psi) = -\frac{n}{2} \ln \left\{ \sum_{t=1}^n [\phi^{-1}(L) \theta(L) \Delta^d y_t]^2 \right\} \quad (3.20)$$

mit  $\psi = \{d, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_q\}$  und Tanaka (1999) zeigt auch für diesen Fall die folgende asymptotische Verteilung auf:

$$\sqrt{n}(\hat{\psi} - \psi) \xrightarrow{a} N(\mathbf{0}, \Xi^{-1}) \quad (3.21)$$

$$\text{mit} \quad \Xi = \begin{pmatrix} \pi^2/6 & \boldsymbol{\kappa}' \\ \boldsymbol{\kappa} & \Phi \end{pmatrix}, \quad \boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_p, \lambda_1, \dots, \lambda_q)$$

$$\text{und} \quad \kappa_i = \sum_{j=1}^{\infty} \frac{1}{j} c_{j-i}, \quad \lambda_i = -\sum_{j=1}^{\infty} \frac{1}{j} d_{j-i}.$$

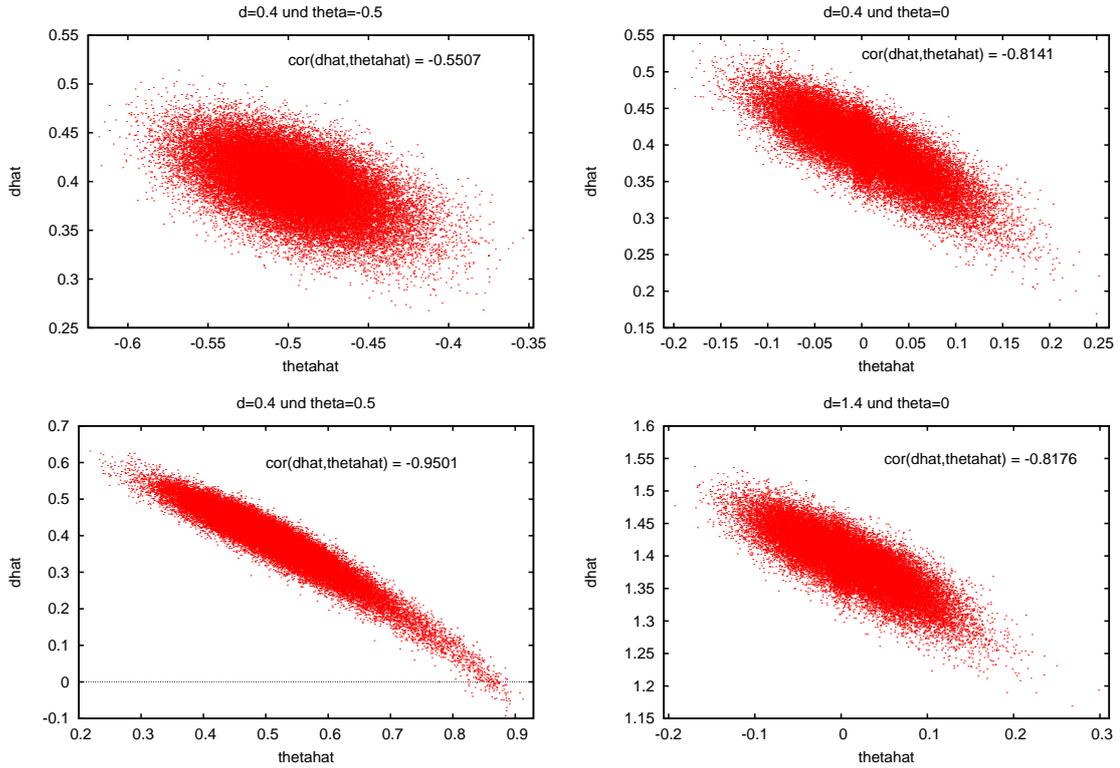


Abb. 3.3: Streudiagramme der Schätzergebnisse  $\hat{d}$  und  $\hat{\theta}$

Dabei stellen  $c_j$  und  $d_j$  die Koeffizienten von  $L^j$  der Entwicklungen von  $\theta(L)^{-1}$  und  $\phi(L)^{-1}$  dar und  $\Phi$  ist die Fisher-Informationsmatrix von  $\theta = \{\theta_1, \dots, \theta_p\}$  und  $\phi = \{\phi_1, \dots, \phi_q\}$ . Bezogen auf die Schätzung des Integrationsparameters bedeutet dies, dass mit

$$\sqrt{n}(\hat{d} - d) \xrightarrow{a} N(0, (\pi^2/6 - \kappa' \Phi^{-1} \kappa)^{-1}) \quad (3.22)$$

die Schätzung des Integrationsparameters ineffizienter wird, falls AR- und MA-Bestandteile berücksichtigt werden, da  $\hat{d}$  und  $\hat{\theta}_1, \dots, \hat{\theta}_p, \hat{\phi}_1, \dots, \hat{\phi}_q$  asymptotisch korreliert sind. Auch wird die Schätzung ineffizienter, wenn die Prozessordnung überspezifiziert wird.

Um dies zu verdeutlichen, simulieren wir nun ARFIMA(1,  $d$ , 0)-Prozesse mit  $\theta \in \{-0,5; 0; 0,5\}$ . Für den Integrationsparameter wählen wir die Situationen  $d = 0,4$  und  $d = 1,4$  und belassen den Stichprobenumfang  $n = 1000$ . Der Störterm sei standardnormalverteilt und wir führen insgesamt 100.000 Wiederholungen durch. Hierzu stellen wir die geschätzten Dichtefunktionen der Schätzergebnisse  $\hat{d}$  und  $\hat{\theta}$  in Abbildung 3.2 dar, in der wir erkennen können, dass die Verteilung der Schätzergebnisse sich deutlicher von einer Normalverteilung unterscheiden. Die Verteilungsparame-

ter sind für  $d = 0,4$  in Tabelle 3.2 und für  $d = 1,4$  in Tabelle 3.3 abgetragen. Im Mittel werden zwar die wahren Parameter gut getroffen, sind aber wesentlich ineffizienter als für den Fall des fraktionalen Rauschens. Für  $\theta = 0$  haben wir zwar im Endeffekt ein solches Rauschen, aber wenn wir eine ARFIMA(1,  $d$ , 0)-Struktur berücksichtigen, führt die Schätzung von  $\theta$  zu Effizienzeinbußen. In der Abbildung 3.3 befinden sich noch Streudiagramme für die geschätzten Parameter  $\hat{d}$  und  $\hat{\theta}$ . Hier kann man erkennen, dass für einen steigenden AR(1)-Koeffizienten  $\theta$  die Korrelation immer stärker ausgeprägt ist, was auch dazu führt, dass die Schätzung von  $d$  auch ineffizienter wird.

### 3.1.5 Der Whittle-Schätzer

Das Schätzverfahren nach Whittle stellt in der Praxis eines der am meisten verwendeten Verfahren zur Bestimmung von  $d$  dar.<sup>23</sup> Es basiert auf der Berechnung des Periodogramms auf Basis der gemittelten schnellen Fourier-Transformation und der Verwendung der Whittle-Approximation der gaussianischen Log-Likelihood-Funktion. Da die schnelle Fourier-Transformation eine relativ hohe Konvergenzgeschwindigkeit von  $O(n \log_2(n))$  besitzt, liefert die Whittle-Schätzung schnelle Parameterschätzungen für  $d$ .

Nehmen wir an, dass ein Vektor  $\mathbf{y} = \{y_1, \dots, y_n\}$  normalverteilt ist mit Mittelwert 0 und eine Kovarianzmatrix  $\mathbf{\Gamma}_\psi$  hat. Dann können wir die logarithmierte Likelihood-Funktion angeben mit

$$L(\psi) = -\frac{1}{2n} \ln \det \mathbf{\Gamma}_\psi - \frac{1}{2} \mathbf{y}' \mathbf{\Gamma}_\psi^{-1} \mathbf{y}. \quad (3.23)$$

Die Elemente der Kovarianzmatrix lassen sich mit Hilfe der inversen Fourier-Transformation der Spektraldichte  $f_\psi(\lambda)$  des Prozesses bestimmen als

$$(\mathbf{\Gamma}_\psi)_{ij} = \gamma_\psi(i - j) \quad \text{mit} \quad \gamma_\psi(k) = \int_{-\pi}^{\pi} f_\psi \exp(i\lambda k) d\lambda. \quad (3.24)$$

Um nun an die für die Whittle-Schätzung relevante logarithmierte Likelihoodfunktion zu gelangen, führen wir zwei Approximationen durch. Da für

$$\frac{1}{n} \ln \det \mathbf{\Gamma}_\psi \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[2\pi f_\psi(\lambda)] d\lambda \quad (3.25)$$

---

<sup>23</sup>Vgl. hierzu Palma (2007) ab S. 78.

gilt, können wir für  $n \rightarrow \infty$  den ersten Teil der logarithmierten Likelihood-Funktion annähern mit

$$\frac{1}{2n} \ln \det \mathbf{\Gamma}_\psi \rightarrow \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln[2\pi f_\psi(\lambda)] d\lambda \quad (3.26)$$

Den zweiten Teil der logarithmierten Likelihood lässt sich wie folgt approximieren

$$\begin{aligned} \frac{1}{2\pi n} \mathbf{y}' \mathbf{\Gamma}_\psi^{-1} \mathbf{y} &\approx \sum_{l=1}^n \sum_{j=1}^n y_l \left\{ \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\psi^{-1}(\lambda) \exp[i\lambda(l-j)] d\lambda \right\} y_j \quad (3.27) \\ &= \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\psi^{-1}(\lambda) \sum_{l=1}^n \sum_{j=1}^n y_l y_j \exp[i\lambda(l-j)] d\lambda \\ &= \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\psi^{-1}(\lambda) \left| \sum_{j=1}^n y_j \exp(i\lambda j) \right|^2 d\lambda \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\psi(\lambda)} d\lambda \end{aligned}$$

$$\text{wobei } i = \sqrt{-1} \text{ und } I(\lambda) = \frac{1}{2\pi n} \left| \sum_{j=1}^n y_j \exp(i\lambda j) \right|^2$$

das Periodogramm des Prozesses  $y_t$  darstellt. Somit erhalten wir als Approximation folgende logarithmierte Likelihood-Funktion:

$$L(\psi) = -\frac{1}{4\pi} \left[ \int_{-\pi}^{\pi} \ln f_\psi(\lambda) d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\psi(\lambda)} d\lambda \right]. \quad (3.28)$$

Um nun die Anwendbarkeit dieser Funktion zu erleichtern, können wir die Integrale durch Riemann-Summenausdrücke ersetzen. Dies erspart die relativ rechenintensive Integration:

$$\int_{-\pi}^{\pi} \ln f_\psi(\lambda) d\lambda \approx \frac{2\pi}{n} \sum_{j=1}^n \ln f_\psi(\lambda_j) \quad \text{und} \quad \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\psi(\lambda)} d\lambda \approx \frac{2\pi}{n} \sum_{j=1}^n \frac{I(\lambda_j)}{f_\psi(\lambda_j)}, \quad (3.29)$$

wobei  $\lambda_j = \frac{2\pi j}{n}$  die bereits erwähnten Fourierfrequenzen darstellen. Folglich können wir eine diskrete Version der logarithmierten Likelihood-Funktion angeben mit

$$L(\boldsymbol{\psi}) = -\frac{1}{2n} \left[ \sum_{j=1}^n \ln f_{\boldsymbol{\psi}}(\lambda_j) + \sum_{j=1}^n \frac{I(\lambda_j)}{f_{\boldsymbol{\psi}}(\lambda_j)} \right]. \quad (3.30)$$

Diese gilt, wie eingehend erwähnt, für den Fall, dass es sich bei  $y_t$  um ein Prozess mit normalverteilten Störtermen handelt. Heben wir diese Annahme auf und stellen die folgende auf, dass  $y_t$  ein stationärer Prozess mit der Wold-Darstellung

$$y_t = \sum_{j=0}^{\infty} \varphi_j(\boldsymbol{\psi}) \varepsilon_{t-j}, \quad (3.31)$$

wobei  $\varepsilon_t$  einen *IID*-Störterm darstellt mit endlichem vierten Moment und  $\sum_{j=0}^{\infty} \varphi_j(\boldsymbol{\psi}) < \infty$ . Unter dieser Annahme konnten Giraitis und Surgailis (1990) die Konsistenz und asymptotische Normalverteilung für den Whittle-Schätzer nachweisen.

Der geschätzte Vektor  $\hat{\boldsymbol{\psi}}$ , der die Log-Likelihood-Funktion maximiert, ist dann konsistent und  $\sqrt{n}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) \rightarrow N[\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\psi}_0)^{-1}]$ , falls  $n \rightarrow \infty$ . Bemerkenswert ist hierbei, dass auch die Annahme der Normalverteilung für die Realisationen des Prozess  $y_t$  fallen gelassen werden kann.

### 3.1.6 Vergleich der parametrischen Verfahren

In den folgenden Simulationsstudien werden wir nun die vorgestellten approximierenden Methoden mit der exakten Schätzweise von ARFIMA-Prozessen vergleichen. Hierzu messen wir in den einzelnen Simulationen die durchschnittliche Verzerrung des geschätzten Integrationsparameters  $\hat{d}$  als

$$\text{Verzerrung} = \frac{\sum_{i=1}^N (\hat{d}_i - d)}{N} \quad (3.32)$$

und die Wurzel des durchschnittlichen quadratischen Fehlers (Root Mean Squared Error, RMSE) mit

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\hat{d}_i - d)^2}{N}}, \quad (3.33)$$

wobei  $N$  die Anzahl der Wiederholungen jedes einzelnen Experiments und  $d$  der wahre Integrationsgrad eines simulierten Prozesses darstellt. Als datengenerierende Prozess legen wir folgende Strukturen fest:

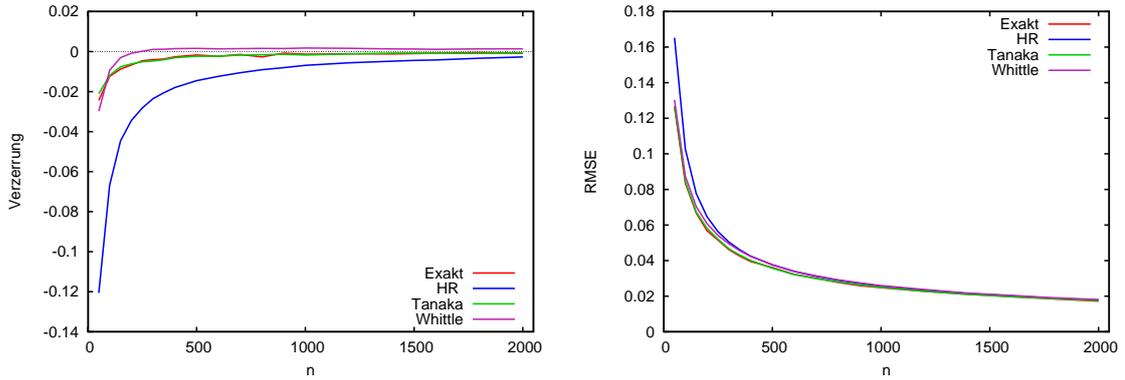


Abb. 3.4: Verzerrung und RMSE von  $\hat{d}$  für ARFIMA(0,0,0)-Prozesse in Abhängigkeit von  $n$

Für ARFIMA(0,  $d$ , 0)-Prozesse:

$$\Delta^d y_t = \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim N(0; 1) \quad \text{für} \quad t = 1, \dots, n. \quad (3.34)$$

Für ARFIMA(1,  $d$ , 0)-Prozesse:

$$\Delta^d(1 - \theta L)y_t = \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim N(0; 1) \quad \text{für} \quad t = 1, \dots, n. \quad (3.35)$$

Für ARFIMA(1,  $d$ , 1)-Prozesse:

$$\Delta^d(1 - \theta L)y_t = (1 - \phi L)\varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim N(0; 1) \quad \text{für} \quad t = 1, \dots, n. \quad (3.36)$$

In den ersten Simulationen variieren wir zunächst den Beobachtungsumfang  $n$  und fixieren die Ordnung ARFIMA(0,  $d$ , 0) mit dem einfachsten Fall  $d = 0$ , d.h. wir simulieren einfache Weißes Rauschen und schätzen die Integrationsordnung. Wir nehmen dabei an, dass die AR-Ordnung  $p = 0$  und MA-Ordnung  $q = 0$  bekannt sein sollen und auch für die folgenden Simulationen keine Unsicherheit bezüglich der Ordnung der Prozessgleichung besteht. Des Weiteren wiederholen wir alle Simulationen in diesem Abschnitt 50.000-mal, für die exakte Schätzmethode führen wir auf Grund der Rechenintensität lediglich 10.000 Wiederholungen durch.

Als Ergebnis lässt sich in der Abbildung 3.4 hierfür festhalten, dass die Schätzmethoden den Integrationsgrad in kleinen Stichproben unterschätzen. Die HR-Approximation weist eine wesentlich stärkere Unterschätzung als die anderen Verfahren auf, die alle auf einem ähnlichen Niveau liegen. Die Whittle-Approximation hat schon bei kleineren Stichproben eine geringere Verzerrung als die exakte Berechnungsweise, überschätzt aber bei größeren Stichproben den Integrationsgrad auf relativ kleinem Niveau. Bei der Effizienz der Schätzverfahren lässt sich aussagen, dass ab einem

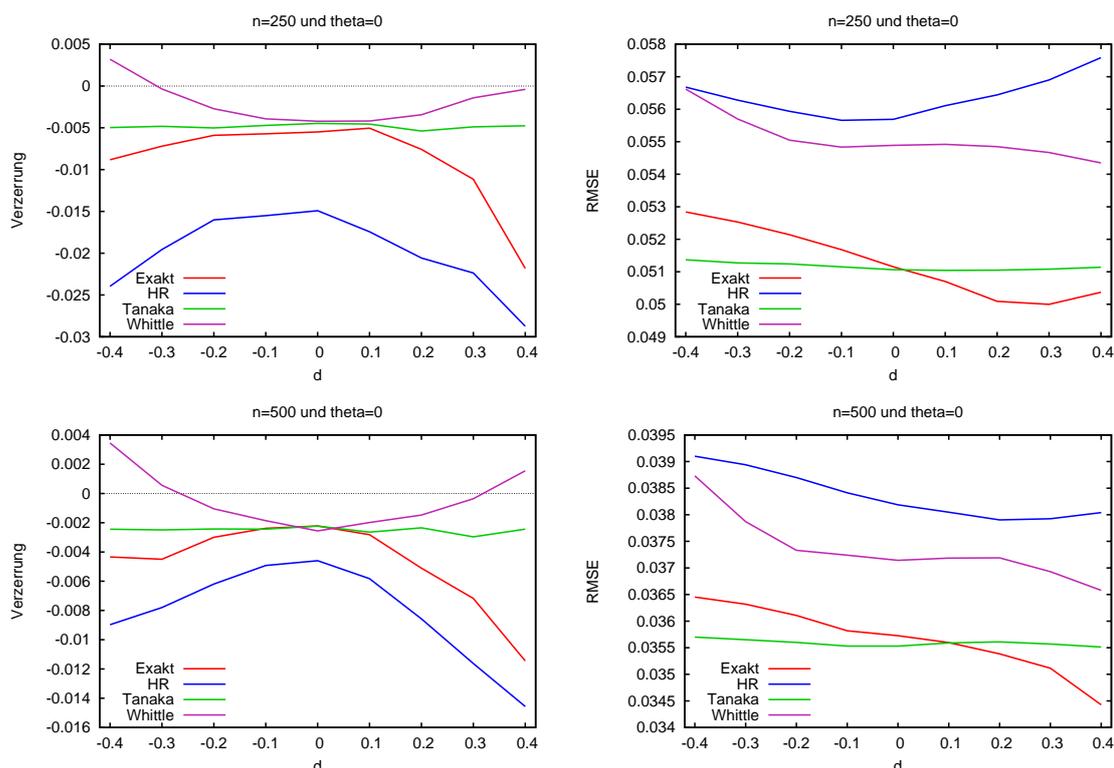


Abb. 3.5: Verzerrung und RMSE von  $\hat{d}$  für ARFIMA(0, $d$ ,0)-Prozesse in Abhängigkeit von  $d$

Stichprobenumfang von  $n = 500$  kaum Unterschiede zwischen den Verfahren auftreten. Bei kleineren Stichproben schneidet auch hier die HR-Approximation etwas schlechter ab.

In der nächsten Simulationsstudie wählen wir wieder einfache fraktionale Rauschprozesse mit ARFIMA(0,  $d$ , 0)-Ordnung und variieren den Parameter  $d$  im Bereich  $[-0,4; 0,4]$  für Stichprobenumfänge  $n = 250$  und  $n = 500$ . Die Grenzfälle  $d = -0,5$  und  $d = 0,5$  werden ausgelassen, da sich diese Werte genau auf der Grenze des Optimierungsbereiches für alle Verfahren außer der Tanaka-Approximation befinden.

In der Abbildung 3.5 kann man zu dem Ergebnis kommen, dass die HR-Approximation die stärkste Unterschätzung von  $d$  aufweist, die aber für  $n = 500$  wesentlich kleiner wird und auch relativ gesehen näher an der Verzerrung der anderen Verfahren liegt. Das HR-Verfahren ist zudem auch meist ineffizienter als die restlichen Verfahren. Die Tanaka-Approximation weist hingegen eine konstante Unterschätzung und RMSE im Bezug auf den wahren Parameter  $d$  auf, während die exakte Berechnungsweise und die Whittle-Approximation Strukturen in  $d$  aufweisen. Diese lassen sich durch die Grenzen im Optimierungsbereich für das Schätzergebnis  $\hat{d}$  erklären, da z.B. im Fall einer einzelnen Überschätzung das Schätzergebnis maximal 0,499 be-

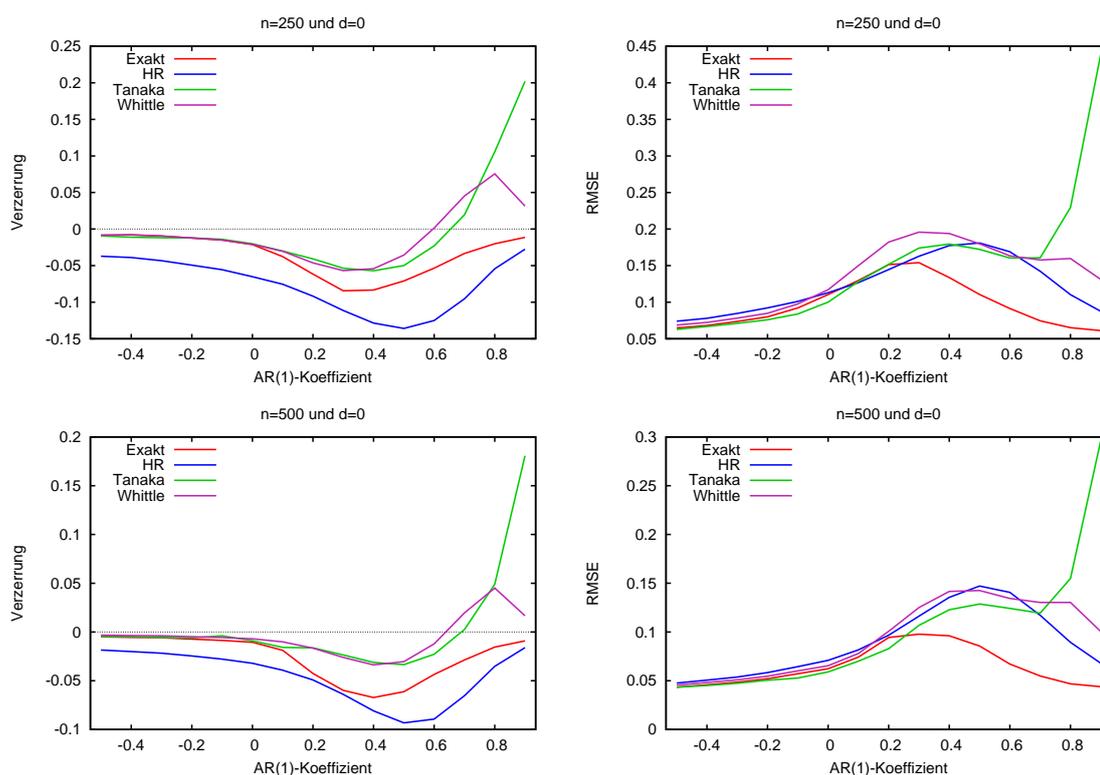


Abb. 3.6: Verzerrung und RMSE von  $\hat{d}$  für ARFIMA(1,0,0)-Prozesse in Abhängigkeit von  $\theta$

trägt und somit Unterschätzungen nur im begrenzten Ausmaß ausgeglichen werden können. Interessanterweise scheint dies aber nicht für die Whittle-Approximation zu gelten, da wir für datengenerierende Prozesse mit  $d = 0,4$  trotzdem eine durchschnittliche Überschätzung des Integrationsgrads messen, was daraufhin hindeutet, dass womöglich mit relativ großer Wahrscheinlichkeit ein Integrationsgrad  $\hat{d} > 0,4$  geschätzt wird.

Die Abbildung 3.6 zeigt die Simulationsergebnisse bei der Verwendung von ARFIMA-(1,  $d$ , 0)-Prozessen mit  $d = 0$  in Abhängigkeit des AR(1)-Parameters  $\theta \in [-0,5; 0,9]$ . Auffällig ist hierbei der merkwürdige Verlauf der Verzerrung und des RMSE für alle Verfahren außer der Tanaka-Approximation, die zunächst eine stärkere Unterschätzung und schwächere Effizienz für steigende  $\theta$ -Parameter aufweisen und dann für größere  $\theta$ -Werte aber wieder abnehmen. Eine Erklärung hierfür könnte wiederum die Beschränkung im Optimierungsalgorithmus für  $d$  sein, der diesen Parameter beschränkt. Für die Tanaka-Approximation haben wir keine derartige Restriktion  $|\hat{d}| \leq 0,5$ , so dass wir für  $\theta = 0,9$  auch relativ starke Abweichungen erhalten können. Ansonsten ist das Verhalten dem anderen Verfahren ähnlich.

Als letzte Simulationsstudie beschäftigen wir uns mit ARFIMA(1,  $d$ , 1)-Prozessen

Tab. 3.4: Schätzeigenschaften von  $\hat{d}$  bei ARFIMA(1,0,1)-Prozessen mit  $n = 250$ 

	$\phi$	Verzerrung				RMSE			
		-0,3	0	0,4	0,6	-0,3	0	0,4	0,6
$\theta = 0$	Exakt	-0,0032	-0,0046	-0,0031	-0,0026	0,0378	0,0515	0,0414	0,0357
	HR	-0,0266	-0,0282	-0,0278	-0,0392	0,0536	0,0563	0,0524	0,0645
	Whittle	-0,0010	-0,0071	-0,0124	-0,0276	0,0516	0,0542	0,0505	0,0621
	Tanaka	-0,0091	-0,0051	-0,0212	-0,0345	0,0301	0,0520	0,0748	0,0697
$\theta = 0,3$	Exakt	-0,0555	-0,0300	-0,0823	-0,0292	0,0592	0,0652	0,0683	0,0739
	HR	-0,0396	-0,0543	-0,0718	-0,0636	0,0914	0,1103	0,1348	0,1591
	Whittle	-0,0520	-0,0569	-0,0436	-0,0599	0,1269	0,1458	0,1595	0,1917
	Tanaka	0,0926	-0,0536	0,0672	0,1497	0,1261	0,1740	0,1880	0,1915
$\theta = 0,5$	Exakt	-0,0813	-0,0709	-0,1001	-0,1012	0,0941	0,1106	0,1184	0,1266
	HR	-0,1018	-0,0913	-0,0853	-0,1173	0,1326	0,1401	0,1554	0,1628
	Whittle	-0,0345	-0,0355	-0,0308	-0,0660	0,1391	0,1662	0,1761	0,1798
	Tanaka	0,1808	-0,0698	0,1701	0,2466	0,1763	0,1922	0,2163	0,3446
$\theta = 0,8$	Exakt	-0,0987	-0,0844	-0,0907	-0,0947	0,1370	0,1523	0,1541	0,1590
	HR	-0,1219	-0,1159	-0,1079	-0,0994	0,1673	0,1694	0,1773	0,1810
	Whittle	0,0551	0,0755	0,0999	0,0885	0,1442	0,1597	0,1651	0,1868
	Tanaka	0,2134	0,1057	0,2701	0,3856	0,2295	0,2877	0,3419	0,5002

 Tab. 3.5: Schätzeigenschaften von  $\hat{d}$  bei ARFIMA(1,0,1)-Prozessen mit  $n = 500$ 

	$\phi$	Verzerrung				RMSE			
		-0,3	0	0,4	0,6	-0,3	0	0,4	0,6
$\theta = 0$	Exakt	-0,0002	-0,0017	-0,0016	-0,0011	0,0276	0,0212	0,0251	0,0361
	HR	-0,0098	-0,0195	-0,0111	-0,0179	0,0286	0,0296	0,0376	0,0392
	Whittle	0,0013	-0,0016	-0,0002	0,0001	0,0398	0,0343	0,0454	0,0482
	Tanaka	-0,0017	-0,0024	-0,0102	-0,0167	0,0277	0,0357	0,0738	0,0738
$\theta = 0,3$	Exakt	-0,0223	-0,0155	-0,0539	-0,0614	0,0484	0,0469	0,0580	0,0586
	HR	-0,0240	-0,0307	-0,0483	-0,0510	0,0718	0,0894	0,1021	0,1133
	Whittle	-0,0228	-0,0261	-0,0144	-0,0156	0,1093	0,1151	0,1240	0,1244
	Tanaka	0,0721	-0,0236	0,0333	0,0785	0,0997	0,1068	0,1252	0,1876
$\theta = 0,5$	Exakt	-0,0341	-0,0399	-0,0421	-0,0521	0,0678	0,0823	0,0837	0,1077
	HR	-0,0365	-0,0540	-0,0582	-0,0650	0,1083	0,1163	0,1169	0,1201
	Whittle	-0,0283	-0,0304	-0,0290	-0,0302	0,1190	0,1303	0,1309	0,1587
	Tanaka	0,1218	-0,0335	0,0464	0,1303	0,1589	0,1688	0,1693	0,1812
$\theta = 0,8$	Exakt	-0,0492	-0,0512	-0,0620	-0,0855	0,0980	0,1056	0,1100	0,1119
	HR	-0,0419	-0,0832	-0,0872	-0,0771	0,1362	0,1472	0,1532	0,1594
	Whittle	0,0303	0,0451	0,0518	0,0484	0,1226	0,1425	0,1714	0,1922
	Tanaka	0,0854	0,0902	0,1699	0,1922	0,1307	0,1549	0,2346	0,2925

in der Form (3.36). Hier simulieren wir auch wieder einen Beobachtungsumfang von  $n = 250$  und  $n = 500$  und nehmen vereinfachend  $d = 0$  an. Für den AR(1)-Parameter  $\theta$  beschränken wir uns auf einen positiven Wertebereich  $\{0; 0,3; 0,5; 0,8\}$  und für den MA(1)-Parameter  $\phi$  den Bereich  $\{-0,3; 0; 0,4; 0,6\}$ . Aus den Ergebnissen, die man für  $n = 250$  der Tabelle 3.4 und für  $n = 500$  der Tabelle 3.5 entnehmen kann, lässt sich schließen, dass die Schätzung des Integrationsparameters effizienter bzw. weniger volatil wird, wenn ein negativer MA-Bestandteil vorliegt. Die exakte Schätzweise hat generell die besten Schätzeigenschaften und ist anscheinend bei den vorliegenden Parameterkonstellationen am effizientesten. Die HR- und Whittle-Approximation liegen in Beziehung auf den RMSE auf einem ähnlichen Niveau, wobei der Whittle-Schätzer häufiger schlechter abschneidet als die Approximation nach HR. Der Unterschied im RMSE zur exakten Schätzweise ist in manchen Fällen für beiden Approximation recht groß, aber meistens nicht sehr deutlich. Im Bezug auf die durchschnittliche Verzerrung ergeben sich aber für sämtlich Verfahren keine großen Unterschiede außer der Tanaka-Approximation, wobei wir stets mit einer tendenziellen Unterschätzung des Integrationsgrad rechnen müssen. Die Tanaka-Approximation überschätzt den Integrationsgrad bei Existenz von AR(1)-Bestandteilen und ist für  $\theta = 0,5$  und  $\theta = 0,8$  wesentlich ineffizienter als die restlichen Verfahren. Selbst bei größeren Stichprobenumfängen weist dieses noch relativ starke Verzerrungen bei einer hohen Volatilität auf.

### 3.1.7 Parameterwahl und Informationskriterien

Die vorangegangenen Simulationen zeigten ausschließlich Ergebnisse unter der Prämisse, dass die wahre Parameterordnung der datengenerierenden Prozesse bekannt sei. Dies ist natürlich, gerade im Hinblick auf eine empirische Anwendung, keine realistische Annahme. Deswegen wollen wir uns in diesem Abschnitt mit Methoden beschäftigen, wie wir die Ordnungen  $p$  und  $q$  von ARFIMA( $p, d, q$ )-Prozessen festlegen können.

Im ersten Schritt untersuchen wir die Effekte einer Fehlspezifizierung auf die Schätzeigenschaften bzgl.  $d$  der parametrischen Schätzverfahren. Hierzu simulieren wir fraktionale Rauschen, also ARFIMA( $0, d, 0$ )-Prozesse, mit  $n = 500$  und schätzen jeweils den Integrationsgrad mit den Ordnungen  $p, q \in \{0; 1; 2; 3; 4\}$  wiederum mit 5000 Wiederholungen. Die Ergebnisse hierfür tragen wir in der Tabelle 3.6 mit der Fixierung  $d = 0$  unter der Verwendung der Whittle-Approximation, da diese selbst bei hohen Ordnungen von  $p$  und  $q$  noch recht schnell berechnet werden kann. Es zeigt sich, dass der Integrationsgrad mit steigender Überspezifizierung stärker unter-

Tab. 3.6: Schätzeigenschaften in Abhängigkeit der Wahl  $p$  und  $q$  für ARFIMA(0,0,0)-Prozesse

$p/q$	Verzerrung					RMSE				
	0	1	2	3	4	0	1	2	3	4
0	-0,0011	-0,0260	-0,0397	-0,0516	-0,0625	0,0653	0,1061	0,1382	0,1639	0,1858
1	-0,0324	-0,0700	-0,0787	-0,0917	-0,0826	0,1204	0,1984	0,2104	0,2137	0,2225
2	-0,0529	-0,0809	-0,0884	-0,0968	-0,0979	0,1710	0,2266	0,2329	0,2336	0,2458
3	-0,0658	-0,0738	-0,0826	-0,0951	-0,1035	0,2135	0,2340	0,2392	0,2439	0,2626
4	-0,0808	-0,0895	-0,0930	-0,1118	-0,1235	0,2436	0,2476	0,2501	0,2508	0,2762

Tab. 3.7: Schätzeigenschaften in Abhängigkeit der Wahl  $p$  und  $q$  für ARFIMA(1,0,0)-Prozesse mit  $\theta = 0,5$ 

$p/q$	Verzerrung					RMSE				
	0	1	2	3	4	0	1	2	3	4
0	0,3830	0,3483	0,2420	0,1423	0,0700	0,4447	0,4055	0,2911	0,2251	0,2144
1	-0,0292	-0,0646	-0,0833	-0,1005	-0,1100	0,1419	0,2338	0,2480	0,2518	0,2623
2	-0,1018	-0,1244	-0,1404	-0,1624	-0,1757	0,2015	0,2671	0,2832	0,2877	0,2996
3	-0,1127	-0,1426	-0,1849	-0,1867	-0,1953	0,2516	0,2758	0,2924	0,2970	0,3093
4	-0,1301	-0,1477	-0,1588	-0,1781	-0,1939	0,2871	0,2918	0,3094	0,3143	0,3273

geschätzt wird und auch wesentlich volatiler wird. Dies ist nicht überraschend, da sich der fraktionale Bestandteil als ein  $AR(\infty)$ - bzw.  $MA(\infty)$ -Prozess darstellen lässt und eine zusätzliche Hinzunahme von  $AR$ - oder  $MA$ -Prozessgliedern sich vermindert auf die Schätzung des Integrationsgrads auswirkt, da hierdurch der persistente Effekt des Long Memory zum Teil aufgefangen wird.

Die Tabelle 3.7 zeigt nun die Ergebnisse bei Verwendung von ARFIMA(1,  $d$ , 0)-Prozessen mit  $\theta = 0,5$  und  $d = 0$ . Hierbei ergeben sich zum Teil noch erheblichere Abweichungen als für den Fall des einfachen Weißen Rauschens. Besonders, falls  $p = 0$  fixiert und somit keine autokorrelierte Struktur berücksichtigt wird, kann man eine starke Überschätzung des Integrationsgrads feststellen. Wächst aber nun die  $MA$ -Ordnung  $q$  sukzessive an, dann verringert sich diese Überschätzung, da ein  $AR(1)$ -Prozess als  $MA(\infty)$  dargestellt werden kann. Ansonsten ergibt sich bei einer Überschätzung der Ordnung eine wachsende Unterschätzung von  $d$  bei sehr volatilen Schätzergebnissen. Aus diesen Gründen versuchen wir nun die Möglichkeiten zu ergründen, die Ordnung eines Prozesses möglichst datengetrieben zu gestalten. Hierfür wollen wir das Akaike Informationskriterium (Akaike Information Criterion, AIC) betrachten, das sich als

$$AIC = -2L(\hat{\psi}) + 2K \quad (3.37)$$

ergibt, mit  $L(\hat{\psi})$  als der maximalen logarithmierten Likelihood eines zu schätzenden ARFIMA( $p, d, q$ )-Prozesses und mit  $K = p+q+1$ , was die Anzahl der zu schätzenden

Parameter darstellt. Da wir hier mit approximativen Schätzmethoden arbeiten und die exakte logarithmierte Likelihood  $L(\hat{\psi})$  nicht bestimmen werden, verwenden wir die Berechnungsweise

$$AIC = \ln(\hat{\sigma}_\varepsilon^2) + 2\frac{K}{n}, \quad (3.38)$$

wobei  $\hat{\sigma}_\varepsilon^2$  die geschätzte minimale Varianz der Residuen ist. Unter den als möglich erachteten Modellspezifikationen wird dann diejenige gewählt, welches die niedrigste Informationskriterium aufweist. Das AIC wird aber zumeist kritisiert, dass es eine Präferenz für überspezifizierte Modellordnungen hat. Deswegen verwenden wir neben dem AIC das Schwarz Informationskriterium (Bayesian Information Criterion, BIC):

$$BIC = \ln(\hat{\sigma}_\varepsilon^2) + \frac{K}{n} \ln n \quad (3.39)$$

und das Hannan-Quinn Informationskriterium (HQC):

$$HQC = n \ln(\hat{\sigma}_\varepsilon^2) + 2K \ln \ln n. \quad (3.40)$$

In den nachfolgenden Simulationen werden wir nun für die Modelle (3.34), (3.35) und (3.36) untersuchen, mit welcher relativen Häufigkeit die einzelnen Informationskriterien welche Modelle finden können. Wir nehmen hierbei an, dass die Parameter  $p$  und  $q$  im Bereich  $\{0; 1; 2; 3\}$  liegen und führen für jede Modellspezifikation 5000 Wiederholungen durch. In den Tabellen 3.8–3.11 fixieren wir zunächst den Integrationsgrad  $d = 0$  für Stichprobenumfänge von  $n = 250$  und  $n = 500$ . Zusätzlich geben wir dann auch die durchschnittliche Verzerrung und RMSE für das jeweils präferierte Modell bzgl. des geschätzten Integrationsparameters, um die Schätzeigenschaften, die sich durch die Verwendung der Modellselektionskriterien ergeben, zu bewerten. Als Schätzmethode wählen wir die beiden approximativen Schätzer nach Whittle und HR. Für die exakte Berechnungsweise haben Lardic und Mignon (2004) sehr ähnliche Simulationsstudien durchgeführt und kamen dabei zum Ergebnis, dass das AIC das wahre Modell schlechter findet als das BIC, aber für tendenziell positive Werte des Integrationsparameters  $d$  bessere Schätzeigenschaften aufweist, wenn das nach AIC bevorzugte Modell verwendet wird. In ihrer Studie werden aber keine ARFIMA(1,  $d$ , 1)-Prozesse untersucht, sondern nur fraktionale Rauschen und ARFIMA(1,  $d$ , 0)-Prozesse. Auf eine Anwendung für die Tanaka-Approximation verzichten wir in den folgenden Simulationsstudien, da Modelle mit höheren Ordnungen

Tab. 3.8: Gewählte Parameterordnungen  $p$  und  $q$  für  $n = 250$  unter Verwendung der Whittle-Approximation

Wahre Ordnung ist ARFIMA(0,0,0)														
Parameterwahl mit AIC				Parameterwahl mit BIC				Parameterwahl mit HQC						
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	<b>23,7%</b>	2,2%	1,2%	1,0%	0	<b>93,8%</b>	0,5%	0,2%	0,0%	0	<b>72,3%</b>	2,3%	0,9%	0,2%
1	2,6%	2,0%	1,6%	0,6%	1	3,2%	0,5%	0,0%	0,0%	1	5,1%	2,6%	0,4%	0,0%
2	1,4%	2,4%	24,3%	8,7%	2	0,9%	0,2%	0,4%	0,4%	2	3,5%	0,5%	5,8%	2,5%
3	1,2%	1,4%	9,5%	15,8%	3	0,0%	0,0%	0,0%	0,0%	3	0,9%	0,4%	1,4%	1,2%

Wahre Ordnung ist ARFIMA(1,0,0) mit $\theta = 0,5$														
Parameterwahl mit AIC				Parameterwahl mit BIC				Parameterwahl mit HQC						
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	0,5%	5,3%	4,9%	3,5%	0	6,3%	12,5%	1,0%	0,4%	0	1,9%	10,5%	3,6%	1,6%
1	<b>38,2%</b>	4,9%	2,8%	1,0%	1	<b>75,4%</b>	0,7%	0,0%	0,0%	1	<b>63,9%</b>	3,0%	0,5%	0,2%
2	10,1%	2,1%	3,9%	5,4%	2	3,4%	0,0%	0,4%	0,0%	2	7,7%	1,0%	1,6%	1,3%
3	2,8%	0,5%	7,0%	7,1%	3	0,0%	0,0%	0,0%	0,0%	3	0,7%	0,5%	1,3%	0,6%

Wahre Ordnung ist ARFIMA(1,0,1) mit $\theta = 0,5$ und $\phi = -0,3$														
Parameterwahl mit AIC				Parameterwahl mit BIC				Parameterwahl mit HQC						
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	0,0%	2,4%	7,0%	4,9%	0	0,0%	20,5%	9,0%	0,9%	0	0,0%	8,7%	8,5%	2,1%
1	0,1%	<b>43,0%</b>	5,3%	1,0%	1	0,3%	<b>52,4%</b>	0,2%	0,0%	1	0,2%	<b>53,2%</b>	1,7%	0,1%
2	10,0%	4,9%	1,3%	4,1%	2	14,6%	0,5%	0,0%	0,1%	2	13,2%	3,2%	0,7%	1,3%
3	7,9%	2,1%	1,4%	4,5%	3	1,4%	0,1%	0,0%	0,0%	3	5,4%	0,7%	0,2%	0,7%

Tab. 3.9: Gewählte Parameterordnungen  $p$  und  $q$  für  $n = 500$  unter Verwendung der Whittle-Approximation

Wahre Ordnung ist ARFIMA(0,0,0)														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3	
0	<b>26,8%</b>	1,8%	1,2%	0,9%	0	<b>96,6%</b>	1,2%	0,2%	0,0%	0	<b>78,1%</b>	3,4%	0,6%	0,4%
1	3,7%	3,2%	1,2%	0,5%	1	1,4%	0,0%	0,0%	0,0%	1	3,4%	1,4%	0,2%	0,2%
2	3,2%	1,8%	20,3%	9,5%	2	0,6%	0,0%	0,0%	0,0%	2	1,2%	0,6%	5,3%	1,4%
3	2,5%	1,1%	9,0%	13,4%	3	0,0%	0,0%	0,0%	0,0%	3	1,0%	0,2%	1,0%	1,4%
Wahre Ordnung ist ARFIMA(1,0,0) mit $\theta = 0,5$														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3	
0	0,0%	1,4%	4,7%	2,9%	0	0,5%	4,4%	1,0%	0,0%	0	0,0%	2,9%	3,2%	0,8%
1	<b>51,0%</b>	5,5%	4,4%	0,7%	1	<b>90,6%</b>	1,9%	0,2%	0,0%	1	<b>80,2%</b>	3,4%	0,9%	0,2%
2	5,3%	4,1%	1,6%	3,3%	2	1,3%	0,0%	0,0%	0,0%	2	3,7%	0,7%	0,2%	0,5%
3	2,7%	0,3%	6,6%	5,5%	3	0,0%	0,0%	0,0%	0,1%	3	0,8%	0,2%	1,5%	0,7%
Wahre Ordnung ist ARFIMA(1,0,1) mit $\theta = 0,5$ und $\phi = -0,3$														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3	
0	0,0%	0,0%	4,6%	6,6%	0	0,0%	6,0%	6,3%	0,6%	0	0,0%	1,4%	6,1%	3,1%
1	0,0%	<b>54,8%</b>	3,1%	1,3%	1	0,0%	<b>75,6%</b>	0,3%	0,0%	1	0,0%	<b>70,9%</b>	1,5%	0,2%
2	5,0%	4,9%	0,7%	2,9%	2	7,0%	1,3%	0,0%	0,0%	2	6,4%	2,7%	0,0%	0,4%
3	9,3%	2,2%	1,0%	3,6%	3	2,8%	0,1%	0,0%	0,0%	3	6,3%	0,4%	0,2%	0,4%

Tab. 3.10: Gewählte Parameterordnungen  $p$  und  $q$  für  $n = 250$  unter Verwendung der HR-Approximation

Wahre Ordnung ist ARFIMA(0,0,0)														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	<b>14,1%</b>	0,8%	1,2%	1,3%	0	<b>33,6%</b>	0,6%	0,2%	0,1%	0	<b>26,5%</b>	1,1%	0,8%	0,4%
1	1,3%	2,0%	1,3%	1,4%	1	0,8%	1,2%	0,9%	0,8%	1	1,4%	1,6%	1,0%	0,9%
2	1,2%	1,4%	17,5%	9,1%	2	0,2%	1,0%	15,9%	9,1%	2	0,8%	1,0%	15,5%	8,8%
3	1,5%	1,5%	9,5%	34,9%	3	0,1%	0,9%	8,6%	26,1%	3	0,6%	1,0%	8,6%	29,9%
Wahre Ordnung ist ARFIMA(1,0,0) mit $\theta = 0,5$														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	0,0%	1,2%	2,6%	2,5%	0	0,3%	4,1%	1,7%	0,2%	0	0,1%	3,1%	2,9%	1,4%
1	<b>20,7%</b>	2,3%	1,6%	2,4%	1	<b>49,5%</b>	0,4%	0,1%	0,8%	1	<b>38,7%</b>	1,3%	0,5%	1,1%
2	4,2%	5,3%	2,9%	7,4%	2	1,8%	2,5%	1,7%	5,5%	2	3,9%	3,5%	1,9%	5,7%
3	3,0%	3,5%	21,3%	19,0%	3	0,2%	2,0%	16,4%	12,8%	3	1,7%	2,2%	16,8%	15,3%
Wahre Ordnung ist ARFIMA(1,0,1) mit $\theta = 0,5$ und $\phi = -0,3$														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	0,0%	0,3%	3,2%	2,9%	0	0,0%	5,0%	6,6%	0,9%	0	0,0%	1,4%	5,4%	2,3%
1	0,0%	<b>22,5%</b>	3,3%	2,4%	1	0,3%	<b>39,2%</b>	0,5%	0,5%	1	0,1%	<b>34,6%</b>	2,0%	1,0%
2	6,2%	2,5%	5,6%	5,0%	2	12,7%	0,7%	2,8%	2,9%	2	10,3%	1,7%	3,4%	3,4%
3	7,3%	3,4%	5,5%	30,1%	3	2,2%	1,3%	3,3%	21,3%	3	6,0%	1,8%	3,8%	23,0%

Tab. 3.11: Gewählte Parameterordnungen  $p$  und  $q$  für  $n = 500$  unter Verwendung der HR-Approximation

Wahre Ordnung ist ARFIMA(0,0,0)														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	<b>23,9%</b>	1,2%	1,5%	1,2%	0	<b>56,1%</b>	0,4%	0,1%	0,0%	0	<b>45,0%</b>	1,2%	0,9%	0,4%
1	1,7%	3,3%	1,5%	1,6%	1	0,9%	1,0%	0,7%	0,7%	1	1,7%	1,9%	0,8%	0,9%
2	1,3%	1,5%	14,8%	7,1%	2	0,1%	0,7%	8,6%	5,5%	2	0,9%	0,9%	9,6%	5,8%
3	1,3%	1,6%	7,5%	28,9%	3	0,0%	0,7%	5,5%	18,8%	3	0,5%	1,0%	5,7%	22,7%

Wahre Ordnung ist ARFIMA(1,0,0) mit $\theta = 0,5$														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	0,0%	0,7%	2,7%	2,4%	0	0,0%	2,6%	1,3%	0,2%	0	0,0%	1,9%	2,6%	1,2%
1	<b>29,9%</b>	3,2%	2,2%	2,1%	1	<b>69,8%</b>	0,5%	0,0%	0,3%	1	<b>56,7%</b>	1,8%	0,6%	0,6%
2	4,7%	6,2%	2,7%	3,9%	2	1,5%	1,6%	1,3%	1,7%	2	3,9%	3,0%	1,3%	2,1%
3	3,3%	3,3%	19,8%	13,0%	3	0,2%	1,4%	10,5%	7,3%	3	1,7%	1,6%	12,0%	8,9%

Wahre Ordnung ist ARFIMA(1,0,1) mit $\theta = 0,5$ und $\phi = -0,3$														
Parameterwahl mit AIC			Parameterwahl mit BIC			Parameterwahl mit HQC								
$p/q$	0	1	2	3	$p/q$	0	1	2	3	$p/q$	0	1	2	3
0	0,0%	0,0%	2,0%	2,9%	0	0,0%	1,0%	5,6%	0,7%	0	0,0%	0,2%	4,1%	2,2%
1	0,0%	<b>31,7%</b>	4,3%	2,1%	1	0,0%	<b>63,6%</b>	0,7%	0,1%	1	0,0%	<b>53,8%</b>	2,6%	0,5%
2	3,2%	3,6%	6,0%	3,7%	2	8,5%	0,5%	1,8%	1,5%	2	6,7%	2,1%	2,9%	1,9%
3	8,9%	2,8%	3,5%	25,2%	3	2,5%	0,3%	1,7%	11,4%	3	6,6%	1,0%	2,0%	13,4%

von  $p$  und  $q$  hier sehr häufig zu Optimierungsfehlern führen und somit eine Anwendung von Informationskriterien nahezu unmöglich macht.

Für die Whittle-Approximation wurde das Verhalten der Informationskriterien bereits von Schmitt und Tschernig (1995) überprüft, geben aber keine Schätzeigenschaften bezüglich  $d$  an und beschränken sich zudem auf eine kleinere Anzahl von Modellen. Sie berücksichtigen aber im Gegensatz zur hier vorliegenden Arbeit auch das modifizierte AIC und BIC und kamen zu dem Ergebnis, dass das BIC wesentlich besser abschneidet als das AIC und das HQC bei der Treffsicherheit zwischen diesen beiden Kriterien liegt. Diese Ergebnisse können wir prinzipiell bestätigen, wie wir den Tabellen 3.8 und 3.9 entnehmen können. Das BIC und das HQC teilen aber das Problem, dass bei ARFIMA(1,  $d$ , 0)- und ARFIMA(1,  $d$ , 1)-Prozessen mit einer relativ hohen Häufigkeit ein unterspezifiziertes Modell bei  $n = 250$  gewählt wird. Diese sinkt zwar tendenziell für einen größeren Stichprobenumfang von  $n = 500$ , kann aber zu überschätzten Integrationsparametern führen, was zu einer erhöhten Annahme der Long-Memory-Eigenschaft führen kann. Etwas überraschend ist, dass wir dieses Problem bei Anwendung der HR-Approximation nicht sehen. Das BIC dominiert hier zwar auch das HQC und das AIC, es wird aber weitaus weniger wahrscheinlich ein unterspezifiziertes Modell selektiert. Das AIC zeigt hingegen bei der Whittle-Approximation das Problem, dass sehr stark überspezifizierte Modelle gewählt werden. Dies beobachten wir aber bei der HR-Approximation auch für das BIC und das HQC, was wiederum volatile Schätzergebnisse für den Integrationsparameter erwarten lassen könnte. Die Erkenntnisse zur HR-Approximation decken sich mit den Ergebnissen der Studie von Crato und Ray (1996), die auch bereits das Verhalten der Informationskriterien bei Anwendung der HR-Approximation untersucht haben. Sie legen den Fokus ihrer Arbeit auf den Effekt von Vorhersagen von ARFIMA-Modellen, die mit Hilfe des AIC und des BIC spezifiziert werden. Diesbezüglich kommen sie aber zur Ansicht, dass für Vorhersagen keines der Informationskriterien dominiert und kritisieren, dass die Wahl der ARFIMA-Ordnung mit Hilfe der Kriterien zu sehr unzuverlässigen Ergebnissen führt. In den Tabellen 3.10 und 3.11 kommen wir aber zum Teil zu anderen Trefferwahrscheinlichkeiten, können aber trotzdem schließen, dass das wahre Modell bei der HR-Approximation mit einer wesentlich geringeren Häufigkeit getroffen wird als bei der Whittle-Approximation.

Da wir in den Tabellen 3.8 bis 3.11 lediglich Ergebnisse für  $d = 0$  angeben, zeigen wir in 3.12 die relativen Selektionshäufigkeiten des wahren Modells für die einzelnen Modelle mit  $n = 250$  und in 3.13 für  $n = 500$ . Wir belassen dabei die gleichen

Tab. 3.12: Relative Selektionshäufigkeiten des wahren Modells bei  $n = 250$ 

Modell	$d$	Whittle-Approximation			HR-Approximation		
		AIC	BIC	HQC	AIC	BIC	HQC
ARFIMA(0, $d$ , 0)	-0,3	31,9%	87,7%	80,2%	19,0%	31,4%	29,4%
	-0,1	25,8%	91,6%	75,3%	15,4%	32,8%	27,6%
	0	23,7%	93,8%	72,3%	14,1%	33,6%	26,5%
	0,2	14,4%	94,1%	67,4%	15,2%	37,1%	29,1%
	0,4	18,0%	94,5%	66,4%	18,7%	46,1%	35,9%
ARFIMA(1, $d$ , 0)	-0,3	41,1%	68,8%	64,9%	22,3%	45,1%	39,2%
	-0,1	40,2%	69,7%	63,0%	21,7%	45,7%	38,1%
	0	38,2%	75,4%	63,9%	20,7%	49,5%	38,7%
	0,2	35,4%	70,7%	57,3%	19,9%	50,7%	39,8%
	0,4	33,5%	79,2%	66,8%	21,2%	53,2%	41,1%
ARFIMA(1, $d$ , 1)	-0,3	46,7%	49,2%	50,0%	24,4%	36,8%	32,6%
	-0,1	49,0%	51,6%	52,6%	25,6%	38,6%	34,3%
	0	43,0%	52,4%	53,2%	22,5%	39,2%	34,6%
	0,2	36,5%	55,7%	51,1%	24,0%	43,4%	37,3%
	0,4	33,1%	67,0%	52,6%	23,4%	48,5%	39,4%

Tab. 3.13: Relative Selektionshäufigkeiten des wahren Modells bei  $n = 500$  in Abhängigkeit von  $d$ 

Modell	$d$	Whittle-Approximation			HR-Approximation		
		AIC	BIC	HQC	AIC	BIC	HQC
ARFIMA(0, $d$ , 0)	-0,3	43,2%	92,8%	79,9%	38,6%	53,9%	46,1%
	-0,1	30,3%	94,8%	77,2%	27,0%	55,1%	44,5%
	0	26,8%	96,6%	78,1%	23,9%	56,1%	45,0%
	0,2	17,3%	97,1%	75,4%	25,4%	59,5%	48,1%
	0,4	21,6%	97,2%	74,5%	31,4%	69,1%	56,8%
ARFIMA(1, $d$ , 0)	-0,3	54,9%	89,7%	81,2%	32,2%	52,6%	47,6%
	-0,1	45,1%	85,8%	75,2%	26,5%	50,3%	44,1%
	0	51,0%	90,6%	80,2%	29,9%	69,8%	56,7%
	0,2	44,1%	85,8%	71,3%	29,9%	73,0%	59,1%
	0,4	36,3%	88,7%	79,2%	29,9%	74,5%	61,3%
ARFIMA(1, $d$ , 1)	-0,3	63,1%	63,7%	66,2%	34,4%	59,7%	50,7%
	-0,1	61,0%	80,6%	72,8%	36,1%	62,6%	53,3%
	0	54,8%	75,6%	70,9%	31,7%	63,6%	53,8%
	0,2	56,9%	74,6%	68,1%	31,4%	64,3%	53,7%
	0,4	44,5%	79,6%	69,0%	33,1%	69,6%	56,8%

Tab. 3.14: Durchschnittliche Verzerrung von  $\hat{d}$  bei Verwendung von Modellselektionskriterien  $n = 250$  in Abhängigkeit von  $d$

Modell	$d$	Whittle-Approximation			HR-Approximation		
		AIC	BIC	HQC	AIC	BIC	HQC
ARFIMA(0, $d$ , 0)	-0,3	-0,0337	-0,0098	-0,0135	-0,0471	-0,0584	-0,0626
	-0,1	-0,0417	-0,0094	-0,0144	-0,0583	-0,0595	-0,0637
	0	-0,0452	-0,0091	-0,0149	-0,0633	-0,0581	-0,0622
	0,2	-0,0744	-0,0048	-0,0160	-0,0122	-0,0207	-0,0192
	0,4	-0,0595	-0,0053	-0,0240	-0,0323	-0,0036	-0,0124
ARFIMA(1, $d$ , 0)	-0,3	-0,0701	-0,0460	-0,0688	-0,0612	-0,0538	-0,0561
	-0,1	-0,0680	-0,0453	-0,0709	-0,0627	-0,0530	-0,0578
	0	-0,0754	-0,0419	-0,0699	-0,0658	-0,0491	-0,0570
	0,2	-0,0669	-0,0447	-0,0780	-0,0683	-0,0513	-0,0593
	0,4	-0,0550	-0,0399	-0,0669	-0,0643	-0,0524	-0,0615
ARFIMA(1, $d$ , 1)	-0,3	-0,0713	-0,0558	-0,0768	-0,0691	-0,0576	-0,0630
	-0,1	-0,0718	-0,0532	-0,0730	-0,0659	-0,0515	-0,0599
	0	-0,0774	-0,0524	-0,0723	-0,0750	-0,0533	-0,0593
	0,2	-0,0815	-0,0496	-0,0643	-0,0946	-0,0482	-0,0681
	0,4	-0,0862	-0,0513	-0,0714	-0,0928	-0,0545	-0,0674

Parameter-Konstellationen wie in den Tabellen zuvor, d.h. für ARFIMA(1,  $d$ , 0)-Prozesse setzen wir den AR(1)-Parameter  $\theta = 0,5$  und für ARFIMA(1,  $d$ , 1)-Prozesse  $\theta = 0,5$  und  $\phi = -0,3$ . Eine Variation im Integrationsparameter führt dabei aber zu keiner Rangverschiebung bezüglich der einzelnen Kriterien und der Approximationsmethoden. Das BIC ist immer besser als das HQC und das AIC und die Trefwahrscheinlichkeit ist für die Whittle-Approximation immer höher als für HR.

Interessant sind auch die Schätzeigenschaften für den Integrationsparameter bei der Wahl des jeweils präferierten Modells. In den Tabellen 3.14 und 3.15 finden wir die durchschnittliche Verzerrung für  $n = 250$  und  $n = 500$ . Hierbei können wir erkennen, dass für die Whittle-Approximation das BIC stets den Parameter  $d$  am geringsten unterschätzt. Bei Vorliegen von ARFIMA(1,  $d$ , 0)- und ARFIMA(1,  $d$ , 1)-Strukturen erhalten wir zum Teil deutlichere Abweichungen als für fraktionales Rauschen, wie wir auch schon in den Tabellen 3.4 und 3.5 feststellen konnten. Für negative Werte von  $d$  haben wir auch eine geringere Unterschätzung als beispielsweise für  $d = 0$ . Bei einem Stichprobenumfang von  $n = 250$  haben wir für die Whittle-Approximation nahezu immer eine geringere Unterschätzung als bei der HR-Approximation, die wiederum bezüglich der Informationskriterien fast immer die gleiche Struktur aufweist. Auch hier liefert die Schätzung mit dem BIC eine geringere Unterschätzung als die anderen Kriterien mit Ausnahme des fraktionalen Rauschens, wo das AIC zum Teil eine geringere Unterschätzung aufweist. Für  $n = 500$  verschiebt sich nun das Bild, da wir für die HR-Approximation bessere

Tab. 3.15: Durchschnittliche Verzerrung von  $\hat{d}$  bei Verwendung von Modellselektionskriterien  $n = 500$  in Abhängigkeit von  $d$

Modell	$d$	Whittle-Approximation			HR-Approximation		
		AIC	BIC	HQC	AIC	BIC	HQC
ARFIMA(0, $d$ , 0)	-0,3	-0,0125	-0,0014	-0,0027	-0,0043	-0,0050	-0,0037
	-0,1	-0,0178	-0,0020	-0,0041	-0,0048	-0,0049	-0,0031
	0	-0,0201	-0,0033	-0,0068	-0,0055	-0,0048	-0,0037
	0,2	-0,0312	-0,0039	-0,0071	-0,0399	-0,0051	-0,0141
	0,4	-0,0250	-0,0038	-0,0072	-0,0671	-0,0253	-0,0450
ARFIMA(1, $d$ , 0)	-0,3	-0,0257	-0,0387	-0,0411	-0,0344	-0,0290	-0,0370
	-0,1	-0,0266	-0,0405	-0,0443	-0,0418	-0,0303	-0,0400
	0	-0,0296	-0,0383	-0,0416	-0,0370	-0,0218	-0,0311
	0,2	-0,0285	-0,0363	-0,0370	-0,0369	-0,0209	-0,0299
	0,4	-0,0364	-0,0375	-0,0411	-0,0354	-0,0205	-0,0288
ARFIMA(1, $d$ , 1)	-0,3	-0,0490	-0,0426	-0,0174	-0,0457	-0,0305	-0,0411
	-0,1	-0,0596	-0,0336	-0,0158	-0,0436	-0,0291	-0,0391
	0	-0,0528	-0,0359	-0,0163	-0,0496	-0,0287	-0,0387
	0,2	-0,0457	-0,0363	-0,0169	-0,0501	-0,0283	-0,0388
	0,4	-0,0375	-0,0341	-0,0167	-0,0475	-0,0262	-0,0367

Ergebnisse bezüglich der Verzerrung beobachten.

Wenn wir nun die Effizienz der Schätzverfahren in den Tabellen 3.16 und 3.17 betrachten, können wir feststellen, dass die HR-Approximation einen geringeren RMSE aufweist als bei der Whittle-Approximation. Gerade bei ARFIMA(1,  $d$ , 0)- und ARFIMA(1,  $d$ , 1)-Prozessen erweist sich die Whittle-Approximation als sehr volatil, auch wenn mit dem BIC z.B. die Modellordnung relativ sicher richtig bestimmt werden kann. Somit kann man zu dem Schluss kommen, dass gerade bei größeren Stichproben die HR-Approximation der Whittle-Approximation vorzuziehen ist.

## 3.2 Semiparametrische Schätzverfahren

Grundlage semiparametrischer Schätzverfahren, die nicht einer genau Spezifizierung der AR- oder MA-Ordnungen bedürfen, ist die Approximation der Spektraldichte eines ARFIMA( $p, d, q$ )-Prozesses im Frequenzbereich nahe des Ursprungs  $\lambda = 0$ , die im vorangegangenen Kapitel heuristisch motiviert wurde. Hierbei konnte gezeigt werden, dass die ARMA-Komponente in der Spektraldichte von ARFIMA-Prozessen gegen eine Konstante konvergiert und die Spektraldichte hauptsächlich vom Integrationsgrad  $d$  abhängt. Der große Vorteil dieser Verfahren liegt darin, dass keine Verzerrungen zu erwarten sind, falls die AR- oder MA-Ordnung fehlspezifiziert wurde. Bei diesen Verfahren besteht dann das Hauptproblem darin, die Anzahl der Frequenzen  $m$  festzulegen, in der das Spektrum in Abhängigkeit von  $d$  angepasst wird. Eine wichtige Bedingung für die Wahl von  $m$ , die allen semiparametrischen

Tab. 3.16: RMSE von  $\hat{d}$  bei Verwendung von Modellselektionskriterien  $n = 250$  in Abhängigkeit von  $d$

Modell	$d$	Whittle-Approximation			HR-Approximation		
		AIC	BIC	HQC	AIC	BIC	HQC
ARFIMA(0, $d$ , 0)	-0,3	0,1383	0,0857	0,1327	0,0818	0,0856	0,0934
	-0,1	0,1710	0,0873	0,1414	0,1169	0,0885	0,0967
	0	0,1857	0,0852	0,1472	0,1320	0,0869	0,0956
	0,2	0,2054	0,0771	0,1581	0,1546	0,0819	0,0895
	0,4	0,2443	0,0621	0,1604	0,1637	0,0705	0,0901
ARFIMA(1, $d$ , 0)	-0,3	0,2524	0,2325	0,2354	0,1952	0,1543	0,1797
	-0,1	0,2584	0,2294	0,2425	0,2377	0,1614	0,1939
	0	0,2713	0,2122	0,2389	0,2102	0,1527	0,1819
	0,2	0,2933	0,2263	0,2668	0,2429	0,1614	0,2047
	0,4	0,3101	0,2021	0,2287	0,2955	0,1560	0,1842
ARFIMA(1, $d$ , 1)	-0,3	0,2436	0,2320	0,2543	0,1677	0,2068	0,1984
	-0,1	0,2324	0,2212	0,2417	0,1734	0,1634	0,1806
	0	0,2645	0,2178	0,2393	0,1930	0,1741	0,1854
	0,2	0,3116	0,2051	0,2491	0,1860	0,1764	0,1930
	0,4	0,3436	0,1705	0,2417	0,2379	0,1654	0,1904

Tab. 3.17: RMSE von  $\hat{d}$  bei Verwendung von Modellselektionskriterien  $n = 500$  in Abhängigkeit von  $d$

Modell	$d$	Whittle-Approximation			HR-Approximation		
		AIC	BIC	HQC	AIC	BIC	HQC
ARFIMA(0, $d$ , 0)	-0,3	0,1196	0,0745	0,1147	0,0793	0,0346	0,0840
	-0,1	0,1479	0,0713	0,1223	0,1133	0,0560	0,0870
	0	0,1605	0,0696	0,1273	0,1279	0,0460	0,0860
	0,2	0,1491	0,0694	0,1157	0,1206	0,0401	0,0891
	0,4	0,1213	0,0691	0,0938	0,0976	0,0450	0,0758
ARFIMA(1, $d$ , 0)	-0,3	0,1731	0,1657	0,1747	0,1427	0,1284	0,1339
	-0,1	0,1772	0,1635	0,1800	0,1651	0,1431	0,1444
	0	0,1860	0,1512	0,1773	0,1537	0,1289	0,1404
	0,2	0,1932	0,1476	0,1721	0,1532	0,1233	0,1349
	0,4	0,1819	0,1406	0,1667	0,1464	0,1155	0,1249
ARFIMA(1, $d$ , 1)	-0,3	0,1927	0,1740	0,1924	0,1433	0,1520	0,1515
	-0,1	0,1837	0,1660	0,1829	0,1367	0,1449	0,1440
	0	0,1889	0,1808	0,1811	0,1556	0,1427	0,1426
	0,2	0,1962	0,1764	0,1757	0,1572	0,1412	0,1429
	0,4	0,2038	0,1722	0,1705	0,1490	0,1303	0,1350

Verfahren gemein ist, ist

$$\frac{1}{m} + \frac{m}{n} \rightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (3.41)$$

$m$  muss also unendlich groß werden für unendlich große Stichproben, damit semi-parametrische Schätzverfahren konsistent werden. Gleichzeitig muss aber auch gegeben sein, dass  $m$  mit einer kleineren Rate gegen unendlich geht wie der Beobachtungsumfang  $n$ , um die Approximation der Spektraldichte im Bereich des Ursprungs  $\lambda \rightarrow 0+$  weiterhin zu gewährleisten wie Robinson (1995b) beschreibt. Gewöhnlich werden folgende funktionale Zusammenhänge zwischen  $m$  und  $n$  angenommen, die diese Bedingung erfüllen:

$$m = n^\alpha \quad \text{mit } 0 < \alpha < 1 \quad (3.42)$$

oder

$$m = Kn^\beta \quad \text{mit } K > 0 \quad \text{und } 2/3 < \beta < 1. \quad (3.43)$$

Als semiparametrische Schätzverfahren stellen wir nun in den folgenden Abschnitten das Verfahren nach Geweke und Porter-Hudak (1983) und die lokale Whittle-Schätzung vor, die eine Schätzung des Integrationsgrads von stationären Prozessen erlaubt. Nicht-stationäre Prozesse können mit diesen Verfahren nur eingeschränkt betrachtet werden und so betrachten wir für die lokale Whittle-Schätzung noch die Erweiterung von Phillips und Shimotsu (2005), die eine direkte Schätzung des Integrationsgrads auch von nicht-stationären Prozessen ermöglicht.

### 3.2.1 Das Geweke/Porter-Hudak-Verfahren

Einen semiparametrischen Schätzer haben Geweke und Porter-Hudak (weiterhin GPH, 1983) entwickelt. Er basiert wie der Schätzer nach Whittle auf der Anpassung des theoretischen Spektrums an das empirisch ermittelbare Periodogramm eines Prozesses in Abhängigkeit des Parameters  $d$ . Da dieses logarithmiert aufeinander regressiert wird, wird diese Schätzmethode auch häufig als Log-Periodogramm-Regression bezeichnet. Im vorangegangenen Kapitel hatten wir uns mit der Spektraldichte eines ARFIMA-Prozesses beschäftigt und hatten folgende Erkenntnis daraus gezogen:

$$f(\lambda) = \left[ 2 \sin \left( \frac{\lambda}{2} \right) \right]^{-2d} f_{\text{ARMA}}(\lambda). \quad (3.44)$$

Dabei bezeichnet  $f_{\text{ARMA}}(\lambda)$  die Spektraldichte, die durch mögliche AR- und MA-Bestandteile verursacht wird. Logarithmieren wir nun diesen Ausdruck, addieren und subtrahieren auf der rechten Seite  $\ln f_{\text{ARMA}}(0)$ , dann erhalten wir folgende lineare Gleichung:

$$\ln f(\lambda) = \ln f_{\text{ARMA}}(0) - d \ln \left[ 4 \sin^2 \left( \frac{\lambda}{2} \right) \right] + \ln \frac{f_{\text{ARMA}}(\lambda)}{f_{\text{ARMA}}(0)}. \quad (3.45)$$

Nun fügen wir auf beiden Seiten das Periodogramm ein und bringen  $\ln f(\lambda)$  auf die rechte Seite und indexieren die Fourierfrequenz  $\lambda_j = 2\pi j/n$  für  $j = 1, \dots, m$ :

$$\ln I(\lambda_j) = \ln f_{\text{ARMA}}(0) - d \ln \left[ 4 \sin^2 \left( \frac{\lambda_j}{2} \right) \right] + \ln \frac{f_{\text{ARMA}}(\lambda_j)}{f_{\text{ARMA}}(0)} + \ln \frac{I(\lambda_j)}{f(\lambda_j)}. \quad (3.46)$$

Betrachten wir die einzelnen Terme dieser Gleichung, so lässt sich für den nicht-konstanten Term  $\ln [f_{\text{ARMA}}(\lambda_j)/f_{\text{ARMA}}(0)]$  aussagen, dass dieser gegen 0 geht, falls die Anzahl der berücksichtigten Frequenzen  $m$  so gewählt wird, dass sie die Bedingung (3.41) erfüllt. Der Term  $2 \cdot I(\lambda_j)/f(\lambda_j)$  stellt, wie bereits erwähnt, eine  $\chi_2^2$ -verteilte Zufallsvariable dar. Es lässt sich zeigen, dass der Erwartungswert des natürlichen Logarithmus dieser Zufallsvariable der negativen Eulerschen Konstanten  $C$ , also  $-0,577216\dots$ , entspricht. Setzen wir nun  $a = \ln f_{\text{ARMA}}(0) - C$  und den Störterm  $\eta_j = \ln [I(\lambda_j)/f(\lambda_j)] + C$ , erhalten wir folgende lineare Regressionsgleichung, aus der wir  $d$  schätzen können:

$$\ln I(\lambda_j) = a - d \ln \left[ 4 \sin^2 \left( \frac{\lambda_j}{2} \right) \right] + \eta_j \quad \text{für } j = 1, \dots, m. \quad (3.47)$$

Der Integrationsgrad  $d$  kann mit Hilfe des einfachen KQ-Schätzers für die Regressionsgleichung (3.47) bestimmt werden:

$$\hat{d}_{GPH} = - \frac{\sum_{j=1}^m (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^m (x_j - \bar{x})^2} \quad (3.48)$$

mit  $y_j = \ln I(\lambda_j)$  und  $x_j = \ln [4 \sin^2 (\lambda_j/2)]$ . Um eine Aussage über die Effizienz der Schätzung machen zu können, ist es notwendig, die Verteilungscharakteristika von  $\eta_j$  zu analysieren. Hierzu argumentieren Geweke und Porter-Hudak (1983), dass der Störterm  $\eta_j$  unter der Bedingung eines stationären Prozesses  $y_t$  asymptotisch normalverteilt ist. Kritik an dieser Argumentation üben aber Hassler (1993) und Robinson (1995a), indem sie unabhängig voneinander zeigten, dass für  $\eta_j$  nur im Falle  $d = 0$  die asymptotische Normalverteilung gegeben ist. Des Weiteren kann

Robinson für den Fall  $d \neq 0$  zeigen, dass die Störterme sogar abhängig sind und nicht die gleiche Verteilung besitzen, was eine Überprüfung der Konsistenz erheblich erschwert. Trotzdem kann Robinson (1995a) folgende asymptotische Eigenschaften für  $\hat{d}_{GPH}$  unter der Bedingung  $d \in (-1/2; 1/2)$  aufzeigen:

$$\sqrt{m}(\hat{d}_{GPH} - d) \xrightarrow{a} N(0, \pi^2/24). \quad (3.49)$$

Hurvich, Deo und Brodsky (1998) analysierten das Verhalten von  $\hat{d}_{GPH}$  weiter. Sie konnten die Verzerrung und den mittleren quadratischen Fehler (MSE) der Schätzung von  $d$  in Abhängigkeit der gewählten Bandbreite  $m$  angeben mit

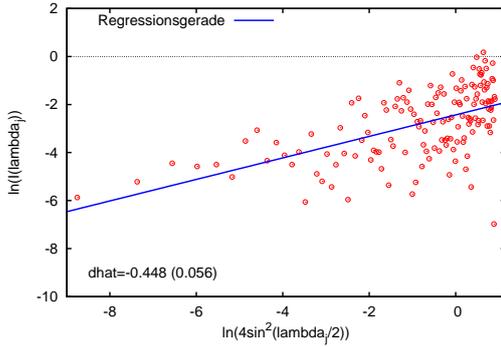
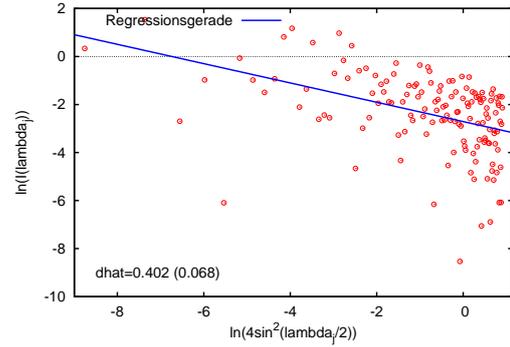
$$E(\hat{d} - d) = -\frac{2\pi^2}{9} \frac{f''_{ARMA}(0)}{f_{ARMA}(0)} \frac{m^2}{n^2} + O\left(\frac{\ln^3 m}{m}\right) \quad (3.50)$$

$$\begin{aligned} \text{MSE}(\hat{d}) = E(\hat{d} - d)^2 &= \frac{4\pi^4}{81} \left[ \frac{f''_{ARMA}(0)}{f_{ARMA}(0)} \right]^2 \frac{m^4}{n^4} \\ &+ O\left(\frac{m(\ln^3 m)}{n^2}\right) + o\left(\frac{m^4}{n^4}\right) + o(m^{-1}). \end{aligned} \quad (3.51)$$

Bemerkenswert ist hierbei, dass wir immer einen konsistenten Schätzer erhalten, wenn  $m^2/n^2 \rightarrow 0$  für  $n \rightarrow \infty$  gilt, was für beide Vorschriften (3.42) und (3.43) gewährleistet ist. Die Konvergenz zum wahren Wert von  $d$  kann aber durchaus sehr lang werden. Deswegen konnten Hurvich, Deo und Brodsky (1998) mit ihren Ergebnissen unter der Bedingung  $f''_{ARMA}(0) \neq 0$  die Bandbreite mit minimalen mittleren quadratischen Fehlern angeben mit

$$m_{GPH} = \left(\frac{27}{128\pi^2}\right)^{1/5} \left[\frac{f''_{ARMA}(0)}{f_{ARMA}(0)}\right]^{-2/5} n^{4/5}. \quad (3.52)$$

Wie die Bandbreite  $m_{GPH}$  empirisch bestimmt werden kann, werden wir in einem der folgenden Abschnitte sehen. Weiterhin wurde auch das Verhalten von  $d_{GPH}$  analysiert, falls der Ausgangsprozess einen nicht-stationären Integrationsgrad  $d \geq 0,5$  aufweist. Hierzu hat Velasco (1999a) festgestellt, dass der Schätzer  $\hat{d}_{GPH}$  asymptotisch normalverteilt ist, falls  $d \in (-1/2; 3/4)$  und immer noch konsistent für  $d \in (-1/2; 1]$ . Für  $d > 1$  ist  $\hat{d}_{GPH}$  nicht mehr konsistent und konvergiert gegen 1. In diesen Fällen kann die Schätzung des Integrationsgrads auf Basis der  $p$ -ten Differenz von  $y_t$  erfolgen, die sicherstellt, dass sich der differenzierte Prozess im stationären Bereich befindet, und durch Zurückaddieren von  $p$  bestimmt werden. Auch ist zu berücksichtigen, dass ein von 0 verschiedener Mittelwert von  $y_t$  die Schätzung mittels  $\hat{d}_{GPH}$  nicht beeinflusst, da die Periodogrammberechnung, die implizit auf

Abb. 3.7: GPH-Schätzung für  $d = -0,4$ Abb. 3.8: GPH-Schätzung für  $d = 0,4$ 

Basis der Autokovarianzfunktion erfolgt, eine Mittelwertzentrierung einschließt.

Die Abbildungen 3.7 und 3.8 zeigen eine beispielhafte Anwendung für zwei fraktional integrierte Rauschen mit Integrationsgraden  $d = -0,4$  und  $d = 0,4$  mit Beobachtungsumfang  $n = 500$ . Die Bandbreite wurde mit  $m = n^{4/5} = 144$  gewählt, was in Abwesenheit von AR- oder MA-Prozessbestandteilen optimal ist. Man sieht hierbei, dass für einen positiven integrierten Prozess die Regressionsgerade fällt und zwar genau mit dem Wert des Integrationsgrads als Steigung. Für negativ integrierte Prozesse gilt dies dann genau umgekehrt, da hier die Testregression steigt. Das GPH-Verfahren erfreut sich in der empirischen Anwendung einer zahlreichen Verwendung, da es sehr einfach mit Hilfe des KQ-Schätzers anwendbar ist.

### 3.2.2 Lokale Whittle-Schätzung

Eine semiparametrische Version des bereits beschriebenen Whittler-Schätzers hat Künsch (1987) vorgeschlagen. Ansatzpunkt des Vorschlags von Künsch (1987) ist wiederum die rechtsseitige Approximation der Spektraldichte für einen stationären ARFIMA-Prozess um den Frequenzbereich  $\lambda = 0$ :

$$f(\lambda) \approx G\lambda^{-2d} \quad \text{für } \lambda \rightarrow 0 + . \quad (3.53)$$

Mit Hilfe dieser Annäherung konnte Künsch (1987) folgende Zielfunktion zur Bestimmung von  $d$  angeben:

$$Q(G, d) = \frac{1}{m} \sum_{j=1}^m \left\{ \ln G\lambda_j^{-2d} + \frac{\lambda_j^{2d}}{G} I(\lambda_j) \right\}. \quad (3.54)$$

Diese Funktion entspricht der negativen, von Whittle approximierten, logarithmierten Likelihood-Funktion, in der die Approximation des Spektrums eingesetzt wurde.

Ähnlich wie bei dem GPH-Schätzverfahren wird eine Anpassung der Spektraldichte nur über  $m$  Fourierfrequenzen durchgeführt, wodurch wiederum die Existenz von ARMA-Bestandteilen berücksichtigt werden kann. Auf Grund dieser Beschränkung wird dieses Verfahren auch lokale Whittle-Schätzung genannt. Die Schätzung für  $d$  bestimmt sich dann durch Minimierung von  $Q(G, d)$  als Quasi-ML-Schätzung, da es sich hierbei um eine negative Likelihoodfunktion handelt:

$$(\hat{G}, \hat{d}_{LW}) = \arg \min_{0 < G < \infty} Q(G, d). \quad (3.55)$$

Alternativ lässt sich  $d$  auch über die Minimierung von  $R(d)$  bestimmen, aus der der Term  $G$  konzentriert wurde:

$$\begin{aligned} \hat{d}_{LW} &= \arg \min R(d), \\ \text{wobei } R(d) &= \ln \hat{G}(d) - \frac{2d}{m} \sum_{j=1}^m \ln \lambda_j \\ \text{mit } \hat{G}(d) &= \frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I(\lambda_j). \end{aligned} \quad (3.56)$$

Robinson (1995b) konnte die Konsistenz der Schätzung für einen stationären ARFIMA-Prozess bei geeigneter Wahl von  $m$  beweisen und die Verteilung für das Schätzergebnis von  $\hat{d}$  angeben mit

$$\sqrt{m}(\hat{d}_{LW} - d) \xrightarrow{a} N(0, 1/4). \quad (3.57)$$

Damit zeigt sich, dass dieses Verfahren eine höhere Effizienz als das GPH-Verfahren bei gleicher gewählter Bandbreite  $m$  aufweist. Ähnlich wie bei dem GPH-Schätzverfahren konnte Velasco (1999b) die asymptotische Normalverteilung von  $\hat{d}_{LW}$  für  $d \in (-1/2; 3/4)$  und die Konsistenz für  $d \in (-1/2; 1]$  bestätigen. Henry und Robinson (1996) führten weitere Analysen zu  $\hat{d}_{LW}$  durch und stellten fest, dass dieser Schätzer mit

$$E(\hat{d}_{LW} - d) = -\frac{2\pi^2}{9} \frac{f''_{\text{ARMA}}(0)}{f_{\text{ARMA}}(0)} \frac{m^2}{n^2} \quad (3.58)$$

die gleiche asymptotische Verzerrung wie  $\hat{d}_{GPH}$  aufweist. Sie konnten auch mit

$$m_{LW} = \left(\frac{3}{4\pi}\right)^{4/5} \left[ \frac{f''_{\text{ARMA}}(0)}{2f_{\text{ARMA}}(0)} - \frac{d}{12} \right]^{-2/5} n^{4/5} \quad (3.59)$$

eine optimale Bandbreite bestimmen, die den mittleren quadratischen Fehler minimiert. Im Gegensatz zu (3.52) hängt beim LW-Schätzer die optimale Bandbreite aber vom Integrationsparameter ab. Dieses Verfahren wurde sukzessive in den letzten Jahren weiterentwickelt. So existiert eine Erweiterung von Andrews und Sun (2004), die sich von der restriktiven Annahme löst, dass die ARMA-Komponente der Spektraldichte konstant ist. Sie passen vielmehr eine polynomiale Struktur der Fourierfrequenz  $\lambda$  an und schätzen die jeweiligen Parameter mit bei der Optimierung.

### 3.2.3 Exakte lokale Whittle-Schätzung

Besonders hervorzuheben ist die Erweiterung von Shimotsu und Phillips (2005) für den Fall, dass es sich bei der zu untersuchenden Zeitreihe  $y_t$  um einen nicht-stationären Prozess handelt. Da die Schätzung dann nicht mehr auf einer oder höheren Differenz von  $y_t$  beruht und im Vorfeld keine Entscheidung über Stationarität gemacht werden muss, benannten sie das Verfahren als exakte lokale Whittle-Schätzung (weiterhin bezeichnet als ELW). Ihr Vorschlag ist der, dass anstatt des Periodogramms des Prozesses  $y_t$  der fraktional gefilterte Prozess  $\Delta^d y_t$  für die Optimierung von  $R(d)$  verwendet wird.

$$\hat{d}_{ELW} = \arg \min_{d \in [\Delta_1; \Delta_2]} R(d), \quad (3.60)$$

wobei  $R(d) = \ln \hat{G}(d) - \frac{2d}{m} \sum_{j=1}^m \ln \lambda_j$

mit  $\hat{G}(d) = \frac{1}{m} \sum_{j=1}^m I_{\Delta^d y}(\lambda_j)$ .

Diese Modifikation erweitert die Anwendbarkeit des Schätzers auch für nicht-stationäre Prozesse und Shimotsu und Phillips (2005) zeigen darüber hinaus, dass  $\hat{d}_{ELW}$  konsistent und asymptotisch normalverteilt ist, falls der wahre Integrationsgrad  $d$  im Bereich  $\Delta_2 - \Delta_1 \leq 9/2$  liegt. Die asymptotische Verteilung von  $\hat{d}_{ELW}$  ergibt sich dann als

$$\sqrt{m}(\hat{d}_{ELW} - d) \xrightarrow{a} N(0, 1/4). \quad (3.61)$$

Shimotsu (2010) entwickelt das ELW-Schätzverfahren weiter, indem er die Möglichkeit zulässt, dass der Prozess  $y_t$  auch einen von 0 verschiedenen Mittelwert  $\mu_0$

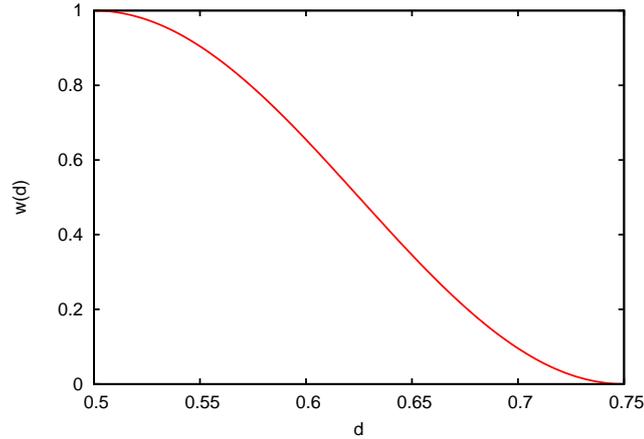


Abb. 3.9: Gewichtungsfunktion  $w(d)$  für  $d \in [1/2; 3/4]$

aufweisen kann. Nimmt man an, dass  $y_t$  erzeugt wird durch

$$y_t = \mu_0 + y_t^0 \quad \text{mit} \quad y_t^0 = (1 - L)^{-d} u_t \quad \text{und} \quad u_t \sim IID(0, \sigma^2), \quad (3.62)$$

dann lässt sich der Parameter  $\mu_0$  einfach als Stichprobenmittel  $\bar{y} = \sum_{t=1}^n y_t$  schätzen. Das Problem ist aber hierbei, dass die Varianz des Stichprobenmittels mit steigendem  $d$  sehr groß werden kann. Adenstedt (1974) argumentiert, dass für stationäre Prozesse, also Prozesse mit Integrationsgrad  $d < 1/2$ , das Stichprobenmittel der beste lineare erwartungstreue Schätzer darstellt. Für  $d > 1/2$  wird die Varianz von  $y_t^0$  aber unendlich groß für  $n \rightarrow \infty$  und dominiert wahrscheinlich die Inkonsistenz, die durch die Existenz von  $\mu_0$  erzeugt wird. Also kann in diesen Fällen darauf verzichtet werden,  $\mu_0$  zu schätzen. Für kleine Stichproben schlägt Shimotsu (2010) im Falle von  $d > 1/2$  vor,  $\mu_0$  einfach durch die erste Beobachtung  $y_1$  zu schätzen. Da nun der Integrationsgrad von  $y_t^0$  vor der Schätzung unbekannt ist, ist laut Shimotsu (2010) eine gewichtete Mittelwertschätzung innerhalb des Schätzverfahrens für  $\hat{d}_{ELW}$  sinnvoll:

$$\tilde{\mu}_0(d) = w(d)\bar{y} + [1 - w(d)]y_1. \quad (3.63)$$

Durch Monte-Carlo-Simulationen fand Shimotsu (2010) heraus, dass folgende Gewichtungsfunktion gute Ergebnisse erzielt:

$$w(d) = \begin{cases} 1 & \text{für } d < 1/2 \\ (1/2)[1 + \cos(4\pi d)] & \text{für } 1/2 \leq d \leq 3/4 \\ 0 & \text{für } d > 3/4 \end{cases} \quad (3.64)$$

Der anwendbare ELW-Schätzer  $\hat{d}_{FELW}$  ergibt sich dann als

$$\begin{aligned}\hat{d}_{FELW} &= \arg \min_{d \in [\Delta_1; \Delta_2]} R_F(d), \\ \text{wobei } R_F(d) &= \ln \hat{G}(d) - \frac{2d}{m} \sum_{j=1}^m \ln \lambda_j \\ \text{mit } \hat{G}(d) &= \frac{1}{m} \sum_{j=1}^m I_{\Delta^d[y - \tilde{\mu}_0(d)]}(\lambda_j).\end{aligned}\tag{3.65}$$

Die Schätzung selbst wird dann iterativ angewandt, indem ausgehend von einem Startwert für  $d$  solange die Schätzung durchgeführt wird, bis eine Konvergenz in  $\hat{d}$  erreicht wird.

### 3.2.4 Vergleich der semiparametrischen Verfahren

In diesem Abschnitt wollen wir zunächst die drei vorgestellten Verfahren nur für simulierte fraktional integrierte Rauschen vergleichen, d.h. wir verzichten auf etwaige AR- und MA-Prozessbestandteile. Bei den Simulationen variieren wir den Stichprobenumfang  $n$  und schätzen den Integrationsgrad  $d$  mit der jeweils gleichen Bandbreite  $m = \lfloor n^{0,8} \rfloor$  zur besseren Vergleichbarkeit. Den Integrationsgrad selbst fixieren wir mit  $d = 0$  und führen insgesamt 50.000 Wiederholungen für jedes  $n$  durch. Die Schätzverfahren bewerten wir wieder anhand der durchschnittlichen Verzerrung im Schätzergebnis und der Wurzel des mittleren quadratischen Fehlers (RMSE), um die Effizienz der jeweiligen Verfahren zu untersuchen. Die Simulationsergebnisse sind in den Graphen der Abbildung 3.10 abgetragen und zeigen, dass das GPH-Verfahren schon bei geringen Stichprobenumfängen eine wesentlich geringere Verzerrung aufweist als das LW- oder ELW-Verfahren. Gerade das ELW-Verfahren weist selbst bei hohen Werten von  $n$  noch eine relative hohe Verzerrung auf und hat auch eine wesentlich schwächere Konvergenz. Auf Basis des RMSE lässt sich aussagen, dass das LW-Verfahren am effizientesten ist und dass selbst das ELW-Verfahren noch eine höhere Effizienz aufweist als das nach GPH.

Für den Fall  $n = 1000$  führen wir nun weitere Simulationen durch, um die Verteilungseigenschaften der geschätzten Parameter zu analysieren. Auch hierfür verwenden wir fraktional integriertes Rauschen ohne zusätzliche AR- und MA-Bestandteile und verwenden die gleiche Regel für die Bandbreite  $m = \lfloor n^{0,8} \rfloor = 251$  für die einzelnen Schätzungen.

In Tabelle 3.18 tragen wir die einzelnen geschätzten Verteilungsparameter ab, nämlich den Mittelwert (MW), die Standardabweichung (SD), die Schiefe (S) und die

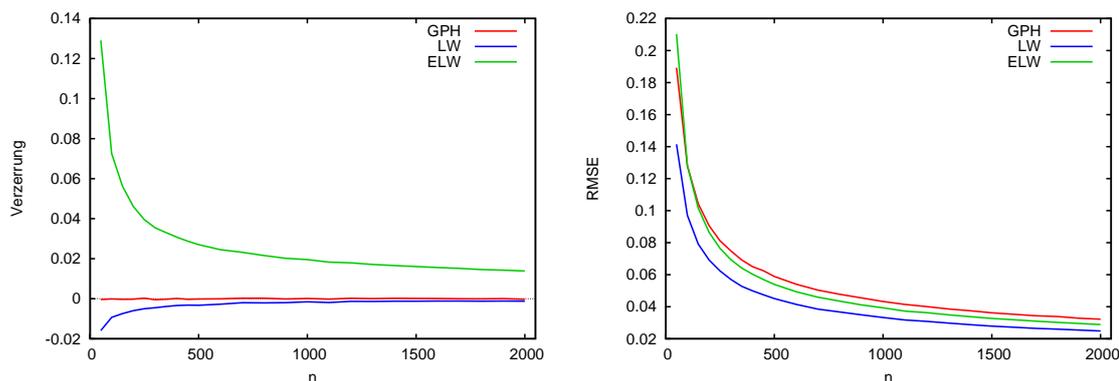


Abb. 3.10: Verzerrung und RMSE in Abhängigkeit von  $n$  mit  $m = n^{0.8}$  und  $d = 0$

Überkurtosis ( $\text{ExK} = \text{Kurtosis} - 3$ ). Man kann hierbei erkennen, dass für  $d \leq 1$  das GPH- und LW-Verfahren den wahren Integrationsgrad im Mittel relativ gut treffen und für  $d > 1$  verzerrt ist, was auch zu erwarten war. Das ELW-Verfahren hingegen kommt auch in diesen Fällen dem wahren Integrationsgrad nahe, überschätzt diesen aber in allen Fällen.

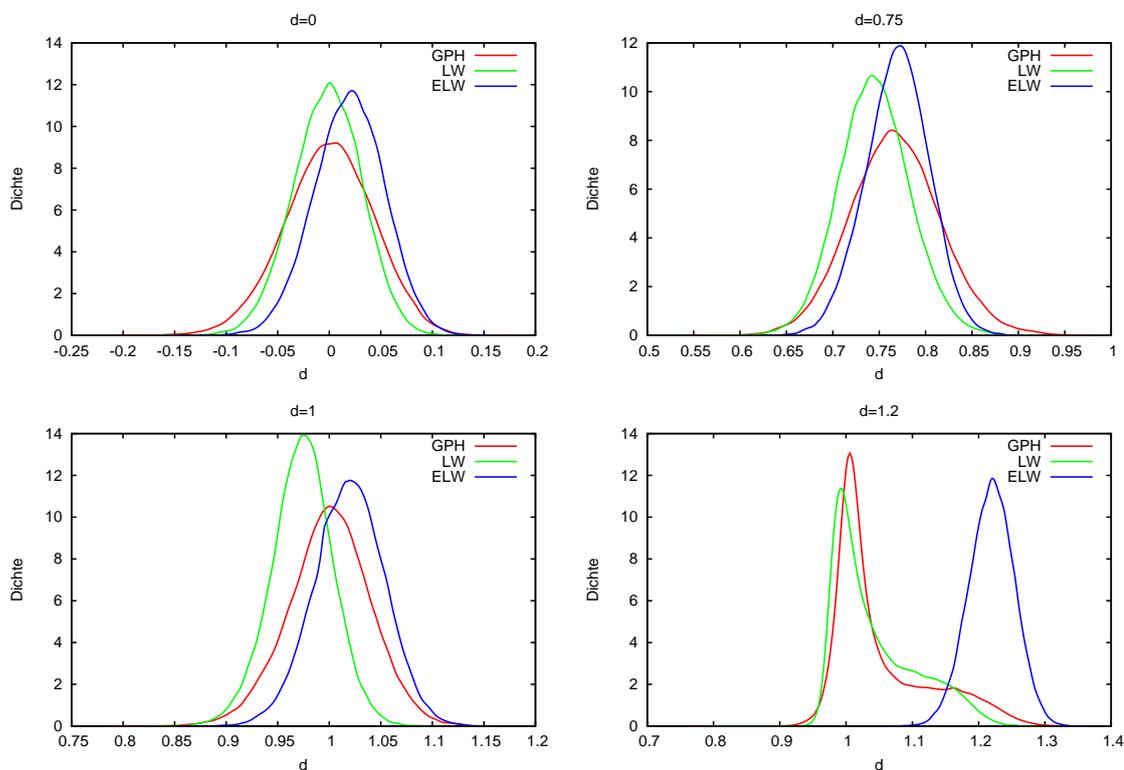
Die geschätzten Standardabweichungen treffen die asymptotisch erwarteten relativ gut. Für das GPH-Verfahren erwarten wir  $\sqrt{\pi^2/(24 \cdot 251)} \approx 0,0405$  und analog für LW und ELW  $\sqrt{1/(4 \cdot 251)} \approx 0,0316$ . Hierbei zeigt sich aber, dass für das ELW-Verfahren die Standardabweichungen durchweg im stationären Bereich über denen des LW-Verfahrens liegen.

Die durchwegs negativ geschätzten Schiefe-Parameter legen nahe, dass die geschätzten Verteilungen tendenziell linksschief bei endlichen Stichproben sind. Auch weisen die Verteilungen eine stärker ausgeprägte Wölbung auf als eine Normalverteilung. Wendet man den bekannten Jarque-Bera-Test an, so ergibt sich für sämtliche Verteilungen eine signifikante Ablehnung der Normalverteilungsannahme zu jedem üblichen Signifikanzniveau. Diese lässt sich erst bei sehr großen Stichprobenumfängen  $n$ , was wiederum einher geht mit großen Bandbreiten  $m$ , annehmen. Für  $d > 1$  ergibt sich im Falle des GPH- und LW-Verfahrens eine Verzerrung zu 1, was die geschätzten Dichtefunktionen in Abbildung 3.11 verdeutlichen.

Im nächsten Schritt vergleichen wir die Verfahren, falls eine ARFIMA-Struktur des zu untersuchenden Prozesses  $y_t$  vorliegt. Für das Schätzverfahren nach GPH und LW haben wir bereits gesehen, dass die Bandbreite mit dem geringsten MSE den Term  $f''_{\text{ARMA}}(0)/f_{\text{ARMA}}(0)$  gemeinsam haben. Dieser Ausdruck lässt sich analytisch

Tab. 3.18: Verteilungsparameter für das GPH-, LW- und ELW-Verfahren für  $n = 1000$ 

$d$		-0,2	0	0,2	0,4	0,6	0,75	1	1,2	1,4
<i>GPH</i>	MW	-0,1997	-0,0002	0,2005	0,4021	0,6078	0,7675	1,0001	1,0521	1,0441
	SD	0,0434	0,0432	0,0434	0,0438	0,0448	0,0489	0,0399	0,0724	0,0923
	S	-0,1384	-0,1334	-0,1283	-0,1461	-0,0755	0,1068	-0,1505	1,1463	2,6405
	ExK	0,0548	0,0504	0,0640	0,0943	0,0935	0,1916	0,2778	0,3870	6,4467
<i>LW</i>	MW	-0,1962	-0,0019	0,1935	0,3902	0,5906	0,7451	0,9735	1,0433	1,0405
	SD	0,0334	0,0332	0,0334	0,0336	0,0352	0,0384	0,0300	0,0625	0,0926
	S	-0,0961	-0,1090	-0,1252	-0,1049	-0,0607	0,1064	-0,1217	0,9380	2,0706
	ExK	0,0046	0,0063	0,0914	0,0570	0,0449	0,1858	0,2519	-0,0854	3,6506
<i>ELW</i>	MW	-0,1810	0,0192	0,2195	0,4199	0,6263	0,7695	1,0195	1,2194	1,4192
	SD	0,0342	0,0341	0,0342	0,0341	0,0329	0,0342	0,0341	0,0343	0,0340
	S	-0,0867	-0,0950	-0,1115	-0,0901	-0,2287	-0,0994	-0,0681	-0,1021	-0,0886
	ExK	0,0072	0,0004	0,0979	0,0854	0,1493	0,0222	0,0280	0,0271	0,0181

Abb. 3.11: Geschätzte Dichtefunktionen für  $\hat{d}_{GPH}$ ,  $\hat{d}_{LW}$  und  $\hat{d}_{ELW}$

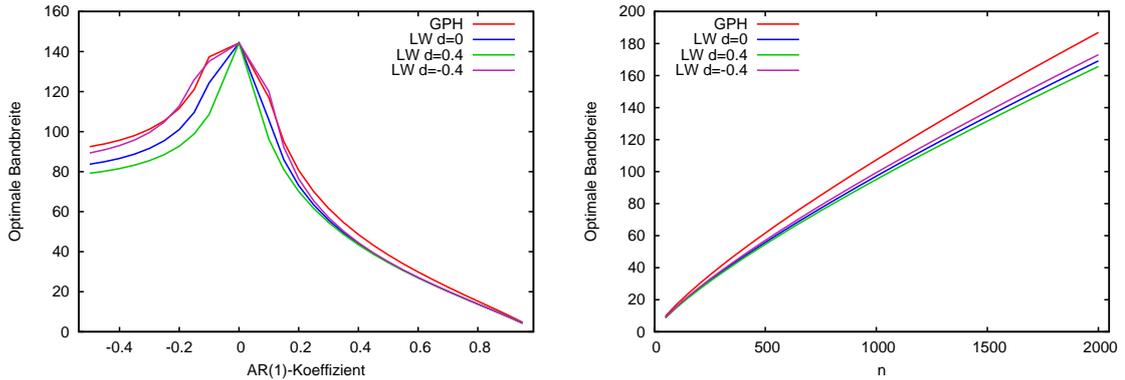


Abb. 3.12: Optimale Bandbreiten in Abhängigkeit des AR(1)-Koeffizienten  $\theta$  und Stichprobenumfang  $n$

für ARFIMA(1, $d$ ,0)-Prozesse angeben mit

$$\frac{f''_{\text{AR}}(0)}{f_{\text{AR}}(0)} = -\frac{2\theta}{(1-\theta)^2} \quad (3.66)$$

und für ARFIMA(1, $d$ ,1)-Prozesse mit dem AR(1)-Parameter  $\theta$  und MA(1)-Parameter  $\phi$  analog

$$\frac{f''_{\text{ARMA}}(0)}{f_{\text{ARMA}}(0)} = -\frac{2\phi(1-\theta)^2 - 2\theta(1-\phi)^2}{(1-\theta)^2(1-\phi)^2}. \quad (3.67)$$

Mit diesen Ergebnissen lässt sich die optimale Bandbreite mit Kenntnis der wahren Parameter bestimmen. Für AR(1)-Prozesse ergibt sich mit einem Beobachtungsumfang  $n = 500$  in Abbildung 3.12 die Struktur der optimalen Bandbreiten. Für  $\theta = 0$  lässt sich mit diesem Ansatz keine optimale Bandbreite bestimmen. Hierfür ergibt sich in Simulationen  $m = n^{0,8}$  als optimale Bandbreite. Die Ergebnisse zeigen, dass der Integrationsgrad mit dem GPH-Verfahren durchweg mit einer höheren Bandbreite geschätzt werden kann als mit dem LW-Verfahren, das bei gleicher Bandbreite effizienter ist.

Bevor wir nun die Eigenschaften der Schätzer bei Vorliegen von ARFIMA-Ordnungen untersuchen, müssen wir noch entscheiden, wie sich die Struktur der optimalen Bandbreiten für das ELW-Verfahren gestaltet. Bisher gibt es hierfür noch keine Überlegungen und es scheint intuitiv plausibel, dass die Strukturen für das LW- und ELW-Verfahren ähnlich sein sollten. Deswegen versuchen wir in Simulationen mit ARFIMA(1, $d$ ,0)-Prozessen in Abhängigkeit von  $\theta$  für das ELW-Verfahren jeweils die Bandbreite zu bestimmen, die den geringsten MSE aufweist. Diese vergleichen wir in Abbildung 3.13 mit der des LW-Verfahrens. Wie zu erwarten war, ergibt sich

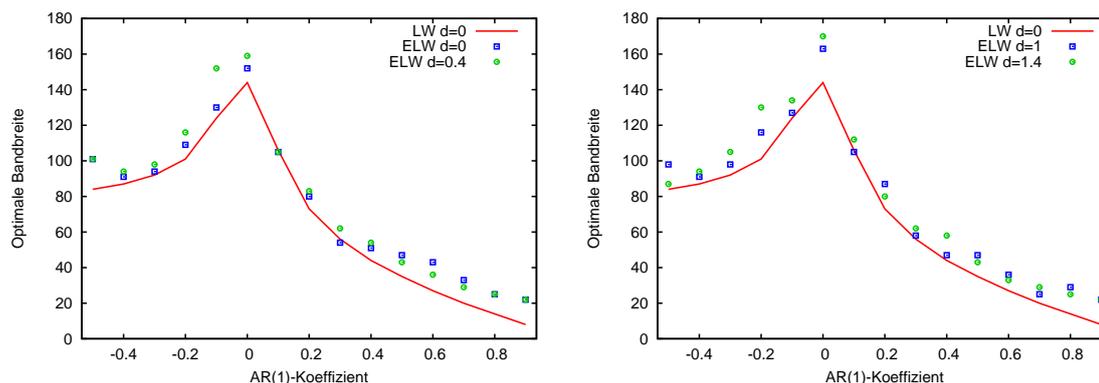


Abb. 3.13: Optimale Bandbreiten in Abhängigkeit des AR(1)-Koeffizienten  $\theta$  für das LW- und ELW-Verfahren mit Stichprobenumfang  $n = 500$

eine ähnliche Struktur mit dem Unterschied, dass diese geringfügig höher liegt als beim LW-Verfahren. Dies stellt zwar keinen Beweis dar, dass wir die gleichen optimalen Bandbreiten haben, aber wir können diese in den folgenden Simulationen als solche verwenden. Wir verzichten aber für das ELW-Verfahren auf den Term  $d/12$  in (3.59), da wir in der Abbildung 3.13 zum einen keinen Zusammenhang zwischen optimaler Bandbreite und dem Integrationsgrad  $d$  erkennen und zum anderen auf den Hinblick, dass das ELW-Verfahren gerade auch für nicht-stationäre Prozesse angewendet werden soll.

In den folgenden Simulationen vergleichen wir nun die Eigenschaften der semi-parametrischen Schätzverfahren für den Fall, dass der datengenerierende Prozess  $y_t$  einer ARFIMA(1,  $d$ , 0) folgt. Wir erzeugen also Zeitreihen der Form

$$\Delta^d(1 - \theta L)y_t = \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim N(0; 1), \quad (3.68)$$

wobei  $\theta$  Werte im Intervall  $[-0,5; 0,9]$  annehmen kann und zur Bestimmung der optimalen Bandbreite für die Schätzverfahren bekannt sein soll. Für  $\theta = 0$  lässt sich mit diesen Ansätzen keine optimale Bandbreite bestimmen und hierfür verwenden wir die Bandbreite  $m = n^{4/5}$ . Als Beobachtungsumfang wählen wir  $n = 500$  und führen, wie auch für die eingangs dieses Abschnitts durchgeführten Simulationen, für jeden Parameterwert von  $\theta$  50.000 Wiederholungen durch. Zusätzlich zur durchschnittlichen Verzerrung und des RMSE berichten wir auch über die Abdeckungshäufigkeit, dass der wahre Parameter  $d$  in einem 95%-Konfidenzintervall um den geschätzten Integrationsgrad liegt. Diese Wahrscheinlichkeit können wir als eine Gegenwahrscheinlichkeit zur Fehlerwahrscheinlichkeit interpretieren, dass wir aufgrund der angenommenen Verteilung des Schätzergebnisses den wahren Integrationsgrad

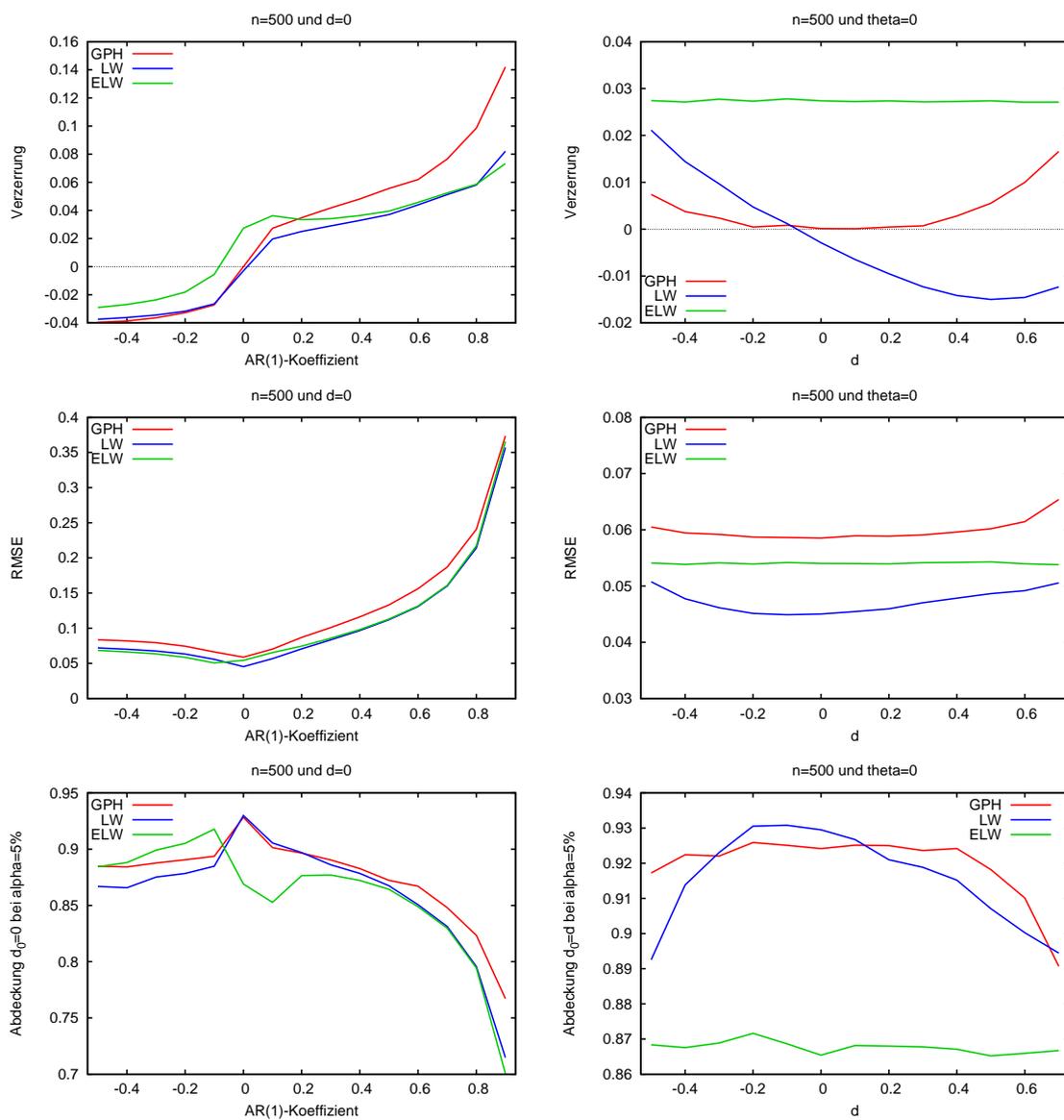


Abb. 3.14: Vergleich der Schätzverfahren für ARFIMA(1,  $d$ , 0)-Prozesse

ablehnen, was einem Fehler erster Art entspricht. Dabei verwenden wir jeweils die asymptotischen Standardabweichungen und nehmen an, dass das Schätzergebnis normalverteilt ist, was zumindest auch asymptotisch gegeben ist. Damit ergeben sich für die beiden Whittle-Schätzer  $d \pm 1,96\sqrt{1/(4 \cdot m)}$  als Ober- und Untergrenze für das 95%-Konfidenzintervall bzw.  $d \pm 1,96\pi\sqrt{1/(24 \cdot m)}$  für den GPH-Schätzer. Die Simulationsergebnisse stellen wir jeweils in Abhängigkeit des Parameter  $\theta$  in der linken Spalte der Abbildung 3.14 graphisch dar. Hierbei zeigt sich, dass das GPH-Verfahren und das LW-Verfahren für  $\theta \leq 0$  ähnlich verzerrt sind. Im Falle stärkerer positiver Autokorrelation ist die Verzerrung geringer für das LW-Verfahren. Das ELW-Verfahren weist eine geringere Verzerrung als das LW-Verfahren im Falle negativer Autokorrelation auf, hat dann aber eine höhere für  $\theta \geq 0$ , ist aber generell auf dem gleichem Niveau. Auffällig ist die größere Verzerrung bei  $\theta = 0$ , die wir schon bei unseren anfänglichen Simulationen in diesem Abschnitt gesehen haben. Auf Basis des RMSE lässt sich aussagen, dass, obwohl beim GPH-Verfahren eine höhere Bandbreite zur Schätzung von  $d$  verwendet wird, dieses durchweg weniger effizient ist als das LW- und ELW-Verfahren, die jeweils einen ähnlichen RMSE aufweisen. Das Schätzergebnis im GPH-Verfahren hat aber eine geringere Wahrscheinlichkeit für einen Fehler erster Art, also dass das wahre  $d$  auf Grund einer Schätzung abgelehnt wird. Gerade bei stark ausgeprägter Autokorrelation wird mit relativ hoher Wahrscheinlichkeit das wahre  $d$  signifikant unterschiedlich bei LW und ELW geschätzt.

In der rechten Spalte der Abbildung 3.14 variieren wir den Parameter  $d$  im Intervall  $[-0,5; 0,7]$  und setzen  $\theta = 0$ . Hier zeigt sich, dass das ELW-Verfahren relativ unabhängig von  $d$  konstant verzerrt ist. Das Verfahren nach GPH hat hingegen eine geringere absolute Verzerrung als das LW-Verfahren. Das LW-Verfahren ist aber für alle verwendeten Werte von  $d$  effizienter und selbst das ELW-Verfahren hat konstant einen besseren RMSE als das GPH-Verfahren. Das GPH-Verfahren hat hingegen wiederum eine geringere Wahrscheinlichkeit für einen Fehler erster Art. Das ELW-Verfahren schneidet diesbezüglich relativ schlecht ab, da es scheinbar eine schwächere Konvergenz hat als das GPH- oder LW-Verfahren, was man an der größeren absoluten Verzerrung erkennen kann. Dieses eignet sich demnach vor allem für größere Stichproben und natürlich auch dann, wenn bzgl. der Stationarität eines zu untersuchenden Prozesses Zweifel bestehen. Ansonsten lassen die Simulationsergebnisse keinen genauen Schluss darüber zu, ob das LW-Verfahren dem GPH-Verfahren vorzuziehen ist. Das LW-Verfahren ist zwar effizienter als jenes nach GPH, dafür hat dieses aber eine bessere Konvergenz und ist somit gerade für kleinere Stich-

proben geeignet. Interessant ist natürlich auch ein Vergleich der Verfahren, falls  $y_t$  einer ARFIMA(1,  $d$ , 1)-Ordnung folgt. Diesen Vergleich werden wir im folgenden Abschnitt ziehen, wenn wir unter anderem annehmen werden, dass die wahren Parameter, die wir zur Bestimmung der optimalen Bandbreite benötigen, unbekannt sind.

### 3.2.5 Wahl der Bandbreite $m$

In diesem Abschnitt wollen wir uns Gedanken machen, wie wir die Anzahl der Frequenzen  $m$  wählen können, wenn die wahre Ordnung und Parameterwerte unbekannt sind. Dazu verfolgen wir zunächst den Ansatz von Hurvich und Deo (1999), die, ausgehend von der Gleichung (3.45), das Periodogramm darstellen als

$$\ln I(\lambda_j) = -d \ln[4 \sin^2(\lambda_j/2)] + \ln f_{\text{ARMA}}(\lambda_j) + \ln \frac{I(\lambda_j)}{f(\lambda_j)}. \quad (3.69)$$

Für die logarithmierte Spektraldichte der ARMA-Bestandteile  $\ln f_{\text{ARMA}}(\lambda_j)$  führen sie eine Taylorreihenentwicklung um  $\lambda = 0$  durch mit

$$\ln f_{\text{ARMA}}(\lambda_j) \approx \ln f_{\text{ARMA}}(0) + \frac{\lambda_j^2}{2} K + \frac{\lambda_j^3}{6} R_j. \quad (3.70)$$

Der Parameter  $K$  entspricht hierbei genau dem Term  $f''_{\text{ARMA}}(0)/f_{\text{ARMA}}(0)$ , der zur Bestimmung der optimalen Bandbreite notwendig ist. Setzt man nun die Gleichung (3.70) in (3.69) ein und formt diese ähnlich wie bei (3.47) um, dann kann der Parameter  $K$  in einer Regression in der Form von

$$\ln I(\lambda_j) = a - d \ln[4 \sin^2(\lambda_j/2)] + \frac{\lambda_j^2}{2} K + \eta_j \quad \text{für } j = 1, \dots, m_1 \quad (3.71)$$

geschätzt werden. Da der relevante Term  $K = f''_{\text{ARMA}}(0)/f_{\text{ARMA}}(0)$  durch Einsetzen in die Gleichung (3.69) gewonnen wird, wird diese Methode auch allgemein als Plugin-Verfahren bezeichnet. Der Term  $R_j$ , so argumentieren Hurvich und Deo (1999), ist gleichmäßig beschränkt im Bereich des Ursprungs und kann zur Schätzung von  $K$  vernachlässigt werden. Zur Durchführung der Regression (3.71) werden  $m_1$  Anzahl Frequenzen berücksichtigt und Hurvich und Deo (1999) schlagen hierfür die Regeln  $m_1 = 0,2n^{6/7}$  und  $m_1 = 0,3n^{6/7}$  vor. Hierbei scheint der Exponent  $6/7$  nicht intuitiv zu sein, da die optimale Bandbreite für die Verfahren nach GPH und LW jeweils proportional zu  $n^{4/5}$  sind. Hurvich und Deo (1999) konnten aber zeigen, dass bei diesen Exponenten der mittlere quadratische Fehler von  $\hat{K}$  minimal ist

und wenden das Plugin-Verfahren zunächst nur für das GPH-Verfahren an. Für das LW-Verfahren ist der Übertrag unproblematisch, da der angenommene Integrationsgrad für den datengenerierenden Prozess der gleiche ist. Für das ELW-Verfahren, das auch nicht-stationäre Prozesse umschließt, ist das nun nicht mehr so einfach, da sich das Periodogramm dann nicht mehr wie in (3.69) darstellen lässt. Wir schlagen deswegen vor, die Schätzung in (3.71) auf Basis des fraktional gefilterten Prozesses durchzuführen. Für die Filtration wählen wir als Integrationsgrad den geschätzten Parameter  $\hat{d}_{ELW}$ , den wir im ersten Schritt z.B. mit  $m = n^{0,8}$  schätzen können. Daraufhin können wir eine weitere Schätzung von  $d$  durchführen und wiederum eine optimale Bandbreite auf der Basis des neu geschätzten Integrationsgrads mit (3.71) bestimmen. Dieses Vorgehen können wir dann genügend häufig wiederholen, bis wir eine Konvergenz erreichen.

Als ein weiteres Verfahren zur Schätzung der optimalen Bandbreite können wir folgendes parametrische Verfahren vorschlagen: Ähnlich wie bei dem Plugin-Verfahren bei Anwendung der ELW-Schätzung können wir eine erste Schätzung mit  $m = n^{0,8}$  durchführen und den Integrationsgrad schätzen. Diesen verwenden wir dann zur fraktionalen Differentiation des zu untersuchenden Prozesses und führen daraufhin eine AR(1)- oder ARMA(1,1)-Schätzung durch. Die daraus resultierenden geschätzten Parameter können wir dann verwenden, um im AR(1)-Fall mit (3.66) bzw. im ARMA(1,1)-Fall mit (3.67) den Parameter  $K$  zu schätzen und damit die optimale Bandbreite zu bestimmen. Mit dieser wiederum führen wir wieder eine Schätzung des Integrationsgrades durch und schätzen die Parameter der AR(1)- bzw. ARMA(1,1)-Darstellung, mit der wir dann wieder eine optimale Bandbreite bekommen. Dieses Vorgehen wiederholen wir dann, bis wir eine Konvergenz des geschätzten Integrationsgrades bzw. der geschätzten optimalen Bandbreite feststellen können. Dieses Verfahren wollen wir nun zur Illustration an einem ARFIMA(1,  $d$ , 0)-Prozess mit  $\theta = 0,5$  und  $n = 500$  in der Tabelle 3.19 darstellen.  $\hat{\theta}_i$  entspricht dem geschätzten AR(1)-Koeffizienten des filtrierte Prozesses  $\Delta^{\hat{d}_i} y_t$  beim  $i$ -ten Schritt. Aus diesem können wir durch Einsetzen in (3.66) den Parameter  $K$  schätzen und wiederum eine neue Bandbreite für den  $(i + 1)$ -ten Schritt gewinnen. Bei diesem Beispiel ist eine Konvergenz bereits nach vier Schritten erreicht.

Der Nachteil des Verfahrens ist dann aber, dass wir eine Annahme über die Parametrisierung des Prozesses (z.B. AR(1) oder ARMA(1,1)) machen müssen, um dann eine semiparametrische Schätzung durchzuführen, was eigentlich den Charakter und die Vorzüge von semiparametrischen Verfahren verfälscht. Nichtsdestotrotz kann dieses Verfahren einen guten Bezugspunkt für das Plugin-Verfahren darstellen,

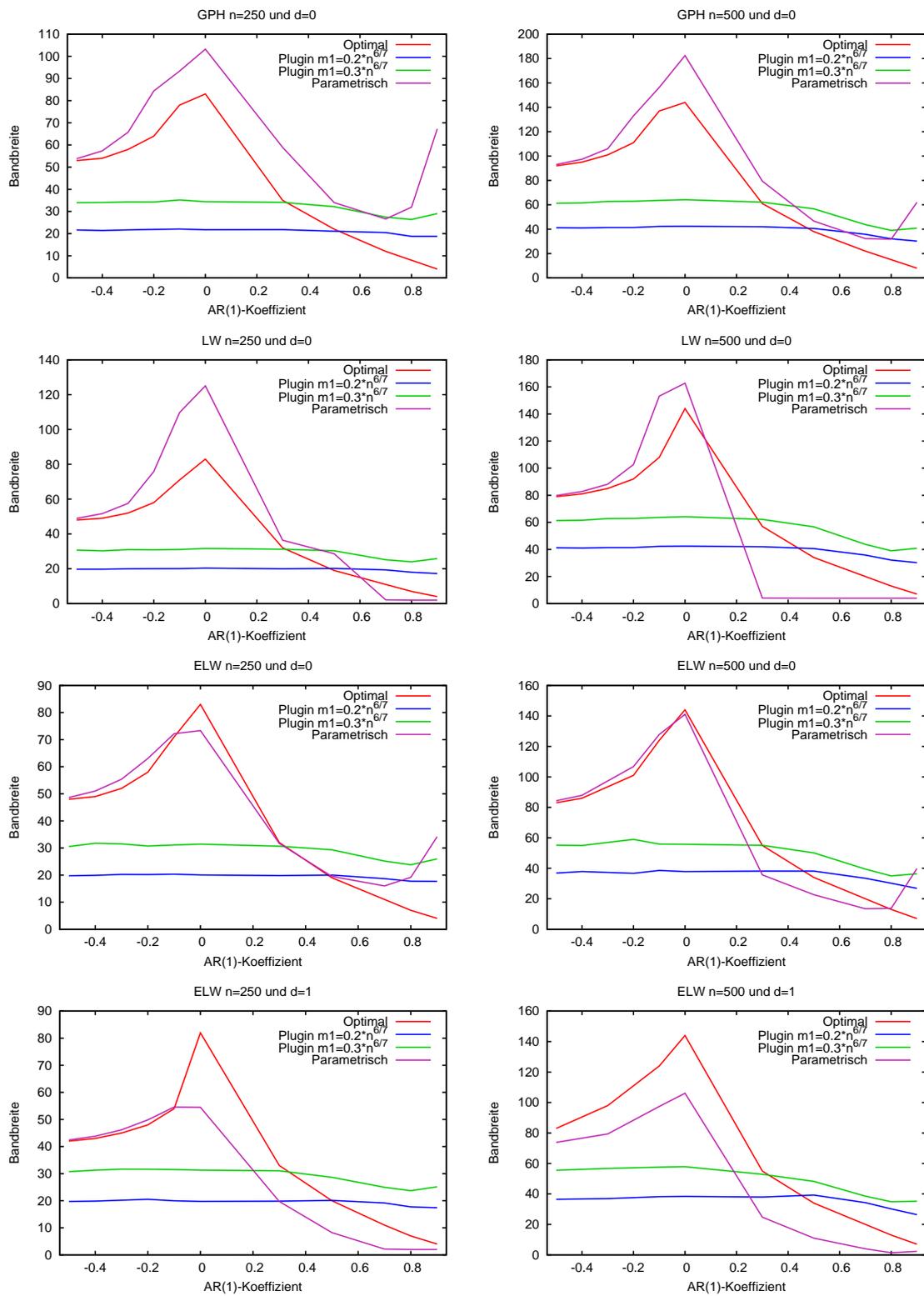
Tab. 3.19: Beispiel der parametrischen Bandbreitenschätzung für das ELW-Verfahren bei einem ARFIMA(1, 0, 0)-Prozess mit  $\theta = 0,5$  und  $n = 500$

Schritt $i$	$m_i$	$\hat{d}_i$ (asy. SD)	$\hat{\theta}_i$	$\hat{K}_i$	$m_{i+1}$
1	144	0,2959 (0,0417)	0,2232	0,4733	68
2	68	0,1217 (0,0606)	0,3758	0,3225	46
3	46	0,0375 (0,0737)	0,4558	0,2676	38
4	38	0,0359 (0,0811)	0,4573	0,2666	38

an dem wir dieses Verfahren in den folgenden Simulationen bewerten können.

Nun betrachten wir das Verhalten der Bandbreitenschätzer für simulierte Prozesse. Zunächst konzentrieren wir uns auf ARFIMA(1,  $d$ , 0)-Prozesse, erzeugt nach (3.68), die wir in Abhängigkeit von  $\theta$  jeweils 10.000-mal simulieren und die jeweiligen Bandbreiten bestimmen, mit denen wir den Integrationsgrad schätzen. In Abbildung 3.15 sind für die jeweiligen Verfahren die durchschnittlichen geschätzten Bandbreiten dargestellt. Als Beobachtungsumfang wählen wir hierbei zunächst  $n = 250$  und  $n = 500$ . Für das GPH- und das LW-Verfahren beschränken wir unsere Betrachtung auf den Fall  $d = 0$ , während wir für das ELW-Verfahren auch  $d = 1$  berücksichtigen. Dabei können wir festhalten, dass die Plugin-Verfahren die Bandbreite für  $\theta < 0,5$  unter- und ab  $\theta \geq 0,5$  überschätzen. Sehr auffällig ist hierbei, dass nahezu eine konstante Bandbreite geschätzt wird, wobei das Plugin-Verfahren mit der höheren Bandbreite  $m_1$  eine noch höhere Variabilität aufweist als mit niedrigem  $m_1$ . Dies führt dazu, dass dieses prinzipiell mit besseren Schätzeigenschaften abschneidet, außer für den Fall sehr starker Autokorrelation. Aus diesem Grund stellen wir in Abbildung 3.16 nur die Verzerrung und RMSE für das Plugin-Verfahren mit  $m_1 = 0,3n^{6/7}$  dar, um eine bessere Übersichtlichkeit zu gewährleisten.

Dieses schneidet nicht schlecht ab, ist aber nicht effizienter, als wenn man eine fixierte Bandbreite von  $m = n^{0,5}$  wählt. Dies gilt sowohl für das GPH- als auch für das LW- und ELW-Verfahren. Das parametrische Verfahren hingegen muss differenzierter betrachtet werden: Es trifft zwar für das GPH-Verfahren die optimale Bandbreite gut, hat aber eine relativ schlechte Effizienz im Vergleich zu den anderen Methoden, falls positive Autokorrelation vorliegt. Für das LW- und ELW-Verfahren wird bereits bei einer moderaten Autokorrelation von  $\theta = 0,3$  die Bandbreite massiv unterschätzt, was zu einer schwachen Effizienz führt. Die durchschnittliche Verzerrung hält sich aber in Grenzen außer beim ELW-Verfahren, wenn wir mit  $d = 1$  nichtstationäre Prozesse behandeln. Prinzipiell scheint eine konservativ fixierte Bandbreite gute Schätzeigenschaften für sämtliche Verfahren zu liefern, da sich hierbei die

Abb. 3.15: Geschätzte Bandbreiten für ARFIMA(1,  $d$ , 0)-Prozesse

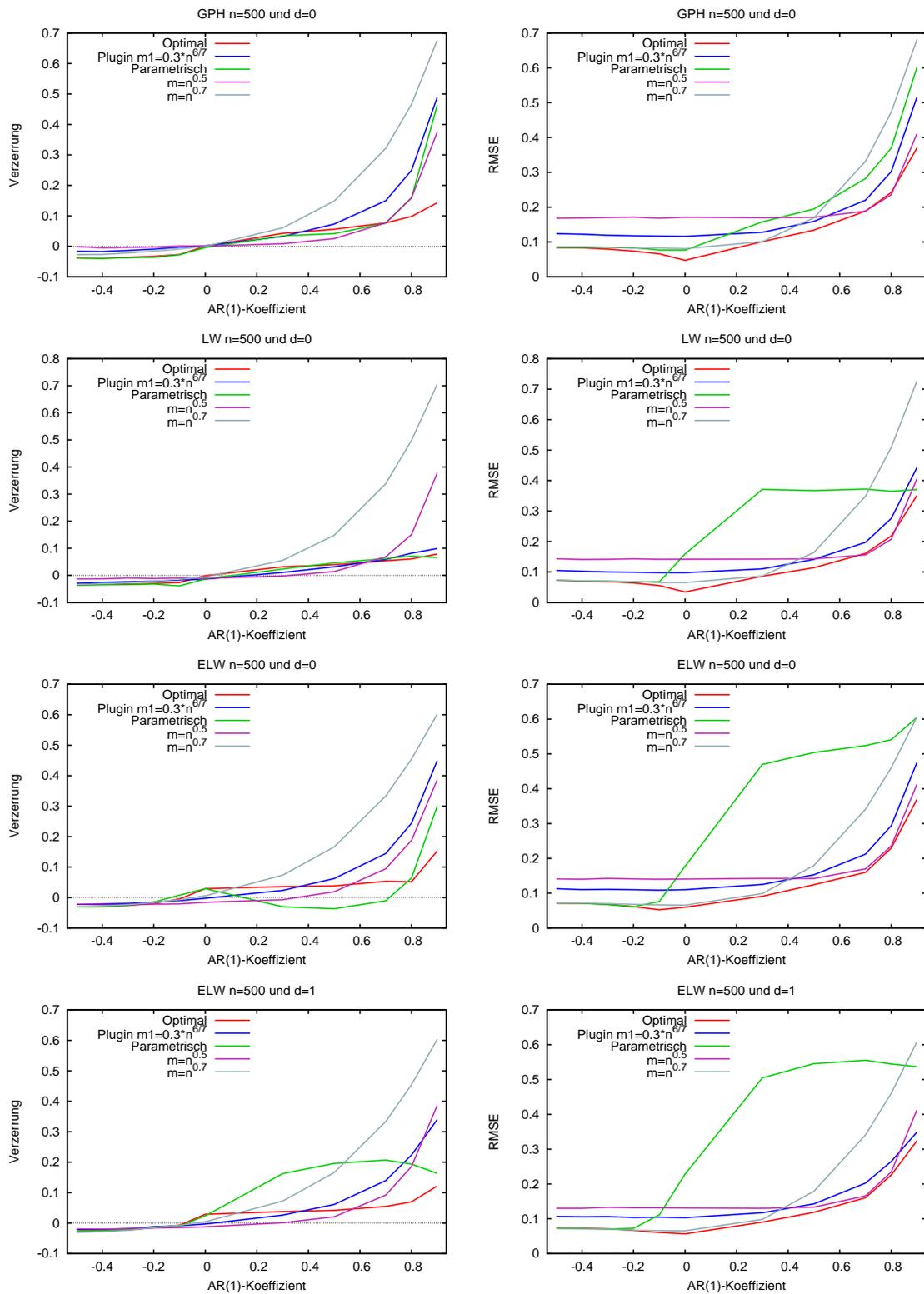


Abb. 3.16: Schätzeigenschaften von GPH, LW und ELW bei geschätzten Bandbreiten für ARFIMA(1,  $d$ , 0)-Prozesse

Verzerrung besonders im Falle starker Autokorrelation in Grenzen hält. Eine weniger konservative Schätzung mit  $m = n^{0,7}$  liefert zwar im Bereich negativer und niedriger positiver Autokorrelation gute Ergebnisse, führt aber zu schwerwiegenden Abweichungen, wenn tatsächlich starke Autokorrelation vorherrscht.

Um diesen Bereich der Arbeit abzuschließen, wollen wir nun auch MA-Bestandteile in den datengenerierenden Prozessen zulassen und ARFIMA(1,  $d$ , 1)-Prozesse der Form

$$\Delta^d(1 - \theta L)y_t = (1 - \phi L)\varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim N(0; 1), \quad (3.72)$$

$\theta \in \{0; 0,3; 0,5; 0,8\}$  und  $\phi \in \{-0,3; 0,4; 0,6\}$  simulieren. Wir betrachten nur positive Autokorrelationskoeffizienten, da diese Konstellationen insbesondere für eine empirische Anwendung relevant sind. Die optimale Bandbreite können wir mit Hilfe von (3.67) bestimmen und als Bandbreitenschätzer wenden wir auch wieder das Plugin-Verfahren mit zwei verschiedenen Bandbreiten und das parametrische Verfahren an. Als Beobachtungsumfang wählen wir wieder  $n = 500$  und führen wie bei den Simulationen zu ARFIMA(1,  $d$ , 0)-Prozessen 10.000 Wiederholungen für jede mögliche Konstellation von  $\phi$  und  $\theta$  durch.

Die Abbildungen 3.17 und 3.18 zeigen jeweils die Struktur der optimalen und geschätzten Bandbreiten an. Hierbei nehmen wir durchgängig den Integrationsgrad  $d = 0$ , da sich auch für das ELW-Verfahren bei  $d = 1$  das gleiche Bild ergibt und wir auf eine Darstellung verzichten können. Zur Beurteilung der Schätzeigenschaften berichten wir aber auch über die Ergebnisse für  $d = 1$ . Ähnlich wie im ARFIMA(1,  $d$ , 0)-Fall kann man sehen, dass die optimalen Bandbreiten für das GPH-Verfahren immer höher liegen als für das LW-Verfahren. Dennoch ist der LW-Schätzer stets effizienter bei einer ungefähr gleichen Verzerrung, wie man den Tabellen 3.20 und 3.21 entnehmen kann. Weitere Parallelen lassen sich auch für die Bandbreitenschätzer ziehen. Das Plugin-Verfahren mit  $m_1 = 0,2n^{6/7}$  weist auch bei ARFIMA(1,  $d$ , 1)-Prozessen nur eine geringe Variabilität auf, während mit höherer Bandbreite  $m_1 = 0,3n^{6/7}$  diese etwas stärker ausgeprägt ist. Auch kann man in den Tabellen 3.20 bis 3.23 ablesen, dass das Plugin-Verfahren mit höherem  $m_1$  meist effizienter ist als mit niedrigem  $m_1$ , außer bei  $\theta = 0,8$ . Das parametrische Schätzverfahren schneidet hingegen beim RMSE stets schlechter ab als beide Plugin-Verfahren, weist aber zum Teil niedrige durchschnittliche Verzerrungen auf, teilweise sogar noch niedriger als bei Verwendung der optimalen Bandbreite. Dies ergibt sich zum Teil aus der Struktur der geschätzten Bandbreiten für hohe Werte von  $\theta$ . Hier wird ähnlich wie im ARFIMA(1,  $d$ , 0)-Fall die Bandbreite stärker unterschätzt, was aber weniger

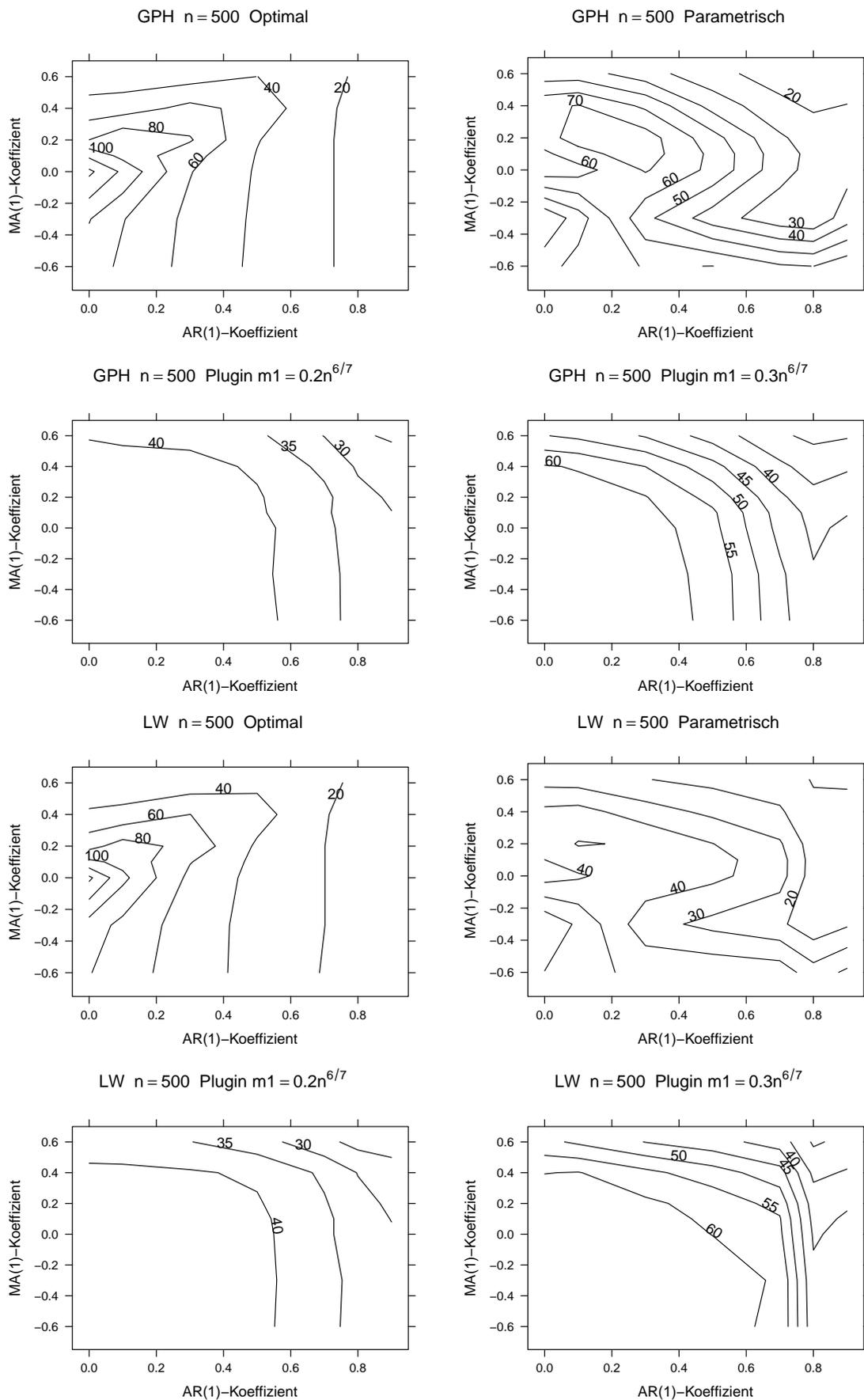


Abb. 3.17: Geschätzte Bandbreiten für ARFIMA(1,0,1)-Prozesse bei GPH und LW

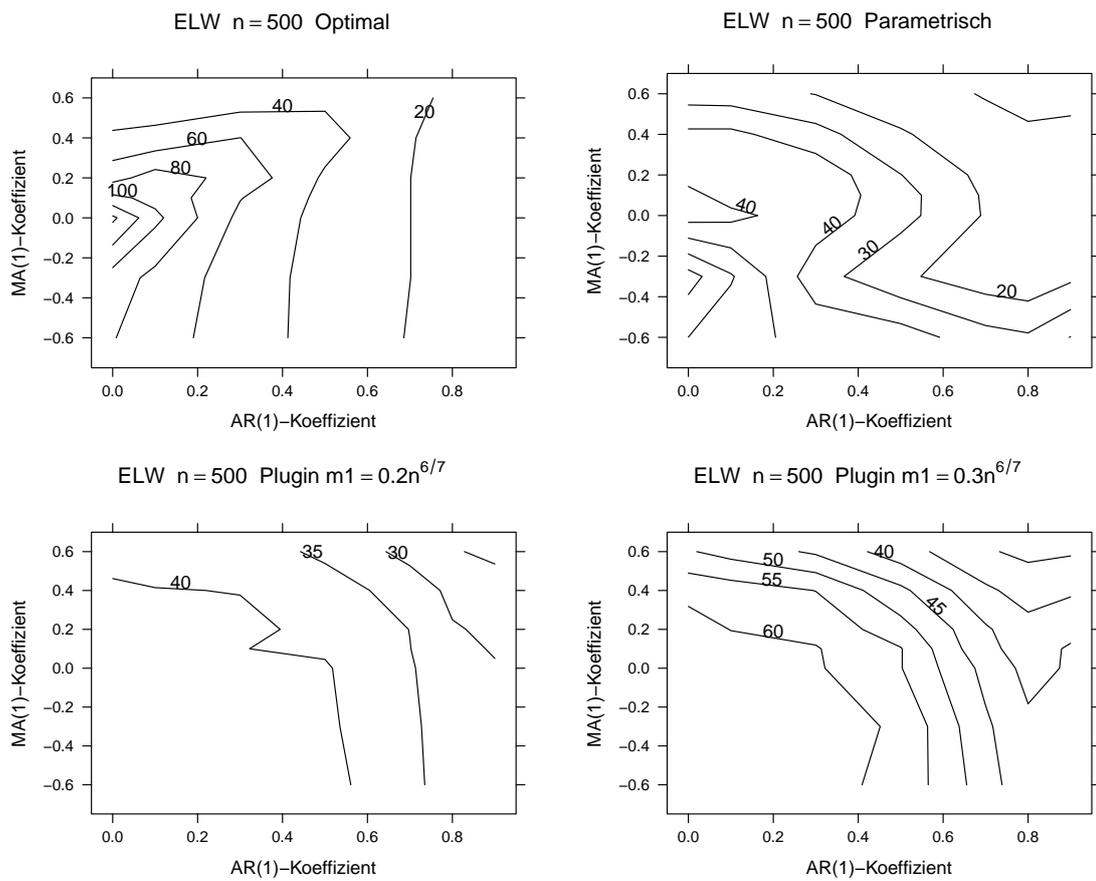


Abb. 3.18: Geschätzte Bandbreiten für ARFIMA(1, 0, 1)-Prozesse bei ELW

Tab. 3.20: Schätzeigenschaften für das GPH-Verfahren bei ARFIMA(1,0,1)-Prozessen mit  $n = 500$

	$\phi$	Verzerrung				RMSE			
		-0,3	0	0,4	0,6	-0,3	0	0,4	0,6
$\theta = 0$	Plugin 1	-0,0061	0,0010	0,0288	0,0711	0,1556	0,1521	0,1617	0,1890
	Plugin 2	-0,0125	0,0020	0,0486	0,0984	0,1215	0,1182	0,1432	0,1816
	$m = n^{0,5}$	-0,0019	0,0005	0,0115	0,0403	0,1720	0,1714	0,1700	<b>0,1746</b>
	$m = n^{0,7}$	-0,0225	0,0015	0,0956	0,2205	<b>0,0845</b>	<b>0,0809</b>	<b>0,1254</b>	0,2347
	Parametrisch	-0,0588	-0,0237	0,0379	0,0463	0,1316	0,1795	0,1742	0,2240
	Optimal	-0,0374	0,0005	0,0369	0,0614	0,0679	0,0472	0,1016	0,1455
$\theta = 0,3$	Plugin 1	0,0107	0,0191	0,0438	0,0825	0,1550	0,1564	0,1759	0,2086
	Plugin 2	0,0185	0,0318	0,0707	0,1077	0,1221	0,1296	<b>0,1689</b>	<b>0,1990</b>
	$m = n^{0,5}$	0,0047	0,0083	0,0188	0,0444	0,1721	0,1697	0,1719	0,1782
	$m = n^{0,7}$	0,0371	0,0601	0,1568	0,2797	<b>0,0894</b>	<b>0,1005</b>	0,1769	0,2912
	Parametrisch	-0,0167	0,0237	0,0548	0,0434	0,1842	0,1569	0,2131	0,2509
	Optimal	0,0231	0,0417	0,0612	0,0684	0,0810	0,0920	0,1533	0,1593
$\theta = 0,5$	Plugin 1	0,0398	0,0486	0,0694	0,1055	0,1648	0,1732	0,1978	0,2268
	Plugin 2	0,0623	0,0728	0,0918	0,1247	<b>0,1475</b>	<b>0,1610</b>	0,1905	0,2081
	$m = n^{0,5}$	0,0225	0,0244	0,0338	0,0658	0,1706	0,1721	<b>0,1735</b>	<b>0,1793</b>
	$m = n^{0,7}$	0,1254	0,1487	0,2453	0,3698	0,1494	0,1694	0,2585	0,3786
	Parametrisch	0,0022	0,0496	0,0361	0,0345	0,1982	0,1988	0,2407	0,2801
	Optimal	0,0285	0,0368	0,0546	0,0816	0,1327	0,1343	0,1577	0,1700
$\theta = 0,8$	Plugin 1	0,1827	0,1859	0,1904	0,2060	0,2702	0,2730	0,2858	0,2986
	Plugin 2	0,2455	0,2479	0,2484	0,2678	0,2959	0,2996	0,3042	0,3219
	$m = n^{0,5}$	0,1562	0,1580	0,1703	0,2030	<b>0,2327</b>	<b>0,2321</b>	<b>0,2411</b>	<b>0,2647</b>
	$m = n^{0,7}$	0,4430	0,4670	0,5621	0,6860	0,4503	0,4738	0,5678	0,6907
	Parametrisch	0,0873	0,1259	0,0954	0,0769	0,3447	0,3332	0,3772	0,4328
	Optimal	0,0914	0,0960	0,1029	0,1335	0,2237	0,2239	0,2341	0,2472

Plugin 1 steht für die Verwendung des Plugin-Verfahrens mit  $m_1 = 0,2n^{6/7}$  und Plugin 2 für  $m_1 = 0,3n^{6/7}$ .

effizient ist.

Prinzipiell lässt sich aber festhalten, dass eine konservativ niedrig gewählte fixierte Bandbreite zumeist eine bessere Schätzansatz darstellt als eine hohe Bandbreite, da das Ausmaß der Überschätzung im Fall hoher Autokorrelation wesentlich schwerwiegender ausfällt als die geringere Effizienz bei niedriger Autokorrelation. Da nun das Plugin-Verfahren durchweg RMSE-Werte liefert, die zwischen den fixierten Bandbreiten liegen, kann man daraus schließen, dass dieses Verfahren ein guter Schätzansatz bei fehlender Kenntnis der Autokorrelationsstruktur darstellt. Aus diesem Grund werden wir es für den empirischen Teil dieser Arbeit anwenden und versuchen, daraus Rückschlüsse für die Bandbreitenwahl zu treffen.

Tab. 3.21: Schätzeigenschaften für das LW-Verfahren bei ARFIMA(1,0,1)-Prozessen mit  $n = 500$ 

$\phi$	Verzerrung				RMSE				
	-0,3	0	0,4	0,6	-0,3	0	0,4	0,6	
$\theta = 0$	Plugin 1	-0,0174	-0,0105	0,0204	0,1118	0,1198	0,1191	0,1307	0,2002
	Plugin 2	-0,0215	-0,0072	0,0460	0,1555	0,0954	0,0917	0,1214	0,2127
	$m = n^{0,5}$	-0,0177	-0,0167	-0,0037	0,0648	0,1401	0,1421	0,1402	<b>0,1570</b>
	$m = n^{0,7}$	-0,0264	-0,0055	0,0939	0,3417	<b>0,0691</b>	<b>0,0645</b>	<b>0,1145</b>	0,3488
	Parametrisch	-0,0469	-0,0385	-0,0079	0,0043	0,1228	0,1879	0,1655	0,2607
	Optimal	-0,0341	-0,0027	0,0318	0,0524	0,0672	0,0448	0,0971	0,1411
$\theta = 0,3$	Plugin 1	0,0017	0,0105	0,0393	0,1239	0,1212	0,1250	0,1506	0,2199
	Plugin 2	0,0133	0,0291	0,0730	0,1561	0,0950	0,1059	0,1565	0,2190
	$m = n^{0,5}$	-0,0091	-0,0057	0,0031	0,0722	0,1401	0,1425	<b>0,1406</b>	<b>0,1590</b>
	$m = n^{0,7}$	0,0332	0,0553	0,1578	0,4111	<b>0,0723</b>	<b>0,0842</b>	0,1708	0,4174
	Parametrisch	-0,0389	-0,0118	-0,0081	-0,0045	0,2005	0,1568	0,1903	0,2986
	Optimal	0,0139	0,0285	0,1048	0,0595	0,0837	0,0826	0,1285	0,1384
$\theta = 0,5$	Plugin 1	0,0340	0,0421	0,0696	0,1350	0,1349	0,1435	0,1790	0,2389
	Plugin 2	0,0627	0,0722	0,0957	0,1644	<b>0,1276</b>	0,1429	0,1794	0,2242
	$m = n^{0,5}$	0,0072	0,0085	0,0207	0,0888	0,1402	<b>0,1395</b>	<b>0,1426</b>	<b>0,1683</b>
	$m = n^{0,7}$	0,1271	0,1495	0,2555	0,5165	0,1430	0,1632	0,2643	0,5219
	Parametrisch	-0,0343	-0,0059	-0,0219	-0,0172	0,2192	0,1868	0,2403	0,3721
	Optimal	0,0302	0,0380	0,1237	0,0888	0,1119	0,1122	0,1509	0,1428
$\theta = 0,8$	Plugin 1	0,1887	0,1933	0,1967	0,2235	0,2561	0,2655	0,2799	0,3024
	Plugin 2	0,2565	0,2579	0,2546	0,3002	0,2951	0,2993	0,3008	0,3430
	$m = n^{0,5}$	0,1471	0,1499	0,1606	0,2305	<b>0,2046</b>	<b>0,2073</b>	<b>0,2160</b>	<b>0,2730</b>
	$m = n^{0,7}$	0,4764	0,5035	0,6201	0,9040	0,4817	0,5088	0,6249	0,9080
	Parametrisch	0,0132	0,0392	0,0094	0,0136	0,3505	0,3265	0,3972	0,5420
	Optimal	0,0589	0,0596	0,0709	0,1386	0,2133	0,2153	0,2089	0,2163

Plugin 1 steht für die Verwendung des Plugin-Verfahrens mit  $m_1 = 0,2n^{6/7}$  und Plugin 2 für  $m_1 = 0,3n^{6/7}$ .

Tab. 3.22: Schätzeigenschaften für das ELW-Verfahren bei ARFIMA(1,0,1)-Prozessen mit  $n = 500$ 

$\phi$	Verzerrung				RMSE				
	-0,3	0	0,4	0,6	-0,3	0	0,4	0,6	
$\theta = 0$	Plugin 1	-0,0156	-0,0086	0,0277	0,0742	0,1204	0,1177	0,1375	0,1748
	Plugin 2	-0,0169	-0,0001	0,0580	0,1143	0,0942	0,0939	0,1367	0,1882
	$m = n^{0,5}$	-0,0183	-0,0174	0,0012	0,0263	0,1425	0,1408	0,1441	<b>0,1438</b>
	$m = n^{0,7}$	-0,0181	0,0029	0,1048	0,2423	<b>0,0674</b>	<b>0,0652</b>	<b>0,1230</b>	0,2519
	Parametrisch	-0,0398	-0,0350	-0,0009	-0,0003	0,1145	0,1806	0,1729	0,2139
	Optimal	-0,0223	0,0271	0,0376	0,0457	0,0631	0,0541	0,0984	0,1316
$\theta = 0,3$	Plugin 1	0,0054	0,0145	0,0492	0,0938	0,1235	0,1287	0,1603	0,2040
	Plugin 2	0,0226	0,0365	0,0835	0,1231	0,1001	0,1169	0,1768	0,2090
	$m = n^{0,5}$	-0,0118	-0,0069	0,0071	0,0356	0,1413	0,1416	<b>0,1424</b>	<b>0,1450</b>
	$m = n^{0,7}$	0,0446	0,0649	0,1701	0,3095	<b>0,0788</b>	<b>0,0911</b>	0,1826	0,3173
	Parametrisch	-0,0352	-0,0082	-0,0020	-0,0093	0,2001	0,1580	0,1903	0,2468
	Optimal	0,0186	0,0332	0,1132	0,0667	0,0854	0,0836	0,1363	0,1354
$\theta = 0,5$	Plugin 1	0,0403	0,0439	0,0791	0,1086	0,1426	0,1538	0,1943	0,2335
	Plugin 2	0,0714	0,0782	0,1072	0,1294	<b>0,1390</b>	0,1566	0,1951	0,2111
	$m = n^{0,5}$	0,0084	0,0043	0,0255	0,0497	0,1417	<b>0,1440</b>	<b>0,1429</b>	<b>0,1490</b>
	$m = n^{0,7}$	0,1368	0,1585	0,2693	0,4145	0,1524	0,1723	0,2781	0,4211
	Parametrisch	-0,0343	-0,0078	-0,0160	-0,0267	0,2200	0,1944	0,2327	0,2983
	Optimal	0,0329	0,0356	0,1292	0,1342	0,1149	0,1124	0,1562	0,1694
$\theta = 0,8$	Plugin 1	0,1934	0,1976	0,2025	0,2093	0,2624	0,2707	0,2851	0,2981
	Plugin 2	0,2675	0,2648	0,2592	0,2730	0,3084	0,3072	0,3037	0,3189
	$m = n^{0,5}$	0,1514	0,1516	0,1648	0,1939	<b>0,2088</b>	<b>0,2077</b>	<b>0,2183</b>	<b>0,2426</b>
	$m = n^{0,7}$	0,4964	0,5222	0,6427	0,8036	0,5018	0,5276	0,6473	0,8079
	Parametrisch	0,0058	0,0386	0,0054	-0,0131	0,3474	0,3272	0,3800	0,4530
	Optimal	0,0599	0,0633	0,0760	0,1013	0,2192	0,2208	0,2119	0,2152

Plugin 1 steht für die Verwendung des Plugin-Verfahrens mit  $m_1 = 0,2n^{6/7}$  und Plugin 2 für  $m_1 = 0,3n^{6/7}$ .

Tab. 3.23: Schätzeigenschaften für das ELW-Verfahren bei ARFIMA(1,1,1)-Prozessen mit  $n = 500$

$\phi$	Verzerrung				RMSE				
	-0,3	0	0,4	0,6	-0,3	0	0,4	0,6	
$\theta = 0$	Plugin 1	-0,0188	-0,0061	0,0338	0,0842	0,1239	0,1143	0,1417	0,1791
	Plugin 2	-0,0203	0,0017	0,0589	0,1202	0,0965	0,0889	0,1420	0,1865
	$m = n^{0,5}$	-0,0232	-0,0188	0,0081	0,0317	0,1409	0,1397	0,1499	<b>0,1445</b>
	$m = n^{0,7}$	-0,0211	0,0045	0,1042	0,2419	<b>0,0679</b>	<b>0,0633</b>	<b>0,1220</b>	0,2515
	Parametrisch	-0,0401	-0,0279	0,0055	0,0068	0,1027	0,1839	0,1491	0,2207
	Optimal	-0,0181	0,0279	0,0205	0,0253	0,0711	0,0531	0,1141	0,1322
$\theta = 0,3$	Plugin 1	0,0100	0,0158	0,0421	0,1005	0,1134	0,1256	0,1651	0,2168
	Plugin 2	0,0170	0,0337	0,0879	0,1220	0,0937	0,1139	0,1942	0,2043
	$m = n^{0,5}$	-0,0078	-0,0083	-0,0027	0,0348	0,1305	0,1429	<b>0,1449</b>	<b>0,1434</b>
	$m = n^{0,7}$	0,0402	0,0670	0,1686	0,3052	<b>0,0778</b>	<b>0,0924</b>	0,1813	0,3138
	Parametrisch	-0,0275	-0,0009	-0,0058	0,0081	0,1694	0,1472	0,1721	0,1853
	Optimal	-0,0006	0,0189	0,0627	0,0348	0,0749	0,0878	0,1158	0,1234
$\theta = 0,5$	Plugin 1	0,0463	0,0469	0,0709	0,1070	0,1380	0,1539	0,1707	0,2226
	Plugin 2	0,0739	0,0808	0,1124	0,1275	<b>0,1314</b>	0,1546	0,2082	0,2066
	$m = n^{0,5}$	0,0099	0,0047	0,0232	0,0428	0,1328	<b>0,1416</b>	<b>0,1430</b>	<b>0,1410</b>
	$m = n^{0,7}$	0,1348	0,1601	0,2667	0,4168	0,1505	0,1735	0,2744	0,4233
	Parametrisch	-0,0207	-0,0026	-0,0095	-0,0217	0,1836	0,1858	0,1793	0,2600
	Optimal	0,0177	0,0156	0,0769	0,0783	0,1234	0,1289	0,1238	0,1348
$\theta = 0,8$	Plugin 1	0,1839	0,1966	0,2155	0,2193	0,2494	0,2668	0,3068	0,2906
	Plugin 2	0,2610	0,2626	0,2588	0,2765	0,3021	0,3052	0,3089	0,3149
	$m = n^{0,5}$	0,1349	0,1524	0,1667	0,1992	<b>0,1916</b>	<b>0,2116</b>	<b>0,2245</b>	<b>0,2438</b>
	$m = n^{0,7}$	0,4924	0,5214	0,6407	0,8008	0,4969	0,5264	0,6457	0,8052
	Parametrisch	-0,0051	0,0404	-0,0065	0,0137	0,3120	0,3097	0,3968	0,4579
	Optimal	0,0245	0,0313	0,0584	0,0706	0,2241	0,2394	0,2341	0,2324

Plugin 1 steht für die Verwendung des Plugin-Verfahrens mit  $m_1 = 0,2n^{6/7}$  und Plugin 2 für  $m_1 = 0,3n^{6/7}$ .

## 4 Spezifikationstests für fraktional integrierte Prozesse

In diesem Kapitel der Arbeit wollen wir auf verschiedene Spezifikationstests unter der Prämisse, dass wir einen fraktional integrierten Prozess vorliegen haben, eingehen. Als erstes wollen wir Testmöglichkeiten bezüglich des Vorliegens von deterministischen Trends untersuchen und zwar insbesondere für die Situation, wenn wir nicht-stationäre Prozesse betrachten. Danach werden wir noch auf übliche Integrationsstests eingehen und deren Erweiterungen für fraktionale Integration untersuchen.

### 4.1 Tests für deterministische Trends

In diesem Abschnitt der Arbeit interessieren wir uns für mögliche deterministische Bestandteile eines beobachteten Prozesses. Dabei nehmen wir zunächst an, dass ein linearer Trend in  $y_t$  auftauchen kann in der Form

$$y_t = \mu_0 + \mu_1 t + \Delta^{-d} \varepsilon_t \quad \text{für } t = 1, \dots, n. \quad (4.1)$$

Die Kenntnis bzw. die Berücksichtigung solcher Trends ist nicht unerheblich bei der Schätzung des Integrationsgrades, da sonst deren Schätzung verzerrt sein kann. Um das Problem kurz zu illustrieren, führen wir wiederum eine Simulationsstudie durch. Wir erzeugen Prozesse integriert vom Grad  $d = 1$  und variieren den Beobachtungsumfang  $n$  und die Steigung des linearen Trends  $\mu_1$ . Danach schätzen wir mit dem ELW-Schätzer den Integrationsgrad mit der Bandbreite  $m = \lfloor n^{0.65} \rfloor$  für das Niveau der simulierten Zeitreihe und mit dem GPH-Schätzer für die erste Differenz mit der jeweils gleichen Bandbreite. Damit haben wir zum einen Schätzergebnisse für einen Prozess mit linearem Trend und beim Fall der ersten Differenz für einen Prozess mit einem Mittelwert von  $\mu_1$ . Die jeweils ermittelten durchschnittlichen Verzerrungen von  $\hat{d}$  wurden auf Basis von 50.000 Wiederholungen ermittelt und sind in den Graphiken der Abbildung 4.1 dargestellt.

Eine Schätzung für einen Prozess mit Mittelwert ist für den GPH-Schätzer tendenziell unproblematisch, da diese Schätzmethode auf dem Periodogramm aufbaut. Bei dessen Berechnung als Fouriertransformation der Autokovarianzfunktion erfolgt bereits eine Mittelwertzentrierung und insofern ist auch keine Verzerrung des Schätzergebnisses  $\hat{d}$  zu erwarten. Das Vorliegen eines linearen Trends hat hingegen bei Verwendung des ELW-Schätzers eine Auswirkung auf die Schätzung von  $d$ , die aber mit steigendem Beobachtungsumfang  $n$  abnimmt. Nichtsdestotrotz sollte besonders

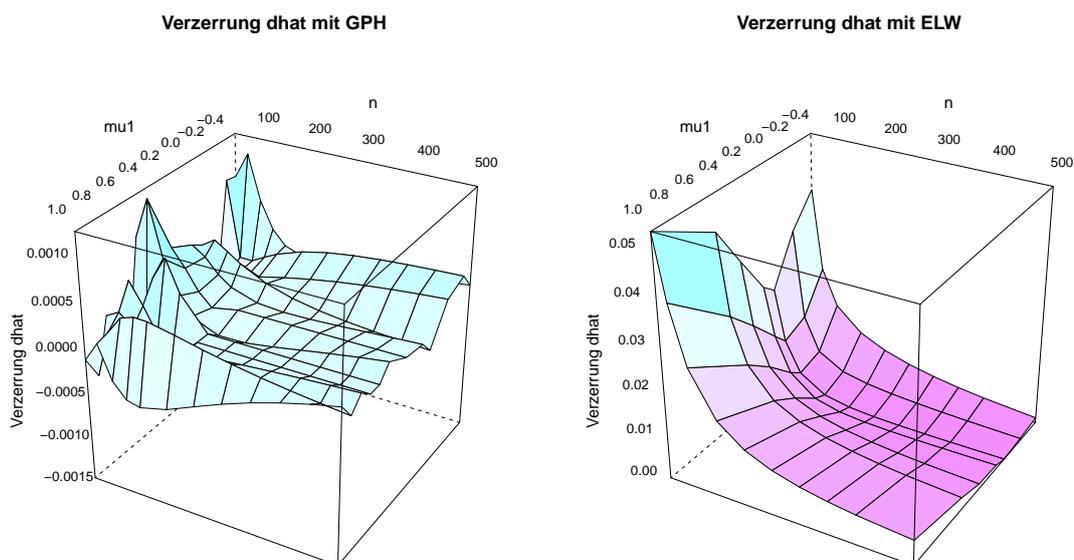


Abb. 4.1: Durchschnittliche Verzerrung bei der Schätzung von  $d$  bei Vorliegen eines linearen Trends

bei kleinen Stichproben auf Vorliegen eines deterministischen Trends geprüft werden und die zu untersuchende Zeitreihe vor Schätzung des Integrationsgrads  $d$  um diesen bereinigt werden.

Weiterhin wollen wir uns auch mit möglichen Strukturbrüchen von linearen Trends beschäftigen. Dazu erwägen wir das von Perron (1989) vorgeschlagene Changing Growth Model:

$$y_t = \mu_0 + \mu_1 t + \mu_2 DT_t(T_b) + \Delta^{-d} \varepsilon_t \quad \text{für } t = 1, \dots, n \quad (4.2)$$

$$\text{mit } DT_t(T_b) = \begin{cases} t - T_b & \text{wenn } t > T_b \\ 0 & \text{sonst.} \end{cases}$$

$T_b$  stellt damit im Endeffekt den Zeitpunkt dar, wann ein weiterer linearer Trend in der Prozessgleichung auftritt. Damit ändert sich in Kombination mit dem bereits existierenden Trend die deterministische Steigung des Prozesses ab dem Bruchzeitpunkt  $T_b$ , weshalb wir in diesem Kontext auch von gebrochenen Trends sprechen können. Wir können auch  $T_b = \lceil \tau n \rceil$  verallgemeinern, wobei dann der Parameter  $\tau$  der relative Bruchzeitpunkt darstellt. Diese Notation hilft uns dann, wenn wir den Beobachtungsumfang  $n$  für Simulationen variabel gestalten wollen.

Für empirische Anwendungen spielt gerade die Erwägung von Trendbruch-Modellen

eine nicht unerhebliche Rolle. Perron (1989) stellt z.B. fest, dass die Vernachlässigung von Trendbrüchen häufig zur falschen Ablehnung der Stationarität von Prozessen führen kann, wenn man übliche Integrationstests anwendet und nur einen linearen Trend annimmt. Bei fraktional integrierten Prozessen gibt es diesbezüglich auch Arbeiten, die sich mit den Auswirkungen der Vernachlässigung beschäftigen. Sibbertsen (2004) untersucht z.B. mehrere mögliche Ausprägungen von nicht-linearen deterministischen Prozessen und kommt dabei zu dem Schluss, dass die Anwendung von semiparametrischen Schätzverfahren verzerrte Schätzergebnisse bzgl. des Integrationsgrades liefern. Im Folgenden führen wir nun weitere Simulationen durch, um das Problem zu illustrieren. Wir variieren wieder den Beobachtungsumfang  $n$  und fixieren den Integrationsgrad  $d = 1$ . Weiterhin nehmen wir an, dass mit  $\mu_1 = 0$  zunächst kein linearer Trend vorliegt und bei  $\tau = 0,5$  sich ein Trendbruch mit der Stärke  $\mu_2$  eintritt. Daraufhin schätzen wir direkt den Integrationsgrad mit der ELW-Methode für das Niveau der simulierten Prozesse und mit der GPH-Methode für die erste Differenz. Betrachten wir nämlich die erste Differenz von (4.2), dann erhalten wir mit

$$\begin{aligned} \Delta y_t &= \mu_1 + \mu_2 DU_t(T_b) + \Delta^{1-d} \varepsilon_t \quad \text{für } t = 2, \dots, n \\ \text{mit } DU_t(T_b) &= \begin{cases} 1 & \text{wenn } t > T_b \\ 0 & \text{sonst.} \end{cases} \end{aligned} \quad (4.3)$$

einen Prozess mit einem Strukturbruch im Mittelwert zum Zeitpunkt  $T_b$ . Damit können wir über die durchschnittlichen Verzerrungen für Trendbruch-Modelle für ELW berichten und für Mittelwertbruch-Modelle, die sich aus der ersten Differenz von (4.2) ergeben und mit der GPH-Methode geschätzt werden. Die Simulationen werden dann 50.000-mal wiederholt und sind in der Abbildung 4.2 abgetragen. Für diese beiden Situationen ergeben sich relativ ähnliche Verzerrungen: Je stärker der Bruchparameter  $\mu_2$  ausfällt, desto stärker wird der Integrationsgrad überschätzt und das zum Teil relativ stark, wenn man die Situation mit dem Fall des linearen Trends vergleicht, bei dem die Überschätzung wesentlich kleiner ausfällt. Für die Mittelwertbruch-Modelle ergibt sich zumindest mit größerem Beobachtungsumfang eine kleinere Überschätzung, die aber asymptotisch scheinbar nicht verschwindet. Bei den Trendbruch-Modellen bleibt hingegen die Verzerrung relativ konstant.

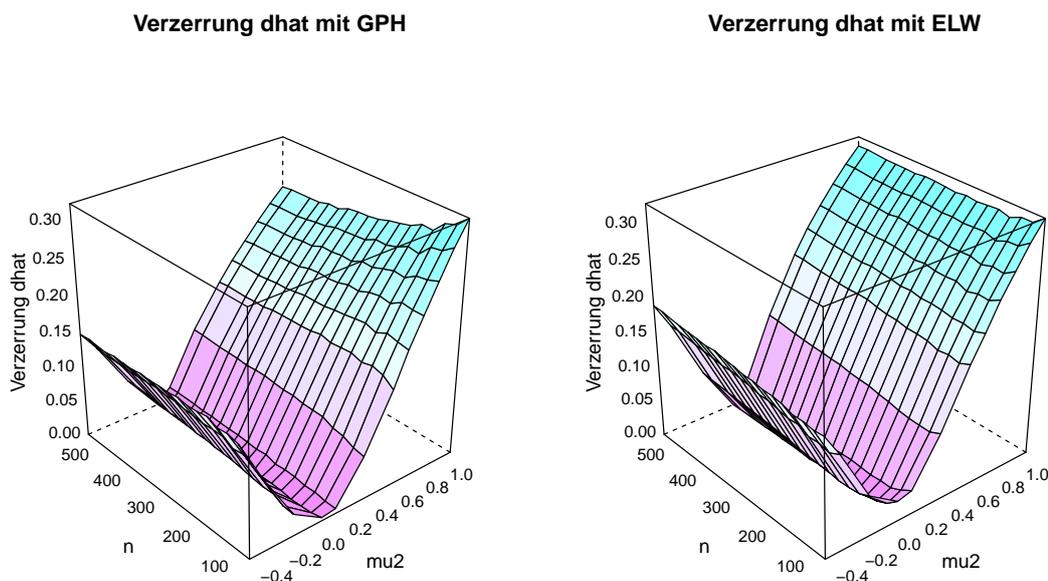


Abb. 4.2: Durchschnittliche Verzerrung bei der Schätzung von  $d$  bei Vorliegen eines gebrochenen Trends (ELW) bzw. gebrochenen Mittelwertes (GPH)

#### 4.1.1 Der Fall für lineare Trends

Nun wollen wir eine Testmöglichkeit für die Existenz von linearen Trends vorschlagen. Wir nehmen dabei an, dass wir einen Prozess der Form (4.1) beobachten und dass dieser einen nicht-stationären Integrationsgrad  $0,5 \leq d \leq 1,5$  aufweist. Für den Fall  $d = 1$  haben Gómez und Ventosa-Santaulària (2011, weiterhin: GVS) einen relativ einfachen Test der Hypothese  $H_0 : \mu_1 = 0$  vorgeschlagen. Dieser besteht darin, eine einfache KQ-Regression des Niveaus der Zeitreihe auf eine Konstante und einem linearen Trend zu regressieren, also

$$\text{Regression}_{LT} : \quad y_t = \mu_0 + \mu_1 t + \xi_t \quad (4.4)$$

mit  $\xi_t = \Delta^{-d} \varepsilon_t$  als Störterm zu schätzen. Für diese Regression lässt sich nun das unkorrigierte Bestimmtheitsmaß

$$R_{LT}^2 = 1 - \frac{\sum_{t=1}^n \hat{\xi}_t^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (4.5)$$

bestimmen, welches zur Entscheidung über die Existenz eines linearen Trends, d.h.  $\mu_1 \neq 0$ , herangezogen werden kann. GVS (2011) konnten für den Fall  $d = 1$  zeigen, dass, falls kein linearer Trend in  $y_t$  vorliegt (d.h.  $\mu_1 = 0$ ),  $R^2$  gegen eine nicht-

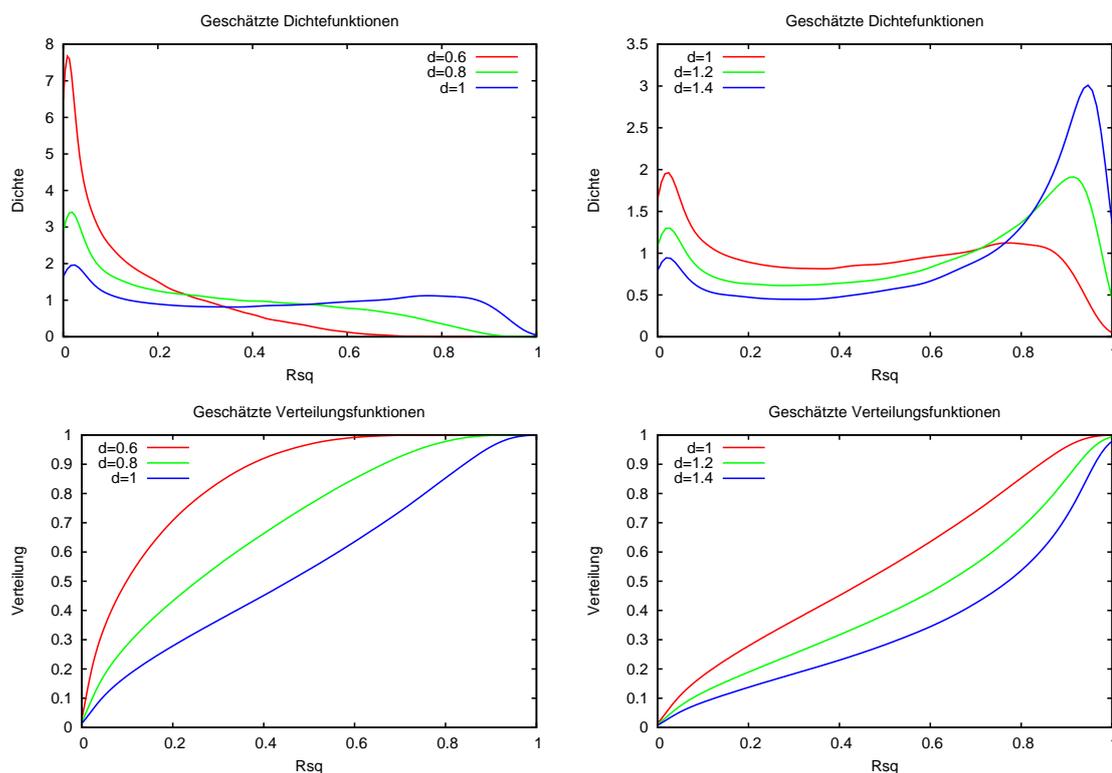


Abb. 4.3: Geschätzte Dichte- und Verteilungsfunktionen für die lineare Teststatistik  $R_{LT}^2$

degenerierte Verteilung konvergiert, die aber keine Standard-Verteilung darstellt. Für den Fall  $\mu_1 \neq 0$  geht das  $R^2$  hingegen asymptotisch gegen 1. Die Verteilung von  $R_{LT}^2$  kann durch Simulation ermittelt werden, was auch GVS (2011) für den Fall  $d = 1$  gemacht haben.

In Abbildung 4.3 stellen wir die geschätzten Dichte- und auch die zugehörigen Verteilungsfunktionen dar, wenn der wahre Prozess  $y_t$  integriert vom Grade  $d \in \{0,6; 0,8; 1; 1,2; 1,4\}$  ist mit einem Beobachtungsumfang  $n = 1000$  und unter Annahme von  $\mu_1 = 0$ . Die jeweiligen Experimente wurden insgesamt 100.000-mal wiederholt. Man kann dabei erkennen, dass die Dichtefunktion keiner bekannten Zufallsvariablen ähnelt und besonders für die Randfälle  $d = 0,6$  und  $d = 1,4$  zwar extrem wird, aber nicht degeneriert. Zudem hängt die Verteilung vom Integrationsparameter ab.

In weiteren Simulation wollen wir nun kritische Werte für Irrtumswahrscheinlichkeiten  $\alpha$  der Testentscheidung, ob ein zu untersuchender Prozess einen linearen Trend enthält, ermitteln. Diese ergeben sich, als das  $(1 - \alpha)$ -Quantil der Verteilung von  $R_{LT}^2$ , simuliert unter der Bedingung  $\mu_1 = 0$ . Wir variieren dafür den Parameter  $d \in$

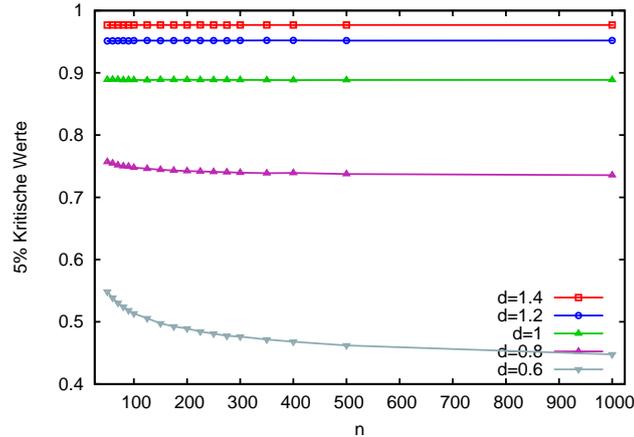


Abb. 4.4: Abhängigkeit der 5%-kritischen-Werte für  $R_{LT}^2$  und  $n$

$\{0,5; 0,52; \dots; 1,48; 1,5\}$  und den Beobachtungsumfang  $n \in \{50; 70; 80; 100; 125; 150; 175; 200; 250; 300; 350; 400; 500; 1000\}$  und wiederholen jede der 714 Parameterkonstellationen insgesamt 1.000.000-mal. In der Abbildung 4.4 kann man die Abhängigkeit der kritischen Werte für  $\alpha = 5\%$  vom Beobachtungsumfang erkennen. Für relativ niedrige Integrationsparameter ergibt sich eine relativ stark ausgeprägte Struktur, während für höhere Werte von  $d$  kaum eine Abhängigkeit erkennbar ist. Wir schlagen deswegen vor, asymptotische kritische Werte zu ermitteln, indem wir für jeden Parameter  $d$ , die ermittelten kritische Werte wie folgt regressieren:

$$R_{LT,(1-\alpha)}^2(d, n) = \gamma_{0,d} + \gamma_{1,d}n^{-1} + \gamma_{2,d}n^{-2} + \gamma_{3,d}n^{-3} + \gamma_{4,d}n^{-4} + \text{Resid} \quad \text{für } d \in [0,5; 1,5]. \quad (4.6)$$

Die Regression führen wir mit der einfachen KQ-Methode durch, obwohl eine Heteroskedastizität in den Residuen aufgrund der unterschiedlichen Beobachtungsumfänge zu erwarten ist. Die einfache KQ-Methode ist hier anwendbar, da wir hier lediglich eine gute Anpassung der kritischen Werte an  $n$  suchen und keine Inferenzaussagen treffen wollen. Das Vorliegen von Heteroskedastizität beeinflusst nicht die Erwartungstreue der jeweiligen geschätzten Koeffizienten. Die Parameter  $\gamma_{0,d}$  lassen sich dann als die asymptotischen kritischen Werte für  $n \rightarrow \infty$  interpretieren und werden in Abbildung 4.5 in Abhängigkeit von  $d$  dargestellt. Man kann dabei erkennen, dass die kritischen Werte eine Struktur in  $d$  aufweisen, die gut erklärbar durch ein Polynom sechsten Grades von  $d$  ist, welches wir auch in der Abbildung angepasst haben. Deswegen kommen wir zu dem Schluss, dass wir die kritischen Werte

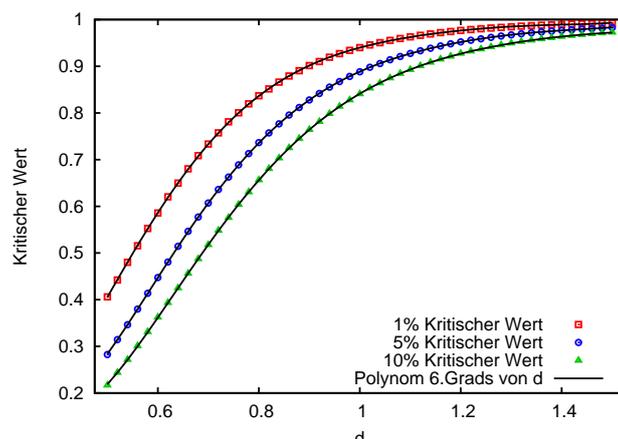


Abb. 4.5: Asymptotische kritische Werte für  $R^2_{LT}$  in Abhängigkeit von  $d$

Tab. 4.1: Response-Surface-Parameter für  $R^2_{LT,(1-\alpha)}$

$(1 - \alpha) \cdot 100\%$	90%	95%	97,5%	99%
$\beta_0$	3,3862	3,6546	3,2455	2,2023
$\beta_1$	-22,974	-26,523	-24,975	-18,851
$\beta_2$	59,715	74,363	74,174	61,206
$\beta_3$	-72,644	-98,346	-102,72	-89,138
$\beta_4$	46,780	69,091	75,119	67,538
$\beta_5$	-15,515	-25,057	-28,237	-26,094
$\beta_6$	2,0949	3,7065	4,3142	4,0764
$R^2$	99,9997%	99,9998%	99,9995%	99,9990%
$\hat{\sigma}_{\text{Resid}}$	$3,78 \cdot 10^{-4}$	$3,04 \cdot 10^{-4}$	$4,00 \cdot 10^{-4}$	$5,20 \cdot 10^{-4}$

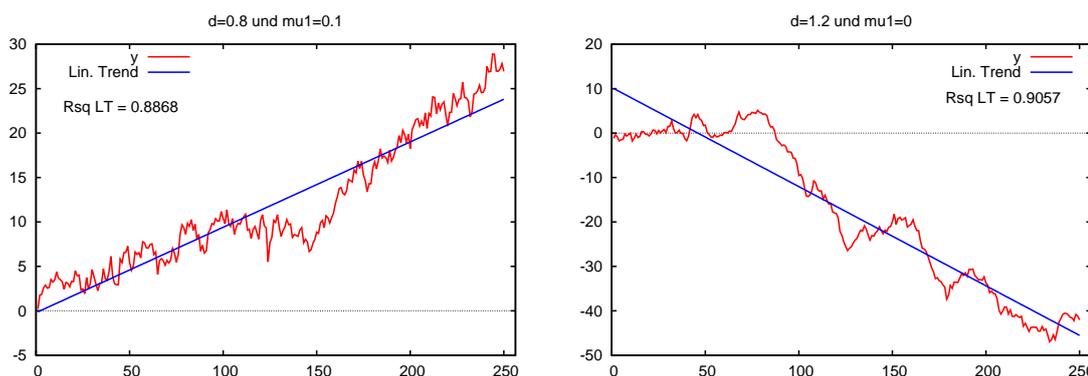
mit folgender Response-Surface-Funktion modellieren wollen:

$$R^2_{LT,(1-\alpha)}(d, \infty) = \beta_0 + \beta_1 d + \beta_2 d^2 + \beta_3 d^3 + \beta_4 d^4 + \beta_5 d^5 + \beta_6 d^6 + \text{Resid.} \quad (4.7)$$

Auch hier wählen wir als Schätzmethode die einfache KQ-Methode, da wir keine Inferenzaussagen treffen wollen, sondern nur eine möglichst gute Anpassung wie bei der Regression zur Extrapolation der kritischen Werte.

Die jeweils geschätzten Koeffizienten befinden sich in Tabelle 4.1. Wir geben hier auch jeweils die Güte der Anpassung  $R^2$  für die Regression (4.7) an, die durchweg auf einem sehr hohen Niveau liegt. Somit kann die Response-Surface-Funktion die asymptotischen kritischen Werte gut abbilden und wir können durch Einsetzen des Parameters  $d$  die kritischen Werte relativ einfach ermitteln.

Nun wollen wir das Vorgehen für den Test auf lineare Trends an zwei Beispielen demonstrieren. In der Abbildung 4.6 haben wir in der linken Graphik einen simu-

Abb. 4.6: Beispiel der Anwendung der linearen Trendstatistik  $R_{LT}^2$ Tab. 4.2: Kritische Werte für  $R_{LT}^2$ 

$(1 - \alpha) \cdot 100\%$	90%	95%	97,5%	99%
$d = 0,8$	0,6561	0,7357	0,7891	0,8365
$d = 1$	0,8412	0,8886	0,9169	0,9400
$d = 1,2$	0,9270	0,9519	0,9659	0,9766

lierten fractional integrierten Prozess mit  $d = 0,8$  und einem linearen Trend mit Steigung  $\mu_1 = 0,1$ . Hierfür ergibt sich bei Anpassung eines linearen Trends nun eine Teststatistik von  $R_{LT}^2 = 0,8868$ , was signifikant bei  $\alpha = 1\%$  für das Vorliegen eines Trends spricht, falls  $d$  bekannt ist. Nimmt man nun die kritischen Werte von GVS (2011), die für  $d = 1$  simuliert wurden, dann ergibt sich immerhin noch bei einem Signifikanzniveau von  $\alpha = 10\%$  die Annahme eines linearen Trends. Als zweiten Prozess beobachten wir im rechten Schaubild der Abbildung 4.6 nun einen fractional integrierten Prozess mit  $d = 1,2$ , der keinen linearen Trend enthält, also mit  $\mu_1 = 0$ . Für diesen sehen wir nun, dass, wenn wir die korrekten kritischen Werte anwenden, keinen Trend annehmen können. Bei Verwendung der kritischen Werte aus GVS (2011) ergibt sich aber bei  $\alpha = 5\%$  eine Ablehnung von  $H_0 : \mu_1 = 0$ . Besonders Prozesse mit hohem Parameter  $d$  weisen stark ausgeprägte stochastische Trends auf, die unter Umständen mit deterministischen Trends verwechselt werden können, wie der mögliche Realisationspfad des Prozesses mit  $d = 1,2$  in der Abbildung 4.6 zumindest andeutet.

Diese Erkenntnis deckt sich nun mit den Ergebnissen der Simulationsstudie, die in Abbildung 4.7 dargestellt ist. Hier nehmen wir nun an, dass der wahre Integrationsparameter unbekannt sei, was auch in den meisten empirischen Anwendungen der Fall sein wird. Demzufolge müssen wir nun den Parameter  $d$  schätzen, wozu sich besonders der bereits im vorherigen Kapitel vorgestellte ELW-Schätzer nach Shi-

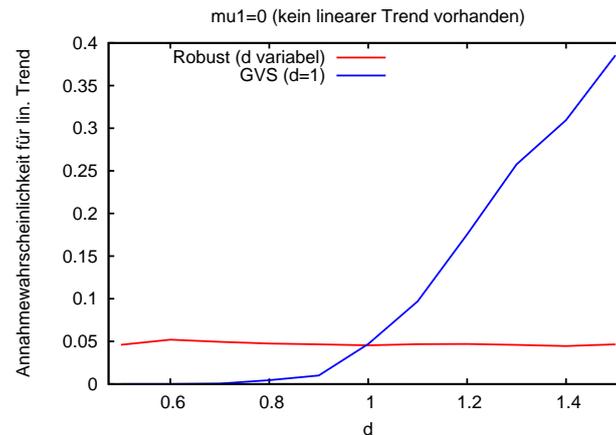


Abb. 4.7: Vergleich der Annahmewahrscheinlichkeiten für einen linearen Trend für  $\mu_1 = 0$  und  $n = 250$

motsu (2010) anbietet. Für die Studie simulieren wir nun Prozesse der Form (4.1) mit einem variablen Integrationsparameter  $d \in [0,5; 1,5]$  und  $n = 250$ . Wir nehmen  $\mu_1 = 0$  an, was bedeutet, dass wir keinen linearen Trend in unseren simulierten Prozessen haben. Für die Störterme verwenden wir eine einfache Standardnormalverteilung und konstruieren keine autokorrelierte Struktur, so dass wir einfache Weiße Rauschen haben. Daraufhin führen wir unsere Testregression (4.4) durch und erhalten unsere Teststatistik  $R_{LT}^2$ . Diese vergleichen wir nun mit den kritischen Werten nach GVS (2011), die vereinfachend  $d = 1$  annehmen, mit einer Irrtumswahrscheinlichkeit von  $\alpha = 5\%$ . Diese ergeben sich auch aus der Response-Surface-Funktion (4.52), wenn wir einfach  $d = 1$  einsetzen. Die Testentscheidung, die wir auf Basis dieser angenommenen Konstellation treffen können, vergleichen wir dann mit der Annahme, dass wir fraktional integrierte Prozesse möglicherweise beobachten. Dazu schätzen wir den Integrationsparameter mit ELW, wobei wir hier die Bandbreite  $m = \lfloor 250^{0,8} \rfloor = 82$  fixieren. Den geschätzten Integrationsparameter verwenden wir dann zur Ermittlung der kritischen Werte und zu unserer Testentscheidung. Die Abbildung 4.7 zeigt nun die relativen Ablehnungshäufigkeiten der Hypothese  $\mu_1 = 0$ , d.h. dass wir auf Basis von  $R_{LT}^2$  einen linearen Trend in den mit  $\mu_1 = 0$  simulierten Trend fälschlicherweise annehmen. Bei Verwendung der kritischen Werte nach GVS (2011) kommen wir zu größeren Verzerrungen bei der Testentscheidung, was auch nicht anders zu erwarten war. Insbesondere, wenn wir einen Prozess mit höherem Integrationsparameter  $d > 1$  beobachten, nimmt die Testentscheidung mit der Restriktion  $d = 1$  mit höherer Wahrscheinlichkeit fälschlicherweise einen linearen Trend an. Nehmen wir hingegen die robusten kritischen Werte, treffen wir zumeist

in 5% aller Fälle eine Fehlentscheidung und nehmen einen linearen Trend an, obwohl wir mit der Unsicherheit konfrontiert werden, dass wir den Integrationsparameter schätzen müssen.

#### 4.1.2 Strukturbruch eines linearen Trends

GVS (2011) beschreiben in ihrer Arbeit nicht nur Testmöglichkeiten für die Existenz eines linearen Trends, sondern beschäftigen sich auch mit Strukturbrüchen des Trends. Sie schlagen nun analog zum Fall des linearen Trends die folgende Testregression vor:

$$\text{Regression}_{TB} : y_t = \mu_0 + \mu_1 t + \mu_2 DT_t(T_b) + \xi_t, \quad (4.8)$$

wobei  $DT_t(T_b)$  bereits in (4.2) beschrieben wurde und  $\xi_t$  als Störterm der Regression angenommen wird und tendenziell nicht-stationär ist. Auch hier könnte man eine Testentscheidung bezüglich des Bestimmtheitsmaßes der Regression  $R_{TB}^2$  treffen, da dieses gegen 1 geht, falls  $\mu_2 \neq 0$  gilt, was ja wiederum bedeutet, dass tatsächlich ein linearer Trend ab dem Zeitpunkt  $T_b$  existiert. Problematisch ist aber nun, dass auch für die Konstellation  $\mu_1 \neq 0$  und  $\mu_2 = 0$  die Teststatistik gegen 1 geht, was GVS (2011) in einem relativ aufwendigen Beweis für  $d = 1$  zeigen konnten. Somit kann man, wenn man nur das Bestimmtheitsmaß  $R_{TB}^2$  als Testgrundlage zu Rate zieht, nicht zwischen der Existenz eines linearen Trends und eines gebrochenen Trends unterscheiden. Deswegen machen GVS (2011) den Vorschlag, als weitere Testgröße die t-Statistik für den geschätzten Parameter  $\hat{\mu}_2$  zu verwenden, da der geschätzte Parameter  $\hat{\mu}_2$  nämlich gegen  $\mu_2$  konvergiert. Damit diese gegen eine feste Verteilung strebt empfehlen GVS (2011), diese Statistik durch die Multiplikation mit  $n^{-1/2}$  zu stabilisieren. Die Abbildung 4.8 zeigt die Diskrepanz der t-Statistik, für den Fall, dass diese nicht normalisiert wird, und den normalisierten Fall. Für die Integrationsparameter  $d \in \{0,8; 1; 1,2\}$  haben wir einfache fraktionale Rauschen mit variablem Beobachtungsumfang  $n$  simuliert und stets angenommen, dass ein Trendbruch zum relativen Zeitpunkt  $\tau = 0,5$  vorliegt und die dementsprechende Testregression (4.8) durchgeführt. Aus den einzelnen Simulationsdurchgängen, die wir 100.000-mal wiederholt haben, haben wir dann die jeweiligen 5%-kritischen-Werte gewonnen. Dazu gibt es zu bedenken, dass der Hypothesentest für  $\hat{\mu}_2$  prinzipiell zweiseitig erfolgen

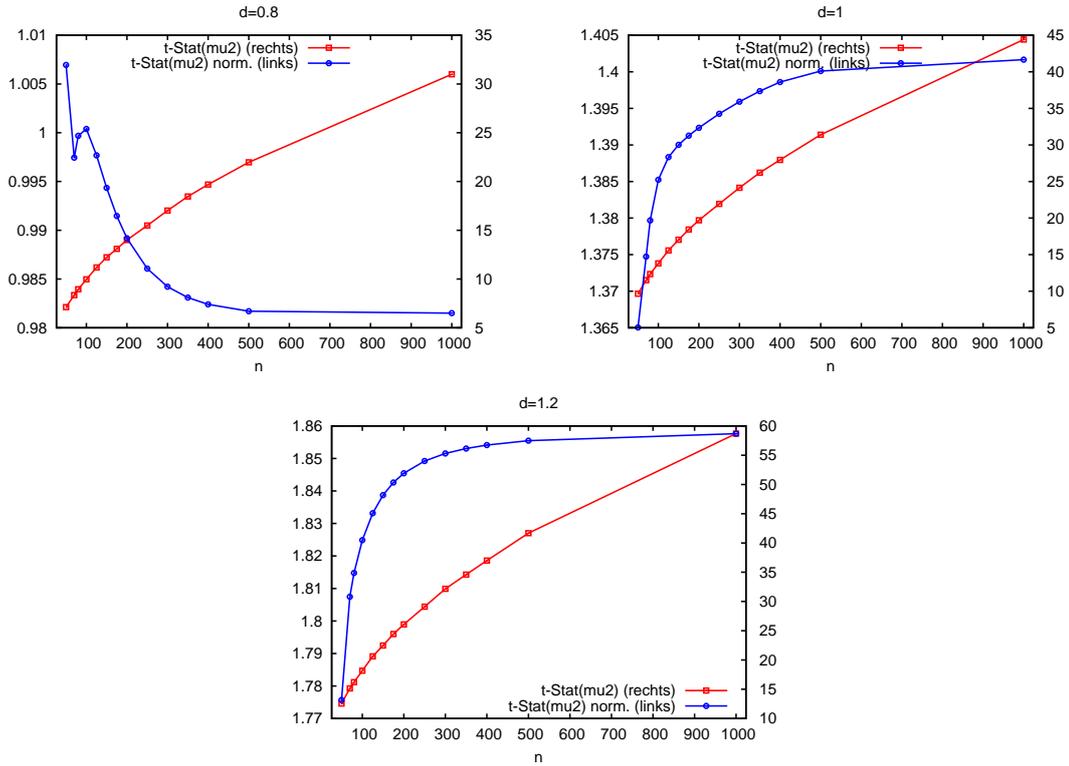


Abb. 4.8: Asymptotisches Verhalten der 5%-kritischen-Werte der t-Statistik für  $\mu_2$

muss. Deswegen definieren wir die normalisierte t-Statistik als

$$t\text{-Stat}(\hat{\mu}_2)_{\text{norm}} = \left| \frac{\hat{\mu}_2}{\sqrt{n\widehat{\text{Var}}(\hat{\mu}_2)}} \right| \quad (4.9)$$

und für die nicht-normalisierte t-Statistik analog  $t\text{-Stat}(\hat{\mu}_2) = \left| \hat{\mu}_2 / \sqrt{\widehat{\text{Var}}(\hat{\mu}_2)} \right|$ . In der Abbildung 4.8 kann man nun gut erkennen, dass für die nicht-normalisierte t-Statistik keine Konvergenz in  $n$  eintritt. Die normalisierte Statistik hat eine solche und somit können wir auch asymptotische kritische Werte für diese bestimmen. Wir fällen damit unsere Testentscheidung, dass ein Prozess einen gebrochenen Trend enthält, wenn gleichzeitig

$$R_{TB}^2 \geq R_{TB,(1-\alpha)}^2 \quad \text{und} \quad t\text{-Stat}(\hat{\mu}_2)_{\text{norm}} \geq t\text{-Stat}(\hat{\mu}_2)_{\text{norm},(1-\alpha)} \quad (4.10)$$

bei einer gegebenen Irrtumswahrscheinlichkeit  $\alpha$  erfüllt ist. In der empirischen Anwendung kann es teilweise vorkommen, dass der wahre Bruchzeitpunkt  $T_b$  unbekannt ist. Dieser kann mit diesem Ansatz auch geschätzt werden, indem die Summe der

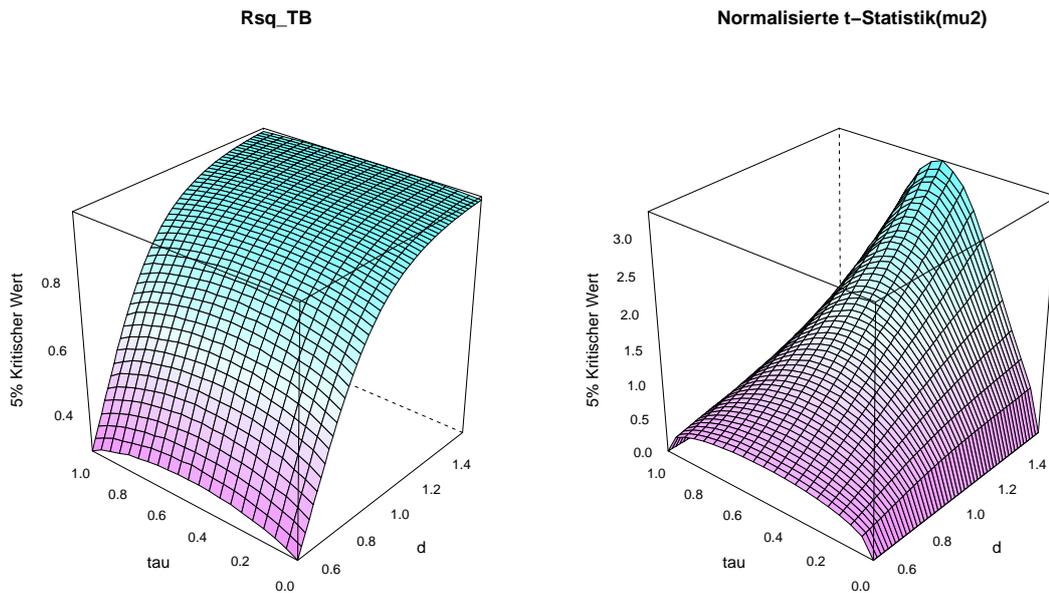


Abb. 4.9: Struktur der 5%-kritischen-Werte für  $R_{TB}^2$  und  $t\text{-Stat}(\hat{\mu}_2)_{\text{norm}}$

quadrierten Residuen (SSR, Sum of Squared Residuals) der Regression (4.8) über eine mögliche Auswahl an Bruchzeitpunkten minimiert wird:

$$\hat{T}_b = \min_{T_b \in [\lfloor \epsilon n \rfloor, \lfloor (1-\epsilon)n \rfloor]} SSR(T_b) \quad \text{mit } 0 < \epsilon < 1 \quad (4.11)$$

Zum gleichen Ergebnis  $\hat{T}_b$  könnte man auch kommen, wenn man die Statistik  $R_{TB}^2(T_b)$  in Abhängigkeit von  $T_b$  maximieren würde. Der Parameter  $\epsilon$  bedeutet in diesem Kontext, dass die ersten und letzten  $\lfloor \epsilon n \rfloor$ -Beobachtungen als Bruchzeitpunkte ausgeschlossen werden. Mögliche Ausprägungen wären z.B.  $\epsilon = 0,05$  oder  $\epsilon = 0,15$ , was bedeutet, dass insgesamt 10% bzw. 30% aller Beobachtungspunkte nicht als Bruchzeitpunkte in Frage kommen.

In der nächsten Studie wollen wir die Struktur der kritischen Werte in Abhängigkeit des relativen Bruchzeitpunkts  $\tau$  und des Integrationsparameters  $d$  betrachten. Hierzu fixieren wir  $n = 1000$  und variieren  $\tau \in \{0,05; 0,1; \dots; 0,9; 0,95\}$  und  $d \in \{0,5; 0,52; \dots; 1,48; 1,5\}$ . Die Abbildung 4.9 zeigt nun die simulierten 5%-kritischen-Werte. Dabei fällt vor allem auf, dass beide kritische Statistiken  $R_{TB}^2$  und  $t\text{-Stat}(\hat{\mu}_2)_{\text{norm}}$  symmetrisch um  $\tau = 0,5$  sind für alle Werte von  $d$  und dort ein Maximum aufweisen. Deswegen entschließen wir uns dazu, die kritischen Werte nur in Abhängigkeit von  $d$  zu modellieren, wie im Falle der linearen Teststatistik  $R_{LT}^2$  und  $\tau = 0,5$  zu fixieren. Dies senkt die Anzahl der zu berücksichtigenden Parameterkonstellationen erheblich

Tab. 4.3: Response-Surface-Parameter für  $R_{TB,(1-\alpha)}^2$ 

$(1 - \alpha) \cdot 100\%$	90%	95%	97,5%	99%
$\beta_0$	3,5074	2,9783	2,2420	1,3202
$\beta_1$	-25,623	-23,241	-19,033	-13,311
$\beta_2$	72,321	69,831	61,380	48,449
$\beta_3$	-95,936	-97,026	-88,894	-74,305
$\beta_4$	67,513	70,957	67,004	58,118
$\beta_5$	-24,514	-26,634	-25,762	-22,962
$\beta_6$	3,6299	4,0603	4,0066	3,6485
$R^2$	99,9998%	99,9996%	99,9993%	99,9986%
$\hat{\sigma}_{\text{Resid}}$	$2,51 \cdot 10^{-4}$	$3,75 \cdot 10^{-4}$	$4,40 \cdot 10^{-4}$	$5,36 \cdot 10^{-4}$

Tab. 4.4: Response-Surface-Parameter für  $t\text{-Stat}(\hat{\mu}_2)_{\text{norm},(1-\alpha)}$ 

$(1 - \alpha) \cdot 100\%$	90%	95%	97,5%	99%
$\beta_0$	0,8529	0,4412	0,6979	0,0941
$\beta_1$	-5,315	-2,4237	-3,9178	0,2040
$\beta_2$	15,560	7,8819	11,726	0,9560
$\beta_3$	-18,597	-7,2951	-11,698	3,5281
$\beta_4$	13,184	4,5128	7,4890	-3,7426
$\beta_5$	-5,1467	-1,7273	-2,7611	1,5288
$\beta_6$	0,8662	0,3213	0,4635	-0,2048
$R^2$	99,9995%	99,9993%	99,9992%	99,9977%
$\hat{\sigma}_{\text{Resid}}$	$1,50 \cdot 10^{-3}$	$2,22 \cdot 10^{-3}$	$2,88 \cdot 10^{-3}$	$5,91 \cdot 10^{-3}$

und wir können die kritischen Werte mit einer höheren Anzahl an Wiederholungen ermitteln. Dadurch verlieren wir an Macht, wenn der Bruchpunkt nicht bei  $\tau = 0,5$  liegt, aber die Unterschiede sind nur wirklich groß, wenn das Bruchdatum relativ früh oder spät im Beobachtungszeitraum liegt, wie wir in der Abbildung 4.9 erkennen können.

Das weitere Vorgehen läuft nun analog wie im Fall der linearen Teststatistik  $R_{LT}^2$ . Wir simulieren die einzelnen Parameterkonstellationen  $d$  mit  $\tau = 0,5$  für die gleichen Beobachtungsumfänge und verwenden diese dann zur Extrapolation  $n \rightarrow \infty$ , um asymptotische kritische Werte zu erhalten. Daraufhin erklären wir deren Struktur mit einem Polynom sechsten Grades von  $d$ . Die jeweils geschätzten Parameter für die Response-Surface-Funktionen befinden sich in den Tabellen 4.3 und 4.4 zusammen mit ihren Anpassungsstatistiken. Auch hier ergibt sich das gleiche Bild: Die Anpassung ist durchweg sehr gut und damit für die empirische Anwendung gut verwendbar.

Auch hier wollen wir eine beispielhafte Anwendung für simulierte Prozesse zeigen. Zum einen verwenden wir wieder einen fraktional integrierten Prozess mit

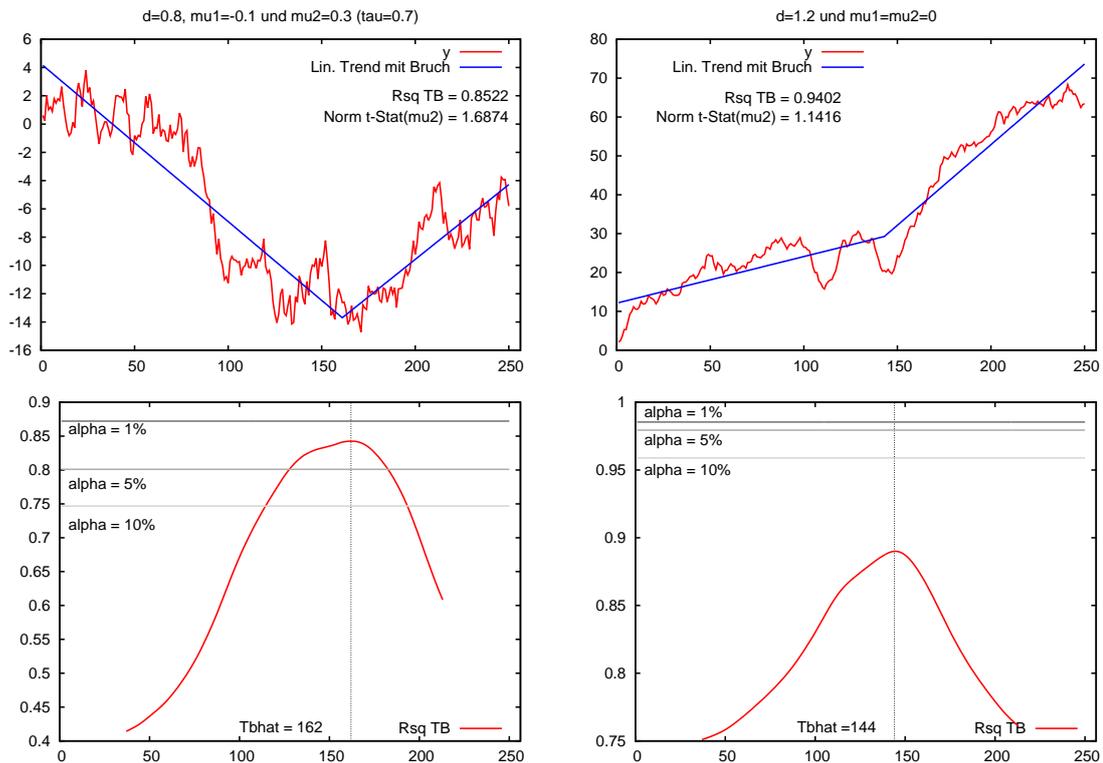


Abb. 4.10: Beispiel des Tests für einen gebrochenen Trend

$d = 0,8$  und  $n = 250$ , der einen gebrochenen linearen Trend enthält, wie in Gleichung (4.2) dargelegt. Dieser ist so spezifiziert, dass er einen linearen Trend mit Steigung  $\mu_1 = -0,1$  bis zum Zeitpunkt  $T_b = 0,7 \cdot 250 = 175$  enthält und ab dann ein weiterer Trend mit Steigung  $\mu_2 = 0,3$  eintritt. Der Bruchzeitpunkt wurde als unbekannt angenommen und über die Maximierung der Teststatistik  $R_{TB}^2$  bestimmt mit  $\epsilon = 0,15$ , was man in der Abbildung 4.10 sehen kann. Das Schätzergebnis  $\hat{T}_b = 162$  liegt zumindest relativ nahe bei dem wahren Datum  $T_b = 175$ , wobei aber ein Konfidenzintervall für den wahren Bruchpunkt auf Basis der Schätzergebnisse leider nicht angebar ist, da dieses u.a. von der Konstellation der Parameter  $d$  und  $\mu_2$  abhängen würde. Der Integrationsparameter  $d$  wird hingegen als bekannt angenommen und man kann somit die kritischen Werte in der Tabelle 4.5 entnehmen. Die kritischen Werte sind auch in der Maximierung von  $R_{TB}^2$  zur Ermittlung des geschätzten Bruchdatums angegeben. Man kann nun zu dem Ergebnis kommen, dass  $R_{TB}^2$  zumindest bei  $\alpha = 5\%$  signifikant abgelehnt werden kann.  $t\text{-Stat}(\hat{\mu}_2)_{\text{norm}}$  ist sogar bei  $\alpha = 1\%$  signifikant von 0 verschieden und man kann damit annehmen, dass der datengenerierende Prozess einen gebrochenen Trend enthält. Für einen weiteren Prozess, der integriert ist vom Grad  $d = 1,2$ , nehmen wir  $\mu_1 = \mu_2 = 0$  an, was be-

Tab. 4.5: Kritische Werte für  $R_{TB}^2$  und  $t\text{-Stat}(\hat{\mu}_2)_{\text{norm}}$ 

$(1 - \alpha)$	$R_{TB}^2$				$t\text{-Stat}(\hat{\mu}_2)_{\text{norm}}$			
	90%	95%	97,5%	99%	90%	95%	97,5%	99%
$d = 0,8$	0,7468	0,8010	0,8381	0,8721	0,9777	1,1782	1,3635	1,5899
$d = 1$	0,8976	0,9258	0,9432	0,9577	1,4033	1,7111	2,0000	2,3637
$d = 1,2$	0,9589	0,9718	0,9794	0,9854	1,8621	2,2958	2,7113	3,2443

deutet, dass keine deterministischen Bestandteile in der Prozessgleichung enthalten sind. Der Algorithmus findet einen Bruchpunkt bei  $\hat{T}_b = 144$ , wo man auch einen Strukturbruch bei grafischer Betrachtung des Realisationspfades vermuten könnten. Dabei sind aber sowohl  $R_{TB}^2$  als auch  $t\text{-Stat}(\hat{\mu}_2)_{\text{norm}}$  nicht signifikant bei einer üblichen Irrtumswahrscheinlichkeit, wenn wir  $d = 1,2$  berücksichtigen. Vernachlässigen wir aber dies und verwenden mit der Restriktion  $d = 1$  die kritischen Werte aus GVS (2011), dann ist zumindest die Statistik  $R_{TB}^2$  signifikant bei  $\alpha = 5\%$ . Die normalisierte t-Statistik des geschätzten Parameters  $\hat{\mu}_2$  ist aber nicht signifikant und so würden wir auch hier keinen gebrochenen Trend annehmen. Tendenziell kann man aber auch beim Testen für gebrochene Trends erwarten, dass die Vernachlässigung der Möglichkeit eines stark persistenten stochastischen Trends in eine gehäufte Annahme von deterministischen Trends führt, was wir bereits bei der Anwendung der linearen Teststatistik  $R_{LT}^2$  gesehen haben.

#### 4.1.3 Vergleich des Tests für Trendbruch mit dem Meanchange-Test nach Shao (2011)

In diesem Abschnitt wollen wir das Verhalten des voran vorgestellten Trendbruch-Tests studieren und diesen mit einer möglichen Alternative, die auch fraktionale Integration von Zeitreihen berücksichtigen kann, vergleichen. Wie wir ja bereits gesehen haben, entspricht die erste Differenz des Trendbruch-Modells einem Mittelwertbruch-Modell. Für dieses hat Shao (2011) einen Meanchange-Test entwickelt, der zumindest für stationäre fraktional integrierte Prozesse anwendbar ist. Er testet dabei in der Nullhypothese, dass die Realisationen  $x_t$  einen konstanten Erwartungswert haben bei der Möglichkeit, dass diese fraktional integriert sind, gegen die Alternative, dass es einen Strukturbruch im Erwartungswert gibt. Nehmen wir nun, dass wir den nicht-stationären Prozess  $y_t$  mit einem Integrationsgrad  $d \in [0,5; 1,5]$  für einen Trendbruch testen wollen, dann können wir auch analog den Prozess  $x_t = \Delta^1 y_t$  auf einen Mittelwertbruch testen, da dieser durch die Bildung der ersten Differenz nun einen Integrationsgrad von  $\tilde{d} = d - 1$  aufweist und im statio-

Tab. 4.6: Response-Surface-Parameter für  $MC(\tilde{d}, \epsilon = 0,15)_{(1-\alpha)}$ 

$(1 - \alpha)$	90%	95%	99%
Konstant	18,544	20,364	23,715
$\tilde{d} + 0,5$	17,558	15,410	43,846
$(\tilde{d} + 0,5)^2$	-42,602	40,754	-158,67
$(\tilde{d} + 0,5)^3$	274,75	-106,79	682,57
$(\tilde{d} + 0,5)^4$	-550,52	259,95	-1057,0
$(\tilde{d} + 0,5)^5$	481,42	-301,00	672,92
$(\tilde{d} + 0,5)^6$	-155,91	123,10	-135,63
$R^2$	99,91%	99,95%	99,98%
$\hat{\sigma}_{\text{Resid}}$	0,6439	0,2750	0,3238

nären Bereich liegt. Der Bruchzeitpunkt  $T_b$  muss dabei nicht bekannt sein und kann endogen bestimmt werden. Hierzu definiert Shao (2011) die Statistik

$$G(T_b) = \frac{\sqrt{n-1} |\bar{x}_{2,T_b} - \bar{x}_{T_b+1,n}|}{(n-1)^{-1} \left[ \sum_{t=2}^{T_b} S_t^2(2, T_b) + \sum_{t=T_b+1}^n S_t^2(T_b + 1, n) \right]^{1/2}}$$

für  $2 \leq T_b \leq n$  (4.12)

mit  $\bar{x}_{j,k} = (k-j+1)^{-1} \sum_{t=j}^k x_t$  und  $S_t(j,k) = \sum_{h=j}^t (x_h - \bar{x}_{j,k})$ . Diese Statistik wird dann über eine Reihe möglicher Bruchzeitpunkte  $T_b \in [\lfloor \epsilon n \rfloor, \lfloor (1-\epsilon)n \rfloor]$  ausgewertet. Die Teststatistik ergibt sich dann als

$$MC(\hat{T}_b) = \max_{T_b \in [\lfloor \epsilon n \rfloor, \lfloor (1-\epsilon)n \rfloor]} G(T_b) \quad \text{mit } 0 < \epsilon < 1. \quad (4.13)$$

Die Verteilung der Teststatistik für  $n \rightarrow \infty$  gegen  $MC(\hat{T}_b) = MC(\tilde{d}, \epsilon)$  selbst hat keine analytisch geschlossene Form und hängt maßgeblich von den Parametern  $d$  und  $\epsilon$  ab. Shao (2011) konnte zeigen, dass sie gegen

$$MC(\tilde{d}, \epsilon) \xrightarrow{a} \max_{r \in [\epsilon, (1-\epsilon)]} \frac{|B_{\tilde{d}}(r)/r - (B_{\tilde{d}}(1) - B_{\tilde{d}}(r))/(1-r)|}{\left[ \int_0^r W_{\tilde{d}}(r'; 0, r)^2 dr' + \int_r^1 W_{\tilde{d}}(r'; r, 1)^2 dr' \right]^{1/2}} \quad (4.14)$$

konvergiert. Dabei bezeichnet  $B_{\tilde{d}}(r)$  eine fraktionale Brownsche Bewegung und  $W_{\tilde{d}}(r; r_1, r_2) = B_{\tilde{d}}(r) - B_{\tilde{d}}(r_1) - (B_{\tilde{d}}(r_2) - B_{\tilde{d}}(r_1))(r - r_1)/(r_2 - r_1)$  für  $r \in [r_1; r_2]$  mit  $0 \leq r_1 < r_2 \leq 1$ . Für einen bekannten Integrationsparameter  $d$  kann die Verteilung der Teststatistik simuliert werden. Shao (2011) selbst beschränkt sich bei der Simulation der kritischen Werte auf  $\epsilon = 0,15$  und wählt einen Beobachtungsumfang von  $n = 5000$ , um asymptotische kritische Werte zu erreichen. Er simuliert Prozesse mit

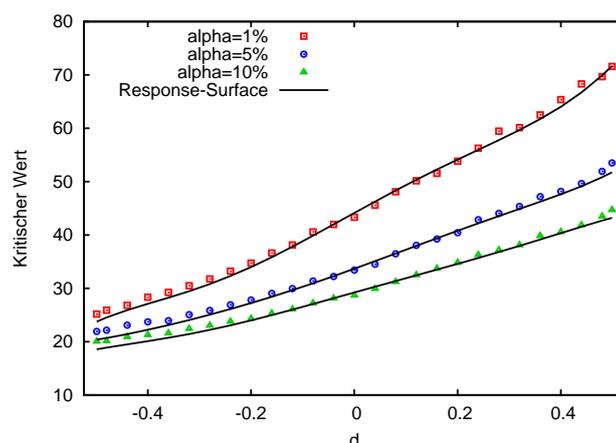


Abb. 4.11: Vergleich der kritischen Werte nach Shao (2011) mit den geschätzten Response-Surface-Funktionen

$\tilde{d} \in [-0,499; 0,499]$  und wiederholt für jeden Integrationsparameter die Testprozedur 10.000-mal. Dies scheint zunächst als sehr wenig, ist aber dem Umstand geschuldet, dass die Berechnung der Teststatistik für  $n = 5000$  mindestens 3500-mal erfolgen muss, um überhaupt einen einzigen Wert für  $MC(\hat{T}_b)$  zu ermitteln. Wir wollen nun kleinere Beobachtungsumfänge wählen und dann die geschätzten kritischen Werte für  $n \rightarrow \infty$  extrapolieren. Dafür können wir dann die Prozedur mit einer größeren Anzahl von Wiederholungen durchführen.

Wir wählen nun die Beobachtungsumfänge  $n_j = 500/j$  für  $j = 1, \dots, 8$  und simulieren hierfür Prozesse integriert vom Grad  $\tilde{d} \in \{-0,499; -0,45; -0,4; \dots; 0,4; 0,45; 0,499\}$ . Dies wiederholen wir dann insgesamt 100.000-mal, woraufhin die Extrapolation der kritischen Quantile für  $\alpha = 10\%, 5\%$  und  $1\%$  folgt. Danach können wir wieder eine Response-Surface-Funktion schätzen, die sich als Polynom sechsten Grades von  $\tilde{d} + 0,5$  ergibt. Die einzelnen geschätzten Parameter befinden sich in Tabelle 4.6.

In den nun folgenden Studien wollen wir den Test auf Trendbruch (weiterhin bezeichnet als TB) mit dem Meanchange (weiterhin: MC) nach Shao (2011) vergleichen.

Dabei wenden wir den TB-Test auf Prozesse der Form (4.2) und den MC-Test, wie beschrieben, auf die erste Differenz an. Generell werden wir annehmen, dass der wahre Zeitpunkt des Strukturbruchs unbekannt sei und werden die jeweiligen Bruchpunkte endogen schätzen. Dazu haben wir erste Ergebnisse für die Beobachtungsumfänge  $n = 100$  in der Tabelle 4.7 und für  $n = 250$  in 4.8, wo wir die jeweils durchschnittlichen geschätzten relativen Bruchzeitpunkte  $\hat{\tau} = \hat{T}_b/n$  darstellen.

Tab. 4.7: Durchschnittliche mit TB und MC geschätzte relative Bruchpunkte  $\hat{\tau}$  für  $n = 100$  mit  $\mu_1 = 0$

$\tau$	$\mu_2 = 0,5$					$\mu_2 = 1$					
	0,3	0,4	0,5	0,6	0,7	0,3	0,4	0,5	0,6	0,7	
$d = 0,8$	TB	0,3186 (0,09)	0,4073 (0,08)	0,5005 (0,07)	0,5957 (0,07)	0,6871 (0,08)	0,3021 (0,03)	0,4000 (0,03)	0,4999 (0,03)	0,5992 (0,03)	0,6983 (0,03)
	MC	0,2752 (0,11)	0,3759 (0,10)	0,4797 (0,10)	0,5853 (0,10)	0,6870 (0,10)	0,2759 (0,03)	0,3774 (0,03)	0,4796 (0,03)	0,5814 (0,03)	0,6837 (0,03)
$d = 1$	TB	0,3750 (0,17)	0,4251 (0,13)	0,5058 (0,12)	0,5806 (0,13)	0,6465 (0,15)	0,3094 (0,07)	0,4035 (0,05)	0,5003 (0,05)	0,5965 (0,05)	0,6931 (0,06)
	MC	0,3406 (0,20)	0,4017 (0,18)	0,4867 (0,17)	0,5678 (0,18)	0,6429 (0,19)	0,2778 (0,07)	0,3768 (0,06)	0,4798 (0,06)	0,5804 (0,06)	0,6837 (0,07)
$d = 1,2$	TB	0,4471 (0,20)	0,4719 (0,18)	0,5045 (0,17)	0,5435 (0,18)	0,5787 (0,20)	0,3554 (0,15)	0,4179 (0,11)	0,5006 (0,10)	0,5874 (0,11)	0,6603 (0,13)
	MC	0,4241 (0,25)	0,4585 (0,23)	0,4926 (0,22)	0,5341 (0,23)	0,5721 (0,24)	0,3234 (0,17)	0,3933 (0,14)	0,4818 (0,14)	0,5761 (0,13)	0,6527 (0,15)
$d = 1,4$	TB	0,4954 (0,21)	0,4951 (0,19)	0,5120 (0,19)	0,5291 (0,19)	0,5418 (0,20)	0,4339 (0,20)	0,4623 (0,17)	0,5081 (0,16)	0,5550 (0,16)	0,5904 (0,19)
	MC	0,4781 (0,25)	0,4904 (0,24)	0,5059 (0,24)	0,5170 (0,24)	0,5380 (0,25)	0,4089 (0,23)	0,4446 (0,21)	0,4956 (0,20)	0,5391 (0,20)	0,5870 (0,22)

Geschätzte Standardabweichungen von  $\hat{\tau}$  befinden sich in Klammern

Tab. 4.8: Durchschnittliche mit TB und MC geschätzte relative Bruchpunkte  $\hat{\tau}$  für  $n = 250$  mit  $\mu_1 = 0$

$\tau$	$\mu_2 = 0,5$					$\mu_2 = 1$					
	0,3	0,4	0,5	0,6	0,7	0,3	0,4	0,5	0,6	0,7	
$d = 0,8$	TB	0,3015 (0,03)	0,4015 (0,03)	0,4995 (0,03)	0,5992 (0,03)	0,6986 (0,03)	0,3007 (0,01)	0,4006 (0,01)	0,5003 (0,01)	0,6000 (0,01)	0,6994 (0,01)
	MC	0,2855 (0,04)	0,3904 (0,03)	0,4913 (0,03)	0,5936 (0,03)	0,6974 (0,04)	0,2909 (0,01)	0,3917 (0,01)	0,4922 (0,01)	0,5926 (0,01)	0,6929 (0,01)
$d = 1$	TB	0,3207 (0,10)	0,4065 (0,07)	0,5011 (0,07)	0,5937 (0,07)	0,6822 (0,09)	0,3030 (0,03)	0,4005 (0,03)	0,5006 (0,03)	0,5990 (0,03)	0,6981 (0,03)
	MC	0,2953 (0,11)	0,3905 (0,10)	0,4925 (0,09)	0,5945 (0,10)	0,6911 (0,10)	0,2898 (0,03)	0,3905 (0,03)	0,4921 (0,03)	0,5935 (0,03)	0,6949 (0,03)
$d = 1,2$	TB	0,4059 (0,19)	0,4453 (0,16)	0,5000 (0,15)	0,5644 (0,16)	0,5996 (0,19)	0,3228 (0,10)	0,4065 (0,08)	0,5015 (0,07)	0,5949 (0,08)	0,6810 (0,10)
	MC	0,3840 (0,22)	0,4340 (0,20)	0,4939 (0,19)	0,5577 (0,20)	0,6065 (0,22)	0,3019 (0,11)	0,3933 (0,09)	0,4933 (0,09)	0,5914 (0,09)	0,6837 (0,11)
$d = 1,4$	TB	0,4802 (0,21)	0,4886 (0,19)	0,5086 (0,19)	0,5194 (0,19)	0,5341 (0,20)	0,4107 (0,19)	0,4493 (0,16)	0,5053 (0,15)	0,5607 (0,16)	0,5963 (0,18)
	MC	0,4725 (0,25)	0,4860 (0,24)	0,5050 (0,23)	0,5233 (0,24)	0,5369 (0,25)	0,3957 (0,22)	0,4412 (0,20)	0,4989 (0,19)	0,5567 (0,19)	0,6004 (0,21)

Geschätzte Standardabweichungen von  $\hat{\tau}$  befinden sich in Klammern

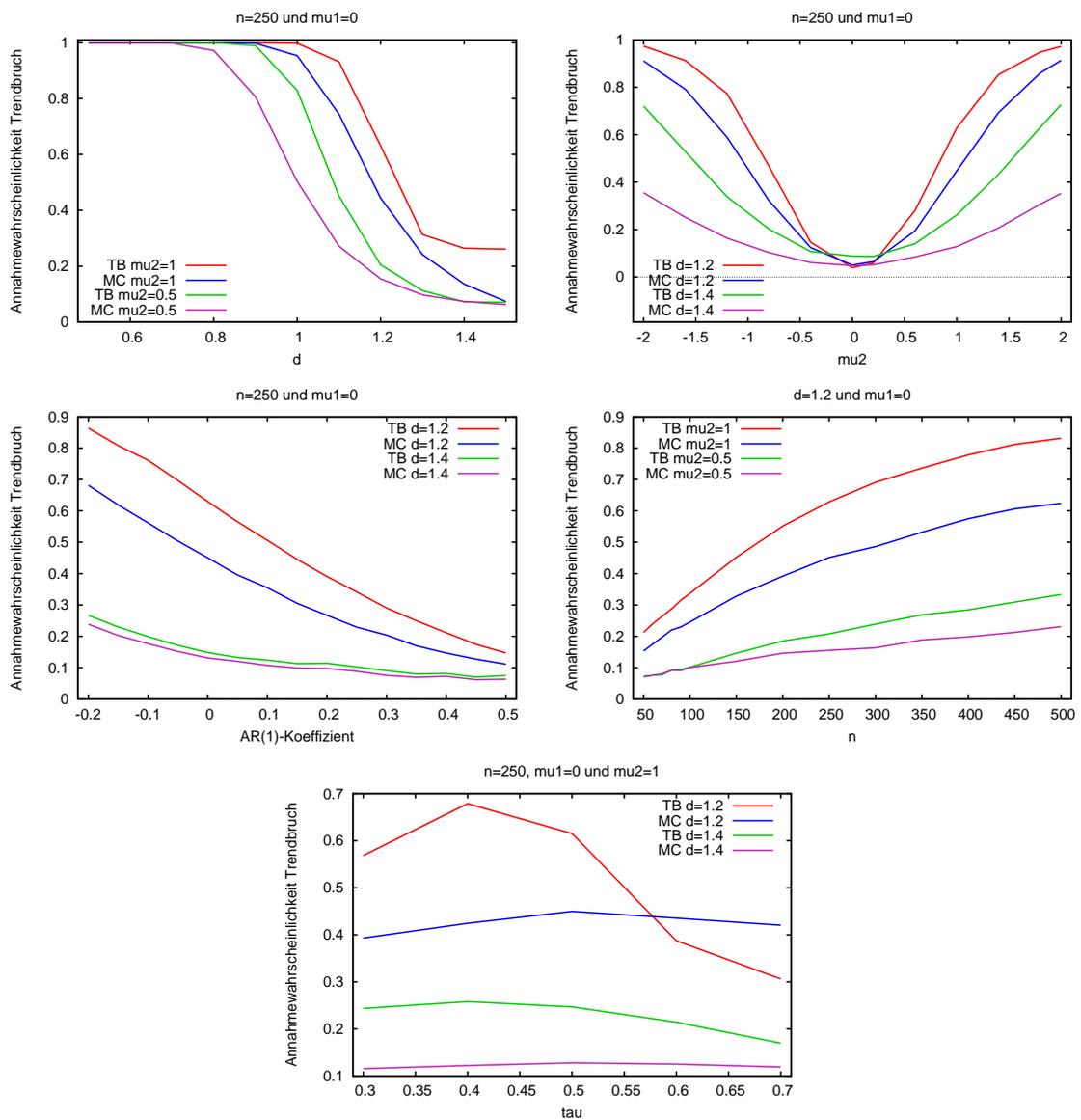


Abb. 4.12: Macht des TB-Tests gegen MC

Die Integrationsparameter  $d$  bzw.  $\tilde{d} = d - 1$  nehmen wir immer als unbekannt und schätzen diese mit dem ELW-Verfahren nach Shimotsu (2010), das für beide Modellspezifikationen anwendbar ist. Dazu verwenden wir immer die optimale Bandbreite, so wie wir sie im vorangegangenen Kapitel bestimmt haben. Weiterhin wiederholen wir jedes der folgenden Experimente 5000-mal und kommen hierbei zum Ergebnis, dass die Bruchzeitpunkte der beiden Verfahren TB und MC schon für relativ kleine Stichproben gut geschätzt werden können. Im Mittel liegen die TB-Werte etwas näher am wahren Bruchpunkt als die MC-Werte, aber die Unterschiede sind relativ gering. Beide Verfahren schätzen den Zeitpunkt schlechter im Bezug auf den Mittelwert und auch auf die Streuung, wenn der Integrationsparameter  $d$  steigt. Das TB-Verfahren ist dabei aber immer leicht effizienter als das MC-Verfahren.

In der Abbildung 4.12 vergleichen wir die Macht der beiden Tests, d.h. wir geben die relativen Häufigkeiten der Ablehnung der Nullhypothesen an, dass ein Prozess keinen Trendbruch hat im Falle des TB-Tests bzw. dass die erste Differenz von diesem keinen Mittelwertbruch aufweist für den MC-Test. Dazu variieren wir den Integrationsparameter  $d$ , die Stärke des Trendbruchs  $\mu_2$ , den Beobachtungsumfang  $n$  und den relativen Bruchzeitpunkt  $\tau$ . Des Weiteren nehmen wir auch eine ARFIMA(1,  $d$ , 0)-Ordnung an, wo wir den AR(1)-Koeffizienten  $\theta$  verändern und die Macht vergleichen. Die Ergebnisse lassen sich dabei zusammenfassen, dass der TB-Test fast immer besser abschneidet als der MC-Test. Die Unterschiede sind zwar teilweise sehr gering, wie z.B. wenn wir  $d = 1,4$  wählen und eine AR(1)-Struktur in den datengenerierenden Prozess aufnehmen. Einzig allein bei der Konstellation  $d = 1,2$  und einen späteren Zeitpunkt des Trendbruchs  $\tau > 0,5$  hat der TB-Test einen unerklärlichen Einbruch. Ein Grund könnte darin liegen, dass mit dem TB-Test der Bruchzeitpunkt evtl. nicht genau genug geschätzt werden kann, aber der MC-Test schneidet in dieser Situation nicht besser ab, wie wir in Tabelle 4.8 sehen können. Alles in allem können wir es leider nicht klären, warum der TB-Test in dieser Situation so schlecht abschneidet, können aber dennoch aussagen, dass dieser eine bessere Macht aufweist als der bereits bestehende und gegenüber Long-Memory-Eigenschaften robuste MC-Test von Shao (2011).

## 4.2 Herkömmliche Einheitswurzeltests

Im Folgenden beschäftigen wir uns mit dem Verhalten der gebräuchlichsten Tests des Integrationsgrads, nämlich dem Dickey-Fuller-Test und dem KPSS-Test. Diese Tests sind primär für die Möglichkeit  $I(1)$  gegen  $I(0)$  zu testen vorgeschlagen

worden, wobei wir aber nun im Rahmen des fraktionalen Kontextes, in dem der Integrationsgrad  $d$  jede beliebige Zahl annehmen kann, das Verhalten dieser Tests studieren wollen.

#### 4.2.1 Der Test nach Dickey-Fuller

Der Test nach Dickey und Fuller (1979) ist darauf ausgelegt, die Nullhypothese zu testen, dass ein Prozess  $y_t$  eine Einheitswurzel hat, also  $I(1)$  ist gegen  $I(0)$ . Dies ist wiederum gleichbedeutend mit der Nicht-Stationarität von  $y_t$ , wenn man keine fraktionale Integrationsparameter zulässt. Es wird angenommen, dass  $y_t$  durch einen Prozess der Form

$$y_t = \beta y_{t-1} + \varepsilon_t \quad (4.15)$$

generiert wird, wobei  $\varepsilon_t$  einen stationären Störprozess darstellt, der womöglich eine autokorrelierte Struktur aufweisen kann. Ergibt sich nun  $\beta = 1$ , dann ist  $y_t$  ein nicht-stationärer Random Walk-Prozess und für  $\beta > 1$  haben wir ebenso einen nicht-stationären Prozess, der aber zudem ein explosives Verhalten aufweist. Dies kann nun nach Dickey und Fuller (1979) mit Hilfe der erweiterten Testregression

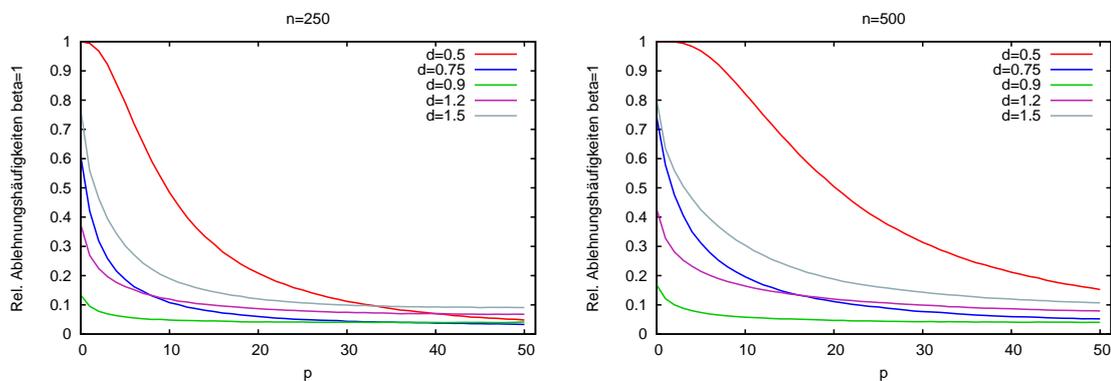
$$\Delta y_t = \underbrace{\phi}_{\beta-1} y_{t-1} + \sum_{j=1}^{p-1} \theta_j \Delta y_{t-j} + \varepsilon_t \quad (4.16)$$

getestet werden, indem eine t-Statistik für die Nullhypothese  $H_0 : \phi = 0$  berechnet wird, was gleichbedeutend mit  $\beta = 1$  ist. Interessiert nun die Gegenhypothese  $H_1 : \phi < 0$  und daraus folgernd die Stationarität von  $y_t$ , dann findet ein einseitiger Test Anwendung, wobei anzumerken ist, dass die t-Statistik keiner t-Verteilung folgt und kritische Werte für die Testentscheidung simuliert werden müssen. Die kritischen Werte ergeben sich dann als die  $\alpha$ -Quantile der simulierten Verteilung der Teststatistik. Alternativ kann auch  $H_1 : \phi \neq 0$  getestet werden, wobei dann ein zweiseitiger Test relevant wird. In der Regression (4.16) tauchen auch insgesamt  $p - 1$  zeitverzögerte Differenzen von  $y_t$  auf, womit eine AR( $p$ )-Struktur in  $y_t$  bzw. auch eine autokorrelierte Struktur in  $\varepsilon_t$  berücksichtigt werden kann. Durch diese Erweiterung ist dieser Test als erweiterter Dickey-Fuller-Test (augmented Dickey-Fuller-Test, weiterhin: ADF-Test) bekannt. Des Weiteren können in der Testregression deterministische Eigenschaften von  $y_t$  berücksichtigt werden, indem sie als weitere Regressoren aufgenommen werden. So kann ein linearer Trend in  $y_t$  modelliert werden, wenn eine Konstante in der Regression (4.16) aufgenommen wird und dementsprechend

Tab. 4.9: Relative Ablehnungshäufigkeiten für  $\beta = 1$  in Abhängigkeit von  $n$ 

$d$		0,5	0,75	0,9	1	1,2	1,5
$n = 100$	$p = 0$	98,5%	39,9%	9,4%	5,3%	29,9%	69,0%
	$p_{\text{short}} = 4$	54,8%	12,7%	5,2%	5,0%	13,3%	22,9%
	$p_{\text{long}} = 12$	12,5%	4,4%	4,1%	4,9%	7,7%	10,1%
	$p = p_{BIC}$	45,5%	18,7%	8,7%	7,3%	16,0%	22,2%
	$\bar{p}_{BIC}$	6,17 (6,38)	5,48 (6,43)	4,76 (6,43)	4,54 (6,48)	5,84 (6,49)	6,77 (6,34)
$n = 250$	$p = 0$	100,0%	60,4%	13,4%	5,1%	37,5%	76,3%
	$p_{\text{short}} = 5$	85,7%	21,5%	6,3%	5,1%	18,0%	34,4%
	$p_{\text{long}} = 15$	33,4%	7,8%	4,3%	4,9%	10,3%	14,0%
	$p = p_{BIC}$	79,7%	30,8%	10,4%	5,4%	21,3%	34,1%
	$\bar{p}_{BIC}$	3,41 (2,62)	2,47 (2,37)	1,15 (1,95)	0,68 (1,73)	2,5 (2,29)	3,85 (2,49)
$n = 500$	$p = 0$	100,0%	74,1%	17,1%	4,8%	43,1%	80,6%
	$p_{\text{short}} = 5$	98,4%	35,0%	8,0%	4,9%	23,2%	46,2%
	$p_{\text{long}} = 17$	61,6%	12,9%	5,0%	4,8%	13,4%	21,2%
	$p = p_{BIC}$	92,8%	41,2%	12,7%	5,0%	25,7%	42,3%
	$\bar{p}_{BIC}$	4,02 (2,26)	2,73 (1,75)	1,07 (1,32)	0,29 (0,89)	2,65 (1,74)	4,43 (2,03)
$n = 1000$	$p = 0$	100,0%	85,2%	21,3%	5,0%	48,1%	83,7%
	$p_{\text{short}} = 7$	99,8%	42,9%	9,1%	5,0%	25,9%	50,6%
	$p_{\text{long}} = 21$	84,0%	19,3%	5,9%	5,0%	16,8%	29,1%
	$p = p_{BIC}$	98,4%	51,6%	14,7%	5,0%	30,0%	49,9%
	$\bar{p}_{BIC}$	5,26 (2,25)	3,5 (1,7)	1,37 (1,12)	0,15 (0,56)	3,25 (1,62)	5,62 (2,03)

ein quadratischer Trend in  $y_t$ , wenn ein linearer Trend in (4.16) mit geschätzt wird. Die Existenz solcher deterministischer Komponenten beeinflusst aber die kritischen Werte und muss bei der Testentscheidung berücksichtigt werden.

Abb. 4.13: Relative Ablehnungshäufigkeiten für  $\beta = 1$  in Abhängigkeit der Wahl von  $p$ 

Das Verhalten von diesem, in der Praxis gebräuchlichsten Test, wurde schon sehr bald nach der Einführung der fraktional integrierten Prozesse untersucht. Hier interessiert z.B die Annahme, dass der Störterm fraktional integriert, aber stationär und invertierbar ist mit

$$y_t = \beta y_{t-1} + \Delta^{-d_\varepsilon} \varepsilon_t \quad \text{mit} \quad -1/2 < d_\varepsilon < 1/2. \quad (4.17)$$

Unter der Nullhypothese der ADF-Statistik gilt nun  $\beta = 1$ , was wiederum impliziert, dass

$$\Delta y_t = \Delta^{-d_\varepsilon} \varepsilon_t \quad \Leftrightarrow \quad y_t = \Delta^{-\overbrace{(1+d_\varepsilon)}^d} \varepsilon_t \quad \text{mit} \quad 1/2 < d < 3/2 \quad (4.18)$$

ist und somit  $y_t$  ein nicht-stationärer Prozess darstellt, der nach einmaliger Differenzenbildung erster Ordnung stationär ist, aber eine Long-Memory-Eigenschaft aufweist. Diebold und Radebusch (1991) untersuchen in diesem Rahmen nun, wie die Macht des ADF-Tests in diesen Fällen ist, um  $\beta = 1$  zu bestätigen. Dazu führen sie zweiseitige Tests durch, da auch positiv integrierte Störterme angenommen werden und stellen fest, dass der ADF-Test sehr häufig die Hypothese  $\beta = 1$  fälschlicherweise ablehnt. Sie vergleichen die Situation der fraktional integrierten Störterme mit dem Fall von AR(1)-Störtermen. Auch hier gilt, dass bei diesen wesentlich seltener die Nullhypothese abgelehnt wird, selbst wenn mit  $\theta$  Autokorrelationsparameter nahe 1 angenommen werden. Hassler und Wolters (1993) argumentieren, dass die Macht des ADF-Tests dadurch eingeschränkt ist, weil tendenziell eine große Anzahl von Lags  $p$  gewählt werden muss, um das Ausmaß der Autokorrelation in  $y_t$  zu berücksichtigen. Sie führen daraufhin Monte-Carlo-Simulationen durch und verwenden zur Lag-Bestimmung die Heuristik nach Schwert (1989)  $p = \lfloor 12(n/100)^{1/4} \rfloor$  und kommen zu dem Ergebnis, dass der Test nicht konsistent ist, wenn stationäre fraktional integrierte Störterme angenommen werden. Krämer (1998) stützt seine Studie mehr auf eine analytische Betrachtungsweise und konnte zeigen, dass bei zweiseitiger Testweise der ADF-Test für  $|d_\varepsilon| < 1/4$  und damit für  $3/4 < d < 5/4$  bei Wahl der Schwert-Heuristik konsistent ist.

In unseren Simulationsstudien wollen wir für die Lag-Wahl neben der Schwert-Heuristik auch eine Lag-Selektion mit Hilfe des Schwarz-Informationskriteriums (BIC) treffen. Dazu simulieren wir nun variable Stichprobenumfänge von  $n \in \{100; 250; 500; 1000\}$  und wählen die Integrationsparameter  $d \in \{0,5; 0,75; 0,9; 1; 1,2; 1,5\}$  für  $y_t$ . Wir wählen  $p_{\max} = 21$  als maximalen Lag, den wir für jeden Prozess annehmen und berechnen die ADF-Statistik mitsamt des Schwarz-Kriteriums für die Testregression bis  $p_{\min} = 0$ . Daraufhin wählen wir den Lag  $p_{BIC}$ , der mit dem niedrigsten BIC einhergeht. Jede der Prozeduren wiederholen wir insgesamt 50.000-mal und ermitteln mit  $\bar{p}_{BIC}$  den Mittelwert des jeweils gewählten  $p_{BIC}$  und geben in Klammern auch deren Standardabweichung an. Die Ergebnisse hiervon sind nun in Tabelle 4.9 angegeben. Zusätzlich geben wir hier die Ablehnungsraten bei der Wahl der Schwert-Heuristiken  $p_{\text{short}} = \lfloor 4(n/100)^{1/4} \rfloor$  und  $p_{\text{long}} = \lfloor 12(n/100)^{1/4} \rfloor$ . Als Irr-

tumsniveau nehmen wir durchweg  $\alpha = 5\%$  an. Für alle Ausprägungen von  $d \neq 1$  ergeben sich steigende Ablehnungsraten mit steigendem Integrationsparameter und zwar für sämtliche Selektionen des Lag-Parameters  $p$ , was auch Hassler und Wolter (1993) als Nicht-Konsistenz des ADF-Tests interpretiert haben. In der Tat müssten die Lag-Parameter  $p$  noch höher gewählt werden, wie die Abbildung 4.13 zeigt. Hier haben wir für  $n = 250$  und  $n = 500$  die Struktur der Ablehnungsraten in Abhängigkeit des gewählten Lags bis zu einem maximalen Lag von  $p = 50$ . Dabei kann man gut erkennen, dass für  $d = 0,5$  und  $d = 1,5$  tendenziell nicht die erwartete 5%-Ablehnungsrate erreicht wird, was sich auch mit den Erkenntnissen von Krämer (1998) deckt. Das Problem ist aber nun, dass die hohen Lag-Parameter in der empirischen Anwendung kaum gewählt werden, da entweder die Schwert (1989)-Heuristiken verwendet werden, die tendenziell zu klein ausfallen, oder es wird nach dem BIC-Kriterium eine Lag-Zahl selektiert, die wiederum im Durchschnitt sogar noch kleiner als bei der Schwert-Heuristik ausfällt. Aus diesem Grund beobachten wir bei der Wahl  $p = p_{BIC}$  noch höhere Ablehnungsraten als bei der Verwendung der Schwert-Heuristik.

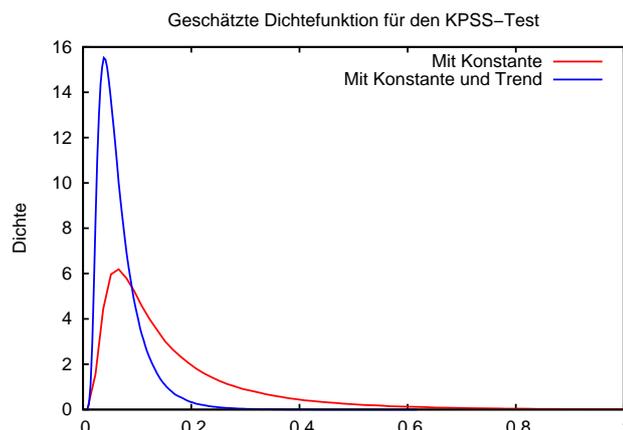
#### 4.2.2 Der KPSS-Test

Ein weiterer oft in empirischen Studien angewandter Test wurde von Kwiatkowski, Phillips, Schmidt und Shin (1992, weiterhin: KPSS) entwickelt. Im Gegensatz zum ADF-Test lautet hier die Nullhypothese, dass Stationarität vorliegt, d.h. ein zu untersuchender Prozess  $y_t$  ist  $I(0)$  und die Gegenhypothese, dass dieser eine Einheitswurzel hat, also  $y_t \sim I(1)$ . KPSS (1992) nehmen die folgenden datengenerierenden Prozesse unter  $H_0$  an:

$$\text{Mit Konstante : } y_t = \mu_1 + \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim I(0) \quad (4.19)$$

$$\text{Mit Trend : } y_t = \mu_1 + \mu_2 \cdot t + \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim I(0) \quad (4.20)$$

Unter Gültigkeit von  $H_0$  lassen sich die Parameter  $\mu_1$  und  $\mu_2$  mit Hilfe einer einfachen KQ-Regression bestimmen. Aus den geschätzten Residuen  $\hat{\varepsilon}_t$  dieser Regressionen

Abb. 4.14: Dichtefunktion des KPSS-Tests für  $n = 2000$ 

haben KPSS (1992) nun einen Lagrange-Multiplikatoren-Test entwickelt:

$$\begin{aligned}
 KPSS &= \frac{1}{n^2} \sum_{t=1}^n \frac{S_t^2}{\sigma^2(l)}, \quad \text{mit } S_t = \sum_{i=1}^t \hat{\varepsilon}_i \\
 \text{und } \sigma^2(l) &= \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2 + 2 \frac{1}{n} \sum_{s=1}^l \omega(s, l) \sum_{t=s+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-s} \\
 \text{mit } \omega(s, l) &= 1 - \frac{s}{l+1}.
 \end{aligned} \tag{4.21}$$

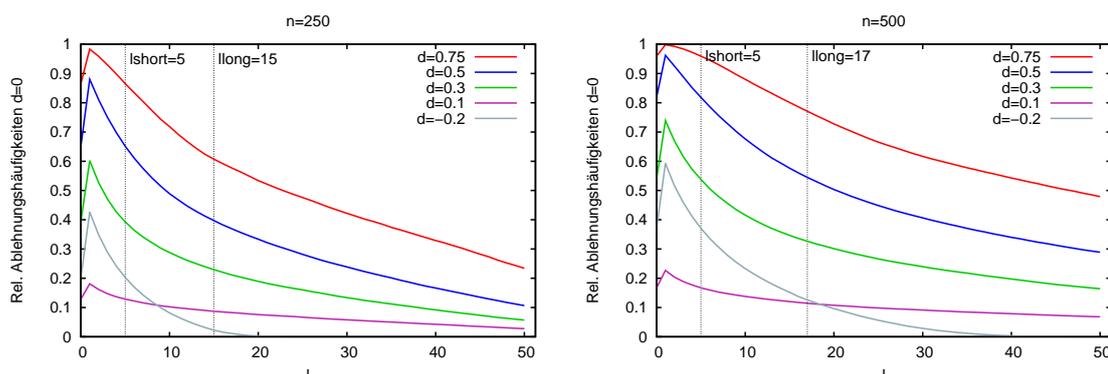
Bei dieser Berechnungsweise wird die langfristige Varianz von  $\varepsilon_t$  gemäß des Vorschlags von Newey und West (1987) berechnet, wobei  $l$  ein frei zu wählender Parameter ist, um eine mögliche Autokorrelation des Störterms  $\varepsilon_t$  zu berücksichtigen. Die Verteilung der Teststatistik hat selbst keine analytisch geschlossene Form und KPSS (1992) haben kritische Werte für den Test durch Simulationen bestimmt, indem sie für einfache Weiße Rauschen mit Beobachtungsumfang  $n = 2000$  die Teststatistik berechnet haben und dies insgesamt 50.000-mal wiederholt haben. Die kritischen Werte ergeben sich dann als  $(1-\alpha)$ -Quantil der simulierten Verteilung, da  $H_0 : d = 0$  gegen  $H_1 : d = 1$  (bzw.  $d > 0$ ) getestet werden soll. Dementsprechend wird die Nullhypothese genau dann abgelehnt, wenn eine ermittelte Teststatistik größer als der jeweilige kritische Wert ist. Für unser Anwendungsgebiet interessiert uns nun, welche Testentscheidungen sich ergeben, wenn wir fraktional integrierte Prozesse der Ordnung  $d$  mit dem KPSS-Test untersuchen. Hierzu stellten Lee und Schmidt (1996) fest, dass der Test durchaus Macht besitzt, um zwischen  $I(0)$ - und  $I(d)$ -Prozessen für  $d > 0$  zu unterscheiden. Dies erfordert aber zum Teil relativ große Stichprobenumfänge, da die Macht geringer wird, wenn der Integrationsgrad nur geringfügig

Tab. 4.10: Asymptotische kritische Quantile des KPSS-Tests

Quantil	0,5%	1%	2,5%	5%	10%	90%	95%	97,5%	99%	99,5%
Mit Konstante KPSS (1992)	0,022	0,025	0,030	0,036	0,046	0,348	0,463	0,583	0,747	0,874
						0,347	0,463	0,574	0,739	
Mit Trend KPSS (1992)	0,015	0,017	0,020	0,023	0,028	0,119	0,148	0,178	0,218	0,249
						0,119	0,146	0,176	0,216	

größer als 0 ist. Sie simulieren zusätzlich in ihrer Studie auch AR(1)-Prozesse mit  $I(0)$ -Störtermen, wo der KPSS-Test tendenziell dazu neigt, die Hypothese  $H_0 : d = 0$  anzunehmen. Damit folgern sie, dass der Test durchaus geeignet ist, um fraktionale Integration von einfacher Autokorrelation zu unterscheiden. Um auch den Fall  $d < 0$  abzudecken, schlägt Marmol (1997) vor, den KPSS-Test zweiseitig anzuwenden. D.h. im Endeffekt, dass die eigentliche Teststatistik nicht verändert wird, aber es müssen auch untere kritische Werte ermittelt werden, die dann den  $\alpha/2$ -Quantilen der simulierten Verteilung der jeweiligen Teststatistiken entsprechen. Diese haben wir auch in einer Simulation bestimmt und die jeweiligen relevanten Quantile in Tabelle 4.10 dargestellt. Hierzu haben wir jeweils standardnormalverteilte Weiße Rauschen mit variablem Stichprobenumfang  $n \in [50; 2000]$  simuliert und die KPSS-Statistik mit Konstante und mit Trend berechnet. Hieraus wurden dann die Quantile, nachdem wir die Prozedur 1.000.000-mal wiederholt hatten, ausgelesen und für  $n \rightarrow \infty$  extrapoliert. In der Tabelle 4.10 sind auch die von KPSS (1992) ermittelten kritischen Werte abgetragen, die bis heute auch noch verwendet werden. Man sieht hierbei, dass diese in der Regel relativ gut getroffen werden. Nur für das 99%-Quantil ergibt sich eine größere Abweichung bei unserer Simulation.

Nun wollen wir in einer Simulationsstudie das Verhalten der KPSS-Statistik bei fraktional integrierten Prozessen untersuchen. Hierzu simulieren wir für Stichprobenumfänge  $n \in \{100; 250; 500; 1000\}$  Prozesse mit einem Integrationsgrad  $d \in \{-0,2; 0,1; 0,3; 0,5; 0,75\}$ . Wir testen für jeden Prozess die Nullhypothese  $d = 0$  zweiseitig mit den kritischen Werten aus Tabelle 4.10, so dass wir auch negative fraktional integrierte Prozesse zulassen. In der Tabelle 4.11 stellen wir die jeweiligen Annahmeraten der Nullhypothese bei einer Irrtumswahrscheinlichkeit von  $\alpha = 5\%$  dar, wobei die Lag-Wahl nach der Schwert-Heuristik getroffen wird, was bei empirischen Anwendungen durchaus üblich ist. Daneben existiert für die KPSS-Statistik eine Reihe von automatisierten Lag-Wahl-Algorithmen, wobei wir auf deren Anwendung hier verzichten wollen. Hierbei zeigt sich, dass der Test durchaus Macht hat, um fraktionale Integration von  $d = 0$  zu unterscheiden. Es überrascht dabei wenig,

Abb. 4.15: Relative Ablehnungshäufigkeiten für  $d = 0$  in Abhängigkeit von  $l$ Tab. 4.11: Relative Ablehnungshäufigkeiten für  $H_0 : d = 0$  bei Verwendung des KPSS-Tests

$d_0/n$	$l_{\text{short}} = \lfloor 4(n/100)^{1/4} \rfloor$				$l_{\text{long}} = \lfloor 12(n/100)^{1/4} \rfloor$			
	100	250	500	1000	100	250	500	1000
-0,2	5,8%	20,9%	37,5%	48,8%	0,0%	2,5%	12,8%	24,6%
0,1	9,6%	13,3%	17,0%	19,2%	5,3%	8,9%	11,6%	13,9%
0,3	26,1%	39,8%	53,4%	62,2%	12,4%	23,3%	32,1%	41,7%
0,5	45,9%	65,4%	81,8%	88,9%	23,1%	40,1%	54,7%	67,8%
0,75	67,4%	86,1%	95,8%	98,4%	40,3%	60,5%	77,2%	87,7%

dass die Macht größer wird, wenn  $d$  betragsmäßig größer von 0 verschieden ist oder der Stichprobenumfang steigt. Interessanter ist das unterschiedliche Verhalten des KPSS-Tests bei unterschiedlichen Integrationsparametern. Dazu haben wir auch in der Abbildung 4.15 für  $n = 250$  und  $n = 500$  den Verlauf der relativen Ablehnungshäufigkeiten in Abhängigkeit der jeweiligen Lag-Wahl  $l$  abgetragen. Hier zeigt sich, dass sich für  $d = -0,2$  ein steilerer Abfall der Ablehnungsrate ergibt, als wenn  $d > 0$  vorliegt.

### 4.3 Lagrange-Multiplikatoren Tests

Nun wollen wir uns mit Teststatistiken zur Überprüfung der Hypothese  $H_0 : y_t \sim I(d)$  beschäftigen, die explizit für die Situation der fraktionalen Integration vorgeschlagen wurden. Dies lässt sich z.B. überprüfen, wenn wir den datengenerierenden Prozess

$$\Delta^{d+\delta} y_t = \varepsilon_t \quad \text{mit} \quad \varepsilon_t \sim IID(0; \sigma_\varepsilon^2) \quad (4.22)$$

auf  $H_0 : \delta = 0$  gegen  $H_1 : \delta \neq 0$  überprüfen. Weiterhin werden wir uns auch für die Gegenhypothese  $H_1 : \delta < 0$  interessieren. Um dies nun zu testen, haben Robinson (1991, 1994) und Tanaka (1999) eine Score-Statistik entwickelt, die auf der logarithmierten Likelihood-Funktion des Prozesses basiert unter der Annahme, dass der Störprozess  $\varepsilon_t$  normalverteilt ist. Dies stellt zwar auf den ersten Blick eine starke Einschränkung dar, aber wie wir bereits bei der Tanaka-Approximation im vorangegangenen Kapitel gesehen haben, weist z.B. die Schätzung des Integrationsparameters auch bei einer Abweichung von dieser Verteilungsannahme eine gewisse Robustheit auf. Der Unterschied zwischen den beiden Arbeiten ist der, dass Robinson (1994b) sich auf logarithmierte Likelihood im Frequenzbereich konzentriert, während Tanaka (1999) sich im Zeitbereich befindet. Setzen wir nun  $x_t \equiv \Delta^d y_t$ , dann konnte Tanaka (1999) eine LM-Statistik auf Basis der gewichteten Autokorrelationskoeffizienten von  $x_t$  vorschlagen und deren asymptotische Verteilung zeigen mit

$$LM = \frac{\sqrt{6n}}{\pi} \sum_{j=1}^{n-1} j^{-1} \hat{\rho}_j \stackrel{a}{\sim} N(0; 1) \quad \text{wobei} \quad \hat{\rho}_j = \frac{\sum_{t=j+1}^n x_t x_{t-j}}{\sum_{t=1}^n x_t^2}, \quad (4.23)$$

wobei zunächst von keinen deterministischen Prozessbestandteilen in  $y_t$  bzw.  $x_t$  ausgegangen wird. Die Hypothese  $H_0 : \delta < 0$  kann damit abgelehnt werden, wenn die LM-Statistik kleiner als  $\pi/\sqrt{6} \cdot c_\alpha$  ist, wobei  $c_\alpha$  das  $\alpha$ -Quantil der Standardnormalverteilung darstellt. Denkbar wäre auch die quadrierte LM-Statistik zu betrachten, die asymptotisch einer  $\chi_1^2$ -Verteilung folgen müsste, also

$$LM_{\text{quad}} = \frac{6n}{\pi^2} \left[ \sum_{j=1}^{n-1} j^{-1} \hat{\rho}_j \right]^2 \stackrel{a}{\sim} \chi_1^2. \quad (4.24)$$

Die gleiche Teststatistik konnte auch von Robinson (1991) hergeleitet werden. Aufbauend auf diesem Resultat schlugen Agiakloglou und Newbold (1994) vor, die LM-Teststatistik als t-Statistik von  $\phi = 0$  aus der Regression

$$x_t = \phi x_{t-1}^* + v_t \quad (4.25)$$

zu berechnen. Für  $LM_{\text{quad}}$  wird dann einfach die quadrierte t-Statistik verwendet. Dabei schlagen sie mit  $x_{t-1}^* = \sum_{j=1}^m j^{-1} x_{t-j}$  eine auf  $m$  Parameter begrenzte Summe einzusetzen, um die Berechnung des Regressors zu vereinfachen. Breitung und Hassler (2002) hingegen schlugen für  $m = t - 1$  vor, um Komplikationen mit einem fehlspezifizierten  $m$  zu vermeiden. Interessanterweise zeigt diese Berechnungsweise

des LM-Tests eine starke Ähnlichkeit zu dem bereits erwähnten Dickey-Fuller-Test. Der Unterschied liegt darin, dass Dickey und Fuller (1979)  $d = 1$  fixieren für die Bildung der Differenz  $x_t = \Delta^d y_t$  und in der Testregression anstatt der gewichteten Summe  $\sum_{j=1}^{t-1} j^{-1} x_{t-j}$  die absolute Summe  $\sum_{j=1}^{t-1} x_{t-j}$  einsetzen. Die Verteilung der Teststatistiken ist aber nur gültig, falls es sich bei dem beobachteten Prozess  $y_t$  um ein fraktionales integriertes Rauschen handelt. Ist  $y_t$  nun aber ein womöglich autokorrelierter Prozess, so haben Agiakloglou und Newbold (1994) und Breitung und Hassler (2002) vorgeschlagen, eine vorherige Bereinigung des Prozesses  $y_t$  durchzuführen. Wir verfolgen hier den Ansatz von Demestrescu, Kuzin und Hassler (weiterhin DKH, 2008) die eine erweiterte Version analog zum erweiterten Dickey-Fuller-Test vorgeschlagen haben. Die zu schätzende Testregression lautet dann

$$x_t = \phi x_{t-1}^* + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + v_t. \quad (4.26)$$

Dabei beschreibt  $p$  die vermutete Ordnung des AR-Teils von  $y_t$ . DKH (2008) fanden in einer Monte-Carlo-Studie heraus, dass, wenn  $p$  unbekannt ist, die von Schwert (1989) vorgeschlagene Heuristik  $p_K = K(T/100)^{1/4}$  mit  $K = 4$  besser abschneidet als  $K = 12$ . Sie war dabei auch weit besser als die Lag-Selektionsverfahren mit den Informationskriterien nach Akaike oder Schwarz.

Eine wichtige Angelegenheit stellt noch die Berücksichtigung von deterministischen Komponenten in  $y_t$  dar. Hier gilt es zu beachten, dass die Teststatistik auf der Betrachtung von  $x_t = \Delta^d y_t$ , wobei  $d$  nicht notwendigerweise ganzzahlig sein muss, basiert und deterministische Komponenten wie einen linearen oder quadratischen Trend nicht einfach als Regressor in der Testregression (4.26) aufgenommen werden können, wie dies z.B. bei der ADF-Statistik der Fall ist. Vielmehr ist hier nötig, die Komponenten bereits für  $y_t$  zu testen und gegebenenfalls zu bereinigen, bevor die jeweilige LM-Statistik angewandt wird. Einen Einfluss der Bereinigung auf die kritischen Werte für die LM-Statistik ist wahrscheinlich vorhanden, lässt sich aber nur schwer berücksichtigen, da für den Parameter  $d$  keinerlei Restriktionen angenommen werden und damit  $y_t$  sowohl stationär als auch nicht-stationär sein kann. Dieser Umstand beeinflusst aber die Art und Weise, wie die deterministischen Komponenten beseitigt werden und hat damit auch wiederum selbst einen Einfluss auf kritische Werte. Deswegen nehmen wir nun im Folgenden an, dass wir mit  $y_t$  um einen korrekt von deterministischen Bestandteilen bereinigten Prozess haben und simulieren nun kritische Werte ohne weitere Behandlung. Als Folge daraus müssen wir damit rechnen, dass die LM-Teststatistiken bei Vorliegen von deterministischen

Komponenten überdimensioniert sein könnten.

Zur Bestimmung von kritischen Werten simulieren wir mit  $x_t = \Delta^{d+\delta} y_t$  einen Prozess, der unter  $H_0 : \delta = 0$  ein  $I(0)$ -Prozess darstellt, falls der Integrationsparameter  $d$  von  $y_t$  bekannt ist. Für diesen berechnen wir die Teststatistiken  $LM$  und  $LM_{\text{quad}}$  und können dann für vorgegebene Irrtumswahrscheinlichkeiten  $\alpha$  die kritischen Werte für  $LM$  als jeweiliges  $\alpha$ -Quantil auslesen und für  $LM_{\text{quad}}$  als  $(1 - \alpha)$ -Quantil. Wie bereits erwähnt, wollen wir mit  $LM$  die Gegenhypothese  $H_1 : \delta < 0$  testen, was wiederum bedeutet, dass der wahre Integrationsgrad von  $y_t$  kleiner als  $d$  ist. Für  $LM_{\text{quad}}$  gilt dann  $H_1 : \delta \neq 0$ , also dass  $y_t$  nicht integriert vom Grad  $d$  ist. In empirischen Untersuchungen werden wir aber zusätzlich mit dem Problem konfrontiert, dass wir den Integrationsgrad  $d$ , den wir mit Hilfe des LM-Statistik testen wollen, gar nicht kennen und diesen mit einem geeigneten Schätzverfahren zunächst bestimmen müssen. Vom Testansatz ist dies eigentlich nicht nötig, aber wir wollen diese Unsicherheit in einer Simulationstudie untersuchen, indem wir zwar mit  $x_t$  einen  $I(0)$ -Prozess simulieren, aber dessen Integrationsgrad noch zusätzlich schätzen mit  $\hat{d}_x$ . Hierzu bietet sich insbesondere das semiparametrische GPH-Verfahren an, da dieses einfach und effizient anwendbar ist und weniger Computerressourcen benötigt als z.B. die Whittle-Schätzer.

Für erste Simulationen generieren wir standardnormalverteilte Störterme als  $x_t$  mit  $n = 1000$  und schätzen für diese den Integrationsgrad  $\hat{d}_x$  mit dem GPH-Schätzer jeweils mit der Bandbreite  $m = n^a$  mit  $a \in \{0,5; 0,6; 0,7; 0,8\}$ . Daraufhin berechnen wir die LM-Statistiken für  $\Delta^{\hat{d}_x} x_t$  und ermitteln zum einen die Dichte und die kritischen Werte für das Irrtumsniveau  $\alpha = 5\%$ . Weiterhin führen wir diese Prozedur auch für  $x_t$  durch, was gleichbedeutend mit der Annahme ist, dass wir  $d_x = 0$  kennen. Sämtliche Simulationen werden dann insgesamt 100.000-mal wiederholt. Die Abbildung 4.16 zeigt die jeweils geschätzten Dichtefunktionen und die Tabelle 4.12 die kritischen Werte mit einigen geschätzten Verteilungsparametern für die geschätzte Verteilung der LM-Teststatistiken. Die asymptotischen Werte in Tabelle 4.12 ergeben sich für  $LM$  durch die Standardnormalverteilung und für  $LM_{\text{quad}}$  durch die  $\chi_1^2$ -Verteilung.

Als Ergebnis kann man festhalten, dass die Verteilung der LM-Statistiken massiv von der Bekanntheit des Integrationsparameters abhängt. Besonders groß sind die Abweichungen, wenn wir  $d$  nur mit geringer Bandbreite schätzen können, was z.B. bei stark autokorrelierten Prozessen der Fall ist. Diese nehmen zwar mit steigender Bandbreite tendenziell ab, haben aber kein einheitliches Niveau, so dass wir uns dazu entscheiden, die kritischen Werte nur in Abhängigkeit des Stichprobenumfangs zu

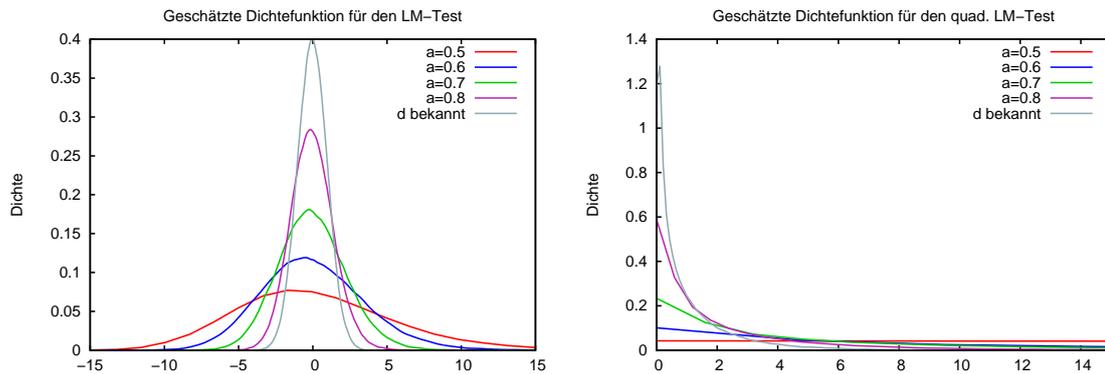


Abb. 4.16: Geschätzte Dichtefunktion der LM-Statistiken

simulieren unter der Annahme, dass wir den wahren Integrationsgrad kennen. Für  $n = 1000$  kann die Normalverteilungsannahme für die einfache LM-Statistik nicht abgelehnt werden, aber es gibt durchaus Anhaltspunkte, dass für kleinere Stichproben größere Abweichungen existieren. Somit kann eine Response-Surface-Funktion in Abhängigkeit von  $n$ , die aus den folgenden Simulationen resultiert, durchaus sinnvoll sein.

Tab. 4.12: Übersicht der Verteilungscharakteristiken und der kritischen Werte für  $\alpha = 5\%$  bei unbekanntem und bekanntem Integrationsparametern

		$a = 0,5$	$a = 0,6$	$a = 0,7$	$a = 0,8$	$d$ bekannt	asymptotisch
$LM$	MW	0,3349	0,0263	-0,0274	-0,0431	-0,0346	0
	SD	5,8564	3,5694	2,2927	1,4354	0,9972	1
	S	1,8339	0,5201	0,2706	0,2124	-0,0036	0
	ExK	26,790	1,1475	0,4191	0,3117	0,0036	0
	$CV_{\alpha=5\%}$	-7,7645	-5,3933	-3,6478	-2,3219	-1,6747	-1,6449
$LM_{quad}$	MW	34,409	12,741	5,2572	2,0595	0,9956	1
	SD	185,40	22,659	8,1547	3,1223	1,4094	1,4142
	S	139,38	9,0934	4,4219	3,9157	2,8419	2,8284
	ExK	25795	223,55	42,296	32,700	12,147	12
	$CV_{\alpha=5\%}$	120,28	47,706	20,129	7,9288	3,8231	3,8415

MW ist hier die Abkürzung für Mittelwert, SD für Standardabweichung, S für Schiefe und ExK für den Exzess (=Kurtosis-3).  $CV_{\alpha=5\%}$  ist der ermittelte kritische Werte für  $\alpha = 5\%$ .

Wir wollen nun für  $LM$  und  $LM_{quad}$  die Response-Surface-Funktion

$$CV_{\alpha}(n) = \gamma_0 + \gamma_1 n^{-1} + \gamma_2 n^{-2} + \gamma_3 n^{-3} + \gamma_4 n^{-4} + \text{Resid} \quad (4.27)$$

für die kritischen Werte parametrisieren. Dazu simulieren wir mit 1.000.000 Wiederholungen wiederum standardnormalverteilte Störterme bis zu einem maximalen

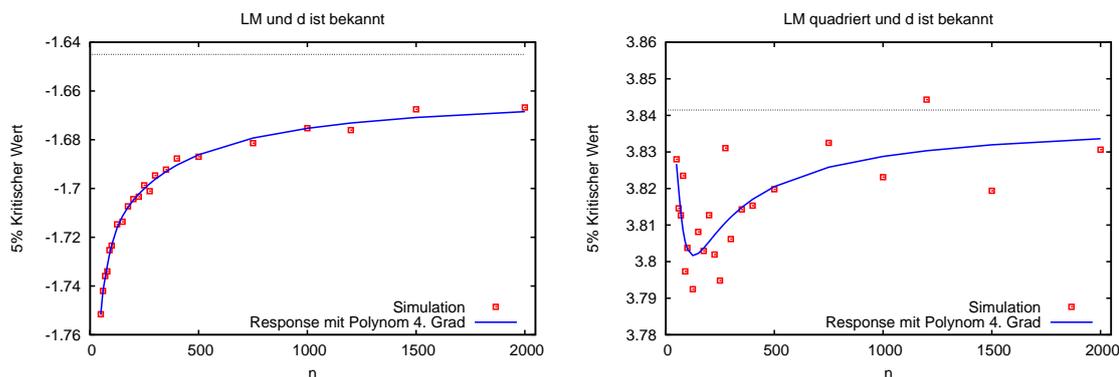


Abb. 4.17: Vergleich der simulierten kritischen Werte für  $\alpha = 5\%$  mit den jeweiligen Response-Surface-Funktionen

Stichprobenumfang  $n = 2000$  unter der Annahme, dass wir mit  $d_x = 0$  den wahren Integrationsgrad des Störterms kennen. In Abbildung 4.17 vergleichen wir nun die Simulationsergebnisse für die kritischen Werte mit  $\alpha = 5\%$  mit der jeweils geschätzten Response-Surface-Funktion. In den beiden Graphen ist auch jeweils das 5%-Quantil der Standardnormal- und der  $\chi_1^2$ -Verteilung als gestrichelte Linie eingezeichnet. Für die LM-Statistik können wir erkennen, dass die kritischen Werte eine Struktur in  $n$  aufweisen, die durch die Response-Surface-Funktion gut erklärt werden kann. Das Bestimmtheitsmaß beträgt in diesem Fall 99,31% und deutet auf eine gute Anpassung hin. Bei der quadrierten LM-Statistik hingegen zeigt sich ein anderes Bild. Die Struktur scheint hier zwar nicht zufällig, kann aber nicht wirklich gut durch eine Response-Surface-Funktion erklärt werden. Hier beträgt die Güte der Anpassung nur noch 56,65% und würde auf 60,27% steigen, wenn wir ein Polynom sechsten Grades von  $n^{-1}$  als Response-Surface-Funktion wählen würden. Ein noch höheres Polynom könnte auch nicht berücksichtigt werden, da dann die Regressor-matrix keinen vollen Spaltenrang mehr besitzt und somit die einzelnen Parameter nicht mehr bestimmbar sind. Aus diesem Grund und weil die einzelnen simulierten kritischen Werte sehr nahe am 5%-Quantil der  $\chi_1^2$ -Verteilung sind, wählen wir für die quadrierte LM-Statistik die kritischen Werte aus der asymptotischen Verteilung. Dies gilt auch für die anderen kritischen Quantile, auf deren Darstellung wir hier aber verzichten wollen.

Wir beschließen diesen Abschnitt mit den Parametern der Response-Surface-Funktion für die einseitige LM-Statistik, die in Tabelle 4.13 abgetragen sind. Die kritischen Werte ergeben sich einfach durch Einsetzen des jeweils vorliegenden Stichprobenumfangs  $n$  in die Gleichung (4.27) mit den Parametern aus der Tabelle 4.13.

Tab. 4.13: Response-Surface-Parameter für LM

$\alpha$	$\gamma_0$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$R^2$	$\hat{\sigma}_{\text{Resid}}$
10%	-1,3020	-11,907	931,59	-36253	473680	99,40%	0,0091
5%	-1,6634	-13,376	1329,8	-73753	1486818	99,31%	0,0108
1%	-2,3509	-7,5926	-163,44	38852	-1330806	98,79%	0,0205

## 4.4 Wald-Tests

Neben den LM-Tests existieren auch eine Reihe von Wald-Tests zur Überprüfung des Integrationsgrads. Im Gegensatz zu diesen liefern Wald-Tests eine Punktschätzung des Integrationsgrads  $d$ , während die LM-Tests unter Gültigkeit einer Nullhypothese berechnet werden. Im Folgenden werden wir zwei Vertreter dieser Gruppe von Tests näher betrachten.

### 4.4.1 Der fraktionale Dickey-Fuller-Test

Dolado, Gonzalo und Mayoral (weiterhin DGM, 2002) schlugen eine Verallgemeinerung des bekannten Dickey-Fuller-Tests vor. Dieser ist unter Vernachlässigung der fraktionalen Integration darauf ausgelegt, die Nullhypothese  $I(1)$  gegen die Alternative  $I(0)$  zu testen. Die Idee von DGM (2002) ist nun, das Testproblem zunächst zu verallgemeinern zu einem Test  $I(d_0)$  gegen  $I(d_1)$ , wobei  $d_0 > d_1$  sein soll. Ihre vorgeschlagene Teststatistik ist dann in Analogie zu Dickey und Fuller (1979) die t-Statistik für  $\phi = 0$  der Testregression

$$\Delta^{d_0} y_t = \phi \Delta^{d_1} y_{t-1} + v_t \quad \text{für } t = 2, \dots, n. \quad (4.28)$$

Unter der Nullhypothese  $H_0 : \phi = 0$  ist  $y_t \sim I(d_0)$  und unter der Alternative  $H_1 : \phi < 0$  haben wir  $y_t \sim I(d_1)$ . Im Hinblick auf eine empirische Anwendung analysieren DGM (2002) die wahrscheinlich interessanteste Konstellation  $d_0 = 1$  und  $d_1 < 1$ . Hiermit lässt sich dann testen, ob  $y_t$  einen Einheitswurzelprozess darstellt gegen die Alternative der fraktionalen Integration vom Grad  $d_1$ . Die Testregression (4.28) wird dann zu

$$\Delta y_t = \phi \Delta^{d_1} y_{t-1} + v_t \quad (4.29)$$

Tab. 4.14: Behandlung von deterministischen Bestandteilen für  $y_t$  für den Test von  $H_0 : d = 1$  gegen  $H_1 : d < 1$

Situation	Annahme für $y_t$	Bereinigung durch Aufnahme von
N	keine determ. Bestandteile vorhanden	keine Bereinigung notwendig.
LT	Linearer Trend	einer Konstanten in die Testregression.
LTB	Linearer Trend mit Bruch in $\tau = T_b/n$	einer Konstanten und der Dummy-Variable $\mathbb{I}(t \geq T_b)$ .
QT	Quadratischer Trend	einer Konstanten und einem linearen Trend in der Testregression.

und die fraktionale Dickey-Fuller-Statistik (weiterhin: FDF-Statistik) berechnet sich als t-Statistik für  $\phi = 0$  folglich mit

$$t_{\phi=0} = \frac{\sum_{t=2}^n \Delta y_t \Delta^{d_1} y_{t-1}}{\hat{\sigma}_v [\sum_{t=2}^n (\Delta^{d_1} y_{t-1})^2]^{1/2}}, \quad (4.30)$$

wobei  $\hat{\sigma}_v$  die geschätzte Standardabweichung der Störterme aus (4.29) darstellt. Für  $t_{\phi=0}$  konnten DGM (2002) die folgenden asymptotischen Eigenschaften aufzeigen:

$$t_{\phi=0} \xrightarrow{a} \frac{\int_0^1 W_{-d_1}(r) dB(r)}{\left(\int_0^1 W_{-d_1}^2(r) dr\right)^{1/2}} \quad \text{für } 0 \leq d_1 < 0,5 \quad (4.31)$$

$$t_{\phi=0} \xrightarrow{a} N(0; 1) \quad \text{für } 0,5 \leq d_1 < 1. \quad (4.32)$$

Die Verteilung der der FDF-Teststatistik hängt damit vom Integrationsgrad  $d_1$  unter der Gegenhypothese ab. DGM (2002) schlagen bei der Wahl des Parameters  $d_1$  für die Testregression vor, dass man einfach den geschätzten Integrationsgrad  $\hat{d}$  von  $y_t$  einsetzen sollte. Diese Sichtweise wird von Lobato und Velasco (2006) jedoch kritisiert. Sie charakterisieren den Parameter  $d_1$  nicht als alternativen Integrationsparameter zu  $d = 1$  für den zu untersuchenden Prozess  $y_t$ , sondern sehen vielmehr darin einen Tuning-Parameter, der die Macht des Tests determiniert. Generell postulieren Lobato und Velasco (2006), dass die Anwender des Tests die Korrelation zwischen  $\Delta y_t$  und  $\Delta^{d_1} y_{t-1}$  maximieren sollten, da dies auch die Macht des Tests maximiert. Sie konnten dabei in Simulationen zeigen, dass dies genau bei der Wahl  $d_1 = 0,717d - 0,03$  der Fall ist. Empirisch lässt sich diese Regel anwenden, wenn wir für  $d = \hat{d}$  wiederum ein Schätzergebnis für den Integrationsgrad von  $y_t$  einsetzen. Eine weiterer Kritikpunkt von Lobato und Velasco (2006) greift die Formulierung der Testregression (4.29) an. Sollte nämlich  $y_t$  durch  $\Delta^d y_t = \varepsilon_t$  mit  $\varepsilon_t \sim IID(0; \sigma_\varepsilon^2)$

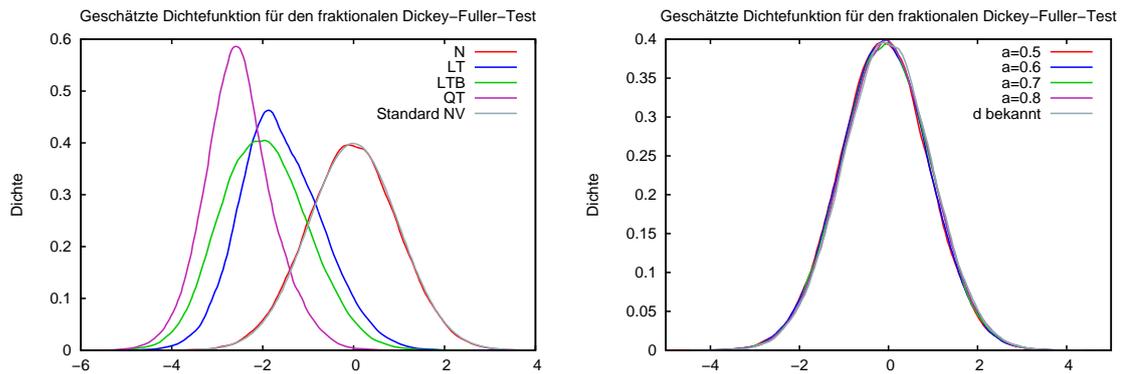


Abb. 4.18: Geschätzte Dichtefunktion des fraktionalen Dickey-Fuller-Test mit bekanntem und unbekanntem  $d$

generiert werden und wenn man annimmt, dass  $d$  bekannt sei und  $d_1 = d$  setzt, dann ergibt sich bei der Betrachtung der ersten Differenz von  $y_t$

$$\begin{aligned} \Delta y_t &= \Delta^{1-d} \varepsilon_t = \varepsilon_t + \underbrace{(d-1)\varepsilon_{t-1}}_{(d-1)\Delta^d y_{t-1}} + 0,5d(d-1)\varepsilon_{t-2} + \dots \\ &= \phi \Delta^d y_{t-1} + \underbrace{\varepsilon_t + 0,5d(d-1)\varepsilon_{t-2} + \dots}_{v_t}, \end{aligned} \quad (4.33)$$

dass mit  $\phi = d-1$  der Störterm  $v_t$  der Regression (4.29) tendenziell autokorreliert ist und damit  $\phi$  nicht mehr effizient mit der einfachen KQ-Methode geschätzt werden kann. Es müsste deswegen eine effizientere Testmöglichkeit für die Überprüfung der Einheitswurzelhypothese existieren, auf die wir im folgenden Abschnitt eingehen werden. DGM (2002) erweitern ihre vorgeschlagene Testregression (4.29) zu

$$\Delta y_t = \phi \Delta^{d_1} y_{t-1} + \sum_{i=1}^p \theta_i \Delta y_{t-i} + v_t, \quad (4.34)$$

um auch eine möglicherweise vorhandene AR-Struktur in den Störtermen  $v_t$  zu berücksichtigen. Weiterhin lassen sich deterministische Komponenten in  $y_t$  bereinigen, indem wir weitere Regressoren in (4.34) einbeziehen. Eine Auflistung der deterministischen Bestandteile und deren Berücksichtigung in der Testregression ist in der Tabelle 4.14 gegeben. Im Folgenden wollen wir nun wieder Testverteilungen simulieren und gegebenenfalls kritische Werte mit einer Response-Surface-Funktion darstellen. Dazu simulieren wir einfache Random-Walk-Prozesse  $\Delta y_t = \varepsilon_t$  mit  $\varepsilon_t \sim N(0; 1)$  für  $t = 1, \dots, n$  und wenden die FDF-Statistik mit den verschiedenen in Tabelle 4.14 genannten deterministischen Bereinigsmöglichkeiten. In Tabelle 4.18 befinden sich

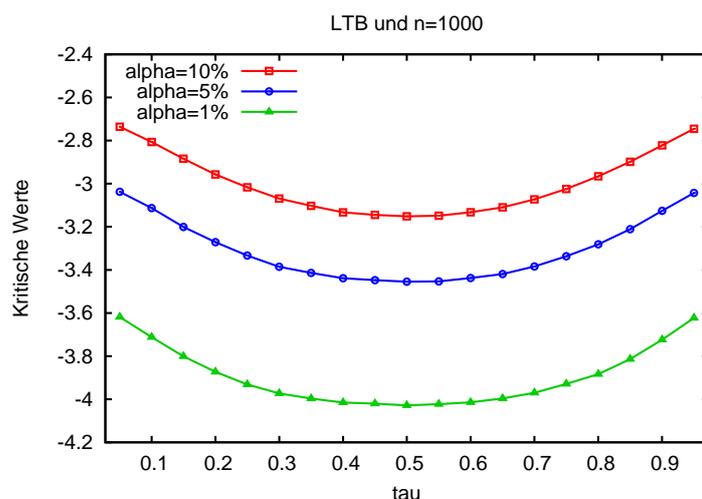


Abb. 4.19: Kritische Werte für die Situation mit Trendbruch in Abhängigkeit von  $\tau$

nun zum einen die einzelnen geschätzten Dichtefunktionen mit Stichprobenumfang  $n = 1000$  für die verschiedenen Trendoptionen bei der Annahme, dass wir den Integrationsgrad von  $y_t$ , also  $d = 1$ , kennen und damit  $d_1 = 0,717 - 0,03 = 0,687$  setzen können. Für die Situation LTB, also den Fall, dass  $y_t$  einen linearen Trend mit Strukturbruch hat, fixieren wir zunächst  $\tau = 0,5$ . Man kann bei den geschätzten Dichtefunktionen gut erkennen, dass sich die einzelnen Verteilungen nach links verschieben und somit auch für die einzelnen Trendoptionen unterschiedliche kritische Werte ergeben, selbst wenn man das Irrtumsniveau  $\alpha$  fixiert. Somit ist es nötig, dass wir für jede der Trendoptionen unterschiedliche kritische Werte bestimmen. Zudem sollte noch bemerkt werden, dass für  $n = 1000$  nur für den Fall ohne jeglichen Trend sich die Normalverteilung bei jeder üblichen Irrtumswahrscheinlichkeit nicht ablehnen lässt. In der linken Graphik von Abbildung 4.18 haben wir auch noch die Dichtefunktion der Standardnormalverteilung eingefügt, die sich relativ gut mit der Dichtefunktion des FDF-Tests ohne Trend deckt. Für diese Situation untersuchen wir auch noch den Fall, dass der Integrationsparameter unbekannt sei und wir diesen mit dem GPH-Schätzer bei Verwendung der Bandbreite  $m = n^a$  schätzen. Die jeweils geschätzten Dichtefunktionen sind im rechten Schaubild der Abbildung 4.18 dargestellt. Hier kann man zur Erkenntnis gelangen, dass die Unsicherheit bezüglich des Integrationsparameters keinen großen Effekt auf die Testverteilung der FDF-Statistik hat. Dies steht tendenziell im Gegensatz zur Verteilung des LM-Tests, wo wir hingegen einen großen Einfluss dokumentieren konnten. Betrachten wir aber nun die einzelnen simulierten kritischen Werte, die in Tabelle 4.15 auch für die Situationen mit Trendbereinigung dargestellt sind, dann kann man aber doch erkennen,

Tab. 4.15: Übersicht der kritischen Werte für  $\alpha = 5\%$  bei unbekanntem und bekannten Integrationsparametern für die FDF-Statistik

	$a = 0,5$	$a = 0,6$	$a = 0,7$	$a = 0,8$	$d$ bekannt
N	-1,7515	-1,7261	-1,7297	-1,6982	-1,6659
LT	-2,9010	-2,9336	-2,9461	-2,9631	-2,9850
LTB	-3,3651	-3,4086	-3,4186	-3,4287	-3,4507
QT	-3,6436	-3,6712	-3,6903	-3,6990	-3,7241

dass die Kenntnis von  $d$  durchaus einen Effekt auf die kritischen Werte. Der Unterschied ist zwar nicht ganz so groß wie bei dem LM-Test, aber durchaus vorhanden. Deswegen entscheiden wir uns auch hier für die Verwendung von bekannten Integrationsparametern zur Simulation von kritischen Werten und versuchen, diese nicht bei der Simulation zu schätzen, da hierfür wiederum die Bandbreite fixiert werden müsste. Bevor wir nun kritische Werte simulieren, können wir uns auch noch Gedanken machen, wie wir den Fall des Trendbruchs evtl. vereinfachen könnten. Es ist nämlich zu erwarten, dass für diesen Fall die kritischen Werte auch vom Zeitpunkt, wann der Trendbruch auftritt, abhängt. Um dies genauer zu untersuchen, sind in der Abbildung 4.19 die simulierten kritischen Werte in Abhängigkeit des relativen Bruchzeitpunkts  $\tau = T_b/n$  für  $n = 1000$  bei jeweils 100.000 Wiederholungen dargestellt. Ähnlich wie für den Fall des Tests auf Vorliegen eines Trendbruchs, können wir hier feststellen, dass wir bei  $\tau = 0,5$  ein Extremum für die kritischen Werte haben. Deswegen scheint es sinnvoll, dass wir für unsere weiteren Simulationen einfach  $\tau = 0,5$  annehmen, was uns zum einen Rechenkapazität erspart, aber zum anderen, was noch wichtiger erscheint, es vereinfacht auch die Modellierung durch eine Response-Surface-Funktion, da wir in dieser dann  $\tau$  unberücksichtigt lassen können. Bei empirischen Anwendungen werden wir dann aber das Problem haben, dass der FDF-Test für den Fall  $\tau \neq 0,5$  unterdimensioniert ist, d.h. wir werden unter Gültigkeit der Nullhypothese bei einem fixierten  $\alpha$  eine kleinere Ablehnungsrate von  $H_0$  als  $\alpha$  haben. Die Testentscheidung wird in diesem Fall evtl. zu konservativ sein. Als Response-Surface-Funktion für den FDF-Test wollen wir ein Polynom fünften Grades von  $n^{-1}$  verwenden in der Form

$$CV_\alpha(n) = \gamma_0 + \gamma_1 n^{-1} + \gamma_2 n^{-2} + \gamma_3 n^{-3} + \gamma_4 n^{-4} + \gamma_5 n^{-5} + \text{Resid.} \quad (4.35)$$

Dies lässt sich in diesem Fall rechtfertigen, wenn wir verschiedene Modellselektionskriterien verwenden, um die Ordnung des Polynoms zu bestimmen. Dazu kon-

Tab. 4.16: Wahl des Polynomgrades von  $n^{-1}$  für die Response-Surface-Funktion der FDF-Statistik ohne Trendbereinigung für  $\alpha = 5\%$

Polynom-grad	$R^2$	$\hat{\sigma}_{\text{Resid}}$	$\hat{\gamma}_0$	99%-Konfidenzintervall für $\gamma_0$	LR-Test	AIC	BIC	HQC
1	82,96%	$4,28 \cdot 10^{-3}$	-1,6692	[-1,6745;-1,6640]		-167,5	-165,4	-167,1
2	94,39%	$2,52 \cdot 10^{-3}$	-1,6640	[-1,6684;-1,6597]	23,351***	-188,9	-185,7	-188,2
3	97,56%	$1,71 \cdot 10^{-3}$	-1,6607	[-1,6637;-1,6578]	17,495***	-204,4	-200,2	-203,5
4	99,08%	$1,09 \cdot 10^{-3}$	-1,6581	[-1,6598;-1,6564]	20,433***	-222,8	-217,6	-221,7
5	99,25%	$1,01 \cdot 10^{-3}$	-1,6568	[-1,6591;-1,6545]	4,3019**	<b>-225,1</b>	<b>-218,8</b>	<b>-223,8</b>
6	99,26%	$1,04 \cdot 10^{-3}$	-1,6573	[-1,6612;-1,6534]	0,3356	-223,4	-216,1	-221,9

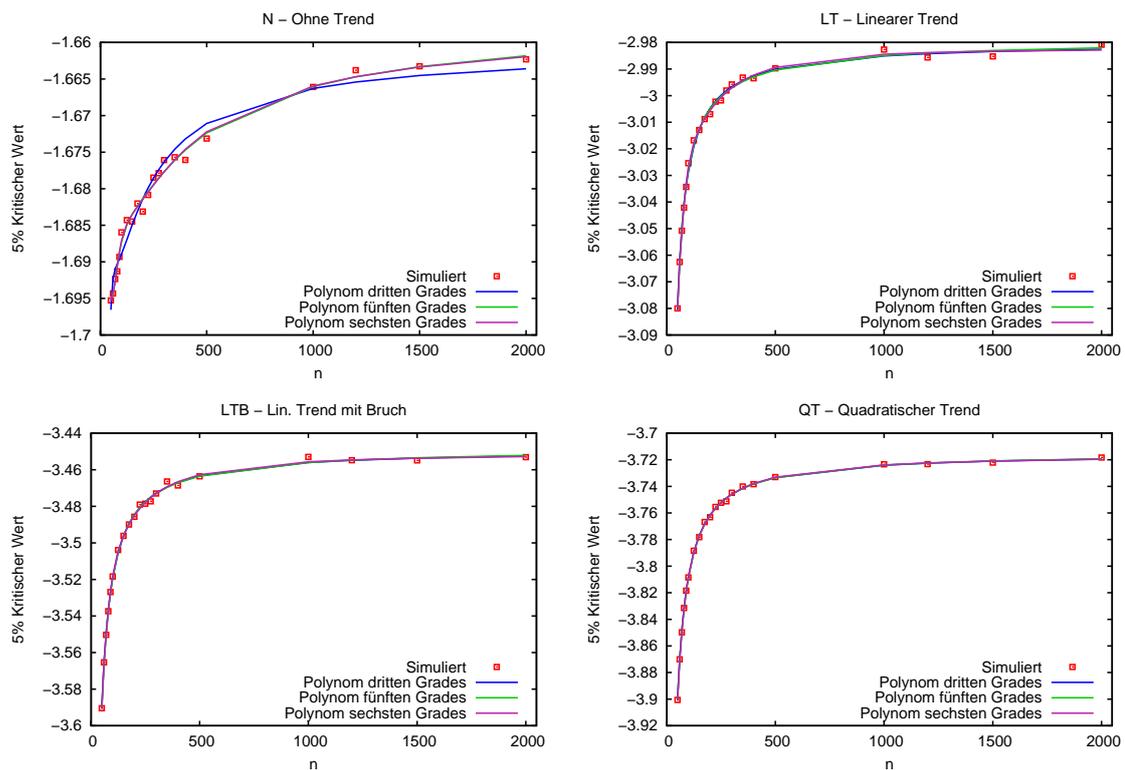


Abb. 4.20: 5% Kritische Werte des fraktionalen Dickey-Fuller-Tests in Abhängigkeit von  $n$

zentrieren wir uns auf die Situation ohne Trendbereinigung und schätzen sämtliche Polynomgrade von  $n^{-1}$  bis zum sechsten Grad für den kritischen Wert für  $\alpha = 5\%$ . Die dazugehörigen Schätzergebnisse sind in der Tabelle 4.16 zusammengefasst. Hier berichten wir über die jeweilige Güte der Anpassung der einzelnen geschätzten Response-Surface-Funktionen und die sich ergebenden asymptotischen kritischen Werte, die den Schätzwerten von  $\gamma_0$  entsprechen. Dazu haben wir noch das geschätzte 99%-Konfidenzintervall für die asymptotischen 5%-Kritische Werte, das mit der Heteroskedastizität-konsistenten geschätzten Standardabweichung von  $\hat{\gamma}_0$  nach MacKinnon und White (1985) bestimmt wurde. Als Modellselektionskriterien verwenden wir das AIC, BIC und das HQC, die allesamt ein Polynom fünften Grades von  $n^{-1}$  präferieren. Auch die Betrachtung der Likelihood-Ratio-Teststatistik (LR-Test) ist dieser Wahl gewogen, da eine Polynom sechsten Grades nicht mehr einem Polynom fünften Grades vorgezogen werden kann. In Tabelle 4.17 geben wir nun die geschätzten Parameter der Response-Surface-Funktion (4.35) an und in Abbildung 4.20 stellen wir die geschätzten Response-Surface-Funktionen für die einzelnen Trendsituationen dar. In jeder der Einzelabbildungen ist außerdem noch die Möglichkeit eines Polynom dritten und sechsten Grades als Response-Surface-Funktion dargestellt. Hier kann man erkennen, dass die Struktur vor allem für die Situation ohne Trendbereinigung einer Modellierung mit einem Polynom fünften Grades bedarf. Für die restlichen Situationen werden auch mit dem Polynom dritten Grades die simulierten kritischen Werte relativ gut getroffen.

#### 4.4.2 Der effiziente Wald-Test nach Lobato und Velasco (2007)

Aufbauend auf der Kritik von Lobato und Velasco (2006) am FDF-Test schlugen Lobato und Velasco (2007) einen effizienten Wald-Test für fraktionale Einheitswurzeln vor. In ihrem Kontext soll wieder unter der Nullhypothese  $d = 1$  gegen eine fraktionale Alternative  $d < 1$  getestet werden. Ausgangspunkt ist die Betrachtung der ersten Differenz eines fraktional integrierten Rauschens  $\Delta^d y_t = \varepsilon_t$  mit  $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$  und einem beliebigen  $d$ :

$$\Delta y_t = (\Delta - \Delta^d)y_t + \underbrace{\Delta^d y_t}_{\varepsilon_t} = (1 - \Delta^{d-1})\Delta y_t + \varepsilon_t. \quad (4.36)$$

Tab. 4.17: Geschätzte Response-Surface-Funktionen für die kritischen Werte des fraktionalen Dickey-Fuller-Tests

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
	N	LT	LTB	QT	N	LT	LTB	QT	N	LT	LTB	QT
$\gamma_0$	-1,2953 (0,0009)	-2,6894 (0,0018)	-3,1460 (0,0012)	-3,4442 (0,0012)	-1,6568 (0,0008)	-2,979 (0,0019)	-3,448 (0,0017)	-3,7145 (0,0013)	-2,3397 (0,0024)	-3,5383 (0,0016)	-4,0161 (0,0025)	-4,2399 (0,0037)
$\gamma_1$	-10,921 (1,3279)	-3,7749 (1,7553)	-6,0861 (1,4489)	-7,484 (1,4865)	-10,987 (1,1255)	-6,2436 (1,9284)	-8,3038 (2,0153)	-9,6427 (1,5726)	-9,4521 (3,4533)	-9,0764 (2,6535)	-10,41 (2,9328)	-15,414 (3,6984)
$\gamma_2$	2115,3 (480,02)	158,71 (549,41)	370,06 (520,85)	216,03 (511,57)	1925,1 (386,73)	310,83 (636,39)	358,17 (658,68)	87,286 (529,16)	1359,4 (1281)	-392,8 (931,25)	-746,15 (975,12)	-161,89 (1238,8)
$\gamma_3$	-2,14398 (66086)	-13765 (71919)	-42188 (75140)	-34007 (70411)	-176944 (50632)	-20950 (86176)	-39915 (85832)	-9708 (70044)	-151776 (182577)	81474 (126842)	100239 (129606)	35874 (163979)
$\gamma_4$	10148946 (3802815)	453430 (4117365)	2163996 (4621884)	2200719 (4148458)	7595595 (2797597)	250972 (5039066)	1966786 (4840512)	412209 (3954876)	8202879 (10923820)	-5827964 (7374475)	-5727501 (7412713)	-2653928 (9230814)
$\gamma_5$	-177891569 (76928228)	-5207925 (84628571)	-40189968 (99984734)	-48418498 (86708964)	-121401551 (54875779)	8467064 (104963061)	-35882434 (98093722)	-5123116 (79653190)	-162668025 (229758746)	133467777 (152433648)	112326855 (151424226)	62571160 (185220578)
$R^2$	98,88%	99,61%	99,81%	99,93%	99,25%	99,74%	99,88%	99,96%	98,63%	99,88%	99,92%	99,92%
$\hat{\sigma}_{Resid}$	$1,08 \cdot 10^{-3}$	$1,31 \cdot 10^{-3}$	$1,31 \cdot 10^{-3}$	$1,19 \cdot 10^{-3}$	$1,01 \cdot 10^{-3}$	$1,63 \cdot 10^{-3}$	$1,62 \cdot 10^{-3}$	$1,24 \cdot 10^{-3}$	$2,54 \cdot 10^{-3}$	$2,17 \cdot 10^{-3}$	$2,38 \cdot 10^{-3}$	$2,94 \cdot 10^{-3}$

N steht für das Vorliegen von keinem Trend, LT einem linearen Trend, LTB einem linearen Trend mit Bruch und QT einem quadratischen Trend. Heteroskedastizität-robust geschätzte Standardabweichungen sind in Klammern.

Lobato und Velasco (2007) motivieren über die Betrachtung

$$(1 - \Delta^{d-1})\Delta y_t = (d-1)\Delta y_{t-1} + \sum_{j=2}^{t-1} \pi_j(d-1)\Delta y_{-j}$$

mit  $\pi_j(d-1) = \frac{\Gamma(j-d+1)}{\Gamma(j+d+1)\Gamma(1-d)},$  (4.37)

dass der Regressor der Testregression

$$\Delta y_t = \varphi_2(\Delta^{d-1} - 1)\Delta y_t + \varepsilon_t \quad (4.38)$$

im Endeffekt frei von  $\Delta y_t$  ist und dass dies sowohl unter  $H_0 : d = 1$  als auch unter  $H_1 : d < 1$  gilt. Unter der Gültigkeit von  $H_0$  geht der Koeffizient  $\varphi_2$  gegen 0 und unter  $H_1$  gegen  $-1$ . Ein möglicher Test zur Überprüfung von  $H_0$  könnte mit dem Einsetzen von  $d = d_1$  die t-Statistik für  $\varphi_2 = 0$  darstellen. Problematisch ist aber hierbei, dass der Regressor für  $d_1 = 1$  nicht existiert und für  $d_1 > 1$  mit einem Vorzeichenwechsel von  $\varphi_2$  verbunden ist. Lobato und Velasco (2007) schlagen viel mehr darüber hinaus vor, die Testregression neu zu formulieren mit

$$\Delta y_t = \phi_2 z_{t-1}(d_1) + u_t \quad \text{wobei} \quad z_{t-1}(d_1) = \frac{(\Delta^{d_1-1} - 1)}{1 - d_1} \Delta y_t, \quad (4.39)$$

um diese Probleme zum Teil zu umgehen. Der Regressor  $z_{t-1}$  existiert zwar auch hier im Fall  $d_1 = 1$  nicht, aber das Vorzeichen von  $\phi_2$  ändert sich nicht für  $d_1 > 1$ . Der Regressor wird mit  $t-1$  indiziert, da er wiederum eine Transformation von  $\Delta y_{t-1}$  darstellt. Interessant ist hierbei der Fall  $d \rightarrow 1$ , da sich gemäß der Regel nach l'Hôpital das Verhältnis  $(\Delta^{d-1} - 1)/(1 - d)$  zum fraktionalen Filter  $(1 - L)^{-\delta}$  an der Stelle  $\delta = 0$  entwickelt. Dies wiederum ergibt den linearen Filter  $-\ln(1 - L) = \sum_{j=1}^{\infty} j^{-1} L^j$ . Somit entspricht die Testregression in diesem Fall genau dem LM-Test in der Version von Breitung und Hassler (2002).

Im Weiteren haben auch Lobato und Velasco (2007) eine Erweiterung des Tests vorgeschlagen, um mögliche ARFIMA-Prozesse untersuchen zu können. Diese Prozesse lassen sich mit der Gleichung  $\theta(L)\Delta^d y_t = \varepsilon_t$  charakterisieren. Untersuchen wir nun äquivalent zum fraktional integrierten Rauschen die absolute Differenz von  $y_t$ :

$$\begin{aligned} \theta(L)\Delta y_t &= \theta(L)(1 - \Delta^{d-1})\Delta y_t + \varepsilon_t \\ \Delta y_t &= \theta(L)(1 - \Delta^{d-1})\Delta y_t + [1 - \theta(L)]\Delta y_t + \varepsilon_t. \end{aligned} \quad (4.40)$$

Auf Basis dieser Gleichung lässt sich ähnlich wie im einfachen Standardfall folgende Testregression motivieren:

$$\Delta y_t = \phi_2[\theta(L)(\Delta^{d-1} - 1)\Delta y_t] + \sum_{j=1}^p \theta_j \Delta y_{t-j} + u_t, \quad (4.41)$$

die wiederum die gleichen Probleme mit sich trägt wie im Standardfall. Auch hier schlagen Lobato und Velasco (2007) die Testregression mit dem bereits definierten Regressor  $z_{t-1}(d)$  vor:

$$\Delta y_t = \phi_2[\theta(L)z_{t-1}(d)] + \sum_{j=1}^p \theta_j \Delta y_{t-j} + u_t. \quad (4.42)$$

Da es sich nun hier um ein nicht-lineares Regressionsmodell handelt, ist die Anwendung in der Realität schwieriger als im Standardfall. Dies gilt besonders dann, wenn die Parameter  $\{\theta_1, \theta_2, \dots, \theta_p\}$  und/oder  $p$  unbekannt sind. Lobato und Velasco (2007) schlagen dabei folgendes Zwei-Stufen-Verfahren vor:

Im ersten Schritt wird folgende KQ-Regression durchgeführt, um die Parameter  $\{\theta_1, \theta_2, \dots, \theta_p\}$  zu schätzen:

$$\Delta^{\hat{d}} y_t = \sum_{j=1}^p \theta_j \Delta^{\hat{d}} y_{t-j} + u_t. \quad (4.43)$$

Diese Schätzung setzt voraus, dass wir eine konsistente Schätzung  $\hat{d}$  für  $d$  haben.

Im zweiten Schritt wird mit den Ergebnissen  $\{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p\}$  aus dem ersten die Testregression

$$\Delta y_t = \phi_2[\hat{\theta}(L)z_{t-1}(\hat{d})] + \sum_{j=1}^p \tilde{\theta}_j \Delta y_{t-j} + u_t \quad (4.44)$$

durchgeführt und der Parameter  $\phi_2$  auf Signifikanz geprüft.

Nun wollen wir uns wieder Gedanken über kritische Werte machen, die wir zur empirischen Anwendung benötigen. Zuerst wollen wir die geschätzten Dichtefunktionen für  $n = 1000$  der effizienten Wald-Statistik in der Abbildung 4.21 betrachten. Hier vergleichen wir, ähnlich wie beim fraktionalen Dickey-Fuller-Test, die Verteilungen bei Berücksichtigung von deterministischen Trends. Diese lassen sich aber nicht wie bei dem FDF-Test durch Hinzunahme als Regressor in die Testregression aufnehmen, sondern müssen in einer Hilfsregression vor Anwendung der Testregression beseitigt werden. Die einzelnen Verteilungen unterscheiden sich nicht so stark bei

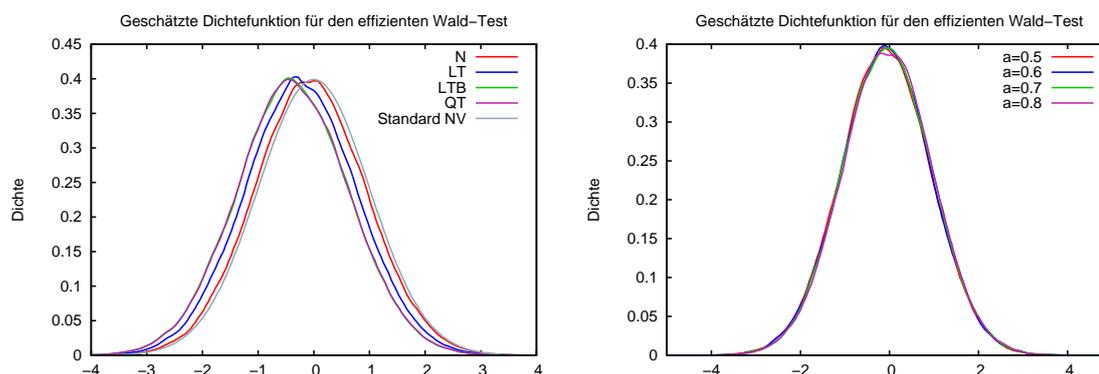


Abb. 4.21: Geschätzte Dichtefunktion der effizienten Wald-Statistik mit unbekanntem  $d$

der FDF-Statistik, sind aber doch unterschiedlich und lassen sich deswegen nicht verallgemeinern. Die Schätzungen selbst basieren wiederum auf der Annahme, dass der wahre Integrationsparameter  $d = 1$  von  $y_t$  nicht bekannt sein soll, sondern mit dem GPH-Schätzer unter Verwendung der Bandbreite  $m = n^{0,8}$  bestimmt werden muss. Im rechten Graphen der Abbildung 4.21 vergleichen wir für den Fall ohne Trend den Effekt der Wahl des Bandbreitenparameters und können dabei erkennen, dass dieser scheinbar keine große Auswirkung auf die Verteilung hat.

Tab. 4.18: Übersicht der kritischen Werte für  $\alpha = 5\%$  bei unbekanntem Integrationsparameter für die effiziente Wald-Statistik und  $n = 1000$

	$a = 0,5$	$a = 0,6$	$a = 0,7$	$a = 0,8$
N	-1,7351	-1,7009	-1,6847	-1,6741
LT	-1,9182	-1,8597	-1,8454	-1,8224
LTB	-2,0945	-2,0111	-1,9701	-1,9435
QT	-2,0873	-2,0029	-1,9687	-1,9421

In Tabelle 4.18 haben wir auch die jeweils resultierenden kritischen Werte für das Irrtumsniveau  $\alpha = 5\%$  für die verschiedenen Trendoptionen und sehen, dass die kritischen Werte größer werden, falls wir die Bandbreite erhöhen und damit den Parameter  $d$  weniger volatil schätzen, aber die Unterschiede recht gering bleiben. Für die Simulationen zur Bestimmung der kritischen Werte fixieren wir die Bandbreite auf einem recht hohen Niveau mit  $m = n^{0,8}$  und schätzen die Teststatistik wiederum bis zum einem maximalen Beobachtungsumfang  $n = 2000$ . Septhon (2009) hat bereits Response-Surface-Funktionen für den effizienten Wald-Test geschätzt und verwendet hierfür aber mit dem exakten lokalen Whittle-Schätzer einen quasi ML-Schätzer, der sehr zeitaufwendig ist und recht viele Computerressourcen

Tab. 4.19: Wahl des Polynomgrades von  $n^{-1}$  für die Response-Surface-Funktion der NEW-Statistik ohne Trendbereinigung

Polynom-grad	$R^2$	$\hat{\sigma}_{\text{Resid}}$	$\hat{\gamma}_0$	99%-Konfidenzintervall für $\gamma_0$	LR-Test	AIC	BIC	HQC
1	83,83%	0,0098	-1,7290	[-1,7449;-1,7131]		-120,0	-118,1	-119,7
2	95,38%	0,0054	-1,7114	[-1,7231;-1,6997]	23,793***	-141,8	-139,0	-141,3
3	99,09%	0,0025	-1,7021	[-1,7121;-1,6921]	30,929***	-170,7	-167,0	-170,1
4	99,24%	0,0023	-1,6951	[-1,7033;-1,6868]	3,369*	-172,1	-167,4	-171,3
5	99,47%	0,0020	-1,6919	[-1,7012;-1,6825]	6,7924***	-176,9	-171,2	-175,9
6	99,56%	0,0019	-1,6882	[-1,7005;-1,6759]	3,766*	<b>-178,7</b>	<b>-172,1</b>	<b>-177,6</b>

Tab. 4.20: Vergleich der asymptotischen kritischen Werte mit Sephton (2009)

	$\alpha = 5\%$			$\alpha = 1\%$		
	Response	Sephton (2009)	Differenz	Response	Sephton (2009)	Differenz
N	-1,6951	-1,6983	0,1871%	-2,3727	-2,3772	0,1886%
LT	-1,8216	-1,8248	0,1756%	-2,5029	-2,5036	0,0294%
QT	-1,9302	-1,9495	0,9875%	-2,6052	-2,6341	1,0988%

in Anspruch nimmt. Er führt deswegen nur 50.000 Wiederholungen für jeden Beobachtungsumfang durch, was tendenziell klein erscheint. Wir wollen, wie bei der Ermittlung der kritischen Werte des FDF-Tests, insgesamt 1.000.000 Wiederholungen durchführen, was aufgrund des leichter berechenbaren GPH-Schätzers möglich ist. Diesen Schätzer wenden wir auf die Differenzen der simulierten  $I(1)$ -Prozesse an, was der Grundannahme folgt, dass wir die Nicht-Stationarität des simulierten Prozesses implizit kennen. Diese Annahme benötigt Sephton (2009) aufgrund des verwendeten ELW-Schätzers nicht.

Sephton (2009) verwendet ein Polynom vierten Grades von  $n^{-1}$  als Response-Surface-Funktion. In Tabelle 4.19 stellen wir die Modellselektionskriterien für die Situation ohne Trend (N) bis zu einer maximalen Polynomordnung von sechs dar. Hieraus können wir entnehmen, dass die Modellselektionskriterien tendenziell eine noch höhere Ordnung wählen würden, was wir in dieser Arbeit nicht machen möchten, da wir bei Betrachtung der simulierten Werte und der Anpassung eines Polynoms vierter Ordnung, was wir für sämtliche Trendoptionen in Abbildung 4.22 darstellen, durchaus zu dem Schluss kommen, dass ein Polynom vierter Ordnung ausreicht, um den Verlauf der kritischen Werte zu beschreiben. In den Graphiken der Abbildung 4.22 sind auch die Response-Surface-Funktion von Sephton (2009) enthalten und wir können dabei sehen, dass für den Fall ohne Trend ein konstanter Unterschied existiert zu unserer geschätzten Funktion. Asymptotisch nähern sich die Funktionen aber weiter an und wir kommen sogar fast zu den gleichen asymptotischen kritischen Werten. Diese wiederum sind nochmal in der Tabelle 4.20 darge-

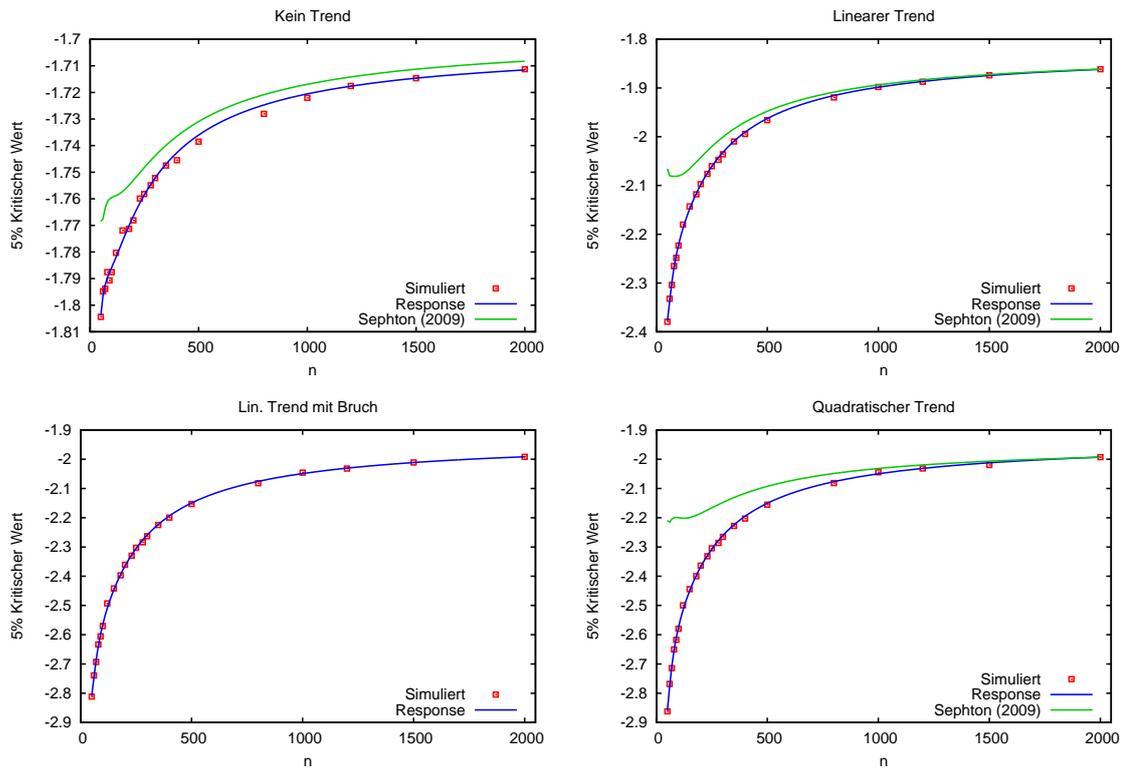


Abb. 4.22: 5%-kritische-Werte des effizienten Wald-Tests in Abhängigkeit von  $n$

stellt. Für den Fall des linearen und quadratischen Trends scheinen sich die beiden Response-Surface-Funktionen besser zu decken. Wir können hier lediglich größere Unterschiede bei kleineren Stichproben ausmachen, da hier die kritischen Werte von Sephton (2009) tendenziell größer geschätzt werden. Wir möchten diesen Abschnitt mit den geschätzten Parametern der Response-Surface-Funktion

$$CV_{\alpha}(n) = \gamma_0 + \gamma_1 n^{-1} + \gamma_2 n^{-2} + \gamma_3 n^{-3} + \gamma_4 n^{-4} + \text{Resid} \quad (4.45)$$

beschließen, die in Tabelle 4.21 enthalten sind.

Tab. 4.21: Geschätzte Response-Surface-Funktionen für die kritischen Werte des effizienten Wald-Tests

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
	N	LT	LTB	QT	N	LT	LTB	QT	N	LT	LTB	QT
$\gamma_0$	-1,3376 (0,0068)	-1,4603 (0,0152)	-1,5656 (0,024)	-1,5652 (0,0228)	-1,6951 (0,0029)	-1,8216 (0,0135)	-1,9292 (0,0246)	-1,9302 (0,0253)	-2,3727 (0,0045)	-2,5029 (0,0142)	-2,6103 (0,0272)	-2,6052 (0,0272)
$\gamma_1$	-20,1090 (3,9325)	-79,3635 (8,764)	-127,2976 (13,722)	-127,7670 (13,178)	-24,241 (3,9114)	-83,8198 (7,9678)	-130,3380 (14,034)	-130,1083 (14,646)	-30,1480 (3,1296)	-88,1381 (8,2941)	-139,3449 (15,095)	-142,9468 (15,297)
$\gamma_2$	2055,9 (730,73)	6836,5 (1615)	11213 (2523,4)	11178 (2447,7)	2537,5 (701,23)	7368,3 (1510,6)	11113 (2570,8)	10978 (2713,1)	3159,4 (625,36)	6917,8 (1541,9)	11590 (2654,6)	12100 (2773,6)
$\gamma_3$	-98338 (53152)	-331024 (116588)	-564117 (182040)	-560697 (178080)	-123008 (48830)	-369722 (111643)	-546841 (183998)	-541568 (196403)	-169377 (45644)	-322168 (110479)	-572426 (182714)	-627929 (197920)
$\gamma_4$	1746414 (1295688)	6104162 (2826987)	10843488 (4411739)	10644098 (4343248)	2159236 (1144123)	7065974 (2751342)	10351399 (4423715)	10090159 (4767551)	3346295 (1086241)	5856642 (2643397)	11024479 (4259261)	12293452 (4760165)
$R^2$	99,77%	99,92%	99,93%	99,93%	99,24%	99,93%	99,92%	99,93%	99,02%	99,91%	99,93%	99,94%
$\sigma_{R_{\text{resid}}}$	0,0022	0,0042	0,0063	0,0063	0,0023	0,0042	0,0069	0,0072	0,0042	0,0055	0,0075	0,0074

N steht für das Vorliegen von keinem Trend, LT einem linearen Trend, LTB einem linearen Trend mit Bruch und QT einem quadratischen Trend. Heteroskedastizität-robust geschätzte Standardabweichungen sind in Klammern.

## 4.5 Varianzverhältnistest-Ansatz

Die bisherigen vorgestellten Tests hatten alle gemeinsam, dass ein Lag-Parameter  $p$  zu bestimmen war, wenn der zu untersuchende Prozess einen ARFIMA-Prozess darstellte. Deswegen handelt es sich bei diesen Tests um parametrische Teststatistiken. Einen nicht-parametrischen Test für das Vorliegen einer Einheitswurzel (d.h.  $d = 1$ ) wurde von Nielsen (2008) vorgeschlagen, welcher von Nielsen (2010) für fraktional integrierte Prozesse erweitert wurde. Dieser basiert auf dem Konzept des Varianzverhältnistests (Variance-Ratio-Test, weiterhin: VR-Test). Man stelle sich einen univariaten Prozess der Form

$$y_t = \Delta^{-d}\varepsilon_t, \quad \text{mit } d > 1/2, \quad t = 1, \dots, n \quad (4.46)$$

vor und dessen fraktionale partielle Summe

$$\tilde{y}_t = \Delta^{-d_1}y_t, \quad \text{mit } d_1 > 0, \quad t = 1, \dots, n. \quad (4.47)$$

Die partielle Summe  $\tilde{y}_t$  hat nach dieser Summation einen Integrationsgrad von  $d+d_1$ . Für den Fall  $d > 1/2$  und unter gewissen Regularitätsvoraussetzungen für  $\varepsilon_t$  gilt folgende Konvergenz für  $n \rightarrow \infty$  und  $\sigma_y > 0$  :

$$T^{1/2-d}y_{[sn]} \Rightarrow \sigma_y W_d(s), \quad 0 < s \leq 1, \quad \sigma_y > 0. \quad (4.48)$$

$W_d(s)$  umschreibt hierbei eine fraktionale Standard-Brownsche Bewegung mit dem Integrationsparameter  $d$ , welche von Marinucci und Robinson (1999) definiert wurde. Der Störparameter  $\sigma_y$  in dieser Darstellung hängt von der jeweiligen ARMA-Spezifikation von  $\varepsilon_t$  ab. Die Konvergenz gilt dementsprechend auch für  $\tilde{y}_t$ . Somit können wir für  $n \rightarrow \infty$  die Verteilung des zweiten unzentrierten Moments von  $y_t$  und  $\tilde{y}_t$  angeben mit:

$$n^{-2d} \sum_{t=1}^n y_t^2 \xrightarrow{a} \sigma_y^2 \int_0^1 W_d(s)^2 ds \quad (4.49)$$

$$n^{-2(d+d_1)} \sum_{t=1}^n \tilde{y}_t^2 \xrightarrow{a} \sigma_y^2 \int_0^1 W_{d+d_1}(s)^2 ds \quad (4.50)$$

Damit können wir einen Varianzverhältnistest mit folgender Form konstruieren:

$$\rho(d_1) = n^{2d_1} \frac{\sum_{t=1}^n y_t^2}{\sum_{t=1}^n \tilde{y}_t^2} \xrightarrow{a} \frac{\int_0^1 W_d(s)^2 ds}{\int_0^1 W_{d+d_1}(s)^2 ds}. \quad (4.51)$$

Die Motivation der Bildung des VR-Tests ist, dass durch die Division der beiden zweiten Momente der Störparameter  $\sigma_y^2$  aus der Verteilung der Teststatistik verschwindet und muss somit bei der Konstruktion der Teststatistik nicht weiter berücksichtigt werden. Diese Idee wurde bereits von Vogelsang (1998) geäußert und ist auch in einem nicht-parametrischen Einheitswurzeltest von Breitung (2002) unter der Annahme  $d = d_1 = 1$  umgesetzt worden. Der Parameter  $d_1$  selbst wiederum stellt einen Tuning-Parameter dar, der wiederum die Verteilung der Teststatistik  $\rho(d_1)$  in (4.51) beeinflusst. Der Test selbst kann dazu verwendet werden, um die Hypothese  $H_0 : d_0 = d$  gegen  $H_1 : d_0 < d$  zu testen. Hierzu wird für eine Wahl  $d_1$  die VR-Statistik  $\rho(d_1)$  berechnet und mit zu bestimmenden kritischen Werten  $CV_\alpha(d, d_1)$  verglichen. Ist nun  $\rho(d_1) < CV_\alpha(d, d_1)$ , dann kann die Nullhypothese nicht abgelehnt werden und man kann zu der Testentscheidung kommen, dass  $y_t$  mindestens integriert vom Grade  $d$  ist, da es sich bei dem VR-Test wie bei den vorangegangenen FDF- und dem effizienten Wald-Test um einen einseitigen Test handelt. Die kritischen Werte  $CV_\alpha(d, d_1)$  ergeben sich als das  $(1 - \alpha)$ -Quantil der unter Gültigkeit der Nullhypothese simulierten Verteilung der VR-Statistik. Des Weiteren gilt es zu beachten, dass die Verteilung wiederum vom Vorliegen von deterministischen Prozessbestandteilen abhängt. Diese gilt es, durch eine Regression der Zeitreihe auf die jeweilige angenommene deterministische Spezifikation vor Berechnung der VR-Statistik zu bereinigen. In unseren Überlegungen nehmen wir an, dass ein zu untersuchender Prozess einen Mittelwert ungleich 0 aufweisen kann oder einen linearen Trend bzw. einen gebrochenen Trend besitzt.

Es stellt sich nun die Frage, welchen Parameter  $d_1$  wir für unsere weiteren Überlegungen annehmen sollten. Hierzu wollen wir in Simulationen untersuchen, wie die Macht des VR-Ansatzes bei verschiedenen angenommenen Werten von  $d_1$  sich darstellt. Dazu produzieren wir anfänglich simulierte Prozesse  $y_t = \Delta^{-d}\varepsilon_t$  mit  $d \in [0,5; 1,5]$  und  $\varepsilon_t \sim N(0; 1)$ . Die jeweilige Stichprobengröße fixieren wir wiederum bei  $n = 1000$  und wiederholen jedes Experiment in Abhängigkeit von  $d$  insgesamt 100.000-mal. Für die Bildung der VR-Teststatistiken bestimmen wir den Tuning-Parameter  $d_1$  im Intervall  $(0; 1]$ . In Abbildung 4.23 stellen wir nun die kritischen Werte für das Irrtumsniveau  $\alpha = 5\%$  dar, welche gerade die 95%-Quantile der simulierten Verteilung

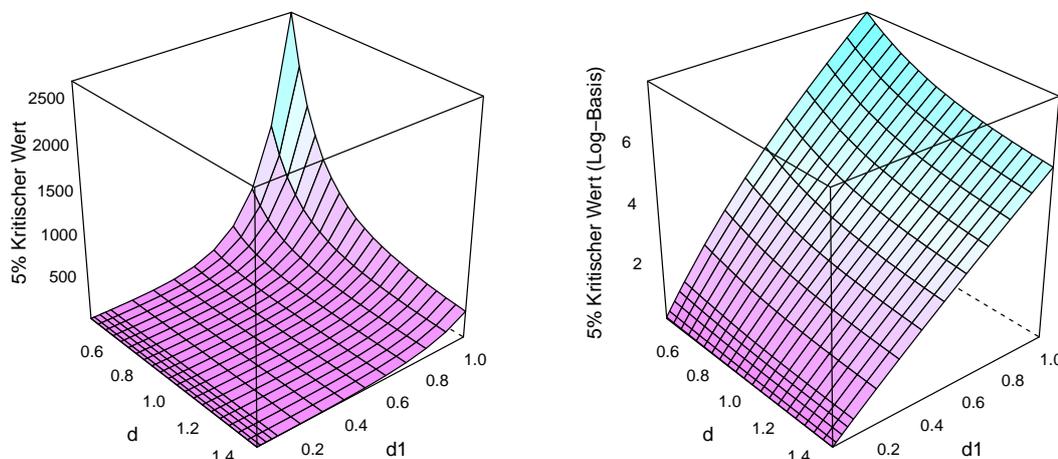


Abb. 4.23: Struktur der 5%-kritische-Werte für VR-Statistik in Abhängigkeit von  $d$  und des Tuning-Parameters  $d_1$  für  $n = 1000$

der VR-Statistik darstellt. Die linke Graphik zeigt die nominalen kritischen Werte, wobei wir erkennen können, dass diese exponentiell mit  $d_1$  ansteigen. Aus diesem Grund stellen wir in der rechten Graphik auch die kritischen Werte auf logarithmierter Basis dar, wo wir nun auch besser erkennen, dass für sämtliche Werte von  $d_1$  die kritischen Werte mit steigendem  $d$  sinken. Mit diesen kritischen Werten können wir nun untersuchen, wie gut die Macht des VR-Tests in Abhängigkeit von  $d_1$  ist. Wir fixieren nun die hypothetischen Parametermöglichkeiten  $d_0 \in \{0,8; 1; 1,2\}$  und wollen kleinere Stichprobenumfänge  $n \in \{100; 250; 500\}$  untersuchen. Dazu simulieren wir Prozesse, die einen Integrationsgrad  $d \leq d_0$  besitzen und untersuchen, welche Ablehnungsraten der Nullhypothese sich bei Verwendung der kritischen Werte aus der Abbildung 4.23 ergeben. Die Ergebnisse stellen wir in Tabelle 4.22 für den Fall, dass der Störterm  $\varepsilon_t \sim N(0; 1)$  ein einfaches Weißes Rauschen darstellt, und in der Tabelle 4.23 für die Situation, dass  $\varepsilon_t = (1 - \theta L)\eta_t$  mit  $\eta_t \sim N(0; 1)$  ein AR(1)-Prozess ist. In den jeweiligen Spalten befinden sich die einzelnen Tuning-Parameter  $d_1$ , während in den Zeilen der wahre Integrationsgrad  $d$  der simulierten Zeitreihe dargestellt ist. Als Ergebnis lässt sich hierbei festhalten, dass bei größeren Stichproben die Macht des VR-Tests größer ist mit kleinerem Tuning-Parameter als bei größeren Ausprägungen von  $d_1$ . So kann z.B. die Hypothese  $d_0 = 1$  für einen Prozess mit wahrem Integrationsgrad  $d = 0,6$  für  $n = 500$  mit  $d_1 = 0,01$  in 96,6% aller Fälle abgelehnt werden,

während bei  $d_1 = 1$  dies nur noch in 53,0% aller Fälle geschah. Es fällt auch auf, dass bei den kleineren Werten von  $d_1 \leq 0,1$  sich keine allzu großen Unterschiede ergeben. Für  $d_1 = 0,01$  ist die Macht zwar für  $n = 500$  am größten von allen angegebenen Tuning-Parametern, dafür fällt diese bei kleineren Stichprobengrößen aber geringer aus. Bei  $n = 100$  und  $d_0 = 0,8$  beobachten wir hingegen, dass wir mit  $d_1 = 1$  eine größere Macht erreichen als bei den kleinen Tuning-Parametern. Wenn wir z.B. Prozesse mit  $d = 0,6$  simulieren, dann können wir  $d_0 = 0,8$  mit einer relativen Häufigkeit von 13,5% ablehnen, während wir mit einer Wahl von  $d_1 = 0,01$  dies nur in 1,3% aller Fälle können. Diese Beobachtungen werden bei Vorliegen von AR(1)-Störtermen noch weiter bestätigt, wobei hier dann ein größerer Tuning-Parameter  $d_1$  noch häufiger vorteilhaft ist. Zusammenfassend kann man sagen, dass für kleinere Stichproben und kleinere Integrationsparameter ein großes  $d_1$  am besten ist, während in den sonstigen Fällen zumeist ein kleineres  $d_1$  die besten Ergebnisse im Hinblick auf die Macht des Tests erzielt. Bei autokorrelierten Störtermen wiederum ist sogar noch häufiger ein großes  $d_1$  besser. Aus diesem Grund entschließen wir uns, für unsere Studien die beiden Tuning-Parameter  $d_1 = 0,1$  und  $d_1 = 1$  weiter zu untersuchen.  $d_1 = 0,01$  hat zwar zumeist eine bessere Macht als  $d_1 = 0,1$ , diese ist aber häufig nicht viel größer und wir müssen damit rechnen, dass wir in kleineren Stichproben schlechtere Ergebnisse als mit  $d_1 = 0,1$  erzielen. Nielsen (2008) schlägt ebenfalls die Wahl  $d_1 = 0,1$  vor, gewinnt diese Erkenntnisse aber nicht dadurch, dass er die Macht bei Vorliegen von fraktional integrierten Prozessen untersucht, sondern lediglich bei autokorrelierten Prozessen. In Nielsen (2010) finden hingegen keine genaueren Untersuchungen über die Wahl von  $d_1$  statt. Nun können wir noch untersuchen, die Verteilung genauer zu spezifizieren bzw. eine mögliche Approximation zu finden. Eine Approximation über eine Normalverteilung scheidet hier prinzipiell aus, da die Teststatistik durch Konstruktion als Division der beiden zweiten Momente nur positive Zahlen annehmen kann. Eine Möglichkeit, diesen Umstand zu berücksichtigen, wäre eventuell die Modellierung mit Hilfe der Gammaverteilung, die sich nur für positive Zufallsvariablen ergibt. Diese bietet sich dafür auch aus einem anderen Grund an: Wie wir später sehen werden, ist der VR-Testansatz relativ eng verwandt mit dem Johansen-Kointegrationstest. Für diesen hat Doornik (1998) vorgeschlagen, dass sich die jeweilige asymptotische Verteilung gut mit Hilfe der Gammaverteilung approximieren lässt. Aus diesem Grund wollen wir eine mögliche Approximation der Testverteilung auch hier in Erwägung ziehen.

Hierzu simulieren wir wieder die jeweilige Testverteilung der VR-Statistik mit

Tab. 4.22: Ablehnungsraten für  $H_0 : d_0 = d$  bei Verwendung von  $\rho(d_1)$  mit  $\alpha = 5\%$ 

$d/d_1$	$n = 100$			$n = 250$			$n = 500$			
	0,01	0,05	1	0,01	0,05	1	0,01	0,05	1	
$H_0 : d_0 = 0,8$ gegen $H_1 : d_0 < 0,8$										
0	99,9%	99,9%	98,9%	100%	100%	100%	100%	100%	100%	100%
0,2	82,4%	87,8%	81,8%	100%	100%	100%	100%	100%	100%	99,8%
0,4	21,5%	28,5%	42,4%	96,2%	96,1%	95,8%	86,6%	65,8%	77,6%	77,6%
0,6	1,3%	2,4%	13,5%	29,9%	31,9%	33,2%	30,3%	21,3%	59,3%	25,4%
0,8	0,0%	0,1%	2,9%	1,2%	1,5%	2,0%	4,1%	4,4%	2,8%	4,4%
$H_0 : d_0 = 1$ gegen $H_1 : d_0 < 1$										
0	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
0,2	100%	100%	96,0%	100%	100%	100%	100%	99,8%	100%	100%
0,4	95,7%	95,5%	72,5%	100%	100%	100%	98,4%	87,4%	100%	93,4%
0,6	58,5%	59,1%	37,6%	89,1%	87,7%	85,9%	67,7%	46,5%	96,6%	53,0%
0,8	15,8%	16,7%	14,7%	32,6%	32,2%	31,4%	23,7%	17,2%	42,2%	18,2%
1	2,4%	2,8%	4,7%	4,6%	4,9%	5,1%	5,9%	5,3%	5,1%	5,0%
$H_0 : d_0 = 1,2$ gegen $H_1 : d_0 < 1,2$										
0	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
0,2	100%	100%	99,2%	100%	100%	100%	100%	100%	100%	100%
0,4	99,9%	99,8%	86,3%	100%	100%	100%	99,8%	95,0%	100%	98,0%
0,6	90,4%	89,6%	58,1%	98,5%	98,1%	97,2%	85,6%	66,1%	99,7%	70,0%
0,8	53,6%	52,9%	29,4%	69,4%	67,3%	64,6%	46,8%	32,1%	76,3%	33,6%
1	19,0%	19,3%	12,5%	24,3%	23,7%	22,9%	17,5%	13,2%	26,0%	13,1%
1,2	4,9%	5,1%	5,2%	5,5%	5,6%	5,8%	5,6%	5,0%	5,4%	5,6%

Tab. 4.23: Ablehnungsraten für  $H_0 : d_0 = d$  bei Verwendung von  $\rho(d_1)$  mit AR(1)-Störtermen ( $\theta = 0,5$ ) und  $\alpha = 5\%$

$d/d_1$	$n = 100$				$n = 250$				$n = 500$						
	0,01	0,05	0,1	0,5	1	0,01	0,05	0,1	0,5	1	0,01	0,05	0,1	0,5	1
$H_0 : d_0 = 0,8$ gegen $H_1 : d_0 < 0,8$															
0	15,2%	23,8%	35,1%	69,7%	63,3%	100%	100%	100%	100%	98,2%	100%	100%	100%	100%	100%
0,2	0,7%	1,6%	3,7%	24,9%	27,7%	86,8%	88,8%	90,2%	86,6%	71,1%	100%	100%	100%	98,9%	90,5%
0,4	0,0%	0,0%	0,1%	4,8%	8,7%	14,1%	18,1%	22,4%	34,0%	28,6%	73,4%	74,2%	74,4%	64,5%	46,8%
0,6	0,0%	0,0%	0,0%	0,6%	1,9%	0,3%	0,6%	1,1%	5,6%	7,0%	6,4%	8,0%	9,6%	14,8%	12,7%
0,8	0,0%	0,0%	0,0%	0,0%	0,4%	0,0%	0,0%	0,0%	0,6%	1,6%	0,1%	0,2%	0,3%	1,7%	2,7%
$H_0 : d_0 = 1$ gegen $H_1 : d_0 < 1$															
0	99,1%	99,2%	99,2%	97,5%	89,8%	100%	100%	100%	100%	99,8%	100%	100%	100%	100%	100%
0,2	76,3%	77,7%	78,8%	76,2%	62,8%	100%	100%	99,9%	98,9%	92,0%	100%	100%	100%	98,7%	98,7%
0,4	26,7%	29,2%	31,6%	37,5%	31,5%	90,0%	89,3%	88,5%	77,2%	59,9%	99,6%	99,5%	99,2%	92,5%	75,7%
0,6	4,9%	6,0%	7,1%	12,2%	12,4%	38,9%	39,5%	39,9%	35,9%	27,2%	69,2%	67,9%	66,1%	51,3%	36,6%
0,8	0,6%	0,7%	1,0%	3,6%	4,9%	6,2%	6,9%	7,7%	10,6%	9,9%	16,7%	17,1%	17,4%	16,6%	13,2%
1	0,1%	0,1%	0,1%	0,9%	1,7%	0,6%	0,8%	1,1%	2,6%	3,5%	1,6%	2,0%	2,4%	3,6%	3,8%
$H_0 : d_0 = 1,2$ gegen $H_1 : d_0 < 1,2$															
0	100%	100%	100%	99,8%	97,2%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
0,2	98,3%	98,2%	97,8%	92,7%	80,2%	100%	100%	100%	99,9%	97,4%	100%	100%	100%	100%	99,8%
0,4	76,9%	76,6%	75,8%	66,6%	52,7%	99,3%	99,0%	98,6%	92,0%	77,3%	100%	100%	100%	98,1%	88,4%
0,6	38,1%	38,7%	38,9%	35,3%	28,4%	79,2%	78,0%	76,3%	61,6%	46,1%	93,5%	92,3%	90,6%	75,1%	56,0%
0,8	12,3%	12,9%	13,4%	14,7%	13,0%	35,2%	35,0%	34,3%	28,0%	21,2%	52,3%	51,0%	49,2%	37,1%	26,7%
1	3,5%	3,8%	4,2%	5,8%	5,9%	11,1%	11,6%	11,7%	11,5%	9,7%	15,8%	15,9%	15,7%	13,5%	10,9%
1,2	0,9%	1,1%	1,3%	2,4%	2,6%	2,6%	2,8%	3,0%	4,0%	4,3%	4,2%	4,5%	4,7%	5,3%	5,0%

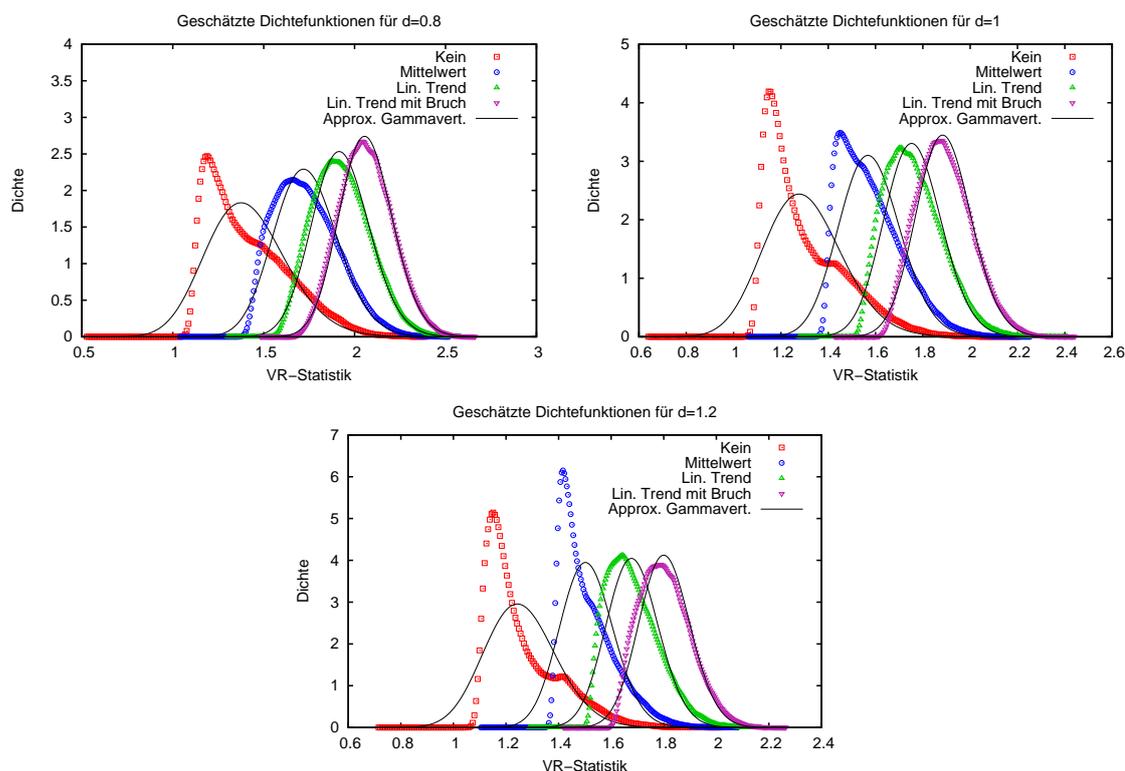


Abb. 4.24: Anpassung der Gammaverteilung für die VR-Statistik mit  $d_1 = 0,1$

einem fixierten Beobachtungsumfang von  $n = 1000$  und jeweils 100.000 Wiederholungen für den Integrationsgrad  $d \in \{0,8; 1; 1,2\}$  für die beiden Tuning-Parameter  $d_1 = 0,1$  und  $d_1 = 1$ . Dann schätzen wir mit einem Kerndichtenschätzer die jeweilige Dichtefunktion der Testverteilung für die einzelnen Trendoptionen und passen jeweils eine approximierende Gammaverteilung an, was wir für  $d_1 = 0,1$  in Abbildung 4.24 zeigen. Für  $d_1 = 1$  verzichten wir auf eine Darstellung, da die Wertebereiche der Teststatistik sehr groß sind und man grafisch nicht feststellen kann, ob eine gute Approximation vorliegt. Beiden Approximationen ist gemein, dass, wenn man z.B. den Kolmogorow-Smirnow-Anpassungstest auf Übereinstimmung mit der Gammaverteilung durchführt, diese signifikant zu allen üblichen Irrtumswahrscheinlichkeiten abgelehnt werden. Allein grafisch betrachtet ergibt sich für den Fall ohne Trendbereinigung und mit Mittelwertbereinigung eine relativ starke Abweichung von der approximierten Verteilung. Für den Fall des linearen Trends und mit Trendbruch hingegen scheint eine moderate Übereinstimmung zu existieren, die aber bei Anwendung von Anpassungstests signifikant abgelehnt werden kann. Für eine Testentscheidung ist es nun nicht unbedingt wichtig, dass eine Verteilung in ihrem gesamten Wertebereich nachempfunden wird. Es ist vielmehr entscheidend, dass die Verteilung

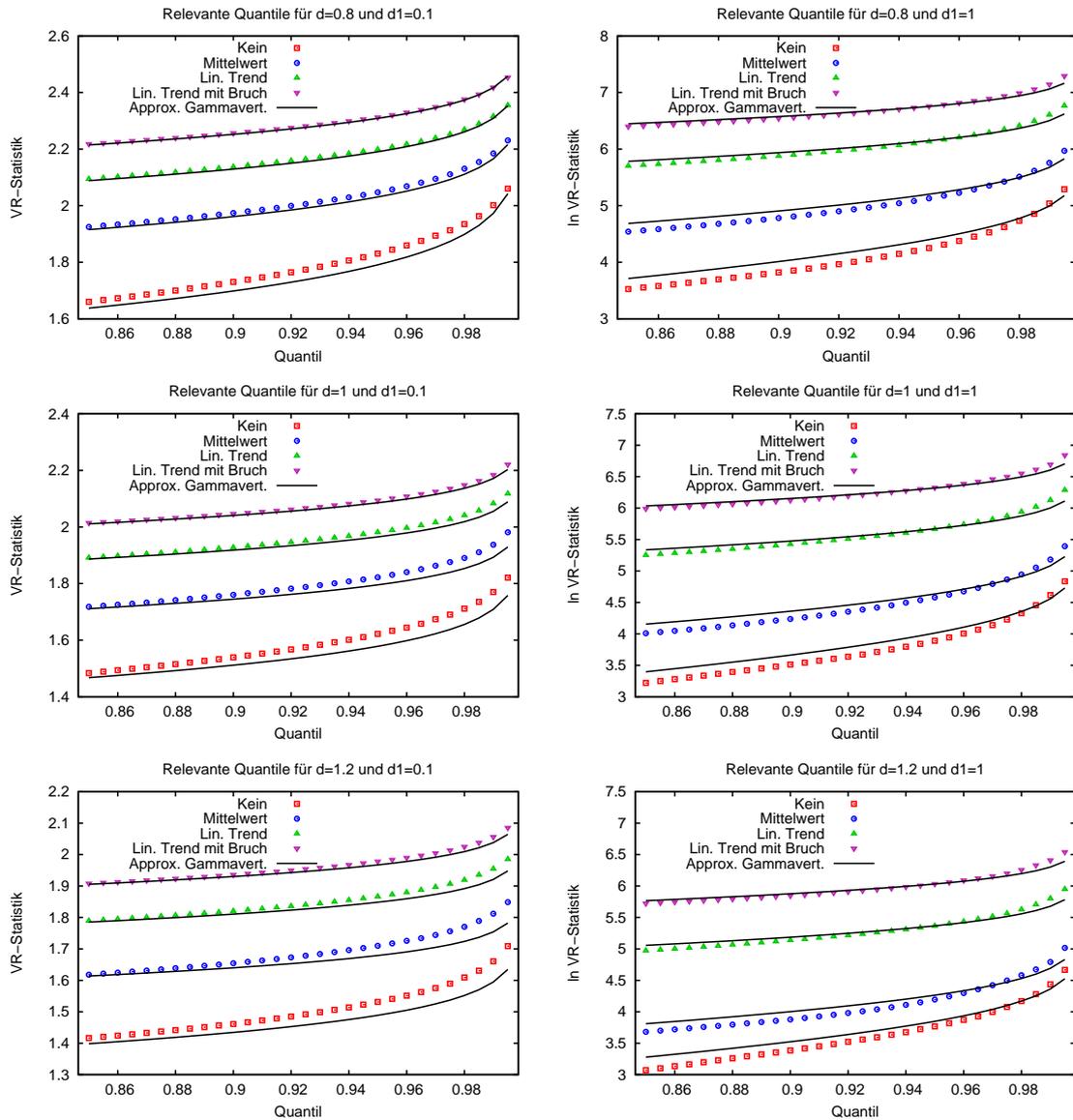


Abb. 4.25: Anpassung der Gammaverteilung für die VR-Statistik im relevanten kritischen Bereich (für  $d_1 = 1$  auf logarithmierter Basis)

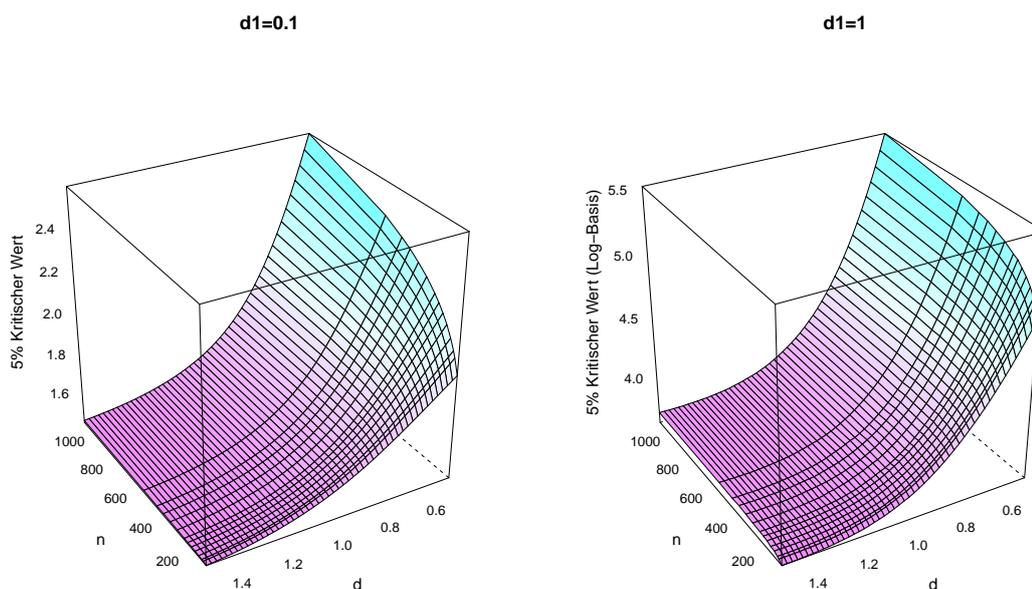


Abb. 4.26: Struktur der 5%-kritische-Werte für die VR-Statistik ohne Trendbeseitigung in Abhängigkeit von  $d$  und des Beobachtungsumfangs  $n$

im relevanten kritischen Bereich gut getroffen wird, welcher für die VR-Statistik im  $(1 - \alpha)$  der Verteilung der Teststatistik liegt. Für diesen Zweck haben wir in der Abbildung 4.25 die jeweils simulierten kritischen Werte mit denen durch die Gamma-verteilung approximierten kritischen Werte für  $d_1 = 0,1$  und für  $d_1 = 1$  dargestellt. Hier können wir erkennen, dass die jeweiligen für eine Testentscheidung relevanten kritischen Werte, insbesondere für den Fall des linearen Trends und Trendbruchs, gut getroffen werden. Hier ergibt sich aber für den Fall ohne Trend und mit Mittelwertbereinigung eine scheinbar systematische Unterschätzung für  $d_1 = 0,1$  und eine teilweise Überschätzung für  $d_1 = 1$ , wodurch wir zu dem Schluss kommen, dass wir eine Approximation mit der Gamma-verteilung nicht weiterverfolgen möchten. Der VR-Test lässt sich zum Testen von Kointegrationsbeziehungen heranziehen, was auch schon Nielsen (2010) beschrieben hat. Für diesen Fall arbeitet unter Berücksichtigung linearer Trends und Trendbrüche Dechert (2014) Approximationsweisen der Testverteilungen auf Basis der Gamma-verteilung heraus, worauf wir aber in dieser Arbeit nicht mehr eingehen wollen.

Im nächsten Schritt werden wir versuchen, nun die kritischen Werte für übliche Irrtumsniveaus von  $\alpha$  zu modellieren. Eine Response-Surface-Funktion in Abhängigkeit von  $1/n$ , wie bei den vorangegangenen Testprozeduren, wollen wir hier nicht

Tab. 4.24: Geschätzte Response-Surface-Funktionen für die kritischen Werte des univariaten VR-Tests mit  $d_1 = 0,1$  und  $d_1 = 1$ 

$d_1 = 0,1$	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$						
	N	M	LT	LTB	N	M	LT	LTB	N	M	LT	LTB	
$\beta_0$	1,7876	1,6536	1,9678	2,3644	1,1126	0,9675	2,1363	2,5897	2,1277	2,5748	1,9064	2,7418	
$\beta_1$	14,991	14,190	11,401	8,328	18,764	18,205	10,004	6,960	11,288	6,9919	11,3902	5,9376	
$\beta_2$	-56,666	-47,842	-37,986	-28,155	-63,254	-56,281	-32,812	-24,193	-39,553	-23,629	-34,970	-20,435	
$\beta_3$	85,068	66,184	51,225	36,837	89,831	75,484	42,693	31,398	54,015	28,276	44,674	25,766	
$\beta_4$	-64,581	-46,725	-35,276	-24,445	-65,313	-52,560	-28,165	-20,629	-37,349	-16,336	-29,637	-16,615	
$\beta_5$	24,774	16,717	12,349	8,2563	24,051	18,714	9,3927	6,9293	13,113	4,5430	10,154	5,5864	
$\beta_6$	-3,8315	-2,4135	-1,7507	-1,1348	-3,5662	-2,7042	-1,2607	-0,9532	-1,8679	-0,4751	-1,4306	-0,7886	
$R^2$	99,9997%	99,9999%	99,9998%	99,9997%	99,9994%	99,9996%	99,9998%	99,9997%	99,9991%	99,9976%	99,9992%	99,9994%	
$\hat{\sigma}_{\text{Resid}}$	$5,47 \cdot 10^{-4}$	$3,45 \cdot 10^{-4}$	$4,36 \cdot 10^{-4}$	$5,02 \cdot 10^{-4}$	$7,72 \cdot 10^{-4}$	$6,67 \cdot 10^{-4}$	$4,56 \cdot 10^{-4}$	$5,39 \cdot 10^{-4}$	$9,99 \cdot 10^{-4}$	$1,70 \cdot 10^{-4}$	$9,11 \cdot 10^{-4}$	$7,74 \cdot 10^{-4}$	
$d_1 = 1$	LT			LTB	M			LT	M			LT	LTB
$\beta_0$	2661,6	6909,0	13173	18919	4129,0	8714,7	17975	24853	9393,7	13034	23042	34583,9	
$\beta_1$	-13396	-33403	-59698	-78972	-20690	-40488	-82989	-107041	-47651	-57151	-96943	-146115	
$\beta_2$	29023	70184	119527	146476	44773	81830	169204	205910	104695	109723	180495	275914	
$\beta_3$	-33913	-80167	-131352	-149294	-52381	-90099	-189220	-218638	-124728	-115269	-185299	-287070	
$\beta_4$	22390	51936	82552	87103	34655	56452	120872	133272	84207	69113	109707	170832	
$\beta_5$	-7885,1	-17976	-27928	-27383	-12229	-18978	-41509	-43847	-30367	-22275	-35386	-54606	
$\beta_6$	1154,2	2586,1	3954,3	3607,3	1792,8	2664,8	5958,4	6049,6	4553,2	3002,3	4845,9	7278,8	
$R^2$	99,9971%	99,9992%	99,9996%	99,9993%	99,9968%	99,9990%	99,9989%	99,9996%	99,9906%	99,9957%	99,9981%	99,9983%	
$\hat{\sigma}_{\text{Resid}}$	0,1696	0,3132	0,5631	1,2489	0,2939	0,4956	1,1814	1,1562	1,1587	1,9101	2,5328	3,7241	

N steht für das Vorliegen von keinem Trend, M für einen Mittelwert, LT einem linearen Trend und LTB einem linearen Trend mit Bruch.

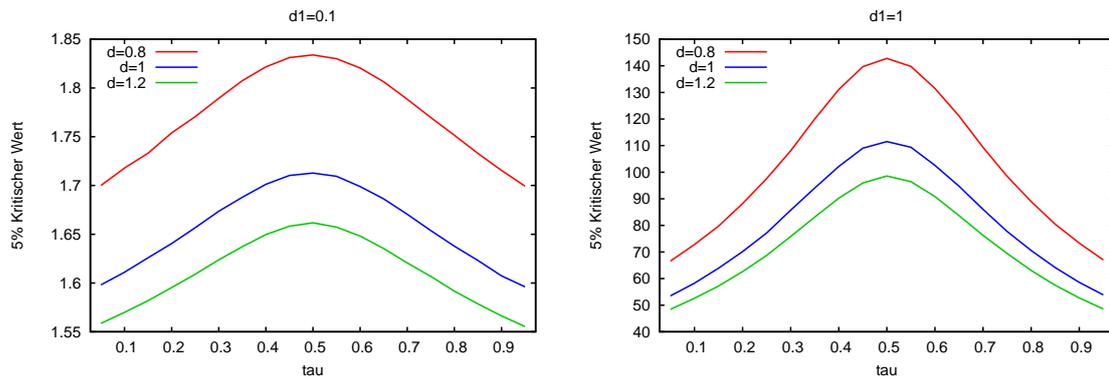


Abb. 4.27: Struktur der 5%-kritischen-Werte der VR-Statistik in Abhängigkeit von  $\tau$  für  $d_1 = 0,1$  und  $d_1 = 1$

angeben, da sich der VR-Test besonders dafür eignet, verschiedene Integrationsparameter flexibel zu testen. Um nun die asymptotischen kritischen Werte für das Irrtumsniveau  $\alpha$  zu bestimmen, simulieren wir wiederum die Testverteilungen für erzeugte Prozesse mit dem Integrationsgrad  $d \in [0,5; 1,5]$  bis zu einem maximalen Beobachtungsumfang von  $n = 1000$  mit jeweils insgesamt 1.000.000 Wiederholungen. Diese sind für die Irrtumswahrscheinlichkeit  $\alpha = 5\%$  in Abbildung 4.26 für  $d_1 = 0,1$  und  $d_1 = 1$  dargestellt. Die asymptotischen kritischen Werte können durch Extrapolation mit einer Regression von Polynomen des Terms  $1/n$  ermittelt werden.

Für die asymptotischen kritischen Werte für den Test der Hypothese  $H_0 : d_0 = d$  gegen  $H_1 : d_0 < d$  schlagen wir eine Response-Surface-Funktion in Abhängigkeit eines Polynom sechsten Grades von  $d$ , ähnlich wie für den Test auf Vorliegen eines linearen Trends in Abschnitt 4.1.1, vor:

$$CV_\alpha(d, d_1) = \beta_0 + \beta_1 d + \beta_2 d^2 + \beta_3 d^3 + \beta_4 d^4 + \beta_5 d^5 + \beta_6 d^6 + \text{Resid.} \quad (4.52)$$

Es ergeben sich damit für die beiden Tuning-Parameter  $d_1 = 0,1$  und  $d_1 = 1$  und jeweiligen vier Trendoptionen insgesamt acht verschiedene Response-Surface-Funktionen für jedes Irrtumsniveau  $\alpha$ , die wir kompakt in der Tabelle 4.24 darstellen. Auch hier ergibt sich wie bei den Tests auf linearen Trend und Trendbruch durchweg eine sehr gute Anpassung der Response-Surface-Funktion an die simulierten kritischen Werte. Dies mag womöglich auch daran liegen, dass wir in diesen Fällen asymptotische kritische Werte modellieren, die bereits in einer vorangegangenen Regression geglättet worden sind. Bei der Simulation der kritischen Werte für den Fall des gebrochenen Trends verwenden wir wieder wie bei den vorangegangenen Tests lediglich die Situation  $\tau = 0,5$  als relativen Bruchzeitpunkt, da sich hier

der maximale kritische Wert der VR-Statistik für den Fall des gebrochenen Trends einstellt. Um dies zu zeigen, stellen wir in der Abbildung 4.27 die kritischen Werte für  $\alpha = 5\%$  für  $d_1 = 0,1$  und  $d_1 = 1$  dar mit  $d \in \{0,8; 1; 1,2\}$ . Hier ist noch anzumerken, dass für  $d_1 = 1$  die kritischen Werte nicht logarithmiert dargestellt werden, da die Variabilität in  $\tau$  nicht allzu groß ist. Auch die Response-Surface-Funktion für  $d_1 = 1$ , dargestellt in Tabelle 4.24, erklärt nicht die logarithmierten kritischen Werte, sondern die jeweiligen absoluten.

## 4.6 Vergleichende Monte-Carlo-Simulation

In der folgenden Simulationsstudie wollen wir nun insbesondere die Macht der fraktionalen Integrationstests miteinander vergleichen. Hierzu betrachten wir simulierte Prozesse  $y_t = \Delta^{-d}\varepsilon_t$  mit festgelegtem Integrationsgrad  $d \in \{0; 0,2; 0,4; 0,6; 0,8; 1\}$  und testen jeweils mit den vorgestellten Testverfahren die Hypothese  $H_0 : d = 1$  gegen die Alternative  $H_1 : d < 1$  bei einem fixierten Irrtumsniveau  $\alpha = 5\%$ . Den Beobachtungsumfang halten wir variabel mit  $n \in \{100; 250; 500\}$ , wobei wir aber, um eine bessere Vergleichsbasis zu erreichen, asymptotische kritische Werte verwenden wollen, da wir für den VR-Ansatz keine in  $n$  angepasste Response-Surface Funktion angeben haben. Den Störterm  $\varepsilon_t$  spezifizieren wir als einen AR(1)-Prozess der Form

$$(1 - \theta L)\varepsilon_t = \eta_t \quad \text{mit} \quad \theta \in \{0; 0,3; 0,5; 0,8\} \quad \text{und} \quad \eta_t \sim N(0; 1). \quad (4.53)$$

Jedes einzelne Experiment wiederholen wir insgesamt 10.000-mal und stellen die jeweiligen Ergebnisse in den Tabellen 4.25 und 4.26 dar. Hier wenden wir mit dem effizienten Wald-Test (NEW), dem fraktionalen Dickey-Fuller-Test (FDF) und dem einfachen Lagrange-Multiplikator-Test (LM) in der Version von Breitung und Hassler (2002) parametrische Testverfahren an, wo wir eine Lagzahl  $p$  spezifizieren müssen. Dabei nehmen wir an, dass wir zum einen mit  $p = 0$  für  $\theta = 0$  und mit  $p = 1$  für  $\theta \neq 0$  die wahre Prozessordnung des Störterms kennen und zum anderen wenden wir mit  $p_{\text{short}} = \lfloor 4(n/100)^{1/4} \rfloor$  die Untergrenze der bekannten Schwert-Heuristik an. DKH (2008) ziehen zwar die Verwendung von Informationskriterien in der Testregression von LM- und FDF-Test in Erwägung, um die Lagzahl  $p$  zu bestimmen, kommen aber zu dem Ergebnis, dass diese der Schwert-Heuristik in Bezug auf die Macht der Tests unterlegen sind und nur sehr unzuverlässige Ergebnisse liefern. Den Varianzverhältnistest (VR) verwenden wir mit den beiden Tuning-Parametern

Tab. 4.25: Ablehnungsraten für  $H_0 : d = 1$  bei Verwendung von fraktionalen Integrationstests für  $\theta = 0$  und  $\theta = 0,3$ 

$\theta = 0$										
$d$	NEW		FDF		LM		VR		$m = n^{0,65}$	
	$p = 0$	$p = p_{\text{short}}$	$p = 0$	$p = p_{\text{short}}$	$p = 0$	$p = p_{\text{short}}$	$d_1 = 0,1$	$d_1 = 1$		
$n = 100$	0	100%	81,6%	99,4%	75,9%	100%	100%	100%	100%	99,8%
	0,2	100%	82,5%	99,8%	80,5%	100%	100%	100%	96,0%	99,6%
	0,4	99,8%	76,1%	99,8%	76,0%	100%	100%	95,2%	72,6%	98,3%
	0,6	93,6%	50,3%	92,7%	48,0%	99,8%	99,7%	59,0%	37,4%	86,5%
	0,8	43,1%	18,7%	44,4%	20,5%	74,8%	73,2%	17,5%	14,7%	44,7%
	1	5,5%	6,5%	5,6%	6,0%	5,2%	5,2%	3,3%	4,7%	9,4%
$n = 250$	0	100%	95,8%	100%	95,8%	100%	100%	100%	100%	100%
	0,2	100%	98,3%	100%	98,6%	100%	100%	100%	99,7%	100%
	0,4	100%	99,2%	100%	99,1%	100%	100%	100%	87,5%	100%
	0,6	100%	86,7%	100%	87,5%	100%	100%	85,9%	46,3%	99,0%
	0,8	82,3%	35,8%	81,6%	33,4%	98,7%	98,5%	31,3%	17,0%	65,2%
	1	5,2%	5,7%	5,4%	5,7%	5,1%	5,2%	5,1%	5,4%	7,4%
$n = 500$	0	100%	99,4%	100%	99,7%	100%	100%	100%	100%	100%
	0,2	100%	99,9%	100%	100%	100%	100%	100%	99,9%	100%
	0,4	100%	100%	100%	100%	100%	100%	99,8%	93,7%	100%
	0,6	100%	99,4%	100%	99,7%	100%	100%	94,4%	53,1%	100%
	0,8	98,6%	60,7%	98,4%	59,4%	100%	100%	38,9%	18,2%	82,6%
	1	5,1%	5,2%	5,2%	5,5%	5,2%	5,1%	5,5%	5,1%	5,8%
$\theta = 0,3$										
$d$	NEW		FDF		LM		VR		$m = n^{0,65}$	
	$p = 1$	$p = p_{\text{short}}$	$p = 1$	$p = p_{\text{short}}$	$p = 1$	$p = p_{\text{short}}$	$d_1 = 0,1$	$d_1 = 1$		
$n = 100$	0	100%	99,9%	100%	100%	100%	100%	100%	93,9%	100%
	0,2	100%	98,9%	100%	97,9%	100%	100%	87,3%	76,1%	99,7%
	0,4	99,2%	85,4%	98,6%	81,7%	96,0%	85,4%	57,0%	47,9%	95,6%
	0,6	80,6%	48,0%	78,0%	44,8%	44,8%	33,8%	28,0%	22,5%	66,4%
	0,8	30,4%	18,9%	28,2%	16,7%	13,5%	8,1%	7,7%	8,8%	21,4%
	1	5,3%	5,9%	5,2%	5,8%	4,0%	2,7%	1,3%	2,9%	2,8%
$n = 250$	0	100%	100%	100%	100%	100%	100%	100%	99,9%	100%
	0,2	100%	100%	100%	100%	100%	100%	99,9%	95,1%	100%
	0,4	100%	99,6%	100%	99,9%	100%	100%	93,1%	70,9%	100%
	0,6	99,7%	85,9%	99,4%	83,5%	84,5%	84,1%	58,3%	34,9%	95,8%
	0,8	61,4%	32,9%	60,1%	30,6%	41,1%	19,0%	17,2%	12,8%	42,4%
	1	5,4%	5,7%	5,3%	5,6%	4,8%	3,2%	2,7%	4,2%	2,1%
$n = 500$	0	100%	100%	100%	100%	100%	100%	100%	100%	100%
	0,2	100%	100%	100%	100%	100%	100%	100%	99,2%	100%
	0,4	100%	100%	100%	100%	100%	100%	99,5%	82,8%	100%
	0,6	100%	99,6%	100%	98,8%	99,2%	91,1%	77,3%	43,2%	99,9%
	0,8	88,9%	55,1%	88,0%	52,8%	79,3%	30,6%	26,0%	15,2%	67,4%
	1	5,1%	5,2%	5,2%	5,4%	5,2%	4,7%	3,6%	4,3%	2,2%

Tab. 4.26: Ablehnungsraten für  $H_0 : d = 1$  bei Verwendung von fraktionalem Integrationstests für  $\theta = 0,5$  und  $\theta = 0,8$ 

$\theta = 0,5$										
$d$	NEW		FDF		LM		VR		LW $m = n^{0,65}$	
	$p = 1$	$p = p_{\text{short}}$	$p = 1$	$p = p_{\text{short}}$	$p = 1$	$p = p_{\text{short}}$	$d_1 = 0,1$	$d_1 = 1$		
$n = 100$	0	100%	99,5%	100,0%	99,8%	99,9%	99,8%	99,3%	89,7%	99,8%
	0,2	99,6%	96,5%	99,3%	93,9%	97,8%	92,9%	78,9%	62,8%	97,6%
	0,4	90,7%	76,3%	89,0%	70,6%	70,6%	55,3%	31,4%	31,6%	78,4%
	0,6	54,9%	41,3%	53,4%	37,0%	47,9%	31,8%	7,3%	12,4%	33,7%
	0,8	18,5%	16,4%	17,0%	14,1%	11,7%	9,4%	1,1%	4,7%	5,4%
	1	4,1%	4,8%	4,6%	4,2%	3,8%	2,0%	0%	1,4%	0,4%
$n = 250$	0	100%	100%	100%	100%	100%	100%	100%	99,9%	100%
	0,2	100%	100%	100%	100%	100%	100%	100%	91,9%	100%
	0,4	100%	98,6%	100%	99,4%	79,1%	68,7%	88,3%	59,9%	99,8%
	0,6	93,4%	79,5%	92,5%	75,1%	66,1%	53,5%	39,8%	27,2%	79,5%
	0,8	37,4%	28,6%	36,9%	25,8%	21,9%	15,6%	7,6%	9,5%	13,3%
	1	4,6%	5,4%	4,7%	4,0%	4,5%	2,6%	1,2%	3,5%	0,3%
$n = 500$	0	100%	100%	100%	100%	100%	100%	99,9%	100%	100%
	0,2	100%	100%	100%	100%	100%	100%	99,9%	99,0%	100%
	0,4	100%	100%	100%	100%	100%	98,1%	99,2%	75,8%	100%
	0,6	99,9%	98,3%	99,9%	96,8%	97,7%	87,3%	66,4%	36,7%	98,8%
	0,8	64,7%	49,1%	66,1%	46,2%	40,2%	32,0%	17,4%	13,3%	35,6%
	1	5,1%	5,6%	4,6%	4,8%	4,6%	3,7%	2,4%	3,9%	0,3%
$\theta = 0,8$										
$d$	NEW		FDF		LM		VR		LW $m = n^{0,65}$	
	$p = 1$	$p = p_{\text{short}}$	$p = 1$	$p = p_{\text{short}}$	$p = 1$	$p = p_{\text{short}}$	$d_1 = 0,1$	$d_1 = 1$		
$n = 100$	0	88,7%	87,3%	85,4%	79,5%	76,8%	65,5%	98,7%	83,7%	76,6%
	0,2	54,1%	62,8%	51,0%	53,5%	38,0%	27,9%	66,1%	42,9%	32,9%
	0,4	19,9%	35,7%	20,0%	28,6%	12,1%	9,3%	6,6%	9,9%	6,1%
	0,6	7,4%	15,1%	7,1%	13,2%	6,0%	3,4%	3,2%	6,7%	0,5%
	0,8	6,1%	3,3%	2,8%	5,1%	0,9%	0%	0%	3,0%	0,1%
	1	1,3%	1,7%	4,3%	1,5%	0%	0%	0%	1,1%	0%
$n = 250$	0	100%	99,9%	99,9%	99,8%	96,7%	96,4%	100%	99,7%	99,6%
	0,2	92,8%	98,1%	90,6%	94,3%	51,1%	56,2%	99,8%	87,3%	84,6%
	0,4	46,5%	80,8%	48,5%	70,4%	20,5%	23,4%	81,6%	43,4%	23,2%
	0,6	10,1%	43,8%	16,1%	34,7%	7,0%	9,6%	12,3%	15,6%	0,9%
	0,8	5,5%	12,6%	5,7%	11,7%	3,2%	6,7%	3,5%	5,5%	0%
	1	2,6%	3,0%	4,2%	1,6%	1,0%	1,4%	0,3%	2,4%	0%
$n = 500$	0	100%	100%	100%	100%	100%	100%	100%	100%	100%
	0,2	99,9%	99,9%	99,9%	100%	97,5%	99,7%	100%	97,9%	99,8%
	0,4	82,3%	98,8%	84,1%	95,9%	53,9%	56,7%	98,7%	65,1%	74,2%
	0,6	21,2%	75,6%	36,5%	64,8%	14,7%	32,5%	49,2%	26,8%	4,4%
	0,8	4,2%	26,8%	12,4%	21,2%	5,7%	12,0%	4,5%	10,2%	0%
	1	3,0%	4,2%	4,0%	3,1%	2,4%	3,3%	0,7%	3,1%	0%

$d_1 = 0,1$  und  $d_1 = 1$  und des Weiteren soll der Integrationsgrad auch mit dem lokalen Whittle-Schätzer (LW) mit der fixierten Bandbreite  $m = n^{0,65}$  direkt bestimmt werden. Diesen wenden wir auf die erste Differenz der simulierten Prozesse an, da wir mit diesen Verfahren nur bei einem Integrationsgrad von  $d < 3/4$  ein asymptotisch normalverteiltes Schätzergebnis erhalten. Die Hypothese  $H_0 : d = 1$  bzw.  $d_{\Delta y_t} = 0$ , wenn wir die erste Differenz des Ausgangsprozesses  $y_t$  betrachten, lässt sich gegen die Alternative  $H_1 : d < 1$  oder  $d_{\Delta y_t} < 0$  ablehnen, wenn das Schätzergebnis  $\hat{d}_{\Delta y_t} < c_{5\%}(4/m)^{1/2}$  ist, wobei  $c_{5\%}$  das 5%-Quantil der Standardnormalverteilung darstellt.

Als Ergebnis der durchgeführten Simulationen können wir festhalten, dass der effiziente Wald-Test meist besser abschneidet als der fraktionale Dickey-Fuller-Test, da er tendenziell eine leicht bessere Macht aufweist. Prinzipiell liegen die Ablehnungsraten der Nullhypothese auf einem ähnlichen Niveau, doch fallen diese zu Gunsten des effizienten Wald-Tests besser aus. Wenig überraschend ist das Ergebnis, dass es bei der Wahl der Schwert-Heuristik zu Einbußen bei der Macht kommt, da die Anzahl der zu schätzenden Parameter in der Testregression ansteigt. Selbst dann schneidet aber der effiziente Wald-Test besser ab. Der LM-Test hingegen hat die beste Macht, wenn keine Autokorrelation der Störterme (d.h.  $\theta = 0$ ) vorliegt. In den anderen Fällen ist dieser aber selbst dem fraktionalen Dickey-Fuller-Test unterlegen. Die Testentscheidungen für den lokalen Whittle-Schätzer fallen nur bei  $\theta = 0$  gut aus, sind aber auch nicht entscheidend besser als bei den parametrischen fraktionalen Integrationstests. Die Wahl der Bandbreite hat hierbei natürlich eine große Bedeutung und verhindert ein besseres Abschneiden. Die Bandbreite wurde in dieser Studie mit  $m = \lfloor n^{0,65} \rfloor$  fixiert, da diese insbesondere bei empirischen Anwendungen sehr häufig gewählt wird und zeigt in den Fällen höherer Autokorrelation eine sehr schwache Macht auf, auch wenn der Beobachtungsumfang mit  $n = 500$  relativ groß ist. Selbst hier wird die Hypothese  $d = 1$  für nicht-stationäre fraktional integrierte Prozesse (d.h.  $d > 1/2$ ) fälschlicherweise nicht abgelehnt. Der VR-Test hingegen ist den parametrischen Verfahren von der Macht her zumeist unterlegen und selbst auch dann, wenn die Schwert-Heuristik angewandt wird. Lediglich für den Fall  $\theta = 0,8$  kann man eine ähnlich gute Macht erzielen wie die parametrischen Verfahren. In dieser Situation ist aber anzumerken, dass es scheinbar zu einer sonderbaren Situation kommt: Die Macht der parametrischen Verfahren steigt tendenziell bei der Wahl der Schwert-Heuristik als Lagzahl gegen die Wahl  $p = 1$ . Der Parameter  $\theta = 0,8$  steht zwar für einen relativ stark persistenten Störterm auf Grund der hohen Autokorrelation, nichtsdestotrotz sollte die Macht aber nicht steigen, wenn die Lagzahl in den

Testregressionen erhöht wird.

Insgesamt lässt sich festhalten, dass bei der empirischen Anwendung der Integrationstests die Verwendung des effizienten Wald-Tests die verlässlichsten Ergebnisse zu erwarten sind. Deswegen werden wir uns im folgenden Kapitel speziell auf diesen Test konzentrieren, wenn wir Integrationstests zum Überprüfen von Kointegrationsbeziehungen heranziehen wollen. Der VR-Ansatz hat seinen großen Vorzug, dass er nicht-parametrisch ist und keinerlei Spezifikation des untersuchenden Prozesses bedarf, auch wenn er bei der Wahl von Heuristiken schlechter abschneidet als parametrische Verfahren. Aus diesem Grund werden wir auch diesen Ansatz beim Testen von Kointegration verwenden.

## 5 Fraktionale Kointegration

### 5.1 Begriff der fraktionalen Kointegration

Einen großen Schritt für die Entwicklung der Zeitreihenanalyse stellte die Arbeit von Engle und Granger (1987) über Kointegration dar, die auch letztendlich 2003 mit dem Nobelpreis bedacht wurde. Grob gesprochen nennt man zwei oder mehrere Zeitreihen als kointegriert, wenn die Linearkombination von nicht-stationären  $I(1)$ -Prozessen wiederum stationär, also integriert vom Grad 0 ist. Dieser Begriff wurde im Lauf der Zeit immer weiter entwickelt, um mehrere empirische Gegebenheiten zu berücksichtigen. Wir wollen den folgenden zunächst bivariaten Modellrahmen verwenden, wobei  $x_t$  und  $y_t$  jeweils skalare Zeitreihen darstellen:

$$\begin{aligned} y_t &= \beta x_t + \Delta^{-(d-b)} \varepsilon_{1t} \\ x_t &= \Delta^{-d} \varepsilon_{2t} \quad \text{mit } d \geq b > 0. \end{aligned} \tag{5.1}$$

Der Prozess  $y_t$  wird somit von dem stochastischen Trend  $x_t$  getrieben, der wiederum integriert vom Grad  $d$  ist. Dies wiederum lässt sich in einem ökonomischem Kontext so interpretieren, dass die beiden beobachteten Prozesse  $x_t$  und  $y_t$  eine Art Gleichgewicht bilden. Zudem existiert in  $y_t$  ein Kointegrationsfehler  $\Delta^{-(d-b)} \varepsilon_{1t}$ , der einen Integrationsgrad von  $d - b$  aufweist. Der Parameter  $b$  stellt die Kointegrationsstärke des Systems dar, da je größer dieser Parameter ist, desto kleiner wird der Integrationsgrad des Kointegrationsfehlers. Dies wiederum bedeutet, dass dieser Fehlerterm und damit die Abweichung von  $y_t$  vom Gleichgewicht weniger persistent wird und somit ein relativ starker Zusammenhang zwischen  $x_t$  und  $y_t$  besteht. Des Weiteren werden wir in unseren Studien Endogenitätsprobleme annehmen, was sich durch eine Korrelation zwischen den Störtermen  $\varepsilon_{1t}$  und  $\varepsilon_{2t}$  ungleich 0 zeigt. Dies hat insbesondere einen Effekt auf die Schätzung der Kointegrationsbeziehung, was wir im folgenden Abschnitt thematisieren werden. Setzen wir nun  $d = b = 1$  sind wir in dem von Engle und Granger (1987) definierten Fall. Von dieser Restriktion werden wir uns aber lösen, indem wir annehmen, dass  $d$  und  $b$  beliebige fraktionale Werte annehmen können. Die Definition der fraktionalen Kointegration stellt somit nur eine Verallgemeinerung der herkömmlichen Kointegration dar.

Aus den verschiedenen Parameterkonstellationen von  $d$  und  $b$  ergeben sich für das Gebiet der fraktionalen Kointegration neue Problemfelder, die wir hier zumindest kurz benennen wollen. Zum einen ist es möglich, dass wir  $d - b \geq 1/2$  beobachten können, was zur Folge hat, dass der Kointegrationsfehler nicht-stationär ist. Diese

Situation nennen wir daher *schwache* Kointegration, da in diesem Fall stark persistente Abweichungen mit unendlich großer Varianz vom implizierten Gleichgewicht auftreten können. Zum anderen ziehen wir auch  $d - b < 1/2$  in Erwägung, wo wir stationäre Kointegrationsfehler haben, was wiederum auch den klassischen Kointegrationsbegriff mit einschließt. Dieser Fall beschreibt dann *starke* Kointegration. In einer weiteren Überlegung können wir uns auch von der Nicht-Stationarität des stochastischen Trends vom kointegrierten System lösen, wenn wir  $0 < d < 1/2$  annehmen, da es auch bei dieser Situation zu sogenannten Scheinzusammenhängen kommen kann. Dieser Fall der *stationären* Kointegration soll aber nicht Gegenstand dieser Arbeit sein.

Um uns mit den Implikationen der schwachen und starken Kointegration zumindest grafisch vertraut zu machen, haben wir in Abbildung 5.1 simulierte Datenbeispiele. Für jede der Situationen verwenden wir den gleichen stochastischen Trend  $x_t$  und erhalten durch Hinzufügen von  $\Delta^{-(d-b)}\varepsilon_{1t}$  eine kointegrierte Zeitreihe  $y_t$ , wobei wir  $d = 1$  für sämtliche Grafiken fixieren. Man kann hierbei erkennen, dass wir für den Fall  $b = 0,6$  (d.h. der Kointegrationsfehler weist einen Integrationsgrad von 0,4 auf und ist somit stationär) eine stärkere persistente Abweichung der kointegrierten Zeitreihe von ihrem stochastischen Trend erhalten als im klassischen Fall. Diese Persistenz wird stärker für  $b = 0,4$ , da wir hier bereits einen nicht-stationären Kointegrationsfehler beobachten. Nichtsdestotrotz tendiert die kointegrierte Zeitreihe wieder zu ihrem stochastischen Trend im Gegensatz zu dem Fall, wo wir keine Kointegration haben. Hier sind die Berührungen bzw. der Zusammenlauf von  $y_t$  und  $x_t$  rein zufällig.

Engle und Granger (1987) schlagen das folgende sequentielle Vorgehen zum Testen von (klassischen) Kointegrationsbeziehungen vor:

1. Testen der (Nicht-)Stationarität der Zeitreihen  $y_t$  und  $x_t$ .
2. Schätzen der Kointegrationsbeziehung und damit Bestimmung des Kointegrationsfehlerprozesses.
3. Testen der Stationarität des Kointegrationsfehlers.

An diesen Vorgehen werden wir uns im Folgenden orientieren. Methoden zur Bestimmung des Integrationsgrades, welche wir im ersten Schritt benötigen, haben wir bereits im vorangegangenen Kapitel besprochen. Im folgenden Abschnitt werden wir uns mit Möglichkeiten der Schätzung der Kointegrationsbeziehung auseinandersetzen und daran anschließend die Möglichkeiten zur Bestimmung des Integrationsgrades des Kointegrationsfehlerprozesses analysieren. Hierbei gilt es zu beachten, dass

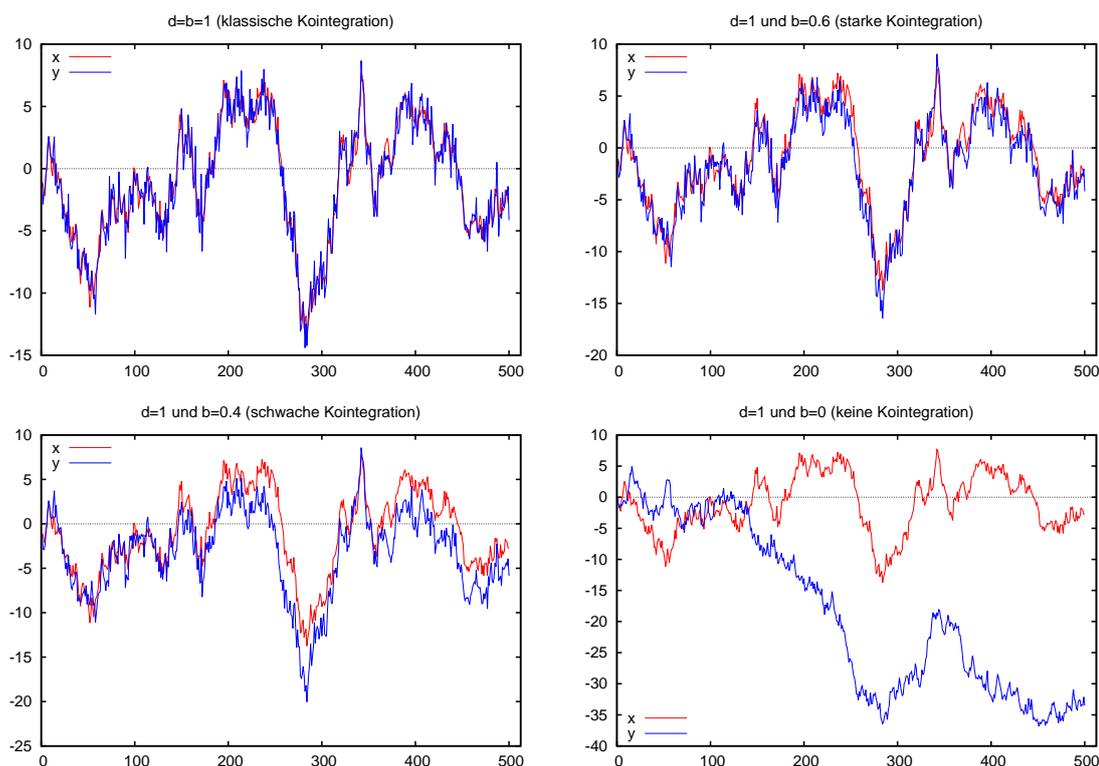


Abb. 5.1: Beispiele für bivariate fraktionale Kointegrationsbeziehungen

der Prozess nicht beobachtbar ist und mit geschätzten Parametern erzeugt wird. Dies wiederum kann Auswirkungen auf die Verteilung der geschätzten Integrationsgrade aufweisen, was wir näher betrachten werden. In den folgenden Abschnitten werden wir uns zunächst auf den Fall  $d = 1$  mit variablem  $b$  konzentrieren, da dies empirisch am häufigsten anzutreffen ist und die Modellierung mit nicht-fixierter Kointegrationsstärke eine Klassifizierung von Kointegrationsbeziehungen ermöglicht.

## 5.2 Schätzung der Kointegrationsbeziehung

Ansatzpunkt für diesen Abschnitt ist der Vorschlag von Robinson (1994a), den einfachen KQ-Schätzer (weiterhin auch synonym bezeichnet als OLS für Ordinary Least Squares) in den Frequenzbereich zu übersetzen, um dessen Effizienz zu erhöhen. Hierzu betrachten wir das Kreuzperiodogramm, das sich analog zum Periodogramm als die Fouriertransformation der Kreuzautokovarianz der beiden Reihen  $x_t$  und  $y_t$  ergibt. Zum Kreuzperiodogramm ist aber anzumerken, dass es im Gegensatz zum Periodogramm nicht notwendigerweise reellwertig ist, sondern neben einem Real- auch einen Imaginärteil umfasst. Im Folgenden betrachten wir den durchschnittli-

chen Realteil des Kreuzperiodogramms zwischen  $x_t$  und  $y_t$  mit

$$F_{xy}(k, l) = \frac{2\pi}{n} \sum_{j=k}^l \operatorname{Re}[I_{xy}(\lambda_j)] \quad \text{mit} \quad I_{xy}(\lambda_j) = w_y(\lambda_j) \overline{w_x(\lambda_j)} \quad (5.2)$$

und setzen für  $k = 1$  und  $l = n/2$ , dann liefert uns diese Größe die Summe aller Kreuzperiodogrammwerte, was gemäß des Parsevalschen Theorem der empirischen Kovarianz entspricht:

$$F_{xy}(1, n/2) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \quad (5.3)$$

Analog gilt für das durchschnittliche Periodogramm der Reihe  $x_t$

$$F_{xx}(1, n/2) = \frac{2\pi}{n} \sum_{j=1}^{n/2} I_x(\lambda_j) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.4)$$

Auf dieser Basis lässt sich der KQ-Schätzer für  $\beta$  mit den Methoden der Frequenzanalyse angeben mit

$$\hat{\beta}_{\text{OLS}} = F_{xy}(1, n/2) F_{xx}^{-1}(1, n/2). \quad (5.5)$$

Robinson und Marinucci (2003) analysieren den einfachen KQ-Schätzer unter anderem im Kontext der fraktionalen Kointegration und konnten die folgenden Konvergenzgeschwindigkeiten in Abhängigkeit der Parameterkonstellationen von  $d$  und  $b$  herleiten:

$$\begin{aligned} \text{Fall 1:} & \quad \hat{\beta}_{\text{OLS}} - \beta = O(n^{-b}) & \text{falls} & \quad 2d - b > 1 \\ \text{Fall 2a:} & \quad \hat{\beta}_{\text{OLS}} - \beta = O(n^{1-2d}/\log n) & \text{falls} & \quad 2d - b = 1 \\ \text{Fall 2b:} & \quad \hat{\beta}_{\text{OLS}} - \beta = O(n^{-1}) & \text{falls} & \quad d = b = 1 \\ \text{Fall 3:} & \quad \hat{\beta}_{\text{OLS}} - \beta = O(n^{1-2d}) & \text{falls} & \quad 2d - b < 1 \end{aligned}$$

Diese Konvergenzgeschwindigkeiten gelten nur für den Fall der nicht-stationären Kointegration, d.h.  $d > 1/2$ . Robinson und Marinucci (2003) beschäftigen sich auch mit dem Fall der stationären Kointegration, aber diese Möglichkeit wollen wir hier außen vor lassen. Die Idee von Robinson (1994a) ist ähnlich wie bei der Anwendung des lokalen Whittle-Schätzers, nicht die volle Bandbreite der verfügbaren Fourierfrequenzen zu verwenden, sondern lediglich ein kleineres Band am Ursprung und  $l = m$  zu setzen. Somit werden die höheren Frequenzen bei der Schätzung weggelassen und es ergibt sich damit der Schätzer

$$\hat{\beta}_{\text{NBLS}} = F_{xy}(1, m) F_{xx}^{-1}(1, m), \quad (5.6)$$

wobei hier NBLs für Narrow Band Least Squares steht, weil nur noch die niedrigen Fourierfrequenzen zur Schätzung herangezogen werden. Implizit kann auch eine Konstante als Bestandteil der Kointegrationsbeziehung berücksichtigt werden, da hier die Fouriertransformation der Autokovarianz zur Berechnung von  $\hat{\beta}_{\text{NBLs}}$  verwendet wird und somit eine Mittelwertbereinigung durchgeführt wird. Die Konstante selbst kann dann als Mittelwert von  $y_t - \hat{\beta}_{\text{NBLs}}x_t$  geschätzt werden. Ebenfalls in Robinson und Marinucci (2003) werden auch die Konvergenzgeschwindigkeiten für den NBLs-Schätzer angegeben:

$$\begin{array}{ll}
 \text{Fall 1:} & \hat{\beta}_{\text{NBLs}} - \beta = O(n^{-b}) \quad \text{falls } 2d - b > 1 \\
 \text{Fall 2a:} & \hat{\beta}_{\text{NBLs}} - \beta = O(n^{1-2d}/\log m) \quad \text{falls } 2d - b = 1 \\
 \text{Fall 2b:} & \hat{\beta}_{\text{NBLs}} - \beta = O(n^{-1}) \quad \text{falls } d = b = 1 \\
 \text{Fall 3:} & \hat{\beta}_{\text{NBLs}} - \beta = O(n^{1-2d}/m^{2d-b-1}) \quad \text{falls } 2d - b < 1
 \end{array}$$

Interessanterweise haben wir für  $2d - b > 1$ , was für  $d = 1$  nahezu sämtliche Kointegrationsfälle außer dem klassischen Fall umschließt, theoretisch die gleiche Konvergenzrate und somit können wir keine besseren Eigenschaften bei der Verwendung des NBLs-Schätzers erwarten. Dabei ist aber anzumerken, dass diese nicht den exakten Raten entsprechen, sondern lediglich Grenzwerte darstellen, da exakte Raten hier nicht herleitbar sind. Für den Fall  $2d - b < 1$  ergeben sich nur Verbesserungen, wenn die Bandbreite geeignet gewählt wird. Hierzu existiert bisher keine Theorie und es wurden keine Arbeiten dazu veröffentlicht. Für  $2d - b = 1$  ist OLS sogar immer besser als NBLs, da zwingend  $m < n/2$  sein muss.

Nach Robinson und Marinucci (2003) bekommt der NBLs-Schätzer insbesondere seine Bedeutung durch seine empirische Anwendbarkeit. Der NBLs-Schätzer schneidet nämlich besser als OLS ab, wenn eine Korrelation des Störterms der stochastischen Trends  $\varepsilon_{2t}$  und des Kointegrationsfehlers  $\varepsilon_{1t}$  vorliegt, was häufig als Endogenitätsproblem angesehen wird und was aus empirischer Sicht begründbar ist. In diesem Fall ist der OLS-Schätzer verzerrt und Robinson und Marinucci (2003) konnten zeigen, dass der NBLs dies zwar auch ist, aber in einem wesentlich geringeren Umfang. Deswegen kann der NBLs-Schätzer herangezogen werden, um einen Anhaltspunkt für dieses Endogenitätsproblem zu erhalten. Unterscheiden sich die jeweiligen Schätzergebnisse relativ stark, so empfehlen Robinson und Marinucci (2003) eher die Verwendung von NBLs zur Schätzung der Kointegrationsbeziehung. Um uns diese Thematik zu veranschaulichen, führen wir nun eine Simulationsstudie durch. Ähnlich wie für die Abbildung 5.1 simulieren wir stochastische Trends  $x_t$  mit Integrationsgrad  $d = 1$  und durch Hinzufügen von  $\Delta^{b-1}\varepsilon_{1t}$  kointegrierte Prozesse  $y_t$ . Für

Tab. 5.1: Durchschnittliche Verzerrung und RMSE für  $\hat{\beta}$  bei Anwendung des NBL- und OLS-Schätzers für  $n = 500$  und  $d = 1$

$b$	$\rho$	Verzerrung					RMSE				
		NBL				OLS	NBL				OLS
		$a = 0,4$	$a = 0,5$	$a = 0,6$	$a = 0,7$		$a = 0,4$	$a = 0,5$	$a = 0,6$	$a = 0,7$	
0,2	-0,8	-0,3087	-0,3130	-0,3161	-0,3184	-0,3215	<b>0,3414</b>	0,3439	0,3462	0,3481	0,3508
	-0,4	-0,1555	-0,1576	-0,1591	-0,1603	-0,1618	0,2624	0,2591	0,2578	0,2573	<b>0,2571</b>
	0	0,0012	0,0012	0,0012	0,0012	0,0012	0,2290	0,2229	0,2197	0,2179	<b>0,2162</b>
	0,4	0,1549	0,1570	0,1586	0,1598	0,1613	0,2621	0,2589	0,2575	0,2570	<b>0,2568</b>
	0,8	0,3092	0,3135	0,3166	0,3189	0,3221	<b>0,3415</b>	0,3441	0,3464	0,3482	0,3509
0,4	-0,8	-0,1146	-0,1189	-0,1224	-0,1253	-0,1300	<b>0,1305</b>	0,1341	0,1374	0,1402	0,1450
	-0,4	-0,0570	-0,0591	-0,0609	-0,0623	-0,0647	0,0981	<b>0,0977</b>	0,0979	0,0985	0,0997
	0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0854	0,0828	0,0814	0,0807	<b>0,0800</b>
	0,4	0,0570	0,0592	0,0609	0,0624	0,0647	0,0983	<b>0,0979</b>	0,0981	0,0986	0,0999
	0,8	0,1141	0,1184	0,1219	0,1248	0,1295	<b>0,1298</b>	0,1334	0,1367	0,1395	0,1443
0,6	-0,8	-0,0384	-0,0414	-0,0441	-0,0468	-0,0523	<b>0,0472</b>	0,0498	0,0525	0,0552	0,0612
	-0,4	-0,0191	-0,0206	-0,0219	-0,0233	-0,0260	<b>0,0381</b>	0,0381	0,0386	0,0394	0,0413
	0	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001	0,0348	0,0337	0,0331	0,0327	<b>0,0324</b>
	0,4	0,0190	0,0205	0,0219	0,0233	0,0260	<b>0,0381</b>	0,0382	0,0387	0,0394	0,0413
	0,8	0,0384	0,0414	0,0442	0,0469	0,0524	<b>0,0473</b>	0,0499	0,0526	0,0554	0,0613
0,8	-0,8	-0,0101	-0,0116	-0,0133	-0,0153	-0,0211	<b>0,0154</b>	0,0166	0,0181	0,0201	0,0265
	-0,4	-0,0050	-0,0058	-0,0066	-0,0077	-0,0105	0,0152	<b>0,0151</b>	0,0154	0,0159	0,0180
	0	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001	0,0150	0,0145	0,0143	0,0142	<b>0,0140</b>
	0,4	0,0050	0,0058	0,0066	0,0076	0,0105	0,0152	<b>0,0152</b>	0,0154	0,0160	0,0181
	0,8	0,0101	0,0116	0,0133	0,0153	0,0211	<b>0,0154</b>	0,0166	0,0181	0,0201	0,0265
1	-0,8	-0,0004	-0,0008	-0,0014	-0,0026	-0,0086	0,0050	<b>0,0050</b>	0,0051	0,0058	0,0118
	-0,4	-0,0002	-0,0004	-0,0007	-0,0014	-0,0043	0,0065	0,0063	<b>0,0063</b>	0,0064	0,0082
	0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0069	0,0067	0,0067	0,0066	<b>0,0066</b>
	0,4	0,0002	0,0003	0,0007	0,0013	0,0042	0,0064	0,0063	<b>0,0063</b>	0,0064	0,0081
	0,8	0,0004	0,0008	0,0014	0,0027	0,0086	0,0050	<b>0,0050</b>	0,0051	0,0058	0,0118

den vektoriellen Störterm  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})'$  verwenden wir eine multivariate Normalverteilung der Form

$$\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}) \quad \text{mit} \quad \boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \text{für} \quad t = 1, \dots, n. \quad (5.7)$$

Für den Korrelationskoeffizienten zwischen  $\varepsilon_{1t}$  und  $\varepsilon_{2t}$  nehmen wir  $\rho \in \{-0,8; -0,4; 0; 0,4; 0,8\}$  an und den Beobachtungsumfang fixieren wir mit  $n = 500$ , was eine relativ große Stichprobe darstellt. Kleinere Stichprobenumfänge liefern substantiell die gleichen Ergebnisse bezüglich der Vorteilhaftigkeit von OLS und NBLs und wir wollen deswegen auf ihre Darstellung verzichten. Auch nehmen wir an, dass keine Autokorrelation für die Störterme selbst existiert. Diese wiederum hat zwar einen Einfluss auf das Ausmaß der Verzerrung und den RMSE, bewirkt aber nicht eine unterschiedliche Rangfolge der einzelnen Schätzsituationen. Bei der Anwendung des NBLs-Schätzers verwenden wir die Bandbreitenregel  $m = \lfloor n^a \rfloor$  und wählen  $a \in \{0,4; 0,5; 0,6; 0,7\}$  und stellen die jeweilige ermittelte durchschnittliche Verzerrung und RMSE bei 100.000 Wiederholungen der einzelnen Parameterkonstellationen neben denen für OLS in der Tabelle 5.1 dar. Als Ergebnis kann man hierbei festhalten, dass der RMSE für die jeweiligen Schätzer symmetrisch um  $\rho$  ist. Das Vorzeichen der Korrelation hat aber einen Einfluss auf die Richtung der Verzerrung, so haben wir für negative Korrelation eine Unterschätzung der Kointegrationsbeziehung und für positive eine Überschätzung. Diese wird auch immer ausgeprägter, wenn die Kointegrationsstärke  $b$  abnimmt. Auch steigt der RMSE, wenn  $b$  kleiner wird. Interessant ist die Erkenntnis, dass bei der Anwendung von NBLs niedrige Bandbreiten ein besseres Ergebnis erzielen, wenn die Kointegrationsstärke niedrig und die Korrelation groß ist. Steigt nun  $b$  an, dann ist prinzipiell immer eine größere Bandbreite von Vorteil. OLS ist dem NBLs-Verfahren nur vorzuziehen, wenn keine Korrelation vorherrscht oder die Kointegrationsstärke sehr gering ist. Dabei ist aber anzumerken, dass die Unterschiede in den RMSE verhältnismäßig gering sind und der NBLs-Schätzer auch verzerrt ist, wenn auch in einem kleineren Umfang. Nielsen und Frederiksen (2011) waren imstande das Ausmaß der Verzerrung zu bestimmen und konnten, aufbauend auf dieser Erkenntnis, einen modifizierten NBLs-Schätzer mit einem Korrekturterm entwickeln. Dieser wiederum hängt von den Parametern  $d$  und  $b$  ab und es muss eine weitere Bandbreite  $m_1$  bestimmt werden, mit der der Korrekturterm geschätzt wird, was aber eine empirische Anwendung weiter erschwert. Auf eine Darstellung des Schätzers wollen wir hier verzichten und werden ihn auch im empirischen Teil dieser Arbeit nicht anwenden, da der Schätzer nur

für die Situation der stationären Kointegration anwendbar ist, die durch die Parameterkonstellation  $0 < b \leq d < 1/2$  charakterisiert werden kann. Vielmehr werden wir später noch andere Möglichkeiten zur Schätzung von Kointegrationsbeziehungen kennenlernen, die in der Lage sind, sämtliche Kointegrationsparameter  $d$ ,  $b$  und  $\beta$  in einem einzigen Schritt zu schätzen.

### 5.3 Bestimmung des Integrationsgrads von Kointegrationsfehlern

Nun wollen wir uns mit der Schätzung des Integrationsgrads von Kointegrationsfehlern  $u_t = \Delta^{-(d-b)}\varepsilon_{1t}$  und damit indirekt mit der Schätzung der Kointegrationsstärke  $b$  beschäftigen. Hier haben wir das Problem, dass wir in der Regel den Kointegrationsfehler nicht beobachten können und somit schätzen müssen mit

$$\hat{u}_t = y_t - \hat{\beta}x_t, \quad (5.8)$$

wobei  $\hat{\beta}$  mit einem geeigneten Schätzverfahren bestimmt wird. Da wir nun die Berücksichtigung des Endogenitätsproblems zunächst vernachlässigen wollen, können wir hierfür durchaus den OLS-Schätzer verwenden. Für den Fall der klassischen Kointegration existieren seit längerem Studien, die sich mit den Auswirkungen der geschätzten Kointegrationsbeziehungen auf Einheitswurzeltests beschäftigen, wie z.B. die Arbeit von MacKinnon (1991), der sich mit der Anwendung des ADF-Tests beschäftigt. Er kommt dabei zu dem Ergebnis, dass eigens simulierte kritische Werte für den Kointegrationsfall verwendet werden sollen, da sonst mit einer größeren Wahrscheinlichkeit als eine angenommene Irrtumswahrscheinlichkeit fälschlicherweise Kointegrationsbeziehungen angenommen werden.

Im Kontext der fraktionalen Kointegration existieren bereits einige Studien, die sich mit den Auswirkungen der Schätzung von Kointegrationsbeziehungen auseinandersetzen. Dittmann (2000) thematisiert den GPH-Schätzer für den Integrationsgrad der Kointegrationsresiduen bei Fixierung von  $d = 1$  in seiner Studie und kommt dabei zu dem Schluss, dass das Schätzen von  $\beta$  einen Einfluss auf die Verteilung der Schätzergebnisse hat und der Integrationsgrad der Residuen tendenziell unterschätzt wird. Auf Basis des t-Tests für den GPH-Schätzer gibt Dittmann (2000) eigene kritische Werte in Abhängigkeit des Beobachtungsumfangs  $n$  und der gewählten Bandbreite  $m$  an und vergleicht die Macht des GPH-t-Tests mit klassischen Integrationstests wie dem ADF- und dem Phillips-Perron-Test, wobei diese für den klassischen  $I(1)/I(0)$ -Kointegrationsfall ausgelegt sind. Dabei kommt er erstaunli-

cherweise zum Ergebnis, dass die Testentscheidungen auf Basis des GPH-Schätzers wesentlich schwächer zu Gunsten der fraktionalen Kointegration ausfallen als bei den klassischen Tests, obwohl er auch nicht-stationäre Kointegrationsfehler in seinen Simulationen mit einschließt.

Hassler, Marmol und Velasco (weiterhin: HMV, 2006) können hingegen zeigen, dass der GPH-Schätzer erwartungstreu den Integrationsgrad von Kointegrationsresiduen schätzen kann und sogar asymptotisch normalverteilt für den Fall normalverteilter Störterme ist, falls ein geeigneter Schätzer wie OLS oder NBLs für  $\beta$  gewählt wird. HMV (2006) untersuchen auch den Fall, dass ein kointegriertes System von mehr als einem stochastischen Trend getrieben wird, empfehlen aber für diesen Fall nicht ein sequentielles Verfahren zum Testen der Kointegrationsbeziehung, sondern favorisieren multivariate Schätzverfahren, auf die wir auch in einem späteren Teil dieser Arbeit eingehen werden.

Gil-Alana (2003) untersucht mit dem bereits vorgestellten LM-Test einen fraktionalen Integrationstest im Zusammenhang mit fraktionaler Kointegration und folgert, dass die Verteilung der Testgröße durch die Schätzung der Kointegrationsbeziehung beeinflusst wird. Er gibt deswegen auch eigene simulierte kritische Werte in Abhängigkeit von  $d$  und  $n$  an, verzichtet aber bei deren Darstellung auf Response-Surface-Funktionen, was somit deren Anwendung bei empirischen Arbeiten erschwert. Im Folgenden wollen wir simulationsbasiert analysieren, ob das Schätzen von Kointegrationsbeziehungen einen Einfluss auf die Verteilung von Integrationstests oder Integrationsgradschätzungen haben. Dazu betrachten wir im ersten Schritt den exakten lokalen Whittle-Schätzer, da er den Vorzug hat, dass wir bei der Schätzung des Integrationsgrades keine Annahme über den Wertebereich des Parameters machen müssen und im zweiten Schritt den effizienten Wald-Test, da dieser mitunter die beste Macht der fraktionalen Integrationstests aufweist, wie wir in Abschnitt 4.6 gesehen haben.

### 5.3.1 Der ELW-Schätzer im Kointegrationsfall

Um das Verhalten des ELW-Schätzers im Kointegrationsfall zu untersuchen, simulieren wir für einen Beobachtungsumfang von  $n = 1000$  zwei nicht-kointegrierte Prozesse  $x_t$  und  $y_t$  mit Integrationsgrad  $d = 1$  und schätzen zum einen eine Kointegrationsbeziehung durch eine OLS-Regression von  $x_t$  und  $y_t$ . Daraufhin schätzen wir den Integrationsgrad der Residuen mit ELW unter Verwendung der Bandbreite  $m = \lfloor 1000^{0,8} \rfloor = 251$ . Zum anderen nehmen wir an, dass  $\beta = 0$  bekannt ist, was wiederum bedeutet, dass keine Kointegration zwischen  $x_t$  und  $y_t$  vorliegt. Für die-

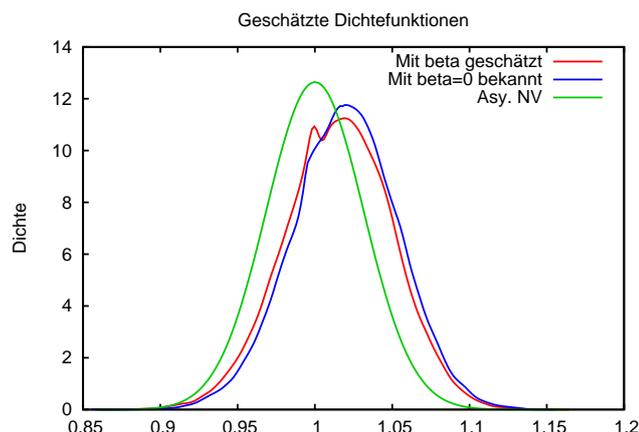


Abb. 5.2: Geschätzte Dichtefunktionen des ELW-Schätzergebnisse der Kointegrationsresiduen für  $\beta = 0$

sen Fall schätzen wir den Integrationsgrad mit der gleichen Bandbreite direkt für den simulierten Prozess  $y_t$ . In der Abbildung 5.2 stellen wir die geschätzte Dichtefunktion für die beiden Situationen dar und vergleichen dies mit der impliziten asymptotischen Normalverteilung des ELW-Schätzergebnisses mit Erwartungswert 1 und Standardabweichung  $\sqrt{1/(4 \cdot 251)} \approx 0,0316$ . Hieraus können wir erkennen, dass der Integrationsgrad in beiden Situationen leicht überschätzt wird, was wir auch aus den Ergebnissen des vorangegangenen Kapitels erwarten konnten. Da sich beide Dichtefunktionen sehr ähneln, können wir annehmen, dass ein geschätzter Kointegrationsparameter  $\hat{\beta}$  keinen Effekt auf die Verteilung des geschätzten Integrationsgrads der Residuen hat. Um dies näher zu betrachten simulieren wir nun diese Situationen für weitere Stichprobengrößen und geben jeweils das 5%- und das 95%-Quantil der Schätzergebnisse in Tabelle 5.2. Des Weiteren betrachten wir hier nicht nur die fixierte Bandbreite  $m = n^{0,8}$ , sondern wählen auch eine kleinere Bandbreite  $m = n^{0,6}$ . Aus den Ergebnissen können wir nun den Schluss ziehen, dass die

Tab. 5.2: Simulierte Quantile für den geschätzten Integrationsgrad der Kointegrationsresiduen

n	5%-Quantil				95%-Quantil			
	$m = \lfloor n^{0,6} \rfloor$		$m = \lfloor n^{0,8} \rfloor$		$m = \lfloor n^{0,6} \rfloor$		$m = \lfloor n^{0,8} \rfloor$	
	unbek.	bek.	unbek.	bek.	unbek.	bek.	unbek.	bek.
100	0,6357	0,7094	0,8658	0,8907	1,2526	1,2815	1,2275	1,2414
250	0,7408	0,7827	0,9176	0,9299	1,1640	1,1868	1,1373	1,1450
500	0,8083	0,8325	0,9410	0,9490	1,1274	1,1411	1,0985	1,1034
1000	0,8548	0,8732	0,9573	0,9624	1,0993	1,1102	1,0712	1,0746

Hier steht bek. für die Annahme, dass die Kointegrationsresiduen  $y_t - \beta x_t$  bekannt seien, und unbek., dass wir  $\beta$  mit OLS schätzen.

Tab. 5.3: Durchschnittliche Verzerrung und RMSE für  $\hat{b}$  bei Anwendung des ELW-Schätzers für die Kointegrationsresiduen

$n$	$b$	Verzerrung				RMSE			
		$m = \lfloor n^{0,6} \rfloor$		$m = \lfloor n^{0,8} \rfloor$		$m = \lfloor n^{0,6} \rfloor$		$m = \lfloor n^{0,8} \rfloor$	
		unbek.	bek.	unbek.	bek.	unbek.	bek.	unbek.	bek.
250	0	-0,0368	-0,0080	0,0293	0,0397	0,1281	0,1218	0,0669	0,0651
	0,2	-0,0200	-0,0015	0,0422	0,0400	0,1265	0,1144	0,0686	0,0655
	0,4	-0,0187	-0,0061	0,0433	0,0398	0,1314	0,1172	0,0643	0,0599
	0,6	-0,0337	-0,0059	0,0333	0,0435	0,1376	0,1290	0,0718	0,0697
	0,8	-0,0401	-0,0074	0,0285	0,0397	0,1314	0,1237	0,0676	0,0652
	1	-0,0379	-0,0078	0,0286	0,0396	0,1282	0,1220	0,0672	0,0649
500	0	-0,0264	-0,0071	0,0212	0,0276	0,0975	0,0936	0,0477	0,0469
	0,2	-0,0136	-0,0046	0,0322	0,0274	0,0973	0,0895	0,0490	0,0467
	0,4	-0,0086	-0,0015	0,0324	0,0258	0,0987	0,0897	0,0458	0,0434
	0,6	-0,0230	-0,0013	0,0234	0,0290	0,1027	0,0976	0,0486	0,0473
	0,8	-0,0280	-0,0059	0,0209	0,0276	0,0984	0,0929	0,0478	0,0465
	1	-0,0268	-0,0061	0,0207	0,0276	0,0978	0,0938	0,0476	0,0464
1000	0	-0,0195	-0,0057	0,0149	0,0191	0,0744	0,0720	0,0347	0,0342
	0,2	-0,0071	-0,0048	0,0252	0,0194	0,0763	0,0715	0,0363	0,0341
	0,4	-0,0030	-0,0090	0,0248	0,0163	0,0741	0,0689	0,0337	0,0325
	0,6	-0,0164	-0,0024	0,0167	0,0200	0,0769	0,0739	0,0347	0,0340
	0,8	-0,0196	-0,0044	0,0151	0,0193	0,0751	0,0717	0,0348	0,0340
	1	-0,0200	-0,0056	0,0148	0,0191	0,0750	0,0723	0,0347	0,0341

Hier steht bek. für die Annahme, dass die Kointegrationsresiduen  $y_t - \beta x_t$  bekannt seien, und unbek., dass wir  $\beta$  mit OLS schätzen.

Quantile für den Fall der größeren Bandbreite sich tendenziell zwar unterscheiden, wenn wir  $\beta$  schätzen, was aber eher an einer leicht höheren Volatilität der Schätzergebnisse liegen könnte. Größere Unterschiede können wir nur bei Verwendung kleinerer Bandbreiten bei kleinen Stichproben beobachten. Ansonsten nähern sich die jeweils geschätzten Quantile immer näher, wenn  $n$  und respektive  $m$  steigt.

Nun weiten wir unsere Analyse noch weiter aus und untersuchen im letzten Schritt die Schätzeigenschaften von ELW noch in Abhängigkeit der Kointegrationsstärke  $b$ . Diesen Parameter können wir bestimmen, indem wir den Integrationsgrad des Kointegrationsfehlers  $u_t$  mit ELW schätzen. Dieser beträgt im Modellrahmen der fraktionalen Kointegration  $d - b$ , wobei wir  $d = 1$  fixieren und zur Schätzung der Kointegrationsstärke als bekannt annehmen. Somit ergibt sich  $\hat{b} = 1 - \hat{d}_u$ , wobei  $\hat{d}_u$  den geschätzten Integrationsgrad von  $u_t$  darstellt, als geschätzte Kointegrationsstärke. In Tabelle 5.3 vergleichen wir die durchschnittliche Verzerrung und RMSE bei 20.000 Wiederholungen je Parameterkonstellation mit den Bandbreitenregeln  $m = n^{0,6}$  und  $m = \lfloor n^{0,8} \rfloor$  und betrachten wiederum die beiden Situationen, dass  $\beta$  bekannt sei oder geschätzt werden muss. Hier bestätigen sich zum Teil unsere Vermutungen: Für die geschätzten Kointegrationsresiduen ergeben sich durchweg volatile Schätzergebnisse bei jeder Bandbreite und Stichprobenumfang. Etwas überrascht hier aber, dass die Differenz im Ausmaß der Streuung nicht allzu groß ausfällt und

scheinbar asymptotisch zu verschwinden scheint. Bei großen Stichprobenumfängen und großer Bandbreite liegt der RMSE nur noch geringfügig über der Situation mit bekannten Residuen. Ähnlich verhält es sich für die durchschnittliche Verzerrung: Bei kleinen Stichproben und Bandbreiten wird die Kointegrationsstärke stärker unterschätzt, wenn die Residuen unbekannt sind. Diese wiederum lässt den Schluss nahe, dass bei unbekanntem Residuen bei Anwendung des ELW-Schätzers tendenziell leicht häufiger eine Kointegrationsbeziehung angenommen wird. Die Verzerrungen verschwinden aber für größere Stichproben und größere Bandbreiten, so dass wir durchaus zu dem Ergebnis kommen können, dass die asymptotische Verteilung für das ELW-Schätzergebn auch im Kointegrationsfall standhält und wir keine separaten kritischen Werte bzw. Verteilungen simulieren brauchen.

### 5.3.2 Der effiziente Wald-Test im Kointegrationsfall

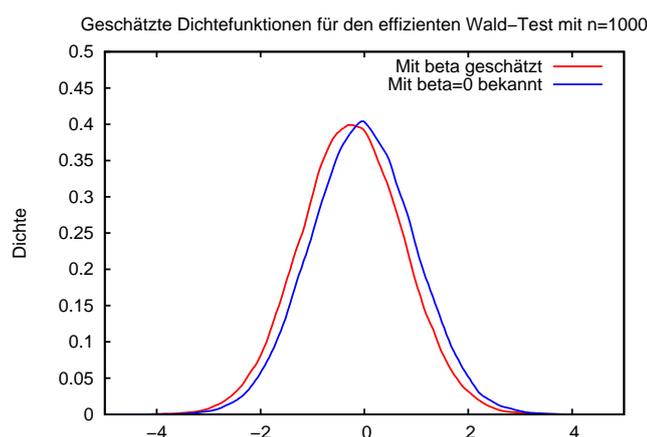


Abb. 5.3: Geschätzte Dichtefunktionen für den effizienten Wald-Test im Kointegrationsfall

In diesem Abschnitt wollen wir nun die Eigenschaften des effizienten Wald-Tests im Kointegrationsfall, d.h. wenn der zu untersuchende Prozess  $\hat{u}_t$  durch eine geschätzte Regression zustande kommt, näher betrachten. Das Vorgehen der Analyse läuft dabei analog wie für den ELW-Schätzer und wir betrachten zunächst zwei nicht-kointegrierte Prozesse  $x_t$  und  $y_t$  der Integrationsordnung  $d = 1$ , wobei wir auch hier wieder eine Kointegrationsbeziehung mit OLS schätzen bzw. wir  $\beta = 0$  als bekannt annehmen. Daraufhin testen wir mit dem effizienten Wald-Test für die Residuen aus dieser Beziehung die Hypothese  $H_0 : d_u = 1$  gegen  $H_1 : d_u < 1$ . Auch hier fixieren wir den Beobachtungsumfang  $n = 1000$  und nehmen keine zusätzliche Autokorrelation in Form von  $AR(p)$ -Bestandteilen in den Störtermen von  $x_t$  und  $y_t$

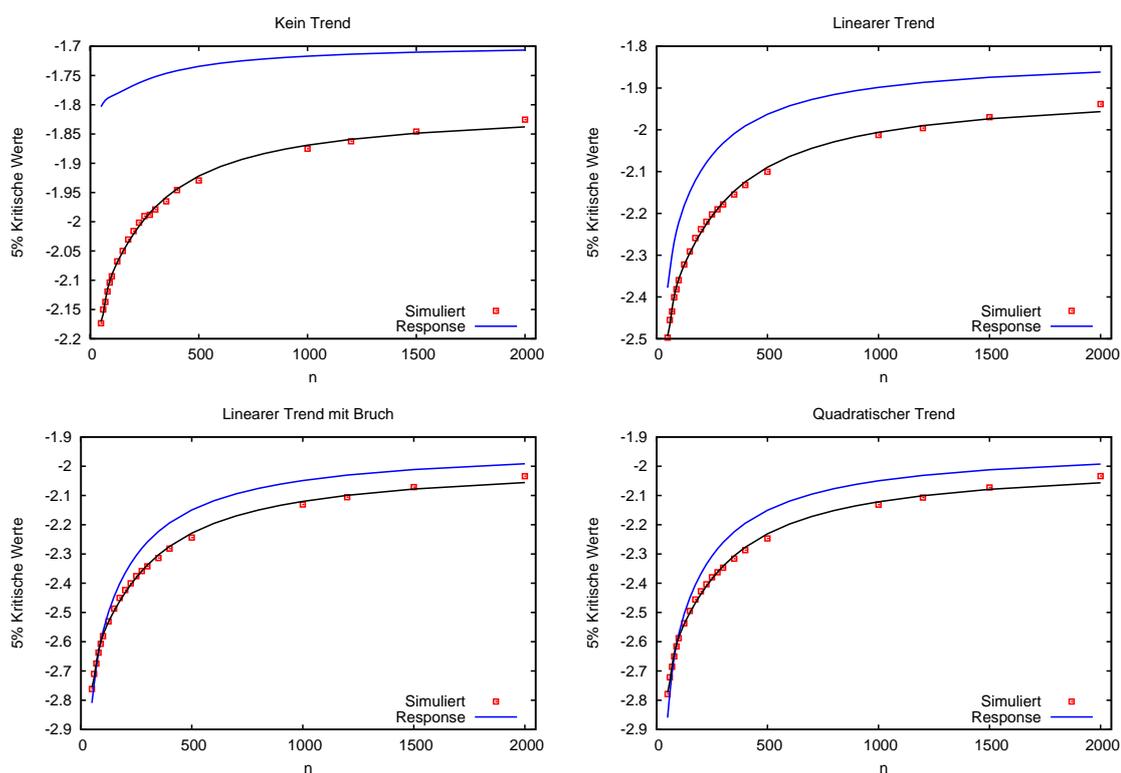


Abb. 5.4: Kritische Werte für den effizienten Wald-Test im Kointegrationsfall im Vergleich zum univariaten Integrationsfall

an, so dass wir die Lagzahl  $p = 0$  setzen können. In Abbildung 5.3 betrachten wir zunächst die geschätzte Dichtefunktion für den Fall, dass keinerlei deterministischen Bestandteile in  $u_t$  vorliegen für die beiden Situationen, dass  $\beta$  geschätzt wurde oder bekannt ist. Dabei ist anzumerken, dass die Situation mit bekanntem  $\beta = 0$  genau der univariaten Anwendung des effizienten Wald-Tests in Abschnitt 4.4.2 entspricht.

Hier lässt sich bereits erkennen, dass wir eine Verschiebung der Verteilung haben und dass wahrscheinlich eine Anpassung der kritischen Werte, die wir im Abschnitt 4.4.2 angegeben haben, notwendig wird. Um nun die Verschiebung genauer zu betrachten, stellen wir in Abbildung 5.4 die simulierten 5%-kritischen-Werte für den Kointegrationsfall für sämtliche angenommenen Trendoptionen gegenüber den Response-Surface-Funktionen aus der Tabelle 4.21 in Abhängigkeit von  $n$  dar. Diese Betrachtung lässt wiederum den Schluss nahe, dass wir auch asymptotisch unterschiedliche kritische Werte verwenden müssen, um größere Verzerrungen bei der Entscheidung über eine Kointegrationsbeziehung zu vermeiden. Falls wir die univariaten kritischen Werte im Kointegrationsfall anwenden, werden wir mit einer größeren Wahrscheinlichkeit als einer vorher fixierten Irrtumswahrscheinlichkeit ei-

Tab. 5.4: Geschätzte Response-Surface-Funktionen für die Kritische Werte des effizienten Wald-Tests für den Kointegrationsfall

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
	N	LT	LTB	QT	N	LT	LTB	QT	N	LT	LTB	QT
$\gamma_0$	-1,4419 (0,0078)	-1,5396 (0,0111)	-1,6223 (0,0136)	-1,6238 (0,0139)	-1,8028 (0,0073)	-1,9010 (0,0107)	-1,9836 (0,0131)	-1,9838 (0,0136)	-2,4865 (0,0063)	-2,5814 (0,0102)	-2,6655 (0,0148)	-2,6648 (0,0136)
$\gamma_1$	-75,811 (5,4281)	-119,12 (7,8782)	-155,51 (9,611)	-156,32 (9,8077)	-73,881 (5,3303)	-117,00 (7,9114)	-152,42 (9,3356)	-153,68 (9,8589)	-70,343 (4,7626)	-113,01 (7,0002)	-146,63 (10,1177)	-148,36 (9,2365)
$\gamma_2$	8458,7 (1201,5)	12967 (1756,5)	17167 (2141,7)	17229 (2182,2)	8039,7 (1192,5)	12737 (1827,2)	16707 (2107,9)	16766 (2251,8)	7471,9 (1137,6)	11932 (1479,8)	15699 (2169,7)	15741 (1948,3)
$\gamma_3$	-471175 (99124)	-708198 (145244)	-950178 (176436)	-956667 (179649)	-442126 (98620)	-700836 (154440)	-926843 (175771)	-928097 (188973)	-417398 (97338)	-651706 (117138)	-876084 (174514)	-872163 (154181)
$\gamma_4$	9609429 (2629932)	14193785 (3857764)	19214319 (4677164)	19409101 (4760488)	8940029 (2618382)	14119718 (4148872)	18791580 (4692678)	18757854 (5065016)	8494236 (2626959)	12974915 (3029686)	17859902 (4561997)	17657801 (3986075)
$R^2$	99,78% $4,67 \cdot 10^{-3}$	99,80% $5,91 \cdot 10^{-3}$	99,83% $7,24 \cdot 10^{-3}$	99,84% $7,28 \cdot 10^{-3}$	99,73% $5,24 \cdot 10^{-3}$	99,81% $5,71 \cdot 10^{-3}$	99,85% $6,94 \cdot 10^{-3}$	99,85% $7,13 \cdot 10^{-3}$	99,86% $4,05 \cdot 10^{-3}$	99,88% $4,85 \cdot 10^{-3}$	99,89% $6,10 \cdot 10^{-3}$	99,87% $6,89 \cdot 10^{-3}$

N steht für das Vorliegen von keinem Trend, M für einen Mittelwert, LT einem linearen Trend und LTB einem linearen Trend mit Bruch.

Tab. 5.5: Vergleich der asymptotischen kritischen Werte für den effizienten Wald-Test im Integrations- und Kointegrationsfall

	$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$		
	Int.	Koint.	Rel. Diff.	Int.	Koint.	Rel. Diff.	Int.	Koint.	Rel. Diff.
N	-1,3376	-1,4419	7,8%	-1,6951	-1,8028	6,4%	-2,3727	-2,4865	4,8%
LT	-1,4603	-1,5396	5,4%	-1,8216	-1,9010	4,4%	-2,5029	-2,5814	3,1%
LTB	-1,5656	-1,6223	3,6%	-1,9292	-1,9836	2,8%	-2,6103	-2,6655	2,1%
QT	-1,5652	-1,6238	3,7%	-1,9302	-1,9838	2,8%	-2,6052	-2,6648	2,3%

ne Kointegrationsbeziehung fälschlicherweise annehmen. In der Abbildung 5.4 sind auch jeweils Polynome vierten Grades von  $n^{-1}$  zur Erklärung der simulierten kritischen Werte im Kointegrationsfall aufgenommen, welche wir auch zur Beschreibung der Response-Surface-Funktion

$$CV_{\alpha}(n) = \gamma_0 + \gamma_1 n^{-1} + \gamma_2 n^{-2} + \gamma_3 n^{-3} + \gamma_4 n^{-4} + \text{Resid} \quad (5.9)$$

für die kritischen Werte verwenden wollen. Die hierzu geschätzten Parameter sind in der Tabelle 5.4 dargestellt. Für die Situation des linearen Trends mit Bruch verwenden wir wieder wie in Abschnitt 4.4.2 zur Ermittlung der kritischen Werte die Fixierung des relativen Bruchzeitpunkts  $\tau = 0,5$ , da sich bei Betrachtung der kritischen Werte auch hier wieder ein Minimum ergibt und dies uns die Modellierung mit Hilfe der Response-Surface-Funktion vereinfacht. Abschließend vergleichen wir in Tabelle 5.5 die asymptotischen kritischen Werte des effizienten Wald-Tests, die wir den Response-Surface-Funktionen entnehmen können und geben die relative Differenz zwischen diesen an. Die univariate Variante aus Abschnitt 4.4.2 entspricht einem reinen Integrationstest, da hier alleine der Integrationsgrad einer zu untersuchenden Zeitreihe geschätzt wird, während wir im Kointegrationsfall mit dem Integrationsgrad des geschätzten Kointegrationsfehlers gleichzeitig das Vorliegen einer Kointegrationsbeziehung testen. Die kritischen Werte unterscheiden sich insbesondere im Fall ohne deterministische Komponenten am stärksten, während bei Aufnahme von verschiedenen Trendoptionen die kritischen Werte sich annähern.

### 5.3.3 Vergleichende Monte-Carlo-Studie

In diesem Abschnitt wollen wir die Mächtigkeit des ELW-Verfahrens und des effizienten Wald-Tests im Falle der fraktionalen Kointegration untersuchen. Unter der Fixierung  $d = 1$ , die wir in dieser Studie als bekannt voraussetzen, können wir eine Kointegrationsbeziehung annehmen, wenn wir den Integrationsgrad des Kointegrationsfehlers  $d_u$  signifikant kleiner als 1 testen können. Als datengenerierendes System

Tab. 5.6: Ablehnungsraten für keine Kointegration ( $b = 0$ ) bei Anwendung verschiedener Integrationstests und fixiertem Irrtumsniveau  $\alpha = 5\%$

$n$	$b$	$\theta = 0$						$\theta = 0.5$								
		ADF			NEW			ADF			NEW			ELW		
		$p = 0$	$p = p_s$	$p = p_1$	$p = 0$	$p = p_s$	$p = p_1$	$p = 1$	$p = p_s$	$p = p_1$	$p = 1$	$p = p_s$	$p = p_1$	$m_1$	$m_2$	$m_3$
100	0	5,9%	4,0%	1,6%	6,7%	13,2%	9,9%	3,1%	3,8%	1,5%	6,9%	12,9%	8,3%	7,7%	1,9%	0,4%
	0.2	36,7%	8,3%	2,5%	61,3%	21,7%	13,1%	9,4%	6,8%	2,4%	18,9%	22,0%	12,5%	21,7%	4,4%	0,3%
	0.4	92,0%	20,6%	4,0%	99,0%	40,0%	16,8%	25,8%	13,6%	4,0%	46,4%	38,6%	21,3%	56,3%	27,4%	2,6%
	0.6	99,9%	47,0%	6,8%	100%	62,7%	23,3%	61,8%	29,6%	5,5%	80,8%	64,0%	34,8%	88,7%	78,6%	36,1%
	0.8	100%	80,9%	11,3%	100%	82,7%	31,3%	94,0%	55,1%	9,2%	97,1%	89,2%	52,9%	98,7%	98,4%	90,1%
	1	100%	97,4%	19,4%	100%	93,5%	40,6%	99,9%	82,7%	15,0%	99,9%	98,6%	74,6%	99,9%	100%	99,8%
250	0	5,5%	4,6%	2,9%	5,8%	13,5%	14,4%	3,6%	4,3%	2,8%	3,3%	13,1%	12,0%	6,6%	1,1%	0,1%
	0.2	59,1%	14,7%	5,6%	95,3%	34,6%	21,1%	17,4%	11,9%	5,4%	27,1%	30,7%	21,2%	37,0%	9,0%	0,1%
	0.4	99,8%	50,0%	12,9%	100%	73,7%	35,8%	65,6%	36,3%	11,4%	82,2%	67,4%	39,8%	89,3%	75,7%	10,3%
	0.6	100%	93,0%	29,7%	100%	95,6%	53,4%	99,3%	78,5%	24,4%	99,5%	96,1%	68,0%	99,8%	99,8%	92,1%
	0.8	100%	100%	58,0%	100%	99,6%	71,4%	100%	99,1%	49,0%	100%	99,9%	92,5%	100%	100%	100%
	1	100%	100%	86,7%	100%	99,9%	85,1%	100%	100%	78,3%	100%	100%	99,7%	100%	100%	100%
500	0	5,2%	4,8%	3,6%	5,4%	13,1%	15,9%	4,1%	4,7%	3,5%	1,9%	11,2%	12,6%	6,7%	0,8%	0,1%
	0.2	74,7%	24,7%	9,3%	99,9%	54,3%	30,5%	28,9%	19,9%	8,8%	43,5%	44,2%	27,6%	58,3%	22,9%	0,0%
	0.4	100%	82,7%	28,8%	100%	96,6%	57,1%	92,9%	70,2%	25,7%	98,5%	93,0%	60,3%	99,1%	98,4%	41,1%
	0.6	100%	99,9%	68,9%	100%	99,9%	82,6%	100%	99,5%	62,5%	100%	99,9%	93,3%	100%	100%	99,9%
	0.8	100%	100%	97,1%	100%	100%	95,2%	100%	100%	94,4%	100%	100%	99,9%	100%	100%	100%
	1	100%	100%	99,9%	100%	100%	99,0%	100%	100%	99,9%	100%	100%	100%	100%	100%	100%

ADF umschreibt den erweiterten Dickey-Fuller-Test, NEW den effizienten Wald-Test und ELW den exakten lokalen Whittle-Schätzer mit den Bandbreiten  $m_1 = \lfloor n^{0,6} \rfloor$ ,  $m_2 = \lfloor n^{0,7} \rfloor$  und  $m_3 = \lfloor n^{0,8} \rfloor$ .

nehmen wir für die folgende Studie

$$\begin{aligned} y_t - x_t = u_t &= \Delta \overbrace{b^{-1}}^{-d_u} (1 - \theta L) \varepsilon_{1t} \quad \text{mit } \theta \in \{0; 0,5\} \\ x_t &= \Delta^{-1} \varepsilon_{2t} \quad \boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim N(\mathbf{0}; \mathbf{I}_2) \end{aligned} \quad (5.10)$$

an.  $\mathbf{I}_2$  stellt hierbei die Einheitsmatrix der Dimension 2, was wiederum bedeutet, dass wir hier keine Korrelation der Störterme  $\varepsilon_{1t}$  und  $\varepsilon_{2t}$  und damit keine Endogenität annehmen. Den Integrationsgrad des stochastischen Trends  $y_t$  nehmen wir durchweg als bekannt an und variieren den Beobachtungsumfang  $n$  und die Kointegrationsstärke  $b$ . Des Weiteren berücksichtigen wir in unserer Studie auch den ADF-Test als klassischen  $I(1)/I(0)$ -Test, für den wir die kritischen Werte im Kointegrationsfall nach MacKinnon (1991) verwenden. Für das ELW-Verfahren verwenden wir drei verschiedene Bandbreiten zur Schätzung des Integrationsgrads, aus denen sich auch die asymptotischen Standardabweichungen ergeben, die wir für die jeweiligen t-Statistiken für die Hypothese  $H_0 : d_u = 1$  verwenden werden. Für die Lagzahl  $p$ , die wir für den effizienten Wald-Test und den ADF-Test bestimmen müssen, wählen wir zum einen die wahre Lagzahl  $p = 0$  (für  $\theta = 0$ ) und  $p = 1$  (für  $\theta = 0,5$ ) sowie die beiden Schwert-Heuristiken  $p_s = \lfloor 4(n/100)^{1/4} \rfloor$  und  $p_l = \lfloor 12(n/100)^{1/4} \rfloor$ .

Als Ergebnis lässt sich bei Betrachtung der Tabelle 5.6 festhalten, dass der ADF-Test eine überraschend gute Macht auch für den Fall der fraktionalen Kointegration besitzt. Er ist zwar in sämtlichen Belangen dem effizienten Wald-Test unterlegen, bleibt aber zumeist nicht weit hinter diesem zurück. Allenfalls bei kleinen Stichproben bei gleichzeitiger Anwendung einer sehr großen Lagzahl  $p$  fällt der ADF-Test sehr schlecht aus. Die Macht ist auch bei nicht-stationären Kointegrationsfehlern zwar eher schwach, nichtsdestotrotz können auch solche Kointegrationsbeziehungen bei großen Stichproben durchaus gefunden werden. Das ELW-Verfahren schneidet tendenziell besser ab als der effiziente Wald-Test unter der Bedingung, dass die Bandbreite für die Schätzung des Integrationsgrads geeignet gewählt wird. Bei fehlender Autokorrelation des Kointegrationsstörterms kann es aber häufig zur Annahme einer Kointegrationsbeziehung kommen, obwohl keine vorliegt, da der Integrationsgrad bei zu kleiner Bandbreite und endlicher Stichprobe unterschätzt wird. Falls aber Autokorrelation vorliegt, dann existiert das Problem, dass bei zu hoher Bandbreite das ELW-Verfahren nahezu keine Macht besitzt, fraktionale Kointegrationsbeziehungen aufzuzeigen. Deswegen ist es gerade bei diesem Verfahren ratsam, die Schätzergebnisse über mehrere Bandbreiten zu betrachten und zudem auch Testergebnisse des effizienten Wald-Tests zu beachten. Weitere Evidenz für das Vorliegen von Kointe-

grationsbeziehungen kann aber auch zusätzlich durch die Anwendung von klassischen Integrationstests gewonnen werden, da diese auch bei fraktionalen kointegrierten Systemen eine bemerkenswerte Macht besitzen.

## 5.4 Multivariate Schätzverfahren von fraktional kointegrierten Systemen

Als Alternative zu den bereits vorgestellten sequentiellen Verfahren bieten sich multivariate Schätzverfahren an, die die Parameter der fraktionalen Kointegration  $d$ ,  $b$  und  $\beta$  in einem einzigen Schritt schätzen können. Dazu werden wir zunächst mit der multivariaten Erweiterung des exakten lokalen Whittle-Schätzers nach Shimotsu (2012) ein semiparametrisches Verfahren vorstellen und im darauffolgenden Abschnitt mit der verallgemeinerten Fehlerkorrekturdarstellung für fraktional kointegrierte Prozesse nach Johansen und Nielsen (2012a) ein parametrisches Verfahren. Die Vorteile dieser gemeinsamen Ansätze liegen insbesondere darin, dass möglicherweise auftretende Endogenitätsprobleme berücksichtigt und die Kointegrationsparameter konsistent geschätzt werden können. Abschließend werden wir die beiden Verfahren in einer Monte-Carlo-Simulationsstudie miteinander vergleichen.

### 5.4.1 Exakte lokale Whittle-Schätzung im Falle der fraktionalen Kointegration

Ansatzpunkt der Anwendung des multivariaten Whittle-Schätzers ist, ähnlich wie im univariaten Fall, die Approximation der Spektraldichte eines Vektorprozesses  $\mathbf{w}_t$ . Nielsen (2007) diskutiert die Anwendung im stationären Kointegrationsfall (d.h.  $0 < b \leq d < 1/2$ ) und definiert  $\mathbf{w}_t = (x_t, y_t - \beta x_t)'$ . Für diesen Vektorprozess ergibt sich die in der Nähe des Ursprungs approximierte Spektraldichtematrix

$$\mathbf{f}_{\mathbf{w}}(\lambda) = \mathbf{\Lambda}(\lambda_j)^{-1} \mathbf{G} \mathbf{\Lambda}(\lambda_j)^{-1} \quad \text{für } \lambda \rightarrow 0+ \quad (5.11)$$

mit  $\mathbf{\Lambda}(\lambda_j) = \text{diag}(\lambda_j^d, \lambda_j^{d-b})$  und  $\mathbf{G}$  als eine reellwertige symmetrische Matrix. Die Spektraldichte eines Vektorprozesses lässt sich so erklären, dass auf der Hauptdiagonalen die Spektraldichte der einzelnen Komponenten steht und auf der Nebendiagonalen die Kreuzspektraldichte, was wiederum deren Symmetrie gewährleistet. Der Parametervektor  $\boldsymbol{\psi} = (d, b, \beta)'$  lässt sich über die Minimierung der konzentrierten

Zielfunktion

$$R_m(\boldsymbol{\psi}) = \ln \det \hat{\mathbf{G}}(\boldsymbol{\psi}) - 2(2d - b) \frac{1}{m} \sum_{j=1}^m \log \lambda_j$$

mit  $\hat{\mathbf{G}}(\boldsymbol{\psi}) = \frac{1}{m} \sum_{j=1}^m \operatorname{Re}[\boldsymbol{\Lambda}(\lambda_j) \mathbf{I}_{\mathbf{w}}(\lambda_j) \boldsymbol{\Lambda}(\lambda_j)]$  (5.12)

schätzen bei vorheriger Festlegung der Bandbreite  $m$ , die wie im univariaten Fall der Bedingung  $1/m + m/n \rightarrow 0$  für  $n \rightarrow \infty$  genügen muss. Das Periodogramm von  $\mathbf{w}_t$  ergibt sich als  $\mathbf{I}_{\mathbf{w}}(\lambda_j) = (2\pi n)^{-1} (\sum_{t=1}^n \mathbf{w}_t e^{it\lambda_j}) (\sum_{t=1}^n \mathbf{w}'_t e^{it\lambda_j})$ . Wie bereits erwähnt ist diese Schätzweise nur für den Fall der stationären Kointegration anwendbar. Eine Erweiterung für nicht-stationäre Kointegrationsbeziehungen kann entweder durch vorherige Differenzbildung erfolgen, so dass das zu untersuchende System wiederum stationär ist, oder durch Anwendung von Daten-Tapering wie es Velasco (1999) für den univariaten Fall vorgeschlagen hat. In diesem Fall ist der Schätzer zwar konsistent, wird aber ineffizient im Vergleich zur Differenzenbildung. Aus diesem Grund hat Shimotsu (2012) die Verwendung des exakten lokalen Whittle-Schätzers vorgeschlagen, in der nun der differenzierte Vektorenprozess

$$\begin{pmatrix} \Delta^d & 0 \\ 0 & \Delta^{d-b} \end{pmatrix} \mathbf{w}_t = \boldsymbol{\Delta}^\delta \mathbf{w}_t = \mathbf{u}_t \quad (5.13)$$

mit  $\boldsymbol{\delta} = (d, d - b)'$  im Mittelpunkt der Whittle-Approximation steht. Das Schätzproblem gestaltet sich dann als

$$\hat{\boldsymbol{\psi}} = \arg \min R_m^*(\boldsymbol{\psi})$$

$$R_m^*(\boldsymbol{\psi}) = \ln \det \hat{\mathbf{G}}^*(\boldsymbol{\psi}) - 2(2d - b) \frac{1}{m} \sum_{j=1}^m \log \lambda_j$$

mit  $\hat{\mathbf{G}}^*(\boldsymbol{\psi}) = \frac{1}{m} \sum_{j=1}^m \operatorname{Re}[\mathbf{I}_{\boldsymbol{\Delta}^\delta \mathbf{w}}(\lambda_j; \boldsymbol{\psi})]$ , (5.14)

wobei eine gewichtete Mittelwertkorrektur der einzelnen Komponenten von  $\mathbf{w}_t$  wie für den univariaten ELW-Schätzer angewendet werden sollte. Zur Anwendung des Schätzverfahrens schlägt Shimotsu (2012) ein zweistufiges Verfahren vor: Im ersten Schritt sollen mit Hilfe des lokalen Whittle-Schätzers nach Nielsen (2007), dargestellt in Gleichung (5.12) und angewandt mit einem Daten-Taper Startwerte für die zweite Stufe gewonnen werden, in der dann der exakte lokale Whittle-Schätzer (5.14) durch-

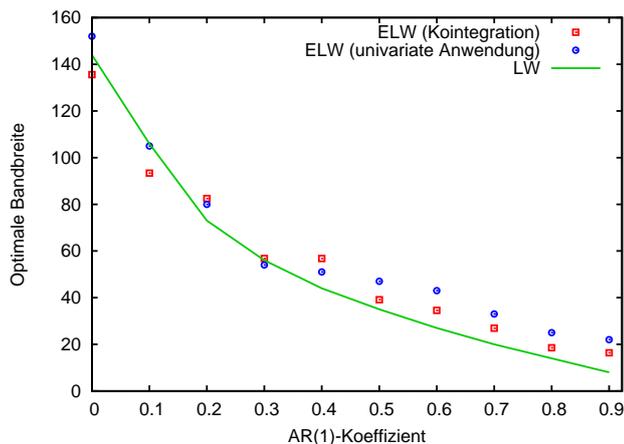


Abb. 5.5: Optimale Bandbreite für den ELW-Schätzer für Kointegration in Abhängigkeit von  $\theta$

geführt wird. Das ganze Verfahren wird dann solange durchgeführt bis eine Konvergenz eintritt. Shimotsu (2012) diskutiert noch weitere technische Details, so wie etwa das Berücksichtigen eines Strafterms in  $R_m^*(\psi)$ , um eine bessere und schnellere Konvergenz der Schätzwerte zu erreichen. Auch schlägt er vor, die ersten Bandbreiten der Periodogrammmatrix  $\text{Re}[\mathbf{I}_{\Delta^{\delta_w}}(\lambda_j; \beta)]$  wegzulassen, da diese das Verhalten des Schätzers im Falle  $b > 1/2$  negativ beeinflussen können. Auch gilt es zu beachten, dass der exakte lokale Whittle-Schätzer lediglich für bivariat kointegrierte Systeme angewendet werden kann, während der lokale Whittle-Schätzer nach Nielsen (2007) auch  $p$ -dimensionale Systeme behandeln kann. Ähnlich wie bei der univariaten Anwendung des exakten lokalen Whittle-Schätzers existiert noch keine Studie bezüglich der Wahl des Bandbreitenparameters  $m$ . Wir wollen hier die Untersuchung relativ kurz halten und konzentrieren uns im Folgenden nur auf die Möglichkeit von AR(1)-Störtermen. Wir betrachten generell wie auch in den vorangegangenen Simulationen nur die Situation  $d = 1$  und nehmen für den AR(1)-Parameter  $\theta \in [0; 0,9]$  lediglich positive Werte an. Wir simulieren zunächst ein nicht-kointegriertes System mit

$$y_t = \Delta^{-1}(1 - \theta L)\varepsilon_{1t} \quad \text{und} \quad x_t = \Delta^{-1}\varepsilon_{2t} \quad \text{für} \quad t = 1, \dots, n \quad (5.15)$$

und fixieren den Beobachtungsumfang  $n = 500$ . Für den Vektor  $\mathbf{z}_t = (y_t, x_t)'$  schätzen wir dann die Parameter  $\beta, d$  und den Integrationsgrad  $d_u = d - b$  der geschätzten Kointegrationsbeziehung. Dies wiederum wiederholen wir insgesamt 5000-mal, was wesentlich weniger ist als bei der univariaten Variante des ELW-Schätzers in Abschnitt 3.2.4. Der Grund für diese wesentlich geringere Wiederholungszahl liegt vor allem darin, dass die Anwendung des multivariaten ELW-Schätzers relativ re-

chenintensiv ist, da mehrere Schritte in der Optimierung durchgeführt werden müssen. Für den uni- und multivariaten ELW-Schätzer hat Katsumi Shimotsu auf seiner Internetseite<sup>24</sup> Programmcodes für Matlab zur Verfügung gestellt, die wir aber nicht verwenden werden, sondern eine eigenständige Umsetzung für die Statistik-Software Gretl. Interessant sind hierbei die Schätzeigenschaften für den Integrationsgrad des fälschlicherweise angenommenen stochastischen Trends  $d$  und letztendlich  $d_u$ , womit in einer empirischen Anwendung die Entscheidung über das Vorliegen einer Kointegrationsbeziehung entschieden würde. Da nun der Störterm des stochastischen Trends aber keine weiteren AR(1)-Bestandteile in sich trägt, können wir anhand des RMSE für  $\hat{d}_u$  eine optimale Bandbreite bestimmen, die wir in Abbildung 5.5 neben den optimalen Bandbreiten darstellen, die wir bereits bei der univariaten Anwendung des ELW-Schätzers in Abschnitt 3.2.4 ermittelt haben. Dort hatten wir bereits gesehen, dass die optimalen Bandbreiten des ELW-Schätzers relativ gut mit denen des LW-Schätzers zusammenpassen. Dieses Ergebnis wiederum können wir auch hier bestätigen, da sich der Verlauf der optimalen Bandbreiten sehr ähneln. Nun wollen wir noch die Schätzeigenschaften der einzelnen Integrationsparameter näher betrachten und zwar unter anderem auch für den Fall, dass wirklich eine fraktionale Kointegrationsbeziehung vorliegt. Hierzu simulieren wir mit

$$y_t = \beta x_t + \Delta^{b-1} \varepsilon_{1t} \quad \text{und} \quad x_t = \Delta^{-1} \varepsilon_{2t} \quad \text{für} \quad t = 1, \dots, n \quad (5.16)$$

mit  $\beta = 1$  ein kointegriertes System mit der Kointegrationsstärke  $b$  und fixem Integrationsgrad des stochastischen Trends  $d = 1$  ohne AR(1)-Bestandteile in den Störtermen  $\varepsilon_{1t}$  und  $\varepsilon_{2t}$ . Wir wollen aber nun annehmen, dass eine Korrelation  $\rho$  zwischen diesen existiert und nehmen damit eine multivariate Normalverteilung - genau wie wir im vorangegangenen Abschnitt in Gleichung (5.7) angegeben hatten.

In Tabelle 5.7 geben wir die Schätzeigenschaften nun für  $d$  und  $d_u$  für die jeweiligen Ausführungen des ELW-Schätzers an, wobei wir hier nun die Ergebnisse für  $n = 500$  betrachten, da sich für die anderen Stichprobenumfänge die Ergebnisse im Wesentlichen gleichen. Da wir hier keine AR(1)-Bestandteile in den Störtermen betrachten, fixieren wir die Bandbreite bei  $m = \lfloor 500^{0.8} \rfloor$ , wobei wir auch hier anmerken können, dass wir für andere Bandbreiten sehr ähnliche Ergebnisse erhalten. Wir haben hier wiederum insgesamt 5000-mal das datengenerierende System (5.16) für jede der angegebenen Parameterkonstellationen simuliert und die jeweiligen Schätzergebnisse ermittelt. Für die univariate Anwendung des ELW-Schätzers für die

<sup>24</sup>Stand 02.12.2013 unter <http://shimotsu.web.fc2.com/Site/Home.html>

Tab. 5.7: Vergleich der Schätzeigenschaften für den ELW-Schätzer in der multivariaten und univariaten Ausführung für  $n = 500$

$\rho$	$b$	Verzerrung				RMSE			
		$d$		$d_u$		$d$		$d_u$	
		Multi.	Uni.	Multi.	Uni.	Multi.	Uni.	Multi.	Uni.
0	0	0,0275	0,0274	0,0260	0,0215	0,0547	<b>0,0547</b>	0,0542	<b>0,0528</b>
	0,2	0,0271	0,0269	0,0260	0,0334	0,0537	<b>0,0535</b>	0,0588	<b>0,0534</b>
	0,4	0,0270	0,0269	0,0234	0,0318	0,0541	<b>0,0541</b>	0,0560	<b>0,0535</b>
	0,6	0,0267	0,0265	0,0236	0,0238	0,0537	<b>0,0536</b>	0,0546	<b>0,0538</b>
	0,8	0,0273	0,0271	0,0210	0,0203	0,0546	<b>0,0545</b>	0,0531	<b>0,0525</b>
	1	0,0278	0,0277	0,0200	0,0200	0,0536	<b>0,0536</b>	0,0523	<b>0,0519</b>
0,8	0	0,0271	0,0264	0,0265	0,0210	<b>0,0528</b>	0,0540	0,0522	<b>0,0521</b>
	0,2	0,0273	0,0274	0,0252	-0,0102	<b>0,0524</b>	0,0538	0,0518	<b>0,0513</b>
	0,4	0,0037	0,0281	0,0191	-0,0012	0,0591	<b>0,0547</b>	<b>0,0461</b>	0,0478
	0,6	0,0202	0,0280	0,0463	0,0101	<b>0,0522</b>	0,0542	<b>0,0412</b>	0,0484
	0,8	0,0253	0,0267	0,0246	0,0243	<b>0,0445</b>	0,0535	<b>0,0451</b>	0,0517
	1	0,0248	0,0262	0,0230	0,0332	<b>0,0440</b>	0,0538	<b>0,0429</b>	0,0549

Kointegrationsresiduen haben wir angenommen, dass wir diese nicht beobachten können und haben sie mit OLS geschätzt. Die Ergebnisse für die beiden Ausführungen gleichen sich sehr. Wir beobachten ähnliche durchschnittliche Verzerrungen und ähnliche RMSE-Werte für die einzelnen Konstellationen. Falls wir keine Korrelation der Störterme annehmen, sehen wir hingegen, dass wir bei Anwendung des multivariaten ELW-Schätzers tendenziell besser, d.h. weniger volatile Ergebnisse erzielen, da die Verzerrung generell auf dem gleichem Niveau bleibt. Dies wiederum gilt sowohl für den Integrationsgrad der Kointegrationsresiduen als auch für den des identifizierten stochastischen Trends. Ansonsten ist die univariate Variante leicht besser, aber auf sehr ähnlichem Niveau. Somit können wir zu dem Schluss kommen, dass zur Ermittlung der Integrationsgrade beide Varianten durchaus gute Ergebnisse erzielen. Bemerkenswert ist das insofern, da ja für die univariate Schätzvariante OLS-Schätzergebnisse für  $\beta$  verwendet wurden, die besonders für einen hohen Korrelationskoeffizient  $\rho = 0,8$  verzerrt sind.

In Tabelle 5.8 vergleichen wir noch die Eigenschaften der Schätzwerte für den Beziehungsparameter  $\beta$ , den wir mit dem multivariaten ELW-Schätzer und OLS ermitteln können. Hier zeigt sich, dass der ELW-Schätzer unter einer sehr hohen Streuung für den Fall der schwachen Kointegration leidet, wenn wir eine kleinere Bandbreite wählen. Der Koeffizient  $\beta$  lässt sich für  $\rho = 0$  mit OLS oder unter Verwendung einer hohen Bandbreite relativ gut schätzen, auch wenn wir für schwache Ausprägungen von  $b$  mit einer deutlich größeren Unsicherheit rechnen müssen. Nehmen wir nun eine Korrelation zwischen den Störtermen des Kointegrationsfehlers und des stochastischen Trends an, dann zeigt sich immer noch eine bessere Eigenschaft von OLS im Falle der schwachen Kointegration. Für  $b > 0,6$  zeigt sich aber,

Tab. 5.8: Vergleich der Schätzeigenschaften für den ELW-Schätzer bzgl.  $\beta$  im Vergleich zu OLS für  $n = 500$

$\rho$	$b$	Verzerrung			RMSE		
		$m = n^{0,6}$	$m = n^{0,8}$	OLS	$m = n^{0,6}$	$m = n^{0,8}$	OLS
0	0,2	0,0204	-0,0071	-0,0025	6,0620	0,7875	<b>0,2165</b>
	0,4	0,0210	-0,0018	0,0004	4,3657	0,0775	<b>0,0605</b>
	0,6	0,0211	0,0008	-0,0002	1,4069	0,0329	<b>0,0324</b>
	0,8	-0,0001	-0,0001	0,0000	0,0158	0,0151	<b>0,0140</b>
	1	0,0000	0,0000	0,0000	0,0067	0,0067	<b>0,0066</b>
0,8	0,2	3,5087	-0,0699	0,3223	10,9913	8,4997	<b>0,3513</b>
	0,4	3,2079	7,3465	0,1298	9,7064	6,5231	<b>0,1448</b>
	0,6	0,6561	1,8009	0,0521	4,1817	0,9097	<b>0,0609</b>
	0,8	-0,0007	-0,0005	0,0212	0,0123	<b>0,0107</b>	0,0266
	1	0,0000	0,0001	0,0085	0,0045	<b>0,0043</b>	0,0118

Die Schätzergebnisse für den multivariaten ELW-Schätzer sind hier einfach mit der jeweils verwendeten Bandbreite angegeben.

dass  $\beta$  mit ELW besser geschätzt werden kann und das auch mit einer niedrigeren Bandbreite. Diese Ergebnisse decken sich auch mit denen von Shimotsu (2012) und zeigen, dass der multivariate Schätzer insbesondere für Systeme mit stationären Kointegrationsstörtermen geeignet ist. Dies gilt laut Shimotsu (2012) auch für den Fall der stationären Kointegration, was wir aber aus unseren Überlegungen zunächst ausgeschlossen haben. Im Hinblick auf die stark volatilen Schätzergebnisse von  $\beta$  bei Anwendung des ELW-Schätzers scheint es bemerkenswert, dass trotzdem der Integrationsgrad für sämtliche Parameter von  $b$  unter Fixierung von  $d = 1$  relativ gut geschätzt werden kann. Shimotsu (2012) gibt aber zu bedenken, dass eine alleinige Anwendung des multivariaten ELW-Schätzers gefährlich sein kann, da die Entscheidung über das Vorliegen von Kointegrationsbeziehungen stark von der Entscheidung über die verwendete Bandbreite abhängen kann. Die Anwendung des multivariaten ELW-Schätzers sollte somit eine Betrachtung über mehrere Bandbreiten und eine Analyse der Kointegrationsfehlerterme auf eine mögliche ARFIMA-Struktur einschließen.

#### 5.4.2 Das allgemeine fraktionale Fehlerkorrekturmodell und dessen Schätzung

Das klassische kointegrierte Fehlerkorrekturmodell in der  $I(1)/I(0)$ -Situation für einen  $p$ -dimensionalen nicht-stationären Vektorprozess  $\mathbf{z}_t$  stellt sich nach Engle und Granger (1987) in Verbindung mit Johansen (1988, 1995) als folgender VAR dar:

$$\Delta \mathbf{z}_t = \underbrace{\mathbf{A}\mathbf{B}'}_{\mathbf{\Pi}} \mathbf{z}_{t-1} + \sum_{i=1}^k \mathbf{\Gamma}_i \Delta \mathbf{z}_{t-i} + \boldsymbol{\varepsilon}_t \quad \text{für } t = 1, \dots, n. \quad (5.17)$$

$\mathbf{B}$  ist eine  $(p \times r)$ -dimensionale Kointegrationsmatrix, die transponiert und multipliziert mit der ersten verzögerten von  $\mathbf{z}_t$  insgesamt  $r$  Kointegrationsbeziehungen liefert. Im Falle der bivariaten Kointegration können wir  $\mathbf{z}_t = (y_t, x_t)'$  setzen und bei Kenntnis, dass die Linearkombination  $y_t - \beta x_t$  von  $x_t$  und  $y_t$  stationär ist und somit eine Kointegrationsbeziehung vorliegt, ergibt sich die Kointegrationsmatrix als  $\mathbf{B} = (1, -\beta)'$ . Interessant ist nun an diesem Modell, dass die Veränderung von  $\mathbf{z}_t$  in  $t$  durch eine Korrektur der Abweichung vom Kointegrationsgleichgewicht in der Vorperiode  $t - 1$  erklärt werden soll. Dazu existiert die Korrekturmatrix  $\mathbf{A}$ , die die einzelnen Korrekturkoeffizienten enthält, und es werden noch insgesamt  $k$  zeitverzögerte Differenzen von  $\mathbf{z}_t$  in diesem Modell berücksichtigt, um einer möglichen autokorrelierten Struktur von  $\Delta \mathbf{z}_t$  Rechnung zu tragen. Die Multiplikation  $\mathbf{A}\mathbf{B}'$  liefert die sogenannte Langfristmatrix  $\mathbf{\Pi}$ , die die Koeffizienten der VAR-Darstellung des Systems enthält. Bereits Granger (1986) schlägt eine mögliche verallgemeinerte Darstellung des Fehlerkorrekturmodells vor, das auch die Möglichkeit der fraktionalen Kointegration beinhaltet. Johansen (2008) modifiziert dieses aber und stellt dieses Modell als

$$\Delta^d \mathbf{z}_t = \Delta^{d-b} L_b \mathbf{A}\mathbf{B}' \mathbf{z}_t + \sum_{i=1}^k \Gamma_i \Delta^d L_b^i \mathbf{z}_t + \boldsymbol{\varepsilon}_t \quad \text{für } t = 1, \dots, n \quad (5.18)$$

dar. Zentraler Ansatzpunkt ist die Verwendung des fraktionalen Lag-Operators, der sich wie folgt gestaltet:

$$\begin{aligned} L_b = 1 - \Delta^b &= L \sum_{n=0}^{\infty} \binom{b}{n+1} (-L)^n \\ &= bL - \frac{1}{2}b(b-1)L^2 + \frac{1}{6}b(b-1)(b-2)L^3 - \dots \end{aligned} \quad (5.19)$$

Für  $b = 1$  ergibt sich  $L_1 = L$  und somit ist gewährleistet, dass das klassische Fehlerkorrekturmodell einen Spezialfall des allgemeinen fraktional kointegrierten VAR-Modells (weiterhin abgekürzt als FCVAR) darstellt. Die Verwendung von  $L_b$  rechtfertigt Johansen (2008) damit, dass somit die Anwendung der Reduced-Rank-Regression-Technik, die Johansen (1988) im klassischen Fall zur Schätzung des Modells (5.17) anwendet, auch im Falle der fraktionalen (Ko-)Integration möglich wird und die Schätzung des Modells, die von Johansen und Nielsen (2012a) beschrieben und analysiert wird, erheblich vereinfacht. Hierfür wird aber die Untersuchung, ob das geschätzte VAR-System überhaupt stabil ist, erschwert, da es nun nicht mehr möglich ist, die Lösungen der charakteristischen Gleichung auf Lage

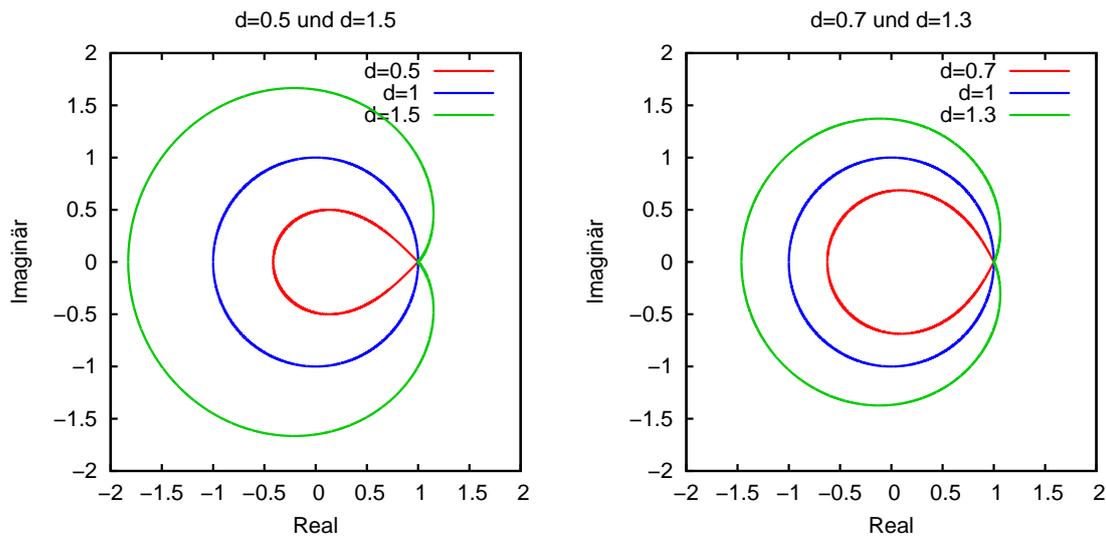


Abb. 5.6: Vergleich der Abbildungen des Einheitskreises nach  $\mathbb{C}_d$

außerhalb des Einheitskreises im Gaußschen Zahlenraum zu überprüfen. Johansen (2008) schlägt deswegen vor, die charakteristische Gleichung in Abhängigkeit des fraktionalen Lag-Operator zu lösen und dann deren Lösungen auf einer transformierten Fläche zum Einheitskreis zu vergleichen. Dies wollen wir nun anhand eines einfachen Beispiels, das auch Johansen (2008) anführte, demonstrieren. Nehmen wir an, dass wir ein univariates Modell (d.h.  $p = 1$ ) beobachten mit  $d = b$  und  $k = 1$ :

$$\Delta^d z_t = \pi \underbrace{(1 - \Delta^d)}_{L_d} z_t + \varepsilon_t \quad \text{für } t = 1, \dots, n$$

oder mit  $\rho = 1 + \pi$

$$z_t = \rho L_d z_t + \varepsilon_t = \rho(1 - \Delta^d) z_t + \varepsilon_t, \quad (5.20)$$

was für  $d = 1$  genau das Testproblem nach Dickey und Fuller (1979) ergibt. Die charakteristische Gleichung des Modells (5.20) ergibt sich als

$$\pi(z) = 1 - \rho(1 - (1 - z)^d). \quad (5.21)$$

Die Stationarität von (5.20) ist nun gegeben, wenn  $\rho^{-1} \notin \mathbb{C}_d$ , wobei  $\mathbb{C}_d$  definiert ist als das Bild des Einheitskreises und der Abbildung  $z \rightarrow 1 - (1 - z)^d$ . Für  $d = 1$  ist  $\mathbb{C}_1$  gerade der Einheitskreis und  $|\rho| < 1$  stellt damit die Stationaritätsbedingung für das Modell (5.20) dar. Für  $d \neq 1$  ist das wiederum komplizierter, da sich hier nun andere Kriterien ergeben. Die Abbildung 5.6 zeigt mit  $\mathbb{C}_d$  die Wertebereiche

für eine stationäre Lösung  $\rho$  für  $d \in \{0,5; 0,7; 1,3; 1,5\}$  unter der Annahme, dass  $\rho$  auch komplexwertig sein kann, im Vergleich zum Einheitskreis. Hier kann man erkennen, dass für  $d > 1$  der Stationaritätsbereich tendenziell kleiner wird und für  $d < 1$  größer wird im Vergleich zur klassischen Situation  $d = 1$ .

Nun wollen wir die Möglichkeit zur Schätzung des FCVAR-Modells näher betrachten, die von Johansen und Nielsen (2012a) diskutiert wurde. Hierbei ist noch anzumerken, dass die Notation zu Johansen und Nielsen (2012a) teilweise vereinfacht wird. In deren Analyse kommt der Definition von Anfangswerten eine große Bedeutung zu, die auch noch in Johansen und Nielsen (2013) tiefergehend beschrieben wird. Prinzipiell wird in den bisher diskutierten Schätzmodellen für fraktionale (Ko-)Integration stets implizit angenommen, dass Beobachtungen für  $t < 1$  null sind und erst für  $t \geq 1$  Realisierung ungleich null möglich sind. Dies ist aber im Hinblick auf eine empirische Anwendung keine realistische Annahme und Johansen und Nielsen (2012a) schlagen deswegen vor, eine Anzahl von Startwerten zu definieren, die zwar bei der Berechnung von fraktionalen Differenzen berücksichtigt werden, aber für die direkte Schätzung ausgeschlossen werden. Somit können mögliche Verzerrungen durch nicht-beobachtbare Realisierungen von  $t < 1$  vermieden werden, die nach Johansen und Nielsen (2012b) insbesondere für Modelle mit nicht-stationären Komponenten auftreten können. Dies wollen wir nun in der Vorstellung des Verfahrens zunächst nicht berücksichtigen, werden dem aber in der empirischen Anwendung im folgenden Kapitel Rechnung tragen.

Zentraler Ansatzpunkt ist wie im klassischen Fall die Annahme, dass der  $p$ -dimensionale Störtermvektor  $\boldsymbol{\varepsilon}_t$  im Modell (5.18) verteilt ist mit  $N(\mathbf{0}; \boldsymbol{\Omega})$ . Hierzu schlagen Johansen und Nielsen (2012a) eine Maximum-Likelihood-Schätzung von  $\boldsymbol{\lambda} = (d, b, \mathbf{A}, \mathbf{B}, \boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_k, \boldsymbol{\Omega})$  vor. Diese wiederum lässt sich wie im klassischen  $I(1)/I(0)$ -Fall mit Hilfe der Reduced-Rank-Regressionstechnik vereinfachen, so dass die Matrizen  $\mathbf{A}, \mathbf{B}, \boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_k$  und  $\boldsymbol{\Omega}$  implizit geschätzt werden können. In diesem Kontext lässt sich aber  $\boldsymbol{\delta}_r = (d, b)'$  nicht konzentrieren und muss somit über ein numerisches Verfahren angepasst werden. Johansen und Nielsen (2012a) definieren hierzu die Residuenvektorprozesse

$$\mathbf{R}_{0t}(\boldsymbol{\delta}_r) = (\Delta^d \mathbf{z}_t | \Delta^d L_b \mathbf{z}_t, \dots, \Delta^d L_b^k \mathbf{z}_t)$$

und

$$\mathbf{R}_{1t}(\boldsymbol{\delta}_r) = (\Delta^{d-b} L_b \mathbf{z}_t | \Delta^d L_b \mathbf{z}_t, \dots, \Delta^d L_b^k \mathbf{z}_t),$$

die aus einer Regression von  $\Delta^d \mathbf{z}_t$  und  $\Delta^{d-b} L_b \mathbf{z}_t$  auf  $\Delta^d L_b \mathbf{z}_t, \dots, \Delta^d L_b^k \mathbf{z}_t$  entstehen. Diese wiederum werden dazu verwendet, um die gemischten Produkte

$$\mathbf{S}_{ij}(\boldsymbol{\delta}_r) = \frac{1}{n} \sum_{t=1}^n \mathbf{R}_{it}(\boldsymbol{\delta}_r) \mathbf{R}_{jt}(\boldsymbol{\delta}_r)' \quad \text{für } i, j = 0, 1 \quad (5.22)$$

zu definieren. Diese werden eingesetzt in das Eigenwertproblem

$$\det(\omega \mathbf{S}_{11}(\boldsymbol{\delta}_r) - \mathbf{S}_{10}(\boldsymbol{\delta}_r) \mathbf{S}_{00}(\boldsymbol{\delta}_r)^{-1} \mathbf{S}_{01}(\boldsymbol{\delta}_r)) = 0, \quad (5.23)$$

welches in den absteigend sortierten Eigenwerten  $1 > \hat{\omega}_1(\boldsymbol{\delta}_r) > \dots > \hat{\omega}_p(\boldsymbol{\delta}_r) > 0$  resultiert. Diese wiederum ergeben eingesetzt in die Profil-Likelihood-Funktion

$$l_r(\boldsymbol{\delta}_r) = \ln \det(\mathbf{S}_{00}(\boldsymbol{\delta}_r)) + \sum_{i=1}^r \ln(1 - \hat{\omega}_i(\boldsymbol{\delta}_r)) = -2n^{-1} L_r^{\max} \quad (5.24)$$

genau das Maximum der logarithmierten Likelihood-Funktion

$$-2n^{-1} \ln L(\boldsymbol{\lambda}) = \ln \det(\boldsymbol{\Omega}) + \text{tr} \left( \boldsymbol{\Omega}^{-1} n^{-1} \sum_{t=1}^n \boldsymbol{\varepsilon}_t(\boldsymbol{\lambda}) \boldsymbol{\varepsilon}_t(\boldsymbol{\lambda})' \right) \quad (5.25)$$

mit  $\boldsymbol{\varepsilon}_t(\boldsymbol{\lambda}) = \Delta^d \mathbf{z}_t - \Delta^{d-b} L_b \mathbf{A} \mathbf{B}' \mathbf{z}_t - \sum_{i=1}^k \boldsymbol{\Gamma}_i \Delta^d L_b^i \mathbf{z}_t$ . Das Maximum-Likelihood-Schätzergebnis  $\hat{\boldsymbol{\delta}}_r$  errechnet sich als Minimum von  $l_r(\boldsymbol{\delta}_r)$  und muss mit Hilfe eines numerischen Verfahrens ermittelt werden, was durchaus rechenintensiv sein kann. Es gilt hier auch zu bedenken, dass die Schätzergebnisse  $\hat{\boldsymbol{\delta}}_r$  von der Anzahl der angenommenen Kointegrationsbeziehungen  $r$  abhängen und bei Unsicherheit über diese für jede einzelne Situation  $r = 0, 1, \dots, p$  Schätzergebnisse ermittelt werden müssen. Johansen und Nielsen (2012a) schlagen zudem die Möglichkeit der Restriktion  $d = b$  in der Schätzung vor, die das ganze Schätzproblem vereinfacht und die Schätzergebnisse robuster macht, sofern die Restriktion zutrifft. Die Motivation für die Betrachtung dieser Möglichkeit liegt in erster Linie darin, dass der stochastische Trend im kointegrierten System durchaus fraktional integriert sein kann, aber der Kointegrationsfehler ist lediglich integriert vom Grad 0. Die zu schätzende Modellgleichung lautet dann

$$\Delta^d \mathbf{z}_t = \mathbf{A} L_d (\mathbf{B}' \mathbf{z}_t + \boldsymbol{\rho}') + \sum_{i=1}^k \boldsymbol{\Gamma}_i \Delta^d L_d^i \mathbf{z}_t + \boldsymbol{\varepsilon}_t \quad \text{für } t = 1, \dots, n \quad (5.26)$$

mit  $\mathbf{A} \boldsymbol{\rho}' = \mathbf{0}$ . Diese Restriktion stellt sicher, dass die  $r$  Komponenten des Vektors

$\rho'$  den jeweiligen Mittelwerten der  $r$  Kointegrationsbeziehungen entsprechen, aber keine Driftparameter für  $\Delta^d \mathbf{z}_t$ . In Modell (5.18) muss eine Konstante nicht notwendigerweise berücksichtigt werden, auch wenn die Kointegrationsresiduen einen Mittelwert ungleich 0 aufweisen, da  $\Delta^a(z_t + \mu) = \Delta^a z_t$  gilt für  $a > 0$ , weil  $\Delta^a \mu = 0$  ist. Somit muss für den Fall  $d > b$  keine Konstante bei der Schätzung berücksichtigt werden. Das Schätzverfahren für die Situation  $d = b$  ändert sich aber nur geringfügig im Vergleich zu der bereits beschriebenen Prozedur für den Fall der restringierten Konstanten: Es muss lediglich bei der Berechnung des Residuenvektorenprozesses  $\mathbf{R}_{1t}$  neben den Regressoren  $\Delta^d L_d^i \mathbf{z}_t$  noch ein Absolutglied  $\boldsymbol{\iota} = (1, \dots, 1)'$  aufgenommen werden.

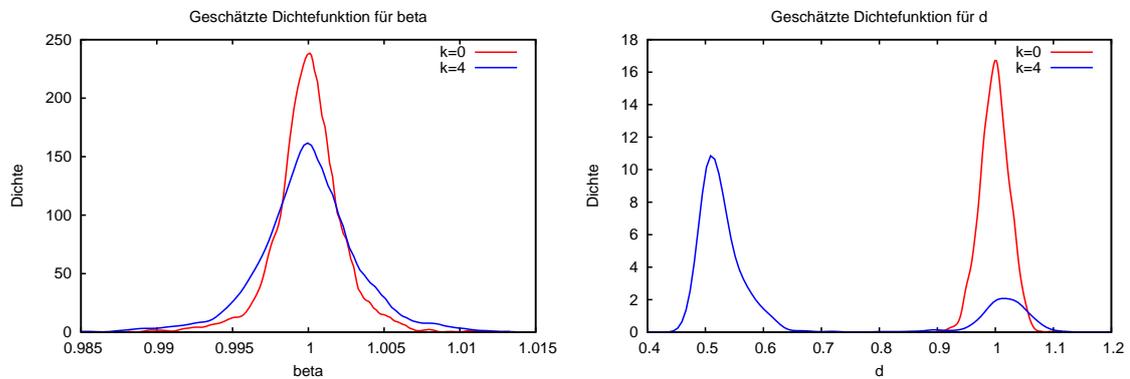
Um nun mit dem Schätzverhalten für das FCVAR-Modell besser vertraut zu werden, simulieren wir ein einfaches bivariates klassisch kointegriertes (d.h.  $d = b = 1$ ) Modell

$$y_t = \beta x_t + \varepsilon_{1t} \quad \text{und} \quad y_t = \Delta^{-1} \varepsilon_{2t} \quad \text{für} \quad t = 1, \dots, n \quad (5.27)$$

und fixieren hierfür  $n = 1000$  und  $\beta = 1$ .  $\varepsilon_{1t}$  und  $\varepsilon_{2t}$  sollen unabhängig und standardnormalverteilt sein. Für dieses Modell schätzen wir nun das FCVAR-Modell (5.18) mit insgesamt 5000 Wiederholungen. Schätzprozeduren für die Software Matlab sind auf der persönlichen Internetseite<sup>25</sup> von Morten Ørregard Nielsen abrufbar. Dazu existiert in Nielsen und Morin (2014) eine ausführliche Beschreibung zur Verwendung der Programme. Wir wollen hier diese Codes aber nicht verwenden, da diese bei fehlenden Konvergenzverhalten des Optimierungsalgorithmus keine guten Eigenschaften aufweisen. Hierzu wurden die Codes für die Software gretl umgeschrieben und mit den beispielhaften Anwendungen von Nielsen und Morin (2014) verglichen. Für den unrestringierten Fall des Modells (5.18) kommt es in ungefähr 10% der Fälle zu keiner Konvergenz und es kann kein Schätzergebnis  $\hat{\boldsymbol{\delta}}_r$  gefunden werden, während dies für die Restriktion  $d = b$  in Modell (5.26) nie auftritt. Weiterhin werden wir die gewählte Lagzahl bei  $k = 0$  und  $k = 4$  fixieren, um zu sehen, wie die Schätzergebnisse sich verhalten, wenn wir ein tendenziell überspezifiziertes Modell schätzen.

In der Abbildung 5.7 stellen wir nun die geschätzten Dichtefunktionen für die Schätzergebnisse  $\hat{\beta}$ , welche wir in diesem einfachen Fall aus der zweiten Komponente der Kointegrationsmatrix  $\hat{\mathbf{B}} = (1, \hat{\beta})'$  ablesen können, und für  $\hat{d}$ . Hierbei können wir anhand der Dichtefunktion des geschätzten Integrationsparameters feststellen, dass

<sup>25</sup>Stand 09.07.2014 unter <http://www.econ.queensu.ca/faculty/mon/research/>

Abb. 5.7: Geschätzte Dichtefunktionen für  $\beta$  und  $d$ Tab. 5.9: Durchschnittliche Verzerrung und RMSE für  $\hat{\beta}$  bei Schätzung des FCVAR und des klassischen CVAR

$n$	$\theta$	$b$	Verzerrung				RMSE			
			FCVAR		CVAR		FCVAR		CVAR	
			$k=0$	$k=4$	$k=0$	$k=4$	$k=0$	$k=4$	$k=0$	$k=4$
250	0	0,2	-0,0142	-0,2782	-0,5972	1,0649	7,1915	13,668	39,474	21,043
		0,4	-0,0069	-0,0318	0,0008	-0,2671	0,0962	0,4028	0,2031	3,9162
		0,6	-0,0019	-0,0062	0,0006	0,0000	0,0430	0,0818	0,0545	0,0890
		0,8	-0,0008	-0,0002	0,0003	0,0002	0,0213	0,0311	0,0215	0,0232
		1	0,0004	-0,0002	0,0000	0,0000	0,0098	0,0139	0,0096	0,0102
500	0	0,2	-0,0379	-0,0155	0,1024	-0,0662	0,4095	8,4480	4,3162	11,784
		0,4	-0,0056	-0,0237	-0,0007	-0,0003	0,0647	0,3104	0,1161	0,5190
		0,6	0,0001	-0,0024	0,0000	-0,0001	0,0273	0,0465	0,0353	0,0376
		0,8	-0,0004	-0,0001	0,0000	0,0000	0,0114	0,0162	0,0123	0,0127
		1	0,0000	-0,0001	0,0000	0,0000	0,0047	0,0066	0,0048	0,0049
$n$	$\theta$	$b$	Verzerrung				RMSE			
			FCVAR		CVAR		FCVAR		CVAR	
			$k=1$	$k=4$	$k=1$	$k=4$	$k=1$	$k=4$	$k=1$	$k=4$
250	0,5	0,2	-0,2539	-1,0703	0,0131	-0,0337	16,007	41,397	27,978	35,515
		0,4	-0,0845	0,2153	0,0614	0,0501	3,2774	13,708	6,6956	3,4372
		0,6	-0,0167	-0,0204	0,0003	0,0016	0,2678	0,2076	0,1342	0,2249
		0,8	-0,0023	-0,0055	0,0000	-0,0001	0,0623	0,0722	0,0455	0,0475
		1	0,0004	-0,0013	-0,0003	-0,0003	0,0255	0,0362	0,0198	0,0204
500	0,5	0,2	0,0270	0,5436	-0,2211	-0,2290	3,4337	23,813	17,054	25,240
		0,4	-0,0534	-0,0503	-0,0472	-0,0033	0,5991	0,3453	4,6266	1,4758
		0,6	-0,0067	-0,0085	0,0011	0,0012	0,0881	0,0880	0,0755	0,0795
		0,8	-0,0022	-0,0018	0,0003	0,0003	0,0322	0,0336	0,0247	0,0252
		1	-0,0002	0,0001	0,0001	0,0001	0,0128	0,0131	0,0097	0,0098

CVAR stellt hier die Anwendung des klassischen Kointegrationsmodells (5.17) nach Johansen (1998,1995) dar, während FCVAR die Schätzergebnisse für das unrestringierte Modell aus Gleichung (5.18) umschreibt.

trotz einer Konvergenz der Prozedur häufiger Werte von  $d$  nahe der Stationaritätsgrenze für  $k = 4$  ausgegeben werden als in der Nähe des wahren Werts. Falls wir  $k = 0$  schätzen, erhalten wir hingegen recht präzise Ergebnisse. Somit bekommen wir schon eine erste Indikation, dass der Wahl des Lagparameters  $k$  eine besondere Bedeutung zukommt und eine falsche Wahl verzerrte Ergebnisse erwarten lassen. In Abbildung 5.7 haben wir auch noch die Dichtefunktion für den geschätzten Beziehungsparameter  $\beta$ , wo wir sehen können, dass eine Überspezifizierung von  $k$  zwar die Streuung der geschätzten Parameter beeinflusst, aber scheinbar nicht seinen Erwartungswert.

In den Tabellen 5.9 und 5.10 können wir die Schätzeigenschaften bzgl.  $\beta$ ,  $d$  und  $b$  für variable Werte des Stichprobenumfangs  $n$  und der Kointegrationsstärke  $b$  ablesen. Auch nehmen wir hier an, dass der Kointegrationsfehler eine autokorrelierte Struktur aufweisen kann in der Form

$$y_t = \beta x_t + \Delta^{b-1}(1 - \theta L)\varepsilon_{1t} \quad \text{und} \quad x_t = \Delta^{-1}\varepsilon_{2t} \quad \text{für} \quad t = 1, \dots, n. \quad (5.28)$$

Auf eine Berücksichtigung einer Korrelation zwischen  $\varepsilon_{1t}$  und  $\varepsilon_{2t}$  - wie für die Simulationen zum multivariaten ELW-Schätzer - wollen wir aber hingegen verzichten. Den Integrationsgrad des stochastischen Trends  $y_t$  fixieren wir weiterhin durchwegs mit  $d = 1$  und den Beziehungsparameter  $\beta = 1$ . Insbesondere beim Vorliegen kleinerer Kointegrationsstärkeparameter  $b$  sind wir bezüglich der Schätzung von  $\beta$  mit sehr volatilen Schätzergebnissen konfrontiert, die aber bei Anwendung des FCVAR-Modells kleiner ausfallen als für das klassische CVAR-Modell. Eine Zunahme der Kointegrationsstärke geht tendenziell mit einer geringeren Volatilität einher und die Verzerrung nimmt auch deutlich ab. Falls wir aber aufgrund von vorliegender Autokorrelation oder sonstiger theoretischer Überlegungen nicht  $k = 0$  wählen können, dann wird eine Schätzung mit dem CVAR-Modell überlegenswert, da wir für  $k \neq 0$  häufig bessere Ergebnisse erwarten können als für das FCVAR-Modell, ausgenommen der Situation  $b = 0,2$ . Hier haben wir aber generell mit sehr volatilen geschätzten Werten zu rechnen, auch wenn wir alternative Schätzverfahren verwenden wie den multivariaten ELW-Schätzer, den wir im vorangegangenen Abschnitt kennengelernt hatten. Nichtsdestotrotz kann man auch bei Vorliegen von fraktionalem Kointegrationsbeziehungen mit guten Schätzergebnissen bzgl.  $\beta$  bei Anwendung des CVAR-Modells rechnen. Betrachten wir nun die Schätzeigenschaften bezüglich der Integrationsparameter  $d$  und  $b$ , die in Tabelle 5.10 angegeben sind: Hier wiederum können wir erkennen, dass insbesondere der Integrationsgrad  $d$  im Fall ohne Autokorrelation und richtiger Wahl des Lagparameters  $k = 0$  sehr gut geschätzt

Tab. 5.10: Durchschnittliche Verzerrung und RMSE für  $\hat{d}$  und  $\hat{b}$  bei Schätzung des FCVAR-Modells

$n$	$\theta$	$b$	Verzerrung				RMSE			
			$k = 0$		$k = 4$		$k = 0$		$k = 4$	
			$\hat{d}$	$\hat{b}$	$\hat{d}$	$\hat{b}$	$\hat{d}$	$\hat{b}$	$\hat{d}$	$\hat{b}$
250	0	0,2	-0,0162	0,2063	-0,1133	0,3642	0,0656	0,3744	0,2681	0,4829
		0,4	-0,0057	0,0561	-0,1013	0,1471	0,0535	0,2049	0,2414	0,3020
		0,6	-0,0052	0,0441	-0,0934	0,0363	0,0521	0,1488	0,2181	0,2191
		0,8	-0,0055	0,0358	-0,1477	-0,0790	0,0518	0,1169	0,2671	0,2250
		1	-0,0063	-0,0307	-0,2845	-0,3219	0,0532	0,0755	0,3658	0,3978
500	0	0,2	-0,0093	0,1024	-0,0993	0,3171	0,0449	0,2585	0,2304	0,4303
		0,4	-0,0046	0,0286	-0,0618	0,1040	0,0373	0,1354	0,1832	0,2359
		0,6	-0,0044	0,0256	-0,0674	0,0092	0,0370	0,0993	0,1772	0,1679
		0,8	-0,0032	0,0230	-0,1379	-0,0858	0,0367	0,0832	0,2493	0,2060
		1	-0,0023	-0,0181	-0,3164	-0,3483	0,0372	0,0496	0,3891	0,4158
$n$	$\theta$	$b$	Verzerrung				RMSE			
			$k = 1$		$k = 4$		$k = 1$		$k = 4$	
			$\hat{d}$	$\hat{b}$	$\hat{d}$	$\hat{b}$	$\hat{d}$	$\hat{b}$	$\hat{d}$	$\hat{b}$
250	0,5	0,2	0,0160	0,4019	0,0081	0,4142	0,1712	0,5744	0,1681	0,5795
		0,4	-0,0548	0,1397	-0,0522	0,1602	0,1974	0,3741	0,1945	0,3849
		0,6	-0,1104	0,0335	-0,1126	0,0323	0,2261	0,2721	0,2305	0,2723
		0,8	-0,1703	-0,1079	-0,1774	-0,0900	0,2816	0,2643	0,2907	0,2552
		1	-0,2271	-0,2715	-0,2458	-0,2869	0,3322	0,3656	0,3451	0,3757
500	0,5	0,2	0,0484	0,3447	0,0404	0,3794	0,1387	0,5529	0,1302	0,5704
		0,4	-0,0084	0,0868	-0,0093	0,1027	0,1218	0,3477	0,1197	0,3510
		0,6	-0,0792	0,0200	-0,0869	0,0211	0,1780	0,2277	0,1828	0,2291
		0,8	-0,1497	-0,0940	-0,1610	-0,0983	0,2563	0,2436	0,2658	0,2441
		1	-0,2743	-0,3105	-0,2764	-0,3104	0,3635	0,3933	0,3648	0,3927

werden kann. Die Verzerrung und der RMSE sind hier sehr gering und sogar auf einem niedrigerem Niveau als für den ELW-Schätzer in dessen uni- und multivariaten Ausführung bei Wahl einer hohen Bandbreite. Die Schätzung der Kointegrationsstärke  $b$  fällt hingegen wesentlich volatiler aus und für die Situation  $b = 0,2$  ist sogar generell mit einer Überschätzung der Stärke und somit eine Unterschätzung des Integrationsgrads des Kointegrationsfehlers zu rechnen. Betrachten wir nun die Ergebnisse für  $k = 4$ : Hier verschlechtern sich die Schätzeigenschaften sehr deutlich, was wir auch aus den Erkenntnissen aus Abbildung 5.7 erwarten konnten. Der Integrationsgrad  $d$  wird insbesondere für  $b = 1$  stark unterschätzt, da die Schätzung des FCVAR-Modells in dieser Situation sehr häufig Schätzergebnisse nahe 0,5 liefert. Für kleinere Werte von  $b$  zeigt sich dieses Problem überraschenderweise weniger deutlich, ist aber durchaus noch vorhanden. Besonders die Zunahme der Streuung der Schätzergebnisse für  $k = 4$  sollte beachtet werden, die in den durchgeführten Simulationen sehr groß ausfällt und auch für größere Stichproben nicht wesentlich sinkt. Für den Fall  $b = 1$  steigt sie sogar noch weiter an. Wenn wir nun weiter AR(1)-Terme in den Kointegrationsfehler hinzunehmen, wird die Schätzung der Integrationsparameter wiederum deutlich volatiler und es ergeben sich auch größere Verzerrungen, da wir

nun für die zu schätzende Modellgleichung zumindest  $k = 1$  wählen müssen. Die geschätzte Kointegrationsstärke scheint hierbei besonders verzerrt zu sein und über alle gewählten Stärkeparameter gegen ungefähr 0,6 zu gehen, was auch für den Fall ohne Autokorrelation zuzutreffen scheint. Die Streuung der Schätzergebnisse ist für den Fall  $\theta = 0,5$  durchwegs sehr groß, lässt sich aber auch durch die Wahl des Lagparameters erklären, der die Schätzung des FCVAR sehr volatil macht. Prinzipiell lässt sich somit aussagen, dass die Schätzung mit dem FCVAR-Modell für  $k \neq 0$  vorsichtig zu interpretieren ist und weitere Test- und Schätzverfahren hinzugezogen werden sollten, um eine sichere Analyse zu treffen.

Analog zu der klassischen Kointegrationsanalyse bietet das FCVAR-Modell auch die Möglichkeit, den Kointegrationsrang und damit die Anzahl der Kointegrationsbeziehungen zu testen. Johansen und Nielsen (2012a) schlagen dabei vor, die Hypothese  $H_r : \text{rg}(\mathbf{\Pi}) \leq r$  gegen  $H_p : \text{rg}(\mathbf{\Pi}) \leq p$  mit Hilfe von Likelihood-Ratio-Tests (weiterhin abgekürzt als LR-Test) zu überprüfen. Für diese Hypothesen existieren wiederum die Maximum-Likelihood-Schätzergebnisse  $\hat{\boldsymbol{\delta}}_r$  und  $\hat{\boldsymbol{\delta}}_p$ , die jeweils in die Profil-Likelihood-Funktion  $l_r(\hat{\boldsymbol{\delta}}_r)$  aus Gleichung (5.24) eingesetzt werden. Hierzu lässt sich die LR-Statistik formulieren mit

$$\begin{aligned} -2 \ln LR(H_r|H_p) &= n \ln \frac{\det[\mathbf{S}_{00}(\hat{\boldsymbol{\delta}}_r)] \prod_{i=1}^r (1 - \hat{\omega}_i(\hat{\boldsymbol{\delta}}_r))}{\det[\mathbf{S}_{00}(\hat{\boldsymbol{\delta}}_p)] \prod_{i=1}^p (1 - \hat{\omega}_i(\hat{\boldsymbol{\delta}}_p))} \\ &= n(l_r(\hat{\boldsymbol{\delta}}_r) - l_r(\hat{\boldsymbol{\delta}}_p)). \end{aligned} \quad (5.29)$$

Die Statistik vereinfacht sich hier nicht wie im klassischen Fall zu einer einfachen Trace-Statistik für welche die ermittelten logarithmierten und transformierten Eigenwerte aufsummiert werden, da hier zum einen die Matrix  $\mathbf{S}_{00}(\hat{\boldsymbol{\delta}}_r)$  als auch die geschätzten Eigenwerte  $\hat{\omega}_i(\hat{\boldsymbol{\delta}}_r)$  von den jeweiligen Schätzergebnissen  $\hat{\boldsymbol{\delta}}_r$  abhängen, die über die einzelnen angenommenen Kointegrationsränge variiert. Für die LR-Statistik konnten Johansen und Nielsen (2012a) in Abhängigkeit des Parameters  $b$  die asymptotische Verteilung angeben mit

$$-2 \ln LR(H_r|H_p) \rightarrow \chi_{(p-r)^2}^2 \quad \text{für } b < 1/2 \quad (5.30)$$

Tab. 5.11: Durchschnittliche Ablehnungsraten für  $r = 0$  bei Anwendung des LR-Tests für (F)CVAR-Modelle

$n$	$b$	$\theta = 0$				$\theta = 0,5$			
		FCVAR		CVAR		FCVAR		CVAR	
		$k = 0$	$k = 4$	$k = 0$	$k = 4$	$k = 1$	$k = 4$	$k = 1$	$k = 4$
100	0	17,6%	36,1%	5,6%	7,1%	19,7%	43,5%	6,9%	9,1%
	0,2	38,1%	33,0%	20,0%	8,7%	23,7%	40,9%	6,7%	7,9%
	0,4	84,9%	30,3%	72,8%	16,3%	34,8%	37,3%	17,5%	12,3%
	0,6	99,5%	30,8%	99,9%	40,7%	41,9%	30,2%	46,6%	28,7%
	0,8	100%	19,2%	100%	77,0%	38,0%	21,6%	89,0%	54,6%
	1	100%	12,4%	100%	97,0%	38,9%	14,2%	99,9%	80,0%
250	0	18,0%	30,1%	5,4%	6,6%	16,5%	29,7%	5,5%	5,9%
	0,2	61,8%	33,9%	35,3%	10,7%	28,2%	35,7%	10,7%	10,0%
	0,4	99,9%	31,4%	95,2%	41,1%	55,8%	36,8%	44,2%	28,2%
	0,6	100%	26,9%	100%	90,0%	68,3%	27,4%	95,6%	73,7%
	0,8	100%	16,9%	100%	100%	73,0%	15,4%	100%	99,4%
	1	100%	11,4%	100%	100%	87,3%	10,5%	100%	100%
500	0	18,1%	31,9%	4,8%	5,3%	16,5%	22,8%	5,9%	5,9%
	0,2	82,2%	38,7%	49,0%	17,4%	35,1%	27,7%	16,2%	13,2%
	0,4	100%	36,5%	99,3%	72,6%	84,3%	34,7%	77,9%	57,1%
	0,6	100%	32,2%	100%	99,7%	90,1%	30,7%	100%	98,1%
	0,8	100%	20,6%	100%	100%	96,4%	19,9%	100%	100%
	1	100%	13,2%	100%	100%	99,9%	12,0%	100%	100%

und für  $b > 1/2$  mit

$$-2 \ln LR(H_r | H_p) \rightarrow \text{tr} \left( \int_0^1 (d\mathbf{B}) \mathbf{B}'_{b-1} \left( \int_0^1 \mathbf{B}_{b-1} \mathbf{B}'_{b-1} du \right)^{-1} \int_0^1 \mathbf{B}_{b-1} (d\mathbf{B})' \right), \quad (5.31)$$

wobei hier  $\mathbf{B}(u)$  eine  $(p - r)$ -dimensionale Brownsche Bewegung darstellt und  $\mathbf{B}_{b-1}(u)$  eine fraktionale Brownsche Bewegung integriert vom Grad  $b - 1$ . Für den Fall  $b > 1/2$  müssen deswegen wiederum Testverteilungen simuliert werden, was für den Fall der restringierten Konstanten und ohne Konstante MacKinnon und Nielsen (2014) bereits gemacht haben. Sie verwenden hierzu nicht wie Doornik (1998) im klassischen Fall Gamma-Verteilungen, um die LR-Tests zu approximieren, sondern wählen eine numerische Approximation, um sogar p-Werte berechnen zu können, ohne eine bestimmte Verteilung anzunehmen. Interessant wäre noch eine Berücksichtigung von deterministischen Komponenten in dem zu untersuchenden System  $\mathbf{z}_t$ , was auch in einer demnächst erscheinenden Arbeit von Johansen und Nielsen vorbereitet wird.

Wir wollen nun diesen Teil damit abschließen, dass wir die Macht des von Johansen und Nielsen (2012a) vorgeschlagenen LR-Tests vergleichen wollen mit dem klassischen Rang-Test nach Johansen (1988), der im Endeffekt dem FCVAR-Modell für

die Restriktion  $d = b = 1$  entspricht. Hierzu simulieren wir auch wieder datengenerierende Systeme nach Gleichung (5.28), wie wir sie auch zur Diskussion der Schätzergebnisse bzgl.  $\beta$ ,  $d$  und  $b$  verwendet haben. Als kritische Werte verwenden wir diejenigen nach MacKinnon und Nielsen (2014) für  $b > 1/2$ , wobei wir aber die Kointegrationsstärke  $b$  als bekannt annehmen und die Irrtumswahrscheinlichkeit  $\alpha = 5\%$  fixieren. Für die Fälle  $b < 1/2$  verwenden wir das 95%-Quantil der  $\chi_4^2$ -Verteilung, da wir in erster Linie die Hypothese  $H_0 : r = 0$  testen wollen bei einem bivariaten System von Zeitreihen. Die jeweiligen Ablehnungsraten für das Vorliegen keiner Kointegration (d.h.  $r = 0$ ) tragen wir in der Tabelle 5.11 ab und können auch hier wieder zu ähnlichen Ergebnissen kommen wie bei der Ermittlung der Schätzeigenschaften des FCVAR-Modells. Die Anwendung des Ansatzes ist prinzipiell problematisch, falls wir nicht  $k = 0$  wählen können und wir deswegen mit sehr volatilen Schätzergebnissen konfrontiert werden. Insbesondere die Entdeckungshäufigkeiten für  $b = 1$  fallen sehr schwach aus und sind sogar wesentlich kleiner als für die restlichen Ausprägungen von  $b$ . Auch wird die erwartete Ablehnungsrate von 5% für  $b = 0$  bei weitem nicht erreicht und es scheint sogar so, dass man sich für steigende Stichprobenumfänge von dieser Zielrate noch weiter entfernt. Überraschend ist hingegen das relativ gute Abschneiden des klassischen CVAR-Modells. Dieses kann auch bei Vorliegen von fraktional integrierten Störtermen relativ gut Kointegrationsbeziehungen aufdecken, solange die Fehlerterme stationär sind. Man sollte hierbei aber auch bedenken, dass wir hier generell Systeme mit  $d = 1$  als Integrationsgrad des stochastischen Trends simulieren und dass damit zumindest die implizite Annahme des CVAR-Modells  $d = b = 1$  nicht schwerwiegender verletzt ist. Würden wir Simulation mit  $d \neq 1$  durchführen, dann würde das CVAR-Modell wahrscheinlich wesentlich schlechter ausfallen als im vorliegenden Fall. Nichtsdestotrotz können die Teststatistiken des CVAR-Modells durchaus zur Überprüfung von fraktionalen stationären Kointegrationsbeziehungen herangezogen werden, solange wir  $d = 1$  nicht ablehnen können.

## 6 Fraktionale Kointegration und die Markterwartungshypothese

Im nun folgenden Abschnitt wollen wir uns mit der empirischen Anwendung der in dieser Arbeit vorgestellten und analysierten Methoden beschäftigen. Einen Überblick über die empirischen Anwendungsmöglichkeiten in den Wirtschaftswissenschaften liefert z.B. die Arbeit von Henry und Zaffaroni (2003). Für eine univariate Modellierung bieten sich insbesondere aggregierte Zeitreihen wie das Bruttoinlandsprodukt und Preisindizes, wie auch schon Granger (1981) vorgeschlagen hatte, an. Hierzu gibt es z.B. die Betrachtungen von Hassler und Wolters (1995), die mit Hilfe des GPH-Schätzers feststellten, dass internationale Inflationszeitreihen einen Integrationsgrad  $0 < d < 1$  aufweisen. Ding, Granger und Engle (1993) untersuchen, ob Aktienrenditen eine Long-Memory-Eigenschaft haben, kommen aber zu dem Ergebnis, dass allenfalls Aktienkursvolatilitäten diese aufweisen. Darauf aufbauend wurden eine Reihe von fraktional integrierte GARCH und stochastische Volatilitätsmodelle entwickelt, um dem Rechnung zu tragen, wie z.B. Breidt, Crato und de Lima (1998). Cheung und Lai (1993) liefern eine der ersten Studien, die sich mit der empirischen Umsetzung der fraktionalen Kointegration beschäftigen. Ihr Untersuchungsgegenstand ist dabei Kaufkraftparitätenhypothese, die größtenteils aufgrund der flexiblen Modellierungsmöglichkeit der fraktionalen Kointegration angenommen werden kann. Als methodischen Unterbau verwenden sie den GPH-Schätzer und stellen fest, dass der Kointegrationsfehlerterm für die meisten Länder einen Integrationsgrad von ungefähr 0,5 aufweist, was Cheung und Lai (1993) auch signifikant unterschiedlich von null und eins testen können. Baillie und Bollerslev (1994) untersuchen auch die Kointegration von Wechselkursen des US-Dollars zu den sieben meistgehandelten Devisen der damaligen Zeit und konnten mit klassischen Kointegrationsmethoden eine Kointegrationsbeziehung zwischen diesen Wechselkursen annehmen. Für den Fehlerterm aus dieser Beziehung konnten Baillie und Bollerslev (1994) eine starke autokorrelierte Struktur feststellen, die sie mit einem nicht-stationären Long-Memory-Prozess erklären konnten, der einen signifikant kleineren Integrationsgrad als eins aufweist. Marinucci und Robinson (2001) untersuchen in ihrer Arbeit das Konsum/Einkommensverhältnis, das tendenziell sogar stark kointegriert ist und teilweise mit klassischen  $I(1)/I(0)$ -Kointegrationsmethoden untersucht werden kann, sowie das Verhältnis zwischen Aktienkursen und Dividendenausschüttungen, wo die Datenlage eher für eine schwache Kointegrationsbeziehung spricht. Darüber hinaus existieren noch weitere Kointegrationsstudien, die hier nicht weiter aufgelistet

werden sollen. Die erwähnten Arbeiten zeigen aber prinzipiell, dass in der Empirie durchaus Evidenz für fraktionale Kointegration existiert, da Kointegrationsfehlerterme nicht notwendigerweise integriert vom Grad null sein müssen bzw. sogar nicht-stationär sein können. In diesem Hinblick wollen wir uns mit der Markterwartungshypothese beschäftigen, die die Existenz von Kointegrationsbeziehungen zwischen Verzinsungen unterschiedlicher Laufzeiten impliziert.

## 6.1 Die Markterwartungshypothese und Kointegration

Hall, Anderson und Granger (weiterhin: HAG, 1992) stellen die Gleichungen von Hicks und Fisher als den zentralen Ansatzpunkt der Markterwartungshypothese dar:

$$R_t(k) = \frac{1}{k} \sum_{j=0}^{k-1} E_t[R_{t+j}(1)] + \Phi(k) \quad \text{für } k = 1, 2, \dots, p, \quad (6.1)$$

wobei  $R_t(k)$  der Zinssatz zum Zeitpunkt  $t$  mit einer Laufzeit  $k$  darstellt. Der Zinssatz  $R_t(1)$  mit Fälligkeit in einer Periode soll für den Zinssatz mit der kürzesten Fälligkeit stehen. Längerfristige Zinssätze können nach (6.1) als Durchschnitt der erwarteten Rendite von zukünftigen kurzfristigen Zinspapieren angegeben werden. Der Erwartungswertoperator  $E_t$  ergibt sich dabei aus den verfügbaren Informationen zum Zeitpunkt  $t$  und  $\Phi(k)$  stellt eine Risikoprämie dar, die die Investoren aufgrund der Unsicherheit der zukünftigen Zinsentwicklung erwarten. Allgemein wird angenommen, dass diese Prämie nicht vom Zeitpunkt  $t$  abhängen soll, sondern nur von der Dauer der Laufzeit  $k$ . HAG (1992) nehmen zwar eine zeitlich abhängige Risikoprämie an, machen aber die Einschränkung, dass diese stationär sein soll. Aufgrund der Ergebnisse von zahlreichen empirischen Studien können wir annehmen, dass Zinssätze i.d.R. nicht-stationäre Zeitreihen sind. Dies wiederum impliziert, dass zwischen den Zinssätzen eine Reihe von Kointegrationsbeziehungen bestehen müssen, falls die Markterwartungshypothese angenommen wird. HAG (1992) zeigen dies, indem sie die Gleichung (6.1) umformen zu:

$$R_t(k) - R_t(1) = \frac{1}{k} \sum_{i=1}^{k-1} \sum_{j=1}^i E_t [R_{t+j}(1) - R_{t+j-1}(1)] + \frac{1}{k} \sum_{j=1}^k \Phi(j). \quad (6.2)$$

Falls nun der Zinssatz  $R_t(1)$  eine  $I(1)$ -Zeitreihe darstellt, dann ist die rechte Seite der Gleichung (6.2) stationär, da dies den erwarteten Zinsveränderungen entspricht zuzüglich dem Mittelwert konstanter Risikoprämien. Daraus folgt, dass die

Zinsdifferenz  $R_t(k) - R_t(1)$  ebenfalls stationär sein sollte und damit die beiden Zinssätze  $R_t(k)$  und  $R_t(1)$  kointegriert sind mit dem Kointegrationsvektor  $(1, -1)'$ . Dies gilt demnach für alle am Markt gehandelten Zinspapiere unter der Gültigkeit der Markterwartungshypothese. Nehmen wir nun an, dass am Markt Zinspapiere mit  $p$  Anzahl verschiedener Laufzeiten  $1, \dots, k_p$  existieren. Dann sollten  $p - 1$  Anzahl Kointegrationsbeziehungen bestehen für den  $p$ -dimensionalen Vektor  $\mathbf{y}_t = [R_t(k_2), \dots, R_t(k_p), R_t(1)]'$ . Wir erwarten dann folgende  $p \times (p - 1)$ -Kointegrationsmatrix für  $\mathbf{y}_t$ :

$$\boldsymbol{\beta}_0 = \begin{pmatrix} \mathbf{I}_{p-1} \\ -\mathbf{1}' \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ -1 & -1 & \cdots & -1 \end{pmatrix}. \quad (6.3)$$

Die  $p - 1$  Kointegrationsbeziehungen bestehen demnach aus Differenz der einzelnen beobachteten Zinssätze zum kurzfristigsten Papier. Somit müssen Systeme mit Zinsen unterschiedlicher Laufzeit kointegriert sein mit  $r = p - 1$ . Es lässt sich weiterhin zeigen, dass sich die Koeffizienten der einzelnen Kointegrationsgleichung zu null addieren müssen.

Diese Implikationen bieten die Basis für die Überprüfung der Markterwartungshypothese in der Empirie. Die meisten Studien kommen dabei zu dem Schluss, dass die Markterwartungshypothese abzulehnen ist. Campbell und Shiller (1987) untersuchen z.B. ein bivariates System aus einmonatigen und 20-jährigen amerikanischen Staatsanleihen und kamen zu dem Ergebnis, dass diese kointegriert sind und der Kointegrationsvektor sehr nahe an dem erwarteten  $(1, -1)'$  liegt. Dennoch lehnen sie die Markterwartungshypothese ab, da sie in einer Testregression feststellten, dass die gegenwärtige Zinsdifferenz zukünftiger Zinsveränderungen signifikant negativ zusammenhängen und somit einen Erklärungsgehalt für diese haben. Dies sahen sie als ein Indiz, dass die Marktteilnehmer auf neue Informationen meistens überreagieren und eine im Durchschnitt zu starke zukünftige Zinsveränderung annehmen. HAG (1992) untersuchen Zinssätze mit mehreren Laufzeiten und wenden dabei das multivariate Johansen-Verfahren an, das u.a. eine Testmöglichkeit für den Kointegrationsraum  $\boldsymbol{\beta}$  liefert. Sie stellen dabei fest, dass die Zinssätze kointegriert sind und dass der erwartete Kointegrationsraum  $\boldsymbol{\beta}_0$  nicht abgelehnt werden kann.

Mit der Möglichkeit von fraktional integrierten Residuen untersuchen bisher Chen und Hurvich (2003) amerikanische Staatsanleihen und Iacone (2009) monatliche Da-

ten für den US-Dollar-Libor. Chen und Hurvich (2003) kamen dabei zu dem Ergebnis, dass der Kointegrationsrang nicht  $p - 1$  betrug, sondern einen geringeren Rang hatte. Dies führten sie u.a. auf das stark persistente Verhalten von langfristigen Zinsdifferenzen zurück. Iacone (2009) hingegen untersucht ein System von drei Zinssätzen mit Fälligkeiten in einem, drei und sechs Monaten mit Hilfe von Rangschätzungs- und Rangselektionsmethoden nach Robinson und Yajima (2002) und konnte einen Rang von zwei ermitteln. Die Parameter des Kointegrationsraumes schätzt Iacone (2009) als Einzelgleichungen mit Hilfe des NBL-Schätzers, da prinzipiell semiparametrische Analysewerkzeuge gewählt werden in dieser Arbeit. Als Integrationsgrad konnte Iacone (2009) mit dem lokalen Whittle-Schätzer für die einzelnen Zinszeitreihen tendenziell einen kleineren Parameter als eins ermitteln, die aber nur in wenigen Fällen auch signifikant sind. Die Kointegrationsfehler konnten als stationär angenommen werden, was wiederum auf starke Kointegrationsbeziehungen hindeutet. Insgesamt kam er dabei zu einer Annahme der Markterwartungshypothese und begründete seinen positiven Befund gegenüber der großen Anzahl der Ablehnungen, die mehrheitlich das Modell der klassischen Kointegration verwendeten, dass der Parameter  $b$  nicht immer auf den Wert 1 vereinheitlicht werden kann und es somit zu Fehlspezifikationen kommt. Anzumerken ist, dass er bei seiner Untersuchung nur Zinssätze mit relativ kurzen Laufzeiten zwischen einem und sechs Monaten untersuchte. Somit konnte er keine Aussage zu längerfristigen Zinssätzen machen. Eine weitere empirische Anwendung im Hinblick auf fraktionale Kointegration für die Markterwartungshypothese lieferte Nielsen (2010) mit dem nicht-parametrischen VR-Test in einer Erweiterung für Kointegrationsbeziehungen. Untersuchungsgegenstand waren amerikanische Staatsanleihen mit Laufzeiten zwischen drei Monaten und zwei Jahren auf täglicher Basis ab dem Jahr 1982 und konnte teilweise die Implikationen der Markterwartungshypothese bestätigen. Nielsen (2010) untersucht im Gegensatz zu Iacone (2009) keine bivariaten Systeme, sondern versucht das gesamte System der Zeitreihen in einem Ansatz zu untersuchen. Er vergleicht seine Ergebnisse auch mit klassischen Kointegrationsmethoden wie den Johansen-Ansatz, mit dessen Hilfe man die Markterwartungshypothese auf Basis der vorliegenden Daten nicht bestätigen kann. Nielsen (2010) untersucht aber nicht den Integrationsgrad der sich ergebenden Kointegrationsfehler und führt auch keine Hypothesentests für den Kointegrationsraum durch, da der multivariate VR-Ansatz zwar einen Schätzer für  $\mathbf{B}$  liefert, aber keinen Testansatz darstellt.

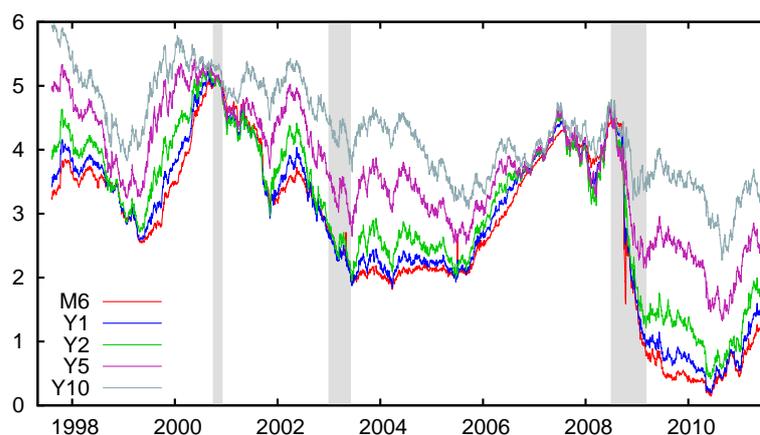


Abb. 6.1: Verlauf der Renditen deutscher Bundesanleihen im Zeitraum August 1997 bis Mai 2011

## 6.2 Überprüfung anhand täglicher Bundesanleihendaten

Um nun die Implikationen der Markterwartungshypothese zu testen, betrachten wir die Verzinsung deutscher Bundesanleihen mit einer Restlaufzeit von sechs Monaten (weiterhin: M6), einem Jahr (Y1), zwei Jahren (Y2) sowie fünf (Y5) und zehn Jahren (Y10) im Zeitraum 7. August 1997 bis 24. Mai 2011 auf täglicher Basis, was in insgesamt 3502 Beobachtungspunkten pro Reihe resultiert. Die frei zugänglichen Daten wurden der Internetseite der Deutschen Bundesbank entnommen und sind in der Abbildung 6.1 dargestellt. Schon eine erste Betrachtung des Verlaufs der Zinsen lässt erkennen, dass sich die Zinssätze gemeinsam bewegen und durch einen gemeinsamen stochastischen Trend getrieben werden könnten. In der Abbildung 6.1 sind Rezessionsphasen der deutschen Wirtschaft schattiert dargestellt, wobei eine Rezession angenommen wird, wenn das Bruttoinlandsprodukt in zwei Quartalen in Folge gesunken ist. Dies wiederum entspricht zwar nicht der gängigen Definition des Sachverständigenrates, liefert aber dennoch einen guten Anhaltspunkt zur grafischen Beurteilung. Es zeigt sich nämlich, dass in Rezessionsphasen die Zinsen für mittlere Fristigkeiten unter den kurzfristigen liegen, da der Markt auf mittlere Sicht Leitzinsänderungen erwartet und in der Bewertung der Bundesanleihen antizipiert. Für langfristige Zinsen wie fünf- und zehnjährige Anleihen stellen wir hingegen ein trägeres Verhalten fest, da deren Anlagehorizont bereits in der Vorstellung der Marktteilnehmer wieder Zinserhöhungsphasen enthalten könnte und somit kurzfristige Zinsänderungen sich relativ langsam auf die höheren Fristen durchschlagen. Gerade diese Persistenz, die wir hier erwarten können, lässt sich gut im Modellrahmen der fraktionalen Kointegration beschreiben und liefert eine Motivation, die

Tab. 6.1: Deskriptive Statistiken der Anleiherenditen

	M6	Y1	Y2	Y5	Y10
Mittelwert	2,7908	2,8965	3,1086	3,6826	4,2990
Standardabw.	1,2888	1,2477	1,1591	0,9404	0,7516
Minimum	0,15 (09.06.2010)	0,19 (08.06.2010)	0,41 (08.06.2010)	1,32 (31.08.2010)	2,27 (27.08.2010)
Maximum	5,12 (24.10.2000)	5,23 (21.08.2000)	5,32 (21.08.2000)	5,42 (19.05.2000)	5,97 (25.08.1997)

Modelle, die in dieser Arbeit vorgestellt wurden, anzuwenden. In Tabelle 6.1 finden wir deskriptive Statistiken zu den beobachteten Fristigkeiten und können daraus folgern, dass der Mittelwert mit der Fristigkeit ansteigt, was wiederum nahe legt, dass die Marktakteure eine positive Prämie für längere Laufzeiten verlangen, was aber nicht überraschen sollte. Interessanter scheint hingegen, dass die Streuung um den Mittelwert für kürzere Laufzeiten stärker ausfällt als für längere, wobei wir aber beachten sollten, dass wir hier tendenziell nicht-stationäre Zeitreihen beobachten, deren Varianzen über alle Grenzen wachsen. Nichtsdestotrotz können wir hieraus bereits erwarten, dass kurzfristige Zinsen volatiler sind als längerfristige, die vielmehr von Erwartungen zukünftiger Zinsniveaus geprägt sind. Aus den Minima und Maxima der Zinsen können wir dies auch zum Teil erkennen, da diese nicht in unmittelbarer Zeitabfolge realisiert wurden. So vergehen nach Erreichen des Minimums von M6 fast drei Monate, bis auch Y5 und Y10 an ihr Minimum gelangen. Bei den Maxima ist die Zeitspanne sogar noch länger, wenn auch in entgegengesetzter Richtung: Nach dem Maximum von Y5 vergehen fünf Monate, bis auch M6 ein Hoch erreicht. Beim Maximum von Y10 müssen wir beachten, dass dieses aus einer anderen Zinshochphase stammt.

### 6.2.1 Univariate Analyse

Im nächsten Schritt schätzen wir den Integrationsgrad mit den semiparametrischen Modellen aus Abschnitt 3.2. Wir verwenden hierzu den GPH- und den lokalen Whittle-Schätzer, den wir auf die ersten Differenzen der Zinssätze anwenden. Den Integrationsgrad der Reihe erhalten wir dann einfach durch Addieren von eins auf den geschätzten Integrationsparameter der Differenzen. Auf die Darstellung der ELW-Schätzergebnisse wollen wir hier verzichten, da es keinen Zweifel an der Nicht-Stationarität der Zeitreihen gibt und da sich diese kaum von den LW-Ergebnissen unterscheiden. Als Bandbreite verwenden wir zum einen fixierte Bandbreiten  $m = \lfloor 3502^a \rfloor$  mit  $a \in \{0,5; 0,6; 0,7; 0,8\}$  und zum anderen wenden wir auch das Plugin-

Tab. 6.2: Geschätzte Integrationsgrade auf Basis der ersten Differenzen

GPH	Fixierte Bandbreite				Plugin 1		Plugin 2	
	$m = 59$	$m = 133$	$m = 302$	$m = 684$	$\hat{m}$	$\hat{d}$	$\hat{m}$	$\hat{d}$
M6	1,2008** (0,088)	1,0324 (0,050)	1,1885*** (0,035)	1,1537*** (0,025)	196	1,0845** (0,039)	130	1,0350 (0,052)
Y1	1,1779** (0,079)	1,1806*** (0,057)	1,1358*** (0,036)	1,0891*** (0,025)	264	1,1489*** (0,039)	279	1,1395*** (0,038)
Y2	1,1273 (0,102)	1,1202** (0,056)	1,0723* (0,037)	1,0289 (0,026)	259	1,0720* (0,039)	362	1,0567* (0,033)
Y5	0,9703 (0,085)	1,0396 (0,053)	0,9907 (0,035)	1,0047 (0,024)	390	1,0071 (0,031)	709	1,0020 (0,023)
Y10	0,8368* (0,097)	0,9405 (0,056)	0,9222*** (0,035)	0,9659 (0,024)	171	0,9439 (0,047)	345	0,9457 (0,033)
LW	$m = 59$	$m = 133$	$m = 302$	$m = 684$	$\hat{m}$	$\hat{d}$	$\hat{m}$	$\hat{d}$
M6	1,2256*** (0,065)	1,0883** (0,043)	1,1778*** (0,029)	1,1261*** (0,019)	177	1,1177*** (0,038)	118	1,0941** (0,046)
Y1	1,2669*** (0,065)	1,1863*** (0,043)	1,1349*** (0,029)	1,0794*** (0,019)	240	1,1344*** (0,032)	254	1,1343*** (0,031)
Y2	1,1868*** (0,065)	1,1697*** (0,043)	1,0715** (0,029)	1,0227 (0,019)	235	1,0750** (0,033)	330	1,0817*** (0,028)
Y5	1,0557 (0,065)	1,0865** (0,043)	1,0092 (0,029)	0,9979 (0,019)	353	1,0219 (0,027)	649	1,0006 (0,020)
Y10	0,9783 (0,065)	1,0363 (0,043)	0,9662 (0,029)	0,9822 (0,019)	155	1,0187 (0,040)	312	0,9702 (0,028)

\*\*\* steht hier für die signifikante Ablehnung der Hypothese  $H_0 : d = 1$  bei einer Irrtumswahrscheinlichkeit von  $\alpha = 1\%$ , \*\* bei  $\alpha = 5\%$  und \* bei  $\alpha = 10\%$ . Die jeweiligen asymptotischen Standardabweichungen zu den geschätzten Integrationsgraden sind in Klammern angegeben. Plugin 1 steht für die Verwendung des Plugin-Verfahrens mit  $m_1 = 0,2n^{6/7}$  und Plugin 2 für  $m_1 = 0,3n^{6/7}$ .

Verfahren zum Schätzen von Bandbreiten an, das wir in Abschnitt 3.2.5 vorgestellt haben.

Hierbei zeigt sich, dass wir für kürzer laufende Zinsen einen höheren Integrationsgrad schätzen als für längere. Y1 weist dabei die tendenziell stärkste Persistenz sogar noch vor M6 auf, wenn wir uns auf die Plugin-Schätzergebnisse konzentrieren. Diese größere Integration der kurzfristigen Zinsen wiederum lässt sich dadurch erklären, dass diese stärker auf Leitzinsentscheidungen reagieren oder dass diese womöglich weitere evtl. sogar nicht-lineare Prozessbestandteile beinhalten und sich diese Nicht-Beachtung in zusätzlich beobachteter Autokorrelation niederschlägt. Mit solchen Modellen, die dies berücksichtigen können, haben wir uns in dieser Arbeit aber nicht beschäftigt und wollen deswegen diese außen vor lassen. Ansonsten können wir festhalten, dass die geschätzten Integrationsgrade nahe bei eins liegen und nur für Y10 bei Anwendung des GPH-Verfahrens für eine fixierte Bandbreite signifikant kleiner geschätzt werden können. Interessant an den Schätzergebnissen ist noch der Umstand, dass für die längerfristigen Zinsen Y5 und Y10 die Schätzergebnisse über die einzelnen Bandbreiten nicht allzu stark variieren, was wiederum den Schluss nahelegt, dass diese Prozesse keine stark ausgeprägte AR- oder MA-

Tab. 6.3: Spezifikationstests für deterministische Bestandteile

	$R_{LT}^2$	$R_{TB}^2$	t-Stat( $\hat{\mu}_2$ ) <sub>norm</sub>	$\hat{T}_b$	$MC$	$\hat{T}_b$
M6	0,3360	0,5350	0,6538	14.04.2008	13,037	15.09.2008
Y1	0,3513	0,5376	0,6346	24.04.2008	12,134	23.09.2008
Y2	0,4023	0,5550	0,5855	14.05.2008	11,478	24.07.2008
Y5	0,5285	0,6041	0,4366	19.06.2008	9,8446	24.07.2008
Y10	0,6125	0,6290	0,2112	29.09.2008	7,0988	23.09.2005

Struktur aufweisen, da sich, falls diese vorhanden wären, diese in stärker variierenden geschätzten Integrationsparametern niederschlagen würden. Dies hingegen können wir aber tendenziell für die kürzerlaufenden Anleiheverzinsungen beobachten. Eine nähere Untersuchung auf möglicherweise vorliegende Short-Memory-Komponenten werden wir aber noch in diesem Abschnitt im Rahmen einer parametrischen Analyse durchführen.

Als nächsten Schritt machen wir uns noch Gedanken über mögliche deterministische Bestandteile in den beobachteten Reihen. Dazu wollen wir testen, ob die Zinssätze einen linearen Trend bzw. einen gebrochenen Trend aufweisen. Die Abbildung 6.1 lässt zumindest die Vermutung zu, dass evtl. ein leicht fallender Trend vorliegen könnte, da die Zinsen im beobachteten Zeitraum tendenziell eher gesunken sind, wenn auch im Zeitablauf Zinshochphasen auftraten. Diesen Umstand können wir mit Hilfe des  $R^2$ -Tests aus Abschnitt 4.1.1 auch bei  $d \neq 1$  robust untersuchen und auch einen Bruch in der Trendkomponente berücksichtigen. Die Ergebnisse der Analyse befinden sich in Tabelle 6.3 und lassen den Rückschluss zu, dass wir für keine der Zinsreihen einen linearen Trend annehmen können. Für  $d = 1$  haben wir für ein Irrtumsniveau von 10% einen kritischen Wert von  $R_{LT,90\%}^2 = 0,8412$ , was von keiner der errechneten Teststatistiken erreicht werden kann. Da nun für  $d > 1$  die kritische Werte noch größer ausfallen, können wir annehmen, dass wir hier keine Evidenz für einen Trend haben. Die Möglichkeit eines gebrochenen Trends können wir auch nicht anhand der  $R_{TB}^2$ -Statistik annehmen, da hier für  $d = 1$  der 10%-kritische-Wert  $R_{TB,90\%}^2 = 0,8976$  von keinen der Reihen übertroffen wird und wir somit die Nullhypothese eines vorliegenden Trendbruchs ablehnen müssen. Die geschätzten Bruchzeitpunkte sind auch in der Tabelle 6.3 enthalten und datieren ungefähr auf das Vorfeld der Insolvenz der US-Investmentbank Lehman Brothers, welche wiederum tiefgreifende Eingriffe der globalen Notenbanken zur Folge hatte. Diese Ergebnisse werden auch von dem Meanchange-Test nach Shao (2011) bestätigt, den wir hier auf die erste Differenz der beobachteten Zinsreihen anwenden. Die geschätzten Bruchzeitpunkte liegen sogar zeitlich noch näher an gravierenden Eingriffen der Notenbanken, wobei die Verzinsung zehnjähriger Anleihen hier eine Ausnahme mar-

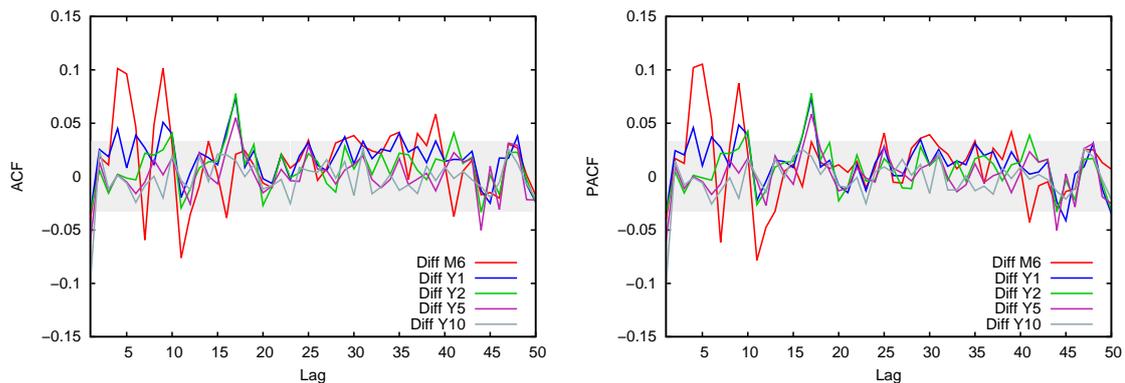


Abb. 6.2: (Partielle) Autokorrelationsfunktion der Zinssätze

kieren. Der kritische Wert auf einem Irrtumsniveau von 10% und für einen Integrationsgrad der ersten Differenzen von  $\tilde{d} = 0$  beträgt laut Response-Surface-Funktion 29,217, was von keiner der in Tabelle 6.3 aufgelisteten Teststatistiken erreicht wird. Auch hier gilt, dass für einen höheren Integrationsgrad der kritische Wert noch ansteigt und somit können wir hier die Hypothese eines gebrochenen Mittelwerts der Zinsveränderung ablehnen.

Als Abschluss der univariaten Analyse wollen wir uns noch an einer parametrischen Modellierung der Zeitreihen versuchen. Dazu wählen wir das BIC-Kriterium, das in unserer Monte-Carlo-Studie in Abschnitt 3.1.7 die besten Ergebnisse erzielt hatte. Als Schätzmethode verwenden wir nur die Whittle-Approximation, da diese auch relativ robust Modelle höherer Ordnungen schätzen kann und da sich auch hier die einzelnen geschätzten Koeffizienten, wenn man sie mit den anderen Methoden schätzt, nur geringfügig unterscheiden. Wir nehmen wiederum an, dass die Zinsreihen durch ARFIMA( $p, d, q$ )-Prozesse der Form

$$\Delta^d(1 - \theta_1 L - \dots - \theta_p L^p)y_t = (1 - \phi_1 L - \dots - \phi_q L^q)\varepsilon_t \quad (6.4)$$

beschrieben werden kann bzw. durch ARIMA( $p, 1, q$ )-Prozesse mit

$$\Delta(1 - \theta_1 L - \dots - \theta_p L^p)y_t = (1 - \phi_1 L - \dots - \phi_q L^q)\varepsilon_t \text{ für } t = 1, \dots, n. \quad (6.5)$$

Generell gehen wir weiterhin davon aus, dass die Zeitreihen nicht-stationär sind und schätzen deswegen die ARFIMA-Modelle für die ersten Differenzen der einzelnen Reihen. Genauso wie bei der semiparametrischen Analyse erhalten wir dann den Integrationsgrad der Zinsreihen durch Addieren von eins auf den jeweils geschätzten Grad der einzelnen Differenzen. Als maximale Ordnung für die AR- und MA-Bestandteile der ARFIMA- und ARIMA-Prozesse wählen wir  $p_{\max} = q_{\max} = 6$  und

Tab. 6.4: Übersicht der nach dem BIC-Kriterium selektierten Modelle bei Verwendung des parametrischen Whittle-Schätzers

	Y1	Y2	Y5	Y10
ARIMA( $p, 1, q$ )	(0, 1, 1)	(0, 1, 0)	(1, 1, 0)	(1, 1, 0)
$\theta_1$			-0,0603 (0,017)	-0,1009 (0,017)
$\phi_1$	-0,0325 (0,017)			
ARFIMA( $p, d, q$ )	(0, $d$ , 1)	(0, $d$ , 0)	(0, $d$ , 0)	(1, $d$ , 0)
$d$	1,1095 (0,028)	0,9958 (0,028)	0,9695 (0,028)	1,0009 (0,028)
$\theta_1$				-0,1018 (0,017)
$\phi_1$	-0,1555 (0,017)			

Die jeweiligen asymptotischen Standardabweichungen zu den geschätzten Koeffizienten sind in Klammern angegeben.

wählen dann das Modell mit dem niedrigsten BIC aus. Die ARIMA-Modelle wurden auch mit dem Whittle-Ansatz gewählt und geschätzt, da es innerhalb des R-Pakets „afmtools“ möglich ist, die Schätzung unter der Restriktion  $d = 1$  bzw.  $d = 0$  für die ersten Differenzen durchzuführen. Die jeweils selektierten Modelle geben wir für alle Zinssätze außer M6 in Tabelle 6.4 und für M6 gesondert in Tabelle 6.5 an. Der Whittle-Ansatz liefert für die jeweils geschätzten Koeffizienten asymptotische Standardabweichungen, die, ausgehend von den jeweils gewählten Modellordnungen und der jeweiligen impliziten Spektraldichte für die einzelnen AR-Parameter  $\theta_i$  mit  $i = 1, \dots, p$  gleich sind. Analog gilt das gleiche für die MA-Parameter  $\phi_j$  mit  $j = 1, \dots, q$ .

Prinzipiell ähneln sich die Ergebnisse außer für die Reihe M6, da generell relativ niedrige Modellordnungen selektiert werden. Für die ARFIMA-Modelle werden zumeist Integrationsgrade nahe bei eins geschätzt außer für die Reihen Y1 und M6. Für Y1 ergibt sich aber aus der Betrachtung der Summe von  $d$  und  $\phi_1$  kein großer Unterschied zu eins, was wiederum nahe legt, dass hier zwei Effekte in der Modellgleichung geschätzt werden, die sich gegenseitig neutralisieren und somit auch tendenziell keine vorhandene Autokorrelation modelltheoretisch erklären können. Wenn wir die empirisch ermittelten Autokorrelationsfunktionen der ersten Differenzen betrachten, welche in Abbildung 6.2 dargestellt sind, können wir feststellen, dass außer für die sechsmonatige Verzinsung kaum signifikante Autokorrelationen auftauchen. Dies können wir daran erkennen, dass in den Abbildungen die jeweils schattierten Flächen das 95%-Konfidenzintervall für einen Autokorrelationsparameter von

Tab. 6.5: Übersicht der nach dem BIC-Kriterium selektierten Modelle für M6 bei Verwendung des parametrischen Whittle-Schätzers

	$d$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\phi_1$	$\phi_2$	$\phi_3$
ARFIMA(4, $d$ , 2)	1,1510 (0,028)	0,1994 (0,058)	-0,7501 (0,058)	-0,1810 (0,058)	0,0112 (0,058)		0,4080 (0,042)	-0,7529 (0,042)	
ARIMA(5, 1, 3)	1 (-)	0,7090 (0,048)	-0,7904 (0,048)	0,4390 (0,048)	0,1192 (0,048)	0,0644 (0,048)	-0,7651 (0,040)	0,8578 (0,040)	-0,4864 (0,040)
ARFIMA(1, $d$ , 1)	1,1991 (0,028)	0,1926 (0,018)					-0,4463 (0,021)		
ARIMA(1, 1, 1)	1 (-)	-0,8427 (0,058)					0,8143 (0,050)		

Die jeweiligen asymptotischen Standardabweichungen zu den geschätzten Koeffizienten sind in Klammern angegeben.

null symbolisiert und wir kaum größere Ausschläge über diesen Bereich verzeichnen. Dies kann natürlich auch erklären, dass wir generell auch nur relativ kleine AR- und MA-Parameter schätzen, wenn diese überhaupt vom BIC selektiert wurden. Neben den empirischen Autokorrelationsfunktionen können wir auch anhand des Periodogramms der Zeitreihen empirisch beurteilen, ob ein geschätztes Modell vorliegende Strukturen erklären kann. Dazu haben wir in Abbildung 6.3 neben den geschätzten Periodogrammen der ersten Differenzen, die jeweils sich aus den geschätzten Modellparametern ergebenden theoretischen Spektren, dargestellt. Zusätzlich ist hier noch für jede Reihe das über ein Bartlett-Fenster geglättete Periodogramm hinzugefügt, um generelle Strukturen im Periodogramm zu entdecken. Hier zeigt sich nun für die Zinszeitreihen, dass das Periodogramm zumeist flach verläuft und die theoretischen Spektraldichten kaum Strukturen erklären, da diese einfach nicht vorkommen. Dies spricht dafür, dass es sich bei den ersten Differenzen um nahezu Weißes Rauschen handelt.

Für die Zeitreihe der sechsmonatigen Zinssätze gestaltet sich nun die Situation etwas schwieriger, da hier das BIC relativ hohe ARFIMA- und ARIMA-Ordnungen präferiert. Als Alternative hierzu geben wir auch die Schätzergebnisse von einfacheren ARFIMA(1,  $d$ , 1)- und ARIMA(1, 1, 1)-Modellen in Tabelle 6.5 an. Hier können wir aber auch feststellen, dass der geschätzte Integrationsgrad der ARFIMA-Modelle signifikant größer als eins geschätzt werden kann. Zur Begründung der Schätzergebnisse können wir auch die Darstellungen der Abbildung 6.3 heranziehen, wo wiederum theoretische Spektraldichte mit dem empirischen Periodogramm verglichen werden. Hier können wir z.B. am geglätteten Periodogramm erkennen, dass dieses keine flache, sondern eine kompliziertere Struktur aufweist als bei den anderen untersuchten Zinszeitreihen. Diese Struktur kann in der Tat von den BIC-selektierten Modellen

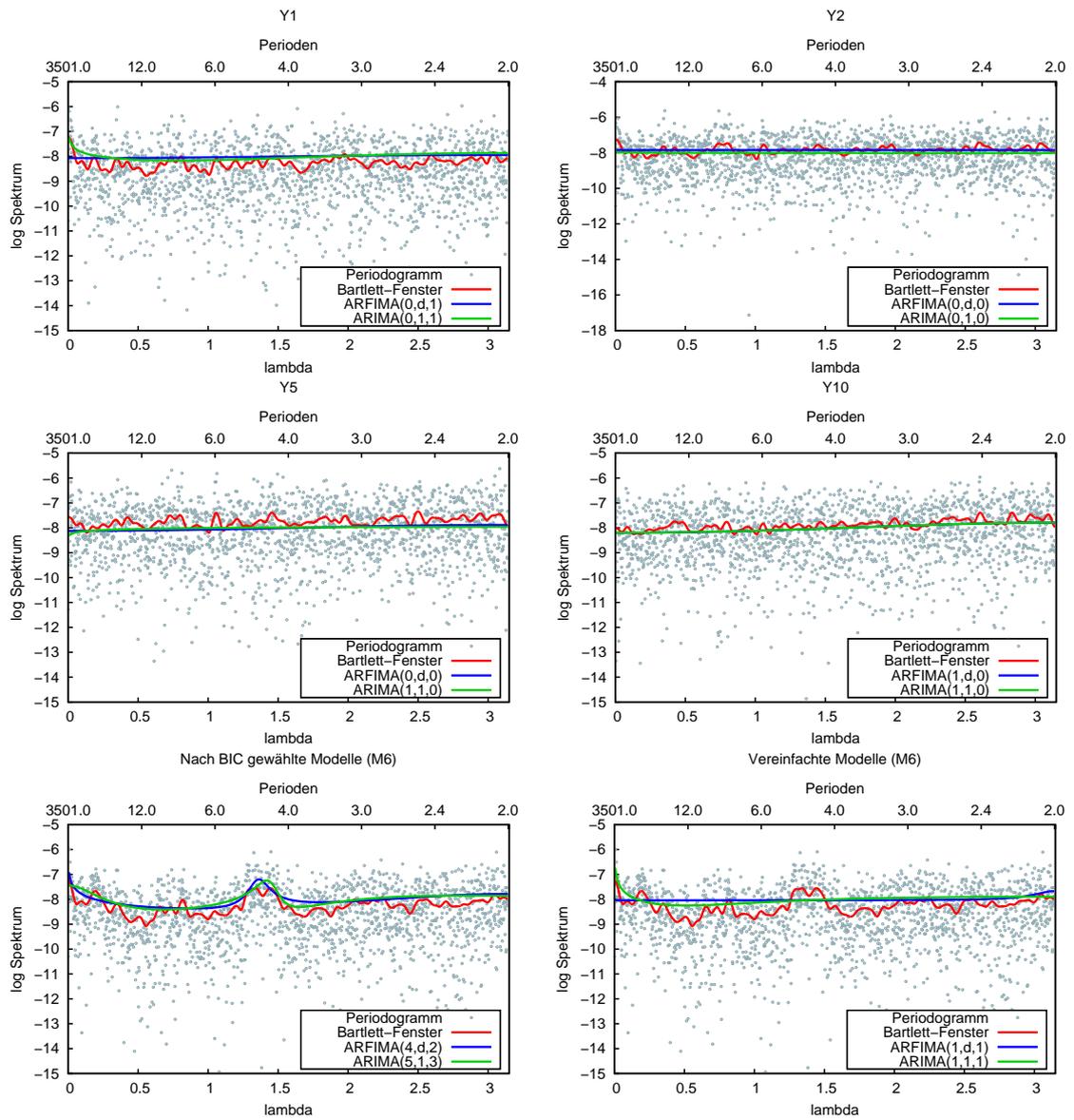


Abb. 6.3: Anpassung der theoretischen Spektren an die geschätzten Periodogramme auf log-Basis

erklärt werden, da diese den Ausschlag des Periodogramms nachempfindet, der sich ungefähr bei der Fourierfrequenz  $\lambda \approx 1,45$  befindet, was einer Periodizität von ungefähr fünf entspricht. Dies wiederum lässt sich gerade durch den Umstand, dass wir Tagesdaten beobachten, begründen und wir anscheinend eine Saisonalität zum Lag fünf haben. Das Interessante wiederum ist nun, wieso wir das gerade bei dem sechsmonatigen Zinssatz beobachten und nicht bei den längeren Fristigkeiten. Hierzu könnte man vermuten, dass gerade die Bundesanleihen mit kürzerer Restlaufzeit womöglich mehr gehandelt werden als längerlaufende Anleihen und somit z.B. der ausbleibende Handel an Wochenenden eine größere Auswirkung hat, da an jedem Montag, der einen Handelstag darstellt, Informationen vom Wochenende nachgeholt werden und wir gerade an diesem Tag eine höhere Volatilität haben als an anderen Tagen. Dieser Effekt fällt wesentlich kleiner aus, wenn generell weniger Handel mit Wertpapieren durchgeführt wird und dies könnte ein Grund dafür sein, dass wir hier diese Anomalie beobachten. Dass nun insbesondere der Whittle-Schätzer mit den nach dem BIC-Kriterium bestimmten Modellen dies gut nachvollziehen kann, ist auch nicht überraschend, da der Whittle-Schätzer eine möglichst gute Anpassung des Spektrums an das vorliegende Periodogramm versucht und somit zu diesem Ergebnis kommt. Die einfacheren ARFIMA- und ARIMA-Modelle scheitern hingegen und schätzen wiederum eine flachere Struktur für die Reihe M6. Nur das ARFIMA-Modell kann aufgrund des signifikant geschätzten Integrationsparameter zumindest die leichte Steigung des Periodogramms am Ursprung beschreiben.

Dass wir diese Steigung aber bei der sechsmonatigen und auch der einjährigen Verzinsung im Gegensatz zu den längerlaufenden Verzinsungen beobachten, scheint zunächst merkwürdig. Ein Grund könnte in der bereits erwähnten zeitlich stärkeren Abhängigkeit zur Leitzinsentscheidung liegen. Bei längeren Anlagehorizonten müssen Marktteilnehmer mehr Erwartungen über zukünftige Zinsentscheidungen treffen, was wiederum tendenziell zu einer Glättung des Zinsniveaus führen kann, da Zentralbanken z.B. auf Rezessionen mit Zinssenkungen reagieren und nach überstandener Rezession die Zinsen wieder angehoben werden. Dies entfällt z.B. bei einem sechsmonatigen Anlagehorizont, da es zumeist nicht zu erwarten ist, dass eine Rezession innerhalb von sechs Monaten ausgestanden ist. Weiterhin könnten auch mehrere Strukturbrüche für diese Schätzergebnisse verantwortlich sein. Wir haben in dieser Arbeit lediglich einen einzigen Bruch in der deterministischen Komponente berücksichtigt, was wiederum in Anbetracht des Beobachtungszeitraums zu restriktiv sein könnte. Die Nicht-Berücksichtigung dieser Bestandteile kann durchaus eine zusätzliche Steigung im Periodogramm verursachen und somit auch zu solchen Schätzer-

Tab. 6.6: Geschätzte Integrationsgrade für M6 und Y10 in verschiedenen Stichproben

LW	Zweijahreszeiträume					
	1998-00	2000-02	2002-04	2004-06	2006-08	2008-10
M6	1,2108*** (0,068)	1,2239** (0,087)	1,0949 (0,072)	1,1467 (0,096)	1,1340 (0,094)	1,1438* (0,081)
Y10	0,9876 (0,08)	1,0566 (0,083)	0,9443 (0,078)	1,0060 (0,081)	1,0403 (0,076)	0,8381** (0,079)
LW	Vierjahreszeiträume					
	1998-02	2002-06	2006-10			
M6	1,2062*** (0,061)	1,1969*** (0,066)	1,0956 (0,068)			
Y10	0,9965 (0,062)	0,9459 (0,059)	1,0042 (0,035)			

\*\*\* steht hier für die signifikante Ablehnung der Hypothese  $H_0 : d = 1$  bei einer Irrtumswahrscheinlichkeit von  $\alpha = 1\%$ , \*\* bei  $\alpha = 5\%$  und \* bei  $\alpha = 10\%$ . Die jeweiligen asymptotischen Standardabweichungen zu den geschätzten Integrationsgraden sind in Klammern angegeben. Die geschätzten Integrationsgrade wurden lediglich mit Plugin 2 (d.h. unter Verwendung  $m_1 = 0,3n^{6/7}$ ) mit dem lokalen Whittle-Schätzers ermittelt.

gebnissen führen. Für lineare, stationäre Zeitreihenmodelle haben Bai und Perron (1998) eine Schätz- und Testprozedur vorgeschlagen, um mehrere Strukturbrüche zu ermitteln. Dies wiederum wurde auch für die einzelnen ersten Differenzen der Zinssätze durchgeführt und führte aber zu dem Ergebnis, dass für alle Reihen kein einziger Strukturbruch im Mittelwert im Beobachtungszeitraum gewählt wurde. Dies sagt aber lediglich aus, dass in einem linearen Modell z.B. für M6 kein Bruch angenommen werden kann und wir hier in dieser Arbeit kein geeignetes nicht-lineares Modell behandelt haben, mit der wir diese größere Persistenz der Zinsreihen erklären könnten.

Ein letzter Punkt, mit der wir die univariate Analyse schließen wollen, wäre noch, dass wir nachprüfen, ob denn der Integrationsparameter im Zeitablauf konstant geblieben ist. Hierzu wollen wir den Beobachtungszeitraum in jeweils Unterstichproben mit zwei bzw. vier Jahren Dauer unterteilen und für diese jeweils den Integrationsgrad testen. Wir wollen uns hier aber nur auf den lokalen Whittle-Schätzer beschränken und diesen nicht auf verschiedene fixierte Bandbreiten anwenden, sondern die Bandbreite datengetrieben mit der Plugin-Methode bestimmen. Auch bei den Zinsreihen wollen wir uns nur auf M6 als Beispiel der kurzlaufenden Anleihenverzinsung und Y10 für die längerfristigen konzentrieren. Die Ergebnisse hierzu befinden sich in Tabelle 6.6 und lassen den Schluss zu, dass für Y10 fast durchgängig der Integrationsgrad nicht signifikant von eins verschieden geschätzt werden kann, außer in der Periode ab 2008, also ungefähr mit Auftreten der globalen Finanzkrise. Bei M6 haben wir hingegen tendenziell meist einen Integrationsgrad größer als eins und insbesondere in der ersten Hälfte unseres gesamten Beobachtungszeitraums einen

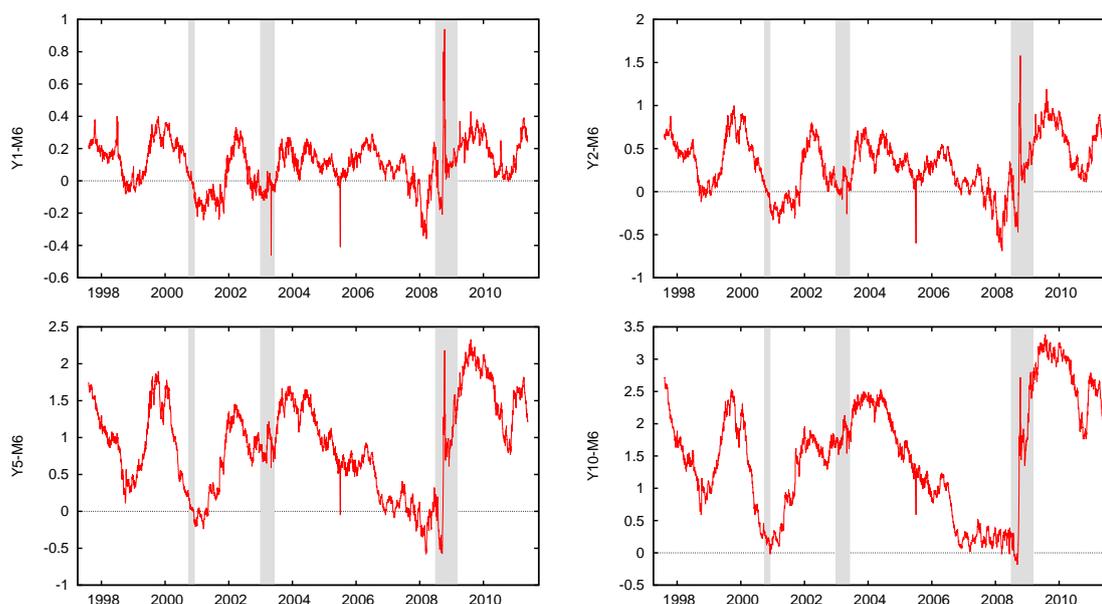


Abb. 6.4: Zinsdifferenzen deutscher Bundesanleihen verschiedener Laufzeiten

höheren Parameter  $d$ . Die Variabilität ist dabei aber auch nicht stark ausgeprägt und der Integrationsgrad geht für die zweite Hälfte des Beobachtungszeitraums näher gegen eins und ist dann auch nicht mehr signifikant davon verschieden.

Zusammenfassend können wir festhalten, dass die beobachteten Zinssätze tendenziell integriert vom Grad eins sind bzw. einen Integrationsgrad aufweisen, der leicht darüber liegt, und keine deterministischen Komponenten aufweisen sowie keine stark ausgeprägte Autokorrelation in den ersten Differenzen besitzen. Der Integrationsgrad weist zwar eine leichte Abhängigkeit im Bezug zur Fristigkeit auf, doch können wir keinen signifikanten Unterschied im Integrationsparameter ausmachen, so dass wir von einem gleichen  $d$  für die einzelnen Zinsreihen ausgehen können. Mit diesen Erkenntnissen können wir nun im Rahmen der multivariaten Analyse beginnen, die Markterwartungshypothese zu testen.

### 6.2.2 Multivariate Modellierung der Zinssätze

Im Folgenden beginnen wir nun mit der multivariaten Analyse der beobachteten Zinssätze, wobei wir zunächst univariate Analysewerkzeuge auf die Zinsdifferenzen anwenden werden. Unter Annahme des klassischen kointegrierten Modells können wir laut Markterwartungshypothese erwarten, dass die einzelnen Differenzen zwischen den Zinsen, die integriert vom Grad eins sind, wiederum stationäre  $I(0)$ -Prozesse sind. In Abbildung 6.4 haben wir nun die Differenz der einzelnen Zinssätze zu M6 im Beobachtungszeitraum dargestellt. Wir wählen hier M6 als Bezugsgröße,

da dies die kürzeste Frist darstellt und somit durch die Zentralbank am ehesten durch geldpolitische Maßnahmen beeinflussbar ist. Aus der Betrachtung des Verlaufs der einzelnen Zinssatzdifferenzen kann man schon vermuten, dass nicht jede der Differenzen das gleiche Ausmaß an Integration aufweist, da die einzelnen Reihen anscheinend unterschiedlich persistent sind. Für die kürzeste Laufzeitdifferenz, also zwischen Y1 und M6, lässt sich womöglich noch ein stationärer Prozess vermuten, während der Verlauf von Y10-M6 eher einem Random Walk ähnelt. Diese Vermutungen werden zunächst auch durch die Betrachtung der Autokorrelationsfunktion bestätigt, da das Ausmaß der Autokorrelation mit Zunahme der Laufzeitdifferenz zunimmt. Aber selbst für Y1-M6 beobachten wir für sehr große Lags immer noch eine große Autokorrelation, die auf jedem Irrtumsniveau signifikant von null verschieden ist. Dies könnte somit auf eine Long-Memory-Eigenschaft dieser Reihe deuten. Die partielle Autokorrelationsfunktion hat hingegen vor allem für die ersten Lags einen deutlichen Ausschlag und bricht danach scheinbar, was wiederum für einen Random Walk der beobachteten Reihe spricht. Für Y1-M6 können wir einen Koeffizienten von 0,9732 und für Y2-M6 0,9886 für den ersten Lag schätzen, was schon sehr nahe eins liegt. Wenn wir aber den klassischen ADF-Test für diese beiden Differenzen durchführen, kommen wir zu dem Ergebnis, dass selbst unter Berücksichtigung einer hohen Lagzahl wir die Nicht-Stationarität im klassischen Sinne ablehnen können. Für Y5-M6 und Y10-M6 hingegen lässt die Anwendung des ADF-Tests die Ablehnung der Nicht-Stationarität nicht zu und wir beobachten hier partielle Autokorrelationskoeffizienten erster Ordnung von 0,9959 bzw. 0,9976, was somit sehr auf eine Nicht-Stationarität der Reihen hindeutet. Diese ersten Betrachtungen lassen schon die Vermutung aufkommen, dass die klassischen Kointegrationsmethoden nicht unbedingt geeignet sind, um die Markterwartungshypothese für die Verzinsung der Bundesanleihen zu testen, da wir hier wahrscheinlich starke Long-Memory-Eigenschaften in den Kointegrationsstörtermen haben oder es gar nur mit schwachen Kointegrationsbeziehungen zu tun haben.

Die Tabelle 6.7 ermöglicht uns wieder mit Hilfe von deskriptiven Statistiken eine grobe Einordnung der einzelnen Zinsdifferenzen, die wir aber zum Teil schon im vorangegangenen Abschnitt gewinnen konnten. So steigt der Mittelwert der Differenz mit der Länge der Laufzeitdifferenzen an, was wiederum eine positive Laufzeitprämie vermuten lässt. Die Standardabweichung steigt auch mit dem Laufzeitunterschied an, was womöglich aber in einer jeweils stärker werdenden Long-Memory-Eigenschaft der einzelnen Zinsdifferenz begründet sein könnte. Die jeweiligen Maxima wurden

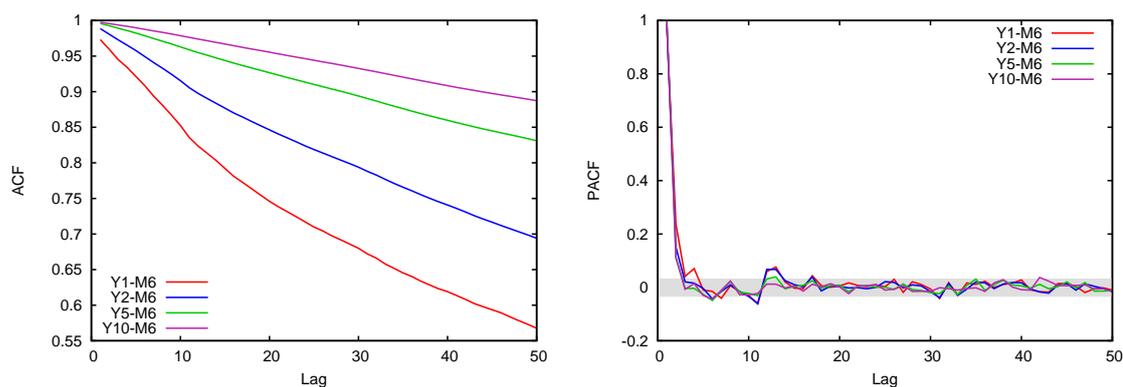


Abb. 6.5: (Partielle) Autokorrelationsfunktion der Zinsdifferenzen

Tab. 6.7: Deskriptive Statistiken der Zinsdifferenzen

	Y1-M6	Y2-M6	Y5-M6	Y10-M6
Mittelwert	0,1056	0,3177	0,8917	1,5081
Standardabw.	0,1414	0,3206	0,6283	0,8899
Minimum	-0,46 (02.05.2003)	-0,69 (17.03.2008)	-0,58 (13.03.2008)	-0,18 (10.09.2008)
Maximum	0,94 (13.10.2008)	1,58 (13.10.2008)	2,33 (10.08.2009)	3,38 (27.07.2009)

durchweg im Vorfeld der anbahnenden Lehman-Insolvenz realisiert und die Minima jeweils zu Beginn des Jahres 2008. Das Minimum von Y1-M6 hingegen wurde durch einen einzelnen Ausreißer im Jahr 2003 markiert, was aber womöglich durch eine fehlerhafte Berechnung der Zinssätze an diesem einzelnen Tag oder durch eine andere Anomalie begründet werden kann.

Nun bietet es sich wieder an, den Integrationsgrad der Zinsdifferenzen zu schätzen, wobei wir hier den exakten lokalen Whittle-Schätzer verwenden sollten, da letztendlich eine Prognose über den zu erwartenden Wertebereich der geschätzten Integrationsparameter schwer fällt. Die ELW-Schätzergebnisse in Tabelle 6.6 sind ähnlich wie im vorangegangenen Abschnitt mit fixierten und mit per Plugin-Verfahren geschätzten Bandbreiten ermittelt worden. Hier können wir erkennen, dass wir für alle Zinsdifferenzen einen Integrationsgrad signifikant über der Stationaritätsgrenze von 0,5 schätzen können, was somit die Vermutung über die Nicht-Stationarität der Differenzen bestätigt. Des Weiteren lässt sich feststellen, dass, je länger der Laufzeitunterschied zwischen den einzelnen Differenzen wird, desto größer auch tendenziell der Integrationsparameter wird. Ein Grund hierfür könnte zum einen in der langsameren Informationsverarbeitung bzw. in der impliziten Glättung der zukünftigen Zinsniveaus durch die Marktteilnehmer liegen, was bereits schon thematisiert wurde.

Tab. 6.8: Geschätzte Integrationsgrade der Zinsdifferenzen

ELW	Fixierte Bandbreite				Plugin 1		Plugin 2	
	$m = 59$	$m = 133$	$m = 302$	$m = 684$	$\hat{m}$	$\hat{d}$	$\hat{m}$	$\hat{d}$
Y1-M6	0,7056*** (0,065)	0,7690*** (0,043)	0,9381** (0,029)	0,9074*** (0,019)	79	0,7814*** (0,056)	112	0,7605*** (0,047)
Y2-M6	0,8358*** (0,065)	0,8502*** (0,043)	0,9962 (0,029)	0,9764 (0,019)	92	0,8385*** (0,052)	114	0,8622*** (0,047)
Y5-M6	0,9466 (0,065)	0,9514 (0,043)	1,0450 (0,029)	1,0376 (0,019)	104	0,9710 (0,049)	159	0,9599 (0,040)
Y10-M6	1,0265 (0,065)	1,0267 (0,043)	1,0791 (0,029)	1,0460 (0,019)	139	1,0116 (0,042)	435	1,0650 (0,024)

\*\*\* steht hier für die signifikante Ablehnung der Hypothese  $H_0 : d \geq 1$  gegen  $H_1 : d < 1$  bei einer Irrtumswahrscheinlichkeit von  $\alpha = 1\%$ , \*\* bei  $\alpha = 5\%$  und \* bei  $\alpha = 10\%$ . Die jeweiligen asymptotischen Standardabweichungen zu den geschätzten Integrationsgraden sind in Klammern angegeben.

Plugin 1 steht für die Verwendung des Plugin-Verfahrens mit  $m_1 = 0,2n^{6/7}$  und Plugin 2 für  $m_1 = 0,3n^{6/7}$ .

Weiterhin lassen die geschätzten Integrationsgrade daraufhin schließen, dass womöglich keine Kointegration zwischen dem fünfjährigen und zehnjährigen Zinssatz zur sechsmonatigen Verzinsung vorliegt, da wir für die entsprechenden Differenzen keinen signifikant kleineren Integrationsgrad als eins schätzen können. Aus diesem Grund geben wir in Tabelle 6.8 das Signifikanzniveau für die Nullhypothese  $H_0 : d \geq 1$  an, im Gegensatz zu Tabelle 6.2, wo wir uns nur für  $H_0 : d = 1$  interessieren. Für die Differenz Y2-M6 haben wir zumindest Indizien, dass wir evtl. allenfalls eine schwache Kointegrationsbeziehung vorliegen haben, da zum einen die Differenz des Integrationsparameters zu eins sehr klein ausfällt als auch, weil bei höheren Bandbreiten die Differenz nicht mehr signifikant geschätzt werden kann. An den Schätzergebnissen fällt im Gegensatz zur univariaten Untersuchung der Zinssätze auf, dass die Variabilität der Schätzungen über die verschiedenen Bandbreiten deutlicher ausgeprägt ist, was wiederum für ein Vorliegen von weiteren autokorrelierten Bestandteilen im datengenerierenden Prozess spricht. Aus diesem Grund haben wir in Abbildung 6.6 sämtliche geschätzten Integrationsgrade für jede Bandbreite zwischen 50 und 1000 mit einem zweiseitigen 95%-Konfidenzintervall um das jeweilige Schätzeresultat graphisch dargestellt. Hier sehen wir gerade für Y1-M6 und Y2-M6 einen erheblichen Anstieg im Integrationsparameter bei den kleineren Bandbreiten, was für eine stark ausgeprägte Autokorrelationsstruktur in den Störtermen spricht. Für Y5-M6 und Y10-M6 können wir feststellen, dass wir nahezu für jede Bandbreite einen größeren Integrationsgrad als eins haben, was eine Kointegrationsbeziehung eher ausschließen lässt.

Als nächsten Schritt bietet sich nun die Anwendung der Integrationsstests aus

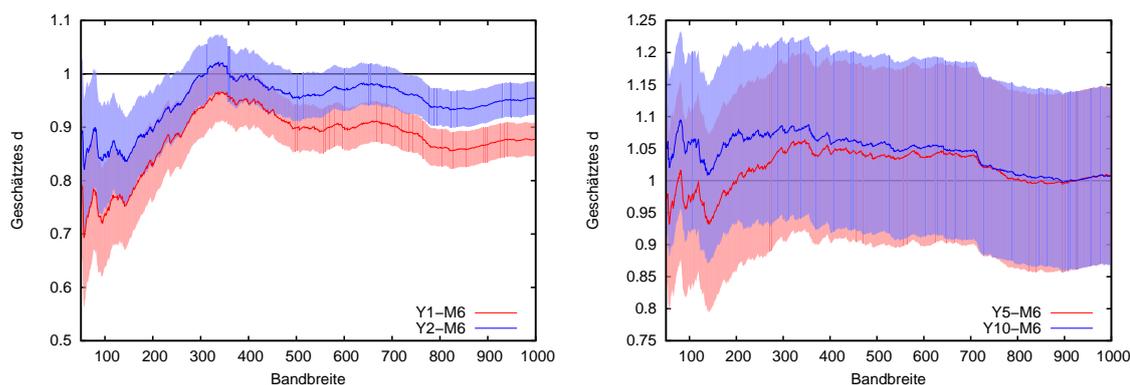


Abb. 6.6: ELW-Schätzergebnisse für den Integrationsgrad der Zinsdifferenzen in Abhängigkeit der Bandbreite  $m$

Tab. 6.9: Integrationstests für die Zinsdifferenzen

	$p = 1$	NEW				VR	
		$p = 5$	$p = 9$	$p = 29$	$d_1 = 0,1$	$d_1 = 1$	
Y1-M6	-5,1991***	-3,6247***	-5,3512***	-2,3259***	2,2012***	410,3***	
Y2-M6	-2,1676**	-2,6320***	-3,8731***	-1,2548	2,0549***	336,58***	
Y5-M6	0,9576	-0,8753	-1,5140*	-0,3568	1,8675**	185,32***	
Y10-M6	1,6012	0,9904	0,2711	0,6157	1,7842*	123,59**	

\*\*\* steht hier für die signifikante Ablehnung der Hypothese  $H_0 : d \geq 1$  gegen  $H_1 : d < 1$  bei einer Irrtumswahrscheinlichkeit von  $\alpha = 1\%$ , \*\* bei  $\alpha = 5\%$  und \* bei  $\alpha = 10\%$ .

Kapitel 4 an, um weitere Evidenz, ob der Integrationsgrad der Zinsdifferenzen kleiner als eins, zu erlangen. Wir verwenden hierzu den effizienten Wald-Test (NEW) aufgrund seiner größten Macht im Vergleich zu den anderen parametrischen Integrationstests und den VR-Test, da dieser nicht-parametrisch angewendet werden kann. Als deterministische Komponente nehmen wir keinen Trend an, da wir bereits in den Zinszeitreihen selbst keinen Trend bestätigen konnten und da die Abbildung 6.4 allenfalls einen positiven Mittelwert vermuten lässt, um den die Zinsdifferenzen schwanken. Dies hat zur Folge, dass wir für NEW keine zusätzliche deterministische Komponente in die Testregression aufnehmen werden und wir bei Anwendung des VR-Tests zunächst eine Mittelwertzentrierung vor Berechnung der Teststatistik durchführen werden. Als Lagzahl wählen wir  $p = 1$  und  $p = 5$  als fixierten Vorschlag, der dem Umstand, dass wir hier Tagesdaten haben, geschuldet ist, und zum anderen nach den Schwert-Heuristiken. Des Weiteren benötigen wir noch zur Anwendung des NEW-Tests den geschätzten Integrationsgrad der Differenzen, wofür wir jeweils den Mittelwert der beiden Plugin-Schätzergebnisse aus Tabelle 6.6 verwenden.

Als Ergebnis der Testanwendung können wir festhalten, dass Y2-M6 und Y1-M6 tendenziell einen kleineren Integrationsgrad als eins aufweisen und wir für Y5-M6 und

Tab. 6.10: ELW-Schätzergebnisse für bivariate Kointegrationsbeziehungen

	$m$	$\beta$		$\hat{d}$	$\hat{d}_u$		ACF( $1, \Delta^{\hat{d}_u} \hat{u}_t$ )
(Y1,M6)	88	0,9753	(-)	1,2782	0,7452	(0,053)	0,0054
	133	0,9658	(0,019)	1,1880	0,7724	(0,042)	-0,0269
	201	0,9548	(0,032)	1,1449	0,8352	(0,034)	-0,0971
(Y2,M6)	88	1,0312	(0,112)	1,1822	0,9277	(0,051)	-0,0906
	133	0,9819	(0,078)	1,1490	0,8767	(0,042)	-0,0302
	201	1,1902	(0,187)	1,0743	0,9527	(0,033)	-0,1203
(Y5,M6)	88	10,691	(3,971)	1,0992	1,0857	(0,038)	-0,1567
	133	-17,771	(6,243)	1,0280	1,0345	(0,031)	-0,0782
	201	-1,0431	(0,738)	1,0198	1,0648	(0,029)	-0,1399
(Y10,M6)	88	-0,5875	(0,303)	1,0172	1,1702	(0,050)	-0,3476
	133	-2,2299	(1,086)	1,0275	1,0618	(0,033)	-0,1395
	201	-0,0129	(0,235)	1,0001	1,1093	(0,035)	-0,2549

Die jeweiligen asymptotischen Standardabweichungen zu den geschätzten Koeffizienten sind in Klammern angegeben. (-) steht dafür, dass keine Standardabweichung bestimmbar war. Die geschätzten Integrationsgrade der stochastischen Trends und des Kointegrationsfehlers weisen die gleiche asymptotische Standardabweichung auf.

Y10-M6 zumindest bei Betrachtung der VR-Statistik eine leichte Evidenz für eine Kointegrationsbeziehung haben. Dies wollen wir nun weitergehend in der Kointegrationsanalyse untersuchen und beschäftigen uns zunächst weiter mit der Betrachtung von bivariaten Systemen.

Hierzu wollen wir den multivariaten ELW-Schätzer aus Abschnitt 5.4.1 verwenden, um die jeweiligen Zinssatzzeitreihen zusammen mit der sechsmonatigen Verzinsung in einer fraktionalen Kointegrationsbeziehung zu schätzen. Die Bandbreite wählen wir wiederum mit  $m = \lfloor 3502^a \rfloor$  und setzen  $a \in \{0,5; 0,55; 0,6\}$  tendenziell eher konservativ, da wir mit einer zusätzlichen autoregressiven Struktur in den Kointegrationsfehlern rechnen. Dies wiederum hat zur Folge, dass wir zwar evtl. weniger effiziente Ergebnisse erhalten, die aber robust geschätzt werden können. Die Ergebnisse fassen wir in Tabelle 6.10 zusammen und bezeichnen hier als  $\hat{d}$  den Integrationsgrad des stochastischen Trends des kointegrierten Systems und  $\hat{d}_u$  als den Integrationsgrad des Kointegrationsfehlers  $u_t = y_t - \beta x_t$  mit M6 als  $x_t$ . Dieser entspricht wiederum modelltheoretisch der Differenz des Integrationsgrads des stochastischen Trends und der Kointegrationsstärke  $b$ . Weiterhin berechnen wir mit  $\text{ACF}(1, \Delta^{\hat{d}_u} \hat{u}_t)$  die Autokorrelation erster Ordnung für den fraktional gefilterten geschätzten Kointegrationsfehler, um eine Aussage über das Ausmaß der Autokorrelation im Kointegrationsstörterm treffen zu können.

Als Ergebnis lässt sich festhalten, dass sich die Vermutungen aus der univariaten Analyse der einzelnen Zinsdifferenzen bestätigen. Die Paare (Y1,M6) und (Y2,M6) scheinen zumindest schwach kointegriert zu sein mit einem nicht von eins signifikant unterschiedlichen Parameter  $\beta$  mit Ausnahme der Schätzergebnisse für die

Bandbreite  $m = 133$  für (Y1,M6). Für die beiden Datenvektoren mit den längeren Laufzeitunterschieden (Y5,M6) und (Y10,M6) können wir zum Teil sogar höhere Integrationsparameter für den Kointegrationsfehler schätzen als für den stochastischen Trend, was sehr stark gegen das Vorliegen einer Kointegrationsbeziehung spricht. Dementsprechend fallen auch die Schätzwerte für  $\beta$  sehr volatil aus. Interessant ist noch die Betrachtung der Autokorrelation der gefilterten Kointegrationsfehlerterme  $\Delta^{\hat{d}_u} u_t$ , die nahezu geschlossen negativ ausfallen. Ein Grund hierfür könnte sein, dass der Integrationsgrad des Kointegrationsfehlers  $d_u$  tendenziell überschätzt wurde und der wahre Grad etwas darunter liegen könnte. Dies wiederum würde dafür sprechen, dass eventuell auch die Paare mit den länger ausfallenden Laufzeitunterschieden doch noch kointegriert sein könnten. Eine Anwendung des multivariaten ELW-Schätzers auf höhere Bandbreiten bringt in diesem vorliegenden zumeist noch höhere Integrationsparameter für den stochastischen Trend und Kointegrationsfehler, so dass wir die weitere Analyse mit einer anderen Schätzmethode fortführen sollten.

Neben den semiparametrischen Methoden steht uns mit dem FCVAR-Modell, das wir in Abschnitt 5.4.2 behandelt haben, noch ein parametrisches Modell zur Schätzung von fraktional kointegrierten Modellen zur Verfügung. Wir untersuchen hiermit die gleichen Datenvektoren wie mit dem multivariaten ELW-Schätzer zuvor und stellen die Schätzergebnisse in der Tabelle 6.11 kompakt dar. In dieser fassen wir die einzelnen Restriktionsmöglichkeiten bzgl. der Integrationsparameter  $d$  und  $b$  zusammen und haben jeweils die geschätzten Koeffizienten für die Situationen  $r = 0$  (d.h. es liegt keine Kointegration vor) und  $r = 1$  (es gibt eine Kointegrationsbeziehung). Für den Fall  $r = 1$  stellen wir noch jeweils die Komponenten der geschätzten Kointegrationsmatrizen  $\hat{\mathbf{A}} = (a_1, a_2)'$  und  $\hat{\mathbf{B}} = (1, -\beta)'$  dar. Wir schätzen für jede Restriktionsmöglichkeit das Modell mit einer restringierten Konstanten, da wir laut Markterwartungshypothese durchaus eine positive Laufzeitprämie erwarten und wir diese als konstanten Mittelwert der einzelnen Kointegrationsbeziehungen betrachten. Dies wiederum beschreiben wir in Tabelle 6.11 mit dem Koeffizienten  $\alpha$  für jede der möglichen Kointegrationsbeziehungen. Des Weiteren müssen wir uns noch Gedanken über die Wahl des Lag-Parameters  $k$  machen. Hierzu bietet es an, sich das BIC zur Lag-Determination heranzuziehen, was für sämtliche Restriktionen und Kointegrationspäarchen  $k = 1$  selektiert. Daneben stellen wir noch die Schätzergebnisse für  $k = 4$  dar, um die Schätzergebnisse abzusichern.

Zusammenfassend von Tabelle 6.11 lässt sich festhalten, dass die Schätzergebnisse für das unrestringierte Modell sehr instabil aussehen. Nichtsdestotrotz können wir

Tab. 6.11: (F)CVAR-Schätzergebnisse für bivariate Kointegrationsbeziehungen

Modell	$k$	$r = 0$			$r = 1$			Komponenten von $\hat{A}$ und $\hat{B}$			
		$L R(r=0)$	$d$	$b$	$L R(r=1)$	$d$	$b$	$\alpha$	$\beta$	$a_1$	$a_2$
FCVAR ( $d$ und $b$ unrestringiert)	1	<i>16,474***</i>	1,2093	0,0004	<i>0,7759</i>	1,1724	0,0001	0,1052	4,8343	-772426	-594193
	4	<i>44,075***</i>	0,7437	0,0001	<i>5,7293*</i>	1,1618	0,6538	0,0879	1,0050	0,0056	0,2103
	1	<i>14,695***</i>	1,2203	0,0001	<i>1,5785</i>	1,1700	0,0001	-0,1043	-4,3877	183189	577954
	4	<i>36,225***</i>	1,3117	0,0052	<i>1,4737</i>	1,1110	0,7167	0,6258	0,8868	-0,0136	0,0627
	1	<i>21,468***</i>	1,2457	0,0001	<i>1,1743</i>	1,1191	0,6060	2,2756	0,3867	-0,0682	0,0384
	4	<i>10,454</i>	0,9884	0,7369	<i>-4,1908</i>	1,0694	0,7231	1,9289	0,5995	-0,0178	0,0291
	1	<i>18,382***</i>	1,2474	0,0001	<i>0,9942</i>	1,1451	0,5069	2,9264	0,1614	-0,1467	0,0358
	4	<i>-2,0544</i>	0,9631	0,7269	<i>11,188***</i>	0,8543	0,0001	-0,0000	-4,2908	-4,43·10 <sup>13</sup>	-9,81·10 <sup>14</sup>
FCVAR ( $d = b$ restringiert)	1	46,500***	1,0157		2,3097	1,0267		0,2087	0,9639	0,0027	0,0280
	4	46,820***	1,0591		4,5785	1,1146		0,1981	0,9660	-0,0040	0,0233
	1	39,057***	1,0306		2,7472	1,0368		0,6709	0,8753	-0,0036	0,0126
	4	41,205***	1,0841		4,8438	1,1267		0,6356	0,8837	-0,0068	0,0110
	1	24,840**	1,0462		3,1177	1,0519		1,8494	0,6540	-0,0048	0,0055
	4	15,190***	0,4254		4,9322	1,1360		1,7900	0,6687	-0,0063	0,0052
	1	14,219	1,0446		2,8493	1,0504		3,0118	0,4393	-0,0051	0,0020
	4	5,8340	0,4211		4,5480	1,1200		2,9743	0,4502	-0,0058	0,0022
CVAR ( $d = b = 1$ restringiert)	1	43,041***			1,6660			0,2192	0,9611	0,0055	0,0287
	4	33,034***			1,7309			0,2176	0,9614	0,0046	0,0251
	1	35,876***			1,8911			0,7083	0,8649	-0,0014	0,0132
	4	28,241***			1,9489			0,6955	0,8682	-0,0013	0,0117
	1	20,230*			1,9357			1,9458	0,6286	-0,0032	0,0058
	4	17,497			1,9465			1,9087	0,6369	-0,0033	0,0052
	1	9,4663			1,7506			3,0758	0,4224	-0,0041	0,0018
	4	9,0561			1,6551			3,0424	0,4262	-0,0042	0,0015

Die kursiv geschriebenen LR-Teststatistiken sind nach Johansen und Nielsen (2012a) asymptotisch  $\chi^2$ -verteilt mit  $(p-r)^2$  Freiheitsgraden, da hier  $\hat{b} < 1/2$  geschätzt wurde. Dementsprechend wurden die Signifikanzentscheidungen anhand dieser Verteilungen getroffen. Für  $\hat{b} > 1/2$  wurden die kritische Werte von MacKinnon und Nielsen (2014) verwendet.

anhand der LR-Statistik bzgl. des Kointegrationsranges festhalten, dass wir in diesem Fall relativ häufig eine Kointegrationsbeziehung annehmen können. Lediglich für (Y5,M6) und (Y10,M6) lässt sich für  $k = 4$  keine Kointegrationsbeziehung annehmen. Für (Y10,M6) kommen wir sogar zu dem sonderbaren Ergebnis, dass die LR-Statistik für  $r = 0$  und für (Y5,M6) für  $r = 1$  negativ sind, was theoretisch ausgeschlossen ist. Dies lässt sich aber dadurch erklären, dass in diesen Situationen lediglich lokale Maxima der Likelihoodfunktion gefunden werden und uns wiederum vor Augen, wie kompliziert sich die Schätzung des FCVAR-Modells insbesondere mit unrestringierter Konstante gestaltet. Für die positiv getesteten Kointegrationsbeziehungen von (Y1,M6) und (Y2,M6) ergeben sich zudem für  $k = 1$  Kointegrationsstärken von nahe null, was eigentlich dafür sprechen würde, dass keine Kointegration zwischen den einzelnen Pärchen vorliegt. Dementsprechend erhalten wir für diese Situationen verwirrende Kointegrationsparameter. Hier ist eher angepasst, die Schätzergebnisse mit  $k = 4$  zu betrachten, wo wir immer noch eine Kointegrationsbeziehung zwischen (Y1,M6) und (Y2,M6) annehmen können. Zwar deutet für (Y1,M6) die LR-Statistik eine leicht signifikante Ablehnung für  $r = 1$  bei einer Irrtumswahrscheinlichkeit von 10% hin, aber der Kointegrationsbeziehungparameter  $\beta$  scheint sehr nahe am erwarteten Wert zu liegen. Für (Y2,M6) hingegen sehen wir diesbezüglich eine stärkere Abweichung, die wir im Folgenden auch noch auf Signifikanz testen werden. Etwas sonderbar scheint, dass die geschätzte Kointegrationsstärke  $b$  für (Y1,M6) kleiner ausfällt als für (Y2,M6), obwohl der Laufzeitunterschied wesentlich geringer ausfällt und die Abbildung 6.4 auf eine wesentlich trägere Anpassung der einzelnen Zinssätze schließt, sowie die geschätzten Integrationsparameter der Zinsdifferenzen mit dem Laufzeitunterschied tendenziell ansteigt. Für (Y5,M6) und (Y10,M6) sehen wir bei Wahl von  $k = 1$  hingegen eine abnehmende Kointegrationsstärke mit dem Laufzeitunterschied, was plausibler erscheint, und zudem noch eine stärkere Abweichung des Parameters  $\beta$  von eins.

Die Restriktion  $d = b$  liefert zumindest stabilere Ergebnisse, wenn auch wir für (Y10,M6) nun keine Kointegrationsbeziehung annehmen können. Vergleichen wir dies aber nun mit den Ergebnissen des klassischen Modells  $d = b = 1$ , so haben wir aber zumindest auch eine signifikante Annahme der Kointegration für den Datenvektor (Y5,M6). Die Frage, die sich nun stellt, lautet dann aber, warum sich dies gerade bei der Restriktion  $d = b$  ergibt, was ja wiederum impliziert, dass die Kointegrationsfehler integriert vom Grad null sind und wir vorangehend ein relativ hohes Ausmaß an Integration für die Zinsdifferenz von Y5 und M6 geschätzt haben, welches sogar im nicht-stationären Bereich lag. Wenn überhaupt, dann bietet sich

Tab. 6.12: LR-Teststatistiken für den Kointegrationsparameter  $\beta = 1$  für bivariate Kointegrationsbeziehungen

	FCVAR ( $d$ und $b$ unrest.)		FCVAR ( $d = b$ rest.)		CVAR ( $d = b = 1$ rest.)	
	$k = 1$	$k = 4$	$k = 1$	$k = 4$	$k = 1$	$k = 4$
(Y1,M6)	(-)	0,010	4,896**	4,138**	7,066***	3,666*
(Y2,M6)	0,426	1,774	8,958***	7,572***	9,718***	6,997***
(Y5,M6)	17,286***	6,824***	10,034***	9,942***	11,305***	8,901***
(Y10,M6)	14,276***	(-)	6,362**	6,920***	8,563***	7,735***

Die Angabe (-) bedeutet hier, dass eine negative LR-Statistik berechnet wurde, was aber theoretisch unmöglich ist.

allenfalls eine sehr starke Autokorrelation der Störterme an, die wir anscheinend nicht mit den semiparametrischen ELW-Schätzern vom Integrationsausmaß trennen konnten. Der Grund, warum wir für das klassische CVAR-Modell für (Y5,M6) keine Kointegrationsbeziehung annehmen konnten, ist hingegen schneller gefunden. Es liegt wohl im Integrationsgrad des stochastischen Trends, den wir für das mit  $d = b$  restringierte FCVAR-Modell leicht höher als eins schätzen konnten und somit uns diese Flexibilität eine Annahme der Kointegrationsbeziehung zwischen Y5 und M6 ermöglicht. Nichtsdestotrotz liegt der Parameter  $\beta$  relativ weit weg von eins, was trotzdem tendenziell gegen die Implikationen der Markterwartungshypothese spricht. Prinzipiell können wir bzgl.  $\beta$  aussagen, dass wir uns für zunehmenden Laufzeitunterschied innerhalb der jeweils untersuchten Datenvektoren immer weiter vom erwarteten  $\beta$  weg entfernen und dass dieser Unterschied für das CVAR-Modell immer größer als für das FCVAR-Modell unter der Restriktion  $d = b$  ist. Auch fällt auf, dass das CVAR-Modell immer leicht höhere Laufzeitprämien schätzt als das FCVAR-Modell. Die Vorzeichen des Parameters  $a_1$  sind auch zumeist plausibel bis auf für (Y1,M6). Dieser Parameter entspricht der Reaktion in der Folgeperiode auf einen Kointegrationsfehler für die erste Komponente des jeweils untersuchten Datenvektors und sollte in den vorliegenden Fällen negativ sein, damit auch wirklich eine Anpassung zum Gleichgewicht, welches Kointegration beschreibt, eintritt. Für das FCVAR-Modell ist der Parameter hingegen schwieriger zu interpretieren, da es nicht unbedingt die Reaktion genau in der Folgeperiode beschreibt und auch der Kointegrationsstärkeparameter  $b$  zu berücksichtigen gilt. Prinzipiell lassen aber die negativen Vorzeichen trotzdem darauf schließen, dass gerade die längerlaufenden Anleiheverzinsungen wieder zum Gleichgewicht zurück tendieren.

Nun wollen wir noch die Testergebnisse der Nullhypothese  $\beta = 1$  für das (F)CVAR-Modell betrachten. Hierzu können wir in den jeweiligen Modellrahmen LR-Statistiken auf Basis der Restriktion  $\beta = 1$  bilden, die im vorliegenden Fall asymptotisch  $\chi^2_1$ -

Tab. 6.13: LR-Teststatistiken für den Kointegrationsrang  $r$  für das alle Renditenreihen umfassende Zeitreihensystem (Y1,Y2,Y5,Y10,M6) im (F)CVAR-Kontext

	$k$	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
FCVAR ( $d$ und $b$ unrestringiert)	1	217,76*** (0,000)	108,56*** (0,000)	64,214*** (0,546)	19,142*** (0,586)	0,7665 (0,594)
	4	70,445*** (0,282)	21,667 (0,185)	13,496 (0,342)	6,8021 (0,441)	0,6146 (0,504)
FCVAR ( $d = b$ restringiert)	1	205,87*** (0,477)	106,07*** (0,467)	44,869*** (0,941)	4,5995 (0,945)	1,5477 (0,948)
	4	67,797*** (0,282)	19,018 (0,185)	21,344* (0,584)	6,9280 (0,600)	0,0553 (0,602)
CVAR ( $d = b = 1$ restringiert)	1	424,90***	176,94***	53,257***	8,3856	2,1416
	4	266,45***	128,29***	40,913**	8,4857	2,5316

Die jeweils geschätzten Kointegrationsstärkeparameter  $b$  sind in Klammern angegeben, da diese die Verteilung der LR-Statistiken definieren. Die kursiv geschriebenen LR-Teststatistiken sind nach Johansen und Nielsen (2012a) asymptotisch  $\chi^2$ -verteilt mit  $(p-r)^2$  Freiheitsgraden, da hier  $\hat{b} < 1/2$  geschätzt wurde. Dementsprechend wurden die Signifikanzentscheidungen anhand dieser Verteilungen getroffen. Für  $\hat{b} > 1/2$  wurden die kritischen Werte von MacKinnon und Nielsen (2014) verwendet.

verteilt sind, da im bivariaten Fall lediglich der Parameter  $\beta$  restringiert wird. Die Ergebnisse sind in Tabelle 6.12 dargestellt und lassen den Schluss zu, dass wir für das restringierte FCVAR- und das CVAR-Modell für alle Datenvektoren  $\beta = 1$  signifikant ablehnen können. Lediglich für das unrestringierte FCVAR-Modell ergibt sich für die Datenvektoren mit kürzeren Laufzeitdifferenzen keine signifikante Ablehnung für  $\beta = 1$ . Dies alles überrascht etwas, da die einzelnen geschätzten Werte für  $\beta$  zumindest für (Y1,M6) recht nahe an eins liegen. Letztendlich muss man festhalten, dass das unrestringierte FCVAR-Modell noch am ehesten in der Lage ist, die Implikationen der Markterwartungshypothese zu bestätigen. Ein Grund liegt darin, dass wir doch einige Evidenz haben, dass die einzelnen Anpassungsprozesse nicht integriert vom Grad null sind und sogar nicht-stationär scheinen, was aber bei Anwendung des restringierten FCVAR- und CVAR-Modell nicht berücksichtigt werden kann. Dem entgegen steht die recht komplizierte Anwendung des unrestringierten FCVAR-Modells, was zum Teil auch unplausible Ergebnisse liefert. Ein weiteres Problem in der Anwendung des FCVAR-Modells liegt in der Identifikation der einzelnen Modelle, wovon wiederum die Schätzergebnisse abhängen. Weiterhin muss man die hier vorliegende Anwendung kritisieren, dass gewisse Charakteristika, die vor allem auf Tagesdaten zutreffen, wie Heteroskedastizität und eine weitere Analyse von nicht-linearen Prozessbestandteilen nahezu unberücksichtigt bleiben. Insbesondere für das CVAR-Modell wurden auch weitere deterministische Modellierungsmöglichkeiten wie eine nicht-restringierte Konstante und lineare Trends überprüft, wobei es aber durchweg zu den gleichen Ergebnissen kam, wie in den vorangegangenen



In Tabelle 6.14 haben wir noch die geschätzten Kointegrationsmatrizen  $\hat{\mathbf{A}}$  und  $\hat{\mathbf{B}}$  für  $k = 1$  und  $r = 4$ , verzichteten aber auf eine Darstellung für den restringierten FCVAR-Fall. Beiden Modellvarianten ist gemein, dass der geschätzte Kointegrationsraum  $\hat{\mathbf{B}}$  erheblich vom implizierten  $\mathbf{B} = (\mathbf{I}_4, -\boldsymbol{\iota})$  abweicht. Für beide Modelle ist prinzipiell eine Angabe einer LR-Statistik bzgl. des Kointegrationsraumes möglich, aber lediglich für das CVAR-Modell kann eine LR-Statistik mit 52,098 berechnet werden, da für das FCVAR-Modell unter der Restriktion  $\mathbf{B} = (\mathbf{I}_4, -\boldsymbol{\iota})$  kein Schätzergebnis aufgrund fehlender Konvergenz möglich war. Aufgrund der  $\chi_4^2$ -Verteilung der LR-Statistik bzgl.  $\hat{\mathbf{B}}$ , da hier insgesamt vier Komponenten restringiert sind, ergibt sich eine hoch signifikante Ablehnung der Markterwartungshypothese, die sich auch bei Betrachtung höherer Lag-Zahlen für  $k$  ergibt. Diese Ergebnisse sowie die Testresultate bzgl. des Kointegrationsranges können aber eine Überraschung darstellen, da sie sich bereits aus der bivariaten Betrachtung der Zinszeitreihen ergeben. Auch hier waren wir nur für das unrestringierte FCVAR-Modell in Teilbereichen in der Lage, die Implikationen der Markterwartungshypothese zu bestätigen. Bei Anwendung des FCVAR-Modell auf das volle System (Y1,Y2,Y5,Y10,M6) machen wir aber vermutlich noch zusätzlich einen Fehler, da wir für sämtliche Kointegrationsbeziehungen die gleiche Kointegrationsstärke  $b$  annehmen. Dies wiederum konnten wir aus den bivariaten Analysen nicht bestätigen und somit müssen wir weiterhin annehmen, dass das FCVAR-Modell für das volle System tendenziell fehlspezifiziert ist.

Zusammenfassend lässt sich aussagen, dass unter Verwendung des vorliegenden Datensatzes wir eher zu einer Ablehnung der Markterwartungshypothese kommen, da nicht für sämtliche Zinspaare eine Kointegrationsbeziehung angenommen werden kann und da die geschätzten Kointegrationsräume von den erwarteten signifikant abweichen. Dies wiederum scheint verwunderlich, da die Abbildung 6.1 einen Gleichlauf der Zinssätze suggeriert. Vielmehr scheinen die Gründe für die Ablehnung darin zu liegen, dass womöglich die Kointegrationsparameter wie die Raumkoordinaten und die Stärke evtl. nicht konstant über den Beobachtungszeitraum zu sein scheinen, was aber für diese Parameter in dieser Arbeit nicht kontrolliert wurde. Auch müssen die Annahmen über die Eigenschaften der Risikoprämie, die Investoren für länger laufende Schuldverschreibungen verlangen, überdacht werden. HAG (1992) hatten hierzu angenommen, dass unter deren Stationarität es zu einer Kointegration zwischen den Zinssätzen kommen muss. Von dieser Sichtweise haben wir uns zwar in dieser Arbeit getrennt, haben aber auch gesehen, dass das Ausmaß der Integration der Zinsdifferenzen mit der Laufzeitdifferenz anwächst und stets im nicht-stationären

Bereich zu liegen scheint. Dies legt die Vermutung nahe, dass die Risikoprämie wohl für lang laufende Zinsdifferenzen einen Integrationsgrad aufweist, der evtl. kaum noch kleiner als der Integrationsgrad des stochastischen Trends ist. In diesem Kontext stellt aber die fraktionale Kointegration keinen geeigneten Modellrahmen mehr dar, sondern man muss auf ein fraktional integriertes VAR-Modell ausweichen, wie es z.B. Tschernig, Weber und Weigand (2013) vorgeschlagen haben. Dieses wurde auf Basis des FCVAR-Modells weiterentwickelt und umfasst neben der Modellierung von fraktionalen Kointegrationsbeziehungen auch die Möglichkeit, strukturelle Beziehungen von fraktional integrierten Zeitreihen zu umfassen.

## 7 Schlussfolgerungen und Ausblick

Das Ziel dieser Arbeit war, einen aktuellen Überblick über den Stand der Forschung über fraktionale Integration und Kointegration zu geben. Neben dem einführenden zweiten Kapitel haben wir im dritten Kapitel im Bereich der parametrischen Schätzansätze eine Reihe von Simulationen durchgeführt, die uns gezeigt haben, dass die Whittle- und die HR-Approximation die besten Approximationsweisen zur Schätzung von ARFIMA( $p, d, q$ )-Prozessen darstellt. Teilweise konnten wir für beide Schätzer über eine bessere Verzerrung als die exakte ML-Berechnungsweise berichten bei nur geringfügig höherer Volatilität. Beide Approximationen weisen ungefähr gleiche Schätzeigenschaften auf, wobei wir mit dem Whittle-Schätzer häufiger einen höheren RMSE gemessen haben. Für weitere Studien im Bereich der parametrischen Modellierung haben wir uns vorzugsweise auf diese beiden Berechnungsweisen gestützt, da sie sehr viel weniger rechenintensiv sind als die exakte Schätzmethode nach Sowell (1992).

Der Problematik der Modellwahl, mit der wir auch im empirischen Teil der Arbeit konfrontiert waren, sind wir mit der Anwendung von Informationskriterien begegnet. Die Wichtigkeit zeigt sich bei ARFIMA-Prozessen darin, dass ein überspezifiziertes Modell zu einem unterschätzten Integrationsgrad  $d$  führt, da zusätzlich zur Schätzung einbezogene AR- und MA-Bestandteile das Vorhandensein der impliziten unendlichen AR- und MA-Darstellung der fraktionalen Integration auffängt und damit das Ausmaß der Persistenz vorweg nimmt. Umgekehrt zeigt sich, dass eine Unterspezifikation zu einem überschätzten  $d$  führt. Aus diesem Grund haben wir die Eigenschaften der einzelnen Informationskriterien im Zusammenhang mit dem Whittle-Schätzer in Monte-Carlo-Simulationen untersucht. Hierzu existieren zwar schon teilweise Untersuchungen, die aber entweder einen anderen Fokus oder einen geringeren Umfang hatten. Prinzipiell konnten wir hier bestätigen, dass die Anwendung des Schwarz-Informationskriteriums die besten Ergebnisse erwarten lässt und die wahre Modellierung mit einer relativ großen Sicherheit treffen kann. Die Wahrscheinlichkeit, das wahre Modell zu treffen, ist bei Verwendung der Whittle-Approximation etwas höher als bei der HR-Approximation, wobei wir unter Verwendung der Kriterien in Kombination mit Modellunsicherheit bei der HR-Approximation immer noch eine leicht niedrigere Streuung haben als bei Verwendung des parametrischen Whittle-Schätzers.

Daraufhin haben wir uns umfangreich mit semiparametrischen Schätzansätzen beschäftigt. Diese sind in der Form des GPH- und des LW-Schätzers für stationäre

Zeitreihen ausgelegt und hängen in ihren Eigenschaften stark davon ab, welche Bandbreite, die die Anzahl der berücksichtigten Frequenzen zur Anpassung des Spektrums an das Periodogramm darstellt, gewählt wird. In diesem Kontext haben wir festgestellt, dass der lokale Whittle-Schätzer eine etwas schlechtere asymptotische Konvergenz als der GPH-Schätzer aufweist, aber gleichzeitig effizienter ist. Der exakte lokale Whittle-Schätzer, der auch für nicht-stationäre integrierte Prozesse anwendbar ist, hat hingegen eine nahezu konstante leichte Überschätzung des Integrationsgrads als Eigenschaft.

Ähnlich wie bei den parametrischen Methoden haben wir uns mit der Wahl der Schätzparameter auseinandergesetzt, die im Falle der semiparametrischen Schätzer die Bandbreite darstellt. Bei stark ausgeprägter Autokorrelation und relativ hoher Bandbreite brauchen wir tendenziell eine hohe Beobachtungsanzahl, um eine Konvergenz zum wahren Integrationsgrad zu erleichtern. Aus diesem Grund betrachten wir optimale Bandbreiten, die wir anhand eines minimalen RMSE definieren. Da für den ELW-Schätzer bisher keine Untersuchungen für optimale Bandbreiten existieren, haben wir hierzu auch Simulationen durchgeführt und dabei festgestellt, dass die optimalen Bandbreiten in Zusammenhang mit ARFIMA(1,  $d$ , 0)-Prozessen sehr den LW-Bandbreiten ähneln.

Diese Erkenntnis haben wir wiederum für Schätzansätze von Bandbreiten verwendet, wobei wir mit dem Plugin-Verfahren ein bereits existierendes semiparametrisches Verfahren untersucht haben. Dieses haben wir auch versucht, für den ELW-Schätzer in einer iterativen Fassung umzusetzen und es in Simulationen mit einem parametrischen Ansatz verglichen. Dieser lässt sich aus der Abhängigkeit der optimalen Bandbreite von der Parameterstruktur motivieren und verwendet die geschätzten Parameter einer ARFIMA-Ordnung, um daraus eine optimale Bandbreite zu ermitteln. Das größte Problem des Ansatzes ist sicherlich, dass hierzu eine bestimmte ARFIMA-Ordnung angenommen werden muss, um die einzelnen Parameter zu schätzen. Dies lässt annehmen, dass, wenn eine falsche Ordnung gewählt wird, dies auch in einer falschen bzw. schlechten Bandbreite resultiert. Diesen Effekt haben wir in dieser Arbeit nicht näher untersucht, da dieses Verfahren bereits schon nicht gut abschneidet, wenn wir die wahre Ordnung kennen. Insbesondere wenn wir eine stark ausgeprägte Autokorrelation in simulierten Prozessen aufweisen, haben wir mit diesem Verfahren eine stark überschätzte Bandbreite und damit auch einen überschätzten Integrationsgrad. Das semiparametrische Plugin-Verfahren kann zwar den Verlauf der optimalen Bandbreite auch nicht gut treffen, schätzt aber zumindest eine niedrigere Bandbreite, falls das Ausmaß der Autokorrelation steigt, was die Gefahr

einer Überschätzung des Integrationsgrads eindämmt. So verwundert es kaum, dass das Plugin-Verfahren wesentlich bessere Schätzeigenschaften als das parametrische Verfahren bzgl. des Integrationsgrads aufweist und somit diesem vorzuziehen ist. Auch im Vergleich zur Schätzung mit fixierten Bandbreiten konnten wir feststellen, dass das Plugin-Verfahren nie von beiden fixierten Bandbreiten dominiert wurde, welche in unserer Simulation eine relativ geringe und eine relativ große waren. Somit stellt das Plugin-Verfahren durchaus eine gute Schätzmöglichkeit der Bandbreite dar, wenn Unsicherheit über das Ausmaß von Short-Memory-Effekten besteht.

Im Bereich der Spezifikationstests haben wir eine Verallgemeinerung des Testverfahrens von Gómez und Ventosa-Santaulària (2010) vorgeschlagen, das auch einen Integrationsgrad  $d \neq 1$  berücksichtigt. Somit haben wir auch die Möglichkeit, einen deterministischen Trend in der nicht-stationären Long-Memory-Modellierung zu testen, sowie auch Trendbrüche, was die Bandbreite an Modellierungsmöglichkeiten erweitert. Auch haben wir die Notwendigkeit der korrekten Spezifikation der deterministischen Prozessbestandteile mit einer Monte-Carlo-Simulation hervorgehoben, da eine Nicht-Berücksichtigung von Trendbrüchen z.B. zu einer scheinbaren Annahme von Long-Memory-Eigenschaften führt, da der Integrationsgrad zum Teil stark überschätzt wird. Diesen Ansatz konnten wir im empirischen Teil dieser Arbeit nicht direkt umsetzen, da wir keine Evidenz für deterministische Bestandteile wie Trends und gebrochene Trends in der Verzinsung deutscher Bundesanleihen gefunden haben. Ein besseres Anwendungsgebiet stellt wohl die Beschreibung von BIP-Daten bzw. Preisindizes dar, was wir aber in dieser Arbeit nicht versucht haben. Den Ansatz haben wir in Bezug auf Trendbrüche mit dem bereits von Shao (2011) vorgeschlagenen Meanchange-Test verglichen, der für stationäre Long-Memory-Prozesse entwickelt wurde, und konnten ihm nahezu durchweg eine bessere Macht attestieren.

Bei den Integrationstests haben wir uns vor allem auf Tests der Hypothese  $d = 1$  gegenüber fraktional integrierten Alternativen konzentriert und eine Reihe von bereits existierenden Testansätzen miteinander verglichen. Zusätzlich haben wir für eine ganze Reihe von Trendoptionen Verfahren zur Ermittlung von kritischen Werte aufgezeigt und somit die Anwendung der Ansätze in der Empirie erleichtert. In vergleichenden Monte-Carlo-Studien konnten wir erkennen, dass vor allem der asymptotische effiziente Wald-Test eine gute Macht hat und haben diesen auch dann für das Kapitel der fraktionalen Kointegration als Integrationstest für geschätzte Kointegrationsresiduen diskutiert.

Im Kapitel über Kointegration haben wir zunächst sequentielle Verfahren zur Bestimmung von Kointegrationsbeziehungen in Anlehnung an die Arbeit von Engle und Granger (1986) betrachtet, die darauf aufbauen, dass wir für geschätzte Residuen aus Regression einen kleineren Integrationsgrad testen als für die in die Regression eingehenden Variablen. Für diesen Zweck haben wir mit dem NBL-Schätzer eine semiparametrische Methode für Kointegrationsräume betrachtet, die insbesondere beim Vorliegen von fraktionalen Kointegrationsbeziehungen und Endogenitätsproblemen vorgeschlagen werden. Die Verwendung des NBL-Schätzers hat im Vergleich zum einfachen OLS tendenziell bei einer Korrelation des Störterms des stochastischen Trends zum Kointegrationsstörterm leicht bessere Schätzeigenschaften, welche aber vergleichsweise gering ausfallen und dennoch eine Verzerrung bei Endogenität aufweisen, die größer wird, je kleiner die Kointegrationsstärke  $b$  ist. Zudem haben wir uns in diesem Fall vor allem darauf konzentriert, dass wir zu meist stochastische Trends beobachten, die integriert vom Grad eins sind und haben aus diesem Grund genau dann eine fraktionale Kointegrationsbeziehung, wenn die Kointegrationsstörtermen einen Integrationsgrad  $d < 1$  aufweisen. Diese Sichtweise wurde auch im sechsten Kapitel gestützt, da die Verzinsungen tendenziell einen Integrationsgrad von eins aufweisen bzw. knapp darüber liegen. In diesem Hinblick haben wir auch in Simulationen gesehen, dass eine Anpassung der kritischen Werte des effizienten Wald-Tests aus dem vierten Kapitel nötig ist im Gegensatz zum univariaten ELW-Schätzer, der robust gegenüber geschätzten Kointegrationsresiduen scheint. Beim univariaten ELW-Schätzer liegt für die Schätzung des Integrationsgrads von geschätzten Störtermen das Problem in der geeigneten Wahl der Bandbreite. Hier haben wir in der vergleichenden Monte-Carlo-Studie gesehen, dass eine niedrig gewählte Bandbreite zu einer häufigen Annahme der Kointegrationshypothese vorkommt, obwohl keine Beziehung simuliert wurde. Dies überrascht ein wenig, da wir im Zusammenhang mit dem ELW-Schätzer im dritten Kapitel gesehen haben, dass dieser Schätzansatz den Integrationsgrad eher überschätzt. Selbst wenn im Fall ohne zusätzliche Short-Memory-Eigenschaften der Störtermen aber eine gute Bandbreite gewählt wird, so reicht die Macht des univariaten ELW-Ansatzes nicht an die des effizienten Wald-Tests heran. Nur wenn weitere AR-Bestandteile im Kointegrationsstörterm spezifiziert werden, so ist die Macht gerade mal auf einem ähnlichen Niveau, aber nur, wenn eine niedrige Bandbreite gewählt wird. Eine zu hohe Bandbreite lässt hingegen Beziehungen übermäßig häufig ablehnen, auch wenn Kointegrationsbeziehungen simuliert wurden. Aus diesem Grund können wir die alleinige Verwendung des ELW-Ansatzes zur Bestimmung von fraktionalen Kointegrations-

beziehungen nicht empfehlen, sondern raten zu einer zusätzlichen Verwendung von Integrationstests.

Neben sequentiellen Verfahren betrachten wir mit dem FCVAR-Ansatz und dem multivariaten ELW-Schätzer Verfahren, die eine integrierte Schätzung von  $d, b$  und des Kointegrationsraumes ermöglicht, die zudem noch die Situation  $d \neq 1$  berücksichtigen können. Beide Ansätze zeigten sich in ihrer Handhabung als nicht unproblematisch, da wir im Zusammenhang mit dem multivariaten ELW-Schätzer wieder vor dem Problem der Wahl der Bandbreite und der Modellordnung für den FCVAR-Ansatz stehen. Des Weiteren weist der ELW-Schätzer bei niedrigen Kointegrationsstärken bzw. nicht-stationären Kointegrationsfehlern schwache Schätzeigenschaften auf und sollte nur in Kombination mit anderen Kointegrationsmethoden verwendet werden. Beim FCVAR-Modell liegt das Problem in der Parameteridentifikation im Schätzalgorithmus. Nichtsdestotrotz weist der FCVAR-Ansatz eine Reihe von Vorzügen wie etwa ein Testansatz für die Anzahl der Kointegrationsbeziehungen und den Kointegrationsraum auf, die ihn für empirische Anwendungen reizvoll machen.

Im sechsten Kapitel der Arbeit haben wir im Rahmen der Markterwartungshypothese die Verzinsung deutscher Bundesanleihen auf fraktionale Kointegrationsbeziehungen untersucht. Den Integrationsgrad konnten wir für die einzelnen Zinsreihen auf nahe eins schätzen, ohne diesen Grad für die meisten Reihen abzulehnen. Für die kürzerlaufenden Zinsen unter einem Jahr haben wir hingegen einen signifikant größeren Integrationsgrad als eins geschätzt und damit eine stärker ausgeprägte Persistenz als für längerlaufende Verzinsungen. Dies haben wir durch eine stärkere Abhängigkeit zu Zinsentscheidungen der Zentralbank wegen der kürzeren Laufzeit erklärt. Die Testergebnisse auf Vorliegen von deterministischen Prozessbestandteilen brachten keinerlei Anhaltspunkte, dass diese Komponenten in den Prozessen vorliegen und haben deswegen auch keine Trendoptionen außer einer restringierten Konstanten in der multivariaten Analyse berücksichtigt. Diese wiederum scheint auf Basis der theoretischen Betrachtung der Markterwartungshypothese geboten, da sie als eine Laufzeitprämie interpretiert werden kann. Eine Betrachtung der Zinssatzdifferenzen ergab bereits Evidenz, dass mit steigender Laufzeitdifferenz der Zinsen auch eine stärker werdende Persistenz einhergeht. Die einzelnen Differenzen können allesamt als nicht-stationär charakterisiert werden, wobei wir für die kürzeren Differenzen eine Mean-Reversion-Eigenschaft feststellen konnten, da die Integrationsgrade signifikant unter eins lagen. Diese konnten wir für die längeren Differenzen nicht mehr bestätigen, was uns schon ein Indiz für keine vorliegende Kointegrationsbeziehung

gab. Die Anwendung der Integrationstests in Form des effizienten Wald-Tests bestätigte diese Betrachtungsweise, wobei aber der VR-Test eine signifikante Ablehnung von  $d \geq 1$  signalisierte. Die fraktionale Kointegrationsanalyse folgte dann unter der Anwendung des multivariaten ELW-Schätzers für bivariate Zinspaare und für höhere Dimensionen im Rahmen des FCVAR-Modells. Hier konnten wir nur schwache Kointegrationsbeziehungen für die kurzen Laufzeitunterschiede ausmachen, während die längeren Differenzen keine Kointegration aufwiesen. Die Wahl der Bandbreite für den ELW-Schätzer gestaltete sich schwierig, da die univariate Schätzung des Integrationsgrads der Zinsdifferenzen auf eine stark autokorrelierte Struktur hinwies. Die Betrachtung der fraktional differenzierten Kointegrationsresiduen ergaben aber keine positive Autokorrelation, sondern wiesen einen negativen Autokorrelationskoeffizienten auf, der umso stärker wurde, je größer die Bandbreite gewählt wurde. Dies deutet tendenziell auf eine Überschätzung des Integrationsgrads hin und ließ somit auch für die längeren Fristunterschiede keine endgültige Ablehnung der Kointegrationshypothese zu. Letztendlich wollten wir uns im Rahmen des FCVAR-Modells weitere Evidenz für Kointegration schaffen und kamen hierbei zum Ergebnis, dass für niedrig gewählte Lagzahlen, die auch vom BIC als Informationskriterium selektiert wurden, durchaus für jede der Zinspaare Kointegrationsbeziehung im unrestringierten Fall vorlägen. Dies scheint aber bei Betrachtung von höheren Lags keinesfalls robust und auch die einzelnen Schätzergebnisse bzgl. des Kointegrationsraums und Kointegrationsstärke sind nicht durchweg plausibel. Auch die Hypothesentests des theoretischen Kointegrationsraums neigen eher zur einen allgemeinen Ablehnung der Markterwartungshypothese. Eine Restriktion der Integrationsparameter führt zwar zu plausibleren Schätzungen des Kointegrationsraums und zu robusteren Ergebnissen, es können aber weniger Kointegrationsbeziehungen gefunden werden und die Kointegrationsräume sind alle signifikant von den theoretischen verschieden.

Im Rahmen dieser Arbeit wurden eine Reihe von Modellierungsmöglichkeiten nicht berücksichtigt, die aber in Zukunft ein interessantes Themengebiet für weitere Forschung darstellen können. Dazu könnte man die Weiterentwicklung von heteroskedastischen Long-Memory-Modellen vorantreiben, worauf schon in Tschernig (1994) verwiesen wurde. Hier wurden u.a. tägliche Wechselkursdaten analysiert, die ähnlich wie die Verzinsungen in dieser Arbeit anscheinend eine Fünf-Tage-Periodizität aufweisen. Hauser und Kunst (1998) schlagen deswegen eine simultane exakte ML-Schätzung von ARFIMA- und ARCH-Komponenten in einem einheitlichen Ansatz vor, der genau solche Tageseffekte berücksichtigen könnte. Diese Schätzung kann-

te sich aber in der empirischen Modellierung kaum durchsetzen und ist deswegen in keinen ökonometrischen Softwarepaketen implementiert. Stattdessen haben sich eher spezielle ARCH-Modelle entwickelt, die eine Long-Memory-Eigenschaft der Volatilität berücksichtigen. Hierzu stellten Baillie, Bollerslev und Mikkelsen (1996) die Grundlage mit fraktional integrierten GARCH-Modellen (auch abgekürzt als FI-GARCH), die die Grundlage für eine Reihe von Weiterentwicklungen darstellten, die bis heute noch nicht abgeschlossen scheinen.

Ein weiterer Ansatzpunkt könnte die Erweiterung des Tests von Gómez und Ventosa-Santaulària (2010) auf mehr als einen Strukturbruch darstellen, was aber im Zusammenhang mit der Modellierung der Zinssätze, die wir in dieser Arbeit angestrengt haben, wenig bringen dürfte. Eine Umsetzung der Schätzalgorithmen könnte auf Basis der Arbeit von Bai und Perron (1998) erfolgen und wäre somit machbar. Schwieriger erscheint hierin aber die Gestaltung der Tests, die auf einer gemeinsamen Basis erfolgen müssten. Auch liegt ein großes Problem in der Identifikation der Strukturbruchanzahl. Hierzu existieren zwar schon Arbeiten im Bereich der klassisch integrierten Prozesse wie Lumsdaine und Papell (1997), aber die Übersetzung für fraktional integrierte Zeitreihen ist komplexer und bedarf einer sorgfältigen Analyse, da evtl. ein größeres Ausmaß an Persistenz, wie es bei Prozessen mit hohem Integrationsgrad der Fall ist, womöglich scheinbare Strukturbrüche signalisiert.

Neben den Strukturbrüchen der deterministischen Komponente von Zeitreihen ist der Ansatz von Kruse und Sibbertsen (2009) vielversprechend, da hier Strukturbrüche im Integrationsparameter  $d$  untersucht werden und auch deren Bruchzeitpunkte geschätzt werden können. Dies könnte z.B. in Zusammenhang mit der Stabilität der Kointegrationsstärke  $b$  erweitert und einen weiteren flexiblen Modellierungsansatz liefern. Gerade die Betrachtung der Zinssätze im sechsten Kapitel könnten evtl. einen konstanten Parameter  $b$  in Frage stellen und die schwierige Schätzung des Parameters über die Zeit hinweg erklären.

Weitere Integrationstests der Einheitswurzelhypothese sind eher nicht nötig, worauf schon Maddala und Kim (1998) hingewiesen haben. In den dreißig Jahren nach der Entwicklung des Dickey-Fuller-Tests sind eine schier überwältigende Anzahl von Tests vorgeschlagen worden, die kaum noch zu überblicken ist. Hier sollte eher der Fokus darauf gerichtet werden, dass bestehende Tests robust gemacht werden sollten wie z.B. gegenüber Heteroskedastizität oder nicht-linearen Trends, wie wir es auch in dieser Arbeit mit Strukturbrüchen linearer Trends umzusetzen versucht haben.

## Literatur

- Adenstedt, R. K. (1974): *On large-sample estimation of the mean of a stationary random sequence*. *Annals of Statistics* **2**, S. 1095–1107.
- Agiakloglou, C. und Newbold, P. (1994): *Lagrange multiplier tests for fractional difference*. *Journal of Time Series Analysis* **15**, S. 253–262.
- Akaike, H. (1974): *A new look at the statistical model identification*. *IEEE Transactions on Automatic Control* **19**, S. 716–723.
- Andrews, D. und Sun, Y. (2004): *Adaptive local polynomial Whittle estimation of long-range dependence*. *Econometrica* **72**, S. 569–614.
- Bai, J. und Perron, P. (1998): *Estimating and testing linear models with multiple structural changes*. *Econometrica* **66**, S. 47–78.
- Baillie, R. T. und Bollerslev, T. (1994): *Cointegration, fractional cointegration, and exchange rate dynamics*. *Journal of Finance* **49**, S. 737–745.
- Baillie, R. T. (1996): *Long memory processes and fractional integration in econometrics*. *Journal of Econometrics* **73**, S. 5–59.
- Baillie, R. T., Bollerslev, T. und Mikkelsen, H. O. (1996): *Fractionally integrated generalized autoregressive conditional heteroskedasticity*. *Journal of Econometrics* **74**, S. 3–30.
- Box, G. E. P. und Jenkins, G. (1976): *Time series analysis: Forecasting and control*. Holden-Day, San Francisco.
- Breidt, F. J., Crato, N. und de Lima, P. (1998): *The detection and estimation of long memory in stochastic volatility*. *Journal of Econometrics* **83**, S. 325–348.
- Breitung, J. (2002): *Nonparametric tests for the unit roots and cointegration*. *Journal of Econometrics* **108**, S. 342–363.
- Breitung, J. und Hassler U. (2002): *Inference on the cointegration rank in fractionally integrated processes*. *Journal of Econometrics* **110**, S. 167–185.
- Brockwell, P. J. und Davis, A. D. (2002): *Time series: theory and methods*. Springer Verlag, New York - Berlin - Heidelberg.
- Campbell, J. Y. und Shiller, R. J. (1987): *Cointegration and tests of Present Value Models*. *Journal of Political Economy* **95**, S. 1062–1088.
- Caporale, G. M. und Gil-Alana, L. A. (2012): *Fractional cointegration in US term spreads*. *Economics Letters* **19**, S. 431–434.
- Chen, W. W. und Hurvich, C. M. (2003): *Semiparametric estimation of multivariate fractional cointegration*. *Journal of the American Statistical Association* **98**, S. 629–642.

- Christensen, B. J. und Nielsen, M. Ø. (2006): *Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting*. Journal of Econometrics **133**, S. 343–371.
- Cheung, Y.-W. und Lai, K. S. (1993): *A fractional cointegration analysis of purchasing power parity*. Journal of Business and Economic Statistics **11**, S. 103–112.
- Crato, N. und Ray, B. K. (1996): *Model selection and forecasting for long-range dependent processes*. Journal of Forecasting **15**, S. 107–125.
- Davidson, J. und Sibbertsen, P. (2009): *Tests of bias in log-periodogram regression*. Economics Letters **102**, S. 83–86.
- Dechert, A. (2014): *Variance ratio testing for fractional cointegration in presence of trends and trend breaks*. Proceedings of the International Work-Conference on Time Series **2**, S. 1055–1068.
- Dechert, A. (2014): *Fraktionale Kointegrationsbeziehungen zwischen Euribor-Zinssätzen*. Würzburg Economic Papers **93**, S. 1–23.
- Demetrescu, M., Kuzin, V. und Hassler, U. (2008): *Long memory testing in the time domain*. Econometric Theory **24**, S. 176–215.
- Dickey, D. A. und Fuller, W. A. (1979): *Distribution of estimators of autoregressive time series with a unit root*. Journal of the American Statistical Association **74**, S. 427–431.
- Dickey, D. A. und Fuller, W. A. (1981): *Likelihood ration tests for autoregressive time series with a unit root*. Econometrica **49**, S. 1057–1072.
- Diebold, F. X. und Rudebusch, G. D. (1991): *On the power of the Dickey-Fuller tests against fractional alternatives*. Economic Letters **35**, S. 155–160.
- Ding, Z., Granger, C. W. J. und Engle, R. F. (1993): *A long memory property of stock market returns and a new model*. Empirical Finance **1**, S. 83–106.
- Dittmann, I. (2000): *Residual-based tests for fractional cointegration: A Monte Carlo study*. Journal of Time Series Analysis **21**, S. 615–647.
- Dittmann, I. (2001): *Fractional cointegration of voting and non-voting shares*. Applied Financial Analysis **11**, S. 321–332.
- Dittmann, I. (2004): *Error correction models for fractionally cointegrated time series*. Journal of Time Series Analysis **25**, S. 27–32.
- Dolado, J. J., Gonzalo, J. und Mayoral, L. (2002): *A fractional Dickey-Fuller test for unit roots*. Econometrica **70**, S. 1963–2206.

- Doornik, J. A. (1998): *Approximations to the asymptotic distribution of cointegration tests*. Journal of Economic Surveys **12**, S. 573–593.
- Dueker, M. und Startz, R. (1998): *Maximum-likelihood estimation Of fractional cointegration with an application to U.S. and Canadian bond rates*. The Review of Economics and Statistics **80**, S. 420–426.
- Engle, R. F. und Granger, C. W. J. (1987): *Co-Integration and error correction: Representation, estimation and testing*. Econometrica **55**, S. 251–276.
- Engsted, T. und Tanggaard, C. (1994): *Cointegration and the US term structure*. Journal of Banking and Finance **18**, S. 167–181.
- Frederiksen, P., Nielsen, F. S. und Nielsen, M. Ø. (2012): *Local polynomial Whittle estimation of perturbed fractional processes*. Journal of Econometrics **167**, S. 426–447.
- Fuller, W. A. (1976): *Introduction to statistical time series*. Wiley, New York.
- Geweke, J. und Porter-Hudak, S. (1983): *The estimation and application of long memory time series models*. Journal of Time Series Analysis **4**, S. 221–238.
- Gil-Alana, L. A. (2003): *Testing of fractional cointegration in macroeconomic time series*. Oxford Bulletin of Economics and Statistics **65**, S. 517–529.
- Giraitis, L. und Surgailis, D. (1990): *A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate*. Probability Theory and Related Fields **86**, S. 87–104.
- Gómez, M. und Ventosa-Santaulària, D. (2011): *Testing for a deterministic trend when there is evidence of unit root*. Journal of Time Series Econometrics **2**, S. 1–26.
- Granger, C. W. J. und Newbold, P. (1974): *Spurious regressions in econometrics*. Journal of Econometrics **2**, S. 111–120.
- Granger, C. W. J. und Joyeux, R. (1980): *An introduction to long-memory time series and fractional differencing*. Journal of Time Series Analysis **1**, S. 15–29.
- Granger, C. W. J. (1980): *Long memory relationships and the aggregation of dynamic models*. Journal of Econometrics **14**, S. 227–238.
- Granger, C. W. J. (1986): *Developments in the study of cointegrated economic variables*. Oxford Bulletin of Economics and Statistics **48**, S. 213–228.
- Hall, A. D., Anderson, H. M. und Granger, C. W. J. (1992): *A cointegration analysis of treasury bill yields*. The Review of Economics and Statistics **74**, S. 116–126.

- Hassler, U. (1993): *Regression of spectral estimators with fractionally integrated time series*. Journal of Time Series Analysis **14**, S. 369–380.
- Hassler, U. und Wolters, J. (1994): *On the power of unit root tests against fractional alternatives*. Economics Letters **45**, S. 1–5.
- Hassler, U. und Wolters, J. (1995): *Long memory in inflation rates: International evidence*. Journal of Business and Economic Statistics **13**, S. 37–45.
- Hassler, U., Marmol F. und Velasco, C. (2006): *Residual log-periodogram inference for long-run relationships*. Journal of Econometrics **130**, S. 165–207.
- Hassler, U., Marmol F. und Velasco, C. (2008): *Fractional cointegration in the presence of linear trends*. Journal of Time Series Analysis **29**, S. 1088–1103.
- Hasslett, J. und Raftery, A. E. (1989): *Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource*. Journal of Applied Statistics **38**, S. 1–50.
- Hamilton, J. D. (1994): *Time series analysis*. Princeton University Press, Princeton.
- Hannan, E. J. und Quinn, B. G. (1979): *The determination of the order of an autoregression*. Journal of the Royal Statistical Society **41**, S. 190–195.
- Hauser, M. A. und Kunst, R. M. (1998): *Fractionally Integrated Models With ARCH Errors: With an Application to the Swiss One-Month Euromarket Interest Rate*. Journal of Econometrics **10**, S. 95–113.
- Henry, M. und Robinson, P. M. (1996): *Bandwidth choice in Gaussian semiparametric estimation of long range dependence*. Athens Conference in Applied Probability and Time Series Analysis, Lecture Notes in Statistics **115**, S. 220–232.
- Henry, M. und Zaffaroni, P. (2003): *The long range dependence paradigm for macroeconomics and finance*. Erschienen in: Doukhan, P., Oppenheim, G. und Taqqu, M. S. (Hrsg.): Long range dependence: Theory and applications, Birkhäuser, Boston / Springer Verlag, New York - Berlin - Heidelberg.
- Hosking, J. R. M. (1981): *Fractional differencing*. Biometrika **68**, S. 165–176.
- Hualde, J. und Robinson, P. M. (2006): *Semiparametric Estimation of Fractional Cointegration*. Faculty Working Papers 07/06, School of Economics and Business Administration, University of Navarra.
- Hualde, J. und Robinson, P. M. (2010): *Semiparametric inference in multivariate fractionally cointegrated systems*. Journal of Econometrics **157**, S. 492–511.
- Hualde, J. und Iacone, F. (2012): *First stage estimation of fractional cointegration*. Journal of Time Series Econometrics **4**, S. 1–32.

- Hurst, H. E. (1981): *Long-term storage capacity of reservoirs*. Transactions of the American Society of Civil Engineers **116**, S. 770–779.
- Hurvich, C. M., Deo, R. S. und Brodsky, J. (1998): *The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series*. Journal of Time Series Analysis **19**, S. 19–46.
- Hurvich, C. M. und Deo, R. S. (1999): *Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series*. Journal of Time Series Analysis **20**, S. 331–341.
- Iacone, F. (2009): *A semiparametric analysis of the term structure of the US interest rates*. Oxford Bulletin of Economics and Statistics **71**, S. 475–490.
- Johansen, S. (1988): *Statistical analysis of cointegration vectors*. Journal of Economic Dynamics and Control **12**, S. 231–254.
- Johansen, S. und Juselius, K. (1990): *Maximum likelihood estimation and inference on cointegration - With applications to the demand for money*. Oxford Bulletin of Economics and Statistics **52**, S. 169–210.
- Johansen, S. (1991): *Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models*. Econometrica **59**, S. 1551–1580.
- Johansen, S. (1995): *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press, Oxford.
- Johansen, S., Mosconi, R. und Nielsen, B. (2000): *Cointegration analysis in the presence of structural breaks in the deterministic trend*. Econometrics Journal **3**, S. 216–249.
- Johansen, S. und Nielsen, M. Ø. (2012a): *Likelihood inference for a fractionally cointegrated vector autoregressive model*. Econometrica **80**, S. 2667–2732.
- Johansen, S. und Nielsen, M. Ø. (2012b): *The role of initial values in nonstationary fractional time series models*. Working Pappers 1300, Queen's University, Department of Economics.
- Krämer, W. (1998): *Fractional integration and the augmented Dickey-Fuller Test*. Economics Letters **61**, S. 269–272.
- Krämer, W. und Hassler, U. (1998): *Limiting efficiency of OLS vs. GLS when regressors are fractionally integrated*. Economics Letters **60**, S. 285–290.
- Krämer, W. und Marmol, F. (2004): *The power of residual-based tests for cointegration when residuals are fractionally integrated*. Economics Letters **82**, S. 63–69.
- Krämer, W. und Sibbertsen, P. (2006): *The power of the KPSS-test for cointegration when residuals are fractionally integrated*. Economics Letters **91**, S. 321–324.

- Kruse, R. und Sibbertsen, P. (2009): *Testing for a break in persistence under long-range dependencies*. Journal of Time Series Analysis **30**, S. 263–285.
- Künsch, H. R. (1986): *Discrimination between monotonic trends and long-range dependence*. Journal of Applied Probability **23**, S. 1025–1030.
- Künsch, H. R. (1987): *Statistical aspects of self-similar processes*. Proceedings of the First World Congress of the Bernoulli Society **1**, S. 67–74.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. und Shin, Y. (1992): *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?* Journal of Econometrics **54**, S. 159–178.
- Lardic, V. und Mignon, S. (2004): *The exact maximum likelihood estimation of ARFIMA processes and model selection criteria: A Monte Carlo study*. Economics Bulletin **3**, S. 1–16.
- Lasak, K. (2008): *Maximum likelihood estimation of fractionally cointegrated systems*. CREATES Research Papers 2008–53, School of Economics and Management, University of Aarhus.
- Lasak, K. (2010): *Likelihood based testing for no fractional cointegration*. Journal of Econometrics **158**, S. 67–77.
- Lee, D. und Schmitt, P. (1996): *On the power of the KPSS test of stationarity against fractionally integrated alternatives*. Journal of Econometrics **73**, S. 285–302.
- Lildholdt, P. M (2000): *Long Memory and ARFIMA modelling*. Masterthesis veröffentlicht im Internet unter [http://mit.econ.au.dk/vip\\_htm/plildholdt/Master\%20thesis.pdf](http://mit.econ.au.dk/vip_htm/plildholdt/Master\%20thesis.pdf), zuletzt abgerufen am 23. Juni 2014.
- Lobato, I. N. und Velasco, C. (2006): *Optimal fractional Dickey-Fuller tests*. Econometrics Journal **9**, S. 492–510.
- Lobato, I. N. und Velasco, C. (2007): *Efficient Wald tests for fractional unit roots*. Econometrica **75**, S. 575–589.
- Lumsdaine, R. L. und Papell, D. H. (1997): *Multiple trend breaks and the unit root hypothesis*. Review of Economics and Statistics **79**, S. 212–218.
- Lyhagen, J. (1998): *Maximum likelihood estimation of the multivariate fractional cointegrating model*. Working Papers in Economics and Finance 233, Stockholm School of Economics.
- MacKinnon, J. G. und White, H. (1985): *Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties*. Journal of Econometrics **29**, S. 305–325.

MacKinnon, J. G. (1991): *Critical values for cointegration tests*. Erschienen als 13. Kapitel in: Engle, R. F. und Granger, C. W. J. (Hrsg.): *Long-run economic relationships: Readings in cointegration*, Oxford University Press, Oxford.

MacKinnon, J. G. (1994): *Approximate asymptotic distribution functions for unit-root and cointegration tests*. *Journal of Business and Economic Statistics* **12**, S. 167–176.

MacKinnon, J. G. (1996): *Numerical distribution functions for unit root and cointegration tests*. *Journal of Applied Econometrics* **11**, S. 601–618.

MacKinnon, J. G. und Nielsen, M. Ø. (2014): *Numerical distribution functions of fractional unit root and cointegration tests*. *Journal of Applied Econometrics* **29**, S. 161–171.

McLeod, A. I. und Hipel, K. W. (1978): *Preservation of the rescaled adjusted range: A reassessment of the Hurst phenomenon*. *Water Resources Research* **14**, S. 491–508.

Maddala, G. S. und Kim, I.-M. (1998): *Unit roots, cointegration and structural change*. Cambridge University Press, Cambridge.

Marinucci, D. und Robinson, P. M. (1999): *Alternative forms of fractional Brownian motion*. *Journal of Statistical Planning and Inference* **80**, S. 111–122.

Marinucci, D. und Robinson, P. M. (2000): *The averaged periodogram for non-stationary vector time series*. STICERD Econometrics Paper Series 2000/408, Suntory and Toyota International Centres for Economics and Related Disciplines, London School of Economics.

Marinucci, D. und Robinson, P. M. (2001): *Semiparametric fractional cointegration analysis*. *Journal of Econometrics* **105**, S. 225–247.

Marinucci, D. und Robinson, P. M. (2003): *Semiparametric frequency domain analysis of fractional cointegration*. Erschienen in: Robinson, P. M. (Hrsg.): *Time series with long memory*, Oxford University Press, Oxford.

Marmol, F. (1997): *Searching for fractional evidence using unit root tests*. mimeo, Universidad Carlos III, Madrid.

Marmol, F. (1998): *Spurious regression theory with nonstationary fractionally integrated processes*. *Journal of Econometrics* **84**, S. 233–250.

Newey, W. K. und West, K. D. (1987): *A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix*. *Econometrica* **55**, S. 703–708.

Nielsen, M. Ø. (2004): *Spectral analysis of fractionally cointegrated systems*. *Economics Letters* **83**, S. 225–231.

- Nielsen, M. Ø. (2005): *Semiparametric Estimation in Time-Series Regression with Long-Range Dependence*. *Journal of Time Series Analysis* **26**, S. 279–304.
- Nielsen, M. Ø. (2007): *Local Whittle analysis of stationary fractional cointegration and the implied-realized volatility relation*. *Journal of Business and Economic Statistics* **25**, S. 427–446.
- Nielsen, M. Ø. und Shimotsu, K. (2007): *Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach*. *Journal of Econometrics* **141**, S. 574–596.
- Nielsen, M. Ø. (2008): *A powerful test of the autoregressive unit root hypothesis based on a tuning parameter free statistic*. *Econometric Theory* **25**, S. 1515–1544.
- Nielsen, M. Ø. (2010): *Nonparametric cointegration analysis of fractional systems with unknown integration orders*. *Journal of Econometrics* **155**, S. 170–187.
- Nielsen, M. Ø. und Frederiksen, P. (2011): *Fully modified narrow band least squares estimation of weak fractional cointegration*. *Econometrics Journal* **14**, S. 77–120.
- Nielsen, M. Ø. und Morin, L. (2014): *FCVARmodel.m: A Matlab software package for estimation and testing in the fractionally cointegrated VAR model*. Working Papers 1273, Department of Economics, Queen’s University.
- Palma, W. (2007): *Long-memory time series*. John Wiley & Sons, Hoboken.
- Perron, P. (1989): *The great crash, the oil price shock, and the unit root hypothesis*. *Econometrica* **57**, S. 1361–1401.
- Phillips, P. C. B. und Ouliaris, S. (1988): *Testing for cointegration using principal components methods*. *Journal of Economic Dynamics and Control* **12**, S. 205–230.
- Robinson, P. M. (1991): *Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression*. *Journal of Econometrics* **47**, S. 67–84.
- Robinson, P. M. (1994a): *Semiparametric analysis of long-memory time series*. *Annals of Statistics* **22**, S. 515–539.
- Robinson, P. M. (1994b): *Efficient tests of nonstationary hypotheses*. *Journal of the American Statistical Association* **89**, S. 1420–1437.
- Robinson, P. M. (1995a): *Log-periodogram regression of time series with long range dependence*. *Annals of Statistics* **23**, S. 1048–1072.
- Robinson, P. M. (1995b): *Gaussian semiparametric estimation of  $\ln$  range dependence*. *Annals of Statistics* **23**, S. 1630–1661.

- Robinson, P. M. und Yajima, Y. (2002): *Gaussian semiparametric estimation of log range dependence*. Journal of Econometrics **106**, S. 217–241.
- Robinson, P. M. und Iacone, F. (2005): *Cointegration in fractional systems with deterministic trends*. Journal of Econometrics **129**, S. 263–298.
- Schlittgen, R. und Streitberg, B. H. J. (1994): *Zeitreihenanalyse*. R. Oldenbourg Verlag, München - Wien.
- Schmitt, M. C. und Tschernig, R. (1995): *The identification of fractional ARIMA models*. SFB 373 Discussion Papers, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Humboldt-Universität Berlin.
- Schwarz, G. E. (1979): *Estimating the dimension of a model*. Annals of Statistics **6**, S. 461–464.
- Schwert, G. W. (1989): *Tests for unit roots: A Monte Carlo investigation*. Journal of Business and Economic Statistics **7**, S. 147–159.
- Sephton, P. S. (2008): *Critical values of the augmented fractional Dickey-Fuller test*. Empirical Economics **35**, S. 437–450.
- Sephton, P. S. (2009): *Critical values of the augmented efficient Wald test for fractional unit roots*. Empirical Economics **37**, S. 615–626.
- Shao, X. (2011): *A simple test of changes in mean in the possible presence of long-range dependence*. Journal of Time Series Analysis **32**, S. 598–606.
- Shimotsu K. und Phillips P. C. B. (2005): *Exact local Whittle estimation of fractional integration*. Annals of Statistics **33**, S. 1890–1930.
- Shimotsu, K. und Phillips, P. C. B. (2006): *Local Whittle estimation of fractional integration and some of its variants*. Journal of Econometrics **130**, S. 209–333.
- Shimotsu, K. (2007): *Gaussian semiparametric estimation of multivariate fractionally integrated processes*. Journal of Econometrics **137**, S. 277–310.
- Shimotsu, K. (2010): *Exact local Whittle estimation of fractional integration with unknown mean and time trend*. Econometric Theory **26**, S. 501–540.
- Shimotsu, K. (2012): *Exact local Whittle estimation of fractionally cointegrated systems*. Journal of Econometrics **169**, S. 266–278.
- Shin, Y. und Schmidt, P. (1992): *The KPSS stationarity test as a unit root test*. Econometrics Letters **38**, S. 387–392.
- Sibbertsen, P. (2004): *Long memory versus structural breaks: An overview*. Statistical Papers **45**, S. 465–515.

- Sowell, F. (1992): *Maximum likelihood estimation of stationary univariate fractionally integrated time series models*. *Journal of Econometrics* **53**, S. 165–188.
- Tanaka, K. (1999): *The nonstationary fractional unit root*. *Econometric Theory* **15**, S. 549–582.
- Tschernig, R. (1994): *Wechselkurse, Unsicherheit und Long Memory*. Physica-Verlag, Heidelberg.
- Tschernig, R., Weber, E. und Weigand, R. (2013): *Long-run identification in a fractionally integrated system*. *Journal of Business and Economic Statistics* **31**, S. 438–450.
- Velasco, C. (1999a): *Non-stationary log-periodogram regression*. *Journal of Econometrics* **91**, S. 325–371.
- Velasco, C. (1999b): *Gaussian semiparametric estimation of non-stationary time series*. *Journal of Time Series Analysis* **20**, S. 87–127.
- Vogelsang, T. J. (1998): *Trend function hypothesis testing in the presence of serial correlation*. *Econometrica* **66**, S. 123–148.
- Whittle, P. (1951): *Hypothesis testing in time series analysis*. Almqvist & Wiksell, Uppsala.
- Xiao, Z. und Phillips, P. C. B. (1999): *Efficient detrending in cointegrating regressions*. *Econometric Theory* **15**, S. 519–548.

## Curriculum Vitae