

Der Einfluß der Benchmarkwahl auf das Ergebnis langfristiger Ereignisstudien*

The benchmark effect in long-run event studies

Olaf Ehrhardt und Ralf Koerstein

Humboldt-Universität zu Berlin, Institut für Bank-, Börsen- und Versicherungswesen,
Spandauer Straße 1, 10178 Berlin, Deutschland (e-mail: ehrhardt@wiwi.hu-berlin.de)

Eingegangen: 30. August 2000 / Angenommen: 31. Mai 2001

Zusammenfassung. Die Durchführung langfristiger Ereignisstudien setzt ein geeignetes Benchmarkmodell voraus, welches die erwartete Rendite der Untersuchungswertpapiere liefert. In den meisten bisherigen Studien wird diese durch die Änderungsrate eines Aktienindexes approximiert. Nach unseren Simulationsergebnissen entstehen bei einer solchen Vorgehensweise jedoch erhebliche Verzerrungen der Mittelwerte und der dazugehörigen Teststatistiken.

In unserer Untersuchung werden Ereignisstudien dadurch simuliert, daß aus einer umfassenden Datenbank über historische Aktienrenditen durch einen Zufallsprozeß einzelne Aktien ausgewählt und analysiert werden. Die wichtigsten Ergebnisse der durchgeführten Simulationen sind: (1) Geeignete Benchmarks sollten mittels einer Simulation ausgewählt werden, um potentielle Verzerrungen in den Ergebnissen bereits bei der Festlegung des Untersuchungsdesigns zu erkennen und zu reduzieren. (2) Die Verwendung speziell konstruierter Vergleichsportefeuilles, z.B. Size-Portefeuilles, führt in Verbindung mit bootstrap-geschätzten Konfidenzintervallen zu genaueren Untersuchungsergebnissen. (3) Verzerrungen der Untersuchungsergebnisse, die auf zeitlich variierende Renditeeffekte zurückzuführen sind, können bei identischen zeitlichen Verteilungen von Untersuchungs- und simulierten Stichproben erfaßt werden.

Abstract. An important step in long-horizon event studies is the choice of the benchmark that is used as a proxy for the expected return of the individual securities. Most existing studies use the rate of return of a stock market index. However, our simulation shows that such a procedure creates a significant bias in the means and in the test statistics.

* Für wertvolle Hinweise danken wir Professor Richard Stehle und zwei anonymen Gutachtern.

Korrespondenz an: O. Ehrhardt

We simulate event studies by randomly choosing stocks out of a large database of historical rates of returns. The main results are: (1) Benchmarks should be selected by a simulation procedure before the design of a study is established in order to reduce potential bias. (2) Using reference portfolios that take account of return anomalies, and testing for long-run abnormal returns with bootstrapped skewness-adjusted t-statistics lead to more precise results. (3) Misspecification due to time-varying return effects can be analyzed by the construction of random samples that share the same time distribution as the real events.

Schlüsselwörter: Simulation – Langfristige Ereignisstudie – Benchmarkwahl – Size-Effekt

Key words: Simulation – Long-horizon event study – Benchmark choice – Size effect

1. Problemstellung

Empirische Untersuchungen zur langfristigen Reaktion von Aktienkursen auf die Bekanntgabe unternehmerischer Entscheidungen setzen ein theoretisches Bewertungsmodell voraus, welches den Erwartungswert für die Rendite der betrachteten Aktie liefert. Zur Überprüfung der Eignung einer Benchmark zur empirischen Schätzung dieses Erwartungswertes bieten sich Simulationen zur Generierung von Stichproben unter der Nullhypothese und Bootstrap-Techniken zur Verifizierung der dazugehörigen Teststatistiken an.

Gegenstand von Ereignisstudien sind beispielsweise Entscheidungen des Managements einer Unternehmung über die Durchführung von Investitionsprojekten, über eine Fusion mit einem anderen Unternehmen, über den Kauf und Verkauf von Unternehmensteilen und die Wahl zwischen bestehenden alternativen Formen ihrer Finanzierung. Sobald diese Entscheidungen den Marktteilnehmern bekannt werden, sollte eine Reaktion des Kurses der betreffenden Aktie beobachtet werden, wenn sich mit der Managemententscheidung auch die erwarteten Zahlungsströme der Anteilseigner verändern. Die Frage nach der Effizienz dieser Informationsverarbeitung erhält einen stark praxisrelevanten Bezug, da signifikant positive oder negative Aktienkursreaktionen auf bestimmte Managemententscheidungen verwendet werden können, um Handlungsempfehlungen für eine optimale und konsequent kapitalmarktorientierte Unternehmenspolitik abzuleiten. Inwieweit die Marktteilnehmer die ihnen zur Verfügung stehenden Informationen tatsächlich effizient verarbeiten, zählt auch heute noch zu den ungelösten Fragestellungen im Gebiet Finanzierung.

Die Ausgangshypothese ist, daß in informationseffizienten Märkten nach Abschluß aller mit der angekündigten Entscheidung in Zusammenhang stehenden Maßnahmen keine weiteren systematischen Kursreaktionen beobachtet werden sollten. Unter dieser Prämisse entsprechen in der Folgezeit die Renditen der betreffenden Aktie (d.h. die relativen Aktienkursveränderungen) im Durchschnitt ihren

Erwartungswerten.¹

Empirische Untersuchungen führen jedoch zum Ergebnis, daß Aktienrenditen für Betrachtungszeiträume von drei bis fünf Jahren zum Beispiel nach Börseneinführungen (Ritter, 1991), Kapitalerhöhungen (Loughran u. Ritter, 1995), Aktienrückkäufen (Ikenberry et al., 1995), Dividendenveränderungen (Michaely et al., 1995) und Fusionen (Mitchell u. Stafford, 1997) systematisch von ihren erwarteten Renditen abweichen. Letztere werden i.d.R. mit Hilfe von Aktienindizes, Vergleichsportefeuilles oder -wertpapieren (Benchmarks) geschätzt.

Die beobachteten Abweichungen könnten zum einen mit den bestehenden Schwierigkeiten in Zusammenhang stehen, eine geeignete Benchmark als Schätzwert für die erwartete Rendite der Untersuchungswertpapiere auszuwählen (vgl. Brown u. Warner, 1980; Dimson u. Marsh, 1986; Fama, 1998). Das Problem besteht vor allem darin, daß es trotz intensiver Forschungen bis heute nicht gelungen ist, ein Kapitalmarktgleichgewichtsmodell zu entwickeln, welches zweifelsfrei die Preisbildung auf dem Aktienmarkt erklärt. Neben den aus Kapitalmarktgleichgewichtsmodellen abgeleiteten Bestimmungsfaktoren – insbesondere das mit dem Beta-Faktor gemessene nichtdiversifizierbare Risiko² – gibt es eine Reihe anderer, rein empirisch identifizierter Faktoren mit einem teilweise hohen Erklärungsgehalt für bestehende Renditeunterschiede zwischen einzelnen Aktien (Renditeanomalien). Hierzu zählen die Höhe der Marktkapitalisierung (Size-Effekt),³ der Quotient aus dem Buch- und Marktwert der Aktie⁴ und eine Reihe weiterer wichtiger Faktoren.

¹ Untersuchungen zur Informationseffizienz von Kapitalmärkten setzen ein Bewertungsmodell voraus, welches die erwartete Rendite einer Aktie unter der Annahme schätzt, daß die Marktteilnehmer alle bewertungsrelevanten Informationen in ihrer Erwartungsbildung berücksichtigen. Deshalb können Informationseffizienz und Bewertungsmodell nur gemeinsam einem Test unterzogen werden (Problem der verbundenen Hypothese bzw. Joint-Hypothesis-Problem).

² Der *Beta-Faktor* ist ein Maß für die Sensitivität zwischen der erwarteten Aktienrendite und der erwarteten Rendite des Marktportfolios. Mathematisch handelt es sich um den Quotienten aus der Kovarianz zwischen den beiden genannten erwarteten Renditen und der Varianz der erwarteten Rendite des Marktportfolios.

³ Zur Berechnung der *Marktkapitalisierung* werden die Anzahl ausstehender Aktien mit dem Börsenkurs multipliziert. Mit dem *Size-Effekt* wird eine Tendenz bezeichnet, daß Aktien mit niedrigerer Marktkapitalisierung im Durchschnitt höhere Renditen erzielen als Aktien mit hoher Marktkapitalisierung. Dies bedeutet zugleich, daß kleinere Unternehmen bei Existenz eines Size-Effektes im langfristigen Durchschnitt höhere Kapitalkosten haben.

⁴ Unter dem *Buchwert-Marktwert-Effekt* versteht man eine positive Korrelation zwischen dem Buchwert-Marktwert-Verhältnis einer Unternehmung und der langfristigen Durchschnittsrendite der ausgegebenen Aktien. Zum Buchwert werden die Vermögensgegenstände einer Unternehmung zu ihren Anschaffungs- oder Herstellungskosten abzüglich ihres durch Nutzung entstandenen Wertverlustes (Abschreibungen) bewertet. Im Marktwert werden zusätzlich die Erwartungen über die zukünftigen Wachstumsmöglichkeiten der Unternehmung kapitalisiert.

Zum anderen könnten Verzerrungen der berechneten Durchschnittswerte und der dazugehörigen Teststatistiken mit methodischen Problemen wie schiefen Überrenditenverteilungen (vgl. Barber u. Lyon, 1997) und bestehenden Querschnittskorrelationen (vgl. Brav, 2000) zwischen den Renditen einzelner Aktien in Beziehung stehen.

Erst wenn die Benchmarkwahl und die mit Ereignisstudien korrespondierenden statistischen Probleme als Ursachen für das Entstehen signifikanter Renditeabweichungen als wenig wahrscheinlich einzustufen sind, scheint es aus unserer Sicht zulässig, von Marktineffizienzen oder von einem niedrigen Effizienzgrad zu sprechen. Das Ziel dieser Studie ist deshalb die Ableitung von Empfehlungen zur Benchmarkwahl bei langfristigen Ereignisstudien, um bei Anwendung dieser Untersuchungsmethodik bei nicht simulierten Daten verzerrte Ergebnisse und daraus resultierende fehlerhafte ökonomische Interpretationen zu vermeiden.

Der folgende Abschnitt 2 gibt einen Überblick über die Berechnungsmethodik langfristiger Ereignisstudien, diskutiert erkannte Verzerrungen in den Durchschnittswerten bzw. Teststatistiken und erörtert Möglichkeiten zu ihrer Minderung.

In Abschnitt 3 werden eigene Simulationsrechnungen unter Verwendung verschiedener Benchmarkmodelle vorgestellt, um insbesondere den Fehler abzuschätzen, der bei der Schätzung der erwarteten Rendite der Untersuchungswertpapiere entstehen kann, wenn empirisch identifizierte Bestimmungsfaktoren zur Erklärung von Aktienrenditen nicht oder unzureichend berücksichtigt werden. Wir konzentrieren uns hierbei auf den erstmals von Banz (1981) empirisch festgestellten Einfluß der Marktkapitalisierung (*Size*) auf die Höhe der Aktienrendite (für den deutschen Markt vgl. Domke, 1987; Schnittke, 1989; Beiker, 1993; Oertmann, 1994; Sattler, 1994; Stehle, 1997; Wallmeier, 1997) und zeigen anhand dieses Effektes, daß eine ungenaue Spezifikation der verwendeten Benchmarkmodelle zu systematischen Verzerrungen der Untersuchungsergebnisse langfristiger Ereignisstudien führen kann (vgl. Dimson u. Marsh, 1986). Wir verwenden in unseren Untersuchungen kleine Stichprobengrößen - eine für Ereignisstudien in den meisten europäischen Ländern typische Situation - und überprüfen zusätzlich die Stabilität der Ergebnisse, indem wir durch eine Variation der Verteilungsannahme eine zeitliche Häufung von Ereignissen simulieren.

In Abschnitt 4 leiten wir Empfehlungen zur Wahl von Benchmarkmodellen ab, die es bereits bei der Festlegung des Untersuchungsdesigns ermöglichen, Verzerrungen der Berechnungsergebnisse bei sich im Zeitablauf verändernden Renditeanomalien zu erkennen und zu minimieren. Die vorgeschlagene Methodik ist auch auf Kombinationen mehrerer Renditeanomalien anwendbar und verdeutlicht somit die Vorteile von Simulationstechniken gegenüber anderen Verfahren.

Abschnitt 5 faßt die wichtigsten Erkenntnisse zusammen.

2. Berechnungsmethodik langfristiger Ereignisstudien

2.1. Berechnung des Durchschnittswertes

In der vorliegenden Simulationsstudie untersuchen wir eine langfristige Reaktion des Aktienkurses von zufällig ausgewählten Wertpapieren über einen Zeitraum von

36 Monaten. Durch die Zufallsauswahl ist gewährleistet, daß die Ergebnisse nicht systematisch von bewertungsrelevanten Ereignissen beeinflußt werden, d.h., es gilt die Nullhypothese, daß die Aktienrenditen im Durchschnitt ihren Erwartungswerten entsprechen. Dies versetzt uns in die Lage, die bei langfristigen Ereignisstudien verwendete Berechnungsmethodik –insbesondere die verwendete Benchmark– auf ihre Eignung zu überprüfen.

2.1.1. Berechnung einer Aktienrendite. Die einfachste Möglichkeit, eine langfristige Kursveränderung zu berechnen, würde darin bestehen, den Aktienkurs am Ende der Periode durch den Aktienkurs am Periodenanfang zu teilen. Da aber z.B. Dividendenzahlungen, Bezugsrechte und Nennwertumstellungen die Kurse verändern, werden für die Untersuchung Renditen verwendet, die alle an die Aktionäre fließenden Zahlungen berücksichtigen. Für die Berechnung einer Aktienrendite gilt deshalb:

$$R_{i,t} = \left(\frac{P_{i,t} + D_{i,t} + W_{i,t}}{P_{i,t-1}} - 1 \right) \cdot 100 \quad (1)$$

$R_{i,t}$ = Rendite der Aktie i für die Periode t (Angabe in Prozent);
 $P_{i,t(t-1)}$ = Kurs der Aktie i am Ende der Periode t bzw. $t - 1$;
 $D_{i,t}$ = Dividende der Aktie i in der Periode t ;
 $W_{i,t}$ = andere geldwerte Vorteile der Aktie i in der Periode t .

2.1.2. Berechnung einer Überrendite. Da Aktienkurse langfristig starken Schwankungen unterliegen, ist es erforderlich, sie im Untersuchungszeitraum mit einer Benchmark zu vergleichen, die als Proxy für den Erwartungswert der betreffenden Aktie dient. Auf diese Weise erhält man einen Vergleichswert, den die Aktie wahrscheinlich ohne das Eintreffen des zu untersuchenden Ereignisses aufgewiesen hätte. Die Differenz zwischen der Rendite der betrachteten Aktie und ihrem Erwartungswert wird in der englischsprachigen Literatur als Abnormal return bezeichnet. Um eine – wenn auch oft anzutreffende – Übersetzung als abnormale Rendite zu vermeiden, wird Abnormal return häufig als Überrendite übersetzt. An dieser Stelle sei jedoch ausdrücklich darauf hingewiesen, daß darunter sowohl positive als auch negative Abweichungen der Renditen von ihren Erwartungswerten verstanden werden.

$$AR_{i,t} = R_{i,t} - E(R_{i,t}) \quad (2)$$

$AR_{i,t}$ = Überrendite (Abnormal return) der Aktie i in der Periode t ;
 $E(R_{i,t})$ = zu Beginn der Periode t erwartete Rendite der Aktie i .

Häufig ist bei empirischen Untersuchungen zu beobachten, daß die Renditen der Untersuchungswertpapiere mit der Änderungsrate von Aktienindizes verglichen werden. Wenn jedoch die Marktkapitalisierung einen Einfluß auf die Höhe der zu erwartenden Renditen hat und systematische Unterschiede in der Marktkapitalisierung zwischen den untersuchten und den im Index enthaltenen Aktien bestehen, so kann eine solche Vorgehensweise zu Ergebnisverzerrungen führen.

Verzerrungen können auch dann entstehen, wenn die erwarteten Renditen mittels Zeitreihenanalyse aus Vergangenheitsdaten geschätzt werden. So führen beispielsweise Kapitalerhöhungen, Fusionen oder Unternehmensübernahmen zu einer Veränderung der Marktkapitalisierung. Besteht ein Size-Effekt, so sind allein aufgrund der Änderung der Marktkapitalisierung die erwarteten Renditen vor und nach der unternehmerischen Maßnahme nicht identisch.

Diese methodischen Probleme können dadurch vermieden werden, daß die Renditen der Untersuchungswertpapiere für die Periode nach dem Ereignis mit den Renditen von Portefeuilles verglichen werden, die Aktien mit ungefähr gleicher Marktkapitalisierung enthalten (Size-Portefeuilles). Alternativ kann die Renditeschätzung auch mit Hilfe von einzelnen, nach der Marktkapitalisierung vergleichbaren Aktien erfolgen.

Allgemeiner formuliert, werden bei einer solchen Vorgehensweise der Wahl eines Vergleichsportefeuilles oder –wertpapiers empirisch identifizierte Bestimmungsfaktoren berücksichtigt, die Renditeunterschiede zwischen einzelnen Aktien erklären. Akzeptiert man aufgrund der vergleichbaren Marktkapitalisierung die Annahme, daß im Durchschnitt aller einbezogenen Werte das nicht durch Portfoliodiversifikation zu beseitigende Risiko zwischen den untersuchten und den in den Vergleichsportefeuilles enthaltenen Aktien annähernd identisch ist, so liegt der Schätzung der erwarteten Aktienrendite ein um die Höhe der Marktkapitalisierung und gegebenenfalls weiterer Bestimmungsfaktoren erweitertes (empirisches) Sharpe-Lintner-CAPM⁵ zugrunde.

2.1.3. Mittelwertbildung. Zur Bestimmung langfristiger Überrenditen sind zusätzliche Überlegungen erforderlich, die das Untersuchungskonzept dahingehend präzisieren, auf welche Weise die Mittelwertbildung über alle einbezogenen Wertpapiere und über die betrachtete Zeitperiode erfolgen soll. In der Literatur werden hierzu verschiedene Verfahren vorgeschlagen, wobei es am zweckmäßigsten erscheint, bei der Berechnung langfristiger Überrenditen auf die als „*Buy-and-hold abnormal return (BHAR)*“ bezeichnete Methode zurückzugreifen (vgl. Dimson u. Marsh, 1986; Cusatis et al., 1993; Stehle u. Ehrhardt, 1999).⁶ Beim BHAR-Verfahren wird in einem ersten Schritt für jede ausgewählte Aktie durch Multiplikation der monatlichen Renditen ein Endwert nach T Monaten berechnet, der sich bei einer Kapitalanlage von einer Geldeinheit in t_0 ergeben hätte. Dieser Endwert wird in einem zweiten Schritt mit dem dazugehörigen und auf gleiche Weise berechneten Endwert einer Kapitalanlage in ein alternatives Vergleichswertpapier oder -portfeuille verglichen. In einem dritten Schritt wird die Differenz der berechneten Endwerte gebildet und über alle Werte der Stichprobe das arithmetische

⁵ Nach dem von Sharpe und Lintner entwickelten *Capital Asset Pricing Model (CAPM)* ist die erwartete Rendite einer Aktie eine positive lineare Funktion der Sensitivität der Aktienrendite (ausgedrückt durch den Beta-Faktor) auf Veränderungen der Rendite des Marktportfolios.

⁶ Die genannten Studien gehen auch auf die Vor- und Nachteile einer additiven und multiplikativen Renditeformulierung ein.

Mittel berechnet:

$$\begin{aligned}
 BHAR_T &= \frac{1}{N} \sum_{i=1}^N \left(\left(\prod_{t=1}^T (R_{i,t} + 1) \right) - \left(\prod_{t=1}^T (R_{P,t} + 1) \right) \right) \\
 &= \frac{1}{N} \sum_{i=1}^N (W_{i,T} - W_{P,T}).
 \end{aligned}
 \tag{3}$$

- $BHAR_T$ = Buy-and-hold abnormal return für den Zeitraum $t=0$ bis $t=T$;
- $W_{i,T}$ = Endwert zum Zeitpunkt T einer Kapitalanlage von einer Geldeinheit, die zu Beginn des Untersuchungszeitraumes ($t=0$) in Aktie i investiert wurde;
- $W_{P,T}$ = Endwert zum Zeitpunkt T einer Kapitalanlage von einer Geldeinheit in ein Vergleichsportefeuille oder –wertpapier für die Aktie i ;
- $R_{i(P),t}$ = Rendite der Aktie i (bzw. des dazugehörigen Vergleichsportefeuilles/-wertpapiers) im Monat t .

Eine Berechnung der langfristigen Überrendite nach Gleichung 3 hat den entscheidenden Vorteil, daß die ermittelte langfristige durchschnittliche Überrendite als Ergebnis einer realisierbaren Portefeuillestrategie interpretiert werden kann.

2.2. Bildung von Vergleichsportefeuilles und Renditeunterschiede

Werden die Renditen von Vergleichsportefeuilles als Schätzwerte für die erwartete Rendite der untersuchten Aktien verwendet, so stellt sich die Frage nach der Zusammensetzung und der Gewichtung der in den Portefeuilles enthaltenen Aktien. Für eine erste Schätzung der langfristigen Überrendite dürfte es zweckmäßig sein, bei der Portefeuillebildung zunächst alle an einer Wertpapierbörse oder in einem Marktsegment einer Börse gehandelten Aktien zu berücksichtigen (Vermeidung eines Selection bias). Werden diese Aktien im Portefeuille gleichgewichtet („naive“ Diversifikation), so berechnet sich der Endwert eines gleichgewichteten Portefeuilles zum Zeitpunkt T nach:

$$R_{P,T}^{GW} = \prod_{t=1}^T \left(\frac{\sum_{i=1}^N R_{i,t}}{N} + 1 \right) - 1.
 \tag{4}$$

$R_{P,T}^{GW}$ = Endwert zum Zeitpunkt T einer Kapitalanlage von einer Geldeinheit, die zu Beginn des Untersuchungszeitraumes ($t=0$) in das gleichgewichtete Portefeuille P investiert wurde.

Dabei ist zu berücksichtigen, daß die im Portefeuille enthaltenen Aktien regelmäßig umgeschichtet werden, um die mit Aktienkursveränderungen korrespondierenden Veränderungen der Portefeuillegewichte wieder auszugleichen. Wird der Endwert eines gleichgewichteten Portefeuilles nach Gleichung 3 berechnet, so kann – entgegen der in der Literatur häufig verwendeten Bezeichnung – vom Grundsatz nicht mehr von einer Buy-and-hold-Strategie gesprochen werden.

Alternativ können die im Portefeuille enthaltenen Aktien mit ihren Marktwerten gewichtet werden (marktwertgewichtetes Portefeuille). Zu ähnlichen Ergebnissen gelangt man, wenn die Änderungsraten geeigneter Aktienindizes Verwendung finden (vgl. Stehle et al., 2000). Für die Berechnung der Rendite eines marktwertgewichteten Portefeuilles gilt:

$$R_{P,T}^{MW} = \prod_{t=1}^T \left(\frac{\sum_{i=1}^N R_{i,t} \cdot MW_{i,t}}{\sum_{i=1}^N MW_{i,t}} + 1 \right) - 1. \quad (5)$$

$R_{P,T}^{MW}$ = Endwert zum Zeitpunkt T einer Kapitalanlage von einer Geldeinheit, die zu Beginn des Untersuchungszeitraumes ($t=0$) in das marktwertgewichtete Portefeuille P investiert wurde;

$MW_{i,t}$ = Marktwert der Aktie i zu Beginn des Monats t .

Bei Vorliegen eines Size-Effektes sind die Renditen eines gleichgewichteten Portefeuilles im langfristigen Durchschnitt höher als die eines marktwertgewichteten Portefeuilles, da die Aktien mit geringerer Marktkapitalisierung stärker gewichtet werden. Folglich werden bei einem Vergleich der Renditen der Untersuchungswertpapiere mit dem gleichgewichteten Portefeuille tendenziell niedrigere Überrenditen ausgewiesen.

Die formelmäßig in den Gleichungen (4) und (5) beschriebenen Vorschriften zur Portefeuillebildung sind analog auf Size-Portefeuilles anzuwenden, mit dem Unterschied, daß nur Teilmengen von Aktien wie folgt berücksichtigt werden: Die Aktien werden nach der Höhe ihrer Marktkapitalisierung geordnet und die Bildungsvorschriften entsprechend der gewünschten Anzahl von Size-Portefeuilles beispielsweise auf Quartile oder Quantile der zugrundeliegenden Verteilung bezogen. Aufgrund der im allgemeinen geringeren Marktwertunterschiede der in den Size-Portefeuilles enthaltenen Aktien sind die auf eine unterschiedliche Gewichtung der Aktien zurückzuführenden Renditeunterschiede deutlich niedriger.

Barber u. Lyon (1997) verweisen in Zusammenhang mit den konkreten Bildungsvorschriften der Portefeuilles auf einen möglichen *Rebalancing bias*, welcher insbesondere bei einer gleichgewichteten Portefeuillestrategie relevant ist. Bei einer solchen Strategie sind in jeder Periode Umschichtungen erforderlich, welche die als Folge der Aktienkursentwicklung entstandenen Veränderungen der Portefeuillegewichte ausgleichen. Dies bedeutet, daß Anteile von Aktien mit einer überdurchschnittlichen Renditeentwicklung in der vergangenen Periode verkauft und Anteile von Aktien mit einer unter dem Durchschnitt liegenden Rendite hinzugekauft werden. Diese, die Durchschnittsrendite beeinflussenden Umschichtungen betreffen die als Schätzer des Erwartungswertes verwendeten Portefeuillerenditen, nicht aber die Renditen der untersuchten Aktien (*Rebalancing bias*). Im Hinblick auf die Ursachen möglicher Verzerrungen läßt sich somit festhalten, daß sich bei der Schätzung erwarteter Renditen mit Hilfe gleichgewichteter Portefeuilles systematische Beeinflussungen der durchschnittlichen Überrendite sowohl aus der höheren Gewichtung von Aktien mit niedrigerer Marktkapitalisierung als auch aus

den regelmäßigen Umschichtungen des gleichgewichteten Portefeuilles (*Rebalancing bias*) ergeben können.

Unterschiedliche Kursentwicklungen der Aktien können außerdem dazu führen, daß sich die Marktwerte der in den Portefeuilles enthaltenen Aktien so verändern, daß die Zusammensetzung der Size-Portefeuilles nicht mehr ihren Bildungsvorschriften entsprechen. Hieraus resultierende Veränderungen in der Zusammensetzung der einzelnen Size-Portefeuilles stellen eine weitere Quelle für mögliche Verzerrungen bei der Schätzung der erwarteten Rendite des Untersuchungswertpapiers dar, die ebenfalls dem *Rebalancing bias* zuzurechnen sind.

Um den *Rebalancing bias* zu vermeiden, empfehlen Lyon et al. (1999), die Renditen von sogenannten Buy-and-hold-Vergleichsportefeuilles als Schätzwerte für die Renditen der untersuchten Aktien zu verwenden. Bei dieser Variante wird das dem Untersuchungswertpapier im Monat 0 zugeordnete Vergleichsportefeuille über die gesamte Betrachtungsperiode von T Monaten nicht verändert. Formelmäßig gilt (vgl. auch Blume u. Stambaugh, 1983):

$$R_{P,T}^{BH} = \sum_{i=1}^N \frac{\prod_{t=1}^T (1 + R_{i,t}) - 1}{N} \quad (6)$$

$R_{P,T}^{BH}$ = Endwert zum Zeitpunkt T einer Kapitalanlage von einer Geldeinheit, die zu Beginn des Untersuchungszeitraumes ($t=0$) in das Buy-and-hold Portefeuille P investiert wurde.

Die Verwendung von Buy-and-hold- respektive gleichgewichteten Vergleichsportefeuilles führt gemäß den Bildungsvorschriften der Gleichungen (4) und (6) zu unterschiedlichen Schätzwerten für die erwartete Rendite der betrachteten Aktie. Die Höhe dieser Renditeunterschiede ist von dem Ausmaß bestehender Autokorrelationen von Einzel- und Portefeuillerenditen abhängig. Zur Darstellung der genannten Zusammenhänge wird zur Vereinfachung ein Zwei-Periodenfall mit im Zeitablauf konstanten erwarteten Renditen unterstellt. Abweichungen zwischen den tatsächlich beobachteten und den erwarteten Renditen sind zufällig und werden im Störterm $\varepsilon_{i,t}$ erfaßt. Für die erwartete Rendite eines Buy-and-hold-Vergleichsportefeuilles gilt (vgl. Roll, 1983):

$$\begin{aligned} E(R_P^{BH}) &= E \left[\frac{1}{N} \sum_{i=1}^N [(R_i + \varepsilon_{i,1}) \cdot (R_i + \varepsilon_{i,2})] \right] \\ &= E \left[\frac{1}{N} \sum_{i=1}^N (R_i^2 + R_i \varepsilon_{i,1} + R_i \varepsilon_{i,2} + \varepsilon_{i,1} \varepsilon_{i,2}) \right] \\ E(R_P^{BH}) &= \frac{1}{N} \sum_{i=1}^N R_i^2 + \frac{1}{N} \sum_i \sigma_{\varepsilon_{i,1}, \varepsilon_{i,2}} \end{aligned} \quad (7)$$

Für den Erwartungswert eines gleichgewichteten Portefeuilles gilt:

$$\begin{aligned}
 E(R_P^{GW}) &= E\left[\left[\frac{1}{N}\sum_{i=1}^N(\bar{R}_i + \varepsilon_{i,1})\right] \cdot \left[\frac{1}{N}\sum_{i=1}^N(\bar{R}_i + \varepsilon_{i,2})\right]\right] \\
 E(R_P^{GW}) &= \bar{R}_i^2 + Cov\left[\frac{1}{N}\sum_{i=1}^N(R_i + \varepsilon_{i,1}); \frac{1}{N}\sum_{i=1}^N(R_i + \varepsilon_{i,2})\right] \\
 E(R_P^{GW}) &= \bar{R}_i^2 + \sigma_{\varepsilon_{i,1}; \varepsilon_{i,2}} \quad (8)
 \end{aligned}$$

Die Differenzbildung von (8) und (7) führt zu:

$$\begin{aligned}
 E(R_P^{BH} - R_P^{GW}) &= \left(\frac{1}{N}\sum_{i=1}^N R_i^2 - \left(\frac{1}{N}\sum_{i=1}^N R_i\right)^2\right) \\
 &\quad + \left(\frac{1}{N}\sum_{i=1}^N \sigma_{\varepsilon_{i,1}; \varepsilon_{i,2}} - \sigma_{\bar{\varepsilon}_1, \bar{\varepsilon}_2}\right) \quad (9)
 \end{aligned}$$

Gemäß der Jensenschen Ungleichung gilt aufgrund der Konvexität der quadratischen Funktion für den ersten Klammerausdruck die Relation:

$$\left(\frac{1}{N}\sum_{i=1}^N R_i\right)^2 \leq \frac{1}{N}\sum_{i=1}^N R_i^2. \quad (10)$$

Der Unterschied zwischen beiden Termen ist aber gerade nach dem Verschiebungssatz der mittleren quadratischen Abweichung die Varianz von R_i . Deshalb gilt (vgl. Roll, 1983):

$$E(R_P^{BH} - R_P^{GW}) = \sigma_{R_i}^2 + \left(\frac{1}{N}\sum_{i=1}^N \sigma_{\varepsilon_{i,1}; \varepsilon_{i,2}} - \sigma_{\bar{\varepsilon}_1, \bar{\varepsilon}_2}\right). \quad (11)$$

Bei fehlender Autokorrelation der Residuen ist der Klammerausdruck in Gleichung (11) null. Bei negativer Autokorrelation der einzelnen Wertpapierrenditen oder bei positiver Autokorrelation der Portefeuillerenditen ist der Schätzwert für die erwartete Rendite der Untersuchungswertpapiere in jedem Fall größer, wenn ein gleichgewichtetes Vergleichsportefeuille zugrundegelegt wird (vgl. Roll, 1983). Das Ausmaß der beobachteten Autokorrelation ist zudem von der Höhe der Marktkapitalisierung der Aktien abhängig. So findet Stehle (1997) in einer empirischen Untersuchung für den deutschen Markt, daß monatliche Renditen von Size-Portefeuilles, die Aktien mit niedrigerer Marktkapitalisierung enthalten, im Durchschnitt eine höhere positive Autokorrelation als die Renditen von Size-Portefeuilles mit einem höheren Marktwert aufweisen.

2.3. Berechnung der Teststatistiken

Cowan u. Sergeant (2001), Lyon et al. (1999), Kothari u. Warner (1997) und Barber u. Lyon (1997) finden bei der Simulation langfristiger Überrenditen stark verzerrte Teststatistiken bei der Überprüfung der Nullhypothese,

$$H_0 : \mu(BHAR_T) = \mu_0(BHAR_T). \quad (12)$$

Die Verzerrungen der Teststatistiken werden in erster Linie durch eine schiefe Überrenditenverteilung verursacht, die entsteht, wenn Portefeullerrenditen als Schätzwert für die Renditen der betrachteten Aktien genommen werden (*Skewness bias*). Die beobachtete Schiefe der Überrenditenverteilung würde bei Anwendung des t-Testes dazu führen, daß zu häufig signifikant negative und zu selten signifikant positive Überrenditen ausgewiesen werden. Auch der nichtparametrische Wilcoxon-Test (vgl. Wilcoxon, 1945) oder der Vorzeichentest setzen symmetrische Verteilungen voraus und würden ebenfalls zu verzerrten Teststatistiken führen.

Um die aus einer schiefen Verteilung resultierenden Verzerrungen in den Teststatistiken zu reduzieren, schlagen Cowan u. Sergeant (2001) vor, die Schiefe durch Winsorisation⁷ der berechneten Überrenditen zu reduzieren und die Auswirkungen einer positiven Korrelation zwischen den Aktienrenditen auf die Varianz durch den t-Test für unabhängige Stichproben zu berücksichtigen. Häufig wird auch die Verwendung von einzelnen Vergleichswertpapieren zur Berechnung von Überrenditen empfohlen (vgl. Barber u. Lyon, 1997). Der Vorteil dieser Methode ist sowohl die Vermeidung des *Rebalancing bias* als auch des *Skewness bias*, der Nachteil liegt in einer höheren Standardabweichung der beobachteten Stichprobenmittelwerte (vgl. Cowan u. Sergeant, 2001). Lyon et al. (1999) empfehlen, die Teststatistiken auf der Grundlage eines schiefeangepaßten t-Tests zu ermitteln. Die transformierte Teststatistik wird berechnet durch:⁸

$$t^s = \sqrt{N} \cdot \left[\frac{\mu(BHAR_T)}{\sigma(BHAR_T)} + \frac{1}{3} \cdot \gamma(BHAR_T) \cdot \left[\frac{\mu(BHAR_T)}{\sigma(BHAR_T)} \right]^2 + \frac{1}{6 \cdot N} \cdot \gamma(BHAR_T) \right],$$

⁷ Unter *Winsorisation* wird ein Verfahren zur Begrenzung von Extremwerten verstanden. Es werden alle Beobachtungen, die zum Beispiel drei oder mehr Standardabweichungen vom Mittelwert entfernt liegen, auf den niedrigsten bzw. höchsten der verbliebenen Werte begrenzt.

⁸ Die schiefeangepaßte Teststatistik wurde von Johnson (1978) auf der Grundlage einer Edgeworth-Expansion entwickelt, von Hall (1992) und Sutton (1993) diskutiert. Die Anwendung der schiefeangepaßten Teststatistik bei langfristigen Ereignisstudien geht auf eine Idee von Lyon et al. (1999) zurück.

$$t^s = t + \sqrt{N} \cdot \left[\frac{1}{3} \cdot \gamma(BHAR_T) \cdot \left[\frac{\mu(BHAR_T)}{\sigma(BHAR_T)} \right]^2 + \frac{1}{6 \cdot N} \cdot \gamma(BHAR_T) \right] \quad \text{mit}$$

$$\gamma(BHAR_T) = \frac{\sum_{i=1}^N [BHAR_T - \mu(BHAR_T)]^3}{N \cdot [\sigma(BHAR_T)]^3}. \quad (13)$$

$\mu(BHAR_T)$ ist der Mittelwert der Überrenditen aller Untersuchungswertpapiere einer Stichprobe, $\sigma(BHAR_T)$ die dazugehörige (Querschnitts-)Standardabweichung und $\gamma(BHAR_T)$ der Koeffizient der Schiefe.

In Anlehnung an Ikenberry et al. (1995) schlagen Lyon et al. (1999) vor, die kritischen t-Werte zur Ablehnung der Nullhypothese mit einem Bootstrap-Verfahren zu bestimmen. An dem Einsatz von Bootstrap-Verfahren wird zuweilen Kritik geäußert, da sich die Renditen von Untersuchungs- und Vergleichswertpapieren in ihrer Varianz unterscheiden können. Eine höhere Varianz der Untersuchungswertpapiere begründen Mitchell u. Stafford (1997) damit, daß die betrachteten Unternehmen Entscheidungen zur Investitions-, Finanzierungs- und Ausschüttungspolitik treffen, die Vergleichsunternehmen hingegen nicht. Dem ist entgegenzuhalten, daß nach Abschluß aller mit der angekündigten Managemententscheidung verbundenen Maßnahmen keine systematischen Varianzunterschiede bestehen dürften, wenn die Märkte informationseffizient sind.

Zur Überprüfung, ob die in den Simulationsrechnungen ermittelten Ablehnungshäufigkeiten signifikant von den theoretisch erwarteten Werten abweichen, wird der Binomialtest verwendet. Die Teststatistik berechnet sich als Quotient aus der Differenz zwischen den beobachteten und theoretisch erwarteten Werten und dem Standardfehler der Binomialverteilung (vgl. Hoel, 1984):

$$\left| \frac{k_{obs} - N \cdot p}{\sqrt{N \cdot p \cdot (1-p)}} \right| \geq z_{1-\frac{\alpha}{2}}. \quad (14a)$$

Für die Berechnung der p-Werte gilt:

$$\Pr(k \geq k_{obs}) = \sum_{m=k_{obs}}^N \binom{N}{m} p^m (1-p)^{N-m},$$

$$\Pr(k \leq k_{obs}) = \sum_{m=0}^{k_{obs}} \binom{N}{m} p^m (1-p)^{N-m}. \quad (14b)$$

2.4. Zuordnung von Untersuchungswertpapieren und Size-Portefeuilles

In den nachfolgenden Simulationsrechnungen werden Size-Portefeuilles auf der Grundlage der Quantile der Marktwertverteilung gebildet. Für jedes der zehn Size-

Portefeuilles wird in einem ersten Schritt die durchschnittliche Marktkapitalisierung aller im Portefeuille enthaltenen Aktien bestimmt. In einem zweiten Schritt wird jedem Untersuchungswertpapier im Monat 0 genau das Size-Portefeuille zugeordnet, bei dem die Differenz zwischen dem Marktwert des Untersuchungswertpapiers (MW_i) und der durchschnittlichen Marktkapitalisierung aller in einem Size-Portefeuille enthaltenen Aktien $\overline{MW_{SPj}}$ minimal ist, also:⁹

$$\text{Size-Portefeuille-Nr. } j: j = \min_j |MW_i - \overline{MW_{SPj}}| \text{ mit } j = 1, \dots, 10. \quad (15)$$

2.5. Datenbeschreibung und Stichprobenbildung mittels Simulation

Aus der Grundgesamtheit aller im amtlichen Handel der Frankfurter Wertpapierbörse zwischen 1960 und 1992 notierten Aktien werden 50.000 Beobachtungen mit Zurücklegen gezogen. Zur Erzeugung von ganzzahligen Zufallszahlen verwenden wir einen zellulären Automaten nach Wolfram (1986). Rationale Zufallszahlen werden mit einem Subtraktion-mit-Übertrag-Verfahren nach Marsaglia u. Zaman (1991) generiert. Dadurch können qualitativ gute Zufallszahlen mit (sehr) langen Periodenlängen erreicht werden. Der Auswahl der Aktien und des Ereigniszeitpunktes liegt zunächst die Annahme einer Gleichverteilung zugrunde, die jedoch im weiteren Verlauf der Untersuchungen durch eine periodenspezifische Zufallsauswahl ersetzt wird. Die 50.000 Beobachtungen werden in 1.000 Stichproben zu jeweils 50 Aktien eingeteilt.

Den Zeitpunkt der Zufallsauswahl definieren wir als Monat 0. Dies entspricht in langfristigen Ereignisstudien jenem Zeitpunkt, in dem die mit der Managemententscheidung verbundenen Maßnahmen praktisch abgeschlossen sind. So endet beispielsweise die Durchführung einer Kapitalerhöhung oder einer Börseneinführung mit der Notierung der neu ausgegebenen Aktien. Ausgehend vom Monat 0 berechnen wir für die zufällig ausgewählten Aktien die langfristige Überrendite über den Zeitraum der folgenden 36 Monate. Die Länge der Berechnungsperiode orientiert sich an den in der Literatur üblicherweise verwendeten Zeiträumen zur Berechnung langfristiger Überrenditen. Bedingt durch den Berechnungszeitraum von 36 Monaten werden Renditedaten bis 1995 in die Betrachtung einbezogen.

Für unsere Simulationsstudie stehen für den genannten Zeitraum und das genannte Marktsegment monatliche Aktienrenditen aus einer Datenbank von Prof. Stehle zur Verfügung.¹⁰ Unsere Simulation verbindet also aus empirisch beobach-

⁹ Um Erfahrungen hinsichtlich der Sensitivität der Berechnungsergebnisse zu erhalten, wurde alternativ für jedes der j Size-Portefeuilles das Wertpapier mit dem niedrigsten und dem höchsten Marktwert und somit eine untere und obere Grenze bestimmt. Die zu untersuchende Aktie wurde dann jenem Size-Portefeuille zugeordnet, indem sie sich selbst befindet. Die Simulationsergebnisse sind für beide alternative Zuordnungskriterien fast identisch.

¹⁰ Die vorliegende Simulationsstudie ist nach unserem Wissen die erste Studie, die nicht auf Daten der CRSP-Datenbank basiert. Die Verwendung einer anderen Datenquelle ist auch deshalb von Bedeutung, weil empirische Untersuchungen durch einen „*Ex post selection bias*“ als Folge unvollständiger Datenbanken (Kothari et al., 1995), fehlerhafter Daten (Shumway, 1997) und Verzerrungen der verwendeten Indizes (Canina et al., 1998) auf systematische Weise beeinflusst sein können.

teten Kursen, Dividenden und anderen geldwerten Vorteilen berechneten Renditen mit einem künstlichen Zufallsprozeß der Bildung von „Pseudo“-Untersuchungsstichproben. Die Zufallsauswahl der Wertpapiere gewährleistet dabei, daß die Gültigkeit der Nullhypothese (Gleichung 12) angenommen werden kann und somit eine Überprüfung der Berechnungsmethodik ermöglicht wird.

Die Renditeberechnungen basieren auf Kassakursen am letzten Handelstag des Monats. Es werden Dividenden und andere, den Aktionären zufließenden geldwerten Vorteile (Bezugsrechte, Gratisaktien etc.) berücksichtigt. Wir unterstellen einen individuellen Grenzsteuersatz von 0 Prozent, so daß die Körperschaftsteuergutschrift in voller Höhe angerechnet wird.

Um einen „*Selection bias*“ zu vermeiden, werden auch Aktien in die Untersuchung einbezogen, die im genannten Zeitraum ihre Börsennotierung durch Konkurs, Fusion, Going private oder anderen Gründen einstellten bzw. in Zusammenhang mit einem Going public, einem Börsensegmentwechsel oder der Erstemission einer weiteren Aktiegattung die Notierung aufnahmen. In der vorliegenden Simulationsstudie wird in dem zuerst genannten Fall die kumulierte Überrendite für den Zeitraum bis zur Notizeinstellung berechnet. Dieser Wert wird dann unverändert bis zum 36. Monat fortgeschrieben. Wenn eine Aktie im Monat 0 ausgewählt wird, die im Folgemonat die Börsennotierung einstellt, so stehen keine monatlichen Renditen für die Berechnung zur Verfügung. Dies führt zu einer Reduzierung der in der Stichprobe enthaltenen Aktien um durchschnittlich 0,13 Prozent.

Als Schätzwert für die erwartete Rendite der zufällig ausgewählten Wertpapiere werden folgende Daten verwendet:

1. die Rendite eines gleich- und marktwertgewichteten Portefeuilles aller im amtlichen Handel der Frankfurter Wertpapierbörse notierten Aktien;
2. die Rendite eines der zehn aus den gleichen Daten gebildeten gleich-, marktwertgewichteten und Buy-and-hold Size-Portefeuilles;
3. die Rendite eines nach der Marktkapitalisierung vergleichbaren Wertpapiers.

Bei einem Simulationslauf werden insgesamt 1,85 Millionen Datensätze verarbeitet (50.000×37). Um eine Erhöhung der Stichprobengröße von 50 auf 150 zu erreichen, müßten 5,55 Millionen Datensätze in die Berechnungen einbezogen werden. Zur Reduzierung des Rechenaufwandes wird bei der Erhöhung der Stichprobengröße deshalb erneut auf eine Bootstrap-Methode zurückgegriffen. Als Ergebnis der ersten Simulationsrechnung liegen 50.000 über einen Zeitraum von 36 Monaten berechnete langfristige Überrenditen vor. Aus diesen Daten wird durch Ziehen mit Zurücklegen eine Stichprobe von 150.000 Beobachtungen aufgebaut, die dann in 1.000 Einzelstichproben eingeteilt wird.

2.6. Bestimmung der kritischen *t*-Werte durch Bootstrap-Verfahren

Durch die Asymmetrie der Überrenditenverteilung folgen die berechneten Prüfgrößen, d.h. die konventionellen *t*-Werte und nach den Erfahrungen von Lyon et al. (1999) auch die transformierten schiefeangepaßten t^s -Werte, nicht mehr einer theoretischen Verteilung. Für die berechneten langfristigen Überrenditen existiert

eine unbekannte theoretische Verteilung mit der Prüfgröße $\Theta(t^s)$. Die Bootstrap-Methode (vgl. Efron, 1982) ersetzt die unbekannte theoretische Verteilung $F(BHAR_{T=36})$ durch eine empirische Verteilungsfunktion $\hat{F}_B(BHAR_{T=36})$ mit der dazugehörigen Statistik $\hat{\Theta}(t_{b=1}^s, \dots, t_{B=1000}^s)$.

Zur Bestimmung des Verwerfungsbereiches ordnen wir in einem ersten Schritt jedem der bereits ermittelten 50.000 Untersuchungswertpapiere ein anderes, zufällig ausgewähltes Wertpapier zu, welches zu dem gleichen Size-Portefeuille gehört. Als Ergebnis entstehen somit 1.000 Bootstrap-Stichproben. Für jede Stichprobe werden in einem zweiten Schritt die langfristigen Überrenditen für einen Zeitraum von 36 Monaten berechnet ($BHAR_{T=36, b=1}^*, \dots, BHAR_{T=36, B=1000}^*$). Als Schätzwert für die langfristige Rendite dienen die gleichen Vergleichsportefeuilles, die auch zur Analyse der Untersuchungswertpapiere Verwendung finden. Für jede Bootstrap-Stichprobe wird dann in einem dritten Schritt ein Wert der Statistik $\hat{\Theta}(t_{b=1}^{s*}, \dots, t_{B=1000}^{s*})$, d.h. ein schiefeangepaßter t^s -Wert berechnet. In einem vierten Schritt werden aus der Verteilung der berechneten schiefeangepaßten t^s -Werte die Grenzen des Verwerfungsbereiches geschätzt, indem auf der Grundlage der Percentile der schiefeangepaßten t^s -Wertverteilung die kritischen Werte zur Ablehnung der Nullhypothese ermittelt werden. Wird zum Beispiel ein Signifikanzniveau von fünf Prozent unterstellt, so werden der 25. und der 975. der nach der Größe geordneten schiefeangepaßten t^s -Werte als kritische Werte angenommen, also:

$$\Pr \left[-t_b^s \leq -t_{krit.}^{s*} \right] = \Pr \left[t_b^s \geq t_{krit.}^{s*} \right] = \frac{\alpha}{2}. \quad (16)$$

3. Empirische Ergebnisse

3.1. Untersuchungen für unterschiedliche Stichprobengrößen

3.1.1. Deskriptive Beschreibung der Verteilung der Stichprobenmittelwerte.

Durch die Zufallsauswahl der Untersuchungswertpapiere ist gewährleistet, daß die Aktienrenditen in dem Zeitraum vor Untersuchungsbeginn nicht durch unternehmerische Entscheidungen auf systematische Weise beeinflußt sind. Folglich sollten auch die über den nachfolgenden Zeitraum von 36 Monaten berechneten Renditen der betrachteten Aktien im Durchschnitt nicht von denen der Vergleichsportefeuilles bzw. -wertpapiere abweichen. Ohne Fehlspezifikationen der Untersuchungsmethodik ist zu erwarten, daß die durchschnittliche Differenz der genannten Renditen, d.h. die durchschnittliche langfristige Überrendite, nahe bei null liegt.

In der Tabelle 1 sind die durchschnittlichen Stichprobenmittelwerte, die zugehörigen Standardabweichungen und die Werte für die Schiefe der Verteilung angegeben. Die über einen Zeitraum von 36 Monaten berechneten langfristigen Überrenditen sind hoch signifikant, wenn die Renditen der zufällig ausgewählten Aktien mit den dazugehörigen Renditen eines Portefeuilles aus allen amtlich in Frankfurt gehandelten Aktien verglichen werden. Dieser signifikante Unterschied der berechneten Renditen in Höhe von -1,78 bzw. 3,62 Prozentpunkte deutet damit auf systematische Verzerrungen der Berechnungsergebnisse hin. Eine unterschiedliche Gewichtung der im Vergleichsportefeuille enthaltenen Aktien führt zu einem

Tabelle 1. Mittelwert, Standardabweichung und Schiefe der Überrenditenverteilung einer simulierten langfristigen Ereignisstudie

Die simulierte Studie basiert auf 1.000 Stichproben mit jeweils 50 bzw. 150 Aktien. Für die einzelnen Untersuchungswertpapiere und Vergleichsportefeuilles bzw. -wertpapiere werden über einen Zeitraum von 36 Monaten Überrenditen nach Gleichung 3 berechnet. Mit dem t-Test für abhängige Stichproben wird überprüft, ob sich der Mittelwert der Stichprobenmittelwert-Verteilung signifikant von null unterscheidet

Schätzwert für die langfristige Rendite	Stichprobenmittelwert-Verteilung			Durchschnittliche Schiefe der Stichprobenverteilung
	Mittelwert (%)	t-Wert	Standardabweichung (%)	
Teil A: 1.000 Stichproben mit 50 Aktien				
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>				
gleichgewichtetes Portefeuille	-1,78	-5,17	10,91	1,77
marktwertgewichtetes Portefeuille	3,62	10,26	11,51	1,56
<i>Berücksichtigung des Size-Effektes</i>				
gleichgewichtetes Size-Portefeuille	-0,61	-1,81	10,62	1,61
marktwertgewichtetes Size-Portefeuille	-0,21	-0,62	10,62	1,62
Buy-and-hold Size-Portefeuille	0,88	2,65	10,52	1,55
Vergleichswertpapier	-0,61	-1,28	15,11	-0,09
Teil B: 1.000 Stichproben mit 150 Aktien				
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>				
gleichgewichtetes Portefeuille	-1,82	-9,02	6,38	3,17
marktwertgewichtetes Portefeuille	3,61	17,19	6,34	2,92
<i>Berücksichtigung des Size-Effektes</i>				
gleichgewichtetes Size-Portefeuille	-0,64	-3,26	6,23	2,96
marktwertgewichtetes Size-Portefeuille	-0,25	-1,26	6,24	2,98
Buy-and-hold Size-Portefeuille	0,89	4,52	6,24	2,95
Vergleichswertpapier	-0,27	-0,99	8,50	-0,01

Unterschied von über fünf Prozentpunkten bei der berechneten durchschnittlichen Überrendite nach 36 Monaten.

Da sich bei beiden Berechnungen weder die Untersuchungspapiere noch die in den Vergleichsportefeuilles enthaltenen Aktien unterscheiden, können die festgestellten Unterschiede nur mit der unterschiedlichen Gewichtung der Aktien in beiden Vergleichsportefeuilles bzw. mit den monatlichen Umschichtungen des gleichgewichteten Vergleichsportefeuilles in Zusammenhang stehen.

Das negative Vorzeichen bei Verwendung des gleichgewichteten Portefeuilles deutet auf eine Beeinflussung der Ergebnisse durch den Size-Effekt hin. Aktien mit geringerer Marktkapitalisierung werden bei einer gleichgewichteten Portefeuillestrategie stärker gewichtet, als dies bei einer Gewichtung der Aktien mit ihren Marktwerten der Fall ist. Bei einem bestehenden Size-Effekt ist folglich die langfristige Durchschnittsrendite eines gleichgewichteten Portefeuilles höher und somit die langfristige Überrendite der Untersuchungswertpapiere niedriger.

Neben dem Size- könnte auch der *Rebalancing*-Effekt die unterschiedlichen langfristigen Überrenditen erklären. Zur Aufrechterhaltung der Portefeuillegewichtungen werden bei einer gleichgewichteten Portefeuillestrategie Anteile von Aktien mit überdurchschnittlicher Kursentwicklung in der zurückliegenden Periode verkauft und Aktien mit unterdurchschnittlicher Kursentwicklung hinzugekauft. Diese Strategie läßt (unter Vernachlässigung von Transaktionskosten) im Vergleich zu einer marktwertgewichteten Portefeuillestrategie auf den ersten Blick eine geringere langfristige Durchschnittsrendite vermuten. Werden kurzfristige Schwankungen der Renditen um ihren Erwartungswert beobachtet, existiert ein kurzfristiger Winner-Loser-Effekt¹¹ oder sind die Portefeuillerenditen positiv autokorreliert (vgl. Gleichung 11), so kann eine gleichgewichtete Portefeuillestrategie (ohne Transaktionskosten) zu höheren langfristigen Durchschnittsrenditen führen.

Wird ein marktwertgewichtetes Portefeuille als Schätzwert für die erwartete Rendite der betrachteten Aktien verwendet, so kann die hier ebenfalls beobachtete deutliche Verzerrung des Mittelwertes zumindest teilweise mit der zugrundeliegenden Simulationstechnik in Zusammenhang stehen. Die einzelnen Untersuchungswertpapiere werden gleichverteilt aus der Grundgesamtheit gezogen, während bei der Berechnung der Portefeuillerendite die Aktien entsprechend ihren Marktwerten gewichtet werden. Bei einem bestehenden Size-Effekt ist bei einer solchen Vorgehensweise zu erwarten, daß eine stärkere Gewichtung der Aktien mit höherer Marktkapitalisierung im Vergleichsportefeuille zu positiven langfristigen Überrenditen führt.

Werden die Renditen der zufällig ausgewählten Wertpapiere mit Schätzwerten für die langfristige Rendite verglichen, die den Size-Effekt und somit unternehmensspezifische Unterschiede zwischen den einzelnen untersuchten Aktien berücksichtigen, so ergeben sich mit Werten zwischen 0,88 und -0,61 Prozentpunkten deutlich geringere Verzerrungen des Mittelwertes. Mit Ausnahme der Verwendung von Buy-and-hold Size-Portefeuilles als Benchmarkmodell können keine von null

¹¹ Der *Winner-Loser-Effekt* beschreibt ein durchschnittliches Renditeverhalten, wonach Aktien mit einer unterdurchschnittlichen Renditeentwicklung in der Vergangenheit (Loser-Aktien) positive Überrenditen in der Zukunft und in einer vorhergehenden Zeitperiode als Winner-Aktien klassifizierte Werte zukünftig negative Überrenditen erzielen.

statistisch signifikanten Mittelwertunterschiede festgestellt werden. Die Ergebnisse deuten also darauf hin, daß die Berücksichtigung des Size-Effektes zu unverzerrteren Schätzungen für die langfristige durchschnittliche Überrendite führt, d.h. zu Ergebnissen, die näher am erwarteten Durchschnittswert von 0% liegen.

Werden die Renditen der untersuchten Aktien mit den anfänglichen (Monat 0) gleichgewichteten Buy-and-hold Size-Portefeuillerenditen verglichen, so wäre zunächst zu erwarten, daß die Abweichungen des Durchschnittswertes zwischen einer unteren und oberen Grenze von $-0,61$ bzw. $-0,21$ Prozentpunkten liegen würden, die sich bei einer Berechnung der langfristigen Überrenditen mit einem gleich- bzw. marktwertgewichteten Size-Portefeuille ergeben. Daß dies nicht der Fall ist, könnte wiederum mit dem *Rebalancing*-Effekt zusammenhängen. Die von uns verwendeten gleich- bzw. marktwertgewichteten Size-Portefeuilles werden monatlich neu gebildet, so daß Umschichtungen zwischen den zehn Size-Portefeuilles im Zeitablauf erfolgen. Die vorliegenden Ergebnisse deuten darauf hin, daß die Verwendung von Portefeuilles, die innerhalb der Betrachtungsperiode von 36 Monaten umgeschichtet werden, tendenziell zu einer Überschätzung der erwarteten Rendite der Untersuchungswertpapiere und somit zu einer negativen Überrendite führt.

Ein Renditevergleich zwischen den Untersuchungs- und den nach der Marktkapitalisierung vergleichbaren Wertpapieren führt zu einem nicht signifikanten Durchschnittswert, der nahe bei null liegt. Die Standardabweichung der Stichprobenmittelwerte von $15,11\%$ liegt jedoch bei kleinen Stichproben von 50 Aktien um fast 50 Prozent über den Werten, die sich bei einer Verwendung von Portefeuillerenditen ergeben.¹² Desweiteren kann bei der Verwendung von Vergleichswertpapieren als Benchmarkmodell eine symmetrische Verteilung der berechneten Überrenditen innerhalb einer Stichprobe beobachtet werden. Die Teststatistiken zur Überprüfung der Nullhypothese sind folglich nicht durch den *Skewness*-Effekt verzerrt.

Die in der Tabelle 1 enthaltenen Ergebnisse zeigen ferner, daß mit steigendem Stichprobenumfang die Querschnittsstandardabweichung der Stichprobenmittelwerte beträchtlich sinkt und im Durchschnitt höhere t-Werte ausgewiesen werden. Die Standardabweichung der Stichprobenmittelwerte gibt zudem Aufschluß über die durchschnittliche Abweichung vom „wahren“ Durchschnittswert, mit der im Rahmen **einer** empirischen Untersuchung eines konkreten bewertungsrelevanten Ereignisses gerechnet werden muß. Zudem wird (erwartungsgemäß) eine positive Schiefe der kumulierten langfristigen Überrenditen festgestellt, wenn die Renditen von Einzelwertpapieren mit Portefeuillerenditen verglichen werden. Die Schiefe in den einzelnen Stichproben nimmt mit der Anzahl der einbezogenen Beobachtungen zu, wenn Portefeuillerenditen zur Schätzung der erwarteten Rendite der Untersuchungswertpapiere verwendet werden. Deshalb müssen Verzerrungen bei der Überprüfung der Nullhypothese vermutet werden, wenn parametrische Tests oder nichtparametrische Testverfahren, die die Symmetrieeigenschaft voraussetzen, eingesetzt werden. Werden hingegen die Renditen von Vergleichswertpapie-

¹² Steigt die Standardabweichung der berechneten Überrenditen in einer Stichprobe um 50% , so müßte die Anzahl der Beobachtungen um das 2,25-fache höher liegen, um bei einem unveränderten Mittelwert einen identischen t-Wert zu berechnen (konventioneller t-Test).

ren als Schätzer für die langfristige Rendite verwendet, so ist die Verteilung der kumulierten Überrenditen symmetrisch.

3.1.2. Ablehnungshäufigkeiten für die Nullhypothese. Die Überprüfung der Nullhypothese für die untersuchten 1.000 Stichprobenmittelwerte erfolgt zunächst unter Verwendung parametrischer Testverfahren (Tabelle 2). Aufgrund der beobachteten schiefen Verteilungseigenschaften werden die Teststatistiken bei Verwendung von Portefeullerrenditen mit einem schiefeangepaßten t-Test berechnet. Werden die Renditen von Vergleichswertpapieren den Berechnungen zugrundegelegt, so kann ein konventioneller t-Test Verwendung finden, da die Überrenditenverteilung nach den in der Tabelle 2 angegebenen Werten für die Schiefe annähernd symmetrisch ist.

Die Ablehnungshäufigkeiten für die Nullhypothese sind in der Tabelle 2 dargestellt. Der erste angegebene Wert von 4,1% bedeutet, daß bei Verwendung eines gleichgewichteten Portefeulles als Schätzwert für die erwartete Rendite der analysierten Aktien in 41 der 1.000 Stichproben eine signifikant negative langfristige Überrendite ausgewiesen wird, wenn von einem einprozentigen Signifikanzniveau (zweiseitiger Test) ausgegangen wird. Ist die Methodik zur Berechnung des Mittelwertes und der Teststatistik gut spezifiziert, so dürfte eine signifikant negative Überrendite jedoch nur bei 5 der analysierten 1.000 Stichproben, d.h. in 0,5% der Beobachtungen festgestellt werden. Aufgrund der Symmetrieeigenschaft der theoretischen Wahrscheinlichkeitsverteilung von t^s müßte auch in 5 Fällen (0,5%) eine signifikant positive Überrendite ausgewiesen werden.

Bei Verwendung von Portefeullerrenditen zur Schätzung der Renditen der analysierten Wertpapiere sind nach den vorliegenden Ergebnissen der Tabelle 2 systematische Verzerrungen der Teststatistiken erkennbar. Erstens liegen die Ablehnungshäufigkeiten der Nullhypothese über den theoretisch erwarteten Werten. Wird zur Veranschaulichung erneut auf das erste Wertepaar von 4,1 und 0,0 zurückgegriffen, so sollte für jeweils 5 Stichproben, d.h. in insgesamt 10 Fällen, eine signifikant positive bzw. negative Überrendite festgestellt werden. Der tatsächlich beobachtete Wert von 41 liegt aber deutlich über dem erwarteten Wert von 10. Zweitens sind die Verzerrungen der Teststatistiken für positive und negative Überrenditen unterschiedlich. Es wird zu häufig eine Underperformance, aber zu selten eine Overperformance beobachtet. Für unser Beispiel wird in 36 Fällen zu häufig eine signifikante negative Überrendite ausgewiesen. Eine signifikant positive Überrendite müßte in 5 Stichproben festgestellt werden. Tatsächlich wird jedoch in keiner der 1.000 Stichproben ein signifikant positiver Durchschnittswert ermittelt. Drittens sind die Abweichungen der empirisch beobachteten und der theoretisch erwarteten Ablehnungshäufigkeiten in fast allen Fällen bei einem vorgegebenen Signifikanzniveau von einem Prozent statistisch signifikant.

Die beobachteten Verzerrungen der Teststatistiken dürften vor allem auf die in der Tabelle 1 nachgewiesenen Verzerrungen der Mittelwerte zurückzuführen sein. Weicht der Mittelwert aus den untersuchten 1.000 Stichproben statistisch signifikant vom theoretisch erwarteten Wert von null ab, so kann von dem schiefeangepaßten t-Test auch nicht erwartet werden, daß er die Nullhypothese akzeptiert. Werden die Ergebnisse der Teststatistiken bei einer Bereinigung mit gleichgewichteten Size-

Tabelle 2. Ablehnungshäufigkeiten für die Nullhypothese unter Verwendung parametrischer Testverfahren

Die simulierte Studie basiert auf 1.000 Stichproben mit jeweils 50 bzw. 150 Aktien. Für die einzelnen Untersuchungswertpapiere und Vergleichsportefeuilles bzw. -wertpapiere werden über einen Zeitraum von 36 Monaten Überrenditen nach Gleichung 3 berechnet. Zur Überprüfung der Nullhypothese werden die Teststatistiken bei Verwendung der Portefeuillerenditen mit dem schiefeangepaßten t-Test, bei Verwendung von einem nach der Marktkapitalisierung vergleichbaren Wertpapier mit dem t-Test berechnet. Die Abweichungen der empirisch beobachteten von den theoretisch erwarteten Ablehnungshäufigkeiten werden mit dem Binomialtest bei einem vorgegebenen Signifikanzniveau von 1 % überprüft. Signifikante Abweichungen werden mit * gekennzeichnet

Signifikanzniveau (%)	Ablehnungshäufigkeiten in %					
	1		5		10	
kumulierte Dichtefunktion (%)	0,5	99,5	2,5	97,5	5	95
Teil A: 1.000 Stichproben mit 50 Aktien						
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>						
gleichgewichtetes Portefeuille	4,1*	0,0*	9,2*	0,7*	12,5*	1,8*
marktwertgewichtetes Portefeuille	1,6*	0,5	2,8	4,2*	4,5	8,3*
<i>Berücksichtigung des Size-Effektes</i>						
gleichgewichtetes Size-Portefeuille	3,0*	0,1	6,6*	1,3*	11,0*	3,0*
marktwertgewichtetes Size-Portefeuille	3,2*	0,1	6,6*	1,1*	10,5*	3,0*
Buy-and-hold Size-Portefeuille	2,2*	0,1	4,4*	2,0	7,4*	4,7
Vergleichswertpapier	0,4	0,3	2,3	1,8	5,2	4,0
Teil B: 1.000 Stichproben mit 150 Aktien						
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>						
gleichgewichtetes Portefeuille	4,0*	0,1	9,6*	0,7*	15,5*	1,4*
marktwertgewichtetes Portefeuille	0,5	2,1*	1,6	9,3*	2,7*	17,0*
<i>Berücksichtigung des Size-Effektes</i>						
gleichgewichtetes Size-Portefeuille	3,0*	0,0*	7,5*	1,4	11,5*	3,6
marktwertgewichtetes Size-Portefeuille	2,7*	0,0*	6,9*	1,8	10,4*	4,2
Buy-and-hold Size-Portefeuille	1,3*	0,3	4,2*	2,6	7,4*	5,8
Vergleichswertpapier	0,6	0,0*	2,2	1,5	5,3	3,9

Portefeuilles und Vergleichswertpapieren mit annähernd identischen Mittelwerten (Teil A, Tabellen 1 und 2) gegenübergestellt, so ergeben sich Hinweise darauf, daß die verzerrten Teststatistiken sowohl durch die Lage als auch durch die asymmetrische Form der Verteilung verursacht werden.

Werden nach der Marktkapitalisierung vergleichbare Wertpapiere zur Berechnung der Überrenditen verwendet, so werden bei einer Stichprobengröße von 50 Aktien keine statistisch signifikanten Unterschiede festgestellt. Bei einer Stichprobengröße von 150 Aktien führt ein im Durchschnitt leicht negativer Mittelwert zu einer etwas niedrigeren Ablehnungshäufigkeit für positive Überrenditen, wenn ein zweiseitiger Signifikanztest bei einer vorgegebenen Irrtumswahrscheinlichkeit von einem bzw. fünf Prozent zugrundegelegt wird. Bei einem Signifikanzniveau von zehn Prozent ergeben sich keine signifikanten Abweichungen. Die Asymmetrie bei der Ablehnung der Nullhypothese in beiden Verteilungsenden ist deutlich geringer als bei der Verwendung von Vergleichsportefeuilles.

Nach den vorliegenden Untersuchungsergebnissen führt die asymmetrische Verteilung der langfristigen Überrenditen und die beobachteten Mittelwertverzerrungen (Lage der Verteilung) zu einem Fehler erster bzw. zweiter Art bei der Überprüfung der Nullhypothese. Die beobachteten Verzerrungen in der Lage der Überrenditenverteilung sind in erster Linie durch die Wahl eines geeigneteren Benchmarkmodells zu korrigieren (vgl. Abschnitt 4). Treten bei nicht signifikanten Durchschnittswerten immer noch Verzerrungen in den schiefeangepaßten t-Statistiken auf, so sind die für die Überprüfung der Nullhypothese erforderlichen Konfidenzintervalle mit einem Bootstrap-Verfahren zu schätzen.

Mittels eines Bootstrap-Verfahrens kann die angenommene theoretische Verteilungsfunktion durch eine empirische Verteilung ersetzt werden, welche für ein vorgegebenes Signifikanzniveau die Schätzung des Verwerfungsbereiches zur Ablehnung der Nullhypothese ohne eine spezielle Verteilungsannahme ermöglicht (vgl. Abschnitt 2.6). Die Ablehnungshäufigkeiten auf der Grundlage der Bootstrap-Verteilung sind in Tabelle 3 angegeben.

Die Testergebnisse auf der Grundlage der empirischen Verteilung zeigen nur in wenigen (mit * symbolisierten) Fällen statistisch signifikante Abweichungen zwischen den empirisch beobachteten und den theoretisch erwarteten Ablehnungshäufigkeiten. Zudem sind keine systematischen Unterschiede in den Ablehnungshäufigkeiten für positive und negative langfristige Überrenditen mehr erkennbar.

Für die Durchführung langfristiger Ereignisstudien kann man aus den vorliegenden Untersuchungsergebnissen die Empfehlung abzuleiten, die langfristigen, auf der Grundlage von Portefeuillerenditen berechneten durchschnittlichen Überrenditen mit nichtparametrischen Testverfahren zu analysieren. Die kritischen Werte zur Ablehnung der Nullhypothese können auf Basis von Bootstrap-Verteilungen hinreichend genau geschätzt werden. Die Untersuchungsergebnisse zeigen, daß bei diesen Verfahren die empirisch beobachteten Ablehnungshäufigkeiten für die Nullhypothese sehr nahe bei den theoretisch erwarteten Werten liegen.

Tabelle 3. Ablehnungshäufigkeiten für die Nullhypothese unter Verwendung nichtparametrischer Testverfahren

Die simulierte Studie basiert auf 1.000 Stichproben mit jeweils 50 bzw. 150 Aktien. Für die einzelnen Untersuchungswertpapiere und Vergleichsportefeuilles bzw. -wertpapiere werden über einen Zeitraum von 36 Monaten Überrenditen nach Gleichung 3 berechnet. Die kritischen Werte zur Ablehnung der Nullhypothese werden auf der Grundlage der schiefeangepaßten t-Wert-Verteilung bestimmt. Zur Simulation der Verteilung wird jedem Wertpapier eine andere Aktie aus dem gleichen Size-Portefeuille zugeordnet und die langfristige Überrendite auf identische Weise berechnet. Die Abweichungen der empirisch beobachteten von den theoretisch erwarteten Ablehnungshäufigkeiten werden mit dem Binomialtest bei einem vorgegebenen Signifikanzniveau von 1% überprüft. Signifikante Abweichungen werden mit * gekennzeichnet

Signifikanzniveau (%)	Ablehnungshäufigkeiten in%					
	1		5		10	
kumulierte Dichtefunktion (%)	0,5	99,5	2,5	97,5	5	95
Teil A: 1.000 Stichproben mit 50 Aktien						
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>						
gleichgewichtetes Portefeuille	0,8	0,3	3,0	1,9	5,4	5,1
marktwertgewichtetes Portefeuille	1,0	0,5	2,7	3,0	4,7	4,2
<i>Berücksichtigung des Size-Effektes</i>						
gleichgewichtetes Size-Portefeuille	0,4	0,1	2,7	2,0	4,7	5,2
marktwertgewichtetes Size-Portefeuille	0,4	0,1	2,8	1,8	4,8	5,0
Buy-and-hold Size-Portefeuille	0,4	0,3	2,5	2,7	4,4	5,4
Teil B: 1.000 Stichproben mit 150 Aktien						
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>						
gleichgewichtetes Portefeuille	0,9	0,9	2,3	3,5	6,0	6,2
marktwertgewichtetes Portefeuille	0,6	0,7	1,9	3,6	5,0	7,6*
<i>Berücksichtigung des Size-Effektes</i>						
gleichgewichtetes Size-Portefeuille	0,8	0,5	3,0	4,3*	6,0	7,0*
marktwertgewichtetes Size-Portefeuille	0,8	0,4	3,1	4,2*	6,1	6,7
Buy-and-hold Size-Portefeuille	1,1	0,4	3,1	3,0	5,8	6,8*

Tabelle 4. Mittelwert der Überrenditenverteilung und Ablehnungshäufigkeiten für die Nullhypothese einer simulierten langfristigen Ereignisstudie bei Vorliegen eines kalenderzeitlichen Clusterings

Die simulierte Studie basiert auf 1.000 Stichproben mit jeweils 50 Aktien. Das Kalenderclustering wird über eine Dreiecksverteilung mit einer Spannweite von 73 Monaten (Monat 0 ± 36 Monate) modelliert. Es erfolgen 1.000 gleichverteilte Ziehungen der Mittelpunkte der Dreiecksverteilung mit jeweils 50 Ziehungen unter der Dreiecksverteilung. Für die einzelnen Untersuchungswertpapiere und Vergleichsportefeuilles bzw. -wertpapiere werden über einen Zeitraum von 36 Monaten Überrenditen nach Gleichung 3 berechnet. Zur Simulation der Verteilung wird jedem Wertpapier eine andere Aktie aus dem gleichen Size-Portefeuille zugeordnet und die langfristige Überrendite auf identische Weise berechnet. *, **, *** kennzeichnen signifikante Durchschnittswerte auf dem 10, 5, 1% -Niveau (zweiseitiger Test). Die kritischen Werte zur Ablehnung der Nullhypothese werden auf der Grundlage der schiefeangepaßten t-Wert-Verteilung (Teil A) bzw. auf Basis einer empirischen Bootstrap-Verteilung (Teil B) bestimmt. Bei einer Bereinigung mit Vergleichswertpapieren wird aufgrund der symmetrischen Verteilung der t-Test verwendet. Die Abweichungen der empirisch beobachteten von den theoretisch erwarteten Ablehnungshäufigkeiten werden mit dem Binomialtest bei einem vorgegebenen Signifikanzniveau von 1% überprüft. Signifikante Abweichungen werden mit * gekennzeichnet

Signifikanzniveau (%)	Mittelwert (%)	Ablehnungshäufigkeiten in%					
		1		5		10	
kumulierte Dichtefunktion (%)		0,5	99,5	2,5	97,5	5	95
Teil A: Parametrische Teststatistiken (t-Test bzw. schiefeangepaßter t-Test)							
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>							
gleichgewichtetes Portefeuille	-1,40***	7,0*	0,1	13,4*	2,1	17,6*	4,2
marktwertgewichtetes Portefeuille	5,20***	5,3*	15,7*	9,5*	23,3*	13,5*	29,0*
<i>Berücksichtigung des Size-Effektes</i>							
gleichgewichtetes Size-Portefeuille	-0,56	6,5*	0,3	12,9*	2,4	16,9*	5,5
marktwertgewichtetes Size-Portefeuille	-0,19	5,5*	0,4	11,6*	2,2	15,3*	5,4
Buy-and-hold Size-Portefeuille	0,68*	2,5*	0,1	6,4*	2,2	9,5*	5,3
Vergleichswertpapier	-0,95*	0,4	0,3	3,0	2,2	5,7	4,4
Teil B: Nichtparametrische Teststatistiken auf Basis empirischer Verteilungen							
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>							
gleichgewichtetes Portefeuille		0,0*	0,7	2,2	3,0	5,3	5,4
marktwertgewichtetes Portefeuille		0,2	1,1	1,9	2,2	4,4	4,6

Table 4. (Fortsetzung)

Signifikanzniveau (%)	Mittelwert (%)	Ablehnungshäufigkeiten in%					
		1		5		10	
kumulierte		0,5	99,5	2,5	97,5	5	95
Dichtefunktion (%)							
<i>Berücksichtigung des Size-Effektes</i>							
gleichgewichtetes		0,1	0,5	1,3*	2,5	5,3	5,1
Size-Portefeuille							
marktwertgewichtetes		0,1	0,6	1,9	2,5	5,1	5,0
Size-Portefeuille							
Buy-and-hold		0,4	1,0	2,5	2,6	6,0	5,9
Size-Portefeuille							

3.2. Untersuchungen bei Vorliegen einer zeitlichen Häufung von Beobachtungen

Die im Rahmen von langfristigen Studien zu analysierenden Ereignisse können in einzelnen Jahren des Untersuchungszeitraumes gehäuft auftreten. Eines der bekanntesten Beispiele hierfür sind die von Ritter (1984) beobachteten sogenannten „Hot-issue“-Phasen, die eine deutlich höhere Anzahl von Börseneinführungen in einzelnen Jahren und Branchen bezeichnen. Dies hat zur Folge, daß bei langfristigen Ereignisstudien eine verstärkte Überschneidung der Ereignisfenster entsteht und somit die einzelnen Beobachtungen innerhalb des Untersuchungssamples nicht mehr unabhängig voneinander sind. Welche Auswirkungen sich daraus auf die Untersuchungsergebnisse ergeben, wird mit der nachfolgenden Simulation analysiert.

Für die Simulation einer zeitlichen Häufung von Ereignissen (Kalenderclustering) wird eine Dreiecksverteilung (Simpson-Verteilung) mit einer Spannweite von 73 Monaten (Monat 0 ± 36 Monate) gewählt, deren Mittelpunkt (Scheitelpunkt) über ein Intervall von Januar 1963 bis Dezember 1989 geht. Es werden 1.000 gleichverteilte Ziehungen der Mittelpunkte mit jeweils 50 Ziehungen von einzelnen Aktien unter der Dreiecksverteilung vorgenommen. Mit Ausnahme der skizzierten Variation der Verteilungsannahme bei der Stichprobenbildung wird die gleiche Untersuchungsmethodik gewählt, die den Berechnungen in Abschnitt 3.1 zugrundeliegen.

Im Unterschied zu den Simulationsergebnissen ohne Vorliegen eines kalenderzeitlichen Clusterings sind deutlichere Verzerrungen der aus den simulierten 1.000 Stichproben berechneten Mittelwerte zu beobachten. Mit Ausnahme der gleich- und marktwertgewichteten Size-Portefeuilles als Benchmarkmodell sind statistisch signifikante Mittelwerte festzustellen. Als Folge dieser Verzerrungen entstehen Abweichungen der empirisch beobachteten von den theoretisch erwarteten Ablehnungshäufigkeiten. Die eingesetzten Verfahren, die marktwertbedingte Renditeunterschiede berücksichtigen, führen in Verbindung mit der Bestimmung der Ablehnungshäufigkeiten für die Nullhypothese auf Basis empirischer Verteilungen auch bei Vorliegen eines kalenderzeitlichen Clusterings zu den genaueren Ergebnissen der Teststatistiken.

Tabelle 5. Mittelwert der Überrenditenverteilung und Ablehnungshäufigkeiten für die Nullhypothese einer simulierten langfristigen Ereignisstudie unter Verwendung verschiedener Vergleichswertpapiere zur Schätzung erwarteter Renditen

Die simulierte Studie basiert auf 1.000 Stichproben mit jeweils 50 Aktien. Für die einzelnen Untersuchungswertpapiere und Vergleichswertpapiere werden über einen Zeitraum von 36 Monaten Überrenditen nach Gleichung 3 berechnet. Als Erwartungswert werden Aktienrenditen von Unternehmen mit vergleichbarer Marktkapitalisierung gewählt, d.h. die Aktien mit den fünf nächsthöheren bzw. nächstniedrigeren Marktwerten. Mit dem t-Test für abhängige Stichproben wird überprüft, ob sich der Mittelwert der Stichprobenmittelwert-Verteilung signifikant von null unterscheidet. *, **, *** kennzeichnen signifikante Durchschnittswerte auf 10, 5, bzw. 1% -Niveau. Die Abweichungen der empirisch beobachteten von den theoretisch erwarteten Ablehnungshäufigkeiten werden mit dem Binomialtest bei einem vorgegebenen Signifikanzniveau von 1% überprüft. Signifikante Abweichungen werden mit * gekennzeichnet.

Signifikanzniveau (%)	Mittelwert (%)	Ablehnungshäufigkeiten in%					
		1		5		10	
kumulierte Dichtefunktion (%)		0,5	99,5	2,5	97,5	5	95
<i>Vergleichswertpapiere mit höherer Marktkapitalisierung (steigend MF1 → MF5)</i>							
MF1	0,24	0,4	0,3	2,0	2,5	5,0	5,3
MF2	1,06**	0,4	0,6	1,4*	2,8	3,5	4,7
MF3	1,15**	0,1	0,2	1,6	1,8	4,1	5,1
MF4	1,56**	0,4	0,2	1,3*	1,9	4,3	4,4
MF5	1,93***	0,4	0,1	1,7	1,8	4,2	4,5
<i>Vergleichswertpapiere mit niedrigerer Marktkapitalisierung (fallend MF1 → MF5)</i>							
MF1	-0,40	0,5	0,2	2,2	2,4	4,4	3,6
MF2	-1,12**	0,6	0,1	2,2	2,0	4,8	4,8
MF3	-1,27**	0,3	0,4	2,7	2,0	6,2	4,8
MF4	-1,63***	0,2	0,2	2,3	2,4	5,2	4,7
MF5	-1,55***	0,4	0,3	1,4	1,7	4,9	4,4

3.3. Marktwertunterschiede zwischen Untersuchungs- und Vergleichswertpapieren

In diesem Abschnitt soll die Eignung von Wertpapieren mit vergleichbarer Marktkapitalisierung als Benchmarkmodell getestet und der Einfluß des Size-Effektes auf die Durchschnittswerte und Teststatistiken langfristiger Überrenditen analysiert werden. Dazu werden in einem ersten Schritt für die zufällig ausgewählten Aktien fünf Vergleichswertpapiere mit einem höheren Marktwert ermittelt. Das erste Vergleichswertpapier ist die Aktie mit dem nächsthöheren Marktwert. Es folgt das Wertpapier mit dem zweithöheren Marktwert. Der Auswahlmechanismus wird fortgesetzt, bis fünf Vergleichswertpapiere ausgewählt wurden. In einem zweiten Schritt werden dann fünf Wertpapiere mit niedrigerem Marktwert bestimmt.

Nach den in der Tabelle 5 dargestellten Ergebnissen führt die Verwendung von Vergleichswertpapieren mit dem nächsthöheren bzw. nächstniedrigeren Markt-

wert zu einem Schätzwert für die Aktienrendite der betrachteten Aktien nach 36 Monaten, der sehr nahe am theoretisch erwarteten Durchschnittswert von null liegt. Die beobachteten Mittelwertunterschiede der Durchschnittsrenditen von Untersuchungs- und Vergleichswertpapieren in Höhe von 0,24 bzw. -0,40 Prozentpunkten sind statistisch nicht signifikant.

Andererseits führen schon sehr geringe systematische Unterschiede in der Höhe der Marktkapitalisierung, wie z.B. die Verwendung von Vergleichswertpapieren mit dem zweithöheren bzw. zweitniedrigeren Marktwert, zu statistisch signifikanten Mittelwertunterschieden für Betrachtungszeiträume von 36 Monaten. Auffallend an den Simulationsergebnissen ist die klare Struktur, mit welcher der Size-Effekt zu beobachten ist. Die Vorzeichen der berechneten Mittelwerte entsprechen für alle zehn Simulationsläufe den Erwartungen bei Existenz eines Size-Effektes. Zudem werden die statistisch von null signifikanten Mittelwertunterschiede kontinuierlich größer, wenn die Unterschiede in der Marktkapitalisierung von Untersuchungs- und Vergleichswertpapieren zunehmen.

Die ebenfalls in der Tabelle 5 dargestellten Ablehnungshäufigkeiten liegen für die einbezogenen Vergleichswertpapiere nahe bei den theoretisch erwarteten Werten, wenn der t-Test für abhängige Stichproben Verwendung findet. Die mit zunehmenden Unterschieden in der Marktkapitalisierung zwischen Untersuchungs- und Vergleichswertpapieren beobachteten Veränderungen der Mittelwerte wirken sich jedoch kaum auf die empirisch ermittelten Ablehnungshäufigkeiten aus.

Nach den bisherigen Betrachtungen scheint die Verwendung der Renditen von Vergleichswertpapieren gegenüber Size-Portefeuillrenditen zur Schätzung von Erwartungswerten zu keinen methodischen Verbesserungen zu führen. Für die Verwendung von Size-Portefeuilles bei kleinen Stichproben spricht vor allem, daß der Durchschnittswert der Stichprobenmittelwerte statistisch und ökonomisch nicht von null verschieden und die Standardabweichung der einzelnen Stichprobenmittelwerte deutlich geringer ist als bei der Verwendung von Vergleichswertpapieren. Die Standardabweichung der Stichprobenmittelwerte liegt nach den Angaben von Tabelle 1 für Stichprobengrößen von 50 Aktien bei Verwendung von Vergleichswertpapieren um ca. 50 Prozent über dem Wert, der sich bei einer Schätzung der erwarteten Rendite der untersuchten Aktien mit Hilfe von Vergleichsportefeuilles ergibt. Bei einem solchen Unterschied in der Standardabweichung kann die Verwendung von Vergleichswertpapieren mit dem nächsthöheren bzw. nächstniedrigeren Marktwert lediglich als Ergänzungsrechnung für die Durchführung von Ereignisstudien mit nicht simulierten Daten empfohlen werden.

4. Empfehlungen zur Benchmarkwahl bei nicht simulierten Daten

In Abschnitt 3 wurden verschiedene Benchmarkmodelle dahingehend analysiert, ob der Schätzwert für die erwartete Rendite der zu untersuchenden Aktien die mit der Marktkapitalisierung in Zusammenhang stehenden Renditeunterschiede adäquat widerspiegelt. Um jedoch Empfehlungen für eine unverzerrte Benchmarkwahl zur Durchführung langfristiger Ereignisstudien geben zu können, sind weitere Anforderungen an die verwendete Simulationsmethode zur Auswahl einer Benchmark zu stellen.

Dies betrifft erstens die in der Literatur gefundenen Hinweise, daß es sich bei dem Size-Effekt um ein zeitlich variierendes Phänomen handelt. So ist zwar nach dem heutigen Stand der empirischen Literatur im langfristigen Durchschnitt von einer positiven Risikoprämie für eine niedrigere Marktkapitalisierung auszugehen, für einzelne Zeitperioden kann es sich jedoch genau umgekehrt verhalten (vgl. Beiker, 1993; Oertmann, 1994; Stehle, 1997).

Zweitens wurden in der Literatur weitere empirische Bestimmungsfaktoren identifiziert, welche Renditeunterschiede zwischen einzelnen Aktien erklären können. Zu ihnen zählen beispielsweise der Quotient aus dem Buch- und dem Marktwert einer Aktie, die Dividendenrendite, der Verschuldungsgrad, Renditen der vorhergehenden oder früherer Perioden, das Kurs-Gewinn- und das Cash-flow-Kurs-Verhältnis (vgl. Stehle et al., 1998; Wallmeier, 2000).

Bei der Durchführung einer (langfristigen) Ereignisstudie werden zunächst innerhalb des festgelegten Betrachtungszeitraumes alle stattgefundenen Ereignisse identifiziert. Gegebenenfalls werden weitere Kriterien herangezogen, welche die Stichprobe zum Beispiel auf börsennotierte Unternehmen eines bestimmten Marktsegmentes und/oder einer bestimmten Börse beschränkt.

In einer Vielzahl von Fällen dürften die ermittelten Ereignisse keiner Gleichverteilung im Beobachtungszeitraum unterliegen. Aufgrund eines sich im Zeitablauf verändernden Size-Effektes wird deshalb für jedes identifizierte Ereignis per Zufallsauswahl zum gleichen Zeitpunkt eine andere Aktie ausgewählt. Jede der 1.000 gezogenen Zufallsstichproben ist somit in der zeitlichen Verteilung und in der Anzahl der Elemente mit der Stichprobe der durchzuführenden Ereignisstudie identisch.

Zur Illustration der (modifizierten) Simulationsmethode verwenden wir die zeitliche Verteilung von 79 Börseneinführungen in den amtlichen Handel der Frankfurter Wertpapierbörse zwischen 1960 und 1992. Die nachfolgende Tabelle zeigt die Ergebnisse der 1.000 Zufallsstichproben mit jeweils 79 ausgewählten Aktien. Die Verteilung der Stichprobe ist durch ein besonders ausgeprägtes zeitliches Clustering der Ereignisse ab 1983 zu charakterisieren.

Nach den Ergebnissen der Tabelle 6 sind die berechneten Durchschnittswerte statistisch hoch signifikant. Insbesondere führt die Verwendung eines marktwertgewichteten Portefeuilles als Benchmark zu den am stärksten verzerrten Berechnungsergebnissen. Gleich- bzw. marktwertgewichtete Size-Portefeuilles als Benchmark weisen nach den vorliegenden Simulationsergebnissen die erwartete Rendite der Untersuchungswertpapiere tendenziell zu niedrig aus (positive Überrenditen). Mit der als Buy-and-hold Size-Portfeuille bezeichneten Benchmark wird im Durchschnitt eine erwartete Rendite geschätzt, die ca. einen Prozentpunkt über den tatsächlich zu erwartenden Werten liegt. Aus den Simulationsergebnissen können in unserem Beispiel für die durchzuführende Ereignisstudie folgende Empfehlungen abgeleitet werden:

- Die Ergebnisse der durchzuführenden Ereignisstudie bei Verwendung von gleich- und marktwertgewichteten Size-Portefeuilles und von Buy-and-hold Size-Portefeuilles als Benchmark können als obere und untere Grenze der Berechnungsergebnisse interpretiert werden. Für die Signifikanztests sind die mit

Tabelle 6. Mittelwerte, Ablehnungshäufigkeiten und Konfidenzintervallschätzungen für eine Simulationsverteilung, die der zeitlichen Verteilung der Börseneinführungen in den amtlichen Handel Frankfurt 1960–1992 entspricht

Die simulierte Studie basiert auf 1.000 Stichproben mit jeweils 79 Aktien. Für die einzelnen Untersuchungswertpapiere und Vergleichsportefeuilles werden über einen Zeitraum von 36 Monaten Überrenditen nach Gleichung 3 berechnet. Zur Simulation der Verteilung wird jedem Wertpapier eine andere Aktie aus dem gleichen Size-Portefeuille zugeordnet und die langfristige Überrendite auf identische Weise berechnet. *, **, *** kennzeichnen signifikante Durchschnittswerte auf dem 10, 5, 1% -Niveau (zweiseitiger Test). Die Ablehnungshäufigkeiten werden auf der Grundlage des schiefeangepaßten t-Testes bestimmt. Die Abweichungen der empirisch beobachteten von den theoretisch erwarteten Ablehnungshäufigkeiten werden mit dem Binomialtest bei einem vorgegebenen Signifikanzniveau von 1 % überprüft. Signifikante Abweichungen werden mit * gekennzeichnet

Signifikanzniveau (%)	Mittelwert (%)	Ablehnungshäufigkeiten in%					
		1		5		10	
kumulierte		0,5	99,5	2,5	97,5	5	95
Dichtefunktion (%)							
Teil A: Mittelwerte und Teststatistiken (schiefeangepaßter t-Test)							
<i>Portefeuilles aus allen in Frankfurt amtlich gehandelten Aktien</i>							
gleichgewichtetes	1,91***	0,4	0,6	3,0	4,0*	4,7	7,5*
Portefeuille							
marktwertgewichtetes	7,70***	0,0*	3,5*	0,1*	23,3*	12,0*	23,3*
<i>Berücksichtigung des Size-Effektes</i>							
Portefeuille							
gleichgewichtetes	2,25***	0,6	0,6	2,4	3,9*	3,9	7,4*
Size-Portefeuille							
marktwertgewichtetes	2,49***	0,5	0,6	2,3	4,0*	3,8	7,8*
Size-Portefeuille							
Buy-and-hold	-1,07***	1,7*	0,1	6,0*	1,5	9,0*	2,9*
Size-Portefeuille							
Teil B: Konfidenzintervallschätzung							
gleichgewichtetes		-2,787	2,736	-2,007	2,213	-1,534	1,876
Size-Portefeuille							
marktwertgewichtetes		-2,774	2,721	-1,930	2,233	-1,490	1,925
Size-Portefeuille							
Buy-and-hold		-3,422	2,242	-2,439	1,752	-2,128	1,386
Size-Portefeuille							

dem Simulationsverfahren ermittelten und im zweiten Teil der Tabelle 6 angegebenen Konfidenzintervalle zu verwenden.

- Die beobachteten signifikanten Abweichungen der Durchschnittswerte vom Wert null könnten auf den Einfluß anderer Renditeanomalien hindeuten. Das gleiche Simulationsverfahren sollte alternativ für andere Benchmarkmodelle

(z.B. Buchwert-Marktwert-Portefeuilles) durchgeführt werden. Im Ergebnis dieser Analysen werden jene Benchmarks ausgewählt, die sehr nahe am theoretisch erwarteten Wert von null liegen.

Die vorgeschlagene Methodik der Benchmarkwahl ermöglicht es, die bisher in Ereignisstudien häufig angegebenen Untersuchungsergebnisse für (zahlreiche) alternativ verwendete Benchmarks (vgl. z.B. Sapusek, 2000) auf jene zu reduzieren, die unter der Nullhypothese und der beobachteten zeitlichen Verteilung der Ereignisse zu weitestgehend unverzerrten Berechnungsergebnissen führen.

Die in diesem Abschnitt abgeleiteten Empfehlungen für die Durchführung langfristiger Ereignisstudien mit nicht simulierten Daten reduzieren auftretende Verzerrungen der Berechnungsergebnisse bei einem sich im Zeitablauf verändernden Size-Effekt. Die vorgeschlagene Methodik ist auch allgemein auf andere oder auf Kombinationen von mehreren Renditeanomalien anwendbar und verdeutlicht die Vorteile von Simulationstechniken gegenüber anderen Verfahren. Der Nachteil des Verfahrens liegt im hohen Rechenaufwand. Eine Umsetzung der Methode ist nur bei Zugriff auf größere Datenbanken möglich.

5. Schlußfolgerungen

Nach den bisher vorliegenden Erkenntnissen bereiten bei der Berechnung langfristiger Überrenditen vor allem die Wahl der Benchmark zur Schätzung der erwarteten Rendite einer Aktie, bestehende Querschnittskorrelationen als Folge überlappender Ereignisfenster, eine Konzentration von Ereignissen in einzelnen Zeitperioden oder in bestimmten Branchen sowie eine zum Teil beträchtliche Schiefe der Überrenditenverteilung enorme Probleme bei der Berechnung des Mittelwertes und der dazugehörigen Teststatistiken. Die hohe Standardabweichung sowohl der Überrenditen einzelner Aktien als auch der Stichprobenmittelwerte erschweren erheblich die ökonomische Interpretation der Ergebnisse und die Anwendung langfristiger Ereignisstudien zur Analyse ökonomisch relevanter Fragestellungen.

Besonders gravierend sind die beobachteten Verzerrungen der Ergebnisse, wenn die Renditen der untersuchten Aktien mit den Renditen eines gleich- bzw. marktwertgewichteten Portefeuilles aus allen im amtlichen Handel der Frankfurter Wertpapierbörse notierten Aktien verglichen werden. Dieses Ergebnis unserer Simulationsstudie läßt erwarten, daß die in Deutschland verfügbaren (marktwertgewichteten) Aktienindizes (DAX, DAFOX etc.) in den meisten Fällen nicht als Schätzwert für die langfristige Rendite einzelner Aktien geeignet sind.

Auch eine Bereinigung mit der Rendite einzelner Vergleichswertpapiere ist lediglich als Ergänzungsrechnung zu empfehlen. Der schwerwiegendste Nachteil dieses Verfahrens ist in der höheren Standardabweichung der Stichprobenmittelwerte zu sehen. Positive und negative Abweichungen in Höhe mehrerer Prozentpunkte vom „wahren“ Durchschnittswert können insbesondere bei kleineren Stichproben sowie bei einer zeitlichen Häufung von Ereignissen mit einer höheren Wahrscheinlichkeit auftreten und zu fehlerhaften Interpretationen der Untersuchungsergebnisse führen. Ein weiterer Nachteil ist, daß die berechneten Überrenditen für die einzelnen Untersuchungswertpapiere nicht in Querschnittsuntersuchungen verwendet werden

können, da ein Einfluß unternehmensspezifischer Ereignisse der Vergleichswertpapiere gegeben ist.

Auf der Grundlage der durchgeführten Simulationen empfehlen wir Vergleichsportefeuilles als Benchmark zu verwenden. Die Konstruktion der Vergleichsportefeuilles sollte auf Basis empirisch identifizierter Bestimmungsfaktoren erfolgen, die einen Beitrag zur Erklärung von Renditeunterschieden und somit der Preisbildung auf den Aktienmärkten leisten. Für die Auswahl eines Benchmarkmodells eignen sich Simulationen, um unter alternativen Modellen jenes auszuwählen, welches unter der Nullhypothese unverzerrte Schätzergebnisse liefert. Da der Erklärungsgehalt der empirisch identifizierten Bestimmungsfaktoren im Zeitablauf variieren kann, ist sicherzustellen, daß die simulierten und die im Rahmen einer langfristigen Ereignisstudie zu analysierende Stichprobe eine identische zeitliche Verteilung besitzen. Die für die Überprüfung der Nullhypothese erforderlichen Konfidenzintervalle sind auf Basis einer Bootstrap-Verteilung zu schätzen, da eine Bereinigung von Einzelmit Portefeuillerenditen zu einer schiefen Überrenditenverteilung führt.

References

- Barber BM, Lyon JD (1997) Detecting long-run abnormal stock returns: The empirical power and specification of tests statistics. *Journal of Financial Economics* 43 : 341–372
- Beiker H (1993) Überrenditen und Risiken kleiner Aktiengesellschaften. Müller Botermann, Köln
- Blume ME, Stambaugh RF (1983) Biases in computed returns: An application to the size effect. *Journal of Financial Economics* 12: 387–404
- Brav A (2000) Inference in long-horizon event studies: a re-evaluation of the evidence. *Journal of Finance* 55: 1979–2016
- Brown SJ, Warner JB (1980) Measuring security price performance. *Journal of Financial Economics* 8: 205–258
- Canina L, Michaely R, Thaler R, Womack K (1998) Carveat compounder: a warning about using the daily CRSP equal-weighted index to compute long-run excess returns. *Journal of Finance* 53: 403–416
- Cowan AR, Sergeant AMA (2001) Interacting biases, non-normal return distributions and the performance of tests for long-horizon event studies. *Journal of Banking & Finance* 25: 741–765.
- Cusatis PJ, Miles JA, Woolridge JR (1993) Restructuring through spinoffs: The stock market evidence. *Journal of Financial Economics* 33: 293–311
- Dimson E, Marsh P (1986) Event study methodologies and size effect: the case of UK press recommendations. *Journal of Financial Economics* 17: 113–142
- Domke HM (1987) Rendite und Risiko von Aktien kleiner Börsengesellschaften. Lang, Frankfurt am Main
- Efron B (1982) The Jackknife, the bootstrap and other resampling plans. Pennsylvania
- Fama EF (1998) Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49: 283–306
- Hall P (1992) On the removal of skewness by transformation. *Journal of the Royal Statistical Society* 54: 221–228
- Hoel PG (1984) Introduction to mathematics statistics. Wiley, New York
- Ikenberry D, Lakonishok J, Vermaelen T (1995) Market underreaction to open market shares repurchases. *Journal of Financial Economics* 39: 181–208

- Johnson NJ (1978) Modified t tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association* 73: 536–544
- Kothari SP, Shanken J, Sloan RG (1995) Another look at the cross-section of expected returns. *Journal of Finance* 50: 185–224
- Kothari SP, Warner JB (1997) Measuring long-horizon security price performance. *Journal of Financial Economics* 43: 301–339
- Loughran T, Ritter JR (1995) The new issues puzzle. *Journal of Finance* 50: 23–51
- Lyon JD, Barber BM, Tsai CL (1999) Improved methods for tests of long-run abnormal stock returns. *Journal of Finance* 54: 165–201
- Marsaglia G, Zaman A (1991) A new class of random number generators. *Annals of Applied Probability* 1: 462–480
- Michaely R, Thaler RH, Womack KL (1995) Price reactions to dividend initiations and omissions. *Journal of Finance* 50: 573–608
- Mitchell ML, Stafford E (1997) Managerial decisions and long-term stock price performance. Working paper, University of Chicago
- Oertmann P (1994) Firm-Size-Effekt am deutschen Aktienmarkt. *Zeitschrift für betriebswirtschaftliche Forschung* 14: 229–259
- Ritter JR (1991) The long-term performance of initial public offerings. *Journal of Finance* 46: 3–27
- Ritter JR (1984) The “hot issue” market of 1980. *Journal of Business* 57: 215–240
- Roll R (1983) On computing mean returns and the small firm premium. *Journal of Financial Economics* 12: 371–386
- Sapusek A (2000) Benchmark-sensitivity of IPO long-run performance: an empirical study of Germany. *Schmalenbach Business Review* 53: 374–407
- Sattler RR (1994) Renditeanomalien am deutschen Aktienmarkt. Shaker, Aachen
- Schnittke J (1989) Überrenditeeffekte am deutschen Aktienmarkt. Müller Botermann, Münster
- Shumway T (1997) The delisting bias in CRSP Data. *Journal of Finance* 52: 327–340
- Stehle R (1997) Der Size-Effekt am deutschen Aktienmarkt. *Zeitschrift für Bankrecht und Bankwirtschaft* 9: 237–260
- Stehle R, Ehrhardt O (1999) Renditen bei Börseneinführungen am deutschen Kapitalmarkt. *Zeitschrift für Betriebswirtschaft* 69: 1395–1422
- Stehle R, Ehrhardt O, Przyborowsky R (2000) Long-run stock performance after initial public offerings and seasoned equity issues in the German capital market. *European Financial Management* 6: 173–196
- Stehle R, Sattler RR, Wulff C (1998) Der Size-Effekt am US-amerikanischen Aktienmarkt. Working paper, Humboldt-Universität zu Berlin
- Sutton CD (1993) Computer-intensive methods for tests about the mean of an asymmetrical distribution. *Journal of the American Statistical Association* 88: 802–808
- Wallmeier M (2000) Determinanten erwarteter Renditen am deutschen Aktienmarkt – Eine empirische Untersuchung anhand ausgewählter Kennzahlen. *Zeitschrift für betriebswirtschaftliche Forschung* 52: 27–57
- Wallmeier M (1997) Prognose von Aktienrenditen und –risiken mit Mehrfaktorenmodellen.
- Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics* 1: 80–83
- Wolfram S (1986) Random sequence generation by cellular automata. *Advance in Applied Mathematics* 7: 123–169