Econometric Analysis of Financial Risk and Correlation

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Ulrich-Michael Homm
aus Mühlbach, Rumänien

Bonn 2012

Dekan:	Prof. Dr. Klaus Sandmann
Erstreferent:	Prof. Dr. Jörg Breitung
Zweitreferent:	JProf. Dr. Christian Pigorsch

Tag der mündlichen Prüfung: 13.09.2012

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn

(http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

Acknowledgments

Many people have contributed to this thesis and I owe them an enormous amount of gratitude. First and foremost, I would like to thank my main advisor Jörg Breitung for his guidance, encouragement and steady support. I am very grateful to have jointly worked with him on my first research project, which paved the way for this thesis.

Many thanks also go to my second advisor Christian Pigorsch and his invaluable contribution to the second and third Chapter of this dissertation. Working with him was very fruitful and never lacking a portion of good humor.

After the completion of this dissertation the last Chapter is being further developed in joint work with Matei Demetrescu, whose insights will greatly enhance future versions. I am greatly indebted to him and Norbert Christopeit for their support concerning technical issues.

Last but not least, I would like to thank my fellow PhD students at the Bonn Graduate School of Economics, especially my office mates Thomas Nebeling, Christoph Roling, and Volker Tjaden, who enriched my academic as well as non-academic experiences during my four years as a PhD students. Thank you!

Contents

List of Figures

List of Tables

Introduction

In recent years the world's major financial markets have experienced dramatic downward movements threatening the welfare of many societies. Similar to other episodes of financial turmoil, the recent financial drawdown was preceeded by astonishing price increases, often termed as bubbles, especially in the US housing market. Other factors that played a role in the financial crisis were, among others, newly developed financial products and inappropriate risk assessment (cf. Gennaioli, Shleifer, and Vishny, 2010).

The contribution of this dissertation is threefold. First, econometric procedures to test for the occurence of asset price bubbles ex post and in real time are proposed. Real time monitoring procedures represent an additional tool for financial agents to gauge whether or not a bubble is building up in a financial market at the date of measurement. Second, we consider the problem of risk assessment and performance measurement. Risk assessment is essential for determining the amount of required capital. It is also important to counterbalance expected profits. The focus here lies on the *economic index of riskiness* proposed by Aumann and Serrano (2008). New theoretical properties of the index are established and estimation techniques are proposed. It is brought to application as a counterweight to expected returns to measure the perfomance of mutual funds and hedge funds.

While the previous approaches are designed for application to financial markets data, the last part of the dissertation is of more basic econometric interest. It addresses the issue of the validity of standard inference procedures in fixed effect panel data models. Ordinary least squares inference about model parameters can be misleading if shocks to cross-sectional units are correlated. Existing tests of cross-section error dependence aim at determining whether or not there is cross-section error correlation *per se*. In this dissertation, a procedure is developed that aims at testing whether there is cross-section error correlation that invalidates ordinary least squares inference. The remainder of the introduction gives a more detailed description of the respective chapters.

Chapter 1, joint work with Jörg Breitung, proposes and compares several tests for speculative bubbles in stock markets. The tests build on a simple asset pricing model, where the fundamental price of an asset is determined by the expected value of discounted future dividend payments. Under the null hypothesis of no bubble, asset prices follow a random walk or a random walk with drift. Under the alternative hypothesis that a bubble is present, stock prices show exponentially increasing or explosive behavior. As explosive behavior typically is only present in a part of the time series of a certain asset price, the tests are designed as structral break tests, i.e. they test for a change from a random walk to an explosive process. Moreover, in most cases one has no a priori knowledge of the date of the structural change, which suggests the use of sequential testing procedures.

Bubbles seem to be recurring phenomena in financial markets and econometric testing that pins down their existence has been suggested before. We mainly refer to the work of Phillips, Wu, and Yu (2011), who proposed a forward recursive Dickey-Fuller (DF) test. As competitors we suggest a Chow-type DF-test and adapted versions of tests that stem from the literature on testing the random walk hypothesis against stationary alternatives or vice versa. Complementing these tests we suggest several estimators for the date at which the structural change occurs. In simulation experiments we find that the Chow-type DF-test and the pertaining break date estimator show a very competitive performance. While these tests and estimators are conceived to analyze bubble episodes ex post, we also propose monitoring procedures to detect ongoing bubbles. In the empirical section several bubbles in stock price indices and also in housing prices are detected ex post.

Chapters 2 and 3, joint work with Christian Pigorsch, are concerned with risk and performance measurement. At the core of these chapters lies the *economic index of riskiness* of Aumann and Serrano (2008). Chapter 2 makes use of the one-to-one relationship between the index of riskiness and the *adjustment coefficient* from ruin theory. Existence conditions for general random variables or financial returns are established for the index of riskiness. Furthermore, we give an approximate operational interpretation of the index, which states that it represents the minimum level of required capital ensuring no bankruptcy with a certain probability. CHAPTER 3 uses the index of riskiness to construct a performance measure that generalizes the commonly used Sharpe ratio (Sharpe, 1966). In analogy to the *economic index of riskiness* the proposed performance measure is referred to as the *economic performance measure* (EPM). It is equivalent to the Sharpe ratio when returns are normally distributed. Generalizing the continuity result of Aumann and Serrano (2008) it can also be shown that the EPM converges to a Sharpe ratio equivalent measure as the underlying return distributions converge to the normal distribution. In contrast to the Sharpe ratio, however, the EPM takes into account higher order moments, which the typical investor is likely to care about. This can have sizeable effects on the ranking of investment funds, especially hedge funds, as is demonstrated in the empirical application.

CHAPTER 4 addresses the issue of testing cross-sectional independence of error terms in fixed effect panel data models. If the independence hypothesis is violated, the reliability of ordinary least squares inference is no longer guaranteed. Special interest is put in the case where the time dimension T is small relative to the number of cross-sections N. In this case, the LM test of Breusch and Pagan (1980) fails in that it is heavily size-distorted. Several tests that make up for this defficiency have been suggested in the literature, see for instance Pesaran, Ullah, and Yamagata (2008). While these tests check the null hypothesis of no error cross-section dependence, we propose a test whose underlying null hypothesis is that no error cross-section dependence is present that adversely affects standard parameter tests in the fixed effects model. After all, this is most relevant for inference. The proposed test keeps the nominal size, also when N is large relative to T, and has good power properties, as is shown by means of simulation experiments.

Chapter 1

Testing for speculative bubbles in stock markets: a comparison of alternative methods

1.1 Introduction

Phenomena of speculative excesses have long been present in economic history. Galbraith (1993) starts his account with the famous *Tulipomania*, which took place in the Netherlands in the 17th century. More recently, the so called *dot.com* or *IT*-bubble came to fame during the end of the 1990s. Enormous increases in stock prices followed by crashes have led many researchers to test for the presence of speculative bubbles. Among these are Shiller (1981) and LeRoy and Porter (1981), who proposed variance bounds tests, West (1987), who designed a two-step test for bubbles, and Froot and Obstfeld (1991), who considered intrinsic bubbles. Moreover, Cuñado, Gil-Alana, and Gracia (2005) and Frömmel and Kruse (2011) employed methods based on fractionally integrated models, while Phillips, Wu, and Yu (2011) used sequential unit-root tests. This list is by no means complete. Gürkaynak (2008) provides an overview of different empirical tests on rational bubbles. In this chapter, we adopt the theoretical framework of Phillips et al. (2011) and propose several other test procedures aiming to improve on the testing power. The forward recursive unit-root test of Phillips et al. (2011) is an attempt to overcome the weaknesses of the approach of Diba and Grossman (1988), who argue against the existence of bubbles in the S&P 500. Evans (1991) demonstrates that Diba and Grossman's (1988) tests do not have sufficient power to effectively detect bubbles that collapse periodically. Phillips et al. (2011) also propose to use sequences of Dickey-Fuller statistics to estimate the date of the emergence of a bubble, i.e. to estimate the date of a regime switch from a random walk to an explosive process.

A main objective of this chapter is to provide alternative tests that are more powerful in detecting a change from a random walk to an explosive process. To this end, we modify several tests that have been proposed in a different context and transfer them to the bubble testing framework. Our Monte Carlo simulations suggest that two of the alternative test procedures outperform the recursive unit-root test of Phillips et al. (2011). Moreover, the empirical power of these procedures is quite close to the power envelope. A second objective of this chapter is to suggest reliable estimators for the break date, i.e. the starting date of the bubble. We also look at the problem from a practitioners perspective and suggest a real time monitoring approach to detect emerging bubbles.

The tests that we adapt to the bubble framework originate from the literature on tests for a change in persistence. Kim (2000), Kim, Belaire-Franch, and Amador (2002), and Busetti and Taylor (2004) proposed procedures to test the null hypothesis that the time series is I(0) throughout the entire sample against the alternative of a change from I(0) to I(1), or vice versa.[1] We adapt these test procedures to the context of bubble detection and study their power properties by means of Monte Carlo simulations. Additionally, two other tests will be included in the study. One is based on Bhargava (1986), who tested whether a time series is a random walk against explosive alternatives. Bhargava (1986) did not construct his test as a test for a structural break. However, we will adjust this test to accommodate regime-switches and apply it sequentially to different subsamples. The other test is a version of the classical Chow-test.

When there is only a single regime switch in the sample, the proposed sequential Chow-test and our modified version of Busetti and Taylor's (2004) procedure exhibit the highest power.

[1] In Section 1.3 we will briefly discuss other tests for a change in persistence and the reason why we focus on the tests of Kim (2000), Kim et al. (2002), and Busetti and Taylor (2004).

Moreover, a breakpoint estimator derived from the Chow-test turns out to be most accurate. This does no longer hold, however, if there are multiple regime changes due to bubble crashes. In that case it is more appropriate to make a slight change of perspective and apply statistical monitoring procedures. In principle, all tests can be redesigned as monitoring procedures, but here we will focus on the sequential Dickey-Fuller t-statistic and a simple CUSUM procedure.

The remainder of CHAPTER 1 is organized as follows. In Section 1.2 we present the basic theory of rational bubbles. We introduce a simple model for *randomly starting bubbles* and review Evans' (1991) *periodically collapsing bubbles*. In Sections 1.3 and 1.4, the above mentioned tests and estimation procedures are introduced. Monitoring procedures are considered in Section 1.5. Furthermore, in Section 1.6 the performance of the procedures is analyzed via Monte Carlo methods. Finally, in Section 1.7 the test and estimation procedures are applied to Nasdaq index data and various other financial time series. Section 1.8 concludes.

1.2 Rational bubbles

Speculative bubbles in stock markets are systematic departures from the fundamental price of an asset. Following Blanchard and Watson (1982) or Campbell, Lo and MacKinlay (1997) the fundamental price of the asset is derived from the following standard no arbitrage condition:

$$P_t = \frac{E_t\left[P_{t+1} + D_{t+1}\right]}{1+R}, \tag{1.1}$$

where P_t denotes the stock price at period t, D_{t+1} is the dividend for period t, R is the constant risk-free rate, and $E_t\left[\cdot\right]$ denotes the expectation conditional on the information at time t. Solving Equation (1.1) by forward iteration yields the fundamental price

$$P_t^f = \sum_{i=1}^{\infty} \frac{1}{(1+R)^i}\, E_t\left[D_{t+i}\right]. \tag{1.2}$$

Equation (1.2) states that the fundamental price is equal to the present value of all expected dividend payments. Imposing the transversality condition

$$\lim_{k\to\infty} E_t\left[\frac{1}{(1+R)^k}\, P_{t+k}\right] = 0 \tag{1.3}$$

ensures that $P_t = P_t^f$ is the unique solution of (1.1) and thereby rules out the existence of a bubble. However, if (1.3) does not hold, P_t^f is not the only price process that solves (1.1). Consider a process $\{B_t\}_{t=1}^{\infty}$ with the property

$$E_t\left[B_{t+1}\right] = (1+R)B_t. \tag{1.4}$$

It can easily be verified that adding B_t to P_t^f will yield another solution to Equation (1.1). In fact, there are infinitely many solutions. They take the form

$$P_t = P_t^f + B_t, \tag{1.5}$$

where $\{B_t\}_{t=1}^{\infty}$ is a process that satisfies Equation (1.4). The last equation decomposes the price into two components: the fundamental component, P_t^f, and a part that is commonly referred to as the bubble component, B_t. If a bubble is present in the stock price, Equation (1.4) requires that any rational investor, who is willing to buy that stock, must expect the bubble to grow at rate R. If this is the case and if B_t is strictly positive, this sets the stage for speculative investor behavior: A rational investor is willing to buy an "overpriced" stock, since she believes that through price increases she will be sufficiently compensated for the extra payment B_t. If investors expect prices to increase at rate R and buy shares, the stock price will indeed rise and complete the loop of a self-fulfilling prophecy.

The crucial condition for rational bubbles is given by (1.4). However, this restriction leaves room for a variety of processes. We next present several models for rational bubbles, some of which will also be considered in our Monte Carlo analysis. The simplest example of a process that satisfies (1.4) is the deterministic bubble, given by $B_t = (1+R)^t B_0$, where B_0 is an initial value. A somewhat more realistic example, in which the bubble does not necessarily grow forever, is taken from Blanchard and Watson (1982). The bubble process is given by

$$B_{t+1} = \begin{cases} \pi^{-1}(1+R)B_t + \mu_{t+1}, & \text{with probability } \pi \\ \mu_{t+1}, & \text{with probability } 1-\pi \end{cases} \tag{1.6}$$

where $\{\mu_t\}_{t=1}^{\infty}$ is a sequence of iid random variables with zero mean. In each period, the bubble described in Equation (1.6) will continue, with probability π, or collapse with probability $1-\pi$. As long as the bubble does not collapse, the realized return exceeds the risk-free rate R as a compensation for the risk that the bubble bursts.

Not every process that satisfies (1.4) is consistent with rationality. For instance, given that stock prices cannot be negative, negative bubbles can be excluded: Applying the law of iterated expectations to (1.4) yields $E_t[B_{t+\tau}] = (1+R)^{\tau} B_t$. If at some time t, B_t was negative, then, as τ goes to infinity, the expected bubble tends to minus infinity, implying a negative stock price at some future time period. Furthermore, Diba and Grossman (1988) argue that a bubble process cannot start from zero. Assume that $B_t = 0$ at some time t. Then $E_t[B_{t+1}] = 0$, by (1.4). Assuming nonnegative stock prices, we have $B_{t+1} \geq 0$. Together this implies that $B_{t+1} = 0$ almost surely.

Although rational bubbles cannot start from zero, they can take a constant positive value for some time and start to grow exponentially with some probability π. For instance, consider the following *randomly starting bubble*:

$$B_t = \begin{cases} B_{t-1} + \frac{RB_{t-1}}{\pi}\theta_t, & \text{if } B_{t-1} = B_0 \\ (1+R)B_{t-1}, & \text{if } B_{t-1} > B_0 \end{cases} \qquad \text{for } t = 1, \ldots, T, \tag{1.7}$$

where $B_0 > 0$ is the initial value of the bubble. R is the risk-free rate and $\{\theta_t\}_{t=1}^{T}$ is an exogenous iid Bernoulli process with $Prob(\theta_t = 1) = \pi = 1 - Prob(\theta_t = 0)$, and $\pi \in (0, 1]$. A process generated according to (1.7) starts at some strictly positive value B_0 and remains at that level until the Bernoulli process switches to unity. At that point, the process B_t makes a jump of size RB_0/π and from then on grows at rate R. One can easily verify that such a process satisfies

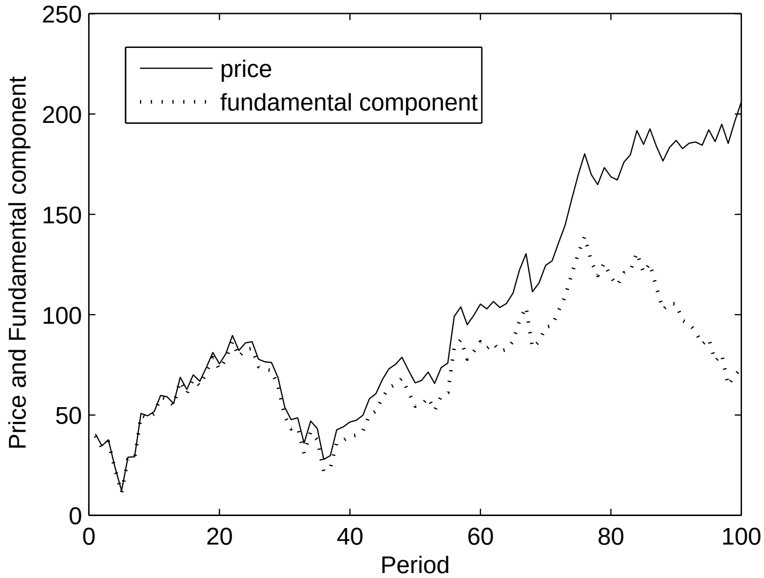

Figure 1.1: Simulated price series with randomly starting bubble component

the no arbitrage condition for rational bubbles, given in (1.4). In our Monte Carlo analysis we will simulate such a bubble process together with a fundamental price P_t^f generated according to (1.10) and (1.11). An example of the resulting price processes $P_t = P_t^f + B_t$ is given in Figure 1.1 (solid line), which also depicts the fundamental price (dotted line). In this example the parameters for the bubble process are $\pi = 0.05$, $R = 0.05$, and $B_0 = 1$. Following Evans (1991) the dividend process in (1.10) is simulated with drift $\mu = 0.0373$, initial value $D_0 = 1.3$ and identical normally distributed disturbances with mean zero and variance $\sigma^2 = 0.1574$.

In his critique of Diba and Grossman's (1988) testing approach Evans (1991) proposed the following model for *periodically collapsing bubbles*:

$$B_{t+1} = \begin{cases} (1+R)B_t u_{t+1} & \text{if } B_t \leq \alpha \\ \left[\delta + \pi^{-1}(1+R)\theta_{t+1}(B_t - (1+R)^{-1}\delta)\right] u_{t+1} & \text{otherwise.} \end{cases} \tag{1.8}$$

Here, δ and α are parameters satisfying $0 < \delta < (1+R)\alpha$, and $\{u_t\}_{t=1}^{\infty}$ is an iid process with $u_t \geq 0$ and $E_t[u_{t+1}] = 1$, for all t. $\{\theta_t\}_{t=1}^{\infty}$ is an iid Bernoulli process, where the probability that $\theta_t = 1$ is π and the probability that $\theta_t = 0$ is $1 - \pi$, with $0 < \pi \leq 1$. It is easy to verify that the bubble process defined in (1.8) satisfies (1.4). Letting the initial value $B_0 = \delta$, the bubble

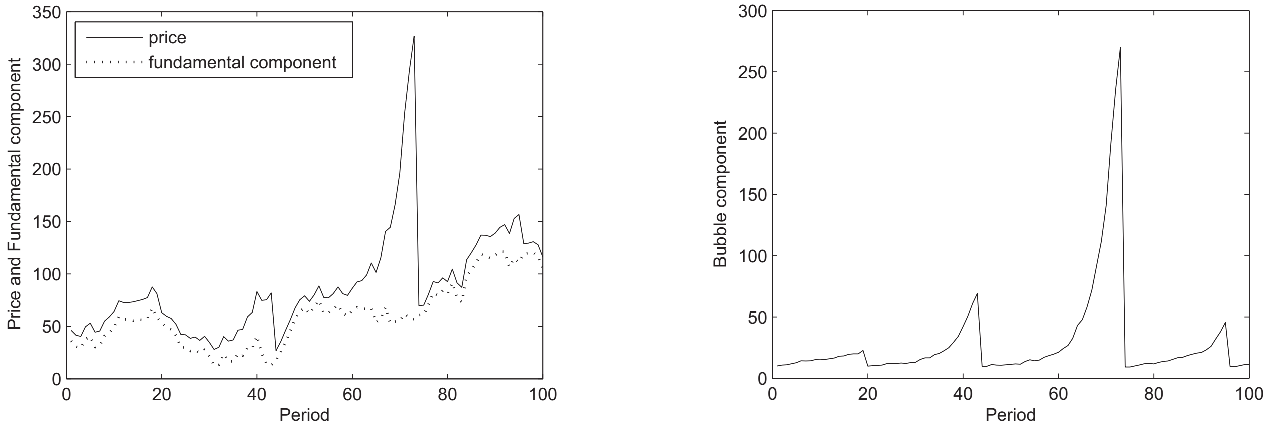

Figure 1.2: Simulated price with fundamental component (left) and bubble component (right)

increases until it exceeds some value α. Thereafter, it is subject to the possibility of collapse with probability $(1-\pi)$, in which case it will return to δu_{t+1} (i.e. to δ in expectation). For our simulations we will follow Evans (1991) and specify $u_t = \exp(\xi_t - \frac{1}{2}\tau^2)$, where $\xi_t \sim$ iid $N(0, \tau^2)$. The parameter values are set to $\alpha = 1$, $\delta = 0.5$, $\tau = 0.05$ and $R = 0.05$. Note that such a periodically collapsing bubble never crashes to zero. Thus, it does not violate Diba and Grossman's (1988) finding that a bubble cannot restart from zero. Evans (1991) demonstrated that Diba and Grossman's (1988) tests lack sufficient power to detect periodically collapsing bubbles, even if the probability of collapse is small. A realization of a periodically collapsing bubble is shown in the right panel of Figure 1.2, where $\pi = 0.85$. The left panel of Figure 1.2 displays the fundamental price (dotted line), which is generated as above, and the observed price (solid line). The observed price is constructed as

$$P_t = P_t^f + 20B_t. \tag{1.9}$$

As in Evans (1991), the bubble process is multiplied by a factor of 20. This is to ensure that the variance of the first difference of the bubble component $\Delta(20B_t)$ is large relative to the variance of the first difference of the fundamental price ΔP_t^f.

An obvious problem is that the fundamental component in (1.5) cannot be directly observed. Therefore, assumptions have to be imposed to characterize the time series properties of the fundamental price P_t^f. A convenient – and nevertheless empirically plausible – assumption is that dividends follow a random walk with drift

$$D_t = \mu + D_{t-1} + u_t, \tag{1.10}$$

where u_t is a white noise process. Under this assumption the fundamental price results as

$$P_t^f = \frac{1+R}{R^2}\mu + \frac{1}{R}D_t, \tag{1.11}$$

(e.g. Evans, 1991). Consequently, if D_t follows a random walk with drift, so does P_t^f. This allows us to distinguish the fundamental price from the bubble process that is characterized by an explosive autoregressive process (see also Diba and Grossmann, 1988).

1.3 Test procedures

The test procedures are based on the time varying AR(1) model

$$y_t = \rho_t y_{t-1} + \epsilon_t \ , \tag{1.12}$$

where ϵ_t is a white noise process with $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$ and $y_0 = c < \infty$. To simplify the exposition we ignore a possible constant in the autoregression. If the test is applied to a series of daily stock prices, the constant is usually very small and insignificant. To account for a possible constant in (1.12), the series may be detrended by running a least-squares regression on a constant and a linear time trend. All test statistics presented in this section can be computed by using the residuals of this regression instead of the original time series.[2]

Under the null hypothesis y_t follows a random walk for all time periods, i.e.,

$$H_0: \quad \rho_t = 1 \quad \text{for } t = 1, 2, \ldots, T. \tag{1.13}$$

[2]In that case, Brownian motions are replaced by detrended Brownian motions in the limiting distribution of the test statistics (see below).

Under the alternative hypothesis the process starts as a random walk but changes to an explosive process at an unknown time $[\tau^* T]$ (where $\tau^* \in (0,1)$ and $[\tau^* T]$ denotes the greatest integer smaller than or equal to $\tau^* T$):

$$H_1: \quad \rho_t = \begin{cases} 1 & \text{for } t = 1, \ldots, [\tau^* T] \\ \rho^* > 1 & \text{for } t = [\tau^* T] + 1, \ldots, T. \end{cases} \tag{1.14}$$

Various statistics have been suggested to test for a structural break in the autoregressive parameter. Most of the work focus on a change from a nonstationary regime (i.e. $\rho_t = 1$) to a stationary regime ($\rho_t < 1$) or *vice versa*. Since these test statistics can be easily adapted to the situation of a change from an I(1) to an explosive process, we first consider various test statistics suggested in the literature.

a) The Bhargava statistic

To test the null hypothesis of a random walk ($\rho_t = 1$) against explosive alternatives $\rho_t = \rho^* > 1$ for all $t = 1, \ldots, T$, Bhargava (1986) proposed the locally most powerful invariant test statistic

$$B_0^* = \frac{\sum_{t=1}^{T} (y_t - y_{t-1})^2}{\sum_{t=1}^{T} (y_t - y_0)^2}. \tag{1.15}$$

Since Bhargava's (1986) alternative does not incorporate a structural break we employ a modified version of the inverted test statistic:

$$B_\tau = \frac{1}{T - [\tau T]} \left(\frac{\sum_{t=[\tau T]+1}^{T} (y_t - y_{t-1})^2}{\sum\limits_{t=[\tau T]+1}^{T} (y_t - y_{[\tau T]})^2} \right)^{-1} = \frac{1}{s_\tau^2 \left(T - [\tau T]\right)^2} \sum_{t=[\tau T]+1}^{T} (y_t - y_{[\tau T]})^2, \tag{1.16}$$

where $s_\tau^2 = (T - [\tau T])^{-1} \sum_{t=[\tau T]+1}^{T} (y_t - y_{t-1})^2$. To test for a change from $I(1)$ to an explosive process in the interval $\tau \in [0, 1 - \tau_0]$, where $\tau_0 \in (0, 0.5)$, we consider the statistic

$$supB(\tau_0) = \sup_{\tau \in [0, 1-\tau_0]} B_\tau. \tag{1.17}$$

The test rejects the null hypothesis for large values of $supB(\tau_0)$. Note that the original Bhargava (1986) test rejects if B_0^* is small. Since all tests presented below reject for large values, our test statistic is inversely related to the original Bhargava statistic.

The statistic (1.17) may be motivated as follows. Assume that we want to forecast the value y_{T^*+h} at period $T^* = [\tau T]$. Since it is assumed that the series is a random walk up to period T^*, the forecast results as $\widehat{y}_{T^*+h|T^*} = y_{T^*}$. The B_τ-statistic is based on the sum of squared forecast errors for $y_{T^*+1}, \ldots, y_T$. If the second part of the sample is generated by an explosive process, then the random walk forecast becomes very poor as h gets large. Therefore, this test statistic is supposed to have good power against explosive alternatives. The supremum of the statistics B_τ is used to cope with the fact that the breakpoint is unknown.

The asymptotic distribution of this test statistic under null hypothesis was not derived in the literature but easily follows from the continuous mapping theorem as

$$supB(\tau_0) \Rightarrow \sup_{\tau \in [0,1-\tau_0]} \left\{ (1-\tau)^{-2} \int_\tau^1 (W(r) - W(\tau))^2 \, dr \right\},$$

where $\Rightarrow$ denotes weak convergence, and W denotes standard Brownian motion on the interval $[0, 1]$.

b) The Busetti-Taylor statistic

Busetti and Taylor (2004) proposed a statistic for testing the hypothesis that a time series is stationary against the alternative that it switches from a stationary to an I(1) process at an unknown breakpoint. Here we propose a modified version of the statistic to test the null hypothesis (1.13) against the alternative (1.14):

$$supBT(\tau_0) = \sup_{\tau \in [0,1-\tau_0]} BT_\tau, \qquad \text{where } BT_\tau = \frac{1}{s_0^2 (T - [\tau T])^2} \sum_{t=[\tau T]+1}^{T} (y_T - y_{t-1})^2. \tag{1.18}$$

The supBT test rejects for large values of $supBT(\tau_0)$. Note that BT_τ employs the variance estimator s_0^2 based on the entire sample, while the inverted Bhargava statistic in (1.16) employs s_τ^2, which uses only the observations starting at $[\tau T]$. Another way to illustrate the difference between the two test statistics is to note that the BT statistic is based on the sum of squared forecast errors of forecasting the final value y_T from the periods $y_{T^*}+1,\dots,y_{T-1}$ by using the null hypothesis that y_t is generated by a random walk. Therefore the BT statistic fixes the target to be forecasted, whereas the Bhargava statistic uses multiple forecast horizons of a fixed forecast interval. The following result for the limiting distribution of $supBT$ can easily be derived:

$$\underset{\tau\in[0,1-\tau_0]}{sup} BT_\tau \Rightarrow \underset{\tau\in[0,1-\tau_0]}{sup} \left\{(1-\tau)^{-2}\int_\tau^1 W(1-r)^2 dr\right\}.$$

Remark: In their work Busetti and Taylor (2004) considered the process $y_t = \beta_0 + \mu_t + \epsilon_t$, where β_0 is a constant, $\epsilon_t \sim$ iid $N(0,\sigma^2)$, and μ_t is a process that is I(0) under the null hypothesis and switches from I(0) to I(1) under the alternative. They proposed the statistic $\varphi(\tau) = \hat{\sigma}^{-2}(T-[\tau T])^{-2}\sum_{t=[\tau T]+1}^{T}\left(\sum_{j=t}^{T}\hat{\epsilon}_j\right)^2$ where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2$ and $\hat{\epsilon}_t$ are the residuals from OLS-regression of y_t on an intercept. To obtain stationary residuals under the null hypothesis (1.13), we use one-step-ahead forecast errors $y_t - y_{t-1}$ instead of OLS-residuals $\hat{\epsilon}_t$, which leads to (1.18).

c) The Kim statistic

Another statistic for testing the $I(0)$ null hypothesis against a change from I(0) to I(1) was proposed by Kim (2000). To transfer the statistic to the bubble-testing framework we apply modifications similar to those in the remark above, which yields the following statistic:

$$supK(\tau_0) = \underset{\tau\in[\tau_0,1-\tau_0]}{sup} K_\tau \qquad \text{with} \quad K_\tau = \frac{(T-[\tau T])^{-2}\sum_{t=[\tau T]+1}^{T}(y_t - y_{[\tau T]})^2}{[\tau T]^{-2}\sum_{t=1}^{[\tau T]}(y_t-y_0)^2}. \tag{1.19}$$

The test rejects for large values of $supK(\tau_0)$. The statistic K_τ is computed over the symmetric interval $[\tau_0, 1-\tau_0]$. It can be interpreted as the scaled ratio of the sum of squared forecast errors. The prediction is made under the assumption that the time series follows a random walk. y_0 is used to forecast $y_1,\dots,y_{[\tau T]}$ (denominator) and $y_{[\tau T]}$ is the forecast of $y_{[\tau T]+1},\dots,y_T$. The limiting distribution is obtained as

$$\underset{\tau \in [\tau_0, 1-\tau_0]}{sup} \quad K_\tau \quad \Rightarrow \quad \underset{\tau \in [\tau_0, 1-\tau_0]}{sup} \left\{ \left(\frac{\tau}{1-\tau} \right)^2 \frac{\int_\tau^1 (W(r) - W(\tau))^2 dr}{\int_0^\tau W(r)^2 \, dr} \right\}.$$

d) The Phillips/Wu/Yu statistic

To test for speculative bubbles, Phillips et al. (2011) suggest to use a sequence of Dickey-Fuller (DF) tests. Let $\hat{\rho}_\tau$ denote the OLS estimator of ρ and $\hat{\sigma}_{\rho,\tau}$ the usual estimator for the standard deviation of $\hat{\rho}_\tau$ using the subsample $\{y_1, \ldots, y_{[\tau T]}\}$.[3] The forward recursive Dickey-Fuller (DF) test is given by

$$supDF(\tau_0) = \underset{\tau_0 \le \tau \le 1}{sup} DF_\tau \qquad \text{with} \quad DF_\tau = \frac{\hat{\rho}_\tau - 1}{\hat{\sigma}_{\rho,\tau}}. \tag{1.20}$$

Usually, the standard Dickey-Fuller test is employed to test H_0 against the alternative $\rho_t = \rho < 1 \quad (t = 1, \ldots, T)$, and the test rejects if DF_1 is small. For the alternative considered here (see (1.14)) we use upper-tail critical values and reject when $supDF(\tau_0)$ is large. Note that the DF statistic is computed for the asymmetric interval $[\tau_0, 1]$. Following Phillips et al. (2011) we will set $\tau_0 = 0.1$ in the simulation experiments. The limiting distribution derived by Phillips at al. (2011) is

$$\underset{\tau_0 \le \tau \le 1}{sup} DF_\tau \Rightarrow \quad \underset{\tau_0 \le \tau \le 1}{sup} \frac{\int_0^\tau W(r) dW(r)}{\sqrt{\int_0^\tau W(r)^2 dr}}.$$

The test procedure does not take into account that both under the null hypothesis (1.13) and under the alternative (1.14) y_t is a random walk for $t = 1, \ldots, [\tau^* T]$. In this sense the supDF test does not exploit all information.

e) A Chow-type unit root statistic for a structural break

The information that y_t is a random walk for $t = 1, \ldots, [\tau^* T]$ under both H_0 and H_1 can be incorporated in the test procedure by using a Chow-test for a structural break in the autoregressive parameter. Under the assumption that $\rho_t = 1$ for $t = 1, \ldots, [\tau T]$ and $\rho_t - 1 = \delta > 0$

[3] In their paper, Phillips et al. (2011) apply augmented Dickey-Fuller tests and use a constant in their regression.

for $t = [\tau T] + 1, \ldots, T$, the model can be written as

$$\Delta y_t = \delta \left(y_{t-1} \mathbb{1}_{\{t > [\tau T]\}} \right) + \varepsilon_t, \tag{1.21}$$

where $\mathbb{1}_{\{\cdot\}}$ is an indicator function that equals one when the statement in braces is true and equals zero otherwise. Correspondingly, the null hypothesis of interest is $H_0 : \delta = 0$, which is tested against the alternative $H_1 : \delta > 0$. It is easy to see that the regression t-statistic for this null hypothesis is

$$DFC_\tau = \frac{\sum\limits_{t=[\tau T]+1}^{T} \Delta y_t y_{t-1}}{\widetilde{\sigma}_\tau \sqrt{\sum\limits_{t=[\tau T]+1}^{T} y_{t-1}^2}}, \tag{1.22}$$

where

$$\widetilde{\sigma}_\tau^2 = \frac{1}{T-2} \sum_{t=2}^{T} \left(\Delta y_t - \widehat{\delta}_\tau y_{t-1} \mathbb{1}_{\{t > [\tau T]\}} \right)^2$$

and $\widehat{\delta}_\tau$ denotes the OLS estimator of δ in (1.21). The Chow-type Dickey-Fuller statistic to test for a change from $I(1)$ to explosive in the interval $\tau \in [0, 1 - \tau_0]$ can be written as

$$supDFC(\tau_0) = \sup_{\tau \in [0, 1-\tau_0]} DFC_\tau. \tag{1.23}$$

The test rejects for large values of $supDFC(\tau_0)$. The test, in fact, corresponds to a one-sided version of the "supWald" test of Andrews (1993), where the supremum is taken over a sequence of Wald statistics. Straightforward derivation yields:

$$supDFC(\tau_0) \Rightarrow \sup_{\tau \in [0, 1-\tau_0]} \frac{\int_\tau^1 W(r) dW(r)}{\sqrt{\int_\tau^1 W(r)^2 dr}}.$$

Note that the limiting distribution is analogous to the one found in (d). In finite samples, the null distribution for both the supDFC and the supDF statistics are affected by the initial value of the time series if the series is not demeaned or detrended. To overcome this problem we suggest to compute the test statistics by using the transformed series $\{\tilde{y}_t\}_{t=1}^{T}$ with $\tilde{y}_t = y_t - y_0$.

Further test procedures and infeasible point optimal tests

The test procedures presented so far fall into two categories: recursive DF t-statistics and tests based on scaled sums of forecast errors. Recursive DF t-tests have originally been proposed to test against stationary alternatives (cf. Banerjee, Lumsdaine, and Stock (1992) or Leybourne, Kim, Smith, and Newbold (2003)). In that case lower-tail critical values are appropriate. In order to test the I(1) hypothesis against explosive alternatives, Phillips et al. (2011) proposed the use of forward recursive DF t-statistics and upper-tail critical values. In (e) we suggested DF t-statistics which are essentially backward recursive. In the literature on tests for a change in persistence several variants of Kim's (2000) and Busetti and Taylor's (2004) tests are available (cf. Taylor and Leybourne (2004) and Taylor (2005)). We have also adapted these tests to the bubble scenario, using the same logic as for the supBT and supK test. Monte Carlo simulations have shown, however, that the resulting procedures perform worse than the supDFC and the supBT tests in terms of power. To save space, these results are not reported here.

In (1.14) we have assumed that the break fraction τ^* is unknown. If, instead, τ^* is known, point-optimal tests can be constructed by using the Neyman-Pearson lemma. This allows to gauge the performance of the tests in *a)* to *e)* relative to the power envelope. Under the additional assumption that the error terms in (1.12) follow a normal distribution with known variance σ^2 and for fixed τ^* and ρ^* in (1.14), the most powerful level-α test of H_0 against H_1 rejects, if

$$PO(\tau^*, \rho^*) = \frac{1}{\sigma^2} \sum_{t=\tau^*+1}^{T} (y_t - y_{t-1})^2 - (y_t - \rho^* y_{t-1})^2 \ > \ k_\alpha(\tau^*, \rho^*), \tag{1.24}$$

where $k_\alpha(\tau^*, \rho^*)$ denotes the critical value with respect to a significance level of α. For a known break date this test is optimal against the alternative $\rho = \rho^*$. Replacing ρ^* with the suitable local alternative $\rho^* = 1 + b/T$ with $b > 0$ and rearranging terms, the asymptotic distribution under H_0 is readily derived as

$$PO(\tau^*, 1 + b/T) \ \Rightarrow \ 2b \int_{\tau^*}^{1} W(r)dW(r) - b^2 \int_{\tau^*}^{1} W(r)^2 dr.$$

By determining the rejection probabilities of the point optimal tests by means of Monte Carlo simulations, we are able to compute the power envelope for our testing problem.

1.4 Estimation of the break date

Assume that the time series under consideration, $\{y_t\}_{t=0}^T$, can be described by (1.12) and (1.14), where τ^* is unknown. First, consider the approach that has been proposed by Phillips et al. (2011). In a simple version, the estimate for the starting date of the bubble, $\hat{\tau}_P$, is given by the smallest value of $\tau \in [\tau_0, 1]$ for which DF_τ is larger than the right-tail 5% critical value derived from the asymptotic distribution of the standard Dickey-Fuller t-statistic. Therefore, the estimator results as[4]

$$\hat{\tau}_P = \inf_{\tau \geq \tau_0} \{\tau : DF_\tau > 1.28\}. \tag{1.25}$$

The next estimator for a change point was proposed by Busetti and Taylor (2004). It supplements their test for the existence of a structural change from $I(0)$ to $I(1)$. The idea is to maximize the ratio of the sample variances of the first and second subsample. Intuitively, for the correct break date the difference between the variance of the first subsample (which is assumed to be $I(0)$) and the second sub-sample (which is assumed to be $I(1)$ under the alternative) should be maximal. This idea can be adapted for estimating the date of a change from an $I(1)$ to an explosive process. The estimator results as

$$\hat{\tau}_{BT} = \underset{\tau \in [\tau_0, 1-\tau_0]}{\operatorname{argmax}} \Lambda(\tau), \quad \text{where} \quad \Lambda(\tau) = \frac{(T - [\tau T])^{-2} \sum\limits_{t=[\tau T]+1}^{T} (\Delta y_t)^2}{[\tau T]^{-2} \sum\limits_{t=1}^{[\tau T]} (\Delta y_t)^2}. \tag{1.26}$$

Finally, we suggest an estimator for the break point that is directly related to the supDFC test. The estimator is given by

$$\hat{\tau}_{DFC} = \underset{\tau \in [0, 1-\tau_0]}{\operatorname{argmax}} DFC_\tau, \tag{1.27}$$

where DFC_τ is as in Equation (1.22). The idea of this estimator is related to that in Leybourne,

[4]To obtain a consistent estimator the significance level has to go to zero as the sample size tends to infinity. In their application Phillips et al. (2011) employ $log(log(\tau T))/100$ as critical values. For a sample size of $T = 400$ this roughly corresponds to a 4% significance level. The results of our simulation experiments remain basically unchanged, if these critical values are used.

Kim, Smith and Newbold (2003). These authors consider the case of a change from I(0) to I(1) and propose a consistent estimator for the unknown break point. Note that this estimator also maximizes the likelihood function with respect to the break date. Bai and Perron (1998) have shown that the maximum likelihood estimator for the break date is consistent.

1.5 Real time monitoring

The test statistics considered in Section 1.3 are designed to detect speculative bubbles within a fixed historical data set. As argued by Chu, Stinchcombe and White (1996) such test may be highly misleading when applied to an increasing sample. This is due to the fact that structural break tests are constructed as "one-shot" test procedures, i.e., the (asymptotic) size of the test is controlled provided that the sample is fixed and the test procedure is applied only once to the same data set (cf. Chu et al., 1996, and Zeileis et al., 2005). To illustrate the problem involved assume that an investor is interested to find out whether the stock price is subject to a speculative bubble. Applying the tests proposed in Section 1.3 to a sample of the last 100 trading days (say) he or she is not able to reject the null hypothesis of no speculative bubbles. If the stock price continues to increase in the subsequent days the investor is interested to find out whether the evidence for a speculative bubble has strengthened. However, repeating the tests for structural breaks when new observations become available eventually leads to a severe over-rejection of the null hypothesis due to multiple application of statistical tests.

Another practical problem is that the tests assume a single structural break from a random walk regime to an explosive process. The results of our Monte Carlo simulations in Section 1.6.4 show that the tests generally lack power if the bubble bursts within the sample, that is, if there is an additional structural break back to a random walk process. The monitoring procedure suggested in this section is able to sidestep the problems due to multiple breaks.

Assume that, when the monitoring starts, a training sample of n observations is available and that the null hypothesis of no structural break holds for the training sample. Then, in each period $n+1, n+2, \ldots$ a new observation arrives. As we will argue below, it is important to fix in advance the maximal length of the monitoring interval $n+1, n+2, \ldots, N = kn$ as the critical value depends on N. Following Chu et al. (1996) we consider two different statistics (detectors):

$$\text{CUSUM:} \qquad S_n^t = \frac{1}{\hat{\sigma}_t} \sum_{j=n+1}^{t} y_j - y_{j-1} = \frac{1}{\hat{\sigma}_t}(y_t - y_n) \quad (t > n) \tag{1.28}$$

$$\text{FLUC:} \qquad Z_t = (\hat{\rho}_t - 1)/\hat{\sigma}_{\rho_t} = DF_{t/n} \quad (t > n) \tag{1.29}$$

where $\hat{\rho}_t$ denotes the OLS estimate of the autoregressive coefficient, $\hat{\sigma}_{\rho_t}$ denotes the corresponding standard deviation, and $\hat{\sigma}_t^2$ is some consistent estimator of the residual variance based on the sample $\{y_0, \ldots, y_t\}$. Note that Chu et al. (1996) suggest a fluctuation test statistic based on the coefficients $\hat{\rho}_t$. Since the coefficient ρ_t is equal to unity under the null hypothesis, the FLUC is essentially similar to the fluctuation statistic advocated by Chu et al. (1996). Also note that both the FLUC detector and the recursive DF test from Section 1.3 make use of standard Dickey-Fuller t-statistics. However, the two procedures apply to different scenarios. While the latter is intended to analyze a given data set with a fixed last observation, the former applies to a data set that increases with the duration of the monitoring. Moreover, the recursive DF *test* is only concerned with whether or not a bubble has emerged within a given data set, while for the FLUC and CUSUM *detectors* it also plays a role how quickly a structural change is detected. Instead of using a constant critical value, one might prefer to use a critical boundary that increases during the monitoring phase. Given that the bubble starts at the beginning of the monitoring, an increasing instead of constant critical boundary should improve chances to detect the bubble quickly.

Under the null hypothesis the functional central limit theorem implies as $n \to \infty$

$$\frac{1}{\sqrt{n}} S_n^{[\lambda n]} \Rightarrow W(\lambda) - W(1) \quad (1 \leq \lambda \leq k)$$

$$Z_{[\lambda n]} \Rightarrow \int_0^\lambda W(r) dW(r) \Big/ \sqrt{\int_0^\lambda W(r)^2 dr} \quad (1 \leq \lambda \leq k),$$

where $W(r)$ is a Brownian motion defined on the interval $r \in [0, k]$. Our CUSUM monitoring is based on the fact that for any $k > 1$ (see Chu et al. 1996)

$$\lim_{n\to\infty} P\left(|S_n^t| > c_t\sqrt{t} \text{ for some } t \in \{n+1, n+2, \ldots, kn\}\right) \leq exp(-b_\alpha/2), \tag{1.30}$$

where

$$c_t = \sqrt{b_\alpha + \log(t/n)} \quad (t > n). \tag{1.31}$$

Since our null hypothesis is one-sided (i.e. we reject the null hypothesis for large positive values of S_n^t) and S_n^t is distributed symmetrically around zero, a one-sided decision rule is adopted as follows. The null hypothesis is rejected if S_t exceeds the threshold c_t the first time, that is,

$$\text{reject } H_0 \text{ if } S_n^t > c_t\sqrt{t} \text{ for some } t > n. \tag{1.32}$$

For a significance level $\alpha = 0.05$, for instance, the one-sided critical value b_α used to compute c_t in (1.31) is 4.6.

Such a test sequence has the advantage that if the evidence for a bubble process is sufficiently large, the monitoring procedure eventually stops before the bubble collapses. Accordingly, such a monitoring procedure sidesteps the problem of multiple breaks due to a possible burst of the bubble.

For the second monitoring using the statistic $DF_{t/n}$ we apply the following rule:

$$\text{reject } H_0 \text{ if } DF_{t/n} > \kappa_t \text{ for some } t = n+1, \ldots, N = kn \tag{1.33}$$

where $\kappa_t = \sqrt{b_{k,\alpha} + \log(t/n)}$.[5] Since the limiting distributions of the CUSUM and the FLUC detectors differ and no theoretical result similar to that in (1.30) is available for the FLUC detector, we determine the critical value $b_{k,\alpha}$ by means of simulation (see Section 1.6.5). This ensures that, under the null hypothesis, the probability of the event $\{DF_{t/n} > \kappa_t,$ for some $t = n+1, \ldots, kn\}$ does not exceed α. It turns out that for the usual significance levels $b_{k,\alpha}$ is monotonically increasing in k, the length of the monitoring period (including the training

[5]It is possible to employ different functional forms of the boundary function κ_t. Our choice is motivated by facilitating comparisons between the performance of CUSUM and FLUC monitoring.

sample) relative to the training sample. Thus, the maximal size of monitoring period has to be fixed before starting the monitoring.

To account for a linear time trend in the data generating process (1.12), the time series can be detrended before computing the detectors. However, it is well known that the DF t-statistic possesses a sizable negative mean and, therefore, the critical values for the detrended FLUC monotoring may be negative. However, our boundary function κ_t is restricted to positive values. To overcome this problem the FLUC detector is computed by using the standardized DF t-statistics $\frac{Z_t - m_{DF}}{\sigma_{DF}}$. The asymptotic first moment m_{DF} and standard deviation σ_{DF} of the DF t-statistic are taken from Nabeya (1999), where $m_{DF} = -2.1814$ and $\sigma_{DF} = 0.7499$. Regarding the CUSUM procedure, instead of using the OLS-detrended series to compute S_t in (1.28) one can replace the forecast error $y_j - y_{j-1}$ with $w_j = \sqrt{\frac{j-1}{j}}(y_j - y_{j-1} - \hat{\mu}_{j-1})$, where $\hat{\mu}_{j-1} = (j-1)^{-1} \sum_{l=1}^{j-1} \Delta y_l$. Note that w_j is the recursive CUSUM residual in the regression of Δy_t on a constant. As is well known, the same asymptotic results for S_n^t hold when this replacement is made. This means that one can proceed as in the case without drift and use the same boundary function $c_t = \sqrt{b_\alpha + \log(t/n)}$ with the same values for b_α.

1.6 Monte Carlo analysis

We start our Monte Carlo analysis within our basic framework in (1.12) – (1.14). In Section 1.6.1 we report critical values and present the results for the power of the tests. In Section 1.6.2 we evaluate the properties of the break point estimators. Furthermore, we consider price processes that contain explicitly modeled bubbles. We investigate the power of the tests to detect *randomly starting bubbles* in Section 1.6.3. *Periodically collapsing bubbles* are considered in Section 1.6.4, where we apply both tests and monitoring procedures. In Section 1.6.5 we present further results for the monitoring procedures.

1.6.1 Testing for a change from I(1) to explosive

We use Monte Carlo simulation to calculate critical values for the test statistics supDF, supB, supBT, supK, supDFC, and for the point-optimal statistics. Here and in the remainder of this

Table 1.1: Large sample upper tail critical values for test statistics

Quantiles	Test statistics				
	supDF	supDFC	supK	supBT	supB
(a) Critical values without detrending					
0.90	2.4152	1.5762	31.4531	1.9317	3.2796
0.95	2.7273	1.9327	43.7172	2.4748	3.9253
0.99	3.3457	2.6285	79.5410	3.8878	5.3746
(b) Critical values with detrending					
0.90	0.5921	0.9436	28.400	1.7374	2.7614
0.95	0.8726	1.3379	38.072	2.2736	3.3472
0.99	1.4176	2.0741	64.863	3.6088	4.6162

Notes: The critical values are estimated by simulation of (1.12) - (1.13) using Gaussian white noise, a sample size of $T = 5000$, and 10,000 replications.

chapter we set $\tau_0 = 0.1$ for all test statistics and estimators. The data is generated according to Equation (1.12) with $\rho_t = 1$ (for all t), an initial value $y_0 = 0$, and Gaussian white noise. To approximate asymptotic critical values we use a sample size of $T = 5000$. The number of replications is $10,000$. We apply the test statistics to the original and to the detrended series, i.e. to the residuals from the regression of y_t on a constant and a linear time trend. The results are reported in Table 1.1. To save space we leave out the critical values for the point-optimal tests. These tests are of little practical use, when the break date is unknown.

To evaluate the empirical power of the tests we generate data according to (1.12) and (1.14) with Gaussian white noise. 2000 replications are performed for the sample sizes $T = 100$, $T = 200$ and $T = 400$. We consider a range of different break points τ^* and growth rates ρ^*. The power of the tests is evaluated at a nominal size of 5%, i.e. a test rejects the null hypothesis when the corresponding statistic is larger than the respective asymptotic 0.95-quantile in Table 1.1. The results are reported in Tables 1.2 and 1.3. The row labeled "actual size" shows that the size of the tests is close to the nominal size, i.e. the asymptotic critical values also apply to finite samples. Only the supB test is somewhat undersized. The actual size of the point-optimal tests depends, apart from the sample size, on τ^* and on the value ρ^* of the autoregressive slope parameter under H_1. The actual size of the point optimal tests ranges between 4.3% and 5.7%.

With regard to testing power, the supDFC test and the supBT test exhibit the best perfor-

Table 1.2: Empirical power: the baseline case

Break point		Test statistics					
		PO(τ^*, ρ^*)	supDF	supDFC	supK	supBT	supB
(a) Power for $T = 100$							
	actual size		0.057	0.050	0.049	0.060	0.023
$\tau^* = 0.7$	$\rho^* = 1.02$	0.347	0.166	0.312	0.085	0.282	0.128
	$\rho^* = 1.03$	0.526	0.293	0.483	0.137	0.466	0.218
	$\rho^* = 1.04$	0.679	0.429	0.634	0.211	0.615	0.342
	$\rho^* = 1.05$	0.780	0.559	0.750	0.316	0.741	0.459
$\tau^* = 0.8$	$\rho^* = 1.02$	0.273	0.107	0.247	0.067	0.214	0.072
	$\rho^* = 1.03$	0.414	0.181	0.379	0.088	0.348	0.118
	$\rho^* = 1.04$	0.545	0.264	0.508	0.117	0.468	0.168
	$\rho^* = 1.05$	0.662	0.369	0.605	0.171	0.589	0.240
$\tau^* = 0.9$	$\rho^* = 1.02$	0.177	0.069	0.169	0.054	0.139	0.031
	$\rho^* = 1.03$	0.276	0.086	0.238	0.061	0.207	0.034
	$\rho^* = 1.04$	0.364	0.112	0.322	0.068	0.288	0.035
	$\rho^* = 1.05$	0.460	0.150	0.397	0.077	0.372	0.034
(b) Power for $T = 200$							
	actual size		0.059	0.054	0.039	0.055	0.031
$\tau^* = 0.7$	$\rho^* = 1.02$	0.694	0.439	0.633	0.216	0.615	0.451
	$\rho^* = 1.03$	0.870	0.673	0.810	0.455	0.802	0.676
	$\rho^* = 1.04$	0.944	0.811	0.905	0.764	0.902	0.813
	$\rho^* = 1.05$	0.973	0.901	0.944	0.894	0.946	0.886
$\tau^* = 0.8$	$\rho^* = 1.02$	0.572	0.271	0.504	0.135	0.467	0.282
	$\rho^* = 1.03$	0.779	0.472	0.698	0.231	0.686	0.478
	$\rho^* = 1.04$	0.876	0.644	0.810	0.430	0.802	0.640
	$\rho^* = 1.05$	0.931	0.761	0.876	0.690	0.873	0.736
$\tau^* = 0.9$	$\rho^* = 1.02$	0.381	0.114	0.328	0.070	0.283	0.088
	$\rho^* = 1.03$	0.560	0.185	0.481	0.095	0.440	0.156
	$\rho^* = 1.04$	0.698	0.303	0.606	0.135	0.584	0.230
	$\rho^* = 1.05$	0.795	0.404	0.705	0.181	0.692	0.303

Notes: The empirical power is computed at a nominal size of 5%. The simulations are conducted with 2,000 replications of (1.12) and (1.14). The true (simulated) breakpoint τ^* is measured relative to the sample size. ρ^* is the autoregressive slope parameter under the explosive regime. In the row labeled *actual size* we report rejection frequencies when the data generating process obeys H_0.

mance among those tests that do not use knowledge of the true break date or of ρ^*. Moreover, these two tests come close to the power envelope computed from the infeasible point-optimal tests. The difference between the power of the supDFC test and the power envelope is never

Table 1.3: Empirical power: the baseline case contd.

Break point		Test statistics					
		PO(τ^*, ρ^*)	supDF	supDFC	supK	supBT	supB
(a) Power for $T = 400$							
	actual size		0.053	0.045	0.040	0.056	0.039
$\tau^* = 0.7$	$\rho^* = 1.02$	0.938	0.824	0.907	0.759	0.904	0.844
	$\rho^* = 1.03$	0.991	0.942	0.978	0.951	0.975	0.949
	$\rho^* = 1.04$	0.999	0.984	0.992	0.986	0.992	0.984
	$\rho^* = 1.05$	1.000	0.995	0.998	0.994	0.998	0.992
$\tau^* = 0.8$	$\rho^* = 1.02$	0.855	0.655	0.817	0.444	0.815	0.694
	$\rho^* = 1.03$	0.954	0.851	0.930	0.824	0.932	0.870
	$\rho^* = 1.04$	0.986	0.934	0.974	0.938	0.973	0.943
	$\rho^* = 1.05$	0.997	0.970	0.990	0.972	0.991	0.969
$\tau^* = 0.9$	$\rho^* = 1.02$	0.687	0.289	0.593	0.115	0.579	0.337
	$\rho^* = 1.03$	0.844	0.503	0.772	0.270	0.771	0.537
	$\rho^* = 1.04$	0.917	0.670	0.862	0.554	0.864	0.673
	$\rho^* = 1.05$	0.960	0.776	0.915	0.728	0.919	0.743

Notes: The empirical power is computed at a nominal size of 5%. The simulations are conducted with 2,000 replications of (1.12) and (1.14). The true (simulated) breakpoint τ^* is measured relative to the sample size. ρ^* is the autoregressive slope parameter under the explosive regime. In the row labeled *actual size* we report rejection frequencies when the data generating process obeys H_0.

larger than 10% and in many cases only about 5% or smaller.[6] The supBT test performs comparably well. Taking into account that the original version of Busetti and Taylor (2004) was constructed to test for a change from I(0) to I(1) and not from I(1) to explosive, the favorable performance is quite remarkable. The power of the supB test is comparable with that of the supDF test if $T \geq 200$. For $T = 100$ supB performs worse than supDF, which is probably due to the fact that supB is undersized.[7] The supK test lacks power if the sample size is small. Note that the supDFC test and the supBT test perform better than the supDF test in all parameter constellations. The advantage over the supDF test tends to increase as the break fraction τ^* increases. For instance, when $T = 400$, $\rho^* = 1.03$ and $\tau^* = 0.7$ the power of the three tests is

[6] For a fixed value of ρ among $1.02, 1.03, 1.04, 1.05$ and known variance σ^2 we also considered statistics of the form $\sup_{\tau \in [0,0.9]} PO(\tau, \rho)$ (cf. (1.24)). None of the resulting feasible tests dominates the *supDFC* test in terms of empirical power. That is, only for a few parameter constellations (T, τ^*, ρ^*) for the data generating process can supDFC be sizeably outperformed, and there are parameter choices where supDFC performs better (not shown).

[7] Further simulations have shown that if finite sample critical values are used, supB performs very similar to supDF for all parameter constellations.

approximately equal, while for $T = 400$, $\rho^* = 1.03$ and $\tau^* = 0.9$ the supDF test rejects the null hypothesis in only 50% of the cases, whereas the supDFC test and the supBT test can reject the null in more than 75% of the cases.

1.6.2 Estimating the break date

In this section we investigate the performance of the estimators from Section 1.4 to identify the date where the time series switches from a random walk to an explosive regime. In the simulation experiments Eqs. (1.12) and (1.14) are used to generate the data. The experiments were conducted for different break points τ^* and for sample sizes $T = 200$ and $T = 400$. The slope parameter ρ^* was set to 1.05. For each parameter constellation $2,000$ replications were performed. For the estimator proposed by Phillips et al. (2011), $\hat{\tau}_P$, we take the 5% right tailed critical value of the standard asymptotic Dickey-Fuller distribution (which equals 1.28) as the signaling threshold for a change from I(1) to explosive (cf. (1.25)). If the threshold is not exceeded for the entire sample, the respective observation is dropped. We set $\tau_0 = 0.1$ for all estimators (cf. Eqs. (1.25) – (1.27)).

Table 1.4 reports the empirical mean and standard deviation (in parentheses) of the break point estimates. The results show that for a sample size of $T = 200$ the accuracy of the estimates is not very high: In several cases the mean value of the estimate is quite far off the true break point and the standard deviations are fairly high. Cases in which the mean value of $\hat{\tau}_P$ or $\hat{\tau}_{BT}$ is close to the true break point seem to be coincidences. The accuracy of the Chow-style break point estimator, $\hat{\tau}_{DFC}$, is more stable across different break points τ^*. When the sample size is increased to $T = 400$, $\hat{\tau}_{DFC}$ is clearly the most reliable estimator. The estimator $\hat{\tau}_P$ tends to indicate the break too early. Among all estimators $\hat{\tau}_P$ has the largest standard deviation. The estimator $\hat{\tau}_{BT}$ tends to lag behind the break date for roughly 10%.

Overall, we observe the best performance for the estimator $\hat{\tau}_{DFC}$. The use of the estimators $\hat{\tau}_P$ and $\hat{\tau}_{BT}$ cannot be recommended in samples with less than 400 observations.

Table 1.4: Estimated break dates when $\rho^* = 1.05$

Break point	Sample Size T=200			Sample Size T=400		
	$\hat{\tau}_P$	$\hat{\tau}_{DFC}$	$\hat{\tau}_{BT}$	$\hat{\tau}_P$	$\hat{\tau}_{DFC}$	$\hat{\tau}_{BT}$
$\tau^* = 0.4$	0.4256	0.4616	0.6564	0.3784	0.4232	0.4901
	(0.2004)	(0.1037)	(0.1334)	(0.1513)	(0.0456)	(0.0601)
$\tau^* = 0.5$	0.4807	0.5582	0.7636	0.4408	0.5207	0.5933
	(0.2376)	(0.0980)	(0.1137)	(0.1929)	(0.0436)	(0.0616)
$\tau^* = 0.6$	0.5298	0.6493	0.8577	0.5009	0.6208	0.6991
	(0.2741)	(0.0862)	(0.0692)	(0.2361)	(0.0441)	(0.0621)
$\tau^* = 0.7$	0.5658	0.7388	0.8967	0.5573	0.7207	0.8080
	(0.3101)	(0.0747)	(0.0149)	(0.2800)	(0.0427)	(0.0548)
$\tau^* = 0.8$	0.5761	0.8184	0.8990	0.6070	0.8143	0.8903
	(0.3432)	(0.0612)	(0.0036)	(0.3208)	(0.0329)	(0.0185)
$\tau^* = 0.9$	0.5129	0.8682	0.8978	0.6122	0.8890	0.8996
	(0.3610)	(0.0887)	(0.0071)	(0.3602)	(0.0512)	(0.0023)

Notes: The table reports mean values and standard deviations (in parentheses) of the break point estimators $\hat{\tau}_P$, $\hat{\tau}_{DFC}$ and $\hat{\tau}_K$. These values as well as the true breakpoint τ^* are expressed as fraction of the sample size T. Data is generated according to Equation (1.12) and (1.14) for the sample size $T = 200$ and $T = 400$. The autoregressive slope parameter under the explosive regime is $\rho = 1.05$. In each case 1000 replications are used.

1.6.3 Randomly starting bubbles

We now investigate how well the tests can detect *randomly starting bubbles*. We use the model presented in Section 1.2. That is we simulate the price process as $P_t = P_t^f + B_t$, where B_t is generated as in (1.7) and P_t^f follows (1.10) and (1.11). The parameters are also specified as in Section 1.2. Since the fundamental price follows a random walk with drift, we detrend the time series before applying the tests. The results are given in Table 1.5. The null and the alternative hypotheses are as in Eqs. (1.12) – (1.14). The nominal size is 5% and the sample sizes are $T = 100$ or $T = 200$. For each parameter constellation $2,000$ replications of the price process $P_t = P_t^f + B_t$ are generated. Cases in which a bubble remains at its initial value for more than 90% of its realizations are excluded from the experiment. Note that for the first row of Table 1.5 no bubble component enters the price series. This suggests that all tests have correct size.

When the sample size equals $T = 100$, the supDFC test and the supBT test have the highest rejection frequencies. In all cases they reject the null hypothesis at least 20% more often than the supDF test or the supB test. As in Section 1.6, the supK test performs worst. When the

Table 1.5: Rejection frequencies in the case of randomly starting bubbles

Initial value	π	Test statistics				
		supDF	supDFC	supK	supBT	supB
Rejection frequencies when the sample size is $T = 100$						
$B_0 = 2$	*no bubble*	0.036	0.049	0.034	0.060	0.020
	0.02	0.419	0.636	0.054	0.647	0.413
	0.05	0.382	0.741	0.051	0.732	0.388
	0.10	0.343	0.825	0.061	0.829	0.365
	0.25	0.288	0.907	0.067	0.907	0.365
	0.50	0.277	0.930	0.072	0.924	0.359
	1.00	0.272	0.931	0.072	0.929	0.367
Rejection frequencies when the sample size is $T = 200$						
$B_0 = 0.05$	*no bubble*	0.037	0.047	0.038	0.060	0.031
	0.01	0.558	0.645	0.104	0.657	0.566
	0.02	0.672	0.770	0.117	0.775	0.678
	0.05	0.863	0.935	0.158	0.938	0.854
	0.10	0.968	0.990	0.169	0.990	0.957

Notes: The tests are applied to series generated as $P_t = P_t^f + B_t$, with P_t^f as in (1.11) and B_t as in (1.7). The number of replications is 2000. The "no-bubble" hypothesis is rejected when the test statistic exceeds the 95% quantile from table 1.1. B_0 is the initial value of the bubble. π is the probability that the bubble enters the phase of explosive growth.

sample size is $T = 200$ and B_0 is 0.05,[8] the differences between the tests are less pronounced. Still, the supK test shows the weakest performance among the tests, while the supDFC test and the supBT test have the highest empirical power.

It is interesting to note that for $T = 100$ and $B_0 = 2$ the power of some tests does not increase if the probability of the bubble starting at the next period increases. This seems odd, since the expected time for the start of the bubble, $1/\pi$, decreases with increasing probability π. However, the parameter π also has another effect on the bubble process. As can be seen in Eq. (1.7), there is a jump in the bubble process at the time when it starts to grow. The jump size is however inversely related to π. These two effects may offset each other, yielding an ambiguous effect on the power. This phenomenon can be observed for the supDF test, the supK test, and the supB test.

Summing up, the simulation experiments with randomly starting bubbles are very much in line with the results of Section 1.6.1, i.e., the supDFC and the supBT tests clearly outperform

[8]When B_0 is set to 2, all tests, except the supK test, have rejection frequencies close to 1 for $\pi \geq 0.05$.

all other tests.

1.6.4 Periodically collapsing bubbles

We now analyze how well the tests are suited to detect Evans' (1991) periodically collapsing bubbles. The price process $P_t = P_t^f + 20B_t$ is generated as described in Section 1.2. Again all tests are applied to detrended data. As in the previous section, rejections at the 5% significance level are considered as evidence for a speculative bubble.

Table 1.6 reports rejection frequencies of both test and monitoring procedures. In the Monte Carlo experiment 2000 replications of the price process P_t have been generated for a sample size of $T = 100$ and different values for the probability π that the bubble continues. First, we discuss the results for the test procedures, whereas results for the monitoring scheme are considered in the next section. Obviously, the power of the tests decreases if the probability π gets smaller. The more striking finding is, however, the superiority of the supDF test relative to all the other tests. For instance, if $\pi = 0.85$ the supDF test rejects in roughly 60% of the cases, while the supK test rejects in 32% of the cases and the other tests have rejection frequencies less than 10%.

There is an intuitive explanation for the superiority of the supDF test. In contrast to the

Table 1.6: Rejection frequencies in the case of periodically collapsing bubbles

π	Test statistics					Detectors	
	supDF	supDFC	supK	supBT	supB	FLUC	CUSUM
no bubble	0.043	0.049	0.033	0.062	0.020	0.045	0.045
0.999	0.803	0.881	0.045	0.934	0.117	0.947	0.942
0.990	0.824	0.589	0.192	0.642	0.119	0.788	0.763
0.950	0.715	0.164	0.371	0.223	0.076	0.618	0.535
0.850	0.593	0.057	0.324	0.072	0.026	0.535	0.379
0.750	0.515	0.040	0.261	0.040	0.018	0.467	0.269
0.500	0.393	0.020	0.235	0.022	0.010	0.342	0.031
0.250	0.248	0.020	0.190	0.023	0.010	0.205	0.015

Notes: The tests and monitoring procedures are applied to series generated as $P_t = P_t^f + 20B_t$, with P_t^f as in (1.11) and B_t as in (1.8). The sample size equals $T = 100$. π is the probability that the bubble does not collapse in the next period. For each case, 2000 replications are performed. The nominal size of the test and monitoring procedure is 5%.

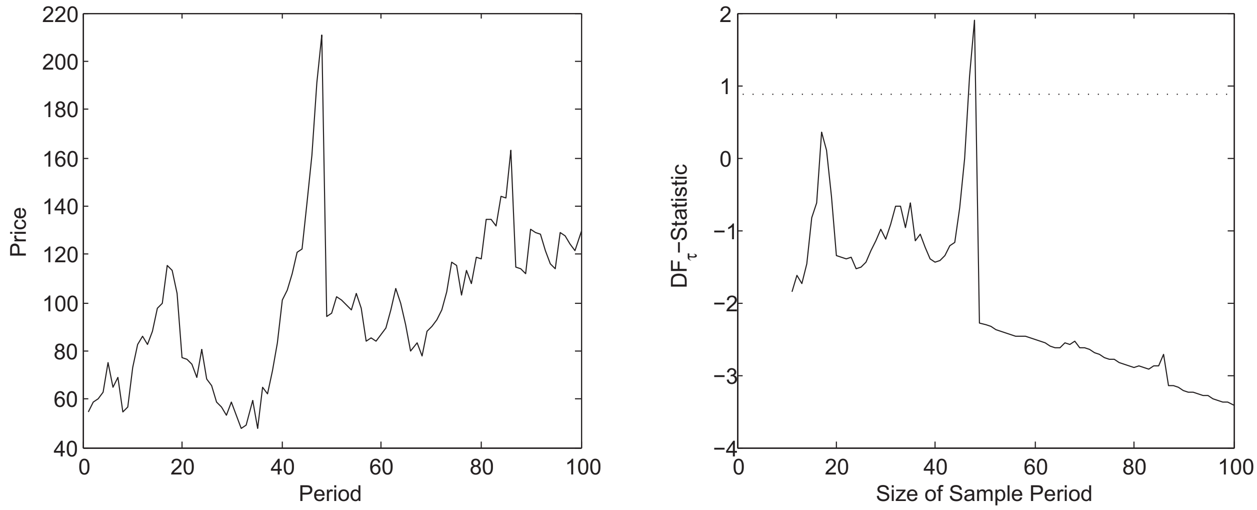

Figure 1.3: Price containing periodically collapsing bubble (left) and DF$_\tau$-statistic (right)

other tests, the supDF test is based on subsamples of the form $\{P_1, \ldots, P_{[\tau T]}\}$. The DF t-statistic for this subsample tends to be large if it contains a long period with a steady growth of the bubble up to the end of the subsample. An obvious case is provided in Figure 1.3. It shows a realization of the price process P_t with $\pi = 0.85$ and a corresponding plot of the sequence of DF_τ statistics.

In contrast to the supDF test, the idea behind the supK test is to compare the first part of the sample $\{P_1, \ldots, P_{[\tau T]}\}$ to the second part $\{P_{[\tau T]+1}, \ldots, P_T\}$. However, if the bubble collapses several times during the whole sample the subsamples mix up periods with and without a bubble. Therefore, the test statistics that are based on a comparison of the pre and post break subsamples have difficulties to indicate a bubble component. Thus, among the alternative tests, the supDF test has the highest power when applied to periodically collapsing bubbles. However, in applications one typically has a clear indication for the end of the explosive regime, just from visual inspection of the time series. If the subsequent observations are excluded from the sample, the supDFC test and the supBT test are most powerful among the tests considered here.

Table 1.7: Critical values for FLUC monitoring

n	k / α	2	3	4	5	6	8	10
i) FLUC monitoring without detrending								
100	0.10	3.05	3.60	3.93	4.15	4.31	4.48	4.57
	0.05	4.50	5.14	5.55	5.69	5.89	6.05	6.26
	0.01	7.76	8.59	9.06	9.48	9.62	9.79	9.99
50	0.10	2.80	3.33	3.62	3.80	3.96	4.14	4.27
	0.05	4.19	4.80	5.11	5.34	5.50	5.72	5.81
	0.01	7.30	8.11	8.43	8.82	8.86	9.25	9.49
20	0.10	2.49	3.12	3.44	3.65	3.78	3.99	4.12
	0.05	3.88	4.56	4.86	5.06	5.19	5.38	5.52
	0.01	7.00	7.84	8.26	8.49	8.66	9.12	9.19
ii) FLUC monitoring with detrending								
100	0.10	6.16	7.55	8.15	8.53	8.98	9.30	9.51
	0.05	8.12	9.82	10.45	10.82	11.20	11.55	11.80
	0.01	12.89	14.22	14.94	15.39	16.22	16.14	16.39
50	0.10	5.47	6.74	7.15	7.69	8.07	8.54	8.80
	0.05	7.29	8.68	9.30	9.62	10.11	10.46	10.87
	0.01	11.54	12.46	13.57	13.81	14.08	14.71	15.44
20	0.10	4.18	5.16	5.71	6.17	6.38	6.78	7.24
	0.05	5.50	6.52	7.24	7.96	8.04	8.45	9.13
	0.01	8.45	9.55	10.82	11.41	11.93	12.35	13.08

Notes: The table shows critical values for FLUC monitoring for different significance levels α, different lengths n of the training sample, and for different lengths k of the monitoring period (measured in multiples of n).

1.6.5 Monitoring

The critical values $b_{k,\alpha}$ in the expression for the boundary function $\kappa_t = \sqrt{b_{k,\alpha} + \log(t/n)}$ from Section 1.5 are reported in Table 1.7 and 1.8 for different values of the significance level α and different sizes of the monitoring sample k (measured relative to the training sample). They have been obtained through application of the monitoring procedures to data simulated according to (1.12) under the random walk null hypothesis with Gaussian white noise innovations. We tabulate critical values for training sample sizes of $n = 20$, $n = 50$, and $n = 100$. It is important to note that $b_{k,\alpha}$ is monotonic increasing in k and the maximal length of the monitoring period as measured by k has to be fixed before starting the monitoring. Table 1.8 reports values $b_{k,\alpha}$ for CUSUM monitoring in finite samples. Using the asymptotic 5%-critical value $b_\alpha = 4.6$ can

Table 1.8: Critical values for CUSUM monitoring

n	k / α	2	3	4	5	6	8	10
i) CUSUM monitoring without drift estimation								
100	0.10	0.92	1.39	1.62	1.80	1.92	2.08	2.19
	0.05	1.51	2.14	2.47	2.73	2.88	3.20	3.36
	0.01	2.86	3.92	4.57	4.94	5.30	5.72	6.02
50	0.10	0.87	1.31	1.55	1.70	1.82	1.97	2.06
	0.05	1.43	2.03	2.34	2.62	2.80	3.03	3.12
	0.01	2.85	3.87	4.25	4.77	5.03	5.47	5.94
20	0.10	0.81	1.18	1.43	1.61	1.75	1.91	2.02
	0.05	1.27	1.96	2.35	2.53	2.72	3.00	3.13
	0.01	2.61	3.91	4.49	4.97	5.16	5.52	5.74
ii) CUSUM monitoring with drift estimation								
100	0.10	0.92	1.38	1.59	1.78	1.90	2.06	2.16
	0.05	1.45	2.11	2.43	2.66	2.80	3.06	3.18
	0.01	2.76	3.78	4.49	4.84	5.12	5.68	5.96
50	0.10	0.91	1.32	1.56	1.69	1.80	1.96	2.08
	0.05	1.43	2.03	2.37	2.57	2.75	2.99	3.14
	0.01	2.71	3.90	4.39	4.74	4.97	5.43	5.62
20	0.10	0.83	1.20	1.43	1.60	1.72	1.86	2.00
	0.05	1.34	1.90	2.24	2.47	2.67	2.88	3.02
	0.01	2.65	3.76	4.42	4.77	5.01	5.53	5.65

Notes: The table shows critical values for CUSUM monitoring for different significance levels α, different lengths n of the training sample, and for different lengths k of the monitoring period (measured in multiples of n).

lead to an empirical size below 1%, when the CUSUM monitoring procedure includes estimation of a drift term.

The results for the FLUC and CUSUM monitoring (with detrending) when applied to periodically collapsing bubbles is reported in the last two columns of Table 1.6. The first $n = 20$ observations are used as training sample and the monitoring ends after 100 observations. The tests are conducted at a nominal size of 5%, i.e. we use the finite sample critical values $b_{5,0.05}$ with $n = 20$. The residual variance for the CUSUM monitoring is estimated as $\hat{\sigma}_t^2 = (t-1)^{-1} \sum_{j=1}^{t} (\Delta y_j - \hat{\mu}_t)^2$, where $\hat{\mu}_t$ is the mean of $\{\Delta y_1, \ldots, \Delta y_t\}$. Similar to the supDF test, the monitoring procedures are more robust against periodically collapsing bubbles. However, except for one case, the monitoring procedures have less power than the supDF test. Moreover,

Table 1.9: Performance of monitoring procedures

	FLUC			*CUSUM*		
Breakpoint	size/ power	delay	σ(delay)	size/ power	delay	σ(delay)
a) size of training sample $n = 40$, *size of total sample* $kn = 200$						
actual size	0.0570	-	-	0.0574	-	-
$\tau^* = 0.2$	0.9766	0.2787	0.1779	0.9744	0.2816	0.1600
$\tau^* = 0.4$	0.9330	0.2454	0.1379	0.9310	0.2415	0.1369
$\tau^* = 0.6$	0.8148	0.2019	0.0969	0.8090	0.2042	0.0982
$\tau^* = 0.8$	0.4696	0.1243	0.0471	0.4588	0.1255	0.0492
b) size of training sample $n = 80$, *size of total sample* $kn = 400$						
actual size	0.0512	-	-	0.0488	-	-
$\tau^* = 0.2$	0.9996	0.1368	0.0977	0.9996	0.1448	0.0906
$\tau^* = 0.4$	0.9974	0.1318	0.0900	0.9972	0.1352	0.0908
$\tau^* = 0.6$	0.9854	0.1250	0.0782	0.9844	0.1319	0.0803
$\tau^* = 0.8$	0.8362	0.0977	0.0474	0.8126	0.1032	0.0484

Notes: The data is generated as a random walk switching to an explosive process at the relative break point τ^* (cf. (1.12) and (1.14)). The autoregressive slope parameter is set to $\rho^* = 1.03$ during the explosive phase. The number of replications is $5,000$. The empirical power is computed at the 5% significance level. In the column *delay* we report the average time needed to detect the regime switch relative to the sample size kn. The column σ(*delay*) contains standard deviations of the detection delay relative to the sample size.

the FLUC monitoring performs better than the CUSUM monitoring. There are two conceptual differences between the supDF test and the FLUC monitoring. First, the critical value of the FLUC monitoring increases in time. Second – and more important – the different detrending methods give rise to the different performance of these two related approaches. Before applying the supDF test the time series is detrended once using the entire sample. In contrast, the monitoring approach uses a sequential detrending scheme. Additional simulations suggest that the empirical power of the FLUC monitoring is virtually identical to that of the supDF test, if the FLUC monitoring uses the same detrending method (i.e. the residuals of a full sample regression) as the supDF test.

We also evaluate the performance of the monitoring procedures in terms of time needed to detect a bubble, a property that is irrelevant for the test procedures. The results are presented in Table 1.9. The data is generated as in (1.12) and (1.14) with slope coefficient $\rho^* = 1.03$ in the explosive regime. The total sample size is set to $N = 200$ and $N = 400$. The size of the training sample is set to $n = 40$ and $n = 80$, respectively. We use 5% critical values $b_{5,0.05}$ with training sample size $n = 50$ and $n = 100$, respectively. The residual variance for the CUSUM monitoring is estimated as $\hat{\sigma}_t^2 = t^{-1} \sum_{j=1}^{t} (\Delta y_j)^2$. The breakpoint τ^* is measured relative to the total sample size. For each case 5000 replications are performed.

The actual size is given in the first row of each panel in Table 1.9. The columns labeled *delay* report the mean delay (measured relative to the sample size) for detecting the bubble. Since a rejection before the bubble starts is spurious, only rejections after the start of the bubble enter the mean delay and standard deviations. Table 1.9 shows that the empirical power of the CUSUM monitoring is comparable to that of the FLUC monitoring. The average time needed to detect the bubble is also very similar for both monitoring procedures. The mean delay and the standard deviation for both procedures decrease as the time of the break date τ^* increases. This is probably due to the fact that the maximum delay is bounded by $1 - \tau^*$.

1.7 Applications

In Section 1.7.1 we start with a detailed discussion of the Nasdaq Composite Index and the so-called dot.com bubble. A range of other financial time series that are often supposed to be affected by speculative bubbles are considered in Section 1.7.2.

1.7.1 The Nasdaq composite index and the dot.com bubble

A price movement for which the term *bubble* is widely used is the development of the Nasdaq composite price index in the late 90s. A mere look at the plot of this time series appears to justify the name "dot.com bubble" (see Figure 1.4). Phillips et al. (2011) found decisive empirical evidence for a speculative bubble during this period (see Shiller, 2001, for a detailed discussion of this speculative bubble). In this section we reproduce their findings and also apply the alternative tests presented in Section 1.3.

We use the same data as in Phillips et al. (2011). The time series of the Nasdaq composite price index and the Nasdaq composite dividend yield are taken from *Datastream International*. Phillips et al. (2011) considered monthly Nasdaq data from February 1973 to June 2005. Figure 1.4 suggests that if there was a bubble it has definitely crashed in March 2000, after

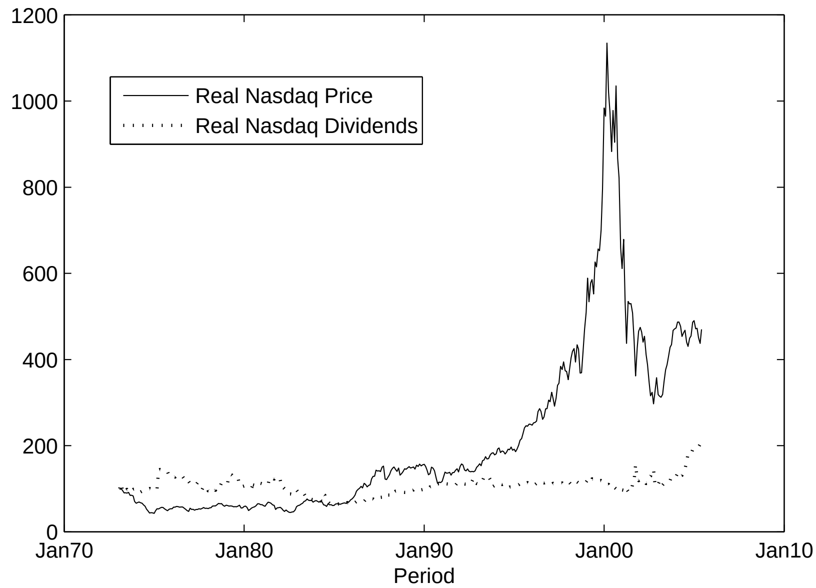

Figure 1.4: Real Nasdaq price and dividends (normalized)

the Nasdaq reached its all time high. If the observations after the peak are included, the chance that the bubble tests detect the bubble will be very low (see Section 1.6.4). Therefore we apply the tests to the restricted sample period from February 1973 to March 2000, which yields $T = 326$ observations. The Nasdaq composite price index is multiplied by the Nasdaq composite dividend yield to compute the time series of the total Nasdaq dividends. By use of the Consumer Price Index from the Federal Reserve Bank of St. Louis, nominal data are transformed to real data. Figure 1.4 shows a plot of the time series for the real Nasdaq price index and the real Nasdaq dividends, where the time series have been normalized to 100 at the beginning of the sampling period. While the price index shows an accelerated increase during the late 1990s followed by a sharp drop, real index dividends are more or less constant.

For the logarithm of these two time series we conduct tests of the random walk hypothesis against the alternative of a switch from I(1) to explosive. If the time series of the Nasdaq composite price index turns out to have changed from I(1) to explosive, while the index dividend series remained constant I(1), this would suggest the presence of bubbles.

In the applications we follow Phillips et al. (2011) more closely and use an augmented Dickey-Fuller (ADF) test that includes an intercept term in the regression, i.e. the regression

Table 1.10: Testing for an explosive root in the Nasdaq index

	Test Statistics				
	supADF	supDFC	supK	supBT	supB
log price index	3.0420	4.5874	12.5203	7.8722	3.0242
log dividends	-0.8563	-0.2086	5.0756	0.1438	2.0057
Upper tail critical values					
1%	2.094	2.6285	79.5410	3.8878	5.3746
5%	1.468	1.9327	43.7172	2.4748	3.9253
10%	1.184	1.5762	31.4531	1.9317	3.2796

Notes: Table 1.10 reports the values of the test statistics applied to the log real Nasdaq price index and dividends. The table also shows the corresponding critical values. Those for the supADF statistic are taken from Phillips et al. (2011). The others correspond to our simulation results.

equation is

$$y_t = \alpha + \rho y_{t-1} + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + \epsilon_t, \quad \text{for } t = 1, \ldots, [\tau T], \tag{1.34}$$

where y_t denotes the logarithm of either the Nasdaq price or dividends and ϵ_t is white noise. The lag order p is determined by a general-to-specific test sequence. (1.34) leads to the augmented DF t-statistic ADF_τ. The supremum of ADF_τ over $\tau \in [0.1, 1]$ is referred to as supADF. The pertaining critical values are taken from Phillips et al. (2011).

The results are shown in Table 1.10. For the dividend series none of the tests rejects the constant I(1) hypothesis at the 5% significance level. Even at the 10% significance level the null hypothesis is not rejected. The application of the supADF test leads to results that are very similar to those obtained by Phillips et al. (2011)[9]. The supDFC test and the supBT test indicate a bubble in the Nasdaq index, while the supB test and the supK test fail to reject the null hypothesis. Most of the results do not change when we consider weekly or daily data, the only exception being the supB test, which becomes significant for higher sampling frequencies (not shown). Thus the supDF test, the supDFC test, the supBT test, and partly the supB test support the view that the Nasdaq price index has changed from I(1) to explosive, while the dividend series remained I(1) throughout the sample. Under the assumptions of the present value model in Eqs. (1.1) and (1.2) one may conclude that a bubble was indeed present in the

[9] They report a test value of 2.894 for the index price and a value of -1.018 for the index dividends.

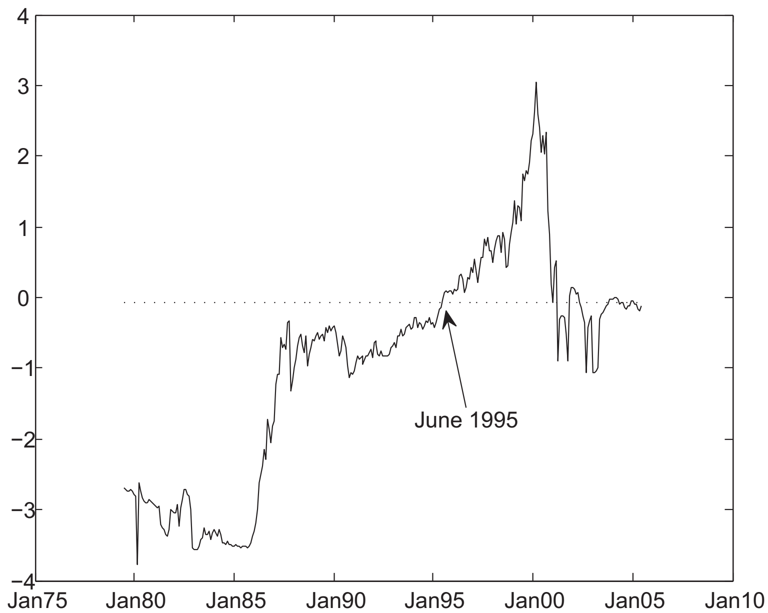

Figure 1.5: ADF_τ Statistics for (log) real Nasdaq prices (solid line), with 5% right tail critical value from asymptotic distribution of the standard Dickey-Fuller t-statistic (dotted line).

Nasdaq index.

This leads to the problem of identifying the starting date of the bubble. First consider the approach of Phillips et al. (2011). Figure 1.5 contains the plots of the ADF_τ statistic for the log real Nasdaq price index (solid line). ADF_τ is the augmented Dickey-Fuller t-statistic for the subsample $\{y_1, \ldots, y_{[\tau T]}\}$. The estimated starting date is June 1995, which is the first time where the ADF_τ statistic is larger than the 5% critical value of the standard Dickey-Fuller t-statistic (-0.08).[10] We also estimate the date of the regime change based on the Chow-style DF-statistics ($\hat{\tau}_{DFC}$). The maximum of DFC_τ is attained at $\hat{\tau}_{DFC} = 72.7\%$, corresponding to October 1992.

1.7.2 Further applications: major stock indices, house prices, and commodities

Besides the Nasdaq Composite index, other financial time series have shown phases of massive growth, often combined with subsequent crashes. Frequently, the term *bubble* has been used

[10] Using $log(log(\tau T))/100$ as critical value for ADF_τ (cf. Phillips et al.(2011)) results in almost the same estimate for the start of the bubble, namely July 1995.

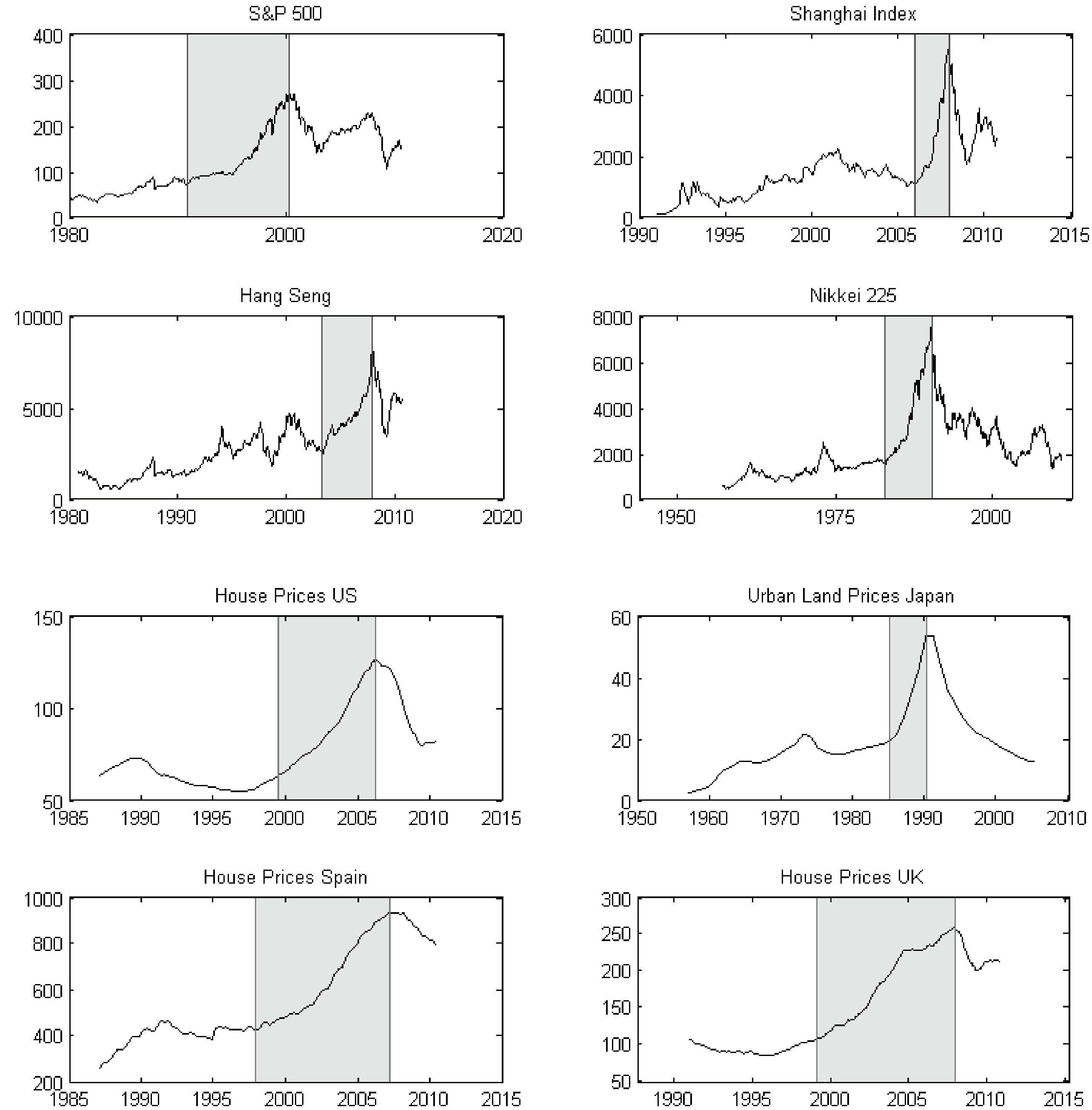

Figure 1.6: The figure shows several stock, house, and land price indices with time on the x-axis and index values on the y-axis.

to describe such phases. Specifically, we consider the Japanese stock market index, Nikkei225, which, during the 80s, exhibited a tremendous increase along with Japanese urban land prices (see Figure 1.6). We also check whether the dot.com bubble has been paralleled by bubbles in the S&P500 and the FTSE100. Furthermore, we analyze more recent upswings in major Chinese stock market indices, the Hang Seng and the Shanghai stock index. During the recent subprime crisis housing markets have received much interest from financial analysts and researchers. We consider the US S&P/Case-Shiller home price index (10-City composite) among other housing indices. Finally, we also test for bubbles in two commodity time series, gold and crude oil. Note that house and commodity prices do not directly fit into the theoretical framework of rational bubbles, in which the fundamental price is based on a stream of future dividend payments. Nevertheless, detecting a change from I(1) to explosive clearly points to excessive speculation.

Our results are presented in Table 1.11. We focus on the supDFC test, which together with supBT performed very well in our Monte Carlo simulations, and on the supDF test. In most cases the supBT test and the supDFC test lead to similar results, so we skip reporting the results for the former. We apply the tests to the logarithm of the inflation adjusted time series at different sampling frequencies: monthly, weekly, and daily. To adjust weekly and daily data for inflation we employ linearly interpolated monthly consumer price indices. Data for house/land price indices were only available at monthly, quarterly or annual frequency. The range of the sampling period for each time series is given in Table 1.11. As in the foregoing section, the price series end at their maximum. With regard to the real gold price we focus on the recent run-up in gold prices and consider the time span from January 1985 to November 2010. Most series were obtained from *Datastream International*. Consumer Price Indices for the US and Hong Kong were downloaded from the websites of the Federal Reserve of St. Louis and the Hong Kong Census and Statistics Department, respectively. The source for the Japanese Urban Land Price Index (6 Major Cities average) is the Statistics Bureau of Japan.

As Table 1.11 shows, there is strong evidence that house prices have run through explosive phases. For the US S&P/Case-Shiller home price index, the UK house price index, and the Spanish house price index both tests reject the random walk hypothesis at the 1% level, indicating bubble-like growth during the time preceding 2006/07. Similar results are obtained for the Japanese land price index, where the explosive phase occurred before 1990. Interestingly, also for the Nikkei225 index, the supDFC and the supDF tests detect explosive behavior prior to January 1990. These results indicate that a land price bubble in Japan was paralleled by a stock market bubble.

Table 1.11: Testing for an explosive root

Series name	monthly data		weekly data		daily data	
	supDFC	supDF	supDFC	supDF	supDFC	supDF
Stock Market Indices:						
S&P 500 (Jan80 - Mar00)	***	*	**	*	***	*
FTSE 100 (Dec85 - Dec99)	*			*	*	**
Nikkei 225 (Jan57 - Jan90)	***	**	***	**	***	***
Hang Seng (Oct80 - Oct07)	**		**		**	
Shanghai (Jan91 - Nov07)	**		***		***	
Commodities:						
Crude Oil (Jan85 - Jul08)						
Gold (Jan85 - Nov10)	*		*		*	
House/ Land Prices:						
US (Jan87 - Mar06)	***	***	-	-	-	-
Spain (87Q1 - 07Q1)	***	***	-	-	-	-
UK (Jan91 - Oct07)	***	***	-	-	-	-
Japan (1957 - 1990)	***	**	-	-	-	-

Notes: The table reports significance levels of the supDFC and supDF test applied to the logarithm of the respective time series at monthly, weekly, and daily frequencies. For the house/ land prices data were not available at a weekly or daily frequency. For Spanish house prices the data frequency is quarterly and for Japanese urban land prices data was available only at an annual frequency.
"*(**/***)" signifies significance at the 10% (5%/ 1%) level.
"-" signifies that the data is not available at the corresponding frequency.

With regard to the S&P 500, both the supDFC and supDF tests detect explosive behavior. The supDFC test rejects at the 5% or 1% level, while the supDF test rejects at the 10% level. Thus, there is evidence that in the 90s not only the Nasdaq but also the S&P 500 was driven by a bubble.

The supDF test detects a bubble in the FTSE 100 series for weekly and daily data. The explosive behavior, however, is not identified during the 90s, but several months before Black Monday 1987. The DF_τ statistic exceeds the critical values only when τ corresponds to July 1987.

Turning to Chinese stock market indices, the results in Table 1.11 show that for the Shanghai Stock Exchange Index, the supDFC test finds clear evidence of explosive growth before November 2007, and for the Hang Seng Index, the supDFC test rejects the no bubble hypothesis at the 5% level (independent of the sampling frequency).

Furthermore, there is no evidence for a bubble in the barrel price of Brent Crude oil. For the price of a troy ounce of gold at the London Bullion Market the evidence is mixed. The supDFC test is significant at the 10% level for monthly and weekly data, while the supDF test is insignificant for all sampling frequencies.[11] Test results are very different, when the real gold price from January 1968 until its all time high in January 1980 is considered. The supDFC and the supDF tests reject the hypothesis of no structural breaks at the 1% level, irrespective of the data frequency.

Finally, the dividend series for the stock market indices were not available in all cases. However, for the S&P 500 and for the Hang Seng we were not able to reject the constant I(1) null hypothesis for the dividend series (not shown), which gives further support to the view, that there has been a bubble in those stock markets.

For those time series where the supDFC detected a change from I(1) to explosive at the 5% level, we use the related estimator $\hat{\tau}_{DFC}$ to estimate the date of the change. The results are reported in Table 1.12. Figure 1.6 shows the plots of the corresponding series. The shaded regions highlight the phase from the estimated start date of the bubble until its presumed collapse. The estimate for the Spanish house price index and for the Japanese house price index should be interpreted with care, since the time series considered are rather short (81 respectively 34 observations). From Table 1.12 one can also see that the break date estimate for a given time series is robust to the choice of the sampling frequency. Also note from the

[11] Interestingly, when we estimate the start of a supposed bubble using $\hat{\tau}_{DFC}$, the result, July 2007, closely coincides with the beginning of the subprime crisis and the failure of several Bear Stearns hedge funds.

Table 1.12: Estimates for the date of change form I(1) to explosive

	data frequency		
Series name	monthly	weekly	daily
Stock Market Indices:			
S&P 500	1990-10-31	1990-10-16	1990-10-11
Nikkei 225	1982-10-03	1982-10-04	1982-10-01
Hang Seng	2003-03-31	2003-04-25	2003-04-25
Shanghai	2005-12-02	2005-12-05	2005-12-05
House/ Land Prices:			
US	1999-15-06	-	-
Spain	1997Q4	-	-
UK	1999-01-15	-	-
Japan	1985	-	-

Notes: The table reports break date estimates for different financial time series using the estimator $\hat{\tau}_{DFC}$ from Section 1.4.

results in Table 1.11 that in most cases, changing the observation frequency has only a minor impact on the p-values of the tests. This seems to mirror the finding of Shiller and Perron (1985) for unit root tests against stationary alternatives. Their theoretical analysis and Monte Carlo simulations suggest that power depends more on the span of the data rather than on the number of observations.

1.8 Conclusion

In this chapter the ability of several tests to detect rational bubbles has been investigated. The main focus was on rational bubbles in stock markets. The basic idea is that a rational bubble gives rise to an explosive component in stock prices. Therefore, a change from a random walk to an explosive regime is considered to be an indication for the emergence of a speculative bubble. Hence, all tests aim at detecting a switch from a random walk to an explosive regime. The sequential Dickey-Fuller test, proposed by Phillips, Wu and Yu (2011) serves as a reference point. We have also adapted various tests for a change in persistence to accommodate a possible change from $I(1)$ to an explosive regime. In a simple simulation framework it was shown that a Chow-type Dickey-Fuller (supDFC) test and our modified version of Busetti and Taylor's

(2004) (supBT) test have higher finite sample power than the test of Phillips et al. (2011), especially when the change from I(1) to explosive occurs late in the sample. This result was confirmed in simulation experiments with randomly starting bubbles.

With regard to the estimation of the date where the bubble process starts, different procedures were considered. It turned out that in typical sample sizes the estimator suggested by Phillips et al. (2011) is downward biased and has a large standard deviation. On the other hand, the estimator that results from maximizing the Chow-type test statistic ($\hat{\tau}_{DFC}$) yields a reliable and roughly unbiased estimator of the starting date.

Since Evans (1991) clearly demonstrated the potential problems of unit root tests to detect periodically collapsing bubbles, we also study the properties of the tests under multiple structural breaks. It turns out that the Phillips, Wu, and Yu (2011) test is much more robust against multiple breaks than all other tests. We also considered sequential Dickey-Fuller t-statistics and CUSUM statistics within a real-time monitoring framework and present critical values for different sizes of the monitoring sample.

The test of Phillips et al. (2011) provides strong evidence for a bubble in the Nasdaq index at the end of the 90s. Their findings are even strengthened by our sequential Chow-type Dickey-Fuller test and the modified version of Busetti and Taylor's (2004) test. The estimation of the starting date of the bubble indicates that the explosive regime emerged in the first half of the 90s. Moreover, our results support the view that bubbles have occurred in several other stock markets. Our sequential Chow-type Dickey-Fuller test and the supDF test of Phillips et al. (2011) also indicate explosive behavior in US, UK, and Spanish house price indices prior to the so-called subprime crisis.

Chapter 2

An operational interpretation and existence of the Aumann-Serrano index of riskiness

2.1 Introduction

In a recent paper Aumann and Serrano (2008) proposed an economic index of riskiness that has several desirable features. For instance, it extends the partial ordering of gambles induced by first and second order stochastic dominance. Foster and Hart (2009) recognized that the economic index of riskiness lacks an operational interpretation and "while attempting to provide an operational interpretation for it, [they] were led instead to the different measure of riskiness" (Foster and Hart, 2009, p. 787, ll. 17ff). According to Foster and Hart (2009) a measure is operational, if it "is defined separately for each gamble and, moreover, has a clear interpretation in monetary terms" (Foster and Hart, 2009, p. 787, ll. 22ff). We show in this chapter that such an operational interpretation can be assigned to the economic index of riskiness of Aumann and Serrano (2008) (in the following abbreviated as AS index).

The key to this result is the so-called adjustment coefficient from ruin theory, see e.g. Grandell (1991). This coefficient has not been considered so far in Aumann and Serrano (2008) or Foster and Hart (2009). However, we show that there is a one-to-one relationship between the AS index and the adjustment coefficient, which allows to provide existence and uniqueness results for the AS index for more general gambles than those considered in Aumann and Serrano (2008).

In the following we briefly restate the definition and some properties of the AS index and the adjustment coefficient. We then combine these to provide an operational interpretation of the AS index as well as existence and uniqueness results in the general case.

2.2 The Aumann-Serrano index of riskiness

Aumann and Serrano's (2008) index of riskiness is an axiomatic approach to assign a meaning to the word *risky*. It enables a decision maker to asses which of two gambles/investments is *riskier* without referring to a specific preference order.

More specifically, let g be a gamble, i.e. a random variable, that takes on values in the real numbers $\mathbb{R}$ (or a subset) with (i) positive probability of negative outcomes ($\mathbb{P}(g < 0) > 0$) and (ii) expected value greater than zero ($\mathbb{E}(g) > 0$). An index Q is defined as a mapping that assigns a positive real number to each gamble that satisfies the assumptions above. Of course, not every index provides a meaningful summary of a gamble's riskiness. Aumann and Serrano (2008) argue that a reasonable risk index should satisfy the following two axioms:

D *Duality:* If i and j are two agents, such that i is uniformly more risk averse than j,[1] and if i accepts a gamble g at wealth w and $Q(g) > Q(h)$, then j accepts the gamble h at wealth w.

H *Homogeneity:* For any positive real number t, it holds that $Q(tg) = tQ(g)$.

Aumann and Serrano (2008) show that any two indexes that satisfy **D** and **H** are positive multiples of each other. Moreover, a specific index that satisfies **D** and **H** can be obtained as the positive solution to the equation

$$\mathbb{E}\left(\mathrm{e}^{-\frac{g}{r}}\right) = 1. \tag{2.1}$$

[1] Agent i is called uniformly more risk averse than agent j, if the fact that i accepts a gamble at some wealth implies that j accepts that gamble at any wealth.

We refer to the solution of (2.1) as the AS index $r_{AS} = r_{AS}(g)$. If the gamble g takes on only finitely many values Aumann and Serrano (2008) show that the assumptions (i) and (ii) are necessary and sufficient for a unique positive solution to (2.1). However, this is no longer the case, if the finiteness assumption on gambles is dropped, as we will discuss below.

2.3 The adjustment coefficient

The adjustment coefficient, denoted by r_{AC}, is a well established quantity in ruin theory (see for instance Grandell, 1991). It plays a crucial role in the assessment of the ruin probability for a wealth process W_j, $j = 0, 1, 2, \ldots$. To describe this process, let $g_0, g_1, g_2, \ldots$ be a sequence of independent and identically distributed gambles and $W_0 > 0$ the initial wealth. The wealth process evolves according to

$$W_n = W_0 + g_0 + g_1 + \ldots + g_{n-1} \qquad \forall\, n \geq 1. \tag{2.2}$$

One can think of W_0 as the starting capital of an insurance company and g_j as the premium payments the company receives minus the claims it has to pay during period j. Ruin or bankruptcy occurs if $W_j < 0$ for some $j \geq 1$. Let $N = \inf\{j \geq 1 : W_j < 0\}$, where conventionally $\inf \emptyset = \infty$. Denote the probability of ruin, that depends on the initial wealth W_0, by $\psi(W_0) = \mathbb{P}(N < \infty)$.

To assess $\psi(W_0)$, we first need to define the adjustment coefficient r_{AC}. For a gamble g it is given by the positive solution of the equation

$$\mathbb{E}\left(\mathrm{e}^{-rg}\right) = 1. \tag{2.3}$$

If r_{AC} exists it is also unique (see below). Now, if r_{AC} is the adjustment coefficient of g_1, then

$$\psi(W_0) = \frac{\mathrm{e}^{-r_{AC}W_0}}{\mathbb{E}\left(\mathrm{e}^{-r_{AC}W_N} \,|\, N < \infty\right)} \leq \mathrm{e}^{-r_{AC}W_0}, \tag{2.4}$$

with $W_0 > 0$.

The similarity between the defining equation for the adjustment coefficient (2.3) and for the AS index (2.1) is striking. We elaborate on it in the next section.

2.4 An operational interpretation of the Aumann-Serrano index of riskiness

In this section we provide an operational interpretation of the Aumann-Serrano index. From the similarity between equations (2.1) and (2.3), it is obvious, that the AS index exists if and only if the adjustment coefficient exists. As both are uniquely determined (see below), it follows that

$$r_{AS} = \frac{1}{r_{AC}}. \tag{2.5}$$

Now assume that we want to ensure that the probability of ruin for a wealth process described by (2.2) does not exceed a certain level, say $p \in (0,1)$. Using (2.4) and (2.5) we have

$$\begin{aligned} \psi(W_0) \leq \mathrm{e}^{-W_0/r_{AS}} &= p \\ \Leftrightarrow \quad W_0 &= \log(1/p) r_{AS}. \end{aligned}$$

Thus, an initial wealth level of $\log(1/p)r_{AS}$ ensures that the probability of bankruptcy is no larger than p. For small p, $\log(1/p)r_{AS}$ is close to the minimal wealth $W_*(p)$ that fulfills the above requirement. To see this, note that the denominator in (2.4) is constant with respect to p. Therefore,

$$\log(1/p)r_{AS} \geq W_*(p) \geq \log(1/p)r_{AS} - O(1)$$

and thus

$$W_*(p)/(\log(1/p)r_{AS}) \to 1$$

as p goes to zero.

For every $p \in (0,1)$, $\log(1/p)r_{AS}$ is an index satisfying **D** and **H**. The AS-index arises as a special case for $p = \mathrm{e}^{-1}$ and can be interpreted as the approximate wealth required to ensure no bankrupty with a probabiltiy of 37%.

Note that the operational interpretation of the AS-index presented here does not depend on a specific preference order. In contrast to that, Aumann and Serrano (2008) provide results that come close to an operational interpretation regarding agents with constant relative risk aversion. Moreover their interpretation is only applicable to bounded gambles, while the interpretation given here only presumes that the AS-index exists.

2.5 Existence of the AS index for non-finite gambles

Aumann and Serrano (2008) assume that the gambles, a decision maker faces, have finitely many outcomes. This excludes many distributions used in financial applications. It is not obvious which conditions are needed to guarantee the existence of the AS index, if we drop the above assumption. Again, the one-to-one relation between the AS index and the adjustment coefficient will prove extremely useful to solve this problem.

To begin with, denote the moment generating function (mgf) of a gamble or a random variable g as

$$M_g(t) = \mathbb{E}\left(\mathrm{e}^{tg}\right).$$

The mgf is a well-known concept in probability theory and statistics to derive the moments (of functions) of random variables. Clearly, if there is a unique $t_R < 0$ such that $M_g(t_R) = 1$, it immediately follows that the AS index is given by $r_{AS} = -1/t_R$ and the adjustment coefficient is given by $r_{AC} = -t_R$. However, it is not clear whether such a t_R exists.

Again,we can draw upon results for the adjustment coefficient established by Mammitzsch (1986). The above arguments show that the conditions for the existence of the AS index are the same. Mammitzsch (1986) has shown that the necessary conditions for the existence of the adjustment coefficients for general gambles are identical to those required by Aumann and Serrano (2008) for bounded gambles with finitely many outcomes

(i) negative outcomes are possible, $\mathbb{P}(g < 0) > 0$,

(ii) the expected outcome of the gamble is positive, $\mathbb{E}(g) > 0$.

However, for continuous distributions these conditions are not sufficient. In fact, if we define the interval of a finite mgf as $I_g = \{t \in \mathbb{R} : M_g(t) < \infty\}$ and require that the lower bound

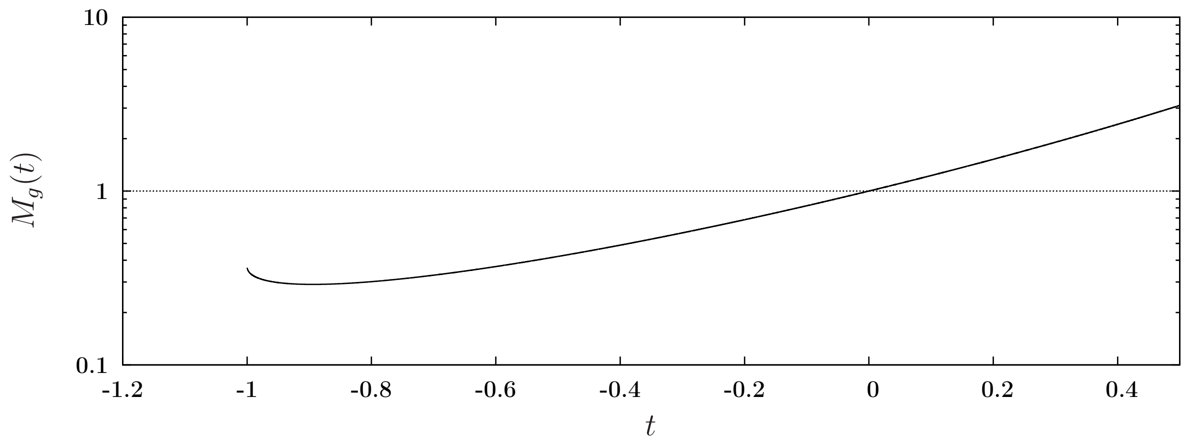

Figure 2.1: This figure shows the mgf of a NIG distributed gamble.

$l = \inf I_g < 0$, then the conditions (i) and (ii) are only sufficient if $l \notin I$. If $l \in I$ the additional condition

(iii) $M_g(l) \geq 1$

must be satisfied. Observe that, if a gamble g has only finitely many outcomes then $I_g = (-\infty, \infty)$ and (i) and (ii) are sufficient for the existence of the AS index.

Provided that the AS index/the adjustment coefficient exists, uniqueness is automatically ensured. Basically, the arguments made by Aumann and Serrano (2008) for gambles with finitely many outcomes also hold for more general gambles.

We illustrate the practical relevance of condition (iii) by considering a normal inverse Gaussian (NIG) distributed gamble. The NIG distribution is frequently used in the field of financial econometrics, financial economics and mathematical finance (cf. Cont and Tankov (2004) and the references therein).

The mgf for a NIG distributed gamble is given by

$$t \mapsto \mathrm{e}^{\mu t + \delta\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + t)^2}\right)},$$

with $0 \leq |\beta| < \alpha$, $\delta > 0$ and $\mu \in \mathbb{R}$. It is finite on the *closed* interval around zero $[-(\alpha + \beta), \alpha - \beta]$ and therefore conditions (i) and (ii) are not sufficient for the existence of the risk index. It will be shown in the next chapter that the additional condition (iii) is satisfied for this gamble whenever $\mu \leq \delta(\alpha - \beta)/\sqrt{\alpha^2 - \beta^2}$. A case where this does not hold is depitced in Figure 2.1, where the parameters are chosen as $\alpha = 1$, $\beta = 0$, $\mu = 0$, and $\delta = 1$. For negative arguments, the mgf stays below one and then discretely jumps to infinity at -1, i.e. it never hits 1 in the negative reals.

2.6 Conclusion

We have shown that the AS index allows for an operational interpretation in terms of wealth and ruin probabilities. The interesting link between the AS index and the Foster and Hart (2009) measure is provided by the adjustment coefficient, a well-known parameter commonly used to asses the ruin probability of an insurance company.

In addition to this new interpretation the so far unnoticed relation between the adjustment coefficient and the AS index has also important implications, as well-known concepts from ruin theory can be transferred to the latter. As an example we use the relation to derive conditions for the existence of the AS index for gambles with uncountably many outcomes.

Chapter 3

Beyond the Sharpe ratio: an application of the Aumann-Serrano index to performance measurement

3.1 Introduction

Performance measures are important tools for management decisions. They induce a total (or partial) order of investment opportunities so that agents can reduce their decisions regarding these investments to a simple comparison of these coefficients. Such decisions are, for instance, concerned with ranking investment opportunities or evaluating money managers, such as fund managers.

The *Sharpe ratio* (introduced as and also called *reward-to-variability ratio*), proposed by Sharpe (1966, 1994), is one of the most prominent performance measures. It is the ratio of the mean over the standard deviation of the expected excess return of an investment opportunity. It thus corrects the expected return by taking into account a specific type of risk taken by the investor. Its justification requires some restrictions on either the distributions of the returns or the investor's preferences. The typical distributional assumption is that all returns under consideration belong to the same location-scale family, see Meyer (1987) and Schuhmacher and Eling (2011) for a recent application of this argument. The most common example is the normal distribution. Moreover, the Sharpe ratio is adequate if investors only care about the mean and variance of an investment. However, it is well known that financial returns very often exhibit non-normal characteristics, such as (negative) skewness and excess kurtosis, which differ between assets (cf. Agarwal and Naik, 2004; Malkiel and Saha, 2005). This rules out the case that return distributions belong to the same location-scale family. Furthermore, empirical and experimental studies show that it is unlikely that investors do not care about these higher order moments (cf. Golec and Tamarkin, 1998; Harvey and Siddique, 2000). A prominent field of application for the Sharpe ratio is fund-ranking. Since the Sharpe ratio ignores differences in higher order moments the question arises as to what extent the ranking changes if one accounts for non-normality.

This has led to the development of various performance measures which take into account these stylized facts of financial returns. Most of them either replace the mean with a different reward measure or they substitute the standard deviation with a different measure of the (relevant) risk taken by the investor, or both. However, most of these measures are proposed in a rather *ad hoc* way and their economic foundation is rather vague. For an overview we refer to Cogneau and Hübner (2009a,b), Eling and Schuhmacher (2007) and Farinelli et al. (2008) and the references therein.

In this chapter we propose a new performance measure that, in contrast to the Sharpe ratio, meets the natural requirement that it is strictly monotone with respect to stochastic dominance and can account not only for the mean and variance but also for higher moments. The performance measure is obtained by dividing the mean of an investment opportunity by its *economic index of riskiness* proposed by Aumann and Serrano (2008) (AS index henceforth). As opposed to the risk measures mentioned above, the AS index is mainly derived from a choice theoretic axiom, namely the axiom of duality. It requires an index to reflect the following natural notion of *less risky*: given that an investment is accepted by some agent, less risk averse agents accept *less risky* investments. As such it is an economically motivated axiom and Aumann and Serrano (2008) therefore termed their index *economic* index of riskiness. To emphasize that our performance measure is based on such an economic risk measure we refer to it as the *economic performance measure* (EPM). If investment returns are normally distributed, the EPM and the

Sharpe ratio produce the same ranking of these investments. The EPM, thus, generalizes the Sharpe ratio with respect to non-normal distributions.

Moreover, we extend the continuity result of Aumann and Serrano (2008) and show that if the distribution of the returns converges to the normal distribution, the EPM converges to two times the squared Sharpe ratio. Thus, the EPM also asymptotically induces the same ranking as the Sharpe ratio. This is especially appealing in connection with the *aggregational Gaussianity* property of financial returns. This property states that for decreasing sampling frequency, e.g. going from daily to monthly and down to yearly returns, the return distribution approximates the normal distribution, see e.g. Cont (2001) and Rydberg (2000). While the Sharpe ratio is appropriate for low frequency returns, the new performance measure is appropriate for both low and high frequency returns, with no disadvantages compared to the Sharpe ratio in the former case.

We propose a parametric and a non-parametric moment estimator for the EPM. For parametric estimation we assume that returns follow a normal inverse Gaussian (NIG) distribution proposed by Barndorff-Nielsen (1997). As the NIG distribution is analytically tractable and has several attractive properties it is widely used in financial applications. It allows to model skewness and semi-heavy tails. We derive a closed form expression for the EPM of NIG-distributed random variables (e.g. excess returns) in terms of the first four moments. This makes explicit the dependence on skewness and kurtosis and provides a moment estimator for the EPM that is virtually as easy to compute as the Sharpe ratio. For non-parametric estimation the crucial idea is to use a moment condition that corresponds to the defining equation of the AS index. Results on asymptotic normality can readily be inferred from the literature on the method of moments. In a simulation study we address the issue of estimation uncertainty. Given that higher moments are important to investors, our results suggest that even for data sets with a limited number of observations, rankings based on the EPM are superior to Sharpe ratio rankings.

We apply our two estimators to rank mutual funds and hedge funds via the EPM and compare the results with a Sharpe ratio ranking. Imposing the parametric assumption of

NIG-distributed returns yields a ranking of the funds that is very similar to the one implied by the non-parametric estimation of the EPM, which indicates that the NIG-distribution is a reasonable choice. While the distributions of the excess returns from the mutual funds are close to Gaussian, the distributions of the hedge funds returns show pronounced skewness and excess kurtosis. As a consequence, the ranking of the mutual funds is very similar under the Sharpe ratio and the EPM. For the hedge funds, however, the two measures yield different rankings. In particular, if a fund's return distribution has relatively low (high) skewness and/ or relatively high (low) excess kurtosis, the fund is typically ranked lower (higher) by the EPM than by the Sharpe ratio.

The remainder of this chapter is structured as follows. The next section introduces the economic performance measure. We derive properties of this new measure and discuss its relation to the Sharpe ratio and other performance measures. In Section 3.3 we suggest estimators for the EPM and conduct a Monte Carlo experiment. Section 3.4 provides an empirical illustration using mutual funds and hedge funds return data. Section 3.5 concludes. The Appendix contains supplementary calculations.

3.2 An economic performance measure

Let $\tilde{r}$ denote the (stochastic) return of an investment portfolio, r^f the (deterministic) risk-free rate and r the resulting (stochastic) excess return. We define the economic performance measure as the expected excess return relative to the AS index of riskiness of this return:

$$\mathsf{EPM}\,(r) = \frac{\mathbb{E}\,(r)}{\mathsf{AS}\,(r)} = \frac{\mathbb{E}\,(\tilde{r}) - r^f}{\mathsf{AS}\,(\tilde{r} - r^f)}. \tag{3.1}$$

Thus, in contrast to the Sharpe ratio, the EPM divides the mean excess returns by its AS index instead of dividing by its standard deviation. This has important implications for the properties of the performance measure. In order to derive these properties we first briefly review the AS index. In Section 3.2.2 we present the properties of the EPM. In particular, we generalize the continuity property of the AS index, see Aumann and Serrano (2008), and extend this property to the EPM. We discuss the relation to the Sharpe ratio and other performance measures in

Section 3.2.3.

3.2.1 The Aumann-Serrano index of riskiness

Aumann and Serrano's (2008) index of riskiness is an axiomatic approach to assign an objective meaning to the word *risky*. It enables a decision maker to assess which of two investments is *riskier* without referring to a specific utility function or preference order. Similar to the standard deviation or *value at risk* (VaR), the AS index summarizes the properties of a *gamble* in a single number, thereby making comparisons very easy.

Note that although the AS index was proposed for gambles in terms of absolute outcomes, it can straightforwardly be applied to excess returns. Indeed, the excess return $\tilde{r} - r^f$ can be regarded as the outcome of a zero investment strategy that consists in borrowing $1 and investing it in a risky asset for a given time span. In the remaining part of this article we will refer to "the gamble" as the "excess return" with this zero investment strategy.

Let an index Q be a mapping that assigns a positive real number to each excess return/ random variable with values in $\mathbb{R}$. Of course, not every index provides a meaningful summary of an investment's "riskiness". Aumann and Serrano (2008) argue that a reasonable risk index should satisfy the following two axioms:

- **D** *Duality:* If i and j are two agents, such that i is uniformly more risk averse than j,[1] and if i accepts an excess return $r^{(A)}$ at wealth w and $Q(r^{(A)}) > Q(r^{(B)})$, then j accepts the excess return $r^{(B)}$ at wealth w.

- **H** *Homogeneity:* For any positive real number t, it holds that $Q(tr^{(A)}) = tQ(r^{(A)})$.

Here, for $t > 0$, $tr^{(A)}$ is the gamble/ excess return that results from $r^{(A)}$ by multiplying every outcome of $r^{(A)}$ by t. It is quite natural to think that if the stakes are doubled, the risk is also doubled, i.e. to require homogeneity. Note that axiom **D** is actually a mild requirement on an index, since the numbers assigned to two gambles only have an implication for an agent's decision if strong preconditions are satisfied. Axiom **D** basically says, that if an agent accepts

[1] Agent i is called uniformly more risk averse than agent j, if the fact that i accepts an excess return at some wealth implies that j accepts that excess return at any wealth.

an excess return $r^{(A)}$, then a less risk averse agent should accept an excess return $r^{(B)}$ that is less risky than $r^{(A)}$ according to the index.

Aumann and Serrano (2008) show that any two indices that satisfy **D** and **H** are positive multiples of each other. Moreover, a specific index that satisfies **D** and **H** can be obtained as the positive solution to the equation

$$\mathbb{E}\left(\mathrm{e}^{-\frac{r^{(A)}}{s}}\right) = 1. \tag{3.2}$$

The value of $s > 0$ that solves (3.2) is referred to as the AS index of $r^{(A)}$, $\mathsf{AS}\left(r^{(A)}\right)$.

The AS index of riskiness is *objective* in the sense that it does not depend on the preferences of an individual agent. From the duality and homogeneity property it follows that the AS index is also *subadditive*. Furthermore, Aumann and Serrano (2008) demonstrate that it is monotone with respect to first and second order stochastic dominance. Since *stochastic dominance* plays an important role, we briefly review it in the next paragraph. For a survey on this topic see Levy (1992).

In the theory of choice under uncertainty, stochastic dominance is a widely acknowledged concept. If an investment A stochastically dominates an investment B, then a "large" group of investors will prefer A over B. More specifically, let F_A and F_B be the distribution functions of the excess returns $r^{(A)}$ and $r^{(B)}$ corresponding to A and B, respectively. A first order stochastically dominates B ($A \overset{①}{\succ} B$), if $F_A(x) \leq F_B(x)$ for all $x \in \mathbb{R}$ and $F_A(x) < F_B(x)$ for some $x \in \mathbb{R}$. If this is the case, then any decision maker with increasing utility function prefers investment A over investment B. Moreover, the converse is also true. Unfortunately, many investments cannot be ordered by first order stochastic dominance. A less restrictive assumption about the relation of two gambles is made by second order stochastic dominance ($\overset{②}{\succ}$). We say that an investment A second order stochastically dominates an investment B ($A \overset{②}{\succ} B$), if $\int_{-\infty}^{y}[F_A(x) - F_B(x)]\,\mathrm{d}\,x \leq 0$ for all $y \in \mathbb{R}$ and strict inequality holds for some $y \in \mathbb{R}$. Second order stochastic dominance is implied by first order stochastic dominance. Similar to first order stochastic dominance, it holds that, $A \overset{②}{\succ} B$ *iff* any investor with increasing and concave utility function prefers A to B. Regarding risk measures, it is natural to require that a risk measure classifies alternative A as less risky than B if all risk averse investors unanimously prefer A over B. The AS index has this property referred to as monotonicity with respect to stochastic dominance.

Finally, it remains to clarify the question about its existence. If the excess return r takes on only finitely many values, Aumann and Serrano (2008) show that the following assumptions are necessary and sufficient for a unique positive solution to (3.2):

(i) possibly negative outcomes, i.e. $\mathbb{P}(r < 0) > 0$, and

(ii) expected value greater than zero, i.e. $\mathbb{E}(r) > 0$.

However, for continuous distributions, i.e. distributions with uncountable many outcomes, the existence of this index is more delicate. The *moment generating function* (mgf) of an excess return r is defined by

$$M_r(t) = \mathbb{E}\left(e^{tr}\right).$$

Furthermore, let $I_r = \{t \in \mathbb{R} : M_r(t) < \infty\}$ and $l = \inf I_r < 0$. In the previous chapter we have shown that in addition to (i) and (ii) the following condition must also be satisfied to ensure the existence of the AS index:

(iii) $l \notin I_r$ or $M_r(l) \geq 1$.

Note that for excess returns with finitely many outcomes it holds that $l = -\infty$ and, thus, (iii) is satisfied.

3.2.2 Properties of the economic performance measure

We now derive useful properties of the EPM. Some of these are easily inferred from the properties of the AS index. Additionally, we also derive a general form of continuity for the AS index and extend it to the EPM. In contrast to Aumann and Serrano (2008), this *generalized continuity* also applies to excess returns with infinitely many possible outcomes, which is typically the case in financial return modeling.

Scale invariance

Both the numerator and the denominator of the EPM (3.1) are homogeneous, so that the EPM is scale invariant. Therefore the scale of the investment can be set to \$1 without loss of generality, identifying the excess return as the zero investment strategy using \$1.

Interpretation

The EPM can be interpreted as a measure of *reward to required capital.* In the foregoing chapter it has been shown that, for $0 < p < 1$ and p close to zero, $\log(1/p)\mathsf{AS}\left(r\right)$ is approximately equal to the minimum required initial wealth that assures no bankruptcy with probability $1-p$ when playing the zero investment strategy with excess return r infinitely often.[2] We do not use the factor $\log(1/p)$, however, since it is irrelevant for ranking purposes.

Stochastic dominance

Beyond the fact that the EPM uses a risk measure that is (strictly) monotone with respect to first ("$\overset{①}{\succ}$") and second order ("$\overset{②}{\succ}$") stochastic dominance, the EPM itself has this property. This follows immediately from the fact that the mean is monotone with respect to stochastic dominance.[3] In other words, if $r^{(A)} \overset{①}{\succ} r^{(B)}$ or $r^{(A)} \overset{②}{\succ} r^{(B)}$, then $\mathsf{AS}\left(r^{(A)}\right) < \mathsf{AS}\left(r^{(B)}\right)$ and $\mathbb{E}\left(r^{(A)}\right) \geq \mathbb{E}\left(r^{(B)}\right)$ and therefore $\mathsf{EPM}\left(r^{(A)}\right) > \mathsf{EPM}\left(r^{(B)}\right)$.

Normally distributed returns

For normally distributed excess returns, $r^{(N)} \sim \mathcal{N}\left(\mu, \sigma^2\right)$, Aumann and Serrano (2008) determine the index of riskiness as $\sigma^2/(2\mu)$. It follows that the EPM is given by

$$\mathsf{EPM}\left(r^{(N)}\right) = \frac{\mu}{\mathsf{AS}\left(r^{(N)}\right)} = \frac{2\mu^2}{\sigma^2}.$$

[2] To be more explicit, let W_0 denote initial wealth and let $\{r_n\}_{n\geq 1}$ a sequence of independent and identically distributed excess returns, for which the AS index exists. Wealth at time n is given by $W_n = W_0 + r_1 + r_2 + \ldots + r_n$. Bankruptcy occurs as soon as $W_n < 0$.

[3] To see this, note that $u(x) = x$ is increasing and concave and thus "$r^{(A)} \overset{①}{\succ} r^{(B)}$" or "$r^{(A)} \overset{②}{\succ} r^{(B)}$" implies $\mathbb{E}\left(r^{(A)}\right) = \mathbb{E}\left(u(r^{(A)})\right) \geq \mathbb{E}\left(u(r^{(B)})\right) = \mathbb{E}\left(r^{(B)}\right)$.

In this case the EPM equals two times the squared Sharpe ratio, $\mathsf{EPM}\left(r^{(N)}\right) = 2\mathsf{SR}^2\left(r^{(N)}\right)$, which means that the two measures produce identical rankings. Hence, for normally distributed excess returns, where the Sharpe ratio is an appropriate performance measure, the EPM suits equally well. But the EPM is also suitable, when other moments of the distribution, such as skewness and kurtosis, are important. In the next paragraph we supplement this result and show that, under some regularity conditions, as the distribution of an excess return r approximates the normal distribution, the EPM of r approximates the EPM of a normally distributed excess return.

Generalized continuity

We show that the continuity property of the AS index also applies to more general cases than those considered by Aumann and Serrano (2008). From this and an additional assumption the continuity of the EPM follows. The above mentioned result concerning the approximation of normally distributed returns is a special case of this *generalized continuity* property of the EPM.

Let r_0 and $\{r_n\}_{n\geq 1}$ be random variables/ excess returns and denote their moment generating function as $M_n(t) = \mathbb{E}\left(\mathrm{e}^{tr_n}\right)$ $(n \geq 0)$. Let "$\xrightarrow{d}$" denote convergence in distribution.

Assumption 3.1. *The economic index of riskiness* $\mathsf{AS}\left(r_n\right)$ *exists for all* $n \geq 0$.

Assumption 3.2. *There exists a real number* $b > 1/\mathsf{AS}\left(r_0\right)$ *such that* $\sup_n M_n(-b) < \infty$.

Assumption 3.2 means that the left tails of the return distributions should not be too heavy. With these two assumptions we can state the following

Proposition 3.3 (Generalized Continuity)**.** *If Assumptions 3.1 and 3.2 hold, then* $r_n \xrightarrow{d} r_0$ *implies* $\mathsf{AS}\left(r_n\right) \to \mathsf{AS}\left(r_0\right)$.

Proof. The proof of Proposition 3.3 is given in the appendix to this chapter. The main tool is Theorem 5.4 in Billingsley (1968).

□

Note that Aumann and Serrano (2008) assume that the excess returns satisfy conditions (i) and (ii), take only finitely many values, and that they be uniformly bounded.[4] These assumptions are stronger and indeed imply Assumptions 3.1 and 3.2. First, requiring that the returns take on only finitely many values and satisfy conditions *(i)* and *(ii)* assures that the AS index exists for the returns under consideration. Second, requiring that the sequence of returns be uniformly bounded implies Assumption 3.2. To see this, let $K > 0$ such that $|r_n| \leq K$ for all n. Then, for $b > 1/\mathsf{AS}(r_0)$, $\mathrm{e}^{-br_n} \leq \mathrm{e}^{bK}$ almost surely for all n. Therefore, $M_n(-b) = \mathbb{E}\left(\mathrm{e}^{-br_n}\right) \leq \mathrm{e}^{bK}$ for all n and Assumption 3.2 holds.

So, we have the proof for *generalized continuity* for the AS index. To achieve continuity for the EPM we have to make an additional assumption:

Assumption 3.4. *The sequence $\{r_n\}_{n\geq 1}$ is uniformly integrable.*[5]

Corollary 3.5 (*Generalized Continuity for EPM*)**.** *If Assumptions 3.1, 3.2, and 3.4 hold, then $r_n \xrightarrow{d} r_0$ implies $\mathsf{EPM}(r_n) \to \mathsf{EPM}(r_0)$.*

Assumption 3.4 together with $r_n \xrightarrow{d} r_0$ imply that $\mathbb{E}(r_n) \to \mathbb{E}(r_0)$ (cf. Billingsley, 1968). Then the corollary immediately follows from *generalized continuity*. Furthermore, note that the requirement of $\{r_n\}_{n\geq 1}$ being uniformly bounded, as in Aumann and Serrano (2008), guarantees that Assumption 3.4 holds.

Corollary 3.5 is not only of theoretical importance but also of practical relevance. Note, that the aggregational Gaussianity property of (excess) asset returns states that the distribution of less frequent sampled returns, e.g. going from daily to monthly and down to yearly returns, approximates the normal distribution. Under the assumptions of this section, this implies that the EPM approximates the EPM of normally distributed returns as the sampling frequency decreases. While the Sharpe ratio is appropriate for low frequency returns, the EPM is appropriate for both low and high frequency returns, with no disadvantages compared to the Sharpe ratio in the former case.

[4] A sequence of excess returns $\{r_n\}_{n\geq 1}$ is uniformly bounded, if there is a $K > 0$ such that, for all n, $|r_n| \leq K$ almost surely. Note that this precludes that the distribution of r_n converges to a normal distribution as $n \to \infty$.

[5] By definition, this means that $\lim_{\alpha\to\infty} \sup_n \mathbb{E}\left(|r_n| \mathbb{I}_{(|r_n|\geq\alpha)}\right) = 0$.

3.2.3 Relation to the Sharpe ratio and alternative performance measures

The Sharpe ratio of an investment opportunity is given by its mean excess return relative to the standard deviation

$$\mathsf{SR}(r) = \frac{\mathbb{E}(\tilde{r}) - r^f}{\sqrt{\mathbb{V}(\tilde{r})}},$$

which is sensible if $\mathbb{E}(\tilde{r}) > r^f$. The Sharpe ratio plays an important role in a mean-variance decision framework. There, the portfolio with the highest Sharpe ratio together with the risk free asset determines the set of (mean-variance) efficient investments, e.g. see Sharpe (1966) and Treynor (1965). The mean-variance framework is appropriate if investors have quadratic utility or if all returns r_i under consideration are equal in distribution to $a_i + b_i r$ for some generic return r and $a_i \in \mathbb{R}$, $b_i > 0$.[6] However, returns typically differ not only in location and scale but also in skewness and kurtosis, for instance. Such aspects are ignored by the Sharpe ratio.[7]

Furthermore, the Sharpe ratio is not monotone with respect to first order stochastic dominance. The following example provides an illustration. Consider two assets A and B, where A yields an excess return $r^{(A)}$ of either -1%, 1% or 5%, with probability 0.1, 0.45 and 0.45, respectively. Asset B yields an excess return $r^{(B)}$ of either -1%, 1% or 3%, with the same probabilities 0.1, 0.45 and 0.45, respectively. Then $\mathsf{SR}\left(r^{(A)}\right) \approx 1.16 < 1.30 \approx \mathsf{SR}\left(r^{(B)}\right)$, although it would be natural to prefer A over B, since A performs strictly better than B in one case and not worse than B in the other cases.

There is a variety of performance measures that aim to overcome the drawbacks of the Sharpe ratio (cf. Cogneau and Hübner, 2009a,b). Typically, they replace the standard deviation in the denominator of the Sharpe ratio by a different risk measure and in some cases use a

[6] Instead of requiring that all r_i belong to a certain location-scale family, one can assume that this holds for $h(r_i)$, where h is some monotonically increasing and concave function. This, however, comes at the cost of restrictions on the utility function u: $u \circ h^{-1}$ has to be a concave function, with h^{-1} denoting the inverse of h. This was investigated by Boyle and Conniffe (2008), who considered two parameter distributions for r_i.

[7] To test the location-scale condition, Meyer and Rasche (1992) consider a generalization, where a_i is replaced by $c + a_i z$, with c a constant and z a random variable that is independent of r. Assuming that $\mathbb{E}(r) = 0$, expected utility can be represented as function of the mean and the variance alone. However, it is an open question under which conditions this function is increasing in the mean or quasi-concave.

different reward measure in the numerator. The EPM also falls into the category of these so-called reward-to-risk measures. Like other measures, the EPM is not in all dimensions superior to all other reward-to-risk measures. First, there is no consensus about *the* relevant properties of a performance measure. These depend rather on the context under which performance evaluation is carried out. Second, the EPM has several attractive features which give it an advantage over other performance measures in *some* respects as outlined below. A complete discussion of how the EPM relates to each single reward-to-risk measure is beyond the scope of this chapter. However, to gain a broad picture we consider relevant groups of reward-to-risk measures and give some examples. For the purpose of the following discussion we distinguish between the *ad hoc* approach, the *axiomatic* approach, and the *economic* approach.

Many reward-to-risk measures replace the standard deviation in the denominator of the Sharpe ratio by another statistical summary measure, for instance by the semi-standard deviation (cf. Sortino and Price, 1994) or *value at risk* (cf. Dowd, 2000). This rather *ad hoc* approach contrasts with the EPM, since the AS index is axiomatically founded on the theory of decision making under risk. Nevertheless, the EPM shares common features with certain *ad hoc* performance measures: The Sortino ratio (cf. Sortino and Price, 1994), the upside potential ratio (Sortino et al., 1999) and the Calmar ratio[8] also yield Sharpe ratio-equivalent rankings when returns are normal (cf. Lien, 2002, Schuhmacher and Eling, 2011). Furthermore, the interpretation of the EPM as reward-to-required-capital relates it to *excess return on value at risk* (cf. Dowd, 2000) and *drawdown-based* performance measures. Although these measures are very popular in practice, their theoretical properties are still subject to research, e.g. see Schuhmacher and Eling (2011).

As opposed to the EPM, *ad hoc* performance measures, in general, use risk measures that are not monotone with respect to stochastic dominance. We illustrate this for the case of the

[8] The Sortino ratio replaces the standard deviation with semi-standard deviation in the denominator of the Sharpe ratio. The upside potential ratio additionally replaces the mean with the so-called upside potential $\mathbb{E}(r\mathbf{1}_{\{r>0\}})$. The Calmar ratio is computed from dividing the mean excess return by the maximum drawdown for a certain period.

excess return on value at risk: The excess return on value at risk is given by

$$\mathsf{ERVaR}_{\alpha} = \frac{\mathbb{E}(r)}{\mathsf{VaR}_{\alpha}(r)},$$

where r denotes the excess return and $\mathsf{VaR}_{\alpha}(r)$ is the threshold that negative excess returns will not exceed with a given probability α. Now, consider the 90% VaR and let A and B be two assets with excess returns $r^{(A)}$ and $r^{(B)}$. Assume that $r^{(A)}$ takes on the values -3%, -1% or 4% with probabilities 5%, 5% and 90%, respectively, while $r^{(B)}$ takes on the values -2% or 4% with probabilities 10% and 90%, respectively. It is easily verified that B second order stochastically dominates A, and thus all profit seeking and risk averse investors prefer B over A. If the risk, instead, is measured with respect to the VaR, then A is preferred over B as $\mathsf{VaR}_{0.9}\left(r^{(A)}\right) = 1\% < 2\% = \mathsf{VaR}_{0.9}\left(r^{(B)}\right)$. Since the mean excess return of both assets is equal, it also follows that $\mathsf{ERVaR}_{0.9}\left(r^{(A)}\right) > \mathsf{ERVaR}_{0.9}\left(r^{(B)}\right)$, which is again at odds with the choice of risk averse investors. However, ERVaR is widely used because of regulatory requirements. This is a case where the context has considerable influence on the choice of the (performance) measure.

We now turn to the axiomatic approach. Artzner et al. (1999) propose so-called *coherent risk measures*. But, their axioms give rise to a whole class of risk measures (instead of determining one) and it is not guaranteed that they are monotone with respect to stochastic dominance either. De Giorgi (2005) proposes an axiomatic definition for both the risk and the reward measure. In both cases he requires that the measures are monotone with respect to second order stochastic dominance. The only reward measure that fulfills all the requirements of De Giorgi (2005) is the expected value. Moreover, to pin down a particular risk measure a distortion of the objective probability measure has to be chosen. The author suggests that this choice should depend on the risk characteristics of the investor, i.e. the utility function. The EPM, on the other hand, uses objective reward and risk measures (independent of individual preferences) and is thus more generally applicable.

Finally, Zakamouline and Koekebakker (2009) consider a simple economic model to define their performance measure. They measure the performance of a risky asset as the maximal expected utility an agent can achieve by allocating his wealth between the risky and the riskless asset. Again, this contrasts with the objectivity of the EPM. Also based on an economic model, Foster and Hart (2009) suggest an *operational measure of riskiness* that, similarly to the AS index, has a unique characterization and is independent of a specific preference relation. It also respects stochastic dominance and is numerically quite close to the AS index. However, since its restriction on the heaviness of the tails is rather strong, it cannot be directly applied to distributions that are commonly used in modeling financial returns.

3.3 Estimation of the economic performance measure

In the following we discuss the estimation of our performance measure. In empirical applications the probability measure $\mathbb{P}$ is not observable but instead we observe n independent and identically distributed (iid) realizations $\{r_1, \ldots, r_n\}$ of the excess return r.[9] Based on these observations we want to estimate the EPM, the ratio of the mean of the excess returns over the corresponding AS index of riskiness. As the estimation of the mean is straightforward we will concentrate on the estimation of the AS index.

3.3.1 Parametric estimation and the normal inverse Gaussian distribution

A typical parametric estimation approach consists of choosing a reasonable parametric distribution, estimating the parameters of this distribution and, finally, computing the AS index based on these estimates. Distributions for which the AS index is available in closed form are for instance the normal distribution (see Aumann and Serrano, 2008) and the normal inverse Gaussian distribution (see below) or the exponential, Gamma, Variance-Gamma, and Poisson distributions (see Schulze, 2010). For other distributions the moment generating function (mgf) might be available in closed form as a function of the distribution parameters, while the AS index itself is not. In that case one can compute the mgf using estimated parameters and then

[9] Although we assume iid random variables in the following this is not essential and the estimator can be straightforwardly generalized to a more general dependence structure.

apply a numerical algorithm to solve for the AS index.[10]

Given its empirical adequacy, we suggest using the normal inverse Gaussian (NIG) distribution for parametric estimation. The NIG distribution is a well established distribution in finance, econometrics and statistics. It is used, for example, to model unconditional as well as conditional return distributions, see e.g. Andersson (2001); Bollerslev et al. (2009) as well as Eriksson et al. (2009). Zakamouline and Koekebakker (2009) use this distribution for performance measurement (cf. Section 3.2.3).

We derive the AS index and the EPM for NIG-distributed returns. Our representation of the EPM in terms of the mean, variance, skewness, and excess kurtosis offers a simple estimation scheme for the EPM. Furthermore, it makes explicit the role of higher order moments, which are neglected in the Sharpe ratio.

A NIG-distributed excess return (random variable), $r^{(NIG)} \sim \mathcal{NIG}(\alpha, \beta, \nu, \delta)$, is characterized by the following density

$$f^{(NIG)}(x; \alpha, \beta, \nu, \delta) = \frac{\alpha\delta}{\pi} \frac{K_1\left(\alpha\sqrt{\delta^2 + (x-\nu)^2}\right)}{\sqrt{\delta^2 + (x-\nu)^2}} \, \mathrm{e}^{\delta\gamma + \beta(x-\nu)} \tag{3.3}$$

with $0 \leq |\beta| < \alpha$, $\delta > 0$, $\nu \in \mathbb{R}$, $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $K_1(y) = (1/2)\int_0^\infty \mathrm{e}^{-(1/2)y(z+z^{-1})}\,\mathrm{d}\,z$ the modified Bessel function of the third kind with index 1. δ is a scaling parameter, ν is a location parameter, β is an asymmetry parameter and $\alpha \pm \beta$ determines the heaviness of the tails. It follows from (3.3) that the moment generating function is given by

$$\mathbb{E}\left(\mathrm{e}^{tr^{(NIG)}}\right) = M^{(NIG)} : [-(\alpha+\beta), \alpha-\beta] \to \mathbb{R} \quad t \mapsto \mathrm{e}^{\nu t + \delta(\gamma - \sqrt{\alpha^2 - (\beta+t)^2})}. \tag{3.4}$$

For the derivation of the EPM and the AS index we first have to ensure that the latter is well defined for NIG-distributed gambles. Note that the two assumptions (i) and (ii) in Section 3.2.1 are obviously satisfied whenever the mean $(\nu + \delta\beta/\gamma)$ is positive. However, (i) and (ii) are not sufficient for the existence of the risk index. For a NIG-distributed random variable condition (iii) is satisfied, if $\nu \leq \delta(\alpha - \beta)/\gamma$. Using the defining Equation (3.2) and Equation

[10] This is in fact very simple, since the problem is one-dimensional and the function under consideration is convex.

(3.4) the AS index and the EPM for NIG-distributed gambles are given by:

$$\mathsf{AS}^{(NIG)}(\alpha,\beta,\nu,\delta) = \mathsf{AS}\left(r^{(NIG)}\right) = \frac{1}{2}\frac{\delta^2+\nu^2}{\beta\delta^2+\gamma\delta\nu} \tag{3.5}$$

$$\mathsf{EPM}^{(NIG)}(\alpha,\beta,\nu,\delta) = \mathsf{EPM}\left(r^{(NIG)}\right) = \frac{2(\nu+\delta\beta/\gamma)(\beta\delta^2+\gamma\delta\nu)}{\delta^2+\nu^2}. \tag{3.6}$$

Details of the derivation are provided in the appendix to this chapter. Due to the parameter restrictions implied by condition (*ii*) the AS index is always positive, as it should be. It also can be easily verified that $-1/\mathsf{AS}^{(NIG)}(\alpha,\beta,\nu,\delta)$ is in the domain of $M^{(NIG)}$.

A representation of the AS index and of the EPM in terms of the moments is given by:

$$\widetilde{\mathsf{AS}}^{(NIG)}\left(\mu,\sigma^2,\chi,\kappa\right) = \left(3\kappa\mu - 4\mu\chi^2 - 6\chi\sigma + 9\sigma^2/\mu\right)/18 \tag{3.7}$$

$$\widetilde{\mathsf{EPM}}^{(NIG)}\left(\mu,\sigma^2,\chi,\kappa\right) = 18\mu/\left(3\kappa\mu - 4\mu\chi^2 - 6\chi\sigma + 9\sigma^2/\mu\right) \tag{3.8}$$

with mean $\mu > 0$, variance $\sigma^2 > 0$, excess kurtosis $\kappa > 0$ and skewness $|\chi| < \sqrt{3\kappa/5}$. The derivation is provided in the appendix to this chapter. Note that assumption (*iii*) can also be rewritten in terms of those moments as $\mu \leq 3\sqrt{\sigma^2/(3\kappa-4\chi^2)}$.

An obvious and efficient way to obtain parameter estimates is maximum likelihood estimation. In this case, (3.6) can be used to compute the EPM. On the other hand, (3.8) suggests estimating the EPM using empirical moments. This is computationally inexpensive compared to maximum likelihood. However, estimation risk will be an issue (cf. Bai and Ng, 2005). An in depth treatment of estimation risk is beyond the scope of this chapter, but we address this issue via a small simulation study in Section 3.3.3.

Lastly, note that for $\chi = 0$ and $\kappa \to 0$ in (3.8), we obtain the EPM for a normally distributed gamble with mean μ and variance σ^2: $(2\mu^2)/\sigma^2$. Thus, the order induced by the EPM for NIG-distributed returns approximates that induced by the Sharpe ratio if skewness and excess kurtosis go to zero. As opposed to the Sharpe ratio, $\widetilde{\mathsf{EPM}}^{(NIG)}$ can easily account for skewness and kurtosis in cases where these are not negligible.

3.3.2 Non-parametric estimation

The choice of a reasonable parametric distribution is not always obvious and in certain cases one might prefer a non-parametric approach to estimate the AS index of riskiness. A natural way to estimate the AS index is given by Method of Moments (MM) (see Hansen, 1982). In particular, set

$$f(x;s) = \mathrm{e}^{-x/s} - 1 \quad (s > 0)$$

and consider the moment equation

$$\mathbb{E}\left(f(r;s_0)\right) = \mathbb{E}\left(\mathrm{e}^{-r/s_0} - 1\right) = 0 \quad (s_0 > 0), \tag{3.9}$$

where r is the excess return/ random variable underlying the realizations $\{r_1, \ldots, r_n\}$. Note that (3.9) corresponds to the defining equation for the AS index (3.2), i.e. $s_0 = \mathsf{AS}(r)$. The MM-estimator $\hat{s}_n$, where we use the subscript n to express the dependence on the sample size, is given by the solution to the empirical counterpart of (3.9):

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{e}^{-r_i/s} - 1 = 0. \tag{3.10}$$

Equation (3.10) has to be solved numerically. A unique positive solution can always be found if some of the realizations r_i are negative and $\frac{1}{n}\sum_{i=1}^{n} r_i > 0$. By the strong law of large numbers, this will almost surely be the case, for a large sample size n, if the generic excess return r satisfies conditions (i) and (ii) in Section 3.2.1.

Within the MM setup for uncorrelated random variables the asymptotic distribution of the estimator is given by

$$\sqrt{n}\left(\hat{s}_n - s_0\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{S}{G_0^2}\right) \tag{3.11}$$

where

$$G_0 = \mathbb{E}\left(\left.\frac{\partial f(r;s)}{\partial s}\right|_{s=s_0}\right) = \frac{1}{s_0^2}\mathbb{E}\left(\mathrm{e}^{-r/s_0} r\right)$$

and

$$S = \mathbb{V}\left(f(r;s_0)\right) = \mathbb{E}\left(f(r;s_0)^2\right) = \mathbb{E}\left(\mathrm{e}^{-2r/s_0}\right) - 1 \tag{3.12}$$

assuming the existence of the necessary moments, i.e. the mgf of r has to exist at $-2/s_0$ (see (3.12)). The variance of $\hat{s}_n$ can be estimated using the empirical counterparts of G_0 and S and replacing s_0 with $\hat{s}_n$.

Some insight can be gained by looking at the above estimation procedure from a different perspective. Given realizations $\{r_1, \ldots, r_n\}$ of the random excess return r one can define the *empirical distribution function*

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(r_i), \tag{3.13}$$

where the indicator function $\mathbb{I}_{(-\infty,x]}(r_i)$ equals 1 if $r_i \leq x$ and 0 otherwise. $\hat{F}_n(x)$ is the distribution of a gamble/ excess return that takes on the values $\{r_1, \ldots, r_n\}$ each with probability $1/n$. The mgf pertaining to $\hat{F}_n(x)$ is the *empirical mgf* and is given by

$$\hat{M}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{e}^{tr_i}.$$

By the Glivenko-Cantelli theorem, $\hat{F}_n(x)$ converges uniformly to the distribution function of r.[11] Therefore, one can hope that solving $\hat{M}_n(-1/s) = 1\,(s > 0)$ will yield a good estimate of the AS index of r. In fact, this estimate is exactly $\hat{s}_n$.

In the following paragraphs we illustrate how Proposition 3.3 can be used to prove that $\hat{s}_n$ is strongly consistent. Furthermore, we demonstrate that mean-variance performance estimation can be considered to be a rough simplification of the non-parametric estimation of the EPM.

Generalized continuity and strong consistency of $\hat{s}_n$

Consistency of $\hat{s}_n$ as an estimator of the AS index can be obtained by applying *generalized continuity*. For this, the requirement that the mgf of r exists at $-2/\mathsf{AS}$ can be replaced by the weaker requirement that M_r exists at some b smaller than $-1/\mathsf{AS}$. Consider the sequence of gambles $f_1, f_2, \ldots$, where f_n follows the distribution $\hat{F}_n(x)$ in (3.13) with outcomes given by the realizations $\{r_1, \ldots, r_n\}$. Almost surely, the AS index of f_n exists for large n and equals $\hat{s}_n$

[11]Uniform convergence of a function F_n to F means that $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0$ as $n \to \infty$.

(assuming (i) and (ii) in Section 3.2.1 hold for r). Moreover, since M_r exists at some b smaller than $-1/\mathsf{AS}$, e^{br} is integrable. Then, by the strong law of large numbers, the mgf of f_n at b, $\hat{M}_n(b)$, converges to $M_r(b)$ almost surely. This implies that the sequence $\{\hat{M}_n(b)\}_{n\geq 1}$ is (almost surely) bounded. Since, by the Glivenko-Cantelli theorem, $f_n \xrightarrow{d} r$, we can apply *generalized continuity* and conclude that $\hat{s}_n \to s_0 = \mathsf{AS}\,(g)$ almost surely.

Truncation and the mean-variance framework

There is an interesting analogy to expected utility theory where mean-variance analysis is justified by assuming either normal returns or that the utility function can be reasonably well approximated by a second order Taylor-expansion. In fact, instead of solving (3.10) one could equally well solve $h(t) = \log[\hat{M}_n(t)] = 0$ $(t < 0)$. Denoting the solution as t^*, $\hat{s}_n$ is given by $-1/t^*$. The second order Taylor-expansion of h can be written as

$$\hat{h}(t) = \hat{\mu}_r t + \frac{1}{2}\hat{\sigma}_r^2 t^2,$$

where $\hat{\mu}_r$ and $\hat{\sigma}_r^2$ are the sample mean and variance of r. The solution of $\hat{h}(t) = 0$ $(t < 0)$ leads to an estimate $\hat{\sigma}_r^2/(2\hat{\mu}_r)$ for the AS index and $2\hat{\mu}_r^2/\hat{\sigma}_r^2$ for the EPM. The latter is the empirical counterpart of the EPM for normal returns. Thus, similarly to truncating the expansion of the utility function, approximating h by a second order Taylor-expansion leads back to the mean-variance decision framework. Second order Taylor series approximations, however, risk a serious loss of accuracy (cf. Loistl, 1976).

Relation to the adjustment-coefficient

There is a remarkable relation between the AS index and the so called *adjustment coefficient* (AC) from ruin theory. In fact one is the reciprocal of the other

$$\mathsf{AS} = \frac{1}{\mathsf{AC}}.$$

Pitts et al. (1996) propose to estimate the adjustment coefficient as the solution of $\hat{M}(-s) = 1\,(s > 0)$, which ends up being the reciprocal of $\hat{s}_n$. Without referring to the methods of moments, they show that their estimator is asymptotically normal with rate $\sqrt{n}$ and asymptotic variance

$$\frac{M_g(-2\mathsf{AC})-1}{M_g'(-\mathsf{AC})^2},$$

provided that the mgf of g, M_g, exists at $-2\mathsf{AC}$. Using the reciprocal relation between the AS index and the adjustment coefficient and applying the so-called Delta-method this could also have been derived from (3.11–3.12), or vice versa.

3.3.3 Estimation uncertainty: a small simulation study

So far we have taken the standpoint that an important drawback of the Sharpe ratio is that it only accounts for first and second moments. However, when it comes to the estimation of distributional characteristics other than the mean and the variance, the induced estimation error may thwart the merits of considering these characteristics in the performance ranking. For example, the estimation of extreme quantiles, as required for the estimation of the VaR, is notoriously challenging. The estimation of skewness and kurtosis also comes at the cost of a higher estimation error than that of the mean or variance. We therefore conduct a small simulation study, which allows us to analyze the sensitivity of the EPM with respect to estimation uncertainty.

We assume that the true ranking is given by the EPM based on complete knowledge of the underlying distribution. Mimicking the estimated NIG distributions for the hedge fund returns in Table 3.3 we simulate a synthetic data set consisting of 25 return series with 250 observations each. For this sample we estimate the Sharpe ratio and the EPM and compute the two corresponding rankings.[12] As the underlying distribution is known, we can compare the rankings based on the simulated data set with the true ranking. To measure the closeness between the true ranking and the estimated rankings based on the Sharpe ratio and the EPM we compute the rank correlation between these rankings. If the estimation error does not dominate the effect of using the correct performance measure we would expect that the rank correlation between the true ranking and the new performance measure is larger than that

[12]Note, that we focus here on the impact of the parametric estimation based on the NIG distribution. However, the results are nearly identical if we use the non-parametric estimation.

between the true ranking and the ranking based on the Sharpe ratio. To minimize the error due to the simulation we use 5000 replications resulting in 5000 rank correlation pairs for each performance measure. In 74.2% of these replications the rank correlation of the estimated EPM is larger than that of the Sharpe ratio. This means that in 74.2% of all cases the ranking based on the new performance measure is closer to the true ranking than the ranking based on the Sharpe ratio. Of course, one would prefer a value that is closer to 100%, but with the small number of observations (250) the estimation error of the additional distributional characteristics is still not negligible. Increasing the sample size, however, yields further improvements in the performance of our measure. In fact, for 1000 observations it outperforms the Sharpe ratio in 80.3% of the cases. We can thus conclude that, even for smaller sample sizes like the one encountered in our empirical analysis, our new performance measure systematically outperforms the Sharpe ratio.

3.4 Empirical illustration

For our empirical illustration we consider two data sets. In our first example we consider the 25 largest-growth mutual funds (as of January 1998 in terms of overall assets managed) and for our second example we use 25 hedge funds.

First, we consider monthly excess returns of mutual funds investments from January 1991 to September 2010 resulting in 237 observations. This is an extension of the data set employed by Bao (2009). However, we drop the period from 1987 to 1990 in order to match the time span for which hedge funds data are available. The excess returns are computed from monthly fund returns and the one-month US Treasury bill rate. Standard tests cannot reject the null hypothesis of no autocorrelation or no ARCH effects indicating that the iid assumption can be maintained in our application.

The estimated values of the different performance measures for the mutual funds are reported in Table 3.1. The second column reports estimates for two times the squared Sharpe ratio. The third and the fourth columns display the values of the EPM for the non-parametric approach and for the maximum likelihood based parametric estimates assuming NIG-distributed

Table 3.1: Performance measures for monthly excess returns of mutual funds (1991-2010).

name	$2SR^2$	EPMNon	EPMNIG	skewness	kurtosis
Amcap	0.0381 (5)	0.0364 (5)	0.0362 (5)	-0.4762 (7)	4.4698 (10)
American Cent-Growth	0.0179 (21)	0.0177 (20)	0.0176 (20)	-0.2420 (3)	3.5575 (1)
American Cent-Select	0.0098 (24)	0.0097 (24)	0.0096 (24)	-0.4662 (6)	4.0436 (3)
Brandywine	0.0245 (15)	0.0235 (15)	0.0234 (15)	-0.5856 (15)	4.4279 (9)
Davis NY Venture A	0.0425 (3)	0.0403 (3)	0.0401 (3)	-0.5110 (9)	4.6986 (15)
Fidelity Contrafund	0.0754 (1)	0.0697 (1)	0.0693 (1)	-0.5892 (16)	4.4079 (8)
Fidelity Destiny I	0.0188 (19)	0.0181 (19)	0.0180 (19)	-0.6161 (17)	4.7790 (17)
Fidelity Destiny II	0.0356 (6)	0.0335 (7)	0.0334 (7)	-0.7119 (22)	4.5791 (11)
Fidelity Growth	0.0322 (8)	0.0319 (8)	0.0317 (8)	-0.0839 (1)	4.7482 (16)
Fidelity Magellan	0.0198 (18)	0.0189 (18)	0.0188 (18)	-0.6961 (20)	5.1119 (21)
Fidelity OTC	0.0262 (12)	0.0256 (13)	0.0255 (13)	-0.3085 (4)	4.1701 (5)
Fidelity Ret. Growth	0.0261 (13)	0.0257 (12)	0.0255 (12)	-0.1506 (2)	5.7421 (22)
Fidelity Trend	0.0166 (22)	0.0157 (22)	0.0156 (22)	-0.9246 (25)	6.7634 (24)
Fidelity Value	0.0394 (4)	0.0367 (4)	0.0365 (4)	-0.5783 (14)	7.4093 (25)
Janus	0.0182 (20)	0.0175 (21)	0.0174 (21)	-0.6700 (19)	4.7802 (18)
Janus Twenty	0.0354 (7)	0.0341 (6)	0.0339 (6)	-0.4069 (5)	4.2764 (7)
Legg Mason Value Prim	0.0251 (14)	0.0242 (14)	0.0241 (14)	-0.5354 (12)	4.2254 (6)
Neuberger & Ber Part	0.0241 (16)	0.0226 (16)	0.0225 (16)	-0.8383 (24)	6.1787 (23)
New Economy	0.0292 (10)	0.0279 (10)	0.0278 (10)	-0.5713 (13)	4.0798 (4)
Nicholas	0.0283 (11)	0.0271 (11)	0.0270 (11)	-0.5118 (10)	4.9975 (20)
Prudential Equity B	0.0204 (17)	0.0195 (17)	0.0194 (17)	-0.7118 (21)	4.5924 (12)
T. Rowe Price Growth	0.0321 (9)	0.0304 (9)	0.0303 (9)	-0.6629 (18)	4.6323 (13)
Van Kampen Pace	0.0144 (23)	0.0140 (23)	0.0139 (23)	-0.4780 (8)	4.6691 (14)
Vanguard U.S. Growth	0.0062 (25)	0.0060 (25)	0.0060 (25)	-0.7414 (23)	4.9873 (19)
Vanguard/Primecap	0.0548 (2)	0.0518 (2)	0.0515 (2)	-0.5126 (11)	4.0078 (2)

Note: This table reports the estimates of the performance measures for different mutual funds based on monthly excess returns from 1991 until 2010. The third and fourth column provide the EPM based on a non-parametric and a parametric estimator, respectively. The second column reports two times the squared Sharpe ratio. The fifth and sixth column show the skewness and kurtosis, respectively.

excess returns, respectively.[13] The rankings generated by the respective performance measure are given in parentheses. The table also reports the sample skewness and kurtosis of the funds along with the corresponding rankings.[14] Sample mean excess returns are strictly positive, as can be inferred from the EPM estimates. If average excess returns were negative for some investment funds, this could be dealt with by setting the Sharpe ratio and the EPM equal to

[13]Note that the moment based parametric estimators of the EPM are nearly identical to the maximum likelihood based estimates.

[14]We intuitively consider low kurtosis to be preferable to high kurtosis and high skewness to be preferable to low skewness. It should be kept in mind, however, that one can find examples where not *all* risk averters agree in this regard (see Brockett and Kahane, 1992).

Table 3.2: Rank correlation (mutual funds)

	Sharpe	EPMNon	EPMNIG
Sharpe	1.0000	0.9800	0.9800
EPMNon	0.9800	1.0000	1.0000
EPMNIG	0.9800	1.0000	1.0000

*Note:*This table presents the rank correlation (Kendall's τ) for the rankings of Table 3.1 based on the Sharpe ratio, the non-parametric estimator of the EPM and the parametric estimator of the EPM assuming NIG-distributed returns.

zero in those cases.

The results show that the rankings induced by the two estimators for the EPM are identical. This is also supported by their rank correlation coefficient (Kendall's τ) which equals 1.0 (cf. Table 3.2).[15] Moreover, the rank correlation between the Sharpe ratio ranking and either of the EPM rankings is close to one, i.e. it is 0.98. At a first glance this might be a surprising result. However, the observed behavior is in line with the generalized continuity property. In particular, the deviations of the mutual funds return distributions from the normal distribution are not substantial, so that the differences between the rankings induced by the Sharpe ratio and the EPM are negligible.

Furthermore, note that the numerical values of two times the squared Sharpe ratio (reported in the second column) are very close to the value of the EPM for this data set indicating the adequacy of the Sharpe ratio as a performance measure for the mutual funds data set. However, this is not always the case as is demonstrated in the next example.

In our second application we estimate the performance measures for different hedge funds over the same period, ranging from January 1991 until September 2010 resulting in 237 observations. In contrast to mutual funds, hedge funds are unconstrained from dynamic and derivative trading strategies. Consequently, the distribution of the excess returns of hedge funds investments can be expected to be different from that of mutual funds, e.g. we expect the corresponding *risk* of hedge investments to differ significantly from the *risk* of mutual funds.

The data set consists of monthly excess returns of all hedge funds and the excess returns are computed from monthly fund returns and the one-month US Treasury bill rate. Note that,

[15] Kendall's τ equals 1 if two rankings perfectly agree, 0 if they are independent, and -1 if they perfectly disagree.

Table 3.3: Performance measures for monthly excess returns of hedge funds (1991-2010).

name	$2SR^2$	EPMNon	EPMNIG	skewness	kurtosis
Aetos Corporation	0.1460 (20)	0.1523 (17)	0.1514 (17)	0.7750 (3)	8.0579 (20)
Aurora Limited Partnership	0.2309 (9)	0.1633 (13)	0.1571 (14)	-1.5496 (25)	9.2812 (21)
Corsair Capital Partners LP	0.1809 (16)	0.1606 (14)	0.1590 (13)	-0.5449 (18)	4.3385 (4)
EACM Multi-Strategy Composite	0.1133 (22)	0.0924 (24)	0.0908 (24)	-1.2210 (22)	7.8312 (19)
Equity Income Partners LP	2.5312 (1)	2.9873 (1)	3.1078 (1)	2.0812 (1)	12.0353 (24)
Gabelli Associates Limited	0.3451 (5)	0.2877 (5)	0.2700 (6)	-0.0616 (10)	7.5461 (17)
GAM Diversity Inc. USD Open	0.1008 (24)	0.1010 (22)	0.1005 (22)	0.2681 (8)	5.6712 (11)
GAM Trading USD	0.2591 (7)	0.2856 (6)	0.2876 (5)	0.5736 (4)	4.3254 (3)
Genesee Balanced Fund Ltd	0.1003 (25)	0.0917 (25)	0.0908 (25)	-0.4793 (17)	5.3130 (6)
High Sierra Partners I	0.3746 (4)	0.3518 (4)	0.3413 (4)	0.4531 (6)	6.9303 (15)
Hudson Valley Partners LP	0.1998 (13)	0.1563 (16)	0.1452 (19)	-0.8134 (20)	10.0152 (23)
KDC Merger Arbitrage Fund LP	0.1972 (15)	0.1506 (19)	0.1476 (18)	-1.1401 (21)	7.6845 (18)
Kingdon Associates	0.2447 (8)	0.2331 (9)	0.2319 (8)	-0.1825 (12)	3.3525 (2)
KS Capital Partners, L.P.	0.2846 (6)	0.2505 (7)	0.2454 (7)	-0.1468 (11)	5.6205 (9)
Libra Fund LP	0.2259 (11)	0.2341 (8)	0.2263 (10)	0.5567 (5)	6.6807 (14)
Millburn MCO Partners LP	0.1098 (23)	0.1001 (23)	0.0993 (23)	-0.4155 (14)	5.4564 (8)
Millennium International Ltd	0.9931 (3)	0.6831 (3)	0.6092 (3)	-0.4611 (15)	5.7819 (12)
Millennium USA LP Fund	1.0139 (2)	0.6964 (2)	0.6259 (2)	-0.4615 (16)	5.6565 (10)
M. Kingdon Offshore Ltd.	0.2261 (10)	0.2138 (11)	0.2129 (11)	-0.2365 (13)	3.3186 (1)
Pan Multi Strategy, LP	0.1993 (14)	0.1690 (12)	0.1669 (12)	-0.7102 (19)	5.0048 (5)
P.A.W. Partners LP	0.2006 (12)	0.2279 (10)	0.2291 (9)	0.8180 (2)	5.3319 (7)
Sandler Associates	0.1553 (19)	0.1517 (18)	0.1519 (16)	0.2948 (7)	6.1078 (13)
SC Fundamental Value Fund	0.1608 (18)	0.1140 (20)	0.1084 (20)	-1.2823 (24)	14.2166 (25)
Summit Private Investments I	0.1356 (21)	0.1081 (21)	0.1064 (21)	-1.2787 (23)	7.5011 (16)
Triumph Master Fund Diversified	0.1775 (17)	0.1598 (15)	0.1557 (15)	0.2651 (9)	9.6181 (22)

Note: The table reports the same estimates as in Table 3.1 for monthly returns ranging from 1991 until September 2010 for selected hedge funds.

hedge funds are not committed to report the return of their funds which reduces the number of available hedge funds to 88. In accordance to the considered number of mutual funds and to save space, we pick from these hedge funds those 25 that have the largest Sharpe ratio.[16]

The estimation results and the implied rankings are reported in Table 3.3. The rankings induced by the two estimators for the EPM are very similar and in most cases identical. This is also supported by the rank correlation coefficient which equals 0.9467 (cf. Table 3.4). The rank correlation between the Sharpe ratio ranking and either of the EPM rankings is considerably

[16]We believe that this provides us with a selection of hedge funds that are most important to investors. This importance could also be measured by the overall assets managed, however, due to the limited regulatory requirements the volume of the hedge fund is often unavailable.

Table 3.4: Rank correlation (hedge funds)

	Sharpe	EPMNon	EPMNIG
Sharpe	1.0000	0.8600	0.8333
EPMNon	0.8600	1.0000	0.9467
EPMNIG	0.8333	0.9467	1.0000

Note: This table presents the rank correlation (Kendall's τ) for the rankings of Table 3.3 based on the Sharpe ratio, the non-parametric estimator of the EPM and the parametric estimator of the EPM assuming NIG-distributed returns.

lower. In particular, hedge funds with a larger than average kurtosis (and/or smaller than average skewness) are penalized more by our performance measure. Since the Sharpe ratio neglects skewness and kurtosis, these measures can serve to explain the difference between the Sharpe ratio ranking and the EPM rankings. In general a fund's EPM ranking deteriorates relative to the Sharpe ratio ranking if the fund's skewness is low and/ or its kurtosis is high, and vice versa. For example, the *Aurora Limited Partnership* fund, which exhibits a large kurtosis and small skewness, now achieves only rank 13 (under the non-parametric estimator and 14 based on the NIG estimator) while under the Sharpe ratio it has been ranked number 9. The *Aetos Corporation* fund instead moves from the Sharpe ratio rank 20 to rank 17 according to our measure, which may be due to the small kurtosis and positive skewness.

The second example highlights a situation where the application of the EPM is preferable as it accounts for empirically important properties of the excess return distribution that go beyond the mean and the variance. The previous example, on the other hand, illustrates that the ranking of the Sharpe ratio is maintained if the distributions are close to normal. A question, that arises, is which of the two examples is more common in practice. Addressing this question, however, goes beyond the scope of this chapter.

3.5 Conclusion

In this chapter we propose a new performance measure (EPM) that generalizes the Sharpe ratio. Instead of standard deviation the EPM employs the Aumann and Serrano (2008) index of riskiness as risk measure. In contrast to the Sharpe ratio, the EPM respects stochastic dominance and accounts for skewness, kurtosis and higher order moments in the return distribution.

If returns are normally distributed, the EPM and the Sharpe ratio induce equivalent rankings. If the distribution of the returns converges to the normal distribution, the EPM converges to two times the squared Sharpe ratio. In this sense, the EPM is asymptotically equivalent to the Sharpe ratio.

We calculate the EPM for returns that follow a normal inverse Gaussian (NIG) distribution, a distribution that is well suited to model financial returns. A representation of the EPM for NIG-distributed returns in terms of the first four moments makes explicit how skewness and kurtosis enter the performance measure. The NIG-distribution, furthermore, provides a parametric way to estimate the EPM, which is virtually as easy as estimating the Sharpe ratio with empirical moments. For non-parametric estimation of the EPM we consider the method of moments with the defining equation of the Aumann-Serrano index as moment equation.

In our empirical illustration we rank mutual funds and hedge funds with the Sharpe ratio and the EPM. The results show that the EPM penalizes investments with significant excess kurtosis (or negative skewness), while the Sharpe ratio neglects these features. Whether EPM- and Sharpe ratio-rankings differ depends on the data set under consideration. In any case, for investors that care about higher moments, the EPM provides an economically justified way to take these moments into account.

Appendix to Chapter 3

Proof of Proposition 3.3

Let r_0 and $\{r_n\}_{n\geq 1}$ be random variables with moment generating function $M_n(t) = \mathbb{E}\left(\mathrm{e}^{tr_n}\right)$ $(n \geq 0)$. Assume that

- the economic index of riskiness $AS(r_n)$ exists for all $n \geq 0$,
- there exists a real number $b > 1/AS(r_0) > 0$ for which $sup_n M_n(-b) < \infty$, and
- $r_n \xrightarrow{d} r_0$

We want to show that $AS(r_n) \to AS(r_0)$.

Proof: From $r_n \xrightarrow{d} r_0$ and the continuous mapping theorem it follows that $\mathrm{e}^{-tr_n} \xrightarrow{d} \mathrm{e}^{-tr_0}$ for all $t \in [0, b]$. Moreover, for $t \in (0, b)$ there exists $\epsilon > 0$ such that $(1+\epsilon)t = b$. Therefore,

$$\begin{aligned} & \sup_n \mathbb{E}\left([\mathrm{e}^{-tr_n}]^{1+\epsilon}\right) = \sup_n \mathbb{E}\left(\mathrm{e}^{-br_n}\right) = \sup_n M_n(-b) < \infty \\ \Rightarrow \quad & \{\mathrm{e}^{-tr_n}\} \text{ is uniformly integrable} \\ \Rightarrow \quad & \mathbb{E}\left(\mathrm{e}^{-tr_n}\right) \to \mathbb{E}\left(\mathrm{e}^{-tr_0}\right) \qquad \text{(cf. Theorem 5.4 in Billingsley (1969).} \end{aligned}$$

So far we have shown that $M_n(-t) \to M_0(-t)$ pointwise in $[0, b)$. To see that this implies $AS(r_n) \to AS(r_0)$, define $L_n(t) \equiv M_n(-t)$ and $\alpha_n \equiv 1/AS(r_n)$ for $n \geq 0$. It suffices to show that $\alpha_n \to \alpha_0$. Observe that the following holds for all $n \geq 0$:

- $L_n(0) = 1$
- $L_n'(0) = -\mathbb{E}\left(r_n\right) < 0$ (where $L_n'(0)$ is the derivative of L_n evaluated at 0)
- $L_n''(t) = \mathbb{E}\left(r_n^2\, \mathrm{e}^{-tr_n}\right) \geq 0$

The second point is true, since $\mathbb{E}\left(r_n\right) > 0$ is necessary for the existence of the AS index. So for each n, $L_n(\cdot)$ is convex. It takes on the value 1 at 0 and then initially decreases. Since

$L_n(\alpha_n) = 1$, L_n must be strictly increasing in a neighborhood of α_n. Now let $\epsilon > 0$ arbitrary. We can assume that $[\alpha_0 - \epsilon, \alpha_0 + \epsilon] \subseteq (0, b)$. Furthermore,

$$L_0(\alpha_0) = 1 \qquad \Rightarrow \qquad L_0(\alpha_0 - \epsilon) < 1 \text{ and } L_0(\alpha_0 + \epsilon) > 1$$

Because of (pointwise) convergence, we can find a $N(\epsilon) \in \mathbb{N}_0$ such that

$$L_n(\alpha_0 - \epsilon) < 1 \text{ and } L_n(\alpha_0 + \epsilon) > 1 \quad \text{for all } n \geq N(\epsilon).$$

But, by the properties of the functions L_n, this implies

$$\alpha_n \in (\alpha_0 - \epsilon, \alpha_0 + \epsilon) \quad \text{for all } n \geq N(\epsilon)$$

□

Derivation of the AS index for NIG-distributed random variables

Let X be a random variable following the normal inverse Gaussian (NIG) distribution with parameters α, β, ν and δ where $0 \leq |\beta| < \alpha$, $\delta > 0$ and $\nu \in \mathbb{R}$. The unique Aumann–Serrano risk index $s > 0$ (if it exists) is implicitly defined as

$$\mathbb{E}\left(\mathrm{e}^{-X/s}\right) = 1. \tag{3.14}$$

Let M_X denote the moment generating function (mgf) of X given by (3.4). Then (3.14) can be equivalently written as

$$M_X(t_s) = 1 \tag{3.15}$$

with $t_s = -1/s$. Solving for t_s we obtain

$$\begin{aligned}
\mathrm{e}^{\nu t+\delta\left(\gamma-\sqrt{\alpha^2-(\beta+t)^2}\right)} &= 1 \\
\Leftrightarrow \nu t + \delta\left(\gamma - \sqrt{\alpha^2 - (\beta+t)^2}\right) &= 0 \\
\Leftrightarrow \nu t + \delta\gamma &= \delta\sqrt{\alpha^2 - (\beta+t)^2} \\
\Rightarrow \nu^2 t^2 + 2\delta\gamma\nu t + \delta^2\gamma^2 &= \delta^2\alpha^2 - \delta^2(\beta^2 + 2\beta t + t^2) \\
\Leftrightarrow (\nu^2 + \delta^2)t^2 + 2(\delta\gamma\nu + \beta\delta^2)t &= 0 \\
\Leftrightarrow t = 0 \vee t &= -2\frac{\beta\delta^2 + \gamma\delta\nu}{\delta^2 + \nu^2} =: t_s.
\end{aligned}$$

For t_s to be a valid solution, we have to check whether (a) t_s is in the domain of M_X, i.e. $t_s \in [-(\alpha+\beta),\, \alpha - \beta]$, (b) t_s indeed solves (3.15) and (c) $t_s < 0$.

As to (a):

$$\begin{aligned}
t_s &\leq \alpha - \beta \\
\Leftrightarrow -2\delta\nu\gamma - 2\delta^2\beta &\leq (\alpha-\beta)(\nu^2 + \delta^2) \\
\Leftrightarrow -2\delta\nu\gamma - \delta^2\beta &\leq (\alpha-\beta)\nu^2 + \alpha\delta^2 \\
\Leftrightarrow (\alpha-\beta)\nu^2 + 2\delta\nu\gamma + (\alpha+\beta)\delta^2 &\geq 0.
\end{aligned}$$

Since $\gamma = \sqrt{\alpha^2 - \beta^2} = \sqrt{(\alpha-\beta)(\alpha+\beta)}$ the last expression is equivalent to

$$(\sqrt{(\alpha-\beta)}\nu + \sqrt{(\alpha+\beta)}\delta)^2 \geq 0.$$

In a similar way one can show that $t_s \geq -(\alpha+\beta)$.

As to (b): Plugging t_s into (3.15) yields (after some manipulations)

$$M_X(t_s) = \mathrm{e}^{\frac{\delta}{\nu^2+\delta^2}(\Psi - \sqrt{\Psi^2})}$$

with $\Psi = \delta^2\gamma - 2\nu\delta\beta - \nu^2\gamma$. Thus, $M_X(t_s) = 1$ iff $\Psi \geq 0$. Ψ can be considered as a downward open second order polynomial in ν with roots $(\delta/\gamma)(-\beta \pm \alpha)$. Therefore, $\Psi \geq 0$ iff $\nu \leq (\delta/\gamma)(\alpha - \beta)$ and $\nu \geq (\delta/\gamma)(-\alpha - \beta)$. Since $\mathbb{E}(X) = \nu + \delta\beta/\gamma$, the latter inequality is fulfilled if X has a positive mean.

As to (c): $t_s < 0$ is equivalent to $\nu > -\delta\beta/\gamma$ (which corresponds to the requirement $\mathbb{E}(X) > 0$).

Summing up, if X is a NIG-distributed random variable with parameters α, β, δ and ν where $0 \leq |\beta| < \alpha$, $\delta > 0$ and additionally $\nu \in \left(-\frac{\delta}{\gamma}\beta, \frac{\delta}{\gamma}(\alpha - \beta)\right]$, then the unique strictly positive solution of (3.14) is given by

$$s^* = \frac{1}{2}\frac{\delta^2 + \nu^2}{\beta\delta^2 + \gamma\delta\nu}.$$

Moment based representations of the AS index

Moments of a NIG-distributed random variable

The following moments of a NIG-distributed random variable can be derived using the moment generating function in (3.4):

$$\text{mean } \mu = \nu + \frac{\delta\beta}{\gamma}; \tag{3.16}$$

$$\text{variance } \sigma^2 = \frac{\delta\alpha^2}{\gamma^3}; \tag{3.17}$$

$$\text{third standardized moment (skewness) } \chi = 3\frac{\beta}{\alpha\sqrt{\delta\gamma}}; \tag{3.18}$$

$$\text{and the fourth standardized moment (excess kurtosis) } \kappa = 3\frac{\alpha^2 + 4\beta^2}{\delta\alpha^2\gamma}. \tag{3.19}$$

The inequality $|\chi| < \sqrt{3\kappa/5}$

To see why $|\chi| < \sqrt{3\kappa/5}$ for χ and κ defined in equations (3.18) and (3.19) note that

$$\frac{\chi^2}{\kappa} = 3\frac{\beta^2}{\alpha^2 + 4\beta^2} = 3\frac{(\beta/\alpha)^2}{1 + 4(\beta/\alpha)^2}.$$

Moreover, as $\alpha > |\beta|$ it holds that $(\beta/\alpha) < 1$ and therefore

$$\frac{\chi^2}{\kappa} < \frac{3}{5} \quad \text{and} \quad |\chi| < \sqrt{3\kappa/5}.$$

The AS index in terms of the first four moments of a NIG-distributed random variable

First note that the system of equations (3.16-3.19) can (uniquely) be solved for α, β, δ and ν (given that $\chi^2 < (3/5)\kappa$):

$$\alpha = 3\frac{\sqrt{3\kappa - 4\chi^2}}{(3\kappa - 5\chi^2)\,\sigma}; \quad \beta = 3\frac{\chi}{(3\kappa - 5\chi^2)\,\sigma}; \quad \nu = \mu - 3\frac{\chi\sigma}{3\kappa - 4\chi^2}$$

$$\text{and} \quad \delta = \frac{3\sqrt{(3\kappa - 5\chi^2)\,\sigma^2}}{3\kappa - 4\chi^2}.$$

This can be used to rewrite the index of riskiness

$$s^* = \frac{1}{2}\frac{\delta^2 + \nu^2}{\beta\delta^2 + \gamma\delta\nu} = \frac{1}{18}\left(3\kappa\mu - 4\mu\chi^2 - 6\chi\sigma + 9\frac{\sigma^2}{\mu}\right)$$

and the condition for the existence

$$\nu \leq \frac{\delta}{\gamma}(\alpha - \beta) \Leftrightarrow \mu \leq 3\sqrt{\frac{\sigma^2}{(3\kappa - 4\chi^2)}}$$

in terms of these moments.

Chapter 4

A directed test of error cross-section independence in fixed effect panel data models

4.1 Introduction

Cross-section dependence between disturbances in panel models can arise for various reasons: unobserved global shocks, be they economic, political or technological in nature, or unobserved shocks affecting only a subset of cross-sectional units to name a few. This dependence can have dramatic effects on the asymptotic properties of the least squares estimator and standard inference procedures (cf. Andrews, 2005).

One way to deal with potential cross-section error dependence is to use robust variance matrix estimation techniques, for instance as proposed by Driscoll and Kraay (1998). A different approach is to test for cross-section dependence and adjust estimation and inference procedures according to the test result. To this end, Breusch and Pagan (1980) proposed a Lagrange multiplier (LM) test. This test is appropriate for panels with large time dimension T and small cross-section dimension N, in which case the seemingly unrelated equation approach of Zellner (1962) is applicable. However, it is known that the LM test is severely oversized, when N is large relative to T, see for instance Pesaran, Ullah, and Yamagata (2008). While the LM test is based on squared Pearson correlation coefficients between the residuals, Frees (1995) suggests the use of Spearman rank correlation coefficients in his test called R^2_{AVE}. In Monte Carlo simulations, when the regression includes non-constant individual explanatory variables and N is large relative to T, R^2_{AVE} also exhibits size distortions. These distortions, however, are by far less pronounced as in the case of the LM test.

Pesaran (2004) suggests another variant of the LM test. Basically, the test employs residual Pearson correlation coefficients without squaring them. Pesaran (2004) provides asymptotic results for both fixed N and large T and for fixed T and large N. The test has correct empirical size in finite samples independent of the relation between N and T. However, it has low power when error correlation is present with correlation coefficients summing up to zero. This can be the case when the disturbances are generated from a factor model where the loadings average to zero. More recently, Pesaran (2012) analyzes the implicit null hypothesis of Pesaran's (2004) test, which, for large N, rather describes weak dependence than indepence of errors between cross-sections.

A direct approach to amend the size distortions of the LM test ist that of Pesaran et al. (2008). These authors compute the expected values and variances of the empirical squared Pearson correlation coefficients under the assumption of normally distributed disturbances. The resulting bias-adjusted LM test (LM_{adj}) is asymptotically normal as $T \to \infty$ first, followed by $N \to \infty$. The empirical size of the test is correct also for large N and small T.[1] The normality assumption does not appear to be restrictive, as simulation experiments with χ^2- and t-distributed errors document (cf. Pesaran et al., 2008, and Baltagi et al., 2011).

The above procedures test for error cross-section correlation in itself. The idea behind our *directed* test is to test for cross-section correlation that impacts the variance matrix estimate for the slope parameters. The procedure determines whether the simple OLS variance matrix estimate differs from a simple robust estimate. In this sense it is akin to the information matrix test of White (1982). The procedure basically augments the correlation coefficients in Pesaran's (2004) statistic with estimates of the cross-section covariance between explanatory variables. This, of course, restricts the set of alternatives against which the test has power to those where

[1] Nonnegligible size distortions can occur, however, when N is very large relative to T (cf. Baltagi et al., 2011)

the regressor covariances differ from zero. However, from the standpoint of variance matrix estimation and standard inference procedures, these are the relevant cases. In these cases the directed test has considerable power, also when residual correlation coefficients approximately sum up to zero. When the correlation coefficients all have the same sign, the directed test inherits the good power properties of Pesaran's (2004) test. A χ^2 limit distribution of the test is obtained for fixed N and $T \to \infty$. The directed test has correct empirical size independent of the size of N relative to T and without presuming a specific distribution for the disturbances.

The remainder of this chapter is structured as follows. In Section 4.2 we introduce the directed test and derive its asymptotic distribution. Before proceeding to extensive Monte Carlo experiments in Section 4.4, we present the existing tests described above in Section 4.3. Concluding remarks are made in Section 4.5. Mathematical details are given in the appendix.

4.2 The directed test

We start from the standard fixed effect panel data model

$$y_i = \iota\alpha_i + X_i\beta + u_i, \quad i = 1, \ldots, N; \tag{4.1}$$

where y_i is a $T \times 1$ vector containing the observations for cross-section i, ι is the $T \times 1$ vector of ones, α_i is a scalar, X_i is a $T \times k$ regressor-matrix with rows x'_{it}, β is a $k \times 1$ vector, and u_i is a $T \times 1$ disturbance vector. Letting $\bar{y}_i$ and $\bar{X}_i$ denote the variables after within-group demeaning and stacking equations, yields

$$\bar{y} = \bar{X}\beta + \bar{u},$$

with $\bar{y} = (\bar{y}'_1, \ldots, \bar{y}'_N)'$ and otherwise obvious notation. In the literature on tests of cross-section independence it is commonly assumed that the errors are independent across time: $\{u_t\}_{t=1}^T$ is an independent sequence, where $u_t = (u_{1t}, \ldots, u_{NT})'$ gathers individual disturbances at period t. So the covariance between u_{it} and u_{js} is given by $\text{cov}(u_{it}, u_{js}) = 0$, for $s \neq t$, and $\text{cov}(u_{it}, u_{js}) = \sigma_{ij}$ if $s = t$. More compactly, we write the variance of $u = (u'_1, \ldots, u'_N)'$ as $\text{var}(u) = \Omega \otimes I_T$, with $\otimes$ denoting the Kronecker-product and $\Omega = [\sigma_{ij}]_{1 \leq i,j \leq N}$. So the variance of the OLS estimator of β is given by

$$\mathrm{var}(\hat{\beta}) = (\bar{X}'\bar{X})^{-1}\bar{X}'(\Omega \otimes\ I_T)\bar{X}(\bar{X}'\bar{X})^{-1}.$$

Under the hypothesis of error cross-section independence, $\sigma_{ij} = 0$ for all $i \neq j$ and Ω can be replaced by

$$\Omega_0 = \begin{pmatrix} \sigma_{11} & 0 & \dots & \dots & 0 \\ 0 & \sigma_{22} & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & \sigma_{NN} \end{pmatrix}.$$

The basic strategy behind the directed test of cross-section independence is to check whether the following equation is significantly violated:

$$\bar{X}'(\Omega \otimes\ I_T)\bar{X} = \bar{X}'(\Omega_0 \otimes\ I_T)\bar{X}.$$

This can be rewritten as

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{ij}\bar{X}_i'\bar{X}_j = \sum_{i=1}^{N}\sigma_{ii}\bar{X}_i'\bar{X}_i$$

$$\text{or} \qquad \sum_{j<i}\sigma_{ij}(\bar{X}_i'\bar{X}_j + \bar{X}_j'\bar{X}_i) = 0.$$

Since $\bar{X}_i'\bar{X}_j + \bar{X}_j'\bar{X}_i$ is symmetric, we can restrict our attention to its lower triangular elements. Let $v_{ij} = \mathrm{vech}\left(\bar{X}_i'\bar{X}_j + \bar{X}_j'\bar{X}_i\right)$ denote the $\frac{1}{2}k(k+1)$-dimensional vector containing these elements. The key component of our test statistic is given by

$$S = \sum_{j<i}\hat{\sigma}_{ij}v_{ij} \tag{4.2}$$

where $\hat{\sigma}_{ij}$ is a consistent estimator for σ_{ij}. It can be obtained from the residuals of the regressions in (4.1) for cross-sections i and j: $\hat{\sigma}_{ij} = (T-k)^{-1}\hat{u}_i'\hat{u}_j$. It will be useful to write $\hat{u}_i$ as a linear function of u_i:

$$\hat{u}_i = M_i u_i \quad \text{with } M_i = I_T - Z_i(Z_i'Z_i)^{-1}Z_i' \quad \text{and } Z_i = [\iota; X_i].$$

The statistic S yet has to be normalized by its variance. Let $u = (u_1', \dots, u_N')'$ and $X = (X_1', \dots, X_N')'$. We make the following assumptions:

Assumption 4.1. *For each i, the disturbances, u_{it}, are serially independent with zero mean and positive variances $\sigma_i^2 = \sigma_{ii} \leq M < \infty$.*

Assumption 4.2. *The regressors X and the errors u are independent.*

Assumption 4.3. *The null hypothesis is specified as*

$$\mathbf{H_0} : u_{it} = \sigma_i \epsilon_{it}, \text{ with } \epsilon_{it} \sim iid(0,1) \text{ across } i \text{ and } t. \tag{4.3}$$

The following lemma states properties of the summands $\hat{\sigma}_{ij}v_{ij}$ of S in Equation (4.2) and provides the resulting variance of S. We use $tr(\cdot)$ to denote the trace of a square matrix.

Lemma 4.4. *Under Assumptions 4.1 - 4.3 the following holds:*

(i) $\mathbb{E}[\hat{\sigma}_{ij}v_{ij}] = 0,$ *for* $i \neq j$,

(ii) $\mathbb{E}[\hat{\sigma}_{ij}v_{ij}\hat{\sigma}_{ik}v_{ik}] = 0,$ *given that the indices i, j, and k differ,*

(iii) $var(\hat{\sigma}_{ij}v_{ij}) = (T-k)^{-2}\sigma_{ii}\sigma_{jj}\mathbb{E}\left[tr(M_iM_j)v_{ij}v_{ij}'\right]$, *for* $i \neq j$, *and*

(iv) $var(S) = (T-k)^{-2}\sum_{j<i}^{N}\sigma_{ii}\sigma_{jj}\mathbb{E}\left[tr(M_iM_j)v_{ij}v_{ij}'\right]$.

Proof. See appendix. □

A natural estimator for the variance of S is given by

$$\hat{V} = (T-k)^{-2}\sum_{j<i}^{N}\hat{\sigma}_{jj}\hat{\sigma}_{ii}\text{tr}(M_jM_i)v_{ij}v_{ij}',$$

where, as above, $\hat{\sigma}_{ij} = (T-k)^{-1}\hat{u}_i'\hat{u}_j$. We can assume that $\hat{V}$ is invertible, at least asymptotically. Otherwise one can reduce the dimension of v_{ij} and consider only independent restrictions.

Standardizing S with its estimated variance leads to the following statistic for the *directed* test of cross-section independence:

$$CD_X^\sigma = S'\hat{V}^{-1}S.$$

The following proposition states that CD_X^σ is approximately χ^2-distributed for fixed N and large T. We use "$\xrightarrow{d}$" to denote convergence in distribution and let $X_t = (x_{1t}, \ldots, x_{Nt})'$.

Proposition 4.5. *Let Assumptions 4.1 - 4.3 and the following regularity conditions hold:*

(I) The sequence $\{X_t\}$ is stationary and ergodic,

(II) The components of X have uniformly bounded second moments,

(III) $\mathbb{E}[x_{it}x_{jt}']$ is non-singular for all $i, j \in \{1, \ldots, N\}$.

Then, for fixed N and as $T \to \infty$

$$CD_X^\sigma \xrightarrow{d} \chi^2(k(k+1)/2),$$

where k is the number of nonconstant regressors in a cross-section.

Proof. See appendix. □

A variant of the directed test is obtained by replacing the estimated covariances in S, cf. Equation (4.2), by estimated correlation coefficients:

$$R = \sum_{j<i}^{N} \hat{\rho}_{ij} v_{ij}, \quad \text{where } \hat{\rho}_{ij} = \frac{\hat{u}_i'\hat{u}_j}{(\hat{u}_i'\hat{u}_i)^{\frac{1}{2}}(\hat{u}_j'\hat{u}_j)^{\frac{1}{2}}}. \tag{4.4}$$

The resulting test, CD_X^ρ, shows a favorable performance in simulation experitments. However, an additional assumption is needed.

Assumption 4.6. *The errors ϵ_{it} in* (4.3) *are distributed symmetrically around zero.*

Under Assumptions 4.1 - 4.6 and following Pesaran (2004) it can be shown that

$$var(R) = \sum_{j<i}^{N} \mathbb{E}[\hat{\rho}_{ij}^2]\mathbb{E}[v_{ij}v_{ij}'].$$

In practice, $var(R)$ can be approximated using

$$\hat{V}_R = \frac{1}{T}\sum_{j<i}^{N} v_{ij}v_{ij}',$$

since $\mathbb{E}[\hat{\rho}^2] \approx T^{-1}$ for large enough T. The resulting statistic is denoted as[2]

$$CD_X^{\rho} = R'\hat{V}_R^{-1}R.$$

The limit distribution of CD_X^{ρ} is given in the following proposition:

Proposition 4.7. *Let Assumptions 4.1 to 4.6 and (I) - (IV) in Proposition 4.5 hold. Then, for fixed N and $T \to \infty$,*

$$CD_X^{\rho} \xrightarrow{d} \chi^2(k(k+1)/2).$$

Proof. See appendix. □

4.3 Alternative test procedures

In this section we describe existing tests that we consider for the purpose of comparison in the simulation experiments in the following section. For completeness, we start with the LM test of Breusch and Pagan (1980), although, because of its known severe size distortions when N is large relative to T (cf. Pesaran et al., 2008, or Moscone and Tosetti, 2009), it is not included in the Monte Carlo section.

LM test:

The LM test of Breusch and Pagan (1980) rests on the statistic

$$LM = \sqrt{\frac{1}{N(N-1)}}\sum_{j<i}^{N}(T\hat{\rho}_{ij}^2 - 1),$$

with $\hat{\rho}_{ij}$ defined as in (4.4). Under the null hypothesis and as $T \to \infty$ first, followed by $N \to \infty$, LM approaches a standard normal distribution. However, for small T, the term $T\hat{\rho}_{ij}^2 - 1$ is not centered at 0. For large N, this can lead to substantial overrejection.

[2]To ensure invertibility of $\hat{V}_R$ a reduction of the length of v_{ij} might be necessary.

Frees' test:

The main component of Frees' (1995) rank correlation test is given by

$$R^2_{AVE} = \frac{2}{N(N-1)} \sum_{j<i}^{N} \hat{r}_{ij}^2,$$

where $\hat{r}_{ij}$ is the Spearman rank correlation coefficient between residual $\hat{u}_i$ and $\hat{u}_j$. Frees (1995) derives the limit distribution of $Q_N = N\left(R^2_{AVE} - (T-1)^{-1}\right)$ for the case that the intercept is the only regressor in (4.1). The resulting limit Q is a weighted sum of two independent χ^2 random variables. Because of dependence on T, critical values are cumbersome to compute. When T is large, Frees (1995) suggests to use the following approximately normally distributed statistic

$$FRE = \frac{Q_N}{\sqrt{var(Q)}}, \quad \text{where } var(Q) = \frac{4(T-2)(25T^2-7T-54)}{25T(T-1)^3(T+1)}.$$

Pesaran's test:

To work around the bias problem of the LM test for finite T Pesaran (2004, 2012) suggests to use $\hat{\rho}_{ij}$ instead of $\hat{\rho}_{ij}^2$ and considers the statistic

$$CD_P = \sqrt{\frac{2T}{N(N-1)}} \sum_{j<i}^{N} \hat{\rho}_{ij}.$$

He shows that under Assumptions 4.1 - 4.6, $\mathbb{E}[\hat{\rho}_{ij}] = 0$ and CD_P approaches a standard normal distribution for $T, N \to \infty$ in any order. Indeed, his test has correct empirical size also when N is large relative to T. A standard critique of the test is that it lacks power against alernatives under which $\sum_{j<i} \hat{\rho}_{ij} \approx 0$.

The adjusted LM test:

Pesaran, Ullah, and Yamagata (2008) compute the exact finite sample expectations and variance of $\hat{\rho}_{ij}^2$ imposing that, in addition to Assumptions 4.1 - 4.6, the errors ϵ_{it} are normally distributed.

Their statistic is given by

$$LM_{adj} = \sqrt{\frac{2}{N(N-1)}} \sum_{j<i}^{N} \frac{(T-k)\hat{\rho}_{ij}^2 - \mu_{Tij}}{\nu_{Tij}},$$

where

$$\mu_{Tij} = \mathbb{E}[(T-k)\hat{\rho}_{ij}^2] = \frac{1}{T-k} tr\left[\mathbb{E}(M_i M_j)\right],$$
$$\nu_{Tij}^2 = Var[(T-k)\hat{\rho}_{ij}^2] = \{tr[\mathbb{E}(M_i M_j)]\}^2 a_{1T} + 2tr\left\{\mathbb{E}[(M_i M_j)^2]\right\} a_{2T},$$
$$a_{1T} = a_{2T} - (T-k)^{-2}, \quad a_{2T} = 3\left[\frac{(T-k-8)(T-k+2)+24}{(T-k+2)(T-k-2)(T-k-4)}\right]^2.$$

As $T \to \infty$ first, followed by $N \to \infty$, LM_{adj} converges to a standard normal distribution. The test keeps the nominal size also when N is large relative to T.

The John test:

A different approach is taken by Baltagi, Feng, and Kao (2011). Their procedure tests for spheric disturbances. That is, apart from independence, the null hypothesis includes homoskedasticity. Although this only allows for a limited comparison with tests of cross-section independence, the John test is included in the Monte Carlo experiments, since Baltagi et al. (2011) found a favorable performance relative to CD_P and LM_{adj} under homoskedasticity. The test statistic is given by

$$J = \frac{T[\frac{1}{n}tr(\hat{\Omega})]^{-2}\frac{1}{n}tr(\hat{\Omega}^2) - T - n}{2} - \frac{1}{2} - \frac{n}{2(T-1)}, \tag{4.5}$$

where $\hat{\Omega} = \frac{1}{T-k}\sum_{t=1}^{T} \hat{v}_t \hat{v}_t'$ and $\hat{v}_t = (\hat{v}_{1t}, \ldots, \hat{v}_{Nt})$ contains the residuals for period t from a fixed effect regression in model (4.1).[3] A crucial assumption of Baltagi et al. (2011) is that the errors are normally distributed. Then, under H_0, the statistic in (4.5) is asymtotically standard normal as $N, T \to \infty$ with $\frac{N}{T} \to c \in [0, \infty)$.

[3] This contrasts with the previous tests, where the residuals are obtained from separate cross-section regressions. These separate regressions ensure that, under H_0, the residuals are independent. One may conjecture that the asymptotic results for those tests are still valid when fixed effect residuals are employed.

4.4 Monte Carlo simulations

In this section, the empirical size and power of the tests for cross-section independence is assessed by means of Monte Carlo experiments. In Section 4.4.1 the simulation scenarios are described. We present the results in Section 4.4.2.

4.4.1 Simulation framework

Similar to Pesaran et al. (2008) and Moscone and Tosetti (2009), we use the following data generating process (for $i = 1, \dots, N$):

$$\begin{aligned} y_{it} &= \alpha_i + \beta_1 x_{1,it} + \beta_2 x_{2,it} + u_{it}, \quad \text{where} \\ u_{it} &= \gamma_i f_t + \sigma_i \epsilon_{it}, \quad \epsilon_{it} \sim iid\, N(0,1),\; f_t \sim iid\, N(0,1), \\ \alpha_i &\sim iid\, N(1,1), \\ \beta_1 &= \beta_2 = 1, \\ \sigma_i^2 &\sim iid\, \chi_2^2/2. \end{aligned} \tag{4.6}$$

Several scenarios for the regressors $x_{l,it}$ and the factor loadings γ_i are considered. In contrast to Pesaran et al. (2008) and Moscone and Tosetti (2009) we simulate regressors, that are correlated across cross-sections. It comes at no surprise that whithout such correlation the CD_X test has little power, since the terms v_{ij} in (4.2) basically contain empirical covariances between regressores in different sections. The point of the CD_X test, however, is to check whether standard inference procedures, such as the OLS t-test on a slope coefficient, that neglect error cross-section correlation are invalid. Unreported Monte Carlo simulations show that the t-test has correct empirical size in the absence of regressor cross-section correlation, even if error cross-section correlation is neglected.

Scenario S0 represents two instances in which the null hypothesis in (4.3) holds: one case in which the errors ϵ_{it} follow a normal distribution as in (4.6) and another case with χ^2-errors. In the second case we simulate the errors as $\epsilon_{it} \sim iid(\chi_1^2 - 1)/\sqrt{2}$. Note that χ^2-errors neither fulfill the symmetry assumption of Pesaran (2004) nor the normality assumption of Pesaran et al. (2008) or Baltagi et al. (2011). Parallel to the errors, the regressors exhibit a factor structure:

S0: $\gamma_i = 0$ for all $i = 1, \ldots, N$, and the regressors are generated as $x_{l,it} = f_{l,t}^{(x)} \gamma_{l,i} + \epsilon_{l,it}$, where $f_{l,t}^{(x)} \sim iid\, N(0,1)$, $\gamma_{l,i} \sim iid\, U(0.1, 0.3)$, and $\epsilon_{l,it} \sim iid\, N(0, 0.1)$.

In the remaining scenarios we investigate the empirical power of the tests. While in scenario S1 all covariances σ_{ij} are positive, they will approximately net out in sceneario S2, such that $\sum_{j<i}^{N} \rho_{ij} \approx 0$. It has been argued (cf. Pesaran et al., 2008) that the CD_P test lacks power in the latter case.

S1: $\gamma_i \sim iid\, U(0.1, 0.3)$, with $U(0.1, 0.3)$ denoting the uniform distribution on the intervall $[0.1, 0.3]$. The regressors are generated as in S0.

In scenario S2 we simulate factor loadings that approximately sum up to zero. This entails that cross-section error correlations approximately cancel out. Moreover, as in Moscone and Tosetti (2009), we consider regressors containing an AR(1)-component:

S2: $\gamma_i \sim iid\, U(-0.4, -0.2)$, for $i = 1, \ldots, \frac{N}{2}$ and $\gamma_i \sim iid\, U(0.2, 0.4)$, for $i = \frac{N}{2} + 1, \ldots, N$. The regressors are generated as $x_{l,it} = z_{l,it} + f_t^{(z)} \gamma_{l,i}$, with $f_{l,t}^{(z)} \sim iid\, N(0,1)$ and where $\gamma_{l,i}$ is generated in the same way as γ_i. $z_{l,it}$ follows an AR(1) process of the form $z_{l,it} = 0.6 z_{l,it-1} + v_{l,it}$, for $t = -49, \ldots, T$, where $z_{l,i-50} = 0$, $v_{l,it} \sim iid\, N(0, \tau_{li}^2(1 - 0.6^2))$, and $\tau_{li}^2 \sim iid\, \chi_6^2/6$. The observations for $t = -49, \ldots, 0$ are discarded to dissipate possible influence of initial values.

Scenario S3 is a case of weak cross-section or spatial dependence. We do not assume any specific metric between the cross-sections, such as geographical distance, that could be exploited.

S3: The disturbances $u_t = (u_{1t}, \ldots, u_{Nt})'$ are generated as $u_t = (I_N + 0.5A)\epsilon_t$, where I_N is the identity matrix, A is a $(N \times N)$ random matrix with elements $a_{ij} \sim iid\, N(0,1)$, and $\epsilon_t \sim iid\, N(0, I_N)$. The regressors are generated as $x_{l,t} = z_{l,t} + \left(I_N + 0.5A^{(x)}\right) \epsilon_t^{(x)}$, with $z_{l,t}$ as in S2 and $A^{(x)}$ and $\epsilon_t^{(x)}$ analogous to (but independent from) A and ϵ_t.

Finally, in scenario S4, we consider the case where disturbances are homoskedastic across the whole panel. This allows for a comparison with the *John test* of spherical distrubances proposed by Baltagi et al. (2011).

S4: As S1, except that σ_i in (4.6) is set to 1 for all i.

For each scenario and varying numbers of cross-sections N and time periods T we employ $2,000$ Monte Carlo replications. All tests are conducted at the 5% nominal level.

4.4.2 Simulation results

The empirical rejection frequencies in scenario S0 are displayed in Tables 4.1 (normal disturbances) and 4.2 (χ^2 disturbances). All numbers are given in percentage points. When errors are normal, the test are, in general, correctly sized. However, a few exceptions exist. The rank-correlation test of Frees (1995), FRE, tends to be oversized as N grows relative to T.

Table 4.1: Scenario S0 - Empirical size, normally distributed disturbances

(T, N)	**10**	**20**	**30**	**50**	**100**	**10**	**20**	**30**	**50**	**100**
	CD_P					CD_X^σ				
10	5.85	5.40	5.90	5.00	4.90	5.40	5.10	5.30	5.85	5.90
20	5.10	5.40	4.70	5.95	4.75	5.45	4.75	5.10	5.60	5.00
30	4.35	5.70	5.45	5.85	5.45	5.05	4.65	5.25	5.95	4.80
50	5.40	4.50	5.50	6.35	5.35	5.75	4.75	4.60	5.80	5.25
100	3.90	4.45	5.40	5.50	5.50	5.00	5.25	5.20	4.75	5.60
	LM_{adj}					CD_X^ρ				
10	5.20	5.00	4.80	5.40	8.45	8.35	7.00	9.00	8.20	6.85
20	5.55	5.10	4.35	6.10	6.25	7.05	5.75	6.80	6.45	7.50
30	4.45	5.40	4.75	4.85	5.55	5.70	6.00	6.20	6.80	6.25
50	4.95	5.50	4.85	5.55	5.60	6.35	5.25	5.40	6.05	5.40
100	4.65	5.45	4.75	5.10	4.60	4.30	4.25	5.70	5.60	6.85
	FRE									
10	6.50	5.95	6.60	9.00	17.85					
20	6.15	6.15	5.75	7.40	7.85					
30	5.85	5.60	5.30	5.35	6.90					
50	6.35	5.70	4.80	6.05	5.85					
100	6.55	5.60	5.00	4.95	5.20					

Table 4.2: Scenario S0 - Empirical size, χ^2 distributed disturbances

(T,N)	**10**	**20**	**30**	**50**	**100**	**10**	**20**	**30**	**50**	**100**
	CD_P					CD_X^σ				
10	5.80	6.20	4.85	5.10	4.85	4.25	5.95	5.15	5.10	5.70
20	5.15	5.10	5.25	4.90	4.85	5.45	5.25	5.65	4.65	5.55
30	5.40	5.05	5.50	6.05	4.75	5.15	4.50	5.75	6.00	6.00
50	4.35	5.00	5.65	4.40	5.60	5.70	4.95	4.70	4.65	4.55
100	4.30	4.45	5.10	5.30	4.60	4.65	4.90	5.25	5.15	4.80
	LM_{adj}					CD_X^ρ				
10	6.05	5.55	6.20	6.75	8.60	8.00	8.90	8.30	8.00	7.95
20	6.25	8.20	7.05	6.40	6.75	6.75	6.25	6.40	6.20	5.85
30	7.00	9.35	7.95	8.40	7.90	5.80	5.50	7.05	6.90	6.55
50	7.40	8.00	8.30	9.15	10.15	4.95	5.55	6.20	5.10	5.55
100	8.45	7.55	7.50	7.95	9.00	4.50	4.95	5.10	5.45	4.35
	FRE									
10	6.80	5.75	5.80	6.75	11.00					
20	5.70	6.15	5.60	5.30	5.10					
30	5.60	5.80	5.55	4.85	4.70					
50	5.85	5.10	4.50	6.15	6.35					
100	5.45	4.45	5.50	5.00	4.70					

Although the empirical rejection rate for $N = 100$ and $T = 10$ is over 17%, this is dwarfed by the original LM test of Breusch and Pagan (1980), where rejection rates in such cases are close to 100% (cf. Moscone and Tosetti, 2009). Also the rejection rates of LM_{adj} have a slight tendency to increase when N grows relative to T. This becomes more apparent for $T = 10$ and $N = 200$. In this case the empirical rejection rate of LM_{adj} is 9.4% . Lastly, note that CD_X^ρ is somewhat oversized when $T = 10$; probably because $\mathbb{E}[\hat{\rho}_{ij}^2] \approx T^{-1}$ is not accurate enough in that case. However, rejection rates remain unharmed by increasing N.

The results by and large do not change when χ^2 errors are used, although CD_P and LM_{adj} assume symmetric or normal errors. However, slight size distortions of the LM_{adj} test can be observed (Table 4.2).

The empirical power of the tests under scenario S1 is reported in Table 4.3. No size adjustments are made. Pesaran's (2004, 2012) test, CD_P, shows the best performance followed by CD_X^ρ. Using estimated error covariances instead of estimated correlation coefficients unambigu-

Table 4.3: Scenario S1 - Empirical power, error cross-section correlations all positive

(T, N)	**10**	**20**	**30**	**50**	**100**	**10**	**20**	**30**	**50**	**100**
	CD_P					CD_X^{σ}				
10	21.35	53.15	74.40	94.15	99.35	9.15	21.40	35.65	48.60	79.90
20	93.35	75.20	95.05	99.45	100.00	64.25	27.25	60.30	79.30	97.30
30	93.95	73.65	98.95	100.00	100.00	48.30	39.45	64.40	97.60	100.00
50	78.25	99.85	100.00	100.00	100.00	34.30	85.20	84.35	98.55	100.00
100	89.30	100.00	100.00	100.00	100.00	48.25	95.40	100.00	100.00	100.00
	LM_{adj}					CD_X^{ρ}				
10	6.25	12.15	17.10	44.00	56.50	17.40	40.10	59.25	85.90	96.55
20	66.15	23.45	45.75	75.15	96.25	85.70	62.05	88.55	98.80	100.00
30	80.80	15.90	72.10	89.75	99.65	86.55	62.50	96.85	100.00	100.00
50	44.60	89.55	97.35	98.70	100.00	68.05	99.15	99.60	100.00	100.00
100	62.60	100.00	100.00	100.00	100.00	81.70	100.00	100.00	100.00	100.00
	FRE					OLS t-test				
10	7.75	12.50	19.10	44.90	66.55	3.20	4.90	5.20	6.60	11.60
20	62.35	21.45	43.15	73.05	95.80	5.80	5.45	6.40	9.70	12.00
30	78.45	16.25	68.30	88.30	99.70	5.80	5.80	6.50	11.30	18.55
50	43.35	86.50	95.90	98.20	100.00	5.40	7.35	6.90	9.45	15.90
100	60.00	100.00	100.00	100.00	100.00	5.40	6.65	9.85	10.80	18.15

Notes: The panel labeled "OLS t-test" reports the empirical size of the standard OLS t-test of the hypothesis $\beta_1 = 1$. The nominal test size is 5%.

ously reduces rejections rates, as the comparison between CD_X^{σ} and CD_X^{ρ} shows. Compared to LM_{adj}, CD_X^{σ} has higher empirical power for relatively small T while the converse is true for larger T. The rejection rates of LM_{adj} and FRE are comparable. The difference when $N = 100$ and $T = 10$ can be explained by the size-distortions of FRE in that case. Table 4.3 also presents rejection frequencies for the OLS t-test of the hypothesis $\beta_1 = 1$. The t-test is based on standard fixed-effect estimation under the null of no cross-section error correlation.[4] The results show that the test is oversized if the number of cross-sections is large.

In Table 4.4 the rejection frequencies for scenario S2 are displayed. In this case, where error cross-section correlations are different from zero but approximately cancel out, CD_P has virtually no power. CD_X^{ρ} outperforms CD_X^{σ}, and LM_{adj} and FRE again exhibit similar

[4]More precisely, we compute the variance of the fixed-effect estimator of the slope parameters as $\left(\bar{X}'\bar{X}\right)^{-1}\left(\sum_{i=1}^{N}\hat{\sigma}_{ii}\bar{X}_i'\bar{X}_i\right)\left(\bar{X}'\bar{X}\right)^{-1}$.

Table 4.4: Scenario S2 - Empirical power, error cross-section correlations approximatively cancel out

(T, N)	**10**	**20**	**30**	**50**	**100**	**10**	**20**	**30**	**50**	**100**
	CD_P					CD_X^σ				
10	2.95	4.20	2.25	2.00	2.50	8.80	17.60	32.50	48.60	80.45
20	5.10	4.25	3.55	3.15	3.00	16.55	32.25	50.65	89.95	99.05
30	4.40	3.60	3.55	4.15	3.40	22.05	45.85	87.50	98.10	100.00
50	4.75	4.95	5.20	6.45	10.45	59.30	64.75	98.00	100.00	100.00
100	15.90	17.40	4.45	14.05	11.25	79.85	99.15	100.00	100.00	100.00
	LM_{adj}					CD_X^ρ				
10	9.80	31.45	47.25	78.00	88.55	17.65	37.25	59.40	80.05	94.75
20	49.75	48.30	86.90	99.60	99.95	33.35	58.05	77.50	99.10	99.95
30	29.65	86.05	99.60	99.90	100.00	30.50	77.30	99.55	99.95	100.00
50	72.75	99.55	100.00	100.00	100.00	86.25	93.95	100.00	100.00	100.00
100	98.45	100.00	100.00	100.00	100.00	97.00	100.00	100.00	100.00	100.00
	FRE					OLS t-test				
10	11.25	32.40	49.00	81.10	94.15	4.30	3.10	5.15	6.40	11.20
20	46.50	45.15	85.45	99.25	99.90	5.30	5.30	5.30	8.80	11.55
30	28.15	82.60	99.25	99.90	100.00	5.60	5.20	7.00	7.55	13.35
50	69.80	99.15	100.00	100.00	100.00	4.60	5.80	7.10	10.95	14.10
100	98.40	100.00	100.00	100.00	100.00	5.50	5.85	7.35	9.20	14.60

Notes: The panel labeled "OLS t-test" reports the empirical size of the standard OLS t-test of the hypothesis $\beta_1 = 1$. The nominal test size is 5%.

rejection rates. CD_X^ρ also performs well compared to the latter two tests. The results for the OLS t-test in this case are similar to those in scenario S1 (cf. Table 4.3).

The results for scenario S3 (weak dependence) are reported in Table 4.5. In this case FRE and LM_{adj} again perform similar to each other and favorably compared to the other tests. The rejection rates of CD_X^ρ and CD_X^σ are comparable and higher than those of CD_P. Moreover, the rejection rates of these tests decrease with increasing N. However, this is no problem in terms of standard OLS t-tests, as can be seen from the bottom right panel of table 4.5. The test clearly maintains the nominal size.

Finally, Table 4.6 reports rejection rates for scenario S4. It differs from S1 only in that the errors are homoskedastic. This allows for a comparison between the test of cross-section independence and Baltagi et al.'s (2011) test for spherical disturbances. The empirical power

Table 4.5: Scenario S3 - Empirical power, weak dependence

(T,N)	**10**	**20**	**30**	**50**	**100**	**10**	**20**	**30**	**50**	**100**
	CD_P					CD_X^{σ}				
10	7.90	6.65	6.65	5.25	4.95	14.45	11.45	8.75	7.80	6.15
20	17.00	9.90	8.80	8.15	5.55	34.45	18.75	16.35	10.15	7.60
30	25.60	16.40	11.80	9.75	6.80	49.55	31.45	22.30	15.80	9.30
50	37.55	25.85	20.25	15.40	9.50	68.40	47.90	37.75	26.15	14.05
100	53.30	39.85	32.30	23.90	14.90	84.35	70.70	59.30	45.60	25.90
	LM_{adj}					CD_X^{ρ}				
10	43.70	52.20	50.30	52.40	53.05	15.65	12.65	10.15	9.55	8.25
20	99.30	99.90	99.95	99.95	100.00	34.45	18.80	16.20	10.05	8.35
30	100.00	100.00	100.00	100.00	100.00	49.05	30.90	22.70	15.35	10.40
50	100.00	100.00	100.00	100.00	100.00	68.40	47.35	36.40	25.70	14.35
100	100.00	100.00	100.00	100.00	100.00	84.15	71.60	59.75	44.55	26.05
	FRE					OLS t-test				
10	41.05	46.95	44.00	44.05	46.70	2.85	3.30	3.25	2.40	2.70
20	97.60	99.15	99.80	99.60	99.70	4.20	3.75	3.40	4.10	3.75
30	99.95	100.00	100.00	100.00	100.00	3.90	4.10	4.40	4.85	3.80
50	100.00	100.00	100.00	100.00	100.00	4.30	5.05	4.40	4.35	4.20
100	100.00	100.00	100.00	100.00	100.00	5.55	5.15	4.50	5.35	4.20

Notes: The panel labeled "OLS t-test" reports the empirical size of the standard OLS t-test of the hypothesis $\beta_1 = 1$. The nominal test size is 5%.

of this test is quite similar to that of FRE and LM_{adj}, with some advantage for the former when N is large.[5] This advantage may be explained by the fact that the $John$ test employs the standard fixed effect estimates to compute residuals, while the other test run the regressions cross-section wise. As in scenario S1, CD_P exhibits again the highest empirical power. Also CD_X^{σ} and CD_X^{ρ} perform well. In fact, CD_X^{σ} and CD_X^{ρ} display very similar rejection frequencies as opposed to scenario S1. This is no big surprise since under homoskedasticity the statistic S in (4.2) is approximately just a multiple of R in (4.4). The muliplicative factor cancels out when normalizing S and R to yield CD_X^{σ} and CD_X^{ρ}, respectively.

[5]The relatively high rejection rate of FRE when $N = 100$ and $T = 10$ is due to the fact that FRE is oversized in that case.

Table 4.6: Scenario S4 - Empirical power, homoskedastic disturbances

(T, N)	**10**	**20**	**30**	**50**	**100**	**10**	**20**	**30**	**50**	**100**
	CD_P					CD_X^{σ}				
10	16.00	28.45	42.85	62.75	87.90	10.55	20.05	28.80	47.65	76.40
20	23.20	49.45	69.90	90.65	99.25	15.10	35.10	53.35	81.65	98.05
30	27.50	64.50	84.10	98.30	100.00	18.65	49.10	71.95	93.90	100.00
40	43.60	83.10	96.55	100.00	100.00	29.65	70.90	91.15	99.55	100.00
50	63.60	97.75	99.95	100.00	100.00	49.35	93.90	99.70	100.00	100.00
	LM_{adj}					CD_X^{ρ}				
10	5.10	5.80	6.90	7.15	12.80	14.40	22.85	31.95	49.90	76.35
20	6.50	7.00	10.05	13.20	28.35	16.35	35.90	54.40	81.20	97.85
30	7.35	10.30	13.15	20.95	45.95	19.55	50.10	72.15	93.65	100.00
50	9.75	15.85	23.55	41.60	80.80	30.20	71.00	91.00	99.50	100.00
100	15.40	36.70	57.20	85.00	99.10	49.25	93.65	99.65	100.00	100.00
	FRE					$John$				
10	6.45	6.45	8.00	11.20	24.65	7.55	9.30	10.35	12.40	18.50
20	8.05	7.50	11.15	14.70	32.15	8.85	9.20	12.10	17.50	35.90
30	8.15	10.75	13.15	23.05	46.80	8.95	11.95	15.40	26.95	54.45
50	10.70	15.75	22.65	39.85	78.55	9.60	17.25	27.50	47.20	84.20
100	16.15	34.55	52.90	81.60	98.25	15.70	36.45	59.45	86.35	99.15
	OLS t-test									
10	3.45	5.35	4.95	6.20	11.10					
20	4.85	6.30	6.90	9.70	12.95					
30	4.65	7.00	7.10	9.30	17.55					
50	5.50	6.60	7.40	11.25	14.70					
100	5.90	6.60	8.15	11.45	16.80					

Notes: The panel labeled "OLS t-test" reports the empirical size of the standard OLS t-test of the hypothesis $\beta_1 = 1$. The nominal test size is 5%.

4.5 Summary

We have introduced two variants of a *directed* test of error cross-section independence. The basic idea is to test, whether error cross-section correlation can be neglegted when computing the variance matrix of the fixed effect slope estimator. Compared to existing testing procedures, this restricts the set of alternative hypotheses. However, from the standpoint of standard inference regarding the slope parameters, these alternatives are most relevant. The asymptotic distribution for $T \to \infty$ and fixed N of the *directed* tests has been shown to be χ^2 with $k(k+1)/2$

degrees of freedom, where k is the number of nonconstant regressors.

In Monte Carlo simulations, the variant CD_X^ρ of the dircted tests exhibits a favorable performance in term of empirical power under strong error cross-section correlation. It comes closest to Pesaran's (2004, 2012) CD_P test, when all correlation coefficients have the same sign. In contrast to CD_P, CD_X^ρ still has high rejection rates, when correlation coefficients approximately add up to zero. Under weak error cross-section dependence, CD_X^ρ has relatively low empirical power. However, if this led a practitioner to ignore error cross-section dependence when conducting, say, t-tests on the slope parameters, the tests would still be correctly sized.

Appendix to Chapter 4

Proof of Lemma 4.4

For notational simplicity we show the results for deterministic X. For stochastic X they can be obtained by conditioning on X.

(i):

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}_{ij} v_{ij}] &= v_{ij}\mathbb{E}[\frac{1}{T} u_i' M_i M_j u_j] \\
&= \frac{1}{T} v_{ij}\mathbb{E}[tr(u_i' M_i M_j u_j)] \\
&= \frac{1}{T} v_{ij} tr(\sigma_{ij} M_i M_j) \\
&\overset{H_0}{=} 0.
\end{aligned}$$

(ii):

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}_{ij} v_{ij} \hat{\sigma}_{ik} v_{ik}'] &= (T-k)^{-2} v_{ij} v_{ik}' \mathbb{E}[(M_i u_i)' M_j u_j (M_i u_i)' M_k u_k] \\
&= 0 \text{ since under the null } u_i,\ u_j, \text{ and } u_k \text{ are independent.}
\end{aligned}$$

(iii) Note that

$$\begin{aligned}
(T-k)^2 \text{var}(\hat{\sigma}_{ij} v_{ij}) &= v_{ij}\mathbb{E}[u_i' M_i M_j u_j u_j' M_j M_i u_i] v_{ij}' \\
&= v_{ij}\mathbb{E}[\text{tr}(u_i' M_i M_j u_j u_j' M_j M_i u_i)] v_{ij}' \\
&= v_{ij}\text{tr}(M_i M_j \mathbb{E}[u_j u_j' M_j M_i u_i u_i']) v_{ij}' \\
&= v_{ij}\text{tr}(M_i M_j \mathbb{E}\left[\mathbb{E}[u_j u_j' M_j M_i u_i u_i' | u_j]\right]) v_{ij}' \\
&= \sigma_{ii}\text{tr}(M_i M_j \mathbb{E}[u_j u_j'] M_j M_i) v_{ij} v_{ij}' \\
&= \sigma_{ii}\sigma_{jj}\text{tr}(M_i M_j M_j M_i) v_{ij} v_{ij}' \\
&= \sigma_{ii}\sigma_{jj}\text{tr}(M_i M_j) v_{ij} v_{ij}'.
\end{aligned}$$

(iv): follows from (ii) and (iii). $\square$

Proof of Proposition 4.5

The proof is conducted in three steps.

Step 1: $\sqrt{T-k}\hat{\sigma}_{ij} = \frac{1}{\sqrt{T-k}} u_i' u_j + o_p(1)$.

Proof: By definition, $\hat{\sigma}_{ij} = \frac{1}{T-k} u_i' M_i M_j u_j$. Therefore,

$$\sqrt{T-k}\hat{\sigma}_{ij} = \frac{1}{\sqrt{T-k}} u_i' M_i M_j u_j.$$

Rewrite $\frac{1}{\sqrt{T-k}} u_i' M_i M_j u_j$ as

$$\begin{aligned} \frac{1}{\sqrt{T-k}} u_i' M_i M_j u_j = & \frac{1}{\sqrt{T-k}} u_i' u_j \\ & - \frac{u_i' X_i}{\sqrt{T-k}} \left(\frac{X_i' X_i}{T-k} \right)^{-1} \frac{X_i' u_j}{T-k} \\ & - \frac{u_i' X_j}{T-k} \left(\frac{X_j' X_j}{T-k} \right)^{-1} \frac{X_j' u_j}{\sqrt{T-k}} \\ & + \frac{u_i' X_i}{\sqrt{T-k}} \left(\frac{X_i' X_i}{T-k} \right)^{-1} \frac{X_i' X_j}{T-k} \left(\frac{X_j' X_j}{T-k} \right)^{-1} \frac{X_j' u_j}{T-k}. \end{aligned} \tag{4.7}$$

The proof of step 1 is completed by combining (4.7) with the following points:

1. $\frac{1}{\sqrt{T-k}} X_i' u_i = \frac{1}{\sqrt{T-k}} \sum_{t=1}^{T} x_{it} u_{it} = O_p(1)$,
2. $\frac{1}{T-k} X_i' X_j = \frac{1}{T-k} \sum_{t=1}^{T} x_{it} x_{jt}' \overset{a.s.}{\to} \mathbb{E}[x_{it} x_{jt}']$, and
3. $\frac{1}{T-k} X_i' u_j = \frac{1}{T-k} \sum_{t=1}^{T} x_{it} u_{jt} = O_p(\frac{1}{\sqrt{T}})$,

where $\overset{a.s.}{\to}$ denotes almost sure convergence. To see the first point, note that $\text{var}(\frac{1}{\sqrt{T-k}} \sum_{t=1}^{T} x_{it} u_{it}) = O(1)$ by the boundedness of the second moments of x_{it} and u_{it}. The second point follows from Theorem 3.34 in White (2001), because $\{x_{it} x_{jt}'\}$ is stationary and ergodic ((I) and Theorem 3.35 in White (2001)) and $\mathbb{E}[|x_{it}^{(l)} x_{jt}^{(j)}|] < \infty$ for $j, l = 1, \ldots, k$ ((II) and Hölder's inequality). Here, $x_{it}^{(l)}$ denotes the l-th component of the vector x_{it}. The third point is similar to the first and follows from the fact that $\text{var}(\frac{1}{T-k} \sum_{t=1}^{T} x_{it} u_{jt}) = O(\frac{1}{T})$.

Step 2: $\frac{1}{\sqrt{T-k}} \sum_{j<i}^{N} \hat{\sigma}_{ij} v_{ij} \overset{d}{\to} N(0, \sum_{j<i}^{N} \sigma_{ii} \sigma_{jj} c_{ij} c_{ij}')$, where c_{ij} contains the stacked lower diagonal elements of $\mathbb{E}[\bar{x}_{it} \bar{x}_{jt}'] + \mathbb{E}[\bar{x}_{jt} \bar{x}_{it}']$ and the vector $\bar{x}_{it}$ contains the regressors x_{it} after fixed

effect demeaning.

Proof: Consider the $\frac{N(N-1)k(k+1)}{4} \times 1$ vector

$$\zeta_T = \frac{1}{\sqrt{T-k}}(\hat{\sigma}_{12}v_{12}', \hat{\sigma}_{13}v_{13}', \hat{\sigma}_{23}v_{23}', \ldots, \hat{\sigma}_{1N}v_{1N}', \ldots, \hat{\sigma}_{N-1,N}v_{N-1,N}')'.$$

We will show that ζ_T is asymptotically normal. Rewrite ζ_T as

$$\zeta_T = \frac{1}{T-k}V_T Z_T,$$

where

$$Z_T = \sqrt{T-k}(\hat{\sigma}_{12}, \hat{\sigma}_{13}, \hat{\sigma}_{23}, \ldots, \hat{\sigma}_{1N}, \ldots, \hat{\sigma}_{N-1,N})',$$

and

$$V_T = \underbrace{\begin{pmatrix} v_{12} & 0 & 0 & \cdots & 0 \\ 0 & v_{13} & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & v_{N-1,N} \end{pmatrix}}_{\left(\frac{N(N-1)k(k+1)}{4} \times \frac{N(N-1)}{2}\right)}.$$

It follows from step 1 that

$$\begin{aligned} Z_T &= \frac{1}{\sqrt{T-k}}(u_1'u_2, u_1'u_3, u_2'u_3, \ldots, u_1'u_N, u_{N-1}'u_N)' + o_p(1) \\ &= \frac{1}{\sqrt{T-k}}\sum_{t=1}^{T}\underbrace{(u_{1t}u_{2t}, u_{1t}u_{3t}, u_{2t}u_{3t}, \ldots, u_{1t}u_{Nt}, \ldots, u_{N-1,t}u_{Nt})'}_{\equiv \xi_t} + o_p(1). \end{aligned}$$

Note that $\{\xi_t\}$ is an i.i.d. sequence of random vectors with zero mean and variance Σ_ξ. Σ_ξ is a diagnoal matrix with diagonal elements given by the $\frac{N(N-1)}{2} \times 1$ vector $(\sigma_{11}\sigma_{22}, \sigma_{11}\sigma_{33}, \sigma_{22}\sigma_{33}, \ldots$ $\ldots, \sigma_{N-1,N-1}\sigma_{NN})'$. Therefore, by standard limit theorems

$$Z_T \xrightarrow{d} N(0, \Sigma_\xi).$$

Regarding V_T, note that, similar to point 2 in step 1, $\frac{1}{T-k}v_{ij} \overset{a.s.}{\to} c_{ij}$. Thus,

$$\frac{1}{T-k}V_T \overset{a.s.}{\to} C \equiv \begin{pmatrix} c_{12} & 0 & 0 & \cdots & 0 \\ 0 & c_{13} & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & c_{N-1,N} \end{pmatrix}.$$

As a consequence,

$$\zeta_T \overset{d}{\to} N(0, C\Sigma_\xi C').$$

Finally, rewrite $\frac{1}{\sqrt{T-k}}\sum_{j<i}^{N} \hat{\sigma}_{ij}v_{ij} = J\zeta_T$, where $J = \iota' \otimes I_{\frac{k(k+1)}{2}}$, ι is a $\frac{N(N-1)}{2} \times 1$ vector of ones, and $\otimes$ denotes the Kronecker product. Obviously,

$$\frac{1}{\sqrt{T-k}}\sum_{j<i}^{N} \hat{\sigma}_{ij}v_{ij} \overset{d}{\to} N(0, JC\Sigma_\xi C'J'),$$

and by simple algebraic computations we obtain $JC\Sigma_\xi C'J' = \sum_{j<i}^{N} \sigma_{ii}\sigma_{jj}c_{ij}c_{ij}'$.

Step 3: $\frac{1}{T-k}\hat{V} \overset{p}{\to} \sum_{j<i}^{N} \sigma_{ii}\sigma_{jj}c_{ij}c_{ij}'$ as $T \to \infty$, where $\overset{p}{\to}$ denotes convergence in probability.

Proof: We have

$$\begin{aligned} \frac{1}{T-k}\hat{V} &= \frac{1}{(T-k)^3}\sum_{j<i}^{N} \hat{\sigma}_{jj}\hat{\sigma}_{ii}\mathrm{tr}(M_jM_i)v_{ij}v_{ij}' \\ &= \sum_{j<i}^{N} \hat{\sigma}_{jj}\hat{\sigma}_{ii}\frac{\mathrm{tr}(M_jM_i)}{T-k}\frac{v_{ij}}{T-k}\frac{v_{ij}'}{T-k}. \end{aligned}$$

From the derivations in step 1 and 2 it obviously follows that $\hat{\sigma}_{ii} \overset{p}{\to} \sigma_{ii}$ and $\frac{1}{T-k}v_{ij} \overset{a.s.}{\to} c_{ij}$. Using results from Yang and Feng (2002), it can be inferred that

$$T - 2(k+1) \leq \mathrm{tr}(M_jM_i) \leq T - (k+1).$$

Hence, $\frac{1}{T-k}\mathrm{tr}(M_jM_i)$ converges almost surely to 1.

Finally, combining step 1 to step 3 with standard results from probability theory completes the proof of Porpositon 4.5.

□

Proof of Proposition 4.7

The proof of Proposition 4.7 is analogous to that of Proposition 4.5. It suffices to establish the following facts:

(i) $\sqrt{T}\hat{\rho}_{ij} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} + o_p(1)$,

(ii) $\frac{1}{\sqrt{T}} \sum_{j<i}^{N} \hat{\rho}_{ij} v_{ij} \xrightarrow{d} N(0, \sum_{j<i}^{N} c_{ij}c_{ij}')$, and

(iii) $\frac{1}{T}\hat{V}_R \xrightarrow{p} \sum_{j<i}^{N} c_{ij}c_{ij}'$.

(i) has been shown by Pesaran (2004) and, similar to step 1 above, also holds under the assumptions made here. (ii) and (iii) can be shown using similar arguments as in step 2 and step 3.

□

Bibliography

Agarwal, V. and N. Y. Naik (2004). Risks and portfolio decisions involving hedge funds. *Review of Financial Studies 17*, 63–98.

Andersson, J. (2001). On the normal inverse Gaussian stochastic volatility model. *Journal of Business & Economic Statistics 19*, 44–54.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Andrews, D. W. K. (2005). Cross-section regression with common shocks. *Econometrica 73*, 1551–85.

Artzner, P., F. Delbeaen, J. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9*, 203–228.

Aumann, R. and R. Serrano (2008). An economic index of riskiness. *Journal of Political Economy 116*, 810–836.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*, 49–60.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Baltagi, B. H., Q. Feng, and C. Kao (2011). Testing for sphericity in a fixed effects panel data model. *Econometrics Journal 14*, 25–47.

Banerjee, A., R. L. Lumsdaine, and J. H. Stock (1992). Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence. *Journal of Business & Economic Statistics 10*(3), 271–287.

Bao, Y. (2009). Estimation risk-adjusted Sharpe ratio and fund performance ranking under a general return distribution. *Journal of Financial Econometrics 7*, 152–173.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics 24*, 1–13.

Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *Review of Economic Studies 53*, 369–384.

Billingsley, P. (1968). *Convergence of probability measures.* John Wiley & Sons.

Blanchard, O. J. and M. W. Watson (1982). Bubbles, rational expectations, and financial markets. In P. Wachtel (Ed.), *Crisis in the economic and financial structure*, pp. 295–315. Lexington, Mass.: Lexington Books.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2009). A discrete-time model for daily S&P 500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics 150*, 151–166.

Boyle, G. and D. Conniffe (2008). Compatibility of expected utility and μ/σ approaches to risk for a class of non location-scale distributions. *Economic Theory 35*, 343–366.

Breusch, T. S. and A. R. Pagan (1980). The lagrange multiplier test and its application to model specification tests in econometrics. *Review of Economic Studies 47*, 239–53.

Brockett, P. L. and Y. Kahane (1992). Risk, return, skewness and preference. *Management Science 38*, 851–866.

Busetti, F. and A. M. R. Taylor (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics 123*, 33–66.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton Univ. Press.

Chu, C.-S. J., M. Stinchcombe, and H. White (1996). Monitoring structural change. *Econometrica 64*, 1045–1065.

Cogneau, P. and G. Hübner (2009a). The (more than) 100 ways to measure portfolio performance. part 1: Standardized risk-adjusted measure. *Journal of Performance Measurement 13*, 56–71.

Cogneau, P. and G. Hübner (2009b). The (more than) 100 ways to measure portfolio performance. part 2: Special measures and comparison. *Journal of Performance Measurement 14*, 56–69.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1*, 223–236.

Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes*. Chapman & Hall.

Cuñado, J., L. A. Gil-Alana, and F. P. de Gracia (2005). A test for rational bubbles in the nasdaq stock index: a fractionally integrated approach. *Journal of Banking and Finance 29*, 2633–2654.

De Giorgi, E. (2005). Reward-risk portfolio selection and stochastic dominance. *Journal of Banking and Finance 29*, 895–926.

Diba, B. T. and H. I. Grossman (1988). Explosive rational bubbles in stock prices? *American Economic Review 78*(3), 520–530.

Dowd, K. (2000). Adjusting for risk: An improved Sharpe ratio. *International Review of Economics & Finance 9*, 209–222.

Driscoll, J. C. and A. C. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics 80*, 549–60.

Eling, M. and F. Schuhmacher (2007). Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking & Finance 31*, 2632–2647.

Eriksson, A., E. Ghysels, and F. Wang (2009). The normal inverse Gaussian distribution and the pricing of derivatives. *The Journal of Derivatives 16*, 23–37.

Evans, G. W. (1991). Pitfalls in testing for explosive bubbles in asset prices. *American Economic Review 81*(4), 922–930.

Farinelli, S., M. Ferreira, D. Rossello, M. Thoeny, and L. Tibiletti (2008). Beyond Sharpe ratio: Optimal asset allocation using different performance ratios. *Journal of Banking & Finance 32*, 2057–2063.

Foster, D. and S. Hart (2009). An operational measure of riskiness. *Journal of Political Economy 117*, 785–814.

Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. *Journal of Econometrics 69*, 393–414.

Frömmel, M. and R. Kruse (2011). Testing for a rational bubble under long memory. Quantitative Finance, forthcoming.

Froot, K. and M. Obstfeld (1991). Intrinsic bubbles: the case of stock prices. *American Economic Review 81*, 1189–1214.

Galbraith, J. K. (1993). *A short history of financial euphoria*. New York: Viking.

Gennaioli, N., A. Shleifer, and R. W. Vishny (2010). Neglected risks, financial innovation, and financial fragility. NBER working paper 16068, http://www.nber.org/papers/w16068.

Golec, J. and M. Tamarkin (1998). Bettors love skewness, not risk, at the horse track. *Journal of Political Economy 106*, 205–225.

Grandell, J. (1991). *Aspects of Risk Theory*. Springer.

Gürkaynak, R. S. (2008). Econometric tests of asset price bubbles: taking stock. *Journal of Economic Surveys 22*(1), 166–186.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Harvey, C. R. and A. Siddique (2000). Conditional skewness in asset pricing tests. *Journal of Finance 55*, 1263–1295.

Kim, J.-Y. (2000). Detection of change in persistence of a linear time series. *Journal of Econometrics 95*, 97–116.

Kim, J.-Y., J. Belaire-Franch, and R. B. Amador (2002). Corrigendum to 'detection of change in persinstence of a linear time series'. *Journal of Econometrics 109*, 389–392.

LeRoy, S. F. and R. D. Porter (1981). The present-value relation: Tests based on implied variance bounds. *Econometrica 49*, 555–574.

Levy, H. (1992). Stochastic dominance and expected utility: Survey and analysis. *Management Science 38*, 555–593.

Leybourne, S., T.-H. Kim, V. Smith, and P. Newbold (2003). Tests for a change in persistence against the null of difference-stationarity. *The Econometrics Journal 6*(2), 291–311.

Lien, D. (2002). A note on the relationships between some risk-adjusted performance measures. *The Journal of Futures Markets 22*, 483–495.

Loistl, O. (1976). The erroneous approximation of expected utility by means of a Taylor's series expansion: Analytic and computational results. *American Economic Review 66*, 904–910.

Malkiel, B. G. and A. Saha (2005). Hedge funds: Risk and return. *Financial Analysts Journal 61*, 80–88.

Mammitzsch, V. (1986). A note on the adjustment coefficient in ruin theory. *Insurance: Mathematics and Economics 5*, 147–149.

Meyer, J. (1987). Two-moment decision models and expected utility maximization. *American Economic Review 77*, 421–430.

Meyer, J. and R. H. Rasche (1992). Sufficient conditions for expected utility to imply mean-standard deviation rankings: empirical evidence concerning the location and scale condition. *The Economic Journal 102*, 91–106.

Moscone, F. and E. Tosetti (2009). A review and comparison of tests of cross-section independence in panels. *Journal of Economic Surveys 23*, 528–61.

Nabeya, S. (1999). Asymptotic moments of some unit root test statistics in the null case. *Econometric Theory 15*, 139–149.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. CESifo Working Paper No. 1229.

Pesaran, M. H. (2012). Testing weak cross-sectional dependence in large panels. working paper.

Pesaran, M. H., A. Ullah, and T. Yamagata (2008). A bias-adjusted LM test of error cross-section independence. *Econometrics Journal 11*, 105–27.

Phillips, P. C. B., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International Economic Review 52*, 201–226.

Pitts, S. M., R. Grübel, and P. Embrechts (1996). Confidence bounds for the adjustment coefficient. *Advances in Applied Probability 28*, 802–827.

Rydberg, T. H. (2000). Realistic statistical modelling of financial data. *International Statistical Review 68*, 233–258.

Schuhmacher, F. and M. Eling (2011). Sufficient conditions for expected utility to imply drawdown-based performance rankings. *Journal of Banking & Finance 35*, 2311–2318.

Schulze, K. (2010). Existence and computation of the Aumann-Serrano index of riskiness. Working Paper, McMaster University. Available at http://www.math.mcmaster.ca/ schulzek/research.html.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business 39*, 119–138.

Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management 21*, 49–58.

Shiller, R. J. (1981). Do stock prices move to much to be justified by subsequent changes in dividends? *American Economic Review 71*, 421–436.

Shiller, R. J. (2001). *Irrational exuberance.* New York: Broadway Books.

Shiller, R. J. and P. Perron (1985). Testing the random walk hypothesis: power versus frequency of observation. *Economics Letters 18*(4), 381–386.

Sortino, F. A. and L. N. Price (1994). Performance measurement in a downside risk framework. *The Journal of Investing 3*, 59–64.

Sortino, F. A., R. van der Meer, and A. Plantinga (1999). The dutch triangle. *The Journal of Portfolio Management 26*, 50–58.

Taylor, A. M. R. (2005). Fluctuation tests for a change in persistence. *Oxford Bulletin of Economics and Statistics 67*(2), 207–230.

Taylor, R. and S. Leybourne (2004). Some new tests for a change in persistence. *Economics Bulletin 3*(39), 1–10.

Treynor, J. L. (1965). How to rate management of investment funds. *Harvard Business Review 43*, 63–75.

West, K. (1987). A specification test for speculative bubbles. *Quarterly Journal of Economics 102*, 553–580.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.

White, H. (2001). *Asymptotic theory for econometricians.* San Diego, CA: Academic Press.

Yang, Z. P. and X. X. Feng (2002). A note on the trace inequality for products of hermitian matrix power. *Journal of Inequalities in Pure and Applied Mathematics 3*. Article 78.

Zakamouline, V. and S. Koekebakker (2009). Portfolio performance evaluation with generalized Sharpe ratios: Beyond the mean and variance. *Journal of Banking & Finance 33*, 1242–1254.

Zeileis, A., F. Leisch, C. Kleiber, and K. Hornik (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics 20*, 99–121.

Zellner, A. (1962). An efficiency method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association 57*, 348–68.