Econometric Analysis of International Financial Markets

Thomas Ernst Herbert Dimpfl

Dissertation zur Erlangung des Grads eines Doktors der Wirtschaftswissenschaft (Dr. rer. pol.) der Universität Erfurt, Staatswissenschaftliche Fakultät

2010

Gutachter: Prof. Dr. Robert Jung

Gutachter: Prof. Dr. Joachim Grammig

Datum der Disputation: 16. Juli 2010

urn:nbn:de:gbv:547-201000562

Zusammenfassung

Die zentrale Fragestellung meines Dissertationsprojektes „Ökonometrische Untersuchung internationaler Finanzmärkte“ ist der Zusammenhang globaler Finanzmärkte in Bezug auf Informations- und Volatilitätsübertragung. Mit Hilfe verschiedener ökonometrischer Methoden werden gezielt Dynamiken offengelegt und einige der in der Literatur als Standard angesehenen Phänomene hinterfragt.

Der erste Teil behandelt die sogenannten Informations- und Volatilitätsspillovers. Von zentraler Bedeutung ist hier die Tatsache, dass aus globaler Sicht der Handel an Börsen als kontinuierlich angesehen werden kann. Aus diesem Grund sollte es möglich sein, Informations- und Volatilitätsspillovers um den Erdball in Übereinstimmung mit der Abfolge aus Öffnen und Schließen der Märkte in Asien, Europa und den USA nachzuvollziehen. Der zweite Teil der Arbeit setzt sich mit Kointegration von Aktienmärkten und den speziellen Herausforderungen von Finanzmarktdatensätzen auseinander. Kointegration ist eine ökonometrische Methode, welche herangezogen wird, um den Integrationsgrad internationaler Finanzmärkte zu messen. Die Ergebnisse sind jedoch sehr heterogen. Wir zeigen, dass internationale Finanzmärkte nicht kointegriert sein können, sofern das „random walk“-Modell für Aktienpreise zutrifft. Mit Hilfe einer Simulationsstudie werden Gründe herausgearbeitet, warum Kointegrationstests andere Schlussfolgerungen nahelegen können. Schließlich widmet sich der letzte Teil der Dissertation der Informationsübertragung von den USA nach Europa zur Zeit der Eröffnung der US-amerikanischen Märkte. Es wird gezeigt, dass Nachrichten aus den USA (welche durch Quantile der Renditeverteilung des S&P 500 identifiziert werden) einen signifikanten Einfluss auf die Renditen und die Volatilität des DAX ausüben und sowohl schnell als auch effizient von deutschen Händlern verarbeitet werden.

Schlagwörter:
Finanzmärkte; Spillover; Kointegration; Volatilität; Ereignisstudie

Summary

The central problem of the dissertation project “Econometric Analysis of International Financial Markets” is the question how financial markets around the globe are linked in terms of information and volatility transmission. Using different econometric techniques some of the dynamics are unraveled and explanations for phenomena taken for granted in the literature so far are proposed.

More precisely, the first aspect covered concerns information and volatility spillovers around the globe, the central aspect being that from a global point of view stock trading is continuous. We therefore state that information and volatility spillovers are traceable around the globe in accordance with the sequence of opening and closing of financial markets in Asia, Europe and the USA. The second subject deals with cointegration of financial markets and the peculiarity of financial data. Cointegration is an econometric technique which is quite frequently used to asses the degree of integration of financial markets. The results are, however, far from being clear-cut. We show that international financial markets are not cointegrated given the commonly used random walk model for stock prices is true. By means of simulation studies we elaborate reasons why the results of cointegration tests can be misleading.

Finally we take a closer look at the information transmission from the USA to Europe at the time when the US markets open. We show that news originating in the USA (which are identified using quantiles of the S&P 500 index return distribution) have a significant impact on the returns and the volatility of the German DAX and are processed rapidly and efficiently by German traders.

Keywords:
financial markets; spillovers; cointegration; volatility; event study

Preface

Iam per complures annos res nummariae in orbe terrarum gestae valde perturbantur. Exemplum recens – metus Graecorum, ne pecunia debita solvi possit – monstrat, quantus timor adhuc perseveret inter conferentes pecuniarum. Ratio systematis monetalis Europaei efficit, ut difficultates consociatorum celerrime in alias civitates transcendant. Quaestionem, quomodo res nummariae omnium gentium inter se conexae et aptae sint, imprimis hoc tempus postulat. Conscriptio huius dissertationis fit igitur medio in hoc gravi discrimine et quaestio iterum atque iterum ad has angustias referet.

Dissertationem meam sine auxilio multorum collegarum et amicorum non perfecerim. Primo omnium educatori et altori meo Robert Jung gratias maximas ago. Multis disputationibus, quae nonnumquam in vehiculo fiebant, is per quadriennium et dux et comes fuit mihi quemque libertatem scientificam concedens. Pro labore et auxilio tuo tibi eo loco sincere gratias agere volo.

Meam magnam gratiam init Joachim Grammig, qui occasionem operis mei ad cathedram suam exponendi et iterum atque iterum cum omnibus sociis operis cathedrae disceptandi dedit. Erga eum gratissimus sum, quod paratus fuit ad sententiam secundam dissertationis meae dicendam.

Deinde gratias ago omnibus, qui me amice comitabantur in arte, praecipue Kerstin Kehrle, Fabian Kleine, Franziska Peter, Henriette Reinhold et Oliver Wünsche. Irenaeo Wolff imprimis gratias dico, quod sine eius accessu magico ad commentarios periodicos directe colligatos verisimile fuisset me desperaturum et ad inopiam pecuniae venturum fuisse. Etiam Achim Ahrens, Cédric André, Robert Fritzsch et Michael Kloß gratias habeo sive pro eorum magno labore sive pro officiis praestatis velut investigatione litterarum et labore indiciorum efficiendorum. Bernd Kroll et Gillian Mansfield gratias persolvo pro auxilio in rebus ad usum legendi pertinentibus.

Denique comiti meo Michael qui non modo inter conscriptionem dissertationis asperitatem meam perferre debebat sed etiam ex tribus annis me hebdomada solum exiente vidit. Gratias pro omnibus!

Erfurt, a. d. IV Kal. Apriles anno post Christum natum bis millesimo decimo

Thomas Dimpfl

Contents

Chapter 1

Introduction

The Asian crisis, the mortgage crisis, the Lehman Brothers bankruptcy: all these events seemingly originated in one global area or even in one country alone, but still ended in turmoil on the stock markets around the globe. This comes as no surprise as both stock and commodity markets all over the world are highly interdependent due to the complex and interwoven network of trade and finance. The benefit of such close relatedness is that traders in all markets, and in stock markets in particular, benefit from a wide range of hedging and diversification opportunities. On the downside, however, any surprising or even shocking event may induce a substantial increase in the volatility of prices and, thus, threaten not only local but also global trading. Furthermore, such events are usually associated with huge losses on the stock markets, which is of course in sharp contrast with the traders' goal of maximizing their profit. It is therefore vital for all traders to be aware of the interdependence of stock markets. Knowledge of not only how, but also how fast and how efficiently information and volatility is transmitted between stock markets around the globe, is therefore beneficial for the individual trader as well as for the market as a whole.

This study uses different econometric approaches to characterize the information transmission mechanisms between global financial markets as well as to describe their interrelatedness in general. Chapter 2 which was written in collaboration with Robert Jung, analyses the transmission of return and volatility spillovers between international financial markets. We are particularly interested in creating a model that captures the characteristic sequence of the opening and closing of financial markets around the globe. For this purpose, we use stock market index futures of three representative indices, namely the Dow Jones Euro Stoxx 50 future, the S&P 500 future and the Nikkei 225 future as proxies for the three major economic regions Europe, USA and Asia.

Returns (which are cleaned for volatility influences) and realized volatilities are developed separately with a structural vectorautoregressive model, thereby accounting for the particular, sequential time structure of opening and closing of the stock markets where the futures are traded. Within this framework, we test hypotheses in the spirit of the Granger-causality tests, investigate the short run dynamics in the three markets using impulse response functions, and identify leadership effects through variance decomposition. Our key results are as follows. Not unexpectedly, return spillovers are found to be weak and short lived, while volatility spillovers are more pronounced and persist. Information from the home market is essential for both returns and volatilities, while the contribution from foreign markets is less pronounced in the case of returns than it is in the case of volatility. Our results are sound with respect to the way the volatility series is computed. Possible gains when applying this modeling strategy as opposed to separate modeling of the time series are illustrated by a short forecast evaluation and an application to the stock market crash on January 14 and 15, 2008.

A further method which is widely used in the empirical financial literature to model the interdependence of financial markets is that of cointegration. The hypothesis is that stock markets are highly interdependent due to the presence of common stochastic trends. More precisely, the long run behavior is assumed to be identical for all stock markets while short run deviations are possible. An issue here is how to identify this relationship. Financial data very often violate the assumptions which are required to derive most cointegration tests. In Chapter 3 we therefore briefly investigate the influence of one particular characteristic of financial data, namely heteroscedasticity, on the Johansen (1991) test for cointegration, the latter being one of the most widely used tests in this context. We use two different cointegration concepts—stationary and stochastic—and evaluate the performance of the Johansen Trace and Maximum Eigenvalue test following some heteroscedasticity and correlation assumptions. We find that the tests in general are quite reliable. However, in some circumstances they seem more apt in detecting cointegration if the data are indeed cointegrated, than in not rejecting cointegration if the data are not cointegrated.

Chapter 4 then revisits the cointegration framework in the context of international financial markets. Although intuitively this econometric technique seems very attractive to model market relationships, we show that international financial markets are not cointegrated if the widely-used random walk model is indeed the appropriate and true model to describe stock prices on a daily basis. We take up and extend previous work by Granger (1986) and Richards (1995) and show that empirical findings are compatible with our theoretical framework. We conclude that results on cointegration of financial markets in previous studies might be due to the lack of power of the testing framework. This is carried out by means of an empirical experiment where we use 28 stock market indices and test for bivariate cointegration. We then simulate indices according to our theoretical model and try to mimic the outcome of the empirical study. We identify common random walk components, correlated innovations and heteroscedasticity as the driving forces behind our empirical results. In particular heteroscedasticity, in conjunction with other features, is a factor which deceives the Johansen cointegration test.

In Chapter 5 we take a closer look at the German stock market and investigate how it is impacted by the opening of the stock markets in the USA. In contrast to the spillover analysis in Chapter 2, we now study the intraday influences from the USA on Germany. The methodological approach here is using an event study framework to study the impact on returns. Volatility will be measured as realized volatility and analyzed with nonparametric techniques. For the purpose of this study, it is necessary to distinguish days with good US news surprises from days with surprising bad US news. We use quantiles of the S&P 500 index return distribution to identify them and to separate them from days when there is no surprising news content. In order to check the adequacy of this selection process, these days are matched with events of macroeconomic importance. We find that the German market reacts to US news announcements which typically precede the opening of the New York Stock exchange. The opening of the market itself and the beginning of trading in the USA is not found to affect German stock prices. On average days, there is no measurable impact on the DAX. Furthermore, once important news is transmitted it is absorbed rapidly into prices. As far as volatility is concerned, we find that the news days identified are marked by significantly higher volatility, both in the morning and in the afternoon, in comparison to days without any news events. Indeed, it is of no importance whether the news is good or bad. Moreover, we can attribute the reported w-shape of volatility (Masset, 2008) in the German stock market to the unexpected news which originate in the USA: on average days, DAX volatility is u-shaped, a feature which is commonly found for stock markets around the globe. If we consider solely the news days, volatility peaks around half past two in the German afternoon trading.

Chapter 6 reviews the results and draws conclusions to the study.

Chapter 2

Financial Market Spillovers Around The Globe

2.1 Introduction

A little while ago the interdependence of international financial markets once again was highlighted by the breakdown of the US mortage financing system. A country-specific peculiarity has spread its effects across the global financial markets. With the burst of the housing bubble and the subsequent decline of the value of mortage assets the so-called mortgage-backed securities (MBS) as well as collateral debt obligations (CDO) deteriorated significantly. International diversification which is usually intended to lower a portfolio's risk position led to the infection of financial markets around the world. Owners of MBSs and CDOs had to face a significant loss. The crisis found (for the time being) its peak in the stock market crash 14th and 15th January 2008. The way it developed, starting in Asia and not even in the USA, shows how interwoven and sensitive financial markets are.

The investigation of these linkages between international financial markets and in particular the transmission of shocks between them has been in the focus of academic researchers and financial practitioners alike for quite some time now. The workhorse in the empirical financial literature for joint modeling of return and volatility transmissions has been the class of (multivariate) Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models which date back to the seminal papers of Engle (1982) and Bollerslev (1986). Important early contributions to this literature are Susmel and Engle (1994) and Lin, Engle, and Ito (1994). Recent papers include Savva, Osborn, and Gill (2005), Baur and Jung (2006), and Wongswan (2006). Typically, these papers concentrate on two financial markets or geographical regions. Moreover, they employ data from (mostly) daily or weekly stock market indices.

The present study deviates from this literature in three important ways. First, we propose separate models for the mean and the volatilities of financial market returns. In particular, we use realized volatilities as suggested by Andersen, Bollerslev, Diebold, and Labys (2001) and estimated daily volatilities as proposed by Bollen and Inder (2002). Second, we seek to model the short run dynamics of financial markets around the globe using structural vectorautoregressive (SVAR) models. This enables us to test various hypotheses in the spirit of Granger-causality testing. Moreover, we can use impulse response functions to analyze short-run dynamics in the system of global financial markets. Finally, we can adopt variance decomposition to identify leadership effects in both the mean and volatility system. Third, we base the empirical analysis on index future data instead of the underlying indices to overcome the widely documented stale quote problem. While some of these issues have been addressed in the literature on financial market linkages before, it is the combination employed in our paper that is novel. To illustrate the possible gains which arise from the combination of the proposed methods we perform a short forecast evaluation.

Global or around-the-clock shock transmission has been employed by Diebold and Yilmaz (2009) who analyze 16 global stock markets using a Garman and Klass (1980) type estimator for volatility. They assume the latter to be stable across one week, an assumption which remains questionable. They then estimate separately a model for the returns and the volatility measure and find that roughly 30% of innovation in returns and volatility is due to foreign markets. In contrast to that, Polasek and Ren (2001) used a multivariate VAR-GARCH-in-mean model, estimated on daily stock index return data, to trace the effects of only three markets on each other: Germany, the USA and Japan. Despite the appealing model the authors seem to ignore the sequence of trading as they allow only lagged influences between the markets. As they use daily data, there should be contemporaneous influence from one market to the next, depending on how the day t is defined. This is due to the fact that trading on the various financial markets around the globe takes place sequentially: when the stock exchanges in Asia close, the European exchanges open and later in the same day the American stock exchanges open. This all happens within the very same trading day and has to be accounted for. We intend to solve this issue by using a structural VAR model instead of a reduced form model only. This allows us to capture the (artificially) contemporaneous effects in the sequence of opening and closing of stock markets. Koutmos and

Booth (1995) recognized this issue and introduced the distinction of calendar time and trading time when analyzing the spillovers between the New York, London and Tokyo stock exchanges. The authors estimate their multivariate EGARCH-model in trading time thus aligning trading around the globe to the same time index. Our intention, however, is to model the spillover effects in calendar time. This allows us to explicitly account for the sequential trading around the globe.

Using a similar methodology like Polasek and Ren (2001) but index future data instead of the underlying stock market indices Pan and Hsueh (1998) examine the linkages between two markets only: the USA and Japan. They perform contemporaneous correlation as well as spillover analysis. Regarding the latter, they find weak, positive mean spillover effects and negative variance spillover effects from the Japanese trading to the USA. In the other direction, they find a negative variance, but no mean spillover effect from the USA to Japan.

When working with intra-day data as we do in this study, Hamao, Masulis, and Ng (1990) introduced the useful distinction between overnight returns (close-to-open) and daytime returns (open-to-close) and, associated with it, contemporaneous correlation and spillover effects. The latter seeks to measure the impact of daytime returns or volatilities of a chronologically upstream market on the daytime returns or volatilities of the following market(s). Thus, spillovers are calculated on the basis of non-overlapping return time spans enabling us to identify possible causal effects in the sense of Granger (1969).

For the subsequent analysis we rely on proxies for three economic regions in the world. As we are interested in around-the-globe information transmission, we select a representative stock index future for Europe, America and Asia which, taken together, almost fully cover 24 hours in terms of trading time. We draw on the Dow Jones Euro Stoxx 50 future as a representative for the European market, the S&P500 future to represent the market in the United States and the Nikkei 225 future as a proxy for the Asian market.

The study continues as follows. Section 2.2 presents the econometric model along with the variables used for estimation as well as the specific time structure of the analysis. Section 2.3 describes the data and section 2.4 presents the empirical results. Section 2.5 presents a short forecast application of the spillover model to the crisis of January 2008. Section 2.7 concludes.

2.2 Econometric Model

In the introduction spillovers were defined as the impact of one market on the chronologically following market. In order to trace these effects we estimate separately a structural vectorautoregressive model (SVAR) of order p for both the standardized log-returns ($\tilde{r}$) and the logarithms of the volatility measures (σ; to be defined below) of the Dow Jones Euro Stoxx 50 future (FESX), the S&P500 future (FSP) and the Nikkei 225 future (FNI). This approach has two important advantages: first, when creating the volatility time series, we use the information available more efficiently then we would when using a GARCH model. The realized volatility measures effectively incorporate more information than using only squared, lagged error terms from the mean equation in the GARCH equation. Second, in the context of a multivariate GARCH model the specific opening and closing sequence of financial markets would require contemporaneous effects of the variances on each other. Such a model, however, is, to the best of our knowledge, not identifiable. Using an SVAR model with volatility measures on the other hand allows straight-forward estimation of the volatility dynamics. Care, however, has to be taken when modeling the return series due to the presence of heteroscedasticity which we will address with an approach similar to weighted least squares estimation.

The return of the individual futures is measured as the difference in the logarithm of the respective transaction prices, that is,

$$r_{t+\Delta} = \ln p_{t+\Delta} - \ln p_t \; . \tag{2.1}$$

This calculation assumes a continuously compounded basis. In the following analysis we need, for example, intraday returns in which case p_t would be the opening price on day t and $p_{t+\Delta}$ would constitute the transaction price at a specific time within the day. The latter is usually the last price fixed at the close of the stock market.

As is well known, return time series suffer from heavy tails and volatility clustering which is also the case here. As we follow an approach which models returns and volatilities separately we have to account for the presence of conditional heteroscedasticity in the return time series. We therefore standardize the returns by their realized volatility which has recently been proposed by Pesaran and Pesaran (2007). This proceeding ensures that the return series are approximately Gaussian and homoscedastic. Pesaran and Pesaran (2007) argue that the interpretation of correlations estimated with non-gaussian returns can be misleading and therefore propose to standardize the returns. They refer to returns standardized by realized volatilities as "devolatized returns". Denote these devolatized returns by $\tilde{r}_t$, then

$$\tilde{r}_t = \frac{r_t}{\sigma_t} \tag{2.2}$$

and σ_t is the square root of the realized volatility measure as defined below.

For the investigation of the volatility linkages we consider two different measures: the realized volatility measure as proposed by Andersen, Bollerslev, Diebold, and Labys (2003) and the daily volatility estimate proposed by Bollen and Inder (2002). Both methods seek to overcome the well documented market microstructure effects present in high-frequency financial data when estimating the unobservable volatility process.

Andersen, Bollerslev, and Diebold (2002) argue that due to market microstructure frictions it is undesirable to sample returns infinitely often as would be required to approach the true underlying volatility. When summing up the squared returns, one would at the same time accumulate the noise present in the market which would lead to non-trivial measurement errors. To overcome this issue the realized volatility of Andersen *et al.* (2003) is, therefore, calculated using returns computed over sufficiently large time intervals Δ. Specifically, they define the daily realized variance on day t as

$$\sigma^2_{t,\Delta} = \sum_{j=1}^{1/\Delta} r^2_{t-1+j\Delta,\Delta} \tag{2.3}$$

where $\frac{1}{\Delta}$ defines the number of intervals used for calculating the volatility measures. In a sample containing observations from 24 hours continuous trading, $\frac{1}{\Delta}$ would be 96 in case that the individual intervals were 15 minutes long. The realized volatility is then given by the square-root of $\sigma^2_{t,\Delta}$. In their application, for example, Andersen *et al.* (2003) use thirty minute returns when computing the realized volatility of exchange rates.

A drawback in using, for example, returns computed over 15 minute intervals is the loss of information contained in the observations within the interval. Bollen and Inder (2002) therefore propose a VARHAC estimator to explicitly account for the different autocorrelation structures in intraday returns induced by market microstructure effects. Specifically, they estimate for each trading day t and for each return series an AR-model

$$r_{\tau,t} = \sum_{j=1}^{p_t} \alpha_{j,t}\, r_{\tau-j,t} + \varepsilon_{\tau,t} \tag{2.4}$$

where τ is the intraday time stamp. The optimal lag length per day p_t is chosen by an information criterion. The purpose of this procedure is to purge the returns from microstructure noise. The estimate of the daily volatility is then computed as

$$\sigma_t^2 = \frac{RSS_t}{\left[1 - \sum_{j=1}^{p_t} \hat{\alpha}_{j,t}\right]^2}, \qquad \text{where } RSS_t = \sum_{j=p_t+1}^{n_t} \left(r_{\tau,t} - \sum_{j=1}^{p_t} \hat{\alpha}_{j,t}\, r_{\tau-j,t} \right)^2 \tag{2.5}$$

and n_t is the number of observations per day. The estimator (2.5) is efficient in the sense that it utilizes all the available high-frequency data.

To model the volatility transmission between the three major financial centres around the globe, we follow Andersen, Bollerslev, Christoffersen, and Diebold (2006) who suggest to treat the derived volatility time series as if it was directly observed. This allows for the straightforward application of standard estimation techniques which are briefly presented in the following.

To trace the spillover effects we suggest to use a structural VAR model on a daily frequency. Let $\boldsymbol{x}_t$ be the (3×3) vector which contains the $\tilde{r}_{i,t}$ or the $\ln(\sigma_{i,t})$, respectively. Then the structural model is given by

$$\begin{aligned} \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ b_{21,0} & 0 & 0 \\ b_{31,0} & b_{32,0} & 0 \end{pmatrix} \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \\ &+ \sum_{i=1}^{p} \begin{pmatrix} b_{11,i} & b_{12,i} & b_{13,i} \\ b_{21,i} & b_{22,i} & b_{23,i} \\ b_{31,i} & b_{32,i} & b_{33,i} \end{pmatrix} \begin{pmatrix} x_{1,t-i} \\ x_{2,t-i} \\ x_{3,t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix} \end{aligned} \tag{2.6}$$

or in matrix notation by

$$\boldsymbol{x}_t = \boldsymbol{a} + \sum_{i=0}^{p} \boldsymbol{B}_i \boldsymbol{x}_{t-i} + \boldsymbol{\varepsilon}_t \tag{2.7}$$

where the index t indicates a trading day and p is the order of the vector autoregression. The matrix $\boldsymbol{B}_0$ will be lower diagonal with zeros on the main diagonal due to the time structure of our analysis. Consider Figure 2.1 which presents the trading times of the stock markets in consideration. Let a particular trading day t start at 23:00 GMT. As we run a regression on a daily basis, anything that happens between 23:00 GMT and 22:59 GMT of the following day will be indexed with t. This restricts the possible causal influence in our SVAR-model: the FNI can only be influenced by the FESX and the FSP of the previous trading day $t-1$ as it is the market which opens first on day t. The FESX on day t, however, may be influenced by the same day FNI (as the Singapore Exchange will be closed again by the time Eurex opens) and the FSP of the previous trading day. Similarly, the FSP on day t may be influenced by the same day FESX and FNI as both markets in Europe and Singapore were already or are still open on that day t.

The described ordering, however, is not unique as there is no natural justification for why a particular trading day t should start at 23:00 GMT. It will, therefore, be useful to shift the beginning of the notational day t to the opening of the Singapore stock market, to the opening of Eurex and to the opening of the Chicago Mercantile Exchange and to estimate the model anew each time. This can be used as a check for robustness of the model: the artificial cut between t and $t-1$ somewhere between 0:00 and 24:00 GMT is not supposed to influence the estimation results.

If the markets were fully efficient in terms of information processing we would expect the matrix $\boldsymbol{B}_1$ to be upper diagonal and the model to be an AR(1)-model only in the case of the return model. This would reflect that markets immediately adjust to new information and that information which is generated in an upstream market is accounted for immediately. In case of the volatility model we do not have any a priori assumptions on how the AR-matrices would be structured. To be able to justify the often described volatility persistence (see, for example, Poterba and Summers, 1986; Kearns and Pagan, 1993) we would expect the order of the autoregressive model p to be greater than one.

The specific structure of the SVAR model, or more precisely the fact that the matrix $\boldsymbol{B}_0$ only contains non-zero elements on the lower diagonal, permits direct, linewise estimation of the model by ordinary least-squares. This circumvents the necessity of Cholesky Decomposition which would otherwise be used to back out the structural parameters after estimation of the reduced form VAR

$$\boldsymbol{x}_t = \boldsymbol{C}\boldsymbol{a} + \sum_{i=1}^{p} \boldsymbol{C}\boldsymbol{B}_i \boldsymbol{x}_{t-i} + \boldsymbol{C}\boldsymbol{\varepsilon}_t \tag{2.8}$$

which results from (2.7) by premultiplication with $\boldsymbol{C} = (\boldsymbol{I} - \boldsymbol{B}_0)^{-1}$. The same structure, however, would only allow for exactly one variable ordering in the Cholesky Decomposition: the sequence of influence will always be from Japan to Europe to the United States to Japan and so on.

In order to trace the linkages between the three stock markets we perform impulse response analysis and variance decomposition (see Hamilton, 1994, for example). For the mean model, their interpretation is straightforward. In case of the volatility model it may seem more complicated at first glance as the variance of a variance measure would be the fourth moment of the original time series already. However, we will follow the hands-on approach of Andersen *et al.* (2006) who use the realized volatilities as if they were an ordinary, i.e. observed time series. Consequently the conclusions drawn from impulse response analysis and variance decomposition are only considered in the context of the volatility model without direct linkages to the mean model.

In the above situation variance decomposition is an additional tool to detect spillovers (both in mean and volatility): It provides an answer to the question of which proportion of an s-step-ahead forecast error variance can be attributed to a shock in any one market. Based on this idea Hasbrouck (1991) introduced a decomposition of the long-run variance of a time series. Its purpose is to derive the contribution of the innovation error in one stock market to the total variation present in the system. This is what ultimately measures the magnitude of the spillover effect: the contribution of one market to the price discovery or the volatility realization of the other markets.

2.3 Institutional Aspects and Data Description

For the subsequent analysis we can exploit the richness of our datasets containing intra-daily transaction prices of the Dow Jones Euro Stoxx 50 future (traded at Eurex), the S&P500 future (traded at the Chicago Mercantile Exchange, CME) as well as the Nikkei 225 future (traded at the Singapore Exchange, SGX). The datasets are obtained from Olsen Financial Technologies and are sampled in minutes. The data cover futures contracts over an almost four years-period from 1st July 2002 to 31st May 2006. All futures are denominated in local currencies.

The data are split into an estimation part covering 1st July 2002 to 29th May 2006 and a holdout part (two days: 30th and 31st May 2006) which we use to illustrate the forecast accuracy of our model.

Previous studies dedicated to spillover analysis like Lin *et al.* (1994) and Baur and Jung (2006) used indices instead of futures. The usage of stock marked indices, however, brings along the so-called stale quote problem. This means that the index when calculated for the first time in the morning of the new trading day might be calculated based on data from the previous trading day and does, thus, not reflect new information. The reason is that when the index is calculated in the morning for some stocks new prices may not yet be available. In this case, the previous day closing price - a stale quote - is used to calculate the index.

To overcome the stale quote problem, it is necessary to use a suitable proxy for the opening quote of the stock index. Proposals in the literature vary from opening plus 5 minutes into the trading day up to opening plus 30 minutes. While such proxies help to overcome the stale quote problem, they deplete the data from vital information necessary to correctly measure the spillover effects we seek to identify. In today's electronic markets new information is rapidly incorporated into quotes and, thus, also reflected in transaction prices. The strategy to approximate the "true" first quote by a quote 5 or 10 minutes after the market's opening might, thus, dilute the results in the same way as the stale quote problem: prices of some underlying stocks might have already changed within these 5 or 10 minutes. The approximative opening quote would then again not reflect the true opening index value.

The use of index future data helps to overcome the stale quote problem without loss of information from the market opening. Index futures are self-contained securities and, thus, the first transaction in the morning of a new trading day is driven only by information available to the market at this point in time. A slight drawback of using futures is that a continuous dataset is not available for a time horizon greater than nine months. So in order to obtain a continuous sample covering the four years period the single future contracts are combined such that the future closest to maturity is selected into the continuous sample. The transition from one future to the next occurs always mid March, June, September, and December. The last trading day is excluded to avoid possible influence of the settlement and to ensure continuity within the single days (especially for the calculation of the realized volatility measure). Earlier transmission as is sometimes advocated in the literature does not seem plausible as the traded volume almost entirely shifts to the new contract after settlement of the previous one (see also Carchano and Pardo, 2009).

An important aspect of our analysis is the creation of a dataset (both for the returns and for the volatilities) containing daily data which are free from overlaps within the day. Throughout the four years, this is not an issue for the trading at the SGX (see Figure 2.1). Our dataset contains data from the Open Outcry Trading period which starts at 7:55 and ends at 14:25 Singapore Time (SGT) with a one hour interruption from 10:15 to 11:15 SGT. These times did not change within the four years where data are available. As there is no overlap in trading times between the SGX and the CME as well as between the SGX and Eurex, we calculate the log-returns for the FNI as open-to-close returns.

The FSP is traded from 8:30 to 15:15 Central Standard Time (CST) throughout the four years. Its return is also calculated as open-to-close return. The trading times at Eurex changed during the four years. Before 21st November 2005 continuous trading started at 9:00 and ended at 20:00 Central European Time (CET). From 21st November 2005 on, Eurex extended trading hours for OTC-trade of their benchmark products from 9:00 to 22:00 CET. So before the extension there was an overlap of 4.5 hours while it extended to 6.5 hours after 20th November 2005. In order to obtain a clean-cut time structure we remove overlapping trading hours of the US and the European market by calculating the FESX return as open to 13:30 CET. We restrict ourselves to this timespan following the idea of Menkveld, Koopman, and Lucas (2007) who suggest to interrupt such a time series according to economically relevant points in time. We choose to truncate the German time series (and not the FSP) keeping in mind the considerations of Susmel and Engle (1994). Applied to the present context the reasoning is as follows: recall again that information can only be transmitted from east to west. In this case the European morning trade should convey information which is interesting for the traders in the United States and accounted for as soon as trading opens. When both markets are open, global information should be processed in both markets equally. So spillovers to the Japanese market should be originating in the US market as it contains additional information as compared to the FESX because its trading hours are up to two hours 15 minutes longer.

Cutting the FESX data at 13:30 CET also ensures that we have the same time of trading in the morning in the European market throughout the sample, even during the one week when the daylight savings time is introduced in the USA already while in Europe it is only introduced one week later.

The return data are sampled such that common days without trading (weekends and common holidays) are excluded from the sample. If at least one market is open for trading the respective day remains in the sample. The market(s) which is (are) closed is (are) assigned a return of zero to indicate that adjustment to new information was not possible on that respective day. This proceeding leaves us with a sample containing 1,019 daytime returns when the FNI is ordered first. In the other cases one observation is lost as the first FNI return (or the first FNI and FESX returns) is (are) dropped when we let the day start in Europe or in the USA, respectively. Table 2.1 provides descriptive statistics of the standardized return series. As can be seen the standardization leaves the time series slightly leptokurtic. The null hypothesis of the Jarque-Bera test that the standardized returns are indeed normally distributed cannot be rejected in two of the three cases. The bottom part of Table 2.1 presents sample correlations between the FNI, FESX and FSP in t with FNI, FESX and FSP in t and $t-1$, respectively. It should be noted that FNI_t and FSP_{t-1} are negatively correlated and that the size of the negative correlation is remarkably high. FESX_t and FSP_{t-1} are also negatively correlated but to a lower extent.

As regards the intraday volatilities we choose to use 5-minute returns for the calculation of the realized variance in Equation (2.3) as done, for example, by Andersen *et al.* (2006) in order to circumvent market microstructure effects. As the futures are not continuously traded but only a few hours a day we restrict the calculation of the daily realized variance to the available time span. This means that we do not include overnight returns in the calculation of the day t realized variance and we calculate the necessary squared returns only while the future is actually traded.

In case of the volatility estimator proposed by Bollen and Inder (2002) we compute the returns on a one-minute basis. As in the case of the mean returns, both measures of the FESX volatility are calculated using only data until 13:30 CET. Although the measure has initially been proposed for transaction data, we can still justify its application with one-minute returns. As we have a dataset available containing transaction data of the FESX we calculated the

VARHAC volatility estimator based on transaction data. It turned out that the Schwarz Information Criterion suggested on average a lag length of 7.4 (varying between 3 and 50). As the average elapsed time between two transactions is 2 seconds, this corresponds, on average, to a 15 seconds lag (varying from 6 to 100 seconds). So when aggregating the data to one minute intervals we should still expect some autocorrelation structure. It seems reasonable to assume that the FNI and FSP show a similar structure and, thus, to apply the same proceeding to these futures, too. When calculating the VARHAC estimator the lag choice of the Schwarz Information Criterion is on average 2 lags for the FNI, 5 lags for the FESX and 2 lags for the FSP. This result is in accordance with the choice of 5-minute intervals for the calculation of the realized volatilities.

Following the example of Andersen *et al.* (2006) we use the log of the realized volatilities σ_t in our estimation. Again, the dataset contains 1,019 observations and days with no trade in all but at least one market are assigned a volatility of zero in the closed markets. Andersen *et al.* (2001) show in an empirical study that the usage of $\ln(\sigma)$ should bring along approximate normality which allows for the straight-forward application of standard estimation techniques. Standard tests for normality, however, are on the edge of rejection of the hypothesis that the data are indeed normally distributed in our case. Tables 2.2 and 2.3 provide the skewness and kurtosis measures along with the Jarque-Bera test statistics and p-values. It should be noted that negative values of the mean, median and minimum are possible due to taking the logarithm of the volatility measure. Further, the modeling of log-volatilities guarantees that forecasts of the realized volatility are positive. The convention to assign a value of zero to a closed market when at least one market is open is carried over to the log-volatilities, too.

The lower part of Tables 2.2 and 2.3 present again sample correlations of the $\sigma_{FNI,t}$, $\sigma_{FESX,t}$ and $\sigma_{FSP,t}$ with their contemporaneous and lagged values. For both volatility measures they are substantially higher than in the case of the returns which suggests already that the interdependence of the volatilities might be more pronounced than dependence among the returns.

2.4 Empirical Results

In the following section we restrict ourselves to the presentation of the estimation results based on one variable ordering only, namely when the day t starts at the opening of the Singapore Exchange, that is at 23:55 GMT. In this case, the variable ordering is FNI - FESX - FSP. This specification is, however, arbitrary. The estimation has therefore been performed with the two other possible orderings (FESX - FSP - FNI, i.e. starting the day when the European markets open, and FSP - FNI - FESX, i.e. starting the day when the New York market opens) as well. The ordering imposes restrictions on the matrix $\boldsymbol{B}_0$ of contemporanous effects. The estimation results proved to be robust to the variable ordering. Neither the coefficient estimates nor the subsequent impulse response analysis and variance decomposition differ qualitatively. Results based on ordering the FESX or FSP first are available from the authors upon request.

In order to evaluate the stability of our results with respect to time, the sample has been split into half and the estimation has been conducted on both subsamples. The estimated parameters changed slightly in absolute value. All in all, the implications deducted from the estimates do not change. The signs of the estimated parameters still point in the same directions despite some of those coefficients which are not significant. So a static model is an appropriate approach to model the time period at hand.

2.4.1 Modeling Daily Returns

The return model is estimated with $p = 1$ lag as suggested by information criteria. As we can rely on approximate normality of the error term in the model we perform a simple parametric bootstrap (see, for example, MacKinnon, 2006) to calculate the standard errors of the parameter estimators. The estimation results are presented in panel 1 of Table 2.4.

The first striking result is the negative and significant estimate for $b_{11,1}$, that is, the influence of the previous day FSP on the FNI. The estimated coefficient of -0.1360 is also quite high and would imply that, on average, if CME closes with a high return, the following trading at SGX realizes a substantially negative return. This result is consistent, however, with the sample correlations presented in Table 2.1. The same negative influence, albeit to a lesser extent, is found for the influence of the FSP on the FESX.

The other results are more in line with expected findings. The influence of the FNI-trade on day t on the return of the subsequent trading of the FESX is positive and significant. The same is true for the influence of the Eurex morning return and the influence of the FNI return on the FSP which are positive, yet not statistically significant. We also find for all three index futures that the influence of the market which precedes directly is greater in magnitude than the influence of the market which is again one step further away. For example the influence of the FESX return on the FSP return ($b_{32,0} = 0.0484$) is greater than the influence of the FNI-return ($b_{31,0} = 0.0185$) which preceded the trading in Europe.

Regarding the signs of the estimates, our results also support the often documented characteristic of negative autocorrelation in return series. The coefficients on the own lag-return of the FNI, FESX and FSP (that is, the coefficients on the main diagonal of $\boldsymbol{B}_1$) are all negative. As regards statistical significance, however, only the $b_{22,1}$ element is significantly different from zero.

The hypothesis that financial markets are efficient and that, thus, there is no influence of trading which lies more than 24 hours back in time is supported by our results. The lower diagonal elements in the $\boldsymbol{B}_1$-matrix are both small in absolute value and not statistically significant.

Consider once again the sign and the absolute value from the perspective of the individual markets. Japan's daytime return is most susceptible to foreign information. This is not only true for the immediately preceding trading in the United States, but also, albeit to a lower extent, for the trading in the European morning which lies 7 hours 45 minutes (and still 5 hours 45 minutes after 20th November 2005) further back than trading at the CME. Moving to Europe, we find a positive significant mean spillover from the same day trading in Japan and a negative and significant mean spillover from the previous trading day in the United States. The overall magnitude as measured by $b_{21,0}$ and $b_{23,1}$ is slightly smaller than for the Japanese market. So the European market seems to be less susceptible to foreign information than the Japanese market. Moving on to the USA, the market there seems to have a very self-sufficient position. There are no (significant) spillovers neither from Europe nor from Japan which would affect US trading.

Impulse response analysis also suggests that markets are efficient. Panels 1-3 in Figure 2.2 show that a shock in one market is indeed perceptible in the subsequent markets, but that its influence dies out quickly. It is usually already the second trading day after the shock where that specific shock is not perceptible any more. The size of the impact of an innovation shock follows the suggested timing structure in two of the three possible cases. Consider panel 1 which presents a shock to the FNI-return in Singapore. Clearly, the reaction is most important for the own return. But then we find the influence dying out through the day, meaning that the reaction of Eurex is more intense than the reaction of the CME. The second panel considers a shock in the morning trade of the FESX. As can be seen the impact on the trading in the USA is lower than the impact on the trading at the SGX which is contrary to what we would have expected. The last panel presents the reaction on a shock in the USA. Again, the impact is greatest on the own return, followed by the impact on the return of the FNI which is traded subsequently. However, as we neglect that the FSP and FESX are traded simultaneously for at least 4.5 hours (the afternoon trading period in Germany), the influence of the American market on the European market might be understated.

The fact that the effect of an innovation shock in one market on day t dies out quickly would also be supported by the cumulative impulse response functions (which are not printed). The reason is that already from $t+1$ to $t+2$ the difference is almost not perceptible any more.

So what we conclude from our analysis is the following. We find small, diminishing and short lived mean spillover effects from the USA and Europe to Japan and from Japan and the USA to Europe in the chronological ordering as expected. The US market turns out to be robust against return spillovers.

2.4.2 Volatility Modeling

The VAR models for the two different volatility measures (Equations (5.8) and (2.5)) are estimated with $p = 4$ lags as suggested by information criteria. As heteroscedasticity is not an issue here (see also Andersen, Bollerslev, Christoffersen, and Diebold, 2005) we use again a parametric bootstrap (see, for example, MacKinnon, 2006) to derive the standard errors. Subsequently, the acronyms ABDL-model and BI-model will be used to refer to the SVAR model based on the realized volatility measure of Andersen *et al.* (2003) and the daily volatility measure of Bollen and Inder (2002), respectively.

It turns out that the estimation results from the ABDL- and the BI-model are not qualitatively different. We conclude from this finding that both measures efficiently account for possible microstructure effects and that our results are robust with respect to the way the volatility series is computed. We therefore restrict ourselves to the presentation and discussion of the results based on the realized volatilities used in the ABDL-model and only highlight striking differences. The parameter estimates are presented in Table 2.5.

The estimation results of the ABDL-model suggest that volatility in one market immediately influences the volatility in the market which is open subsequently. This is reflected by the coefficients in the matrices $\boldsymbol{B}_0$ and $\boldsymbol{B}_1$: in $\boldsymbol{B}_0$ the lower diagonal elements are all positive and (besides the $b_{31,0}$-element) significantly different from zero. As regards the matrix $\boldsymbol{B}_1$, the upper diagonal elements are positive and statistically significant (on a 5 percent significance level) as well. So we conclude that there is a significant volatility spillover effect from one market to the next. The elements on the main diagonal of $\boldsymbol{B}_1$ are positive and statistically significant as well, whereas the elements below the main diagonal are not statistically significant. In the higher order lags only the elements on the main diagonal (with one exception in the matrix $\boldsymbol{B}_2$) are significant. In short our results indicate that there are spillovers from one market to the next which affect the volatility of the upstream market immediately. When looking more than 24 hours back in time, only the volatility in the home market exerts an effect on the respective volatility which supports the notion of volatility persistence.

Considering the relative sizes of the coefficient estimates we find that volatility in Europe is most influenced by the volatility in the two other markets. Also the chronological ordering is reversed as the influence of the US market volatility of the previous trading day is higher than the influence of the Japanese market's volatility which would precede directly. The same reversal is found for the Japanese market which is influenced to a greater extent by the European volatility than by the US volatility. As regards the US market, the chronological order is restored as the influence on its volatility stems in principle from Europe.

The same conclusions can be drawn from the estimation of the BI-model. The signs of the coefficients remain the same for all parameters that were significant in the ABDL-model. The size changes somewhat for the effects of USA and Europe on Japan which are almost equivalent in the BI-model (which is mainly due to a reduction in the European influence). Also the effect of the US volatility on European volatility is more pronounced. In the higher order lags we find the coefficients indicating the effect of Japan onto Europe to be significant for $p = 1$ and $p = 2$. This would imply that there is still an influence from the Japanese market on European volatility after more than 24 hours.

The impulse response functions presented in Figure 2.3 once again support the finding that volatility persists across a few trading days. In panels 1 and 2 which present a one unit shock to the FNI and FESX volatility, respectively, it can be seen that in the home market it takes longer until the effect of a shock dies out. Unfortunately this result is slightly corroborated by the third graph which depicts the reaction to a unit-shock in the FSP. It seems that the reaction of the FESX volatility is somewhat heavier than that of the home index future FSP volatility. All in all, the impulse response analysis suggests that there is volatility persistence as it takes on average 10 to 15 trading days until the impact of a volatility shock is not perceptible any more. This is supported by the cumulative impulse response functions which are presented in Figure 2.4.

2.4.3 Market Leadership

When comparing the results of the mean model and the volatility model we conclude that spillovers are more pronounced in the realized variance of the index futures than in their return itself. This is supported by the decomposition of the long-run variance as suggested by Hasbrouck (1991). Consider panel 2 in Tables 2.4 and 2.5: it turns out that in the long run, the return of a market is to roughly 99% determined by information events which happen in the home market. This is surprisingly also the case for events happening in the United States. As far as the ABDL realized volatilities are concerned it is only the Japanese market which seems quite self-sufficient: its own contribution amounts to 95.82%. If the variables are ordered differently the contribution of the home market even raises slightly up to 97.04% (when the day begins at the opening of Eurex, details not reported). However, there seem to be more important interlinkages between Europe and the United States, a result which one would probably expect due to the political and economic ties. The total variance in FSP-trading is caused to 9.75% by events in Europe. The contribution in the other direction amounts to 12.50%. The findings of the BI-model point in the same direction. The difference is that the US market seems more self-sufficient than the European market, but their ties are still remarkable.

This highlights how interwoven European and US financial markets are compared to their linkages with Asian markets. At the same time it clearly indicates a slight dominance of the US market.

2.5 Model Evaluation

An important aspect when deciding to model returns and volatilities separately instead of using, for example, a GARCH model, was the finding of Andersen *et al.* (2003) that forecasts based on realized volatility were more accurate than those based on other forecast methods. In order to check the joint forecasting ability of our models we also perform a simple forecast evaluation. We evaluate whether an out of sample return forecast based on the estimated SVAR models can compete with a univariate modeling approach forecasting the devolatized return and the realized volatility separately and compare these two to a univariate GARCH(1,1) model-based forecast as well as a forecast based on a univariate AR(1) model. Note that the evaluation is meant to compare a forecast of the log-returns, not the devolatized returns. We therefore undo the devolatization when using the multivariate and univariate models, i.e. we forecast the volatility and the standardized returns separately and combine the results according to Equation (2.2). In order to account for distributional aspects of the log-returns, the GARCH model as well as the univariate AR(1) model are estimated by maximum likelihood assuming t-distributed errors.

To evaluate the accuracy of the forecast we use the Mean Absolute Error (MAE), the Mean Absolute Percent Error (MAPE) and the Mean Percent Error (MPE) measures (e.g. Makridakis, Wheelwright, and Hyndman, 1998) which are defined as

$$MAE = \frac{1}{s}\sum_{t=1}^{s} |r_t - r_t^{\star}| \cdot 100, \tag{2.9}$$

$$MAPE = \frac{1}{s}\sum_{t=1}^{s} \left|\frac{r_t - r_t^{\star}}{r_t}\right| \cdot 100, \tag{2.10}$$

$$MPE = \frac{1}{s} \sum_{t=1}^{s} \frac{r_t - r_t^\star}{r_t} \cdot 100, \tag{2.11}$$

where s is the forecast horizon and $r_t^\star$ is the forecast of r_t.

The evaluation measures are reported in 2.6. Detailed estimation results of the different models are not reported, but are available from the authors upon request. What we find is that the multivariate model always performs better than any of the univariate models. To justify the usage of our estimation preceeding in contrast to the other approaches, consider the differences in MAPE of the one step ahead - forecast between these models. When modeling mean and volatility separately, the forecast of the FNI based on this approach is distinctly better (by almost 16 percentage points) than the forecast based on the GARCH-model and slightly better than the forecast based on the AR(1)-model. In case of the FESX forecast the model is only slightly worse (by 1.5 percentage points) than the GARCH model and performs better than the AR(1)-model. In case of the FSP, the univariate model and the GARCH model are nearly equivalent and perform both better than a univariate AR(1)-process. The picture remains the same for a two step ahead forecast. Note that the SGX was closed on the last day in the sample, so the forecast evaluation measures did not change.

To summarize the findings of the forecast evaluation, we clearly see two advantages in our modeling approach. First, the forecast based on the strategy of separate modeling of returns and variances pays off in terms of forecast accuracy. And second, by this approach we avoid the delicate issues arising when using a multivariate GARCH model within the context of a structural VAR approach, especially the issues concerning the identification of a structural GARCH process.

2.6 Application of the Spillover Model to the Stock Market Crash 14th and 15th January 2008

As the spillover model has been designed to consider the influence of previous markets on the actually open market, it is interesting to evaluate what it can tell about the stock market crash in January 2008. In consequence of the US mortage crisis which came about in summer 2007, the markets heavily reacted to information which accrued over the weekend 12th and 13th January 2008.

The avalanche started in the Asian markets where for example the Nikkei 225 lost 3.9% (calculated as close-to-close-return). It continued its way to Europe where the EuroStoxx 50 lost 7.3%. The US markets being closed on that Monday, there was no reaction so far. The downward movement continued on the following Tuesday in Asia and was slightly reversed in Europe (which was probably due to the announcement of the Federal Reserve Bank in the USA to lower interest rates by 75 basis points). The US market in the following was only slightly hit by the wave which the other markets had to stand the day before. The S&P 500 fell by only 1.1 percent which is far less than the other indices.

To evaluate whether our spillover model is capable of tracing these influence effects we use the model to predict what should have happened during the third week in January 2008 based on data of the preceding week. We forecast both the mean and the volatility model and combine the results according to Equation (2.2) to obtain the log-returns in which we are interested. This is a forecast only, the coefficient estimates are not updated. This means we have an almost four years estimation period where the markets were quite stable. Then there is a gap of one and a half years where the mortage crisis slowly built and finally the event period January 2008. We use daily open, high, low, and close data from the Nikkei 225, EuroStoxx 50 and S&P 500 indices as futures data are not readily available. These data are obtained from finance.yahoo.com. Further, due to the lack of availability of intraday data the volatility is measured by a simple range based estimator (Garman and Klass, 1980) as

$$\sigma^2_{GK,t} = 0.5(\log H_t - \log L_t)^2 - (2\log 2 - 1)(\log C_t - \log O_t), \tag{2.12}$$

where H_t is the day's high, L_t the day's low, and O_t and C_t are open and close prices, respectively. This approximation is motivated by the high correlation of this volatility measure and the realized volatility of Andersen *et al.* (2003). The European morning returns also have to be approximated by open-to-close returns. Although we are only interested in sign forecasts, we nevertheless compute the mean percent error (MPE) to evaluate total model performance. The MPE is given as

$$MPE = \frac{1}{s}\sum_{t=1}^{s}\frac{r_t - r_t^\star}{r_t}\cdot 100, \tag{2.13}$$

where s is the forecast horizon and $r_t^\star$ is the forecast of r_t.

Once we use our model to predict what should have happend during this third week of January 2008, the results are quite encouraging (cp table 2.7). For the week 14th to 18th of January the model is able to predict the correct sign of the returns in 11 out of 15 cases. A forecast based on the random walk assumption should, on the other hand, only deliver the right sign in about 50% of all cases. One case, where the market in the USA is closed on Monday, 14th January, the model has to fail as it has not been designed to explicitly account for holidays. As the model intends to predict effects of uptime markets on the following markets, an indication of the direction in which the market will develop is what we would expect the model to be able to tell us[1]. A prediction of the actual returns should not necessarily be accurate. It turns out that the deviation from the true returns ranges in between 0.0018 (prediction for the Nikkei225 on Tuesday, 15th January 2008) and 0.0857 (prediction for Europe on Monday, 14th January 2008, where also the predicted sign is incorrect). Looking at the model prediction as a whole the model seems quite able to trace the effects of events in previous markets on the following markets. It may therefore support an investor in evaluating his/her gut feeling when it comes to judge rumors in international financial markets.

2.7 Concluding Remarks

Our paper contributes to the fast growing literature in empirical financial economics dedicated to the investigation of international financial market linkages. We propose a new modeling strategy to capture the short-run daytime spillover dynamics of the main financial centres around the globe. Specifically, we employ structural vectorautoregressive models for the mean and the volatilities of the daytime returns which draw their structure from the natural, chronological ordering of the trading in the three markets (Europe, USA and Japan) used in our study. This allows us to provide impulse response and variance decomposition analysis as well as Granger-type causality testing within this well established framework.

For the mean system we find only short lived significant spillovers on Japan and

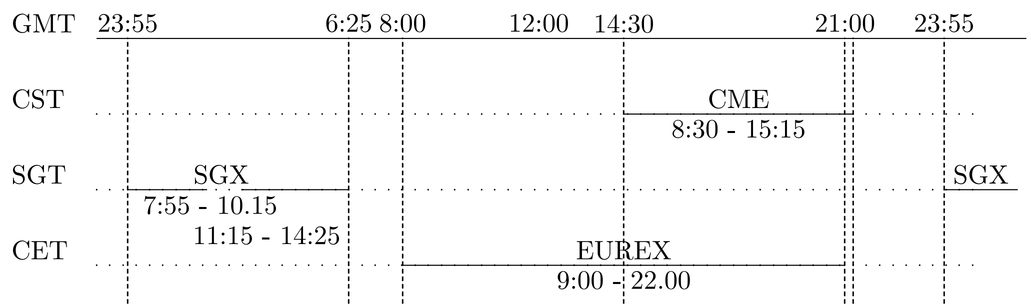
[1] See also Christoffersen and Diebold (2006) who state, inter alia, that "Short-run return forecasting (...) is (...) difficult, and perhaps even impossible. (...) There is substantial evidence that sign forecasting can often be done with surprising success."

Europe, albeit in a small order of magnitude. It turns out that the Japanese market is the most susceptive to foreign information, originating both from Europe and the United States. The European market, on the other hand, only reacts to information spilling over from the Japanese market. This indicates that, while the US and European markets are closed, the markets in Asia efficiently process information which then spill over to Europe, the market which opens first after Asian markets close. The US market, however, seems to have a particular position in that we do not find spillovers neither from Europe nor from Japan to the USA.

As regards volatility spillovers, we find that all markets react more intensely to the volatility of the previous market than in the case of the return spillovers. The effect originating in foreign markets dies out within one trading day, the influence of the home market is persistent, however, across four lags. In contrast to the findings of the mean model the timing seems to be less important for volatility spillovers as it is not always the market which was open before which exerts the greatest influence. Our findings are robust with respect to the way the volatility series is computed.

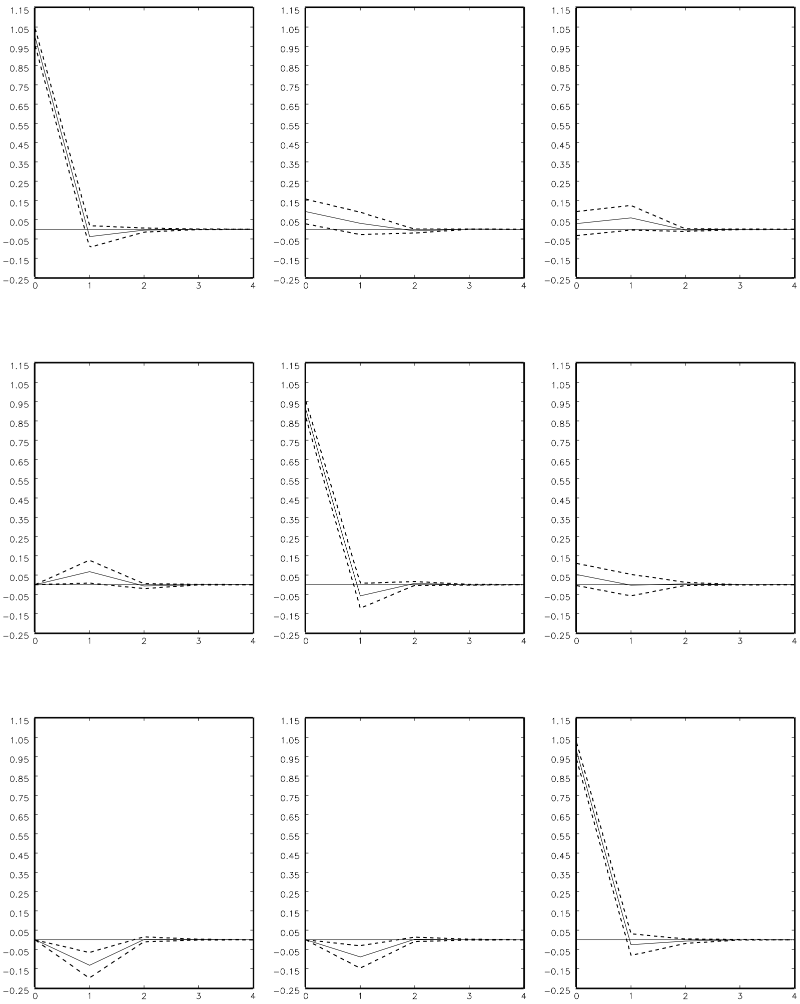
The estimated dynamical systems can ultimately be employed to trace and forecast the impact of a shock in one of the worlds leading markets on the other markets as well as to perform a forecast of the returns in the markets. We find that the contribution of the separate modeling approach in the multivariate context is threefold. First, the multivariate structure allows for a more accurate forecast of the return series than a univariate approach. Second, the (univariate) separation of returns and volatilities and their detached forecast turns out to perform on average better than a univariate forecast based on a GARCH-model or an AR-model. And finally, the application of structural VARs is econometrically better manageable than the usage of multivariate GARCH models within this structural context. The application to the recent financial crisis which has been triggered by the US house crisis also shows encouraging results. The model thus seems able to trace the linkages between international stock markets and highlights once again the interdependence of global financial markets.

Figure 2.1: Trading Times



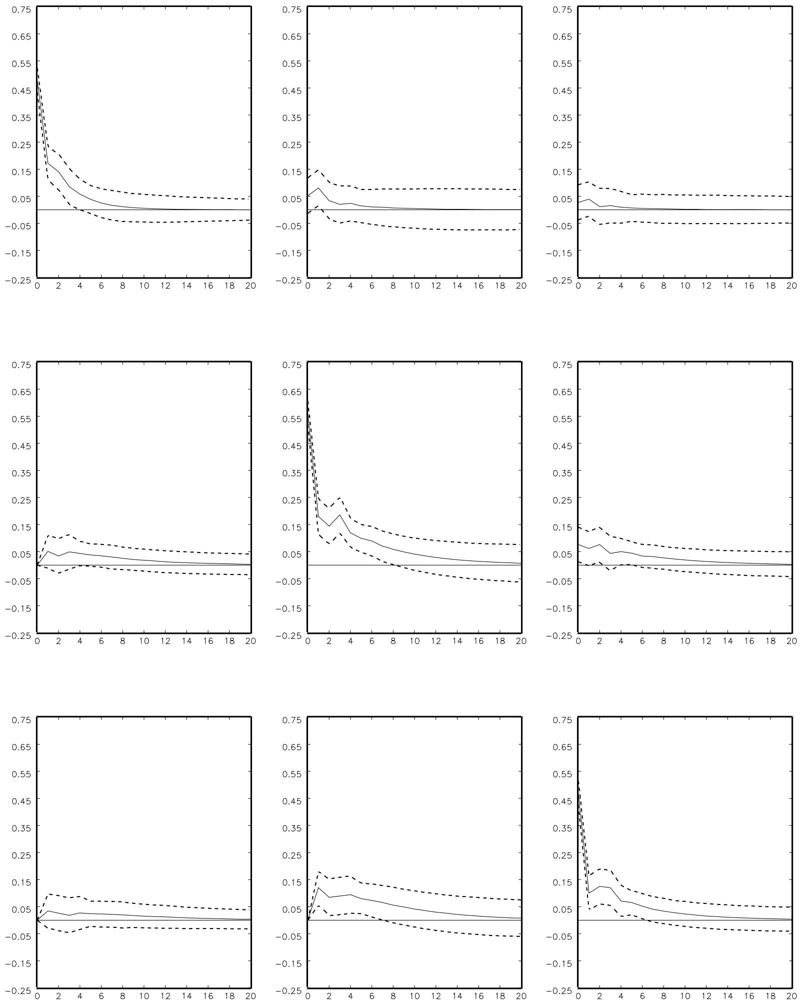
The figure presents the trading times at the different exchanges in Singapore (SGX), Frankfurt (Eurex) and Chicago (CME). GMT is Greenwich Mean Time, CST is Central Standard Time, SGT is Singapore Time and CET is Central European Time. Trading hours are presented as of 1 January 2006.

Figure 2.2: Return Model: Impulse Response



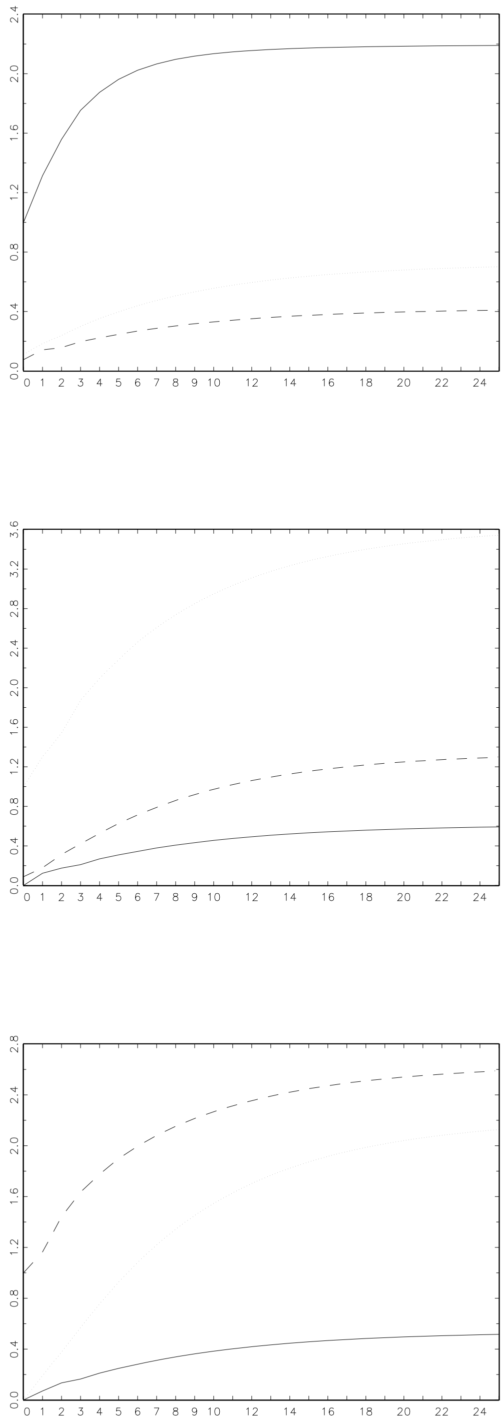
The graphs depict the response of the FNI (left column), FESX (middle column), and FSP returns (right column) to a one standard deviation shock in Singapore (first row), Europe (second row), or the USA (third row), respectively. The dashed lines are two standard error bounds.

Figure 2.3: ABDL Volatility Model: Impulse Response



The graphs depict the response of the FNI (left column), FESX (middle column), and FSP volatilities (right column) to a one standard deviation shock in Singapore (first row), Europe (second row), or the USA (third row), respectively. The dashed lines are two standard error bounds.

Figure 2.4: ABDL Volatility Model: Cumulative Impulse Response



The graphs depict the cumulative impulse response of the FNI (solid line), FESX (dotted line), and FSP volatilities (dashed line) to a one-unit shock in Singapore (panel 1), Europe (panel 2), or the USA (panel 3), respectively.

Table 2.1: Descriptive Statistics of log-Returns

	FNI	FESX	FSP
Mean	0.0108	0.0050	0.0449
Median	0.0000	0.0000	0.0184
Maximum	2.8399	2.5500	2.5358
Minimum	−3.0747	−2.5755	−2.9090
Variance	1.0168	0.8509	0.9693
Skewness	−0.0584	−0.0455	−0.1383
Kurtosis	2.8027	2.4232	2.7461
Jarque-Bera	2.2315	14.4744	5.9862
	(0.3277)	(0.0007)	(0.0501)
Sample Correlations			
FNI_t	1.0000		
$FESX_t$	0.1064	1.0000	
FSP_t	0.0257	0.0604	1.0000
FNI_{t-1}	−0.0372	0.0341	0.0620
$FESX_{t-1}$	0.0634	−0.0593	0.0044
FSP_{t-1}	−0.1290	−0.0991	−0.0244

The table provides descriptive statistics for the devolatized log-returns of the Dow Jones Euro Stoxx 50 future, the S&P500 future and the Nikkei 225 future. Note that the return for the FESX is calculated as open-to-1330. The Jarque-Bera test for normality is presented together with p-values which are given in parentheses.

Table 2.2: Descriptive Statistics of log-Volatilities (ABDL)

	FNI	FESX	FSP
Mean	−0.3675	−1.1244	−0.6467
Median	−0.2794	−1.3877	−0.7613
Maximum	2.2171	3.1176	2.7808
Minimum	−2.8674	−4.4321	−3.5935
Variance	0.5967	1.4915	0.7677
Skewness	−0.2580	0.5371	0.5771
Kurtosis	2.9952	2.6892	3.5959
Jarque-Bera	11.3019	53.1023	71.6503
	(0.0035)	(< 0.0001)	(< 0.0001)
Sample Correlations			
$\sigma_{FNI,t}$	1.0000		
$\sigma_{FESX,t}$	0.5175	1.0000	
$\sigma_{FSP,t}$	0.4559	0.6953	1.0000
$\sigma_{FNI,t-1}$	0.6060	0.4934	0.4467
$\sigma_{FESX,t-1}$	0.5201	0.8036	0.6920
$\sigma_{FSP,t-1}$	0.4557	0.7126	0.6737

The table provides descriptive statistics of the daily volatility measure as proposed by Andersen *et al.* (2003) in logarithms of the Dow Jones Euro Stoxx 50 future, the S&P500 future and the Nikkei 225 future. The Jarque-Bera test for normality is presented together with p-values which are given in parentheses.

Table 2.3: Descriptive Statistics of log-Volatilities (BI)

	FNI	FESX	FSP
Mean	−0.3058	−1.0124	−0.5140
Median	−0.1826	−1.3039	−0.6532
Maximum	2.6407	3.4985	2.7650
Minimum	−2.9700	−3.9332	−2.8722
Variance	0.5246	1.4509	0.6964
Skewness	−0.3663	0.6047	0.6656
Kurtosis	3.4272	2.8087	3.6079
Jarque-Bera	30.5342	63.6552	90.9388
	(< 0.0001)	(< 0.0001)	(< 0.0001)
Sample Correlations			
$\sigma_{FNI,t}$	1.0000		
$\sigma_{FESX,t}$	0.5564	1.0000	
$\sigma_{FSP,t}$	0.4791	0.7432	1.0000
$\sigma_{FNI,t-1}$	0.6723	0.5517	0.4854
$\sigma_{FESX,t-1}$	0.5494	0.8268	0.7267
$\sigma_{FSP,t-1}$	0.4897	0.7528	0.7310

The table provides descriptive statistics of the daily volatility measure as proposed by Bollen and Inder (2002) in logarithms of the Dow Jones Euro Stoxx 50 future, the S&P500 future and the Nikkei 225 future. The Jarque-Bera test for normality is presented together with p-values which are given in parentheses.

Table 2.4: Mean Model

Panel 1: SVAR coefficient estimates			
	FNI	FESX	FSP
a	0.0186	0.0048	0.0477
	(0.5472)	(0.8772)	(0.1300)
$B_{FNI,0}$	0.0000	0.0000	0.0000
	(-)	(-)	(-)
$B_{FESX,0}$	0.1087	0.0000	0.0000
	(0.0006)	(-)	(-)
$B_{FSP,0}$	0.0185	0.0484	0.0000
	(0.5435)	(0.1240)	(-)
$B_{FNI,1}$	-0.0436	0.0810	-0.1360
	(0.1562)	(0.0097)	(<0.0001)
$B_{FESX,1}$	0.0418	-0.0668	-0.0730
	(0.1951)	(0.0296)	(0.0183)
$B_{FSP,1}$	0.0553	-0.0003	-0.0166
	(0.0867)	(0.9914)	(0.6039)
Panel 2: Long-run Variance Decomposition			
	FNI	FESX	FSP
FNI	0.9782	0.0045	0.0173
FESX	0.0110	0.9802	0.0087
FSP	0.0040	0.0025	0.9938

The table provides in panel 1 the structural VAR estimates for the mean model given in Equation (2.7) where the variables are ordered as FNI - FESX - FSP. P-values are given in parentheses. Panel 2 presents the long-run variance decomposition according to Hasbrouck (1991). It is to be read as the proportion in the forecast error variance in row i due to the variance in column j.

Table 2.5: ABDL Volatility Model

Panel 1: SVAR coefficient estimates			
	FNI	FESX	FSP
a	0.0026	-0.0986	-0.0027
	(0.9487)	(0.0187)	(0.9498)
$B_{FNI,0}$	0.0000	0.0000	0.0000
	(-)	(-)	(-)
$B_{FESX,0}$	0.1113	0.0000	0.0000
	(0.0005)	(-)	(-)
$B_{FSP,0}$	0.0568	0.0881	0.0000
	(0.0639)	(0.0052)	(-)
$B_{FNI,1}$	0.2960	0.1169	0.0705
	(< 0.0001)	(0.0002)	(0.0266)
$B_{FESX,1}$	-0.0072	0.2732	0.1864
	(0.8278)	(< 0.0001)	(< 0.0001)
$B_{FSP,1}$	0.0256	0.0433	0.1435
	(0.4480)	(0.1788)	(< 0.0001)
$B_{FNI,2}$	0.1402	-0.0268	0.0087
	(< 0.0001)	(0.3978)	(0.7852)
$B_{FESX,2}$	-0.0233	0.1264	0.0942
	(0.4846)	(0.0001)	(0.0040)
$B_{FSP,2}$	-0.0481	0.0615	0.2241
	(0.1532)	(0.0707)	(< 0.0001)
$B_{FNI,3}$	0.0783	-0.0243	-0.0356
	(0.0191)	(0.4761)	(0.2751)
$B_{FESX,3}$	-0.0090	0.1912	0.0490
	(0.7725)	(< 0.0001)	(0.1370)
$B_{FSP,3}$	0.0019	0.0064	0.0783
	(0.9554)	(0.8458)	(0.0170)
$B_{FNI,4}$	0.1807	0.0127	0.0002
	(< 0.0001)	(0.6826)	(0.9954)
$B_{FESX,4}$	0.0437	0.1006	-0.0042
	(0.1890)	(0.0010)	(0.8960)
$B_{FSP,4}$	0.0042	0.0437	0.1076
	(0.8955)	(0.1545)	(0.0007)
Panel 2: Long-run Variance Decomposition			
	FNI	FESX	FSP
FNI	0.9582	0.0290	0.0137
FESX	0.0173	0.8584	0.1250
FSP	0.0120	0.0975	0.8922

The table provides in panel 1 the structural VAR estimates for the volatility model given in Equation (2.7) where the volatilities are calculated as proposed by Andersen *et al.* (2003) and are ordered as FNI - FESX - FSP. P-values are given in parentheses. Panel 2 presents the long-run variance decomposition according to Hasbrouck (1991). It is to be read as the proportion in the forecast error variance in row i due to the variance in column j.

Table 2.6: Out of sample Forecast Evaluation

Panel 1: one step ahead forecast					
		Mulitvariate Model	Univariate Model	Univariate GARCH(1,1)	Univariate AR(1)
MAE	FNI	0.8274	0.8915	0.9534	0.8938
	FESX	0.8072	0.8508	0.8364	0.8791
	FSP	1.4424	1.5272	1.5254	1.5470
MAPE	FNI	105.1314	113.2727	121.1348	113.5627
	FESX	98.2613	103.5717	101.8124	107.0138
	FSP	96.2996	101.9589	101.8413	103.2784
MPE	FNI	105.1314	113.2727	121.1348	113.5627
	FESX	98.2613	103.5717	101.8124	107.0138
	FSP	96.2996	101.9589	101.8413	103.2784
Panel 2: two steps ahead forecast					
		Mulitvariate Model	Univariate Model	Univariate GARCH	Univariate AR(1)
MAE	FNI	0.8274	0.8915	0.9534	0.8926
	FESX	1.2386	1.2649	1.2587	1.2790
	FSP	1.0578	1.1184	1.1040	1.1042
MAPE	FNI	105.1314	113.2727	121.1348	113.4141
	FESX	98.8679	101.7887	100.9719	103.5083
	FSP	95.5841	100.9861	99.0180	98.2548
MPE	FNI	105.1314	113.2727	121.1348	113.4141
	FESX	98.8679	101.7887	100.9719	103.5083
	FSP	95.5841	100.9861	99.0180	98.2548

The table provides the out of sample forecast evaluation comparison for the separate VAR models for mean and volatility (Multivariate Model), their univariate counterpart (Univariate Model), a univariate GARCH(1,1) model with t-distributed errors (Univariate GARCH(1,1)) and a univariate AR(1) model with t-distributed errors (Univariate AR(1)). Panel 1 contains the evaluation of the one step ahead forecast while panel to contains the two steps ahead forecast. MAE is the mean absolute error, MAPE is the mean absolute percent error and MPE is the mean percent error as defined in section 2.5.

Table 2.7: Forecast of Third Week in January 2008

	Nikkei 225	EuroStoxx 50	S&P 500
Panel 1: 14th January 2008			
true	-0.0278	-0.0742	0.0000
predicted	-0.0073	0.0114	0.0206
MPE	73.5694	115.4070	–
Panel 2: 15th January 2008			
true	-0.0430	0.0170	-0.0019
predicted	-0.1120	0.0521	-0.0001
MPE	-160.6299	-205.8090	97.2642
Panel 3: 16th January 2008			
true	0.0057	-0.0544	0.0213
predicted	0.0366	-0.0106	0.0002
MPE	-543.9730	80.5720	99.0609
Panel 4: 17th January 2008			
true	0.0108	0.0554	0.0089
predicted	-0.0592	0.0108	0.0111
MPE	649.4364	80.5307	-25.3950
Panel 5: 18th January 2008			
true	0.0276	-0.0124	-0.0199
predicted	0.0988	-0.0422	0.0106
MPE	-258.5521	-239.8869	153.2014

The table provides the one step ahead forecasts of the open-to-close-returns of Nikkei 225, EuroStoxx 50 and S&P 500 for the week 14th to 18th January 2008. MPE is the mean percent error as defined in Equation (2.13).

Chapter 3

A Note on the Influence of Heteroscedasticity on the Johansen Cointegration Test

3.1 Introduction

Quite a number of empirical studies in the financial literature use the cointegration framework to explain long-term relationships between asset prices, market indices, interest rates or currencies (Barassi, Caporale, and Hall, 2005; Haug, MacKinnon, and Michelis, 2000; Masih and Masih, 2004, and various others). Obviously, the results crucially hinge on the reliability of cointegration tests. The workhorse in empirical finance nowadays is the Johansen (1988, 1991) methodology to test and estimate cointegrated systems. Financial data quite often, however, violate the assumptions (normality, homoscedasticity; see, for example, Tsay, 2005, ch.3) which were made to derive the tests. Heteroscedasticity is probably the most prominent feature which is still often neglected, albeit some recent theoretical research provides possibilities to account for it (e.g. Wong, Li, and Ling, 2005). This note therefore will address the question whether the identification of cointegration hinges on time varying volatility.

This question is twofold. On the one hand it means whether the presence of heteroscedasticity impedes the detection of a cointegration relationship if it does indeed exist. There is evidence that the traditional cointegration tests perform well under certain conditions. In particular, in the studies of Lee and Tse (1996) or Mantalos (2001) heteroscedasticity is only an issue in the innovations' variance. However, there is also a branch of literature that tries to explicitly account for the presence of heteroscedasticity in the unit root behavior when testing for cointegration, for example McCabe, Leybourne, and Harris (2006). On the other hand, if cointegration is indeed not given, a cointegration test should also be powerful enough as to indicate its absence. Lee and Tse (1996) show that the way volatility is modeled influences the performance of the Johansen cointegration test to some extent and, thus, its reliability.

Although there are more cointegration tests than just the likelihood ratio tests of Johansen (1991)[1], these are the most widely used ones in the recent empirical literature. The system maximum likelihood estimator also provides asymptotically efficient estimates of the cointegrating vector(s) and the adjustment coefficients. Further, Seo (2007) shows (both theoretically and by means of a simulation study) that the maximum likelihood estimator is far more efficient than OLS-based estimation in the context of error correction models with conditional heteroscedasticity. The focus of this note lies on the evaluation of the trace and the maximum eigenvalue test. We use two different cointegration concepts—stationary cointegration in the sense of Engle and Granger (1987) and stochastic cointegration in the sense of Harris, McCabe, and Leybourne (2002)—and different data simulation models to investigate the reliability of the Johansen testing framework under various heteroscedasticity and correlation assumptions. Note that this study is not designed like a typical size and power study where one would calibrate the size of a test under the null hypothesis and then investigate the power under the alternative hypothesis. Here we take the critical values as given, i.e. there is no size adjustment. Our focus lies on the evaluation of the tests if a cointegration test is conducted without taking particular features of financial data into account. In their respective studies Toda (1995) and Haug (1996) also proceed in this way.

We continue as follows. The next section establishes the theoretical cointegration framework and presents Johansen's trace and maximum eigenvalue tests for cointegration. Section 3.3 then describes the different data generating processes. Simulation results are presented in section 3.4. Section 3.5 concludes.

3.2 Cointegration Models and Tests

Before conducting the simulation experiment we briefly establish the general cointegration framework, differentiating between stationary cointegration as introduced by Engle and Granger (1987) and stochastic cointegration which

[1] e.g. residual based tests (Engle and Granger, 1987; Hansen, 1990), tests based on principal components (Harris, 1997) and a number of system tests (Saikkonen, 1992, and others)

has recently been brought forward by Harris *et al.* (2002). The first subsection describes these two different cointegration models. The second subsection then considers the Johansen method to test for cointegration.

3.2.1 Model Framework

An important aspect of two (or more) $I(1)$ variables is that there may exist a stationary, linear combination of these variables. In this case the variables are cointegrated $CI(1,1)$ in the sense of Engle and Granger (1987). The relationship is called stationary cointegration as it requires the combination of the $I(1)$ variables to be strictly stationary. For illustration, consider the VAR(p) model

$$\mathbf{y}_t = \sum_{i=1}^{p} \boldsymbol{A}_i \boldsymbol{y}_{t-i} + \boldsymbol{u}_t \tag{3.1}$$

with $\mathbf{y}_t$ a vector of n $I(1)$ variables, $\boldsymbol{A}_j$ are $(n \times n)$-matrices of parameters and $\boldsymbol{u}_t$ an n-vector of Gaussian errors. If the $I(1)$ variables in $\mathbf{y}_t$ are cointegrated, then by the Granger Representation Theorem (Engle and Granger, 1987) the VAR model in (3.1) can be written in the form of a vector error correction model (VECM)

$$\Delta\mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \sum_{i=i}^{p-1} \boldsymbol{B}_i \Delta\mathbf{y}_{t-i} + \boldsymbol{u}_t \tag{3.2}$$

where $\boldsymbol{\alpha}\boldsymbol{\beta}' = -I + \sum_{i=1}^{p} \boldsymbol{A}_i$ and $\boldsymbol{B}_j = -\sum_{i=j}^{k} \boldsymbol{A}_i$. The characteristic feature of this model is that the VAR in first differences still contains the level $\mathbf{y}_t$.

If the assumption that the variables have a constant unit root is relaxed this is the context of stochastic unit root processes and stochastic cointegration. The notion of stochastic unit root processes has been introduced by Granger and Swanson (1997) and then elaborated in the context of cointegration by Harris *et al.* (2002). Stochastically integrated processes are characterized by a non-constant unit root which is stochastic and varies around unity over time. Such a process can be stationary for some periods and then be mildly explosive for others. In the context of cointegration, the linear combination of stochastically integrated variables will not be strictly stationary any more. Stochastic cointegration (as opposed to stationary cointegration) only requires the absence of

$I(1)$ behavior. McCabe *et al.* (2006) define the following model:

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu} + \boldsymbol{\Pi}\boldsymbol{w}_t + \boldsymbol{u}_t + \mathbf{V}_t\boldsymbol{h}_t \\
\boldsymbol{w}_t &= \boldsymbol{w}_{t-1} + \boldsymbol{\eta}_t \\
\boldsymbol{h}_t &= \boldsymbol{h}_{t-1} + \boldsymbol{v}_t
\end{aligned} \tag{3.3}$$

where $\boldsymbol{u}_t$, $\boldsymbol{\eta}_t$, $\boldsymbol{v}_t$ and $\mathbf{V}_t$ are mean zero stationary processes (which may be correlated), $\mathbf{w}_t$ and $\boldsymbol{h}_t$ are vectors of integrated processes with $\mathbf{w}_0 = \boldsymbol{\eta}_0$ and $\boldsymbol{h}_0 = \boldsymbol{v}_0$. The characteristic feature of this model is the presence of the random term $\mathbf{V}_t\boldsymbol{h}_t$ which causes non-linear shocks in the data generating process of $\mathbf{y}_t$ which, thus, consists of a constant, an integrated process, and a shock term containing additively a linear and a non-linear component. $\mathbf{V}_t\mathbf{h}_t$ is heteroscedastic as it depends on $\mathbf{h}_t$ which is an integrated process. In contrast to an $I(1)$ series, $\Delta\mathbf{y}_t$ is not stationary as it still contains the level $\mathbf{w}_{t-1}$. To illustrate the behavior of the individual time series, consider the i-th element of $\mathbf{y}_t$

$$y_{i,t} = \mu_i + \boldsymbol{\pi}_i'\mathbf{w}_t + u_{i,t} + \mathbf{v}_{i,t}'\mathbf{h}_t \; , \tag{3.4}$$

where $\boldsymbol{\pi}_i'$ and $\mathbf{v}_{i,t}'$ are the i-th row of $\boldsymbol{\Pi}$ and $\mathbf{V}_t$, respectively. If $\boldsymbol{\pi}_i' \neq 0$, $p_{i,t}$ is said to be stochastically integrated. If, in addition, $\mathbb{E}[\mathbf{v}_{i,t}'\mathbf{v}_{i,t}] > 0$, $p_{i,t}$ is heteroscedastically integrated (HI). If, on the other hand, $\mathbf{v}_{i,t}' = \mathbf{0}$, $p_{i,t}$ is simply $I(1)$, so the variance of a change does not depend on t. Hence, the concept of stochastic integration covers both heteroscedastic integration and $I(1)$ behavior. When neglecting the trend term and assuming $\mathbf{v}_{i,t}' = \mathbf{0}$, the representation in (3.4) corresponds to a common stochastic trends representation which is similar to an individual element in Equation (3.1) above.

3.2.2 Johansen Cointegration Test

The Johansen (1988) method is based upon the full-information maximum likelihood estimation of the so-called reduced rank model[2]. Recall the VAR representation of the VECM in Equation (3.2). Under the hypothesis of r cointegration relations, $\boldsymbol{\beta}$ is an $(n \times r)$ matrix containing the r cointegration vectors and $\boldsymbol{\alpha}$ an $(n \times r)$ matrix of adjustment coefficients. In this case, only r distinct linear combinations of the level $\mathbf{y}_t$ appear in Equation (3.2).

[2]Refer to Hamilton (1994) or Lütkepohl (2005) for further details.

For notational simplicity let $\boldsymbol{Z}_t = \left(\Delta\mathbf{y}'_{t-1}, \ldots, \Delta\mathbf{y}'_{t-p+1}\right)'$. Conduct the reduced rank regressions

$$\begin{aligned}\Delta\mathbf{y}_t &= \boldsymbol{\xi}_0 + \boldsymbol{\Xi}\boldsymbol{Z}_t + \boldsymbol{u}_t \\ \mathbf{y}_t &= \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\boldsymbol{Z}_t + \boldsymbol{v}_t\end{aligned}$$

to obtain the residuals $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$. Next calculate their sample variance-covariance matrices as

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{uu}} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}'_t \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{vv}} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{v}}_t\hat{\boldsymbol{v}}'_t \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{uv}} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{v}}'_t = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{vu}}.\end{aligned}$$

Johansen (1988) shows that the maximum likelihood estimator of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is a function of these moments and that it can be found by choosing the eigenvalues $(\lambda_1, \ldots, \lambda_r)$ from the normalized eigenvalues solving the equation

$$|\ \lambda\hat{\boldsymbol{\Sigma}}_{\boldsymbol{vv}} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{vu}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{uu}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{uv}}| = 0$$

which are ordered $\lambda_1 > \lambda_2 > ... > \lambda_n$. Due to the necessary normalization finding the eigenvalues of the above expression and subsequently normalizing them is equivalent to finding the eigenvalues of the matrix

$$M = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{vv}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{vu}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{uu}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{uv}}.$$

The maximum value of the likelihood function under the assumption that there are r cointegration relationships is then given as

$$\begin{aligned}\mathcal{L}^{\star} &= -\frac{Tn}{2}\log(2\pi) - \frac{Tn}{2} - \frac{T}{2}\log|\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{uu}}\ | \\ &\quad -\frac{T}{2}\sum_{i=1}^{r}\log(1-\lambda_i). \qquad (3.5)\end{aligned}$$

Based on the likelihood in Equation (3.5), Johansen (1991) derives two likelihood ratio tests: the so-called maximum eigenvalue test and the trace test. The maximum eigenvalue ($\lambda_{\max}$) test determines under its null hypothesis whether the $(r+1)$th eigenvalue is still different from zero. The alternative hypothesis is that eigenvalues are only different from zero up to λ_r. If the null hypothesis can be rejected, λ_{r+1} as well as the remaining eigenvalues λ_{r+2} to λ_n which are smaller than λ_{r+1} can be considered to be zero. The test statistic is given by

$$-T \log(1 - \lambda_{r+1}) \ .$$

The test, thus, examines the hypothesis of $r+1$ cointegrating vectors against the alternative of (at most) r cointegrating vectors. Usually the test is formulated in terms of the rank of the cointegration matrix $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$. If $\mathbf{\Pi}$ has rank 0 there are n unit roots in the VAR and zero cointegration vectors. If in contrast $\mathbf{\Pi}$ has full rank n there are no unit roots and the data were stationary in the first place. For the maximum eigenvalue test this means that we check the null hypothesis of rank $r+1$, $H_0 : \mathrm{rk}(\mathbf{\Pi}) = r+1$, against the alternative that the rank is smaller than or equal to $r+1$, $H_1 : \mathrm{rk}(\mathbf{\Pi}) \leq r$.

The trace test considers the null hypothesis that $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is of rank r against the alternative of an unrestricted model where $\mathbf{\Pi}$ has full rank n. The test statistic (usually referred to as 'trace statistic', see Johansen and Juselius, 1992) is given as

$$-T \sum_{i=r+1}^{n} \log(1 - \lambda_i).$$

Note that under the null hypothesis of no cointegration ($r = 0$) the eigenvalues converge to zero. If all the eigenvalues indeed are zero, there are n unit roots in the VAR in Equation (3.2). Critical values for both tests have been obtained by means of Monte Carlo Simulations by, for example, Osterwald-Lenum (1992) and MacKinnon, Haug, and Michelis (1999).

In order to determine the number of cointegrating vectors, especially in absence of any *a priori* knowledge about their number, Johansen (1992) suggests to use a general to specific approach using the trace test to avoid underestimation of the number of cointegrating vectors. More precisely, one would start with the null hypothesis that $r = 0$, i.e. that there are zero cointegrating vectors. If this hypothesis is rejected, the next null hypothesis to be tested is $r = 1$. Upon rejection of the null hypothesis, a new one is formed until $r = n$. The first non-rejection of a null hypothesis $r = i$ $(i = 0, \ldots, n)$ calls the procedure to a halt and indicates that there are i cointegrating vectors. In case that the trace test suggests $r = n$ the time series are stationary.

3.3 Simulation Design

The simulation considers different data generating processes which are related to estimations and simulations previously conducted in the literature.

3.3.1 A Bivariate Model

The first model considered is inspired by the simple microstructure model given by Hasbrouck (1995) in his introduction. The variance of an error term here, however, is not necessarily constant over time but may follow a GARCH(1,1) process. In order to investigate the influence of heteroscedasticity on the cointegration test, we need a model which is cointegrated as well as one which is not cointegrated. Both of these models will have the same error terms. The mean equation of the cointegrated model reads as follows:

$$\begin{aligned} x_t &= x_{t-1} + \sigma_{1,t} u_{1,t} \\ y_t &= x_{t-2} + \sigma_{2,t} u_{2,t} \end{aligned} \tag{3.6}$$

where x and y are cointegrated processes, so the difference $x_t - y_t = \sigma_{1,t} u_{1,t} + \sigma_{1,t-1} u_{1,t-1} - \sigma_{2,t} u_{2,t}$ is a stationary process as long as the error processes are stationary. We compare this model with a not cointegrated version where y_t does not depend on x_{t-2} but on its own lagged term:

$$\begin{aligned} x_t &= x_{t-1} + \sigma_{1,t} u_{1,t} \\ y_t &= y_{t-1} + \sigma_{2,t} u_{2,t} \ . \end{aligned} \tag{3.7}$$

In both cases the innovations' variance may follow separate GARCH(1,1) processes which are given as

$$\begin{aligned} \sigma_{1,t}^2 &= a_{1,0} + a_{1,1} u_{1,t-1}^2 + a_{1,2} \sigma_{1,t-1}^2 \\ \sigma_{2,t}^2 &= a_{2,0} + a_{2,1} u_{2,t-1}^2 + a_{2,2} \sigma_{2,t-1}^2 \ . \end{aligned} \tag{3.8}$$

$a_{1,1} + a_{1,2}$ and $a_{2,1} + a_{2,2}$ are restricted to be lesser than 1 in order to assure a stationary GARCH process. u_1 and u_2 are independent white noise processes. We consider three different settings for the GARCH process in Equation (3.8). First, let $a_{1,0} = a_{2,0} = 1$ and the remaining $a_{i,j} = 0$ in which case the errors

u_i are homoscedastic. Second, both x and y exhibit heteroscedastic errors by setting $a_{1,0} = a_{2,0} = 0.1$, $a_{1,1} = 0.04$, $a_{1,2} = 0.94$, $a_{2,1} = 0.05$ and $a_{2,2} = 0.93$. And third, only x exhibits heteroscedastic errors while the errors in y are homoscedastic using the same parameterization as in the previous cases.

3.3.2 VAR-GARCH

The second data generating process is a multivariate vectorautoregressive model with possibly heteroscedastic errors. The mean model is simulated via

$$\boldsymbol{z}_t = \boldsymbol{a}_0 + \boldsymbol{A}\boldsymbol{z}_{t-1} + \boldsymbol{u}_t \tag{3.9}$$

if the variables are not to be cointegrated. If the variables in $\boldsymbol{z}_t$ are to be cointegrated, we simulate the VECM form

$$\Delta\boldsymbol{z}_t = \sum_{i=1}^{2} \boldsymbol{\Gamma}_i \Delta\boldsymbol{z}_{t-i} + \boldsymbol{\Pi}\boldsymbol{z}_{t-1} + \boldsymbol{u}_t \; . \tag{3.10}$$

The errors $\boldsymbol{u}_t$ are assumed to be multivariate normally distributed with mean zero and variance $\boldsymbol{\Sigma}_t$ which is specified as a BEKK model of Engle and Kroner (1995):

$$\boldsymbol{\Sigma}_t = \boldsymbol{C}'\boldsymbol{C} + \boldsymbol{F}'\boldsymbol{u}_{t-i}\boldsymbol{u}_{t-i}'\boldsymbol{F} + \boldsymbol{G}'\boldsymbol{\Sigma}_{t-j}\boldsymbol{G}. \tag{3.11}$$

The models are implemented as follows. For the specification where the data are not cointegrated the unrestricted VAR(1) in Equation (3.9) is simulated with $\boldsymbol{A}_1$ as the identity matrix and the constant $a_0 = \begin{pmatrix} 0.001 & 0.006 & 0.002 \end{pmatrix}'$. The system, thus consists of three independent random walks. The innovations may be correlated (see below). In order to simulate cointegrated data we implement the VECM in Equation (3.10) with

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = \begin{pmatrix} 0.4 & 0.2 & 0.4 \end{pmatrix} \begin{pmatrix} 1 & -0.8 & -0.6 \end{pmatrix}'$$

$$\Gamma_1 = \begin{pmatrix} 0.1676 & -0.020 & 0.022 \\ 0.322 & 0.003 & -0.116 \\ 0.151 & 0.116 & 0.420 \end{pmatrix} \quad \Gamma_2 = \begin{pmatrix} 0.001 & 0 & 0 \\ 0 & 0.006 & 0 \\ 0 & 0 & 0.002; \end{pmatrix}.$$

The rank of $\boldsymbol{\Pi}$ is one. The two models (3.9) and (3.10) are not directly comparable in the sense that they specify the same unit root processes. For the purpose of this study, however, it is only important that we can distinguish a setting with cointegration and a setting without cointegration.

We then implement the BEKK model in Equation (3.11) in the following way. First, to check implementation, the matrix $\boldsymbol{C}$ is the identity matrix and $\boldsymbol{F}$ and $\boldsymbol{G}$ are zero which results in uncorrelated, homoscedastic innovations in the VAR. Second, the rows of $\boldsymbol{C}$ are specified as

$$\begin{pmatrix} 1 & 0 & 0; & 0.2 & 1 & 0; & 0.03 & 0.09 & 1 \end{pmatrix}$$

which induces contemporaneous correlation in the (still homoscedastic) errors. Finally, we fully specify the BEKK (following the empirical example in Lütkepohl, 2005) as

$$\boldsymbol{C} = \begin{pmatrix} 0.04 & 0 & 0 \\ 0.001 & 0.03 & 0 \\ 0.005 & 0.003 & 0.09 \end{pmatrix} \quad \boldsymbol{F} = \begin{pmatrix} 0.25 & 0.004 & 0.030 \\ 0.004 & 0.33 & 0.024 \\ 0.030 & 0.024 & 0.038 \end{pmatrix}$$

$$\boldsymbol{G} = \begin{pmatrix} 0.94 & 0.023 & 0.02 \\ 0.023 & 0.86 & 0.04 \\ 0.02 & 0.04 & 0.90 \end{pmatrix}.$$

3.3.3 The Heteroscedastic Cointegration Model of McCabe, Leybourne, and Harris (2006)

The third model in the simulation experiment is the stochastic cointegration model of Harris *et al.* (2002) in Equation (3.1). It is implemented as a slightly modified version of the data generating process considered by McCabe *et al.* (2006) in the following way:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ (1-d_1) & d_1 \end{pmatrix} \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix} + \begin{pmatrix} v_{x,t} & 0 \\ v_{y,t} & 0 \end{pmatrix} \begin{pmatrix} h_{x,t} \\ h_{y,t} \end{pmatrix} + \begin{pmatrix} u_{x,t} \\ u_{y,t} \end{pmatrix}. \tag{3.12}$$

The difference to the original version of McCabe *et al.* (2006) lies in the matrix $\boldsymbol{\Pi}$ (cp. Equation (3.3)). It contains an element $\pi_{21} = (1 - d_1)$ here whereas it is implemented as $\pi_{21} = 1$ by McCabe *et al.* (2006). There is only a difference when the two processes are not cointegrated: in the case where $\pi_{21} = 1$, the series y_t contains the same random walk component as x_t plus another random walk. If $\pi_{21} = (1 - d_1)$, the two series will be two completely independent random walks. The data generating process was then implemented according to

$$\begin{array}{rclcrcl} u_{x,t} & = & 0.5u_{x,t-1} + \epsilon_{1,t} & & u_{y,t} & = & -0.5u_{y,t-1} + \epsilon_{2,t} \\ v_{x,t} & = & -0.8v_{x,t-1} + 0.3d_2\epsilon_{3,t} & & v_{y,t} & = & 0.8v_{y,t-1} + 0.2d_3\epsilon_{4,t} \\ w_{1,t} & = & w_{1,t-1} + \epsilon_{5,t} & & w_{2,t} & = & w_{2,t-1} + \epsilon_{6,t} \\ h_{x,t} & = & h_{x,t-1} + \epsilon_{7,t} & & h_{y,t} & = & h_{y,t-1} + \epsilon_{7,t} \end{array}$$

with $(\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t}, \epsilon_{4,t}, \epsilon_{5,t}, \epsilon_{6,t}, \epsilon_{7,t}, \epsilon_{8,t})$ a multivariate Gaussian white noise process. Contemporaneous correlation may be induced by $\mathrm{cov}(\epsilon_{2,t}, \epsilon_{4,t}) = \mathrm{cov}(\epsilon_{4,t}, \epsilon_{5,t}) = 0.5$. Whether x_t and/or y_t are $I(1)$ or HI is determined by d_2 and d_3, respectively. If one of them is equal to zero, the respective series will be an $I(1)$ series. If d_2 and/or d_3 are different from zero, x_t and/or y_t are heteroscedastically integrated. Whether the series are cointegrated depends on d_1: if $d_1 \neq 0$ there is no cointegration in any sense whilst x_t and y_t are cointegrated if $d_1 = 0$.

3.3.4 General Simulation Design

All simulations have been conducted in GAUSS using 10,000 replications for each experimental setting. In the simulation of the data, the first 200 observations are discarded to avoid startup effects (cp., *inter alia*, Haug, 2002). The random seed to initiate the random number generator has been set to 746283. Tests are conducted on the $\alpha = 5\%$ significance level for 100 observations and $\alpha = 1\%$ for 1000 observations. The two estimated cointegration models are both specifications without deterministic trends. Model one (CIM 1 in the latter) allows for an intercept in the VAR specification, while for model two (CIM 2) the intercept is moved to the cointegration equation.

Note that the cointegration tests are performed stepwise as suggested by Johansen (1992). The respective rejection rates reported in Tables 3.1 to 3.4 for the hypotheses $r = 1$ and $r = 2$ (where applicable), thus, depend upon rejection of the first null hypothesis $r = 0$. More precisely, if the hypothesis $r = 0$ is rejected in less than 100% of all simulation runs, the following hypothesis tests are based on less than 10,000 replications. Again, we follow an empiricist's approach and conduct the tests as we would treat a single sample. We then check how the tests behave given that we know the true data generating process. In proceeding this way (in accordance to Haug, 1996, for example), however, we can not control the significance level of the test as a whole. This problem, however, is not the focus of this study.

3.4 Simulation Results

The first model setting considered is the design in Equations (3.6) to (3.8). The simulation results are summarized in Table 3.1, an example for a simulated dataset for each setting is provided in Figure 3.1. Cases 1 and 2 are designed such that the assumptions needed for derivation of the test are fulfilled. In Case 1 the data are cointegrated and the errors are homoscedastic. So when using the true cointegration model for the test (which is CIM 1 here), $H_0 : r = 0$ is always rejected. For $H_0 : r = 1$, we find rejection rates which are very close to the chosen significance level. When using the wrong testing model (CIM 2) the data are considered stationary 6 to 10 times more often than would be expected under the given significance level. This is true for both the trace as well as the maximum eigenvalue test. In Case 2 the data are not cointegrated. Here we find rejection rates of the first null hypothesis $r = 0$ close to the significance level when using CIM 1 as testing model. When using CIM 2, it turns out that the maximum eigenvalue statistic is less reliable than the trace statistic to reject cointegration, i.e. to not reject the first null hypothesis that $r = 0$.

In Cases 3 and 4 we introduce heteroscedastic errors as specified in Equation (3.8). If the data are cointegrated (Case 3) we find that the hypothesis that the data are not cointegrated is rejected in as many cases as is suggested by the significance level when using CIM 1. Again, when using CIM 2, the test is more inclined to suggest stationarity of the data. In the absence of cointegration (Case 4) the test still performs well (rejection rates of $H_0 : r = 0$ close to the significance level) when the correct testing model CIM 1 is used. When using CIM 2, rejection rates of the null hypothesis of no cointegration rise similar to Case 2 with homoscedastic errors.

The last setting considered is that only one data series exhibits heteroscedastic errors while the errors of the other data series are homoscedastic. Again, for both the case with cointegration (Case 5) and the case without cointegration

(Case 6) we find that the tests perform well given the correct testing model is used (CIM 1). Otherwise the trace statistic is somewhat less reliable than the maximum eigenvalue statistic. We conclude from this part of the simulation that if the underlying models are of simple structure (in particular, the errors are uncorrelated) the Johansen tests are quite reliable to either identify or reject cointegration. Only the choice of the testing model is crucial in some cases. In an empirical study where the true structure of the data generating process is not known it, thus, seems advisable to specify both models and perform the tests, unless a theoretical model suggests the use of only one particular specification.

The second model considered is the multivariate cointegration model specified in Equations (3.9) to (3.11) which, for convenience, is limited to three data series. In contrast to the model considered before, contemporaneous correlation may be induced via the full specification of a variance-covariance matrix through a multivariate GARCH model in BEKK representation[3]. Simulation results are presented in Table 3.2 and samples of the different data generating processes are depicted in Figure 3.2. For the first case where the data are not cointegrated and the errors are uncorrelated and homoscedastic, we find rejection rates of the first null hypothesis $r = 0$ which are close to the significance level, especially in large samples when using CIM 1, the correct specification of the VAR. When using CIM 2, the rejection rates rise by a factor 2 in case of the trace statistic, but remain close to the significance level when using the maximum eigenvalue statistic. The rejection rates for the hypotheses $r = 1$ and $r = 2$ are not necessarily meaningful any more as they are only tested once the first null hypothesis has been rejected. For Case 2 with cointegrated data we keep the assumptions on the innovations. It turns out that both tests (in both CIM 1 and 2) always reject $H_0 : r = 0$. As we have one cointegration relationship only, the second null hypothesis $r = 1$ should not be rejected. For small samples we find a slightly higher than α rejection rate while in large samples it is close to α. Again, the maximum eigenvalue statistic is more reliable if the underlying test model is misspecified.

In Cases 3 and 4 we induce contemporaneous correlation in the innovations. The results of the testing remain stable compared to Cases 1 and 2. So contemporaneous correlation is not found to affect the cointegration test negatively.

[3] A VECH representation has been specified as well. As the conclusions to be drawn are qualitatively the same as for the BEKK model, results are not reported.

In Cases 5 and 6 the innovations are heteroscedastic and contemporaneously correlated through the specification of a covariance matrix. We find that for the model without cointegration the rejection rate of the first null hypothesis that the data are not cointegrated is rejected more often than expected, even in large samples. This is true for both the trace as well as the maximum eigenvalue test. On the other hand, if the data are indeed cointegrated, both tests reliably reject $H_0 : r = 0$. However, the rejection rates of the second null hypothesis $r = 1$ are higher than they should be given the significance level. So one would be inclined to assume more than one cointegration relationship in too many cases. The following null hypothesis $r = 2$ is then also rejected in more cases than expected. So it seems in general that the test is more capable of detecting cointegration if it is present (which is at least partially due to the way the test is conducted, trying to avoid under-estimation of the number of cointegrating vectors) than to reject it if it is not present.

The third model considered is the heteroscedastic cointegration model by McCabe *et al.* (2006) as given in Equation (3.12). Simulation results are summarized in Table 3.3 for the case with correlated errors and in Table 3.4 for uncorrelated errors. One realization of the data generating process with correlated errors for each parameter setting is given in Figure 3.3. Note again that the difference to the previous models lies in the assumption about the non-stationarity behavior which has been deterministic so far, i.e. for any t the data were non-stationary. In the case of heteroscedastic integration the degree of integration is stochastic and varies around one. When applying the Johansen test to this model framework we find the following. In the first case where $d_1 = 1$, $d_2 = d_3 = 0$, the data are individually integrated of order 1 and not cointegrated, the shocks to the system, however, are correlated. We find that under these circumstances the Johansen test overrejects the null hypothesis of no cointegration about 2-6 times more often than would be acceptable under the respective significance level. This is true for both the trace as well as the maximum eigenvalue test. The choice of CIM 2 even leads to rejection rates of the true null hypothesis $r = 0$ which are 4-14 times higher as the significance level on which tests are conducted. The second null hypothesis $r = 1$ is, at least under CIM 1, rejected in accordance with α. The reason for this behavior is solely due to the contemporaneous correlation in the innovations process. Once $\text{cov}(\epsilon_{2,t}, \epsilon_{4,t}) = \text{cov}(\epsilon_{4,t}, \epsilon_{5,t}) = 0$ the performance of the test is far better in the sense that rejection rates correspond to the significance level.

The second setting is $d_1 = d_2 = d_3 = 0$, i.e. the data are cointegrated, but still not heteroscedastically integrated. In small samples the power of the trace and maximum eigenvalue test to reject the wrong null hypothesis $r = 0$ is weak: rejection rates are well below $(100 - \alpha)\%$. Rejection of $H_0 : r = 1$ is about twice as much as suggested by α under CIM 1 and ten times under CIM 2. In large samples, the test performs well as regards $H_0 : r = 0$ with rejection rates of 100%. However, rejection rates are well above an acceptable rate for $H_0 : r = 1$. Again, this is largely due to the contemporaneous correlations in the $\epsilon_{i,t}$.

In Cases 3 and 4, both ($d_2 = d_3 = 1$) or one ($d_2 = 1$, $d_3 = 0$) of the data series are individually heteroscedastically integrated, but they are not cointegrated ($d_1 = 1$). Here the Johansen test erroneously rejects the null hypothesis of no cointegration far too often. We find rejection rates which are between 40 and 99%, irrespective whether CIM 1 or 2 is used. Contemporaneous correlation in the innovations slightly worsens this effect.

In the last two settings 5 and 6, again either both ($d_2 = d_3 = 1$) or one ($d_2 = 1$, $d_3 = 0$) of the data series are individually heteroscedastically integrated. Now, however, the two series are cointegrated ($d_1 = 0$) as well. Here the influence of contemporaneous correlation is again more important. If we use CIM 1 we find, particularly in large samples, the rejection rates to be close to 100% for $H_0 : r = 0$. Non-rejection of the second null hypothesis that $r = 1$ is well above the significance level, ranging from 30-65% if the errors are correlated and 20-50% if the errors are uncorrelated. Surprisingly, matters are worse in large samples. If CIM 2 is used for testing, the rejection rates for $r = 1$ even raise. It, thus, seems that heteroscedastic integration of time series leads the Johansen tests to the conclusion that the data in question are stationary. If only one of the data series is heteroscedastically integrated while the other one is strictly $I(1)$, the performance of the tests improves substantially. Rejection rates of $H_0 : r = 0$ are at 100% in large samples for both CIM 1 and 2. If the errors are uncorrelated, rejection rates of the second $H_0 : r = 1$ are 2-5 times higher than would be suggested by the significance level. If the errors are correlated, the rate of wrong rejections rises to 13% in small samples and even 12% in large samples.

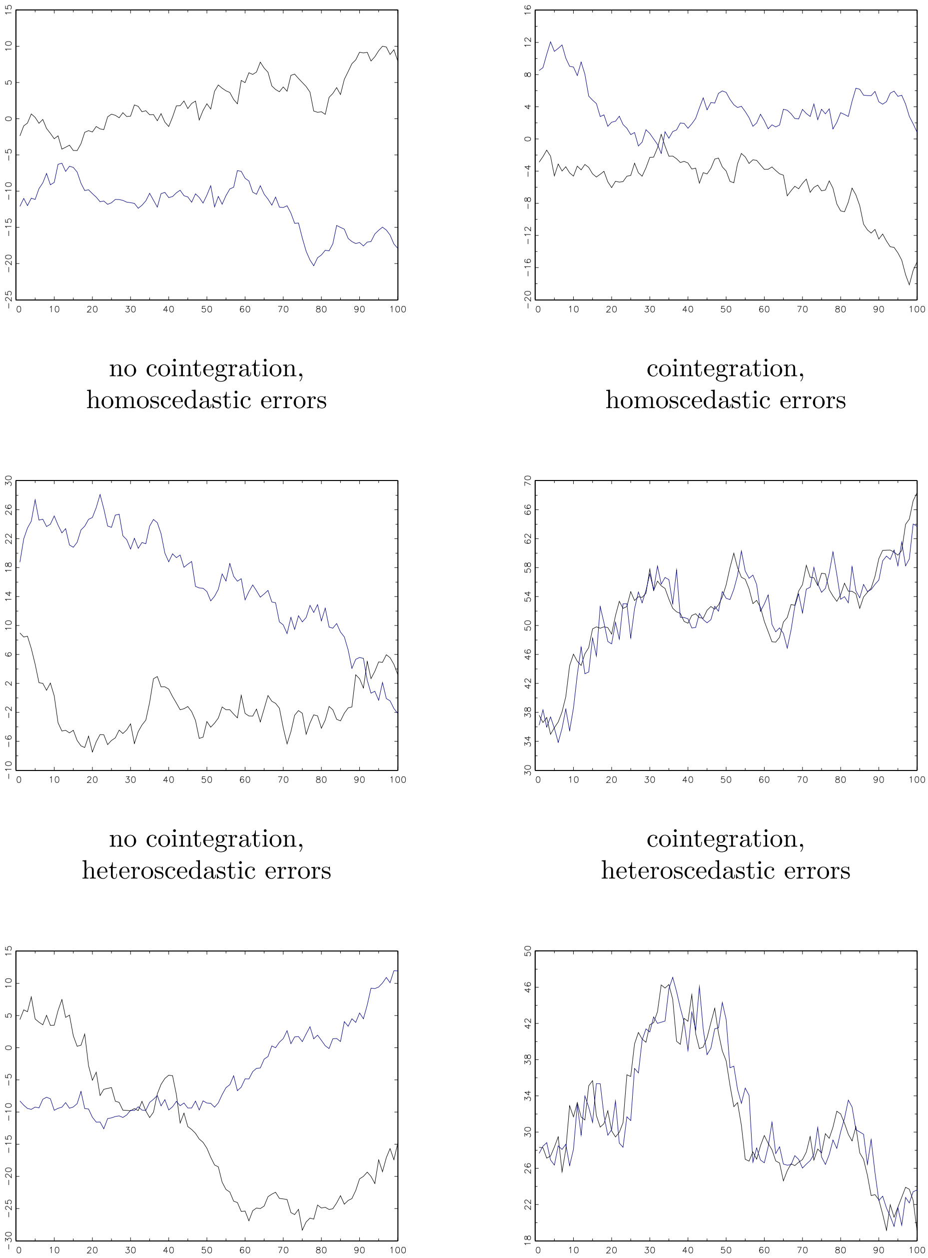
3.5 Concluding Remarks

The previous simulations support the notion that the Johansen cointegration test is largely unaffected by either heteroscedasticity or contemporaneous correlation individually. Only when these features are combined the performance of the tests weakens. Under such circumstances it seems that they are more capable to detect cointegration if it is indeed present (albeit with a tendency to overestimate the number of cointegrating vectors) than to not reject the first null hypothesis of no cointegration if the data are not cointegrated. This finding is in line with the reported tendency to slightly overestimate the number of cointegrating vectors by Ho and Sorensen (1996) in the context of high dimensional cointegrated VARs. It also corresponds to the results of Lee and Tse (1996) who find increasing, albeit (as they say) not serious size distortion if GARCH-type heteroscedasticity is present in the data and the variance tends to explosive behavior.

A crucial point in our simulations is whether the data are truly $I(1)$ or heteroscedastically integrated. In the latter case, the Johansen framework is not reliable enough to detect cointegration. This, however, is purely due to the fact that the data can be locally stationary which is the conclusion the Johansen tests tend to draw more often than would be appropriate. So careful pre-analysis of the data is necessary to determine the appropriate testing framework.

From all the simulations it should become clear that the choice of the test model is crucial. We limited the study to two settings (without trend and an intercept either in the VAR or in the cointegrating equation). However, more settings are possible (inclusion of a deterministic trend together with or without a constant, for example) and inconsiderate application might lead to wrong conclusions. There are, in our view, two ways to avoid erroneous conclusions. First, the tested model can be justified by theory, i.e. theory suggests the inclusion or exclusion of certain model parameters. The cointegration test in this case also coincides with a test of the model. The second option would be to thoroughly test the data (whether they drift or trend) and then to still test more than one model setting to make sure the conclusions are robust.

Figure 3.1: Data from the bivariate model



no cointegration, homoscedastic errors

cointegration, homoscedastic errors

no cointegration, heteroscedastic errors

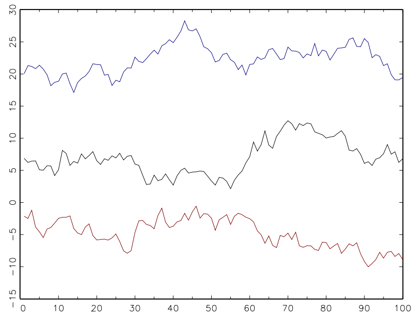
cointegration, heteroscedastic errors

no cointegration, mixed errors

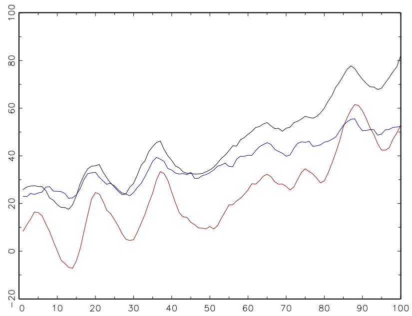
cointegration, mixed errors

The graphics present data series which are simulated according to bivariate data generating process given in Equations (3.6) - (3.8).

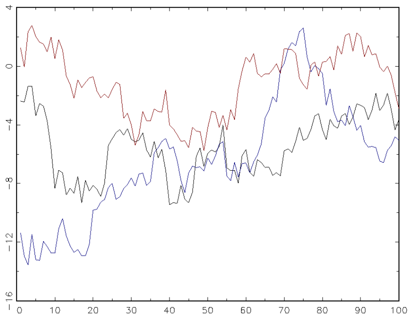
Figure 3.2: VAR-BEKK Data



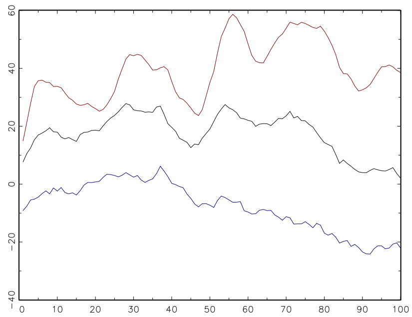
no cointegration,
uncorrelated, homoscedastic errors



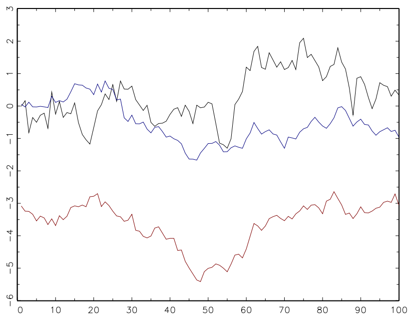
cointegration,
uncorrelated, homoscedastic errors



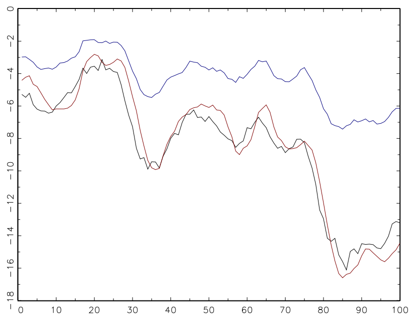
no cointegration,
correlated, homoscedastic errors



cointegration,
correlated, homoscedastic errors



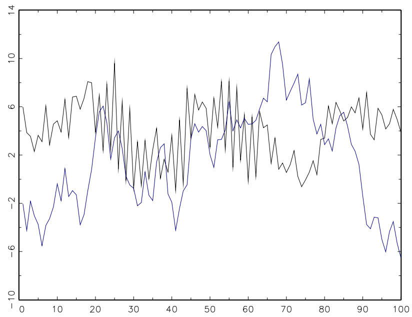
no cointegration,
heteroscedastic errors



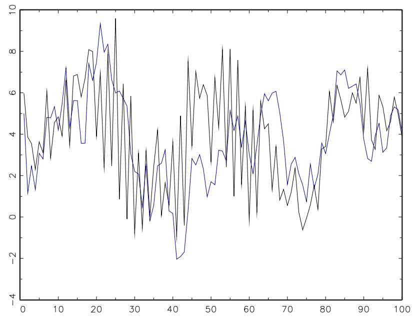
cointegration,
heteroscedastic errors

The graphics present data series which are simulated according to the multivariate data generating process given in Equations (3.9) - (3.11).

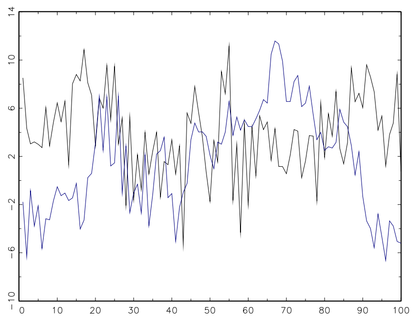
Figure 3.3: MLH Data



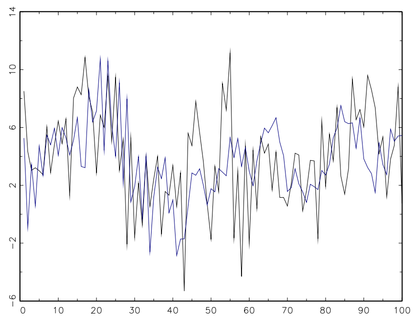
no cointegration,
I(1) data series



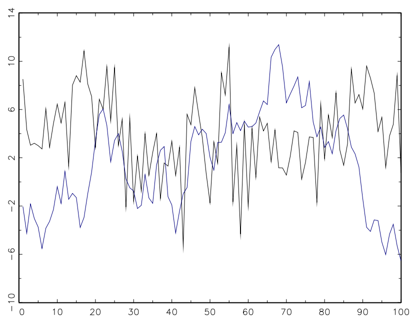
cointegration,
I(1) data series



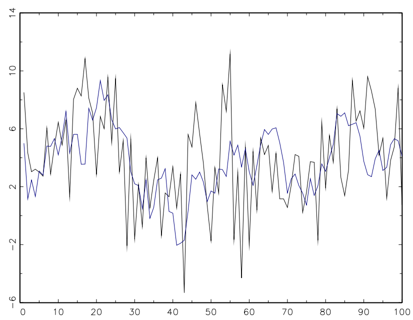
no cointegration,
heteroscedastically integrated data



cointegration,
heteroscedastically integrated data



no cointegration,
mixed data



cointegration,
mixed data

The graphics present data series which are simulated according to the data generating process of McCabe *et al.* (2006) given in Equation (3.12) using correlated innovations.

Table 3.1: Results from the bivariate model

			Trace statistic		$\lambda_{\max}$ statistic	
CIM	obs	$\alpha(\%)$	$r=0$	$r=1$	$r=0$	$r=1$
Case 1: cointegration, homoscedastic errors						
1	100	5	100.00	5.51	100.00	5.51
1	1000	1	100.00	0.89	100.00	0.89
2	100	5	100.00	30.99	100.00	30.99
2	1000	1	100.00	9.98	100.00	9.98
Case 2: no cointegration, homoscedastic errors						
1	100	5	5.87	9.37	5.78	3.29
1	1000	1	0.96	3.12	0.97	0.00
2	100	5	12.86	44.48	7.49	22.43
2	1000	1	2.79	24.73	1.40	5.00
Case 3: cointegration, heteroscedastic errors						
1	100	5	100.00	5.65	100.00	5.65
1	1000	1	100.00	0.88	100.00	0.88
2	100	5	100.00	31.09	100.00	31.09
2	1000	1	100.00	10.04	100.00	10.04
Case 4: no cointegration, heteroscedastic errors						
1	100	5	6.30	8.41	6.04	3.31
1	1000	1	1.04	1.92	1.02	0.00
2	100	5	13.03	43.82	7.50	23.60
2	1000	1	2.73	27.11	1.39	4.32
Case 5: cointegration, mixed errors						
1	100	5	100.00	5.71	100.00	5.71
1	1000	1	100.00	0.91	100.00	0.91
2	100	5	100.00	31.25	100.00	31.25
2	1000	1	100.00	10.02	100.00	10.02
Case 6: no cointegration, mixed errors						
1	100	5	5.98	9.03	5.87	3.24
1	1000	1	0.97	4.12	1.01	0.99
2	100	5	12.91	44.93	7.43	23.82
2	1000	1	2.69	26.02	1.41	4.96

The table reports empirical rejection rates of the respective null hypotheses which are:

- Trace statistic: $H_0 : rg(\Pi) = r$ vs $H_1 : rg(\Pi) > r$;
- Maximum Eigenvalue ($\lambda_{\max}$) statistic: $H_0 : rg(\Pi) = r$ vs $H_1 : rg(\Pi) = r+1$;

the hypothesis $r = 1$ is only tested if the hypothesis $r = 0$ has been rejected.

CIM is the underlying cointegration model. 1 is no trend, intercept in VAR; 2 is no trend, intercept in the cointegration equation. obs is the number of observations and α is the significance level (in per cent).

Table 3.2: Results from the VAR-BEKK model

			Trace statistic			$\lambda_{\max}$ statistic		
CIM	obs	$\alpha(\%)$	$r=0$	$r=1$	$r=2$	$r=0$	$r=1$	$r=2$
Case 1: no cointegration; uncorrelated homoscedastic errors								
1	100	5	7.18	7.66	20.00	7.18	2.92	0.00
1	1000	1	0.94	2.13	0.00	1.04	0.00	0.00
2	100	5	11.19	20.02	60.71	7.67	4.30	18.18
2	1000	1	2.26	4.87	27.27	1.09	0.00	0.00
Case 2: cointegration; uncorrelated, homoscedastic errors								
1	100	5	100.00	7.84	8.42	100.00	7.50	3.60
1	1000	1	100.00	0.90	1.11	100.00	1.03	0.00
2	100	5	100.00	15.11	46.13	100.00	9.44	25.85
2	1000	1	100.00	2.79	25.81	100.00	1.42	8.45
Case 3: no cointegration; correlated, homoscedastic errors								
1	100	5	7.25	7.45	20.37	7.18	2.92	0.00
1	1000	1	0.99	2.02	0.00	1.04	0.00	0.00
2	100	5	11.23	20.04	59.56	7.68	4.30	18.18
2	1000	1	2.18	4.59	30.00	1.16	0.00	0.00
Case 4: cointegration; correlated, homoscedastic errors								
1	100	5	100.00	7.92	10.10	100.00	7.51	3.86
1	1000	1	100.00	0.98	1.02	100.00	1.11	0.00
2	100	5	100.00	15.33	45.34	100.00	9.12	24.01
2	1000	1	100.00	2.75	24.73	100.00	1.43	9.09
Case 5: no cointegration; heteroscedastic errors								
1	100	5	8.60	9.30	13.75	8.14	3.32	3.70
1	1000	1	2.44	4.51	0.00	1.86	100.00	0.00
2	100	5	12.58	17.57	56.11	8.47	4.25	16.67
2	1000	1	1.87	3.74	42.86	1.45	0.00	0.00
Case 6: cointegration; heteroscedastic errors								
1	100	5	99.99	9.63	10.49	100.00	8.84	4.64
1	1000	1	100.00	2.93	4.44	100.00	2.33	3.00
2	100	5	100.00	16.23	45.29	100.00	10.72	25.09
2	1000	1	100.00	2.53	15.02	100.00	1.75	4.00

The table reports empirical rejection rates of the respective null hypotheses which are:

- Trace statistic: $H_0 : rg(\Pi) = r$ vs $H_1 : rg(\Pi) > r$;
- Maximum Eigenvalue ($\lambda_{\max}$) statistic: $H_0 : rg(\Pi) = r$ vs $H_1 : rg(\Pi) = r + 1$;

the hypothesis $r = 1$ is only tested if the hypothesis $r = 0$ has been rejected; the hypothesis $r = 2$ is only tested if the hypothesis $r = 1$ has been rejected;
CIM is the underlying cointegration model. 1 is no trend, intercept in VAR; 2 is no trend, intercept in the cointegration equation. obs is the number of observations and α is the significance level (in per cent).

Table 3.3: Results from the McCabe *et al.* (2006) model with correlated errors

			Trace statistic		λ_{max} statistic	
CIM	obs	$\alpha(\%)$	$r=0$	$r=1$	$r=0$	$r=1$
Case 1: no cointegration, I(1) data						
1	100	5	9.06	7.06	8.77	3.19
1	1000	1	6.09	1.97	6.52	0.92
2	100	5	22.86	51.22	13.54	32.50
2	1000	1	14.82	22.94	9.33	10.18
Case 2: cointegration, I(1) common trend						
1	100	5	55.65	10.21	56.63	7.24
1	1000	1	100.00	2.26	100.00	2.26
2	100	5	81.29	47.52	68.71	41.44
2	1000	1	100.00	17.04	100.00	17.04
Case 3: no cointegration, HI data						
1	100	5	59.36	27.29	55.37	25.05
1	1000	1	98.03	57.95	97.35	58.21
2	100	5	77.25	69.68	63.53	66.47
2	1000	1	99.21	79.80	97.90	80.12
Case 4: no cointegration, HI and I(1) data						
1	100	5	43.11	10.93	42.39	8.04
1	1000	1	91.80	6.49	91.53	6.41
2	100	5	64.59	53.82	51.18	46.72
2	1000	1	95.29	30.93	92.89	30.39
Case 5: cointegration, HI data						
1	100	5	91.41	27.35	91.10	26.47
1	1000	1	100.00	64.61	100.00	64.61
2	100	5	98.07	67.76	95.08	67.55
2	1000	1	100.00	84.79	100.00	84.79
Case 6: cointegration, HI and I(1) data						
1	100	5	78.91	13.69	78.62	12.05
1	1000	1	100.00	12.23	100.00	12.23
2	100	5	93.64	57.01	86.60	55.28
2	1000	1	100.00	40.70	100.00	40.70

The table reports empirical rejection rates of the respective null hypotheses which are:

- Trace statistic: $H_0 : rg(\Pi) = r$ vs $H_1 : rg(\Pi) > r$;
- Maximum Eigenvalue (λ_{max}) statistic: $H_0 : rg(\Pi) = r$ vs $H_1 : rg(\Pi) = r + 1$;

the hypothesis $r = 1$ is only tested if the hypothesis $r = 0$ has been rejected;
the data are simulated according to Equation (3.12), the innovations are correlated;
CIM is the underlying cointegration model; 1 is no trend, intercept in VAR; 2 is no trend, intercept in the cointegration equation. obs is the number of observations and α is the significance level (in per cent).

Table 3.4: Results from the McCabe *et al.* (2006) model with uncorrelated errors

			Trace statistic		$\lambda_{\max}$ statistic	
CIM	obs	$\alpha(\%)$	$r=0$	$r=1$	$r=0$	$r=1$
Case 1: no cointegration, I(1) data						
1	100	5	7.38	8.81	7.38	3.66
1	1000	1	3.97	2.27	4.28	0.93
2	100	5	18.23	47.28	10.97	26.07
2	1000	1	8.92	15.92	5.78	5.71
Case 2: cointegration, I(1) common trend						
1	100	5	55.23	8.17	57.75	5.58
1	1000	1	100.00	1.45	100.00	1.45
2	100	5	81.02	42.94	70.10	37.38
2	1000	1	100.00	13.62	100.00	13.62
Case 3: no cointegration, HI data						
1	100	5	52.16	24.37	49.19	21.61
1	1000	1	96.42	50.73	95.36	51.04
2	100	5	70.62	66.19	57.61	61.99
2	1000	1	97.98	75.87	95.99	76.05
Case 4: no cointegration, HI and I(1) data						
1	100	5	35.54	10.47	35.61	7.27
1	1000	1	85.38	4.93	85.05	4.74
2	100	5	56.95	52.48	43.73	43.75
2	1000	1	90.74	25.80	86.79	24.66
Case 5: cointegration, HI data						
1	100	5	90.34	18.33	91.65	17.52
1	1000	1	100.00	49.06	100.00	49.06
2	100	5	97.75	57.96	95.78	57.53
2	1000	1	100.00	73.80	100.00	73.80
Case 6: cointegration, HI and I(1) data						
1	100	5	77.90	9.23	80.05	7.88
1	1000	1	100.00	5.01	100.00	5.01
2	100	5	93.44	47.77	87.31	45.75
2	1000	1	100.00	27.23	100.00	27.23

The table reports empirical rejection rates of the respective null hypotheses which are:

- Trace statistic: $H_0 : rg(\Pi) = r$ vs $H_1 : rg(\Pi) > r$;
- Maximum Eigenvalue ($\lambda_{\max}$) statistic: $H_0 : rg(\Pi) = r$ vs $H_1 : rg(\Pi) = r + 1$;

the hypothesis $r = 1$ is only tested if the hypothesis $r = 0$ has been rejected;
the data are simulated according to Equation (3.12), the innovations being uncorrelated;
CIM is the underlying cointegration model; 1 is no trend, intercept in VAR; 2 is no trend, intercept in the cointegration equation. obs is the number of observations and α is the significance level (in per cent).

Chapter 4

On Cointegration of International Financial Markets

4.1 Introduction

For many years now financial econometrics has dedicated a lot of effort and resources to the analysis of the linkages between international financial markets. In the context of the present turmoil the question how exactly these markets are linked and how these linkages can be described best is again in the focus of researchers. A great number of empirical studies have already documented that financial markets around the globe are not independent (at the least because of worldwide monetary and commodity flows). The assumption that they even share common stochastic trends is therefore also quite plausible at first sight. This is the reason why cointegration analysis has been one of the dominating tools in the study of interrelatedness of financial markets since the seminal work of Engle and Granger (1987) and Johansen (1988).

Based on the assumption that stock markets in different countries share common stochastic trends, numerous studies have tried to detect those. One of the first was Kasa (1992) who can identify one common stochastic trend for the stock markets of the U.S., Japan, England, Germany, and Canada. He used monthly and quarterly data over a period of almost 16 years which suits the notion that cointegration is a long term concept while short run deviations from the common trend are possible. More recent contributions include Choudhry, Lu, and Peng (2007), Lagoarde-Segot and Lucey (2007), Constantinou, Kazandjian, Kouretas, and Tahmazian (2008) and Valadkhani and Chancharat (2008). These studies have in common that they all identify exactly one common stochastic trend. However, there is no economic or financial theory predicting how many common stochastic trends there should be. Empirically, Click and Plummer (2005), for example, who investigate the relationship between five ASEAN stock markets on a daily basis for four years, find that these markets are cointegrated. The authors can, however, identify only one cointegrating vector. This means that there would be four stochastic trends which influence the cointegration relationship. The authors conclude that in this case the integration of these financial markets is far from being perfect. Empirical work, thus, cannot conclude how many stochastic trends financial markets share. How many cointegrating vectors will be found, therefore, critically depends on how many markets are analyzed.

It is not only unclear how many stochastic trends international financial markets would share. Empirical results whether financial markets do share one ore more stochastic trends at all are mixed. The studies cited above all find evidence for the existence of a cointegration relationship. In contrast, Chan, Gup, and Pan (1997) who analyse 18 stock market indices, find that these markets are not cointegrated. The analysis is conducted using monthly data from 1961 to 1992. Pascual (2003) studies whether the degree of integration between the French, German, and UK stock market increases. He does not find a cointegration relationship using quarterly observations from 1960 to 1999 either. The results of Narayan and Smyth (2005) who investigate the relationship between the stock markets of New Zealand, Australia and the G7 countries, are mixed, depending on which test they use to detect cointegration. Their analysis is based on real monthly observations from 1967 to 2003.

As regards financial theory, the existence of cointegration relationships in the long run would contradict the Efficient Market Hypothesis (EMH). The latter requires that returns and, thus, future prices, be not predictable. A common model frequently used in the literature which captures this behavior of stock or index returns at high frequencies is the random walk model for stock prices. It dates back to work by Fama (1965) and Malkiel (1973) and has ever since frequently been applied (see, *inter alia*, Black, 1986; Richardson, 1993; Lewellen, 2002; Godfrey, Granger, and Morgenstern, 2007) and tested, albeit with mixed results (see, *inter alia*, Bondt and Thaler, 1985; Fama, 1995; Worthington and Higgs, 2009). Cointegration by contrast would allow for some kind of predictability in the long run, even though short run predictions are not possible. This argument is not limited to stock markets. Granger (1986) shows that gold and silver prices are not cointegrated once these prices are generated on an efficient market.

This study suggests that under the assumption that stock prices are generated according to the random walk model, international financial markets are not cointegrated. It follows similar arguments as put forward by Richards (1995) who claims that stock return indices in one stock market cannot be cointegrated if one assumes that excess returns are generated according to the Capital Asset Pricing Model (CAPM). He argues that the company specific shocks of one company would have to be offset by shocks of the other company. However, both of these shock would have to be completely unexpected, but identical in size and direction. He states that this would rule out the possibility that any management decision permanently affected a company's stock price. He summarizes that these company specific shocks "will not translate into a cointegrating relationship between the actual return indices for the two (or more) assets." It seems that this result has been neglected in the literature on cointegration of financial markets since then. This paper will therefore reinforce the argumentation that company specific shocks eventually inhibit the existence of cointegration relations (as defined by Engle and Granger, 1987) between international stock market indices. In contrast to Richards (1995) who seeks to explain the results of Kasa (1992) obtained on low frequencies, our line of argumentation will keep features of high frequency data in mind. Our model will therefore be different from Richards (1995), in that we will not rely on the CAPM, but the more general random walk model for stock prices. It is widely accepted that on high frequencies stock prices are modeled best by a random walk. Further, Richards (1995) attributed some of the results in the literature specifically to a small sample bias in the Johansen (1988) cointegration testing framework. This issue can be regarded as overcome as high frequency data (in particular daily data) are nowadays widely (and even freely) available. However, daily data are marked by other features (e.g. heteroscedasticity) which have to be taken into account when testing international financial markets for cointegration.

The chapter proceeds as follows. Section 4.2 outlines the common random walk model of stock prices and derives the implications for stock indices and cointegration. It shows that stock market indices from different countries are not cointegrated and illustrates the result using a simple example. Section 4.3 presents the results of a cointegration and correlation analysis of 28 stock indices. The features found in this section are modeled by simulation methods in section 4.4 to illustrate the adequacy of our model assumptions. Section 4.5 concludes.

4.2 Stock Prices, Indices and Cointegration

The basic model for stock prices which is widely used in the literature assumes that log-prices follow a random walk. The model can be written as

$$p_{i,t} = p_{i,t-1} + e_t, \tag{4.1}$$

where $p_{i,t}$ is the price of stock i in time t (in logarithms). The error term e_t is a white noise process with $\mathbb{E}[e_t] = 0$ and Variance σ_t^2. Whether the variance is time dependent or not will not influence the theoretical result, so we suppress the time subscript in σ^2 in the subsequent outline. The model may contain a drift term δ_t, but there is an ongoing debate on whether a drift term is compatible with information efficient markets (Malkiel, 1973; Edwards and Magee, 2001). The following results hold irrespective of the inclusion of a drift term which is therefore neglected in the following as well.

Following the idea of latent factor models in finance, we allow the error term in Equation (4.1) to consist of different components, namely a global, a local and an idiosyncratic component (cp. Dungey, Martin, and Pagan, 2000; Jung, Liesenfeld, and Richard, 2010). Thus, e_t is a multivariate white noise process with $\mathbb{E}[e_t] = 0$ and Variance Σ. The error term in Equation (4.1), thus, needs to be written as $\iota' e_t$ where ι is a (3×1)-vector of ones. A common assumption would then be that $e_t \sim N(0, \Sigma)$. The errors are serially uncorrelated ($\mathbb{E}[e_{i,t}, e_{i,t-1}] = 0$), but may be cross-sectionally correlated such that $\mathbb{E}[e_{i,t}, e_{j,t}] \neq 0$.

A stock market index is usually calculated as a weighted and normalized sum of individual stock prices. Without loss of generality we assume that an index X_j is calculated as

$$X_{j,t} = \sum_{i=1}^{n} w_i \cdot p_{i,t}, \tag{4.2}$$

where w_i is the weight for asset i. As X_j is composed of n price series which are assumed to follow a random walk, the index will be a weighted sum of n random walks and, thus, also be non-stationary.

The crucial question which arises is whether any two stock market indices X_1 and X_2 of countries 1 and 2, respectively, are cointegrated in the sense of Engle and Granger (1987). This is the case if and only if $X_1 - \beta X_2$ is stationary[1].

[1] In a bivariate cointegration analysis one of the coefficients in the cointegrating vector is

This will happen if the linear combination of the indices successfully eliminates the stochastic trends which compose the individual stock prices. However, if stock i is an element of X_1 and at the same time not an element of X_2 (for all stocks i), X_1 and X_2 cannot be cointegrated. The reason is that the individual stochastic trends which are contained in the individual stocks do not cancel out, as the random walk contained in stock i will be different from the random walk contained in stock $i^\star$. As in the present framework two stock market indices are weighted averages of distinct $I(1)$-series, no linear combination exists which removes all stochastic trends. So for any β, $X_1 - \beta X_2 \sim I(1)$ and the stationarity requirement is violated. No cointegrating vector exists which would assure that $X_1 - \beta X_2 \sim I(0)$. Therefore, the indices are not cointegrated in the sense of Engle and Granger (1987).

This result also holds for market indices in one country as long as their basis, i.e. the stocks used to calculate them, are not identical. The same is true for the cross-listing of stocks which also does not alter the result. Cross-listing, i.e. the listing of a company on two exchanges in two different countries, would imply that a stock k is contained in both indices X_1 and X_2. A cointegration relationship between these two stocks most likely exists due to the law of one price (e.g. Hasbrouck, 1995; Grammig, Melvin, and Schlag, 2005) which allows only for temporary price deviations, but no fundamental ones. As regards the indices, however, only if all stocks are the same, i.e. one index is the exact reproduction of the other, these indices will be cointegrated. Two such indices are, to the best of our knowledge, not calculated on any stock exchange.

In order to illustrate the result that stock market indices of different countries are not cointegrated in the assumed context, we limit ourselves to two indices which are composed of only two stock prices each. Rewrite these four stock prices as

$$
\begin{aligned}
p_{1,t} &= p_{1,t-1} + g_t + l_{1,t} + \varepsilon_{1,t} &&= \sum_{s=1}^{t} g_s + \sum_{s=1}^{t} l_{1,s} + \sum_{s=1}^{t} \varepsilon_{1,s} \\
p_{2,t} &= p_{2,t-1} + g_t + l_{1,t} + \varepsilon_{2,t} &&= \sum_{s=1}^{t} g_s + \sum_{s=1}^{t} l_{1,s} + \sum_{s=1}^{t} \varepsilon_{2,s} \\
p_{3,t} &= p_{3,t-1} + g_t + l_{2,t} + \varepsilon_{3,t} &&= \sum_{s=1}^{t} g_s + \sum_{s=1}^{t} l_{2,s} + \sum_{s=1}^{t} \varepsilon_{3,s} \\
p_{4,t} &= p_{4,t-1} + g_t + l_{2,t} + \varepsilon_{4,t} &&= \sum_{s=1}^{t} g_s + \sum_{s=1}^{t} l_{2,s} + \sum_{s=1}^{t} \varepsilon_{4,s}, \qquad (4.3)
\end{aligned}
$$

usually normalized to 1 as in $X_1 - \beta X_2$ where the cointegrating vector is (1 -β).

where g_t is the global, $l_{j,t}$ the local and $\varepsilon_{i,t}$ the idiosyncratic innovation in e_t of Equation (4.1). We assume that the initial values $g_0 = l_{j,0} = \varepsilon_{i,0} = 0$. The indices are then constructed as

$$\begin{aligned} X_{1,t} &= w_1 p_{1,t} + (1-w_1) p_{2,t} \\ X_{2,t} &= w_2 p_{3,t} + (1-w_2) p_{4,t}. \end{aligned} \tag{4.4}$$

Substituting the individual prices in (4.4) by the respective stock prices in (4.3) gives

$$\begin{aligned} X_{1,t} &= w_1 p_{1,t-1} + (1-w_1) p_{2,t-1} + g_t + l_{1,t} + w_1 \varepsilon_{1,t} + (1-w_1)\varepsilon_{2,t} \\ &= \sum_{s=1}^{t} g_s + \sum_{s=1}^{t} l_{1,s} + w_1 \sum_{s=1}^{t} \varepsilon_{1,s} + (1-w_1) \sum_{s=1}^{t} \varepsilon_{2,s} \end{aligned} \tag{4.5}$$

$$\begin{aligned} X_{2,t} &= w_2 p_{3,t-1} + (1-w_2) p_{4,t-1} + g_t + l_{2,t} + w_2 \varepsilon_{3,t} + (1-w_2)\varepsilon_{4,t} \\ &= \sum_{s=1}^{t} g_s + \sum_{s=1}^{t} l_{2,s} + w_2 \sum_{s=1}^{t} \varepsilon_{3,s} + (1-w_2) \sum_{s=1}^{t} \varepsilon_{4,s}. \end{aligned} \tag{4.6}$$

In order for X_1 and X_2 to be cointegrated, the linear combination $X_1 - \beta X_2$ would have to eliminate the global, the two local as well as the four stock specific stochastic trends. Denote by $\hat{u}_t$ the residuals of a cointegration regression:

$$\begin{aligned} \hat{u}_t &= X_{1,t} - \beta X_{2,t} \\ &= \sum_{s=1}^{t} g_s - \beta \sum_{s=1}^{t} g_s + \sum_{s=1}^{t} l_{1,s} - \beta \sum_{s=1}^{t} l_{2,s} + w_1 \sum_{s=1}^{t} \varepsilon_{1,s} \\ &+ (1-w_1) \sum_{s=1}^{t} \varepsilon_{2,s} - \beta w_2 \sum_{s=1}^{t} \varepsilon_{3,s} - \beta(1-w_2) \sum_{s=1}^{t} \varepsilon_{4,s}. \end{aligned} \tag{4.7}$$

Cointegration would require $\hat{u}_t$ to be stationary. This is, however, not the case as it still contains random walk components. As can be seen easily, for $\beta = 1$ only the global stochastic trend is eliminated. The local and the stock specific stochastic trends, however, are still present. Thus, $X_1 - \beta X_2$ still contains a combination of stochastic trends and is not stationary. More precisely,

$$\hat{u}_t = \hat{u}_{t-1} + l_{1,t} - \beta l_{2,t} + w_1 \varepsilon_{1,t} + (1-w_1)\varepsilon_{2,t} - \beta w_2 \varepsilon_{3,t} - \beta(1-w_2)\varepsilon_{4,t} \tag{4.8}$$

which is an AR(1) process. The last result holds for any possible $\beta \in \mathbb{R}$. The only difference in $\hat{u}_t$ will be that it also contains $(1-\beta)g_t$, i.e. the global innovation.

One might criticize the assumption about the error term. Dividing it into global, local and idiosyncratic components might artificially and needlessly increase the number of stochastic trends in the price process. Allowing for one innovation term only, however, does not alter the result. As long as there are stock specific innovations which are modeled as a martingale difference sequence, the stock market index will always be a weighted average of stock specific stochastic trends and local innovations $l_{1,t} - \beta l_{2,t}$ would not appear in the process $\hat{u}_t$ in Equation (4.8). In economic terms, this means that there is always stock specific information which affects the share price of one company only and does not affect the share price of another company. It may happen that a certain news event affects the distribution of the innovations of other companies as well, i.e. the innovations are correlated, but they still are not exactly identical. Identity and corresponding weights, however, are what is required for the indices to be cointegrated in the sense of Engle and Granger (1987).

An important feature of financial time series which has been often documented in empirical studies is that return models exhibit heteroscedastic errors. In the theoretical derivation of the result why international financial markets are not cointegrated, the presence of heteroscedasticity does not matter. We show the relationship of returns which exhibit heteroscedasticity and level log-prices in an appendix to this chapter. A time-varying variance does only influence the behavior of the random walk in such a way that it would be more volatile. The important feature, the non-stationarity, is, however, not affected.

This argumentation easily extends to the multivariate case where it will be impossible to find a cointegration vector β such that a linear combination $\beta' x$ with x an $(n \times 1)$ vector of stock market indices, will be stationary. The global trend again may cancel out, but the stock specific innovations do not.

A further property which has frequently been documented in the empirical literature is the high correlation between stock market indices. According to the model framework above, there are possibly two sources which would induce correlation between indices. First, there may be stochastic trends which are common to the individual prices. If there is a global stochastic trend, this very same trend will be contained in both indices. Thus, the indices would not be independent any longer and therefore exhibit some degree of correlation which depends on the variance of the global trend relative to the variance of the idiosyncratic innovations. Second, index correlation could also be induced by cross-sectionally correlated innovations which is not ruled out by the above setting. We will show how these features interrelate in the simulation study in the next but one section.

4.3 An Empirical Example

In order to evaluate the theoretical result in Section 4.2 we first analyse a dataset of 28 stock indices[2] which are taken from finance.yahoo.com. The dataset covers daily close values between 1st March 2001 and 28th February 2009, i.e. we have 2084 observations. In case of a national holiday in one country, the previous closing value has been substituted as the still valid value as is standard in the literature. A necessary condition for any pair of two indices to be cointegrated is that each single index is integrated of order 1. We therefore perform Augmented Dickey-Fuller tests with individual lag-length selection using the Schwarz Information Criterion. The indices are all found to be non-stationary.

As financial time series are usually found to exhibit heteroscedastic errors, we use the Ljung-Box test for autocorrelation (see Ljung and Box, 1978) on the squared levels and the squared log-returns of each index, the squared variables being a crude measure of the variance of the respective time series. The tests are conducted using ten lags. We find that both the stock market indices as well as the respective return time series exhibit time dependence in the variance.[3]

We perform bivariate cointegration tests among all possible combinations of the indices. As we have 28 indices, there are 378 possible index pairs. To perform cointegration tests we rely on the Johansen (1988) methodology. We use the Johansen (1991) test instead of the Engle and Granger (1987) two-step method in order to keep the study comparable to Kasa (1992) and Richards (1995).

[2] AEX (Netherlands), All Ordinaries (Australia), Austrian Traded Index (Austria), Euronext Bel-20 (Belgium), Bovespa (Brazil), BSE Sensex (India), CAC 40 (France), CASE 30 (Egypt), DAX (Germany), Dow Jones Industrial Average (USA), Euro Stoxx 50 (Europe), Financial Times Stock Exchange (UK), Hang Seng (Hongkong), IPC (Mexico), ISEQ 20 (Ireland), Jakarta Composite (Indonesia), FTSE Bursa Malaysia KLCI Index (Malaysia), Madrid General (Spain), MerVal (Argentina), MIB TELEMATICO (Italy), Nasdaq Composite (USA), NZX 50 (New Zealand), Nikkei 225 (Japan), OMX Copenhagen-20 (Denmark), Oslo Exchange All Share (Norway), PSI 20 (Portugal), S&P 500 (USA), S&P TSX Composite (Canada), Seoul Composite (South Korea), S&P 400 (USA), OMXS (Sweden), Straits Times Index (Singapore), SMI (Switzerland), TSEC (Taiwan), Tel Aviv TA-100 (Israel)

[3] To conserve space, the results of the I(1) and heteroscedasticity tests are not printed.

We restrict the analysis to using the trace test as the maximum eigenvalue test leads to the same conclusions. The model used for testing is a simple VAR without intercept and one lagged term. As regards the cointegration relationship we use both the specification with and without intercept. P-values are calculated using the response surface tables of MacKinnon *et al.* (1999).

In the first case without intercept in the cointegration relationship (model 1), the trace test indicates that 46 out of the 378 combinations (i.e. 12.17 %) are cointegrated when performed on a 5% significance level. When adding an intercept in the cointegrating equation (model 2), we find that (based on the trace test) 36 out of 378 (i.e. 9.52 %) stock index combinations are cointegrated. As we have 2084 observations, we repeat the test on a more conservative significance level of 1%. We now find that 17 out of 378 combinations (4.50%) or 9 out of 378 (2.38%) combinations, respectively, are cointegrated. Hence we find a tendency to reject the null hypothesis of no cointegration about twice as often as the significance level would allow given the assumption that indices are not cointegrated is true.

Whether the null hypothesis of no cointegration is rejected or not also depends on the time period which is analyzed. Intuitively, this should not be the case: if a cointegration relationship existed and was identified between t and $t+1000$ days, then it should also be identifiable between $t+5$ and $t+1005$ days. In order to illustrate that this assumption does not hold in the context of international stock market indices, we restrict the dataset and perform the cointegration test on a window of 1042 observations moving through the full dataset in weekly steps (i.e. we have 208 times 378 tests). The result is graphically illustrated in Figure 4.1. The solid line gives the rejection rates based on model 1 while the dashed line represents the rejection rates based on model 2. The first observation depicted corresponds to the sample window starting 1st March 2001 and ending 25th February 2005 while the last corresponds to 24th February 2005 to 25th February 2009. Across all windows we find on average 20% of the stock market indices to be cointegrated, both based on model 1 and model 2. As the graph in Figure 4.1 stresses, the rejection rates are quite volatile when moving through the sample, ranging between 6 and 46%. This range is similar for both models used, but the results are usually quite different as is shown by the deviant pattern of the two lines. In the subsequent simulation study we will show that a possible explanation for this behavior lies in the presence of common stochastic trends, i.e. the global and local shocks, and heteroscedasticity in the error term in Equation (4.1).

In order to calibrate the simulation model in the following section, we also calculate pairwise correlation measures for the stock indices and the return series. We find that correlation between the indices is on average 0.7997 (with a standard deviation of 0.1741). Correlation is lowest for AEX and MerVal (0.1043) and highest for OSEAX and S&P TSX (0.9901). As regards returns, the average correlation is found to be 0.3758 (with a standard deviation of 0.2052). Correlation is lowest between ATS and KLSE returns (0.0366) and highest for DJIA and S&P 500 returns (0.9780).

4.4 A Simulation Experiment

In order to see whether the theoretical considerations in Section 4.2 are in line with the empirical findings in Section 4.3 we conduct a short simulation experiment. We simulate prices according to the model

$$p_{i,t} = p_{i,t-1} + \iota' e_t, \tag{4.9}$$

where e_t is a (3×1)-vector of (un-)correlated global, local, and stock specific innovations. ι is a (3×1)-vector of ones. The elements of e follow a normal distribution with $\mathbb{E}[e_s] = 0$ and $\text{Var}[e_s] = 1$. The magnitude of the variance does not influence the following results. The initial values $p_{i,0}$ are set to zero. The first 200 observations of the simulated price series are discarded.

Indices are then calculated as a weighted sum of the individual price series. In order to follow the simple model in Section 4.3, we use two price series to construct an index (subsequently referred to as two-stocks index). As stock market indices are never composed of two stocks only, we also construct two indices using 30 price series for each index (as in the DAX or the DJIA, for example; we will refer to this index composition as thirty-stocks index). In both cases the weights are $w_i = \frac{1}{n}$ where n is the number of price series used to calculate the index. The study is conducted for sample sizes of $T = 500$ and $T = 1000$ observations. The simulations are run with Gauss using the "KISS + Monster"-based random number generator and 10,000 replications. The Johansen (1991) test is conducted using only a model without drift and critical values are obtained using the response surface tables of MacKinnon *et al.* (1999).

4.4.1 The Benchmark Case

In case that the error term of the model in Equation (4.9) only contains an individual component (i.e. the second and the third element of e_t are zero), the resulting index is a weighted average of two $I(1)$ series. Hence, we expect the cointegration test, testing whether two such indices are cointegrated, to reject the null of no cointegration as often as implied by the significance level. As regards correlation, the two indices should not be correlated as the individual errors of the simulated prices are independent.

The results of the simulation support these assumptions. We first test the null hypothesis that the data have zero common stochastic trends (to which we will refer to subsequently as the null hypothesis of no cointegration) against the hypothesis that there are more than zero common stochastic trends. The rate of rejection of this null hypothesis corresponds to the significance level in all of the four cases (see Table 4.1 for details). The second hypothesis whether there is one common stochastic trend (tested against the alternative that there are more than one common stochastic trends) is only calculated if the first hypothesis is rejected. Rejection of this null hypothesis would indicate that the data are stationary. The rejection rates of this null hypothesis are very low (less than 1%). Our main focus, however, lies on the first hypothesis. As the rejection rates of the second hypothesis are quite low throughout all tests (usually lower than the significance level on which the test is conducted), we will not discuss these results subsequently.

As regards the sample correlation of the simulated indices, we find that it is on average close to zero. As the indices are two independent random variables, this is what we expected. However, this is not in line with the empirically found high correlation of the indices. We therefore relax the assumption on the individual errors and allow for some correlation there. We construct the time series such that the innovations' correlation varies between -0.6 and 0.6. The range is somewhat arbitrary, but we believe it is justifiable. One the one hand, we need high and positive correlation if the correlation between the indices is to go up at all. One the other hand, as the correlation approaches one, the individual prices would be identical. We, thus, require the correlation to be distinctly lower than one. Further, negative correlation has to be allowed as well because it would otherwise rule out any hedging possibilities. If all correlations would be set to exactly 0.6, the index correlation would be 0.6 as well. We believe, however, that this case is not empirically relevant. Table 4.1 (lower panel) reports possible results with random entries in the covariance matrix (which is assured to be positive definite in the simulation). What we see is that on average, correlation is fairly low, especially when using 30 stock prices to construct the index. The cointegration test is not affected by the correlation in the individual errors. Its rejection rates are still close to the chosen significance level. We therefore conclude that correlation between price innovations alone can not explain the empirical findings of high correlation and relatively too high rejection rates of the cointegration test. We believe that this is a strong hint to the presence of common factors which we will elaborate further in the following subsection.

4.4.2 The Model with Common Global and Local Components

We now add the common components to the model in Equation (4.9). As we consider a two-countries case, they represent a global stochastic trend which is common to all stock prices, and an area specific local stochastic trend which only concerns the stocks of one of the indices. The results are summarized in Table 4.2.

With the innovations being uncorrelated we find that the cointegration test still performs well in the sense that rejection rates are close to the chosen significance level. This means that the sheer presence of a stochastic trend which is common to both indices does not mislead the cointegration test. This shows that while the assumption of a common stochastic trend is sensible, the indices are not cointegrated.

Sample correlation of the two indices, however, rises to 0.36 for the two-stocks index or even 0.45 for the thirty-stocks index. The higher correlation is only due to the inclusion of the common random walk component. It basically constitutes a random variable present in all price processes which are therefore dependent. The latter is reflected in the higher correlation compared to the benchmark case.

In a second setting we again allow the errors to be correlated[4]. We assume that the global and local components are moderately positive correlated and

[4]The correlation matrices are constructed such that the fully specified model reflects the empirically found features while still being compatible with empirical findings (eg. Harvey, 1991)

that individual errors within one area j are correlated with each other as well as with the local component. In the latter case, correlation can be negative. The correlation between the individual innovations of different areas is still zero. In case of two constituents of an index, the covariance matrix Σ of the errors $e = (g, l_1, l_2, \varepsilon_{1,1}, \varepsilon_{1,2}, \varepsilon_{2,1}, \varepsilon_{2,2})$ is given as

	g	l_1	l_2	$\varepsilon_{1,1}$	$\varepsilon_{1,2}$	$\varepsilon_{2,1}$	$\varepsilon_{2,2}$
g	1.00						
l_1	0.60	1.00					
l_2	0.45	0.20	1.00				
$\varepsilon_{1,1}$	0.00	0.20	0.00	1.00			
$\varepsilon_{1,2}$	0.00	−0.30	0.00	−0.40	1.00		
$\varepsilon_{2,1}$	0.00	0.00	0.30	0.00	0.00	1.00	
$\varepsilon_{2,2}$	0.00	0.00	0.10	0.00	0.00	0.20	1.00

We assume that global and local components are dependent with covariance $\sigma_{g,l_1} = 0.6$ and $\sigma_{g,l_2} = 0.45$. The idiosyncratic shocks are independent from the global innovations (last four entries in the first column), but covary with the local component (second and third column). In country 1 one stock is negatively correlated with the local component and with the other stock. Individual innovations are independent across countries (e.g. columns four and five contain zeros in rows six and seven). Although the choice of the entries will influence the outcome in this subsection, the general conclusions are not affected. For the thirty-stocks index the structure of the covariance matrix is preserved (i.e. global and local components are still correlated while country 1 innovations and country 2 innovations are not). The entries, however, are random and can be positive or negative.

While the cointegration test still performs as expected, correlation between the two-stocks indices is up to 0.55, which is higher than in the previous model without correlation. For the thirty-stocks indices correlation is now lower by 0.05 to 0.1. The reason for this behavior is that there are two different sources inducing dependence: (1) the common random walk component which is present in every single price series (and, thus, induces correlation); and (2) correlation among the idiosyncratic innovations. So if the innovations were only positively correlated, correlation between the indices would rise. As we allow for negative correlation as well, the correlation between the indices can be lower than in the first case with uncorrelated errors.

Although in the two prices case, correlation is higher than if the errors were not correlated, while being even a little lower in the thirty prices case, the magnitude is still considerably lower than in the empirical study. Of course this is to some extent due to the way the covariance matrix is defined. Higher relative covariance values, however, are not plausible as the series would almost be identical as correlation goes to 1. So the presence of common random walk components as well as correlated errors alone cannot sufficiently explain the empirically found high correlation.

4.4.3 The Model with Individual Heteroscedastic Errors

So far the simulation has not been able to reflect the tendency to overreject the null hypothesis of no cointegration which has been found in the empirical analysis in Section 4.3. A decisive feature of financial time series, namely heteroscedasticity, has been neglected so far as well. We therefore model the individual errors in e of Equation (4.9) as GARCH(1,1) processes according to

$$\begin{aligned} e_{i,t} &= \sqrt{h_{i,t}}\,\nu_{i,t} \\ h_{i,t} &= 0.01 + \gamma e_{i,t-1}^2 + \omega h_{i,t-1} \\ \nu_{i,t} &\sim N(0,1) \quad . \end{aligned} \tag{4.10}$$

The parameters of the GARCH-model vary: ω follows a uniform distribution (between 0.90 and 0.98 and $\gamma = 1-\omega-0.01$). The model thus exhibits the commonly documented pattern of high volatility persistence (eg. Akgiray, 1989) while the variance process itself is assured to be stationary. In the basic setting we simulate the model free of common components and let the errors be independent. As in the benchmark case, we find that correlation is on average close to zero. At the same time the test for cointegration becomes less reliable. The trace test rejects the null hypothesis that the time series are not cointegrated about two to three times as often as suggested by the chosen significance level (cp. Case 1 in Table 4.3). So the presence of heteroscedasticity in the level price series seems to mislead the Johansen test. This is mainly due to the high volatility persistence. However, if we lower ω such that it varies between 0.55 and 0.65, the cointegration test performs well within the expected limits.

In the second setting we add the common global and local components while the errors $\nu_{i,t}$ are still uncorrelated. Again we find the tendency of the cointegration test to slightly overreject the null hypothesis of no cointegration (cp. Case 2 in Table 4.3). This level is also similar to what we found in the empirical example. The correlation between the simulated indices increases due to the common random walk component in both index series. It varies around 0.4, which is higher than in the benchmark case, but still lower than in the empirical example.

In the third setting we allow the individual errors to be correlated as in Subsection 4.4.2. We now find features within the simulated data which are similar to those found in the empirical analysis: the correlation is high (between 0.7 and 0.85) and the rejection rate of the null hypothesis that the indices are not cointegrated is approximately two to three times higher than the significance level would allow for (cp. the lower part of Table 4.3). The rise of the correlation is due to the additional source of dependence among the errors which is induced through the structure of the covariance matrix.

Variation of the parameters in the GARCH-model shows that as ω increases, the rejection rates of the null of no cointegration of the Johansen trace test increase. At the same time, the correlation measure diminishes slightly. So the more persistence there is in the variance equation, the less reliable the Johannsen methodology seems to be.

4.4.4 An Example containing a Drift Term

We now add a drift term δ to the model in Equation (4.9) and simulate prices according to

$$p_{i,t} = \delta + p_{i,t-1} + \iota' e_t \; . \tag{4.11}$$

The magnitude of δ is modeled as one standard deviation divided by 50 times the number of observations in the sample ($\delta = \frac{\sigma}{50T}$). It has been determined empirically using the index data. We estimate an AR(1)-model with drift term and compare the size of this estimate with the standard deviation of the data. The drift term turns out to be on average 65 times smaller than the standard deviation of the sample.

Table 4.4 holds the results for the different settings. Case 1 (uncorrelated innovations and no common components) is not printed to conserve space. The general conclusions concerning cointegration which were presented in the previous subsections, still hold. However, with respect to sample correlation the results are different. The correlation ranges now from 0.9 to 0.99 in most cases which is a lot higher than in the case without drift and also higher than what we found using real world data. We therefore conclude that the market indices in Section 4.3 can best be described by models which do not contain a drift term.

4.5 Concluding Remarks

The Chapter shows that under the assumption that stock prices follow the common random walk model, international financial markets cannot be cointegrated in the sense of Engle and Granger (1987). Cointegration is eventually inhibited by company specific innovations which are permanently absorbed into stock prices. These individual random walk components do not cancel in a cointegration regression.

In a simulation study we model the typical features of financial assets (correlation, heteroscedasticity) in order to replicate the characteristics of real world data. We find hints that the combination of correlated innovations, common random walk components, and heteroscedasticity describes those features best. The common component as well as correlation of the errors mainly drive the empirically found high correlation while heteroscedasticity leads the Johansen cointegration test to slightly overreject the null hypothesis of no cointegration. In the absence of heteroscedastic errors in the simulation, however, correlation is also lower than in the empirical example. So according to our model, global financial markets most probably do share at least one common stochastic trend, the global trend. This trend, however, cannot be identified by means of cointegration analysis. A feature showing that the market indices are interrelated is the relatively high correlation among them. It is considerably higher when common global and local stochastic trends are present than in their absence. High correlation, of course, is due to the fact that the same stochastic component appears in each of the individual stock prices.

Appendix: Heteroscedasticity in Returns and Levels

Financial research suggests that a model for stock returns should account for time dependence of the variance. A possible autoregressive model for stock returns is

$$
\begin{aligned}
r_t &= ar_{t-1} + \varepsilon_t \sqrt{h_t} && (4.12)\\
h_t &= \omega + \alpha \varepsilon_{t-1} + \beta h_{t-1}. && (4.13)
\end{aligned}
$$

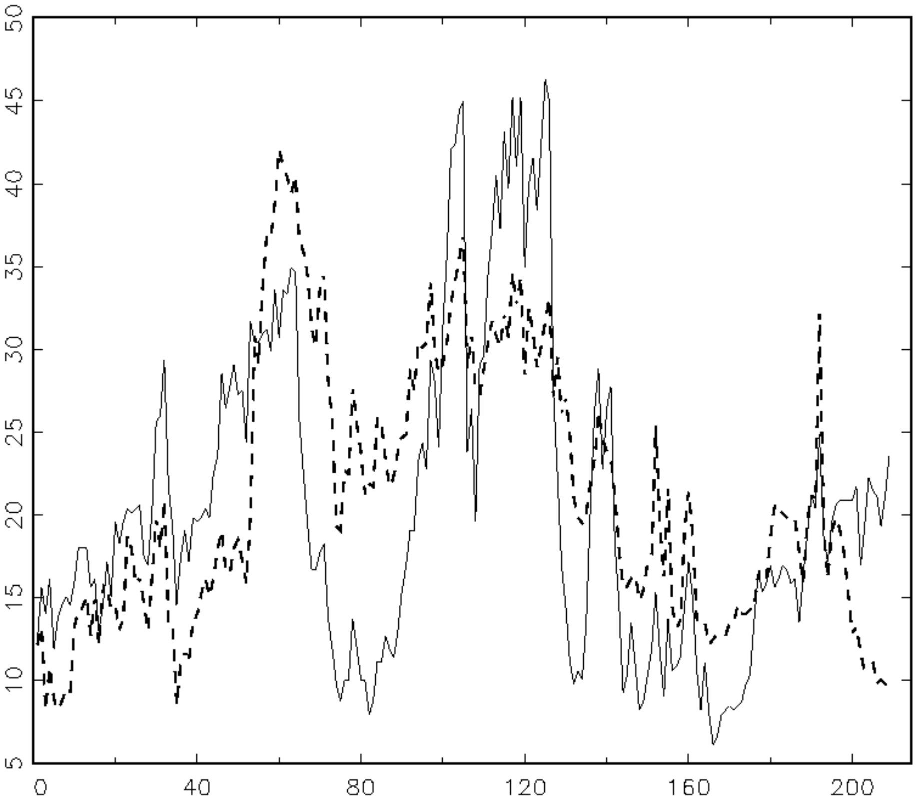
According to the strong form Efficient Market Hypothesis, $a = 0$. As log-returns are calculated as the first difference of the log-prices

$$
r_t = p_t - p_{t-1}, \tag{4.14}
$$

the random walk model for log-prices may also be marked by time varying variance in the error terms:

$$
\begin{aligned}
p_t &= p_{t-1} + r_t && (4.15)\\
&= p_{t-1} + \varepsilon_t \sqrt{h_t} && (4.16)
\end{aligned}
$$

Figure 4.1: Rolling Cointegration Test - Rejection rates of $H_0 : r = 0$



The graphic depicts the rejection rates of the null hypothesis that the rank of the cointegration matrix is zero resulting from the rolling cointegration test. The solid line presents rates based on the model without intercept in the cointegration relationship while the dashed line presents the rates based on the model with intercept in the cointegration relationship. Tests were conducted on a 5% significance level.

Table 4.1: Benchmark case

number of observations	significance level	number of prices in X_j	$H_0 : r = 0$ (percent)	$H_0 : r = 1$ (percent)	sample correlation
Case 1: uncorrelated innovations					
500	0.05	2	5.6400	0.6041	0.0042
500	0.05	30	5.1900	0.5379	-0.0020
1000	0.01	2	1.0900	0.0303	0.0133
1000	0.01	30	0.9300	0.0303	-0.0001
Case 2: correlated innovations					
500	0.05	2	5.3900	0.5813	0.1440
500	0.05	30	5.1200	0.4005	0.0013
1000	0.01	2	1.1500	0.0303	0.3433
1000	0.01	30	0.9600	0.0202	-0.0013

The table presents simulation results for the benchmark case of the model in Equation (4.9). $H_0 : r = 0$ and $H_0 : r = 1$ are the null hypotheses that the rank of the cointegration matrix is zero or one, respectively, in the Johansen trace test; the table reports rejection rates of these hypotheses; the second test is only performed once the first null hypothesis has been rejected.

Table 4.2: Model with global and local component

number of observations	significance level	number of prices in X_j	$H_0 : r = 0$ (percent)	$H_0 : r = 1$ (percent)	sample correlation
Case 1: uncorrelated innovations					
500	0.05	2	4.7900	0.4726	0.3581
500	0.05	30	5.3700	0.5389	0.4504
1000	0.01	2	1.0500	0.0000	0.3685
1000	0.01	30	1.0200	0.0505	0.4525
Case 2: correlated innovations					
500	0.05	2	5.0100	0.4737	0.5428
500	0.05	30	5.0100	0.5158	0.6144
1000	0.01	2	0.9500	0.0505	0.5507
1000	0.01	30	1.1100	0.0202	0.5565

The table presents simulation results for the model in Equation (4.9) where the error term consists of a global, a local, and an idiosyncratic component. They may be uncorrelated (case 1) or correlated (case 2). $H_0 : r = 0$ and $H_0 : r = 1$ are the null hypotheses that the rank of the cointegration matrix is zero or one, respectively, in the Johansen trace test; the table reports rejection rates of these hypotheses; the second test is only performed once the first null hypothesis has been rejected.

Table 4.3: Model with heteroscedastic errors

number of observations	significance level	number of prices in X_i	$H_0 : r = 0$ (percent)	$H_0 : r = 1$ (percent)	sample correlation
Case 1: no common components and uncorrelated innovations					
500	0.05	2	11.0000	5.7079	-0.0027
500	0.05	30	11.1800	6.7327	0.0025
1000	0.01	2	2.7600	0.6787	0.0099
1000	0.01	30	2.5200	0.5950	-0.0097
Case 2: common components and uncorrelated innovations					
500	0.05	2	12.2000	5.9909	0.3552
500	0.05	30	12.4400	5.8931	0.4440
1000	0.01	2	2.8400	0.7616	0.3489
1000	0.01	30	2.9600	0.6801	0.4465
Case 3: common components and correlated innovations					
500	0.05	2	11.9100	6.3571	0.8081
500	0.05	30	11.4900	5.4570	0.5827
1000	0.01	2	3.2000	0.7128	0.8233
1000	0.01	30	3.3500	0.8795	0.7294

The table presents simulation results for the model in Equation (4.9) where the error term consists of a global, a local, and an idiosyncratic component. Each component follows an individual GARCH(1,1) process. They may be uncorrelated (case 1) or correlated (case 2). $H_0 : r = 0$ and $H_0 : r = 1$ are the null hypotheses that the rank of the cointegration matrix is zero or one, respectively, in the Johansen trace test; the table reports rejection rates of these hypotheses; the second test is only performed once the first null hypothesis has been rejected.

Table 4.4: Model with drift

number of observations	significance level	number of prices in X_j	$H_0 : r = 0$ (percent)	$H_0 : r = 1$ (percent)	sample correlation
Case 2: correlated innovations, no common components					
500	0.05	2	5.8000	0.9766	0.9818
500	0.05	30	4.8300	0.6725	0.9985
1000	0.01	2	1.0300	0.1212	0.9911
1000	0.01	30	1.0700	0.0708	0.9994
Case 3: uncorrelated innovations, common components					
500	0.05	2	6.3900	1.6451	0.9011
500	0.05	30	6.0800	1.4267	0.9332
1000	0.01	2	1.2700	0.0912	0.9613
1000	0.01	30	1.0600	0.1011	0.9744
Case 4: correlated innovations, common components					
500	0.05	2	7.4500	2.4203	0.8746
500	0.05	30	6.6200	2.0561	0.8906
1000	0.01	2	1.2300	0.1721	0.9488
1000	0.01	30	1.1300	0.1719	0.9541
Case 5: correlated innovations, common components, heteroscedastic errors					
500	0.05	2	7.9100	2.8885	0.9412
500	0.05	30	7.6300	2.4467	0.9192
1000	0.01	2	1.5100	0.2132	0.9731
1000	0.01	30	1.4500	0.2131	0.8964

The table presents simulation results for the model in Equation (4.11). $H_0 : r = 0$ and $H_0 : r = 1$ are the null hypotheses that the rank of the cointegration matrix is zero or one, respectively, in the Johansen trace test; the table reports rejection rates of these hypotheses; the second test is only performed once the first null hypothesis has been rejected.

Chapter 5

The Impact of US News on the German Stock Market

5.1 Introduction

When looking at a graph plotting high frequency observations of the DAX index, a quite frequently observable feature is a jump of the DAX around 2.30 p.m. or 3.30 p.m. These times correspond to the time when macroeconomic news are usually announced in the USA and the opening of the New York Stock Exchange (NYSE), respectively. Figure 5.1 presents some examples when the DAX dropped or rose quite substantially at these times. However, on a great number of days the behavior of the DAX index is smoother in the sense that it does not exhibit such jumps (see the examples in the last row of Figure 5.1). This paper seeks to show that the different behavior of the DAX index on these days is driven by unexpected news. It relies on event study methodology to evaluate whether a significant influence of US generated information on the German DAX does exist at all and to estimate its size. Further, it addresses the question how and how fast this news is processed by the German market and its impact on the general level of volatility in the market.

In this respect, the paper links two strands in the literature. The first strand deals with information and volatility transmission between financial markets and is usually referred to as spillover literature. This literature analyses the impact of news generated in foreign markets on the home market. It dates back to three important papers by Hamao *et al.* (1990), Susmel and Engle (1994) and Lin *et al.* (1994). Hamao *et al.* (1990) study the interdependence between the Tokyo, London, and New York markets and find, amongst other, evidence of price volatility spillovers from New York to Tokyo and to London. Susmel and Engle (1994) explore spillovers between the London and New York stock markets on an hourly basis using ARCH models. They find that the spillovers are most pronounced around the opening of the NYSE. Lin *et al.* (1994) analyse the relationship between the Tokyo and New York markets and identify weak influence of open-to-close returns on close-to-open returns. These studies rely essentially on GARCH models and use low frequency data (up to hourly returns). More recent contributions to this literature are Booth, Martikainen, and Tse (1997) and Baur and Jung (2006). Diebold and Yilmaz (2009) and Dimpfl and Jung (2007) also analyse mean and volatility spillovers. However, instead of using (G)ARCH-models these authors rely on VAR-models and propose different methods to account for time varying second moments.

The second strand in the literature connected to the present paper is the study of the impact of news announcements. The outstanding characteristic of this literature is the widespread use of event study methodology. The first work associated with this field is Dolley (1933) who examined the price effect of stock splits. Since then, the methodology has been further developed and refined. An early contribution is Ederington and Lee (1993) who investigate the impact of news announcements on interest rates and foreign exchange futures and find that these announcements have an important impact on daily and weekly volatility. More recent studies in this area include Muntermann and Guettler (2007) and Kerl and Walter (2007). The former investigate intraday effects of ad hoc disclosures on German stocks and find that stock prices react within half an hour to the announcements. Kerl and Walter (2007) analyse the impact of personal finance magazines' buy recommendations on German stocks. They find that these recommended stocks earn significant abnormal returns within five days after publication. Adams, McQueen, and Wood (2004) determine the reaction of high frequency stock returns on inflation news. However, they do not rely on the event study technique as such but estimate a return model with time dummies for the event of interest. Surprisingly, they do not find an instantaneous reaction of stock prices to unanticipated news announcements. Hess (2004) also relies on dummy variables to identify the determinants of unanticipated macroeconomic news announcements on T-bond futures. Similarly, Hess, Huang, and Niessen (2008) measure the effect of macroeconomic news on commodity futures. In general, these studies are designed such that they evaluate the impact of local events on local stock markets, individual stocks, futures or commodities.

The present paper contributes to the literature by combining the two strands in order to shed light on the intraday information transmission from the US to the German stock market. We investigate whether the occasional jumps of the DAX in the early afternoon trading are information driven or whether observing jumps at this point in time is merely coincidental. The paper therefore introduces event study techniques to the spillover literature and extends the use of this methodology to the analysis of events which took place in a foreign market on the home market. An important difference to the traditional spillover literature lies in the way the spillover effect is measured. When using GARCH- or VAR-models, the significant parameter estimates indicate the spillover effect and its direction, the absolute value of the estimate indicates its magnitude. In contrast, the present paper seeks to quantify the impact of information spillovers in terms of abnormal, i.e. unexpected index returns as compared to a still to be defined normal, i.e. expected return.

More precisely, we address the following questions: Does the opening of the NYSE per se contain information which is valuable to investors in Germany? This could be the case if German investors await the valuation of news by US investors and act only subsequently. The second question we ask is about the impact of news announcements which on a regular basis take place before the opening of the NYSE. Do German investors take advantage of the fact that the market is open and act immediately after new information is released?

The motivation for the first question is that if important news is announced in the USA, German investors might wait and see how their US counterparts process this information. One might assume that US investors can interpret information about the US economy or US companies more accurately and, thus, judge their price impact more precisely. The reason is that they are closer to the market and therefore have more insight into the functioning of the US economy as well as into the trading mechanism at a US stock exchange. When information of global importance (such as US unemployment figures or interest rate changes, for example) is released, its long-term impact on stock prices needs to be assessed and US trading agents might have an informational advantage. If this is the case it might be rational for cautious German traders to await the reaction of US traders to such news announcements. In the end, such information is still local information albeit its possible global relevance. In this situation the impact of the news announcements should, thus, not be identifiable before the opening of the first US stock market. In other words, the reaction of the German stock market to such news should only take place at or after 3.30 p.m. Central European Time (CET). If this is the case, our results would be in line with findings of King and Wadhwani (1990) who show that the UK stock market does not immediately react to US macroeconomic news announcements. We will refer to this situation as hypothesis (i).

In case that the price impact of the released information is obvious and easily interpreted, German investors should take advantage of the fact that the German market is already open at the time of the news release. This suggests that they would react immediately after the announcement instead of awaiting the actions of their US counterparts. So hypothesis (ii) states that the price impact is measurable immediately after US macroeconomic news is usually released (around 2.30 p.m. CET). Recent findings of Andersen, Bollerslev, Diebold, and Vega (2007) show that this is the case for the UK stock market and, thus, contradict the results of King and Wadhwani (1990). Of course, the two studies differ in terms of the applied methodology.

We believe that the event study methodology is a very appropriate way to analyse this issue for the German stock market because it allows to account for normal or expected reactions of the stock market. The two hypotheses as formulated above imply different abnormal return patterns of the DAX in the Frankfurt afternoon trading. The first hypothesis would implicate that abnormal returns arise very closely around the opening of the NYSE, i.e. around 3.30 p.m. CET. The second hypothesis, however, suggests that abnormal DAX returns will be observed already one hour before the opening of the NYSE, i.e. around 2.30 p.m. CET. Nikkinen and Sahlström (2004) have also addressed the question how valuable US macroeconomic announcements are to German and Finish investors. The authors use the methodology of Ederington and Lee (1993) and find a significant impact on implied volatility which stems from US information whereas local information seems to be unimportant. Although the basic idea of their paper is similar, the implementation is different. We measure the impact of the news release within the trading day on the return distribution of the DAX index whereas Nikkinen and Sahlström (2004) analyse implied volatility of the whole trading day. We find an abnormal return pattern around 2.30 p.m. and, hence, conclude that German investors immediately react to US news announcements which precede the opening of the New York Stock exchange. The opening of the market itself and the beginning of trading in the USA is not found to affect German stock prices. On average days, there is no measurable impact on the DAX.

If the identified news events are unforeseen, there will be valuation insecurity in the market, especially in the early afternoon trading. In this case, the risk of trading is higher and, thus, we expect volatility to be higher than on quiet days. We therefore conclude the study by testing whether volatility is different on days with announcements from days without announcements. We find that volatility is generally higher on announcement days (irrespective of whether good or bad news are transmitted). These days are marked by a generally higher level of volatility also in the morning, but the increase in the afternoon is still significant.

The Chapter is organized as follows. Section 5.2 outlines the event study methodology. Section 5.3 describes the data along with the mechanism used to identify events and to classify them into positive and negative ones. Section 5.4 provides the empirical results and interpretations and section 5.5 concludes.

5.2 Methodology

In order to measure the price impact of the opening of the NYSE and the preceding news announcements on the German stock market, an event study will be designed as follows. First, we define the event depending on the two hypotheses. If the hypothesis that the opening of the NYSE contains information which is valuable for German investors is true, the event takes place at $\tau_{H(i)} = 3.30$ p.m. CET. If, however, the hypothesis that the impact is due to news announcements is true, the event takes place at $\tau_{H(ii)} = 2.30$ p.m. CET. Figure 5.2 illustrates the timing of the event study. In any of the two cases we would expect an abnormal return closely around these times.

An abnormal or unexpected return is defined as the actual return over the event window minus the expected or normal return over this period. So we define

$$\varepsilon_{it} = R_{it} - \mathbb{E}\left[R_{it} \mid X_t\right] \tag{5.1}$$

where ε_{it}, R_{it} and $\mathbb{E}\left[R_{it}\right]$ are abnormal, actual, and normal returns, respectively, on day i at time t. The underlying time unit to conduct the analyses and to estimate the models will be one minute, i.e. the interval $[t, t+1)$ corresponds to one minute. X_t is the conditioning information for the normal performance model. We propose to use either the so-called constant-mean-return model or an autoregressive model of order p (AR(p)) as information set. The first model assumes the return during a specific trading day i to be constant, i.e.

$$R_{it} = \mu_i + \nu_{it} \; . \tag{5.2}$$

This model essentially amounts to estimating the mean return over the estimation window. In the present case the beginning of the estimation window is set to $t = T_0 = 10.30$ a.m. CET only in order to exclude any overnight valuation effects on the DAX return distribution and to exclude volatility effects as implied by the well-documented volatility smile. Asian news events should also be processed and priced by 10.30 a.m. already. The estimation window ends at $t = T_1 = 1.30$ p.m., i.e. two hours before the opening of the NYSE and only 7.30 a.m. EST. This should assure that we include as much information as possible in the estimation of the normal DAX returns while simultaneously avoiding the possible influence of US events. Stopping at 1.30 p.m. also reduces volatility influences which have recently been documented by Masset (2008) who shows the volatility pattern of the DAX to be W-shaped with a spike at 2.30 p.m. (see also Figure 5.5 and the discussion of volatility below). To check the robustness, the estimation window has been extended to $T_1 = 2.00$ p.m. and shortened to $T_1 = 1.00$ p.m. The results presented below are robust to this alteration.

As an alternative to the mean model we also propose an AR(p) model in order to account for possible market microstructure effects:

$$R_{it} = \mu_i + \sum_{k=1}^{p} \beta_k R_{i,t-k} + \nu_{it} \; . \tag{5.3}$$

It is estimated separately for each day i with individual lag length p determined by the Bayesian Information Criterion of Schwarz (1978). Of course, in case that $p = 0$ the two models coincide.

For the estimation to be valid we need to assume trading days to be independent, i.e. ν_{it} and ν_{jt} are independent for all $i \neq j$. We believe that this is not an issue as we are working on high frequency data. If ν_{it} and ν_{jt} were dependent, this would imply an effect taking place every day at exactly time t, say, for example, at 11.32 a.m. We are unaware of any such systematic and regular event in the Frankfurt morning trading[1].

The next step is to measure and analyse abnormal returns. In case of the constant-mean-return model in Equation (5.2) abnormal returns are calculated

[1] The Intraday Auction at 1 p.m. CET is a technical feature of the trading at the Frankfurt Stock Exchange and does not provide any relevant information itself.

as

$$\hat{\varepsilon}_{it} = R_{it} - \hat{\mu}_i \, . \tag{5.4}$$

In the AR(p) model in Equation (5.3) abnormal returns are defined as the difference between actual returns in the event window and predicted returns of an s-steps out-of-sample forecast of the model. Out-of-sample in this context means that the forecast is based solely on data in the estimation window. A dynamic forecast is not suitable because it would mix information from two different information sets. Hence,

$$\hat{\varepsilon}_{it} = R_{it} - \hat{R}_{it} \tag{5.5}$$

where $\hat{R}_{it}$ are appropriately forecasted returns (in contrast to the estimated $\hat{\mu}_i$ in Equation (5.4)). Abnormal returns are then aggregated within a day to calculate cumulated abnormal returns $CAR_i = \sum_t \hat{\varepsilon}_{it}$.

In order to test whether the measured abnormal returns are significantly different from zero, we rely on two most commonly used test statistics which differ in the required statistical assumptions about the abnormal returns. The first is a standard cross-sectional test (note that the cross section in this context are the different days i). In order to be valid it requires that abnormal returns are normally distributed and that there is no cross-sectional dependence in abnormal returns while the event may influence the variance. It is given as

$$t_{cs} = \frac{\frac{1}{N}\sum_{i=1}^{N} CAR_i}{\sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}\left[CAR_i - \frac{1}{N}\sum_{i=1}^{N} CAR_i\right]^2}} \, . \tag{5.6}$$

The second test statistic has been developed by Boehmer, Masumeci, and Poulsen (1991). In contrast to Equation (5.6) it uses standardized cumulated abnormal returns to ensure that all the CARs have unit variance. This procedure allows for consistent estimation of the standard deviation in the denominator if the event induced variance differs across days. Define $SCAR_i$ as the cumulated abnormal returns on day i divided by an estimate of their standard deviation (see, for example, Campbell, Lo, and MacKinley, 1997, for details).

The test statistic is then calculated as follows:

$$t_{BMP} = \frac{\frac{1}{N}\sum_{i=1}^{N} SCAR_i}{\sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}\left[SCAR_i - \frac{1}{N}\sum_{i=1}^{N} SCAR_i\right]^2}} \,. \tag{5.7}$$

Boehmer *et al.* (1991) show that their test statistic is robust to variance changes induced by the event. Both t_{cs} and t_{BMP} are approximately standard normally distributed under the null hypothesis that the event does not have an impact on the return distribution.

In order to conduct the volatility analysis we rely on realized volatilities as introduced by Andersen *et al.* (2003). We use five-minute intervals and calculate realized volatility measures as follows:

$$\sigma^2_{i,\Delta} = \sum_{j=1}^{1/\Delta} R^2_{i-1+t\Delta,\Delta} \tag{5.8}$$

where Δ is the time interval and $R^2_{i-1+t\Delta,\Delta}$ are log-returns on day i (in percent) within the respective time horizon. We construct a measure for the morning (from 10.30 a.m. to 12.30 p.m.) and for the afternoon (from 2.00 to 4.00 p.m.) volatility. As for the event study we exclude the period of overnight insecurity and the end of the trading day as we are only interested in the effect of the opening of the US market. To test whether the measures actually differ, we use the Wilcoxon signed-rank test (see Gibbons and Chakraborti, 2003, pp.196ff) as realized volatilities are by construction not normally distributed and we don't want to impose any assumptions.

5.3 Data and Event Identification

The study is conducted using DAX data obtained from Tick Data. The sample covers high frequency DAX index observations from July 2003 to August 2008. The data have been resampled to one minute intervals in order to compute log-percentage-returns. Subsequently, the sample is split by days and 2.30 p.m. and 3.30 p.m. CET are marked as possible event times. The lead or lag of one hour in spring and autumn when times are switched to or from daylight-saving time, respectively, is taken into account.

The present paper does not rely on an external definition of event dates. To identify these dates and to distinguish between days with good news events and bad news events we rely on the S&P 500 index close-to-open return. It serves both as an indicator that a news event took place and simultaneously as a proxy for the quality of the event. The S&P 500 data are also obtained from Tick Data.

We assume that any kind of news event taking place in the USA while it might have an impact on the German DAX will definitely have an impact on the S&P 500. If the event is indeed of global importance we expect a substantial reaction of the S&P 500 which should translate into a high close-to-open return. As this index contains 500 individual stocks from different sectors it seems reasonably broad to capture globally relevant events while reducing the weight of sector specific local events which would otherwise blur our identification. Further, a strong reaction of the S&P 500 should only be observed if the announced content is surprising. Extreme close-to-open returns should, therefore, occur if and only if a surprise in the information flow has occurred some time before the opening of the NYSE. So by detecting the extreme events mirrored by the S&P 500 index, we seek to identify the arrival of surprising information. In particular, we do not discriminate between the type of information (e.g. political or economic).

While all information which has accrued during the night and in the morning will be reflected in the S&P 500 close-to-open return, we need to identify those days where an extreme return is driven by US news only. Although the US stock markets are largely autonomous in terms of information generation and processing (see, for example, Diebold and Yilmaz, 2009), we need to make sure that information which originates from Europe does not influence the identification procedure. To filter the S&P 500 close-to-open returns we fit an AR(1)-GARCH(1,1) model and include the DAX close-to-open return as an additional explanatory variable in the mean equation. The model reads as follows:

$$
\begin{aligned}
r_{SP,i} &= \mu + \beta_1 r_{SP,i-1} + \beta_2 r_{DAX,i} + e_i \\
e_i &= \sqrt{h_i}\,\varepsilon_i \\
h_i &= \omega + \alpha e_{i-1}^2 + \gamma h_{i-1} \qquad (5.9)
\end{aligned}
$$

where $r_{SP,i}$ and $r_{DAX,i}$ are close-to-open returns of the S&P 500 and the DAX index, respectively, and ε_t follows a $t-$distribution with m degrees of freedom. We use a t-distribution because a test of the $\hat{\varepsilon}_t$ rejected the normality assumption. The estimated degrees of freedom are $1/0.2706 \approx 4$ (see Table 5.2) and, hence, support the choice of a t-distribution. The DAX close-to-open return accounts for any non-US information which has accrued in Asia and in Europe (until the opening of the Frankfurt market at 9 a.m. CET) while stock markets in the USA were closed. So any innovation ε_i in $r_{SP,i}$ should be due to US information only. Estimation results of the GARCH model are summarized in Table 5.2. The estimates are in line with findings of other papers which use GARCH models with financial data. They are not discussed any further as this model is only an auxiliary estimation to identify event days.

Day i is subsequently labelled 'good news day' if the residual $\hat{\varepsilon}_i$ resulting from the estimation of Equation (5.9) exceeds a certain threshold. This threshold is defined as the q-th quantile of the residual distribution. We, thus, identify a good news day if $\hat{\varepsilon}_i > \varepsilon^{(1-q)}$ and a bad news day if $\hat{\varepsilon}_i < \varepsilon^q$. All other days are marked as average with no particular incidents. Days where the US markets were closed while there was trading in Europe were removed from the sample because there might have been a news event but our algorithm cannot identify it. A similar identification strategy has already been applied by Fabozzi, Ma, Chittenden, and Pace (1995). These authors, however, define the threshold explicitely as a return of 2%.

Descriptive statistics of the S&P 500 close-to-open and the DAX intraday returns are given in Table 5.1. Note that the number of S&P 500 close-to-open returns and the number of days in the DAX dataset are different. To determine the conditional return quantiles of the S&P 500 the complete dataset has been used. More precisely, days where the Frankfurt market was closed completely or closed before 14.30 p.m. CET while there was trading in New York are included. For the subsequent event study, days where there was no trading in Germany at 14.30 p.m. were eliminated from the DAX dataset (the last day of the year, national holidays) as well as days without trading in New York. As can be seen in the first panel of Table 5.1, the identified event days differ quite substantially from the average across all days in terms of the S&P 500 close-to-open return (which is higher by a factor 45 on days with positive news announcements and lower by a factor -45 on days with negative news announcements). A similar pattern is true for the DAX. On days with negative announcements the difference is even more pronounced.

In order to check the validity of the event identification procedure we determine events which actually took place on the dates marked good news day or bad news day. It turns out that the identification is quite successful. To mention only a few, reconsider the DAX plots of the introductory example in Figure 5.1. The upper panel shows the development of the DAX value on Friday, 3 October 2003 and on Friday, 2 April 2004. Both dates are marked by the announcement of positive data about the US job market. On 2 April 2004, for example, the general economic outlook turned out to be good. The US Department of Labor announced that the number of jobs created rose considerably more than expected: 308,000 jobs (without agricultural sector) had been created while only 103,000 new jobs had been expected by analysts.

The graphs in the second row show the DAX value on Friday, 6 August 2004 and on Friday, 27 October 2006. On 6 August 2004 the US job market turned out to be less dynamic than expected. The Department of Labor disclosed figures that only 32,000 new jobs had been created while 228,000 had been expected. Further, the oil price reached a new peak and, thus, a slowdown in economic growth became quite likely. The matching procedure marked this day as a bad news day. 27 October 2006 was again characterized by US GDP figures which were disappointing as analysts said. However, it is not marked as a bad news day by our procedure. This is not surprising as the reaction in general was quite weak with the Dow Jones losing 0,60% and the S&P 500 losing 0,85%. In Europe the markets did not react substantially either: the Euro Stoxx 50 closed trading with a loss of 0,25% and the DAX lost 0,34% and closed with 6262,54 points.

For the remaining days identified as news days we are in most cases able to trace back which news was announced. It generally consists of job market data or general economic indicators. In a few cases it was also political information like George W. Bush winning the 2004 presidential election on 3 November 2004. We therefore believe that our approach is valid and viable to identify US American (news) events of global importance. Applying this procedure we are convinced that we do not have an endogeneity problem as we effectively first identify the event (although through means of an empirical identification). It turns out that a day which is not within the $q = 5\%$ quantile but where the DAX still exhibits a sharp increase or decrease around 2.30 or 3.30 p.m. is then in the $q = 10\%$ quantile. However, extending the quantile also includes quite a number of days where a specific and significant event cannot be identified.

This happens scarcely when the quantile is set to $q = 5\%$.

5.4 Empirical Results

In order to test the hypotheses (i) and (ii) as stated in the introduction we calculate abnormal returns for different event windows. The results presented below are based on the AR(p) model. Lag length varies between $p = 1$ and $p = 10$, but short specifications dominate. As regards sensitivity to the modeling, the results of the mean model in general point to the same conclusions although the estimated impact is less pronounced in some cases[2]. This is, of course, due to the short lag length and the fact that the estimated autoregressive parameters are also small in absolute value. So the impact of past returns vanishes quickly and the forecast will converge to the estimated intercept which is similar to the intercept of the mean model.

5.4.1 How does the DAX depend on the US?

The first hypothesis suggests that the opening of the NYSE contains itself information which is valuable to investors in Germany. Under this hypothesis we expect an abnormal return behavior around 3.30 p.m. CET. In a first step we therefore set the event window to 3.30 to 3.35 p.m. CET. The result of this proceeding is summarized in Table 5.3. The outcome does not support the hypothesis. First, for all possible news categories the average cumulative abnormal returns (CARs) are quite small in absolute value. Any of the test statistics suggests that they are not statistically significant. Further, the sign of the estimated CARs on good and bad news days is the opposite of what we would have expected. So we are inclined to reject the hypothesis that the opening of the NYSE per se provides valuable information to German investors.

It might be, however, that some information which is generated during the opening auction of the NYSE is already disseminated and translated into prices in Germany. We therefore redefine the event window to include 10 minutes before and after the opening of the NYSE. The results are given in Table 5.4. CARs are on average negative, negative on good news days and positive on bad news days. Again, they are small in absolute value. The t-statistics suggest

[2]Detailed estimation results of the mean model are available upon request.

that they are significant on average days as well as on days with positive news announcements, but not on days with negative news announcements. Together with the previous results we conclude that the opening of the NYSE does not per se contain valuable information for German investors. The signs of the CARs, however, may be a hint that there is some kind of reversal effect around 3.30 p.m. CET which would be compatible with hypothesis (ii).

The second hypothesis states that the news releases which take place roughly one hour before the opening of the New York market are responsible for the observed jumps in the DAX. In this case we need to address two questions: do the news releases affect trading in Germany? And if so, how fast is the reaction to the information, given that it is not observable any more when the NYSE opens? To address the first issue, we enlarge the event window substantially from 2.30 to 3.30 p.m. CET, the hour before the NYSE opens for trading. The calculated CARs and tests are presented in Table 5.5. We find that on average the cumulative abnormal returns in this period are zero. However, on days with good news announcements we have a significant positive CAR of 0.25 percentage points. On negative days, the CAR is of -0.13 percentage points. As regards the absolute value of the estimates, they may not seem too important, but one has to bear in mind that these figures give abnormal returns within one hour of trading. If this behavior would prevail the whole day, i.e. 8.5 hours, one could expect an average abnormal return of 2.13% on good news days or -1.09% on bad news days. Considering that the average daily DAX return is 0.023% and the average daily absolute return in the sample is 0.68% this is not a neglectable amount. In terms of index points, at a DAX level of 8000 the abnormal return on good announcement days corresponds to an additional gain of 20 index points. On days when negative news are announced, the associated abnormal loss is 10 index points at an index level of 8000 points. On average, the abnormal return is zero.

We are aware of the fact that for an event study analysis to be precise, an event window of 60 minutes would be too large in the present context. In the event study literature event windows of less than five minutes are generally considered to create meaningfull and concise results. Further, we expect the greatest effect immediately at the beginning of the news release time. We therefore repeat the analysis for a shorter window ranging from 2.30 to 2.32 p.m. Assuming that news are generally released at that time and if markets react rationally, i.e. without delay, we would expect abnormal returns to be significant even within this short time frame. This is indeed what we find. Table 5.6 summarises these results. Although comparatively small in absolute value, the abnormal returns are significant for both good and bad news releases. This is supported by any of the computed t-statistics. As regards the size, approximately 50% of the cumulative abnormal returns are realized within the first three minutes after the announcement on good news days and even 80% on bad news days. The fact that the cumulative abnormal returns still grow until 3.30 p.m. can be ascribed to the difference between the announcement time set by the model and the actual announcement where the important information might be released by officials directly in their first sentence or just a bit later. As news announcements do never carry an exact time stamp it is impossible to exactly define the time it took place. Therefore the time set here is in our view a sensible approximation, but still only an approximation which leads to the described pattern of abnormal returns.

What do the results presented so far mean for the visually observed jumps at 3.30 p.m.? We conclude that there is no systematic dependence of the German investors on the US markets. Specifically, German investors don't seem to wait for their US counterparts before they start trading. On the contrary, they rationally incorporate any information as soon as it is available. The occasional jumps observable at the time of the NYSE opening, thus, seem coincidental. Reconsidering again the results presented in Table 5.4, the signs of the CARs indicate a slight adjustment.

Setting the news quantile to $q = 10\%$, i.e. allowing 20% of the days in the sample to be good or bad news days, respectively, does not alter the results qualitatively. We still find that the CARs are significant around 2.30 to 2.32 p.m. In absolute value, however, they are slightly smaller. This is perfectly compatible with the methodological approach as we possibly include more days with news of weaker global importance which should lower the estimated impact and thus the size of the CARs. At the opening of the NYSE we still do not find significantly abnormal return behavior. Only around the opening (between 3.20 and 3.40 p.m. CET) are effects found to be statistically significant. Again, the calculated CARs carry the inverse sign, i.e. they are estimated to be positive on bad news days and vice versa. and are small in absolute value (≤ 0.042). Allowing only for $q = 2\%$ of the days in the sample to be associated with good or bad news announcements renders most of the CARs insignificant. The absolute values are more pronounced for good news days. For bad news days, however, the effect sometimes almost vanishes.

5.4.2 *Speed of Reaction*

The above results in general suggest that the average German investor reacts immediately to news announcements in the USA rather than to the opening up of the US markets and, thus, implicitly to valuation suggestions of US investors. The question which arises naturally in this context is how fast the reaction of German traders is. In other words, how long does it take until prices in Germany fully reflect the US information? To answer this question we have a closer look at the event time around 2.30 p.m. First, we allow for an event window of 10 minutes. If the information is absorbed sufficiently quickly after the beginning of the announcement we would expect significant cumulative abnormal returns between 2.30 and 2.40 p.m. while they should be neglectable in size and probably not be significant in the following 10 minutes. Table 5.7 summarises the estimation results for the first 10 minute interval. We find that both positive and negative news days exhibit significant excess returns compared to the average day. As regards the size of the coefficients, we find that roughly 70% of the cumulative abnormal returns which have been documented for the time period 2.30 to 3.30 p.m. are already realized within the first ten minutes of this time window on good news days. On bad news days, we even find that they are completely realized within these ten minutes. It turns out that repeating the calculation of the CARs 10 minutes later, i.e. between 2.40 and 2.50 p.m., renders all results insignificant. So we conclude that any relevant information generated in the process of news announcements in the wake of the opening of the NYSE is absorbed quickly into the German market.

To illustrate the speed and the time of the news transmission we plot the cumulative abnormal returns between 2.25 and 3.45 p.m. CET. A graph based on the AR(p) model is given in Figure 5.3. On an average day there are no abnormal returns measurable during the time span of interest (the solid line). The plot supports the second hypothesis and shows a sharp increase or decrease around 2.30 p.m. which we interpret as the incorporation of US news into the German market. Good news positively influence the German DAX (dashed line) while negative news announcements negatively impact on the DAX value (dotted line).

5.4.3 Stability in light of the Financial Crisis

The study was carried out using data until August 2008. The reason is that we did not want the events in September and October 2008 to influence the results. In some cases, the reaction of the stock markets during the turmoil in September and October 2008 did not seem to be news driven only. At some stage it seemed more like a race to the bottom. Further, there is no common agreement in the literature yet on how to handle the present crisis. If the study is repeated[3] including data until 20 October 2008, the general results are not altered. However, as the S&P 500 return was large negative on quite a number of days, the quantile search technique would prefer these days as bad news days, even if, for the above mentioned reasons, not all days might have revealed truly new information (at least not unexpected information). The same might be true already for the rest of the sample. However, there the influence of an occasional misclassification error should not be as important as here where we are to include only eight more weeks compared to 5 years before. On average, however, the results still suggest that the reaction of the DAX takes place at 2.30 p.m. already and not at 3.30 p.m. So even in periods of crisis German investors behave rationally in the sense that they process information as soon as it is available.

5.4.4 On the Difference between Positive and Negative Announcements

Inclusion of the time period until 20th October 2008 also helps to explain the surprisingly different size of a reaction on positive and negative news days as reported in Figure 5.3. During the period July 2002 to August 2008 the number of days with positive returns and probably positive news announcements outnumbered those with negative returns and/or news announcements. Thus, the events included in the analysis favor more important positive news as compared to negative announcements, i.e. the selection of positive news announcement days is more strict than the selection of negative announcement days which are simply less numerous. Therefore, even weaker reactions are considered negative reactions. Extending the sample and thus including more negative information mass puts more weight on the negative side. For the returns from 2.30 to 2.40 p.m., for example, the negative average CARs are

[3]Detailed results are available on request.

-0.181 as compared to -0.126. The CARs on positive days are also slightly higher (0.00191 instead of 0.00178) which is probably also due to the inclusion of this highly volatile period. In general, positive and negative announcement days are more similar than when using the period up to August 2008 only.

Another way to further investigate this issue is to use absolute conditional returns of the S&P 500 index. We thereby set the threshold such that the innovations $\hat{\varepsilon}_i$ (cp. Equation (5.9)) within a range $-\varepsilon^c < \hat{\varepsilon}_i < \varepsilon^c$ are marked as normal. ε^c is the critical value of the distribution of absolute residuals $\hat{\varepsilon}_i$, e.g. the 90% quantile. Returns which are outside this interval are then considered unusual and the day is marked as positive or negative event day, respectively. When again allowing 10% of the days to be news days, the threshold is now ± 0.1699, i.e. it is slightly lower for positive days and almost unchanged for negative days (cp. Table 5.1). Using this approach on the dataset from July 2002 to August 2008, we find that a difference between positive and negative news still prevails although to a slightly lesser extent. Figure 5.4 summarizes the results. In general, the cumulative abnormal returns are lowered, especially for the days marked as good news days. We therefore conclude that in the end there is no real difference between the reaction to positive and negative announcements. The difference found above is due to the modeling strategy. Still, the conclusions drawn in sections 5.4.1 and 5.4.2 are not affected.

5.4.5 Volatility Analysis

The above analysis as well as the descriptive statistics in Table 5.1 strongly suggest that announcement days do not only differ in terms of returns, but also in terms of volatility. A plot of squared five-minute returns as a proxy for instant volatility supports this assumption. As can be seen in Figure 5.5, the volatility graph of announcement days always lies above the graph which corresponds to no news days. Volatility peaks at around 2.30 p.m. (which is compatible with the findings of Masset, 2008) and then seems to remain at a higher level than in the morning. The peak at 2.30 p.m., however, seems to be largely due to high volatility on announcement days.

The results of the formal comparison of realized volatility in the morning and the afternoon are summarized in Table 5.8. We find that realized volatility is in general significantly higher on announcement days, both in the morning and in the afternoon. On average, afternoon volatility on good news days is about twice as high than on days without news announcements while it is about three times as high on days with bad news announcements. The difference between good and bad news days, albeit economically not negligible, is not found to be statistically significant. In the morning, volatility is 1.5 to two times higher on announcement days than on quiet days. The difference between good and bad news days is smaller than in the afternoon and again not statistically significant. So the important conclusion to draw is that volatility in both conditions is higher than usual. It, thus, seems that German investors are somewhat apprehensive in expectation of some still unknown news announcement. This strengthens the view that the identified days really are informative and provide surprising information.

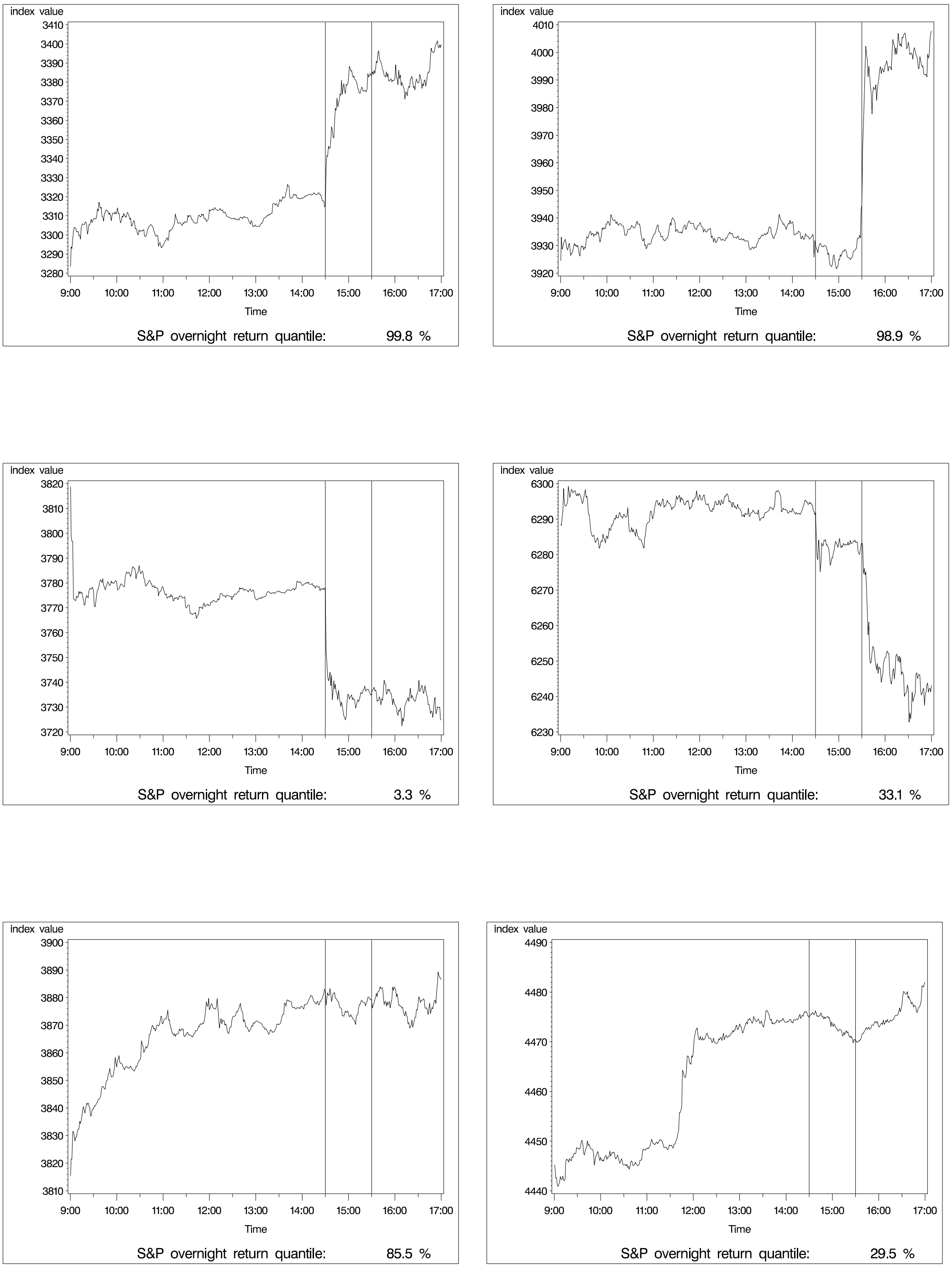
Volatility is generally higher in the afternoon than in the morning. The increase is greater on days with announcements (factor 2 to 2.5) than on normal days (factor 1.5). So even though German investors do have a timing advantage, valuation insecurity seems to persist to some degree. This finding may also help to explain the seemingly odd result in Table 5.4 where we found significant negative abnormal returns for average and positive announcement days while they were found to be positive (albeit not significant) on negative announcement days from 3.20 to 3.40 p.m. It seems that once the NYSE opens, there may be some adjustment or correction needed. This does not contradict the results presented above: while German investors exploit their timing advantage, they need to react again once they observe the reaction of US investors. In case that these investors judge the news differently, there may arise a need to adjust to the new circumstances. If we again consider the graph of cumulated abnormal returns in Figure 5.3, it seems that German investors are slightly overconfident on days with good news announcements. Hence, there is a negative abnormal return once the NYSE opens. On days with negative announcements, the (inverse) pattern is there as well, but CARs are not significant.

5.5 Concluding Remarks

This paper studies the behaviour of the DAX index in the early afternoon trading in Frankfurt. It shows that it is most likely that surprising news events which take place in the US before the opening of the NYSE are the reason for the occasionally observable jumps. Using event study methodology the paper shows that around 2.30 p.m. abnormal returns are possible which is one hour before the first US market opens and the time when macroeconomic news are usually announced. It is further found that the opening of the stock market itself does not (or only negligibly) alter the DAX return pattern. Once US information of global importance is publicly available it is quickly absorbed into prices in Germany. The outcome of the study is compatible with rational behaviour. If there is good or bad news which concerns German investors as well as US investors, the reaction of the German market takes place one hour earlier than in the US for the simple reason that the US market is still closed. Still, volatility on announcement days is generally higher than average, especially in the afternoon. So while there is a timing advantage, valuation insecurity still persists due to the still unknown precise reaction of US investors to the same news event. A small adjustment effect is found once the US stock market finally opens.

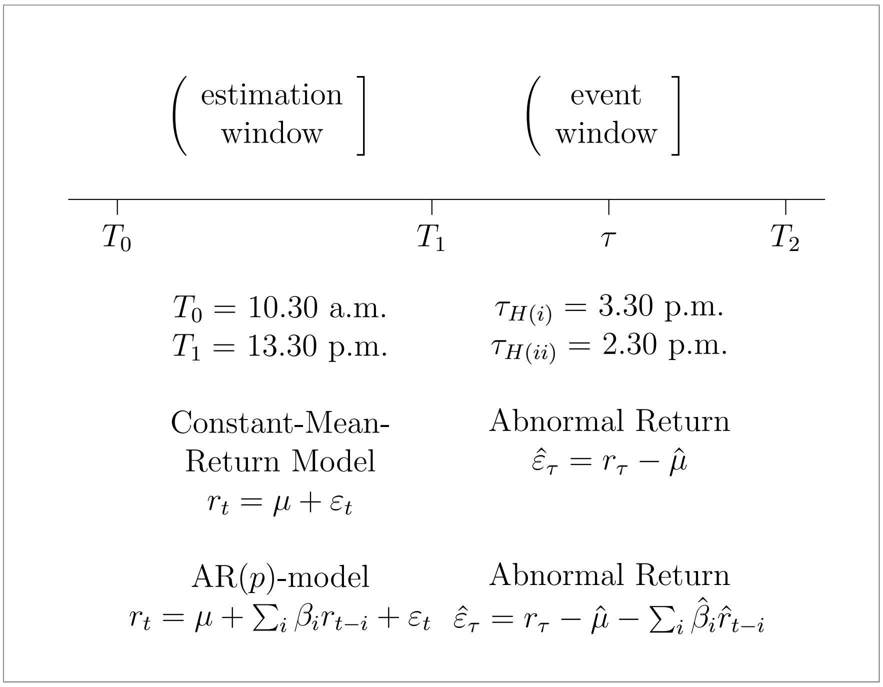
The design of the study, unfortunately, is as such that there is no possibility to exploit arbitrage gains from the knowledge of a possible DAX reaction. The reason is that the proxy which has been used here to distinguish good news days from days with bad or no news is not availably yet at 2.30 p.m.

Figure 5.1: Plot of DAX Index Value



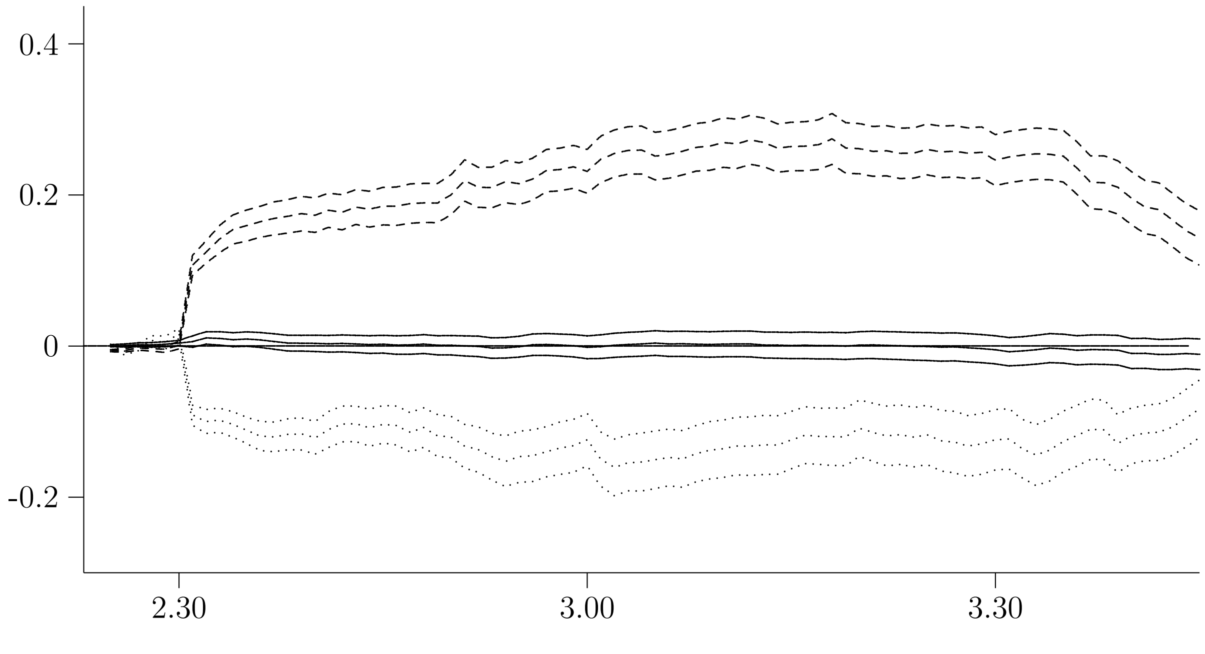
The graphs depict the value of the DAX index on 3 October 2003 (top left), 2 April 2004 (top right), 6 August 2004 (middle left), 27 October 2006 (middle right), 29 July 2004 (bottom left), and 30 May 2005 (bottom right).

Figure 5.2: Event Study Timeline



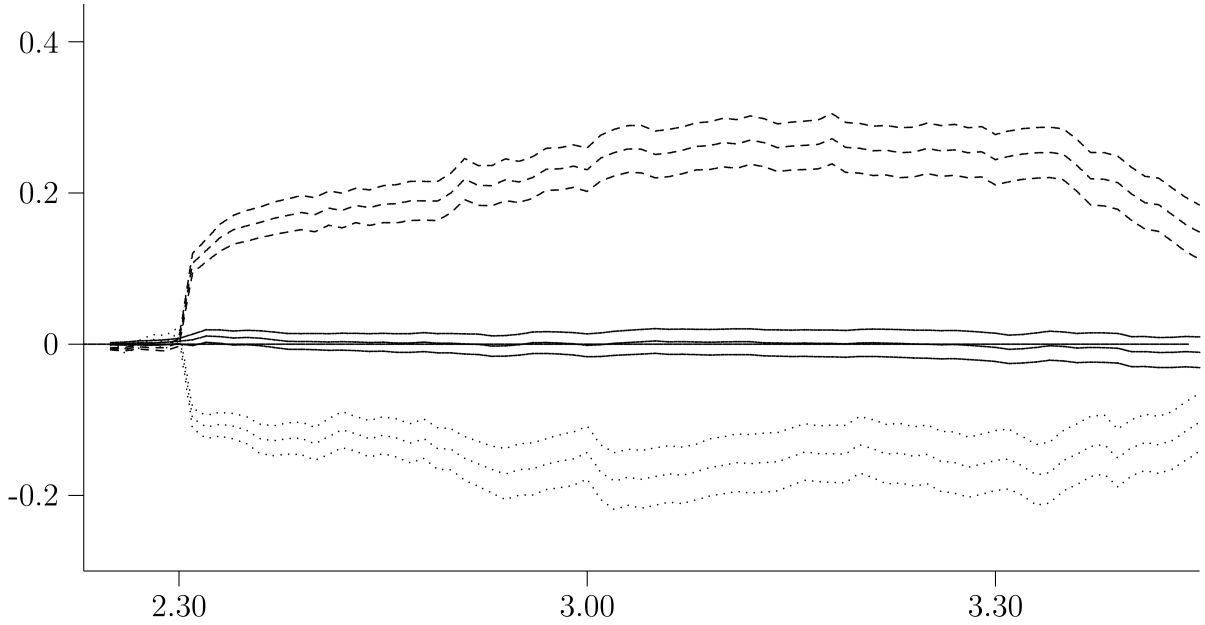
The graphic illustrates the event study time line along with the estimated models. T_0 and T_1 define the estimation window start and end time. $\tau_{H(i)}$ is the time the event takes place under Hypothesis (i) and $\tau_{H(ii)}$ is the time the event takes place under Hypothesis (ii).

Figure 5.3: DAX CAR plot (AR(p) model and conditional S&P 500 returns)



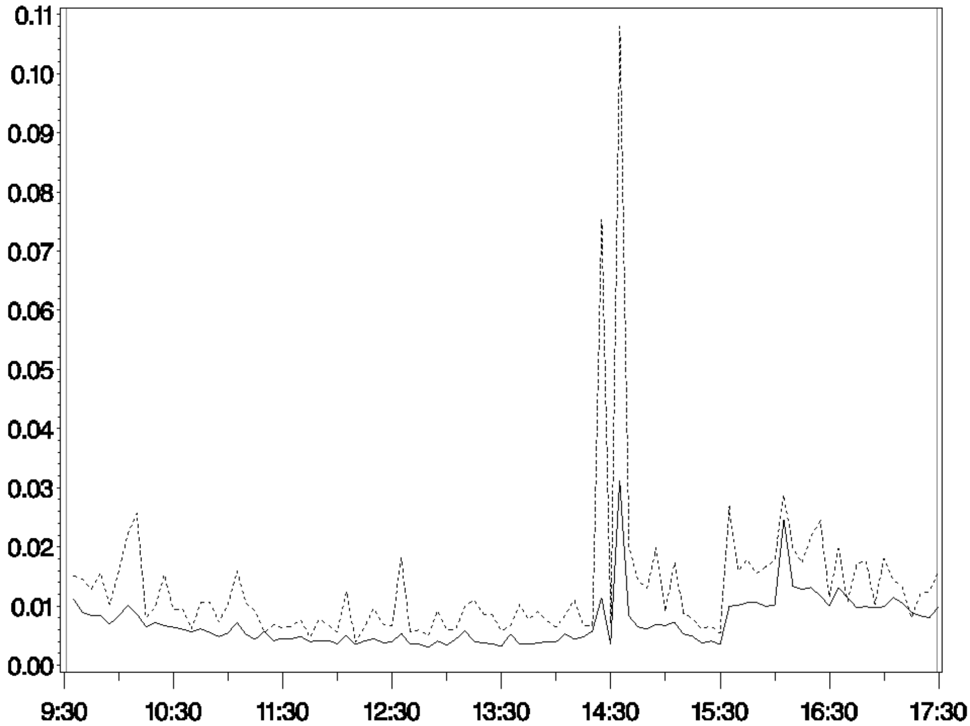
The graphic depicts the cumulative abnormal returns of the DAX index between 2.25 and 3.45 p.m. CET as calculated based on the AR(p) model. The solid line marks the average across all days, the dashed line represents days with good news events while the dotted line represents days with bad news events.

Figure 5.4: DAX CAR plot (AR(p) model and conditional absolute S&P 500 returns)



The graphic depicts the cumulative abnormal returns of the DAX index between 2.25 and 3.45 p.m. CET as calculated based on the AR(p) model. Good and bad news events are distinguished using the distribution of conditional absolute S&P 500 returns. The solid line marks the average across all days, the dashed line represents days with good news events while the dotted line represents days with bad news events. The distinction between days is based on absolute returns of the S&P 500 index.

Figure 5.5: DAX volatility plot



The graph depicts the pattern of squared log-returns (calculated over 5 minute intervals) as a proxy for intraday volatility. The solid line represents days without news announcements while the dashed line represents the days with news announcements (both positive and negative).

Table 5.1: Descriptive Statistics

S&P 500 return	ALL	POS	NEG
number of days	1301	66	65
mean	0.0059	0.2668	−0.2514
minimum	−0.9347	0.0802	−0.9347
maximum	0.6745	0.6745	−0.1469
standard deviation	0.1120	0.1095	0.1107
threshold (p=0.05)		0.1721	−0.1671

DAX	ALL	POS	NEG
number of days	1280	65	64
avg. morning return	0.0002	0.0028	-0.0049
avg. morning RV	0.0049	0.0069	0.0094
avg. afternoon return	0.0000	0.0026	-0.0009
avg. afternoon RV	0.0078	0.0143	0.0233

The table provides descriptive statistics for the S&P 500 close-to-open returns (upper panel) and the DAX returns (lower panel). ALL is the average across all days in the sample, POS is days with positive news announcements and NEG is days with news announcements of negative content. Returns in the lower panel are averages across one minute intervals. RV is the realised volatility of Andersen *et al.* (2001) calculated on five-minute intervals. Morning is the estimation window from 10.30 a.m. to 12.30 p.m. CET, afternoon is from 2.00 to 4.00 p.m. CET., encompassing all considered event windows.

Table 5.2: GARCH model estimates for S&P 500 close-to-open returns

Variable	Estimate	SE	t-value	p-value
μ	0.0016	0.0021	0.7688	0.4422
β_1	0.0427	0.0254	1.6811	0.0936
β_2	0.0521	0.0060	8.7240	<0.0001
ω	1.098E-8	6.056E-9	1.8127	0.0699
α	0.0390	0.0127	3.0709	0.0022
γ	0.9561	0.0133	71.8872	<0.0001
$1/m$	0.2706	0.0322	8.4037	<0.0001
Log Likelihood		7977.4991	R^2	0.0587

The table provides the estimation results of the GARCH model in Equation (5.9). SE is the standard error of the estimate.

Table 5.3: CAR at the NYSE opening

	ALL	POS	NEG
M(CAR)	−0.0003	−0.0046	0.0033
M(SCAR)	−0.0219	−0.1011	−0.0616
t_{cs}	−0.1111	−0.3891	0.2603
	(0.9116)	(0.6985)	(0.7955)
t_{BMP}	−0.5277	−0.6559	−0.3970
	(0.5978)	(0.5142)	(0.6927)

The table provides the average cumulative abnormal returns (CAR) and standardised CARs along with the appropriate test statistics. CARs are calculated around the opening of the NYSE (3.30 - 3.35 p.m. CET) based on the AR(p) model. P-values are given in parentheses. ALL is the average across all days in the sample, POS is days with positive news announcements and NEG is days with news announcements of negative content.

Table 5.4: CAR around the NYSE opening

	ALL	POS	NEG
M(CAR)	−0.0123	−0.0660	0.0012
M(SCAR)	−0.1113	−0.4860	−0.0289
t_{cs}	−2.9330	−3.3615	0.0607
	(0.0034)	(0.0013)	(0.9518)
t_{BMP}	−3.1130	−3.8377	−0.2397
	(0.0019)	(0.0003)	(0.8113)

The table provides the average cumulative abnormal returns (CAR) and standardised CARs along with the appropriate test statistics. CARs are calculated around the opening of the NYSE (3.20 - 3.40 p.m. CET) based on the AR(p) model. P-values are given in parentheses. ALL is the average across all days in the sample, POS is days with positive news announcements and NEG is days with news announcements of negative content.

Table 5.5: CAR during the news release time

	ALL	POS	NEG
M(CAR)	-0.0004	0.2508	-0.1284
M(SCAR)	-0.0785	1.0993	-0.5199
t_{cs}	-0.0387	3.4469	-1.6132
	(0.9692)	(0.0010)	(0.1116)
t_{BMP}	-1.5942	3.6160	-1.7356
	(0.1112)	(0.0006)	(0.0875)

The table provides the average cumulative abnormal returns (CAR) and standardised CARs along with the appropriate test statistics. CARs are calculated for the news release time (2.30 - 3.30 p.m. CET) based on the AR(p) model. P-values are given in parentheses. ALL is the average across all days in the sample, POS is days with positive news announcements and NEG is days with news announcements of negative content.

Table 5.6: CAR at the beginning of the news release time (1)

	ALL	POS	NEG
M(CAR)	0.0089	0.1293	-0.1040
M(SCAR)	0.1285	2.4912	-2.1651
t_{cs}	2.0180	4.2523	-2.9472
	(0.0438)	(0.0001)	(0.0045)
t_{BMP}	1.2774	4.1465	-3.2393
	(0.2017)	(0.0001)	(0.0019)

The table provides the average cumulative abnormal returns (CAR) and standardised CARs along with the appropriate test statistics. CARs are calculated for the beginning of the news release time (2.30 - 2.32 p.m. CET) based on the AR(p) model. P-values are given in parentheses. ALL is the average across all days in the sample, POS is days with positive news announcements and NEG is days with news announcements of negative content.

Table 5.7: CAR at the beginning of the news release time (2)

	ALL	POS	NEG
M(CAR)	0.0037	0.1779	−0.1260
M(SCAR)	−0.0015	1.8337	−1.3438
t_{cs}	0.6123	3.7411	−2.7948
	(0.5405)	(0.0004)	(0.0069)
t_{BMP}	−0.0207	3.6314	−3.0564
	(0.9835)	(0.0006)	(0.0033)

The table provides the average cumulative abnormal returns (CAR) and standardised CARs along with the appropriate test statistics. CARs are calculated for the beginning of the news release time (2.30 - 2.40 p.m. CET) based on the AR(p) model. P-values are given in parentheses. ALL is the average across all days in the sample, POS is days with positive news announcements and NEG is days with news announcements of negative content.

Table 5.8: Volatility Measures and Tests

Tested samples	test statistic	p-value
average vs abnormal days' afternoon volatility	8.6545	< 0.0001
average vs positive days' afternoon volatility	5.6156	< 0.0001
average vs negative days' afternoon volatility	6.9495	< 0.0001
positive vs negative days' afternoon volatility	1.2153	0.1121
average vs abnormal days' morning volatility	6.1078	$<.0001$
average vs positive days' morning volatility	3.6091	0.0002
average vs negative days' morning volatility	5.2614	$<.0001$
positive vs negative days' morning volatility	0.9704	0.1659
morning vs afternoon volatility on normal days	-9.1793	$<.0001$
morning vs afternoon volatility on positive days	-3.4061	0.0003
morning vs afternoon volatility on negative days	-3.8268	$<.0001$
volatility difference on average vs abnormal days	4.5075	$<.0001$

The table provides test statistics and one-sided p-values for the Wilcoxon signed-rank test. Positive and negative days are the days identified as announcement days in Section 5.3. P-values are approximated using the normal distribution.

Chapter 6

Summary and Conclusion

We are interested in the way international financial markets are linked. To this end, econometrics offers a vast toolbox that enables us to investigate and model different, yet related aspects of the interplay of financial markets on a worldwide scale. Understanding this interdependence structure is vital to investors and politicians alike as was painfully experienced in 2007 and 2008, two years which saw a great financial crisis associated with huge losses on stock markets all over the world. With only slowly recovering economies and a barely averted refinancing problem of Greece the crisis is probably far from being over. The focus of this study is on how information and shocks in general are spread and processed around the world, and it is as such linked to the broad field of market efficiency (in terms of efficient information processing).

As a first step, we trace global information and volatility transmission. Nowadays stock trading is only discontinuous if we limit ourselves to one stock market. However, from a global point of view, there is always an open market with investment opportunities. Information is, therefore, not only generated but also processed continuously. Keeping this thought in mind we model return and volatility spillovers between the three major financial centres USA, Europe and Asia, thereby covering almost 24 hours of trading activity. We find that dependence in the mean returns is weak and short-lived whereas dependence in the volatility dynamics is much more pronounced. We can thus conclude that it is beneficial to look back in time, especially for markets in Europe that are not only dependent on the behaviour of the US markets, but also respond to events in the Asian markets.

We then approach the aspect of long-term relationships between stock markets and therefore reconsider the context of cointegration in international financial markets. First of all we investigate the properties of the Johansen cointegration test in order to check that the influence of time varying volatility does not affect the results of the test. We find, however, that in certain circumstances volatility is an issue when testing for cointegration. Taking this knowledge into account, we show in a second stage that if the underlying true model for stock prices is the random walk model, cointegration is not a suitable framework with which to describe the interdependence of international financial markets. Stock specific individual innovations (or information) are the driving forces behind our theoretical result and provide the reason why stock market indices cannot be cointegrated. We show with the help of both an empirical and a simulation experiment that stock market indices most likely share a common stochastic trend component which, however, cannot be identified within the cointegration framework. This way we can explain both the heterogeneous results regarding stock market cointegration reported in the literature as well as the often documented comovement and high correlation between stock markets which is probably driven by a global common factor.

Finally, we move on to an intra-daily investigation of the dependence of the German stock market on US news surprises. We show that news which is generated abroad (at a time when the US markets are still closed) is immediately and efficiently incorporated into prices in the German stock markets. At the same time, these surprises cause a peak in volatility in the early afternoon trading in Germany, leading to a w-shape of the intraday volatility pattern. On average days without news from the US this shape is u-formed as it is for most other stock markets. Once the stock market opens in the USA we find only minor adjustments taking place in Germany.

The primary goal of the different studies is to highlight the interplay between stock markets around the globe and to offer suitable models to describe it. We show that information generated in one market may have a global impact on prices and volatility—spilling over from one country to the next and/or being absorbed in a common world factor. Although the way of measuring this impact differs, the central message behind the studies is identical: financial markets around the globe are highly interdependent. Furthermore, with respect to information processing, we conclude that both in terms of speed and timing, stock market agents behave rationally and that markets are information efficient.

References

Adams, Greg, Grant McQueen, and Robert Wood (2004): The Effects of Inflation News on High Frequency Stock Returns. *The Journal of Business*, 77(3): 547–574. 83

Akgiray, Vedat (1989): Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. *The Journal of Business*, 62(1): 55–80. 73

Andersen, Torben G., Tim Bollerslev, Peter F. Christoffersen, and Francis X. Diebold (2005): Volatility Forecasting. NBER Working Paper No. 11188. 19

—— (2006): Practical Volatility and Correlation Modeling for Financial Market Risk Management. In Mark Carey and René M. Stulz (editors), *The Risks of Financial Institutions*, chapter 17. University of Chicago Press, Chicago, Illinois, pp. 513 – 548. 10, 12, 15, 16

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold (2002): Parametric and Nonparametric Volatility Measurement. In Yacine Aït-Sahalia and Lars Peter Hansen (editors), *Handbook of Financial Econometrics*. Elsevier Science Ltd, Amsterdam, pp. 67–138. 9

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys (2001): The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 93(453): 42–55. 6, 16, 106

—— (2003): Modeling and Forecasting Realized Volatility. *Econometrica*, 71(2): 529–626. 9, 20, 22, 24, 32, 35, 89

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega (2007): Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, 73(2): 251–277. 85

BARASSI, MARCO R., GUGLIELMO MARIA CAPORALE, AND STEPHEN G. HALL (2005): Interest Rate Linkages: Identifying Structural Relations. *Applied Financial Economics*, 15(14): 977–86. 38

BAUR, DIRK AND ROBERT C. JUNG (2006): Return and volatility linkages between the US and the German stock market. *Journal of International Money and Finance*, 25(4): 598–613. 5, 13, 83

BLACK, FISCHER (1986): Noise. *The Journal of Finance*, 41(3): 529–543. 61

BOEHMER, EKKEHART, JIM MASUMECI, AND ANNETTE B. POULSEN (1991): Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, 30(2): 253–272. 88, 89

BOLLEN, BERNARD AND BRETT INDER (2002): Estimating daily volatility in financial markets utilizing intraday data. *Journal of Empirical Finance*, 9: 551–562. 6, 9, 15, 20, 33

BOLLERSLEV, TIM (1986): Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3): 307–327. 5

BONDT, WERNER F. M. DE AND RICHARD THALER (1985): Does the Stock Market Overreact? *The Journal of Finance*, 40(3): 793–805. 61

BOOTH, GEOFFREY G., TEPPO MARTIKAINEN, AND YIUMAN TSE (1997): Price and volatility spillovers in Scandinavian stock markets. *Journal of Banking and Finance*, 21: 811–823. 83

CAMPBELL, JOHN Y., W. LO, AND A. CRAIG MACKINLEY (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey. 88

CARCHANO, ÓSCAR AND ÁNGEL PARDO (2009): Rolling over stock index futures contracts. *Journal of Futures Markets*, 29(7): 684 – 694. 14

CHAN, KAM C., BENTON E. GUP, AND MING-SHIUN PAN (1997): International Stock Market Efficiency and Integration: A Study of Eighteen Nations. *Journal of Business Finance & Accounting*, 24(6): 803–813. 61

CHOUDHRY, TAUFIQ, LIN LU, AND KE PENG (2007): Common stochastic trends among Far East stock prices: Effects of the Asian financial term crisis. *International Review of Financial Analysis*, 16(3): 242–261. 60

Christoffersen, Peter F. and Francis X. Diebold (2006): Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics. *Management Science*, 52(8): 1273–1287. 25

Click, Reid W. and Michael G. Plummer (2005): Stock market integration in ASEAN after the Asian financial crisis. *Journal of Asian Economics*, 16(1): 5–28. 60

Constantinou, E., A. Kazandjian, G.P. Kouretas, and V. Tahmazian (2008): Common stochastic trends among the Cyprus stock exchange and the ASE, LSE and NYSE. *Bulletin of Economic Research*, 60(4): 327–349. 60

Diebold, Francis X. and Kamil Yilmaz (2009): Measuring Financial Asset Return and Volatility Spillovers, With Application to Global Equity Markets. *The Economic Journal*, 119(534): 158–171. 6, 83, 90

Dimpfl, Thomas and Robert C. Jung (2007): Financial Market Spillovers Around the Globe. Working paper vailable at: http://ssrn.com/abstract=1001615. 83

Dolley, James C. (1933): Characteristics and Procedure of Common Stock Split-Ups. *Harvard Business Review*, 11(3): 316–326. 83

Dungey, Mardi, Vance L. Martin, and Adrian R. Pagan (2000): A Multivariate Latent Factor Decomposition of International Bond Yield Spreads. *Journal of Applied Econometrics*, 15(6): 697–715. 63

Ederington, Louis H. and Jae Ha Lee (1993): How Markets Process Information: News Releases and Volatility. *The Journal of Finance*, 48(4): 1161–1191. 83, 85

Edwards, Robert D. and John Magee (2001): *Technical Analysis of Stock Trends*, 8th edition. CRC Press, Boca Raton, Florida. 63

Engle, Robert F. (1982): Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4): 987–1007. 5

Engle, Robert F. and Clive W. J. Granger (1987): Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55: 251–276. 39, 40, 60, 62, 63, 64, 66, 67, 75

ENGLE, ROBERT F. AND KENNETH F. KRONER (1995): Multivariate Simultaneous Generalized Arch. *Econometric Theory*, 11(1): 122–150. 45

FABOZZI, FRANK J., CHRISTOPHER K. MA, WILLIAM T. CHITTENDEN, AND R. DANIEL PACE (1995): Predicting intraday price reversals. *Journal of Portfolio Management*, 21(2): 42–53. 91

FAMA, EUGENE F. (1965): The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1): 34–105. 61

—— (1995): Random Walks in Stock Market Prices. *Financial Analysts Journal*, 51(1): 75–80. 61

GARMAN, MARK B. AND MICHAEL J. KLASS (1980): On the Estimation of Security Price Volatilities from Historical Data. *The Journal of Business*, 53(1): 67–78. 6, 24

GIBBONS, JEAN DICKINSON AND SUBHABRATA CHAKRABORTI (2003): *Nonparametric Statistical Inference*, 4th edition. Marcel Dekker, Inc., New York. 89

GODFREY, MICHAEL D., CLIVE W. J. GRANGER, AND OSKAR MORGENSTERN (2007): The Random-Walk Hypothesis of Stock Market Behavior. *Kyklos*, 17(1): 1–30. 61

GRAMMIG, JOACHIM, MICHAEL MELVIN, AND CHRISTIAN SCHLAG (2005): Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects. *Journal of Empirical Finance*, 12(1): 139–164. 64

GRANGER, CLIVE W. J. (1969): Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3): 424–438. 7

—— (1986): Developments in the Study of Cointegrated Economic Variables. *Oxford Bulletin of Economics and Statistics*, 48(3): 213–228. 3, 61

GRANGER, CLIVE W.J. AND NORMAN R. SWANSON (1997): An introduction to stochastic Unit Root Processes. *Journal of Econometrics*, 80(1): 35–62. 40

HAMAO, YASUSHI, RONALD W. MASULIS, AND VICTOR NG (1990): Correlations in Price Changes and Volatility across International Stock Markets. *The Review of Financial Studies*, 3(2): 281–307. 7, 82

Hamilton, James D. (1994): *Time Series Analysis*. Princeton University Press, Princeton, New Jersey. 12, 41

Hansen, Bruce E. (1990): A Powerful, Simple Test For Cointegration Using Cochrane- Orcutt. RCER Working Papers No. 230, University of Rochester. 39

Harris, David (1997): Principal Components Analysis of Cointegrated Time Series. *Econometric Theory*, 13(3): 529–557. 39

Harris, David, Brendan McCabe, and Stephen Leybourne (2002): Stochastic cointegration: estimation and inference. *Journal of Econometrics*, 111(2): 363–384. 39, 40, 46

Harvey, Campbell R. (1991): The World Price of Covariance Risk. *The Journal of Finance*, 46(1): 111–157. 71

Hasbrouck, Joel (1991): The Summary Informativeness of Stock Trades: An Econometric Analysis. *The Review of Financial Studies*, 4(3): 571–595. 12, 21, 34, 35

—— (1995): One Security, Many Markets: Determining the Contributions to Price Discovery. *The Journal of Finance*, 50(4): 1175–1199. 44, 64

Haug, Alfred A. (1996): Tests for cointegration a Monte Carlo comparison. *Journal of Econometrics*, 71(1-2): 89–115. 39, 48

—— (2002): Testing Linear Restrictions on Cointegrating Vectors: Sizes and Powers of Wald and Likelihood Ratio Tests in Finite Samples. *Econometric Theory*, 18(2): 505–524. 47

Haug, Alfred A., James G. MacKinnon, and Leo Michelis (2000): European Monetary Union: a cointegration analysis. *Journal of International Money and Finance*, 19: 419–432. 38

Hess, Dieter (2004): Determinants of the relative price impact of unanticipated information in U.S. macroeconomic releases. *Journal of Futures Markets*, 25(7): 609–629. 83

Hess, Dieter, He Huang, and Alexandra Niessen (2008): How Do Commodity Futures Respond to Macroeconomic News? *Financial Markets and Portfolio Management*, 22(2): 127–146. 83

Ho, Mun S. and Bent E. Sorensen (1996): Finding Cointegration Rank in High Dimensional Systems Using the Johansen Test: An Illustration Using Data Based Monte Carlo Simulations. *The Review of Economics and Statistics*, 78(4): 726–732. 52

Johansen, Søren (1988): Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3): 231–254. 38, 41, 42, 60, 62, 67

Johansen, Soren (1991): Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6): 1551–80. 2, 38, 39, 42, 67, 69

—— (1992): Determination of Cointegration Rank in the Presence of a Linear Trend. *Oxford Bulletin of Economics & Statistics*, 54(3): 383–397. 43, 47

Johansen, Søren and Katarina Juselius (1992): Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics*, 53: 211–244. 43

Jung, Robert, Roman Liesenfeld, and Jean-Francois Richard (2010): Dynamic Factor Models for Multivariate Count Data: An Application to Stock-Market Trading Activity. *Journal of Business & Economic Statistics*, forthcoming. 63

Kasa, Kenneth (1992): Common stochastic trends in international stock markets. *Journal of Monetary Economics*, 29(1): 95–124. 60, 62, 67

Kearns, Philip and A. R. Pagan (1993): Australian stock market volatility: 1875-1987. *Economic Record*, 69(205): 163–178. 11

Kerl, Alexander G. and Andreas Walter (2007): Market Responses to Buy Recommendations Issued by Personal Finance Magazines: Effects of Information, Price-Pressure, and Company Characteristics. *Review of Finance*, 11(1): 117–141. 83

King, Mervyn A. and Sushil Wadhwani (1990): Transmission of Volatility between Stock Markets. *The Review of Financial Studies*, 3(1): 5–33. 85

Koutmos, Gregory and G. Geoffrey Booth (1995): Asymmetric volatility transmission in international stock markets. *Journal of International Money and Finance*, 14(6): 747–762. 6

Lagoarde-Segot, Thomas and Brian M. Lucey (2007): Capital Market Integration in the Middle East and North Africa. *Emerging Markets Finance and Trade*, 43(3): 34–57. 60

Lee, Tae-Hwy and Yiuman Tse (1996): Cointegration tests with conditional heteroskedasticity. *Journal of Econometrics*, 73(2): 401–410. 38, 39, 52

Lewellen, Jonathan (2002): Momentum and Autocorrelation in Stock Returns. *The Review of Financial Studies*, 15(2): 533–564. 61

Lin, Wen-Ling, Robert F. Engle, and Takatoshi Ito (1994): Do Bulls and Bears Move Across Borders? International Transmission of Stock Returns and Volatility. *The Review of Financial Studies*, 7(3): 507–538. 5, 13, 82, 83

Ljung, G. M. and G. E. P. Box (1978): On a measure of lack of fit in time series models. *Biometrika*, 65(2): 297–303. 67

Lütkepohl, Helmut (2005): *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. 41, 46

MacKinnon, James G. (2006): Bootstrap Methods in Econometrics. *Economic Record*, 82(s1): S2–S18. 17, 19

MacKinnon, James G., Alfred A. Haug, and Leo Michelis (1999): Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration. *Journal of Applied Econometrics*, 14(5): 563–577. 43, 68, 69

Makridakis, Spyros, Steven C. Wheelwright, and Rob J. Hyndman (1998): *Forecasting - Methods and Applications*, third edition. John Wiley & Sons, Inc., New York. 22

Malkiel, Burton G. (1973): *A Random Walk Down Wall Street*. W. W. Norton & Company, Inc., New York, London. 61, 63

Mantalos, Panagiotis (2001): ECM-Cointegration test with GARCH(1,1) Errors. *InterStat: Statistics on the Internet*. Available at interstat.statjournals.net. 38

Masih, R. and A.M.M. Masih (2004): Common stochastic trends and the dynamic linkages driving European stock markets: Evidence from pre- and post-October 1987 crash eras. *European Journal of Finance*, 10(1): 81–104. 38

MASSET, PHILIPPE (2008): Properties of High Frequency DAX Returns: Intraday Patterns, Jumps and their Impact on Subsequent Volatility. Working paper available at SSRN: ssrn.com/abstract=1123766. 3, 87, 98

MCCABE, BRENDAN, STEPHEN LEYBOURNE, AND DAVID HARRIS (2006): A Residual-based Test for Stochastic Cointegration. *Econometric Theory*, 22(3): 429–456. 38, 41, 46, 50, 55, 58, 59

MENKVELD, ALBERT J., SIEM JAN KOOPMAN, AND ANDRÉ LUCAS (2007): Modeling Around-the-Clock Price Discovery for Cross-Listed Stocks Using State Space Methods. *Journal of Business & Economic Statistics*, 25(2): 213–225. 14

MUNTERMANN, JAN AND ANDRE GUETTLER (2007): Intraday stock price effects of ad hoc disclosures: the German case. *Journal of International Financial Markets, Institutions and Money*, 17(1): 1–24. 83

NARAYAN, PARESH KUMAR AND RUSSELL SMYTH (2005): Cointegration of Stock Markets between New Zealand, Australia and the G7 Economies: Searching for Co-Movement under Structural Change. *Australian Economic Papers*, 44(3): 231–247. 61

NIKKINEN, JUSSI AND PETRI SAHLSTRÖM (2004): Scheduled domestic and US macroeconomic news and stock valuation in Europe. *Journal of Multinational Financial Management*, 14(3): 201–215. 85

OSTERWALD-LENUM, MICHAEL (1992): A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics. *Oxford Bulletin of Economics and Statistics*, 54(3): 461–472. 43

PAN, MING-SHIUN AND L. PAUL HSUEH (1998): Transmission of Stock Returns and Volatility between the U.S. and Japan: Evidence from the Stock Index Futures Markets. *Asia-Pacific Financial Markets*, 5(3): 211–225. 7

PASCUAL, ANTONIO GARCIA (2003): Assessing European stock markets (co)integration. *Economics Letters*, 78(2): 197–203. 61

PESARAN, BAHRAM AND M. HASHEM PESARAN (2007): Modelling Volatilities and Conditional Correlations in Futures Markets with a Multivariate t Distribution. Cambridge Working Papers in Economics number 0734. 8

Polasek, Wolfgang and Lei Ren (2001): Volatility analysis during the Asia crisis: a multivariate GARCH-M model for stock returns in the U.S., Germany and Japan. *Applied Stochastic Models in Business and Industry*, 17(1): 93–108. 6, 7

Poterba, James M. and Lawrence H. Summers (1986): The Persistence of Volatility and Stock Market Fluctuations. *The American Economic Review*, 76(5): 1142–1151. 11

Richards, Anthony J. (1995): Comovements in national stock market returns: Evidence of predictability, but not cointegration. *Journal of Monetary Economics*, 36(3): 631–654. 3, 62, 67

Richardson, Matthew (1993): Temporary Components of Stock Prices: A Skeptic's View. *Journal of Business & Economic Statistics*, 11(2): 199–207. 61

Saikkonen, Pentti (1992): Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation. *Econometric Theory*, 8(1): 1–27. 39

Savva, Christos, Denise R. Osborn, and Len Gill (2005): Volatility, Spillover Effects and Correlations in US and Major European Markets. Discussion Paper, Money Macro and Finance (MMF) Research Group Conference 2005, available at repec.org/mmfc05/paper23.pdf. 5

Schwarz, Gideon (1978): Estimating the Dimension of a Model. *Annals of Statistics*, 6(2): 461–464. 87

Seo, Byeongseon (2007): Asymptotic distribution of the cointegrating vector estimator in error correction models with conditional heteroskedasticity. *Journal of Econometrics*, 137(1): 68–111. 39

Susmel, Raul and Robert F. Engle (1994): Hourly volatility spillovers between international equity markets. *Journal of International Money and Finance*, 13: 3–25. 5, 14, 82

Toda, Hiro Y. (1995): Finite Sample Performance of Likelihood Ratio Tests for Cointegrating Ranks in Vector Autoregressions. *Econometric Theory*, 11(5): 1015–1032. 39

Tsay, Ruey S. (2005): *Analysis of Financial Time Series*, 2nd edition. John Wiley & Sons, Hoboken, New Jersey. 38

VALADKHANI, A. AND S. CHANCHARAT (2008): Dynamic linkages between Thai and international stock markets. *Journal of Economic Studies*, 35(5): 425–441. 60

WONG, HEUNG, W. K. LI, AND SHIQING LING (2005): Joint Modeling of Cointegration and Conditional Heteroskedasticity with Applications. *Annals of the Institute of Statistical Mathematics*, 57(1): 83–103. 38

WONGSWAN, JON (2006): Transmission of Information across International Equity Markets. *Review of Financial Studies*, 19(4): 1157–1189. 5

WORTHINGTON, ANDREW C. AND HELEN HIGGS (2009): Efficiency in the Australian stock market, 1875-2006: a note on extreme long-run random walk behaviour. *Applied Economics Letters*, 16(3): 301 – 306. 61

Curriculum Vitae

Thomas Ernst Herbert Dimpfl

geboren am 19. September 1980 in Furth im Wald, Deutschland

Ausbildung

09/1991 – 06/2000	Abitur Robert-Schuman-Gymnasium, Cham Dreimonatiger Aufenthalt am Lycée Lavoisier, Mayenne (F)
10/2001 – 03/2006	Studium der Internationalen Betriebswirtschaftslehre mit Abschluss Diplom-Kaufmann Eberhard-Karls-Universität, Tübingen Zweisemestriger Aufenthalt an der University of Port Elizabeth, Port Elizabeth (ZA)
04/2006 – 06/2007	Promotionsstudium im Graduiertenkolleg „Unternehmensentwicklung, Marktprozesse und Regulierung in dynamischen Entscheidungsmodellen“ Eberhard-Karls-Universität, Tübingen
06/2007 – 03/2010	Bearbeitung der vorliegenden Doktorarbeit zum Thema „Econometric Analysis of International Financial Markets“ unter der Betreuung von Prof. Dr. Robert Jung Universität Erfurt, Erfurt

Berufserfahrung

09/2000 – 07/2001	Zivildienst an der Schule zur individuellen Lebensbewältigung St. Gunther, Cham
09/2002 – 02/2003	Wissenschaftliche Hilfskraft am Lehrstuhl für Statistik, Ökonometrie und Empirische Wirtschaftsforschung Eberhard-Karls-Universität, Tübingen
01/2004 – 04/2004	Praktikum bei der Lufthansa Cargo AG, Frankfurt/Main, im Bereich Marketing/Pricing
04/2004 – 01/2006	Wissenschaftliche Hilfskraft am Lehrstuhl für Statistik, Ökonometrie und Empirische Wirtschaftsforschung Eberhard-Karls-Universität, Tübingen
06/2007 – 07/2010	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Ökonometrie Universität Erfurt, Erfurt