Sascha Mergner

Applications of State Space Models in Finance

An Empirical Analysis of the Time-varying
Relationship between Macroeconomics,
Fundamentals and Pan-European Industry Portfolios

Universitätsverlag Göttingen

Sascha Mergner

Applications of State Space Models in Finance

Sascha Mergner

Applications of State Space Models in Finance

An Empirical Analysis of the
Time-varying Relationship
between Macroeconomics, Fundamentals
and Pan-European Industry Portfolios

Universitätsverlag Göttingen
2009

Sascha Mergner
e-mail: sascha.mergner@gmx.de

To Leyla

Contents

List of figures

List of tables

Notation and conventions

The outline throughout this thesis relies on notation and conventions that are in common use. The numerous mathematical symbols introduced in the theoretical part of this thesis are based on the standard Greek and English alphabets. As a consequence, the same symbol might have different connotations. By defining all employed symbols and abbreviations in the text, the different meaning should generally become obvious from the context. The following notation and conventions are adopted:

- Vectors, denoted by lower-case letters, and matrices, denoted by upper-case letters, are printed in **bold** type.

- If a vector $\boldsymbol{x}$ is identically and independently distributed (IID) with mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{H}$, we write $\boldsymbol{x} \sim \text{IID}(\boldsymbol{\mu}, \boldsymbol{H})$.

- The expression "log" generally refers to the natural logarithm.

- Phrases printed in *italics* are particularly important in the context of the respective section.

- Expressions either referring to software packages, to web links or to mnemonics are printed in `typewriter` style.

- In connection with the use of trade and service marks, the ™ and ℠ symbols are placed in the top right-hand corner of the mark, respectively.

- Throughout this thesis, the symbol "§" is used as a reference to a chapter, a section or a subsection.

- In the list of references, any links to the internet are given together with their respective access date (reported in parentheses).

Used abbreviations and symbols

Abbreviations

AC	autocorrelation
AIC	Akaike information criterion
APT	Arbitrage Pricing Theory
ARCH	autoregressive conditional heteroskedasticity
ARIMA	autoregressive integrated moving average
BEKK	Baba-Engle-Kraft-Kroner
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BIC	Bayesian information criterion
BV	book value
CAPM	Capital Asset Pricing Model
CCC	constant conditional correlation
cf.	compare (Latin: confer)
CUSUMSQ	cumulative sum of squares
DCC	dynamic conditional correlation
DF	degrees of freedom
DJ	Dow Jones
EGARCH	exponential GARCH
EM	expectation-maximization
FIBOR	Frankfurt Interbank Offered Rate
GARCH	generalized ARCH
GJR	Glosten, Jagannathan and Runkle
GLS	generalized least squares
GMM	generalized methods of moments
HMMs	hidden Markov models
ICAPM	Intertemporal Capital Asset Pricing Model
ICB	Industry Classification Benchmark
i.e.	that is (Latin: id est)
IGARCH	integrated GARCH
IID	identically and independently distributed
IR	information ratio
JB	Jarque-Bera
KF	Kalman filter

LM	Lagrange multiplier
MAE	mean absolute error
MGARCH	multivariate GARCH
MCL	Monte Carlo likelihood
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
MMR	moving mean reverting
MR	mean reverting
MS	Markov switching
MSE	mean squared error
MSM	Markov switching market
MSV	multivariate stochastic volatility
MV	market value
NIC	news impact curve
NSB	negative sign bias
OLS	ordinary least squares
p.	page
p.e.d.	prediction error decomposition
p.e.v.	prediction error variance
PE	price-earnings ratio
pp.	pages
PSB	positive sign bias
QML	quasi-maximum likelihood
RC	random coefficient
RLS	recursive least squares
RR1	one-year rolling regression
RR5	five-year rolling regression
RW	random walk
SB	sign bias
SSE	sum of squared errors
SV	stochastic volatility
t-GARCH, tG	GARCH model with t-distributed errors
TMI	Total Market Index
TMT	Technology, Media and Telecommunications
WLS	weighted least squares

Greek letters

$\boldsymbol{\alpha}_t$	vector of forward probabilities
α	intercept term of a regression model
$\boldsymbol{\beta}_t$	vector of conditional regression coefficients;
	vector of backward probabilities
$\tilde{\boldsymbol{\beta}}_{t\vert T}$	ML estimate of $\boldsymbol{\beta}_t$ given the complete sample up to time T

$\boldsymbol{\beta}_i$	vector of unconditional factor loadings for the i-th asset
β_k	coefficient of k-th regressor
$\bar{\beta}$	mean coefficient
$\boldsymbol{\Delta}$	coefficient of lagged variances in an MGARCH process
$\boldsymbol{\delta}$	diffuse vector
$\boldsymbol{\delta}_t$	combined vector of the two system vectors $\boldsymbol{c}_t$ and $\boldsymbol{d}_t$
δ_i	coefficient related to lagged variances of a (G)ARCH process
δ	stochastic discount factor
$\boldsymbol{\epsilon}_t$	observation disturbances
$\hat{\boldsymbol{\epsilon}}_t$	smoothed observation disturbances
$\tilde{\boldsymbol{\epsilon}}_t$	vector of standardized residuals
ϵ_t	residual in the mean equation of a time series process
ϵ_t^{*2}	independent variable in sign bias test
$\boldsymbol{\Phi}_t$	matrix of combined system matrices $\boldsymbol{T}_t$ and $\boldsymbol{Z}_t$
ϕ_t	rescaled forward probabilities
ϕ	persistence parameter in the variance equation of a SV model
$\boldsymbol{\Gamma}$	transition probability matrix;
	coefficient of lagged shocks in an MGARCH process
γ_i	coefficient related to lagged shocks of a (G)ARCH process
γ_{ij}	probability of changing from state i to state j
γ_t	intercept term in a cross-sectional regression
$\boldsymbol{\eta}_t$	state disturbances
$\hat{\boldsymbol{\eta}}_t$	smoothed state disturbances
$\boldsymbol{\iota}$	N-dimensional column vector every element of which is unity
$\kappa_t(i)$	auxiliary variable in the context of the Viterbi algorithm
κ	positive scalar as part of a diffuse prior
$\boldsymbol{\lambda}_t$	vector of factor risk premia
λ_k	risk premium for the k-th risk factor
μ	unconditional mean
$\boldsymbol{\nu}_t$	state disturbances;
	$N \times 1$ disturbance term of a cross-sectional regression
$\boldsymbol{\nu}$	vector of nuisance parameters of a GARCH process
$\nu_t(i)$	auxiliary variable in the context of the Viterbi algorithm
ν	degrees of freedom of the Student-t distribution
$\boldsymbol{\pi}_s$	vector of stationary transition probabilities
$\boldsymbol{\pi}(1)$	probability distribution of the initial state vector in HMMs
π_i	weight of the i-th component in a mixture model
$\boldsymbol{\Omega}$	unconditional covariance matrix of standard regression model
$\boldsymbol{\Omega}_f$	unconditional covariance matrix of factor realizations $\boldsymbol{f}_t$
$\boldsymbol{\Omega}_t$	covariance matrix of combined observation and state errors
Ω_{t-1}	information set up to and including time $t-1$
ω	constant term in (G)ARCH process
$\boldsymbol{\theta}$	finite dimensional parameter vector of a GARCH model
θ	latent stochastic volatility process with $\theta = (h_1, \ldots, h_T)'$

$\theta^{(i)}$	i-th draw of θ
ϑ	coefficient related to EGARCH or GJR-GARCH component
ρ	correlation coefficient
ρ_t^S	Spearman's rank correlation coefficient
$\boldsymbol{\Sigma}$	matrix of combined state vector and initial covariance matrix
σ^2	unconditional standard deviation;
	prediction error variance
σ_t	conditional standard deviation
$\tilde{\sigma}^2$	approximated p.e.v. in terms of unstandardized GLS residuals
σ_*^2	positive scaling factor to which the variance is proportional
$\tilde{\sigma}_*^2(\boldsymbol{\psi}_*)$	ML estimator of σ_*^2 for given $\boldsymbol{\psi}_*$ without explanatory variables
$\tilde{\sigma}^{+2}$	p.e.v. for models with fixed regression coefficients
$\tilde{\sigma}_*^{+2}$	ML estimator of σ_*^2 that depends on GLS residuals
ς_t	state disturbances
τ	time index
$\boldsymbol{\upsilon}_t$	combined vector of state and observation disturbances
$\boldsymbol{\xi}_t$	state vector
$\bar{\boldsymbol{\xi}}$	mean state vector
$\boldsymbol{\xi}^*$	mean-corrected state vector
$\boldsymbol{\xi}_t^\dagger$	extended state vector
$\hat{\boldsymbol{\xi}}_t$	smoothed state vector
$\boldsymbol{\Psi}$	quadratic matrix used in the context of GLS
$\boldsymbol{\psi}$	vector of unknown parameters
$\boldsymbol{\psi}_*$	concentrated vector of unknown parameters
$\boldsymbol{\zeta}_t$	state disturbances

Other letters

$\boldsymbol{A}$	selection matrix
$\boldsymbol{a}_t$	mean state vector
$\boldsymbol{a}_1$	mean vector of the initial state vector
$\boldsymbol{a}$	constant term as part of the initial state vector
a_t	location parameter
$\boldsymbol{B}$	matrix of unconditional factor loadings
b_t	scaling parameter
BMR_t	benchmark residual factor
$\boldsymbol{c}_t$	system vector in state equation
c_t	consumption at time t
d	number of diffuse elements in the state vector
$\boldsymbol{D}_t$	matrix of conditional standard deviations
$D_{i,t}$	rank difference
$\boldsymbol{d}_t$	system vector in observation equation
$\boldsymbol{F}_t$	variance matrix of the one-step ahead prediction error

$\bar{\boldsymbol{F}}_{T+l}$	minimum MSE variance matrix for a l-period ahead forecast
$\boldsymbol{f}_{t-1}$	vector of lagged factor realizations
f_t	variance of one-step ahead prediction errors (univariate case)
$f_{k,t}$	realizations of the k-th risk factor
$\bar{f}$	steady-state value of f_t
$\tilde{f}_{k,t}$	mean zero realizations of the k-th risk factor
fib_t	FIBOR rate
FX_t	exchange rate factor
$\boldsymbol{H}_t$	variance-covariance matrix of observation disturbances
$h^{(i)}$	i-th draw for h obtained from an importance density
h_t	conditional variance in a GARCH model;
	log-volatility process in a SV model
h	unconditional variance of observation errors
i_T^*	the most probable state at time T in the Viterbi algorithm
$\boldsymbol{J}_\Phi$	index matrix
$\boldsymbol{K}_t$	Kalman gain matrix
$\bar{\boldsymbol{K}}$	steady-state value of the Kalman gain matrix
K_ϵ	kurtosis for the unconditional distribution of ϵ_t
K_z	kurtosis for the unconditional distribution of z_t
k	number of explanatory variables
ku	kurtosis of a series
$\boldsymbol{L}_t$	element of the covariance matrix of the state vector
l	order of test statistics;
	lead time
lev	correlation between $R_{i,t}^2$ and $R_{i,t-1}$
M	number of draws from a simulation
m	dimension of the state vector;
	number of components in a mixture model
m_{t+1}	stochastic discount factor
$\boldsymbol{N}_t$	variance matrix of the weighted sum of future innovations
N_t	number of pairs of sector ranks
N	dimension of a time series vector
OIL_t	oil factor
$\boldsymbol{P}_1$	covariance matrix of the initial state vector
$\boldsymbol{P}_*, \boldsymbol{P}_\infty$	elements of the covariance matrix of the initial state vector
$\boldsymbol{P}_t$	covariance matrix of the state vector
$\bar{\boldsymbol{P}}$	steady-state value of $\boldsymbol{P}_t$
$\boldsymbol{p}^*$	likelihood of the most probable path in the Viterbi algorithm
p	number of lagged conditional variance terms of a GARCH model
p_t	price of an asset at time t
$\boldsymbol{Q}_t$	variance-covariance matrix of state errors;
	variance-covariance matrix of standardized residuals
$\boldsymbol{Q}_0$	initial variance-covariance matrix of state errors
$\boldsymbol{Q}_t^\dagger$	variance-covariance matrix of state errors (extended state vector)

$\boldsymbol{Q}_t^{\beta}$	block matrix determining the nature of regression coefficients
Qsq	Box-Ljung test statistic based on squared (excess) returns
q	row dimension of diffuse vector $\boldsymbol{\delta}$;
	the order of an ARCH process
$\boldsymbol{R}_t$	system matrix in state equation;
	multivariate vector of excess returns
$\boldsymbol{R}_0$	selection matrix
$\tilde{\boldsymbol{R}}$	matrix of unconditional variances of standardized residuals
$R_{0,t}$	excess log-return of the market proxy
$R_{i,t}$	excess log-return of sector i
R^2	coefficient of determination
$\bar{R}^2$	adjusted coefficient of determination
$r_{i,t}$	log-return of sector i
$\boldsymbol{r}_t$	vector of weighted sum of future innovations
r_t^f	risk-free interest rate
r	column dimension of the system matrix $\boldsymbol{R}_t$
S_{t-1}^-, S_{t-1}^+	measures of asymmetry in tests for asymmetric GARCH effects
S_t	Markov chain
$S^{(t)}$	history of the Markov chain S_t up to time t
s^2	p.e.v. for models with time-varying regression coefficients
s_*^2	ML estimator of σ_*^2 depending on generalized recursive residuals
SIZ_t	size factor
sk	skewness of a series
$\boldsymbol{T}_t$	system matrix in state equation
T	number of dates included in the sample
TS_t	term structure factor
$\boldsymbol{u}_t$	smoothing error
$\boldsymbol{V}_t$	smoothed state variance matrix
$\boldsymbol{v}_t$	one-step ahead prediction error
$v_t^{\dagger}$	recursive residuals
$\tilde{v}_t^{\dagger}$	generalized recursive residuals
v_t^+	least squares residuals
$\tilde{v}_t^+$	generalized least squares residuals
VGS_t	value growth spread factor
w_t	sum of forward probabilities;
	weighting factor in the context of WLS
w	dimension of $\boldsymbol{\psi}$
X_t	sequence of observations
$X^{(t)}$	history of the sequence of observations X_t up to time t
$X^{(-u)}$	sequence of random variables X_t with X_u being excluded
$x^{(-u)}$	observations x_t with x_u being excluded
x_t	realized observation at time t;
	asset payoff at time t
$\boldsymbol{Y}_t$	set of observations up to time t

$\boldsymbol{y}_t$	multivariate time series vector of observations
$\boldsymbol{y}$	stacked vector of observations with $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$
$\bar{\boldsymbol{y}}_{T+l}$	minimum MSE forecast of $\boldsymbol{y}_{T+l}$ given $\boldsymbol{y}$
y_t	time series of observations
$\bar{y}$	unconditional mean of y_t
y_t^*	mean-corrected univariate time series
z_t	IID process with mean zero and variance unity
$\boldsymbol{Z}_t$	system matrix in observation equation

Mathematical symbols

$Corr(X, Y)$	correlation between X and Y
$Cov(X, Y)$	covariance between X and Y
$\text{diag}(\cdot)$	diagonal matrix operator
$\exp(x)$	the number e raised to the x power
$F(\cdot)$	empirical cumulative distribution function
$E(X)$	expectation of X
$f(\cdot)$	probability density function
$\boldsymbol{g}$	gradient vector
$g(\cdot)$	conditional Gaussian density function
$\odot$	Hadamard product (element-by-element multiplication)
$\boldsymbol{H}$	Hessian matrix
∞	infinity
$\boldsymbol{I}_n$	n-dimensional identity matrix
$I(\cdot)$	indicator function
$L(\cdot)$	likelihood function;
	lag operator with $L(x_t) := x_{t-1}$
$L_c(\cdot)$	concentrated likelihood function
$L_d(\cdot)$	diffuse likelihood function
$\log(x)$	natural logarithm of x
$N(\cdot)$	normal density function
$P(A)$	probability that event A occurs
$p(\cdot)$	probability function
$\mathbb{R}$	the set of all real numbers
S	discrete random variable
$U(\cdot), u(\cdot)$	utility function
$Var(X)$	variance of X
$\text{vech}(\cdot)$	vech operator
X	random variable (discrete or stochastic)
$\boldsymbol{X}'$	transpose of the matrix X
$(\boldsymbol{X})_{\bullet j}$	j-th column of the matrix X

Preface

The work on this book began in 2004, when equity markets had just overcome the burst of the dotcom bubble and entered a very different market regime, characterized by high returns and low volatility. The idea to have a closer look at applications of state space models in finance grew out of my experience as portfolio manager and strategist for European equities. While conducting research on a quantitative sector allocation model, it soon became obvious that traditional time series models with constant coefficients are not able to capture the dynamics of financial markets.

The key to handling models with time-varying sensitivities is the state space representation of a dynamic system. The interest in using state space models in fields away from control engineering and speech recognition, where techniques such as the Kalman filter and Markov regime switching have originated, has increased in recent years. However, a transportation of these techniques from theoretical work into applied research in the context of widely used concepts in finance is still underdeveloped. The presented material will be useful for financial economists and practitioners who are interested in taking time-variation in the relationship between financial assets and key economic factors explicitly into account. As a distinctive feature, the empirical part includes a comprehensive analysis of the ability of time-varying coefficient models to explore the conditional nature of systematic risks for European industry portfolios.

This book has been accepted as doctoral thesis by the Faculty of Economics and Business Administration of the Georg-August-Universität Göttingen for award of the degree Doctor of Economics in 2008. The original title "Applications of Advanced Time Series Models to Analyze the Time-varying Relationship between Macroeconomics, Fundamentals and Pan-European Industry Portfolios" was changed upon publication via the Göttingen University Press.

Sascha Mergner

Acknowledgements

Chapter 1

Introduction

> *"Economists study the economy both for the sheer intellectual pleasure of trying to understand the world in which they live and with the hope that improved knowledge will lead to better economic policy and performance."*
>
> Blanchard and Fischer (1989, p. 614)

A central conception of finance is the generally accepted trade-off between risk and expected return. The assessment of risk and the required risk premium is usually modeled by an asset pricing model, in which the common variation in returns is accounted for by a possibly multivariate set of risk factors. The first and still widely used pricing model is the single-factor Capital Asset Pricing Model (CAPM), proposed by Sharpe (1964) and Lintner (1965). It implies a linear relationship between an asset's expected return and its systematic risk, also referred to as beta.[1] In testing the validity of the static CAPM, various studies have demonstrated the possibility to earn risk-adjusted excess returns by forming portfolios according to fundamental attributes, such as firm size or valuation. These market anomalies, so-called because the abnormal returns related to these patterns cannot be explained under the CAPM, motivate the alternative use of pricing models that allow for multiple sources of risk. Depending on the choice of variables, important variants include fundamental and macroeconomic multifactor models. Irrespective of the number of considered systematic factors, all of these pricing models share one common property: in their basic representation, the beta coefficients are assumed to be constant over time.

In an inherently dynamic world that is characterized by changing relationships between economic agents over time (cf. Chow 1984) the paradigm of beta constancy has to be questioned. Given theoretical arguments and empirical stylized facts of financial return series, the true degree of beta can be assumed to depend on the available information at any given date. This thesis addresses the explicit modeling of time-varying

---

[1] As the focus of this thesis is on the econometric modeling of time-varying financial sensitivities and not on portfolio theory, it is not intended to provide a comprehensive derivation of factor pricing models and their assumptions. For a summary of the basic foundations of asset pricing theory and further references, see Appendix A.

sensitivities in financial markets by application of advanced contemporary time series techniques. It contributes an empirical analysis of the time-varying relationship between macroeconomics, fundamentals and pan-European industry portfolios.

1.1   The modeling of change

Motivated by theoretical arguments that systematic risks depend on micro- and macroeconomic factors, the assumption of beta stability over time has been rejected, among others, by Fabozzi and Francis (1978), Bos and Newbold (1984) and Collins et al. (1987). Jagannathan and Wang (1996) demonstrated that a conditional CAPM is better able to explain the cross-section of returns than a specification with constant coefficients; anomalies are left with little explanatory power. According to Leusner et al. (1996) this might be due to omitted variables whose impact can be captured by the conditionality of beta. It is well recognized that parameters in economic and financial settings change over time and that "the case for modeling change is compelling" (Kim and Nelson 1999, p. 4). Nevertheless, only minor efforts have been made to explicitly model systematic risk as a stochastic process and to systematically compare different modeling techniques to carry out the task. In this respect applied finance, with a focus on time-varying second moments and conditional heteroskedasticity models, is broadly lagging other disciplines, where the modeling and forecasting of conditional sensitivities has long been established.

Since the 1960s control engineers and physicists have been modeling the continuous change attached to a regression coefficient by means of an unobserved state variable as part of a so-called *state space model*. The unknown parameters in such a system of equations can be estimated via the *Kalman filter*, a very powerful and flexible recursive algorithm, named after Kalman (1960, 1963), which plays a central role in the modeling and estimation of change throughout this thesis. Engineering applications also led to the introduction of *Markov regime switching*, another class of latent variable time series models, where the observation-generating distribution depends on unobserved, discrete states modeled as a Markov chain. In economics and financial applications we are also regularly confronted with gradually or structurally shifting time series without actually observing the time-varying dynamics. However, it was not until the works of Harvey (1981), Meinhold and Singpurwella (1983) and Hamilton (1988) that applied economists and financial econometricians began to apply the Kalman filter and the Markov regime switching framework, respectively. The reason for this hesitation is twofold: notational discomfort and computational complexity. Hence, the presentation of these concepts in a "language, notation and style" (Meinhold and Singpurwella 1983, p. 123) that is familiar to economists can be considered as the most important contribution of the pioneering authors cited above. With the computational capacities nowadays offered by personal computers in combination with powerful statistical software, the computational complexity required for implementing modern time series techniques can be handled.

1.2   Research objectives

Although these advances spurred the interest in applying advanced time series techniques such as the Kalman filter and Markov regime switching models in economic and financial analysis in recent years, a transportation of these concepts from theoretical work into applied research on a broader scale is still underdeveloped. Inspired by the quote at the beginning of this chapter, this thesis analyzes the relative merits of selected elaborate econometric methods to model change in the context of widely used concepts in finance. The exploration of the dynamics of financial markets is aimed at an improved understanding and modeling of real-world phenomena. The following three research objectives are addressed:

1. Provide a notationally conformable introduction of Gaussian state space models, the Markov regime switching framework and conditional heteroskedasticity models.

2. Analyze which modeling technique is best able to model and forecast time-varying systematic beta risk as a stochastic process.

3. Evaluate the practical relevance of taking time-variation in factor sensitivities explicitly into account.

As the different contemporaneous time series models originate from different disciplines, very different notation and terminology is commonly employed to outline the respective theory behind these concepts. The first objective of this thesis is to introduce the theory of the different models at hand in a unified notational framework: linear Gaussian state space models and the Kalman filter, the Markov regime switching framework, as well as two of the most prominent models for time-varying volatility, namely autoregressive conditional heteroskedasticity (ARCH) and stochastic volatility models. This will allow the applied researcher to adopt the various concepts without having to deal with different notation that is typical for the disciplines in which the models were originally employed. It is intended to provide the methodology for the modeling of time-varying relationships in a way that is as compact and intuitive as possible and as comprehensive as necessary. The outline of the respective basic ideas and estimation procedures in Chapters 3–5 illustrates that both Markov regime switching and stochastic volatility models are closely related to the linear Gaussian state space framework and the Kalman filter.

The second contribution of this thesis is a systematic and comprehensive analysis of the ability of the different techniques under consideration to model and forecast the time-varying behavior of systematic market risk. The rationale behind starting the empirical analysis with an application of the selected time series techniques to the single-factor CAPM, is motivated by the fact that the CAPM beta is widely established in practice. It is used, for example, to calculate the cost of capital, to identify mispricings and to estimate an asset's sensitivity to the broad market. As discussed by Yao and Gao (2004) betas, while traditionally employed in the context of single stocks, are particularly useful at the sector level. However, in spite of various studies dealing with the modeling of conditional sector betas in other regions of the world, similar work in a pan-European context, where the advancement of European integration and the introduction of a single currency increased the importance of the sector perspective over recent years, is still

missing. Chapter 6 aims to close this gap by investigating the time-varying behavior of systematic risk for eighteen pan-European sectors, also referred to as industry portfolios. Using weekly data over the period 1987–2005, time-variation in beta is modeled by a bivariate t-GARCH(1,1) model, a bivariate stochastic volatility model estimated via efficient Monte Carlo likelihood, four Kalman filter based approaches as well as two Markov switching models. The main purpose of the chapter is to compare the performance of these modeling techniques including non-standard procedures such as Markov switching and stochastic volatility. Overall, the respective ex-ante forecast performances of the various modeling approaches indicate that the random walk process combined with the Kalman filter is the most accurate in describing and forecasting the time-varying behavior of sector betas in a European context.

The analysis of the dynamics in the relationship between financial variables represents an exciting area of research, motivated by both economic arguments and empirical findings. However, it is not yet clear to what degree the modeling of change is relevant in practice. The third contribution of this thesis is an empirical analysis of the practical relevance of taking time-variation in factor loadings explicitly into account. A multi-factor pricing framework that relates pan-European industry portfolios to systematic macroeconomic and fundamental risks is considered. Chapter 7 focuses on the question whether an explicit consideration of the changing importance of macroeconomics and fundamentals on European sectors can be exploited in a profitable way, either from a risk management or from a portfolio management perspective. The selection of common factors is based on the latest findings of the anomalies literature. Covering an out-of-sample period of ten years, the proposed conditional multifactor pricing model employs a market factor, three macroeconomic factors and two fundamental factors, which are motivated by the well-known three-factor model by Fama and French (1993). The stochastic process of conditional betas is modeled as individual random walks. Estimation of the model over time using the Kalman filter yields series of estimated conditional factor loadings. These are used as explanatory variables in a cross-sectional regression approach that follows Fama and MacBeth (1973). Traditionally, the Fama-MacBeth approach utilizes constant coefficients as instruments. In order to get an indication of the importance of conditional betas for the pricing of risk, in this thesis a modified formulation with time-varying sensitivities as instruments is employed. The question whether employing conditional factor loadings to derive return forecasts leads to more profitable portfolios is examined based on a series of backtests. The main results of this chapter indicate that time-varying betas carry implications for the prediction of returns, and less so for the analysis of risk.

1.3 Organization of the thesis

This thesis deals with the modeling of change in the context of widely used concepts in finance. To link the theoretical and practical parts, Chapter 2 introduces the set of sector return series and reviews their major empirical properties. The stylized facts reveal that the paradigm of stable sensitivities has to be questioned. This motivates the use of more advanced time series concepts to analyze the time-varying importance attached

to common systematic risks for pan-European industry portfolios. The stylized facts serve as a guide toward selecting the modeling techniques to be employed in the course of this thesis. Chapter 3 gives a general treatment of the class of linear Gaussian state space models and introduces the Kalman filter and smoother. Chapter 4 introduces the theoretical conception of Markov regime switching, in which the hidden dynamics are modeled as being discrete. Both modeling approaches, Kalman filtering and Markov switching, are employed to model changing coefficients directly. Chapter 5 reviews the two most important concepts of capturing conditional heteroskedasticity, which are used to derive indirect estimates of conditional betas: ARCH and stochastic volatility. Various simulation-based procedures to estimate the latter are discussed. Chapter 6 applies the selected time series concepts to model and forecast time-varying market betas for pan-European industry portfolios. Chapter 7 analyzes the practical relevance of explicitly considering conditionality in factor loadings by applying the Kalman filter to a multifactor pricing model with macroeconomic and fundamental variables. The chapter introduces a synthesis of the classical Fama-MacBeth approach with time-varying betas and conducts a series of backtests on which the evaluation is based. Chapter 8 summarizes the main results and offers suggestions for future research.

Chapter 2

Some stylized facts of weekly sector return series

A sound theoretical background of how advanced time series concepts can be applied to model changing relationships over time constitutes a prerequisite to conduct empirical research on the modeling of change. As a wide spectrum of modeling techniques is available, the first step in carrying out research is to decide *how* a problem should be approached. In this respect, the selection and specification of a model is generally driven by empirical stylized facts of the series at hand. This chapter reviews some of the properties that are inherent to the sector return series to be used throughout this thesis. Describing the characteristics of the data will emphasize the necessity of applying appropriate time series techniques that are capable of modeling the series' empirical distributional and temporal regularities.

In the financial econometrics literature, it is well documented that financial time series share a number of common features; see, for example, Ghysels et al. (1996) or Pagan (1996) for an overview. In particular, following Palm (1996), these include the following issues:

- Asset prices are usually nonstationary while returns are usually stationary.

- Returns tend to show an erratic behavior with large outlying observations occurring more frequently than may be expected under the assumption of a normal distribution.

- Returns are usually not autocorrelated, but squared returns are autocorrelated, a phenomenon known as volatility clustering.

- Some return series exhibit the so-called leverage effect, where large negative returns are followed by periods of high volatility.

- Volatilities of different assets often move together indicating potential linkages.

By employing standard statistical and graphical methods, these properties are illustrated for some selected weekly return series.

This thesis aims to analyze the time-varying importance of market, macroeconomic and fundamental factors for pan-European industry portfolios. The two major objectives

of this chapter are to motivate the modeling of change and to link the theoretical and empirical parts. The stylized empirical properties are expected to point toward the modeling techniques from the available econometric toolbox that our attention should be focused upon.

2.1   The data

The empirical analyses throughout this thesis are based on pan-European sector return series. Over the last few years, the level of integration in European equity markets highly increased. This development was spurred by the general trend toward deregulation of security markets, a growing harmonization of economic, fiscal and monetary policies as well as technical advances in information technology. The practice of top-down asset allocation experienced a radical change: to add alpha to their equity portfolios, European investors shifted their focus away from countries toward selecting sectors. The general emergence of a sector approach is reflected by the fact that many brokerage firms and institutional investors reorganized their research departments and investment processes toward sectors. The volumes traded in cross-country investment funds, sector exchange tradable funds and sector futures increased sharply in recent years (cf. Isakov and Sonney 2004).

In spite of the recent coherence of sectors across international markets and the emerging role of sector allocation in the investment process of an institutional investor, so far only relatively little work has been done analyzing European industry portfolios. This thesis contributes toward filling this gap by introducing a new data set of pan-European sectors to the literature. The main motivation to employ sector data in the empirical parts of this thesis is of statistical nature. The use of returns on aggregated sector data instead of individual stock returns largely eliminates idiosyncratic variation (cf. Moskowitz and Grinblatt 1999). This is likely to lead to a reduction of potential errors-in-variables problems, which should result in better estimates of (conditional) betas; see, for example, Chen et al. (1986). Besides, following Lo and MacKinlay (1988) systematic influences can be expected to be uncovered more readily in the context of grouped data with idiosyncratic noise being subdued.

The data set consists of weekly excess return series calculated from the total return indices for eighteen pan-European industry portfolios. The choice of a weekly frequency represents a compromise between too noisy daily data and too sluggish monthly data. All sector indices are from Stoxx Ltd. (2004), a joint venture of Deutsche Boerse AG, Dow Jones & Company and the SWX Group that develops a global free-float weighted index family, the Dow Jones (DJ) Stoxx℠ indices. The DJ Stoxx℠ Broad return index, which consists of the 600 largest European stocks by market capitalization, serves as a proxy for the overall market. All indices are expressed in Euros as common currency. With the exception of three sectors, the sample period runs from 2 December 1987 until 2 February 2005 (897 weekly observations). In September 2004, Stoxx Ltd. switched its sector definitions from the DJ Stoxx℠ Global Classification Standard to the Industry Classification Benchmark (ICB). The sectors Cyclical Goods & Services, Non-Cyclical Goods & Services and Retail (old) were replaced by the new sectors Travel & Leisure,

Table 2.1: *The DJ Stoxx^{SM} sector classification.*

Industries	Supersectors
Basic Materials	Basic Resources
	Chemicals
Consumer Goods	Automobiles and Parts
	Food and Beverages
	Personal and Household Goods
Consumer Services	Media
	Retail
	Travel and Leisure
Financials	Banks
	Financial Services
	Insurance
Health Care	Health Care
Industrials	Construction and Materials
	Industrial Goods and Services
Oil and Gas	Oil and Gas
Technology	Technology
Telecommunications	Telecommunications
Utilities	Utilities

Personal & Household Goods and Retail (new), respectively. The history for the newly formed sectors begins 31 December 1991, such that for these three series only 683 weekly return observations are available. Table 2.1 presents the first two tiers of the ICB sector structure.

Weekly returns, $r_{i,t}$, between period t and $t-1$ for index i are compounded continuously as

$$r_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1}), \tag{2.1}$$

for $t = 1, \ldots, T$, and $i = 0, \ldots, 18$, where $i = 0$ refers to the market proxy and $i > 0$ refers to the sector series; $p_{i,t}$ is Wednesday's index closing price in week t and log is the natural logarithm. Returns in excess of a risk-free interest rate, so-called *excess returns*, denoted as $R_{i,t}$, are defined as

$$R_{i,t} = r_{i,t} - r_t^f, \tag{2.2}$$

where r_t^f is the risk-free rate of return, which is calculated from the 3-month Frankfurt Interbank Offered Rate (FIBOR). As the FIBOR yields (fib_t) are in percentage per annum, they are converted to weekly rates as $r_t^f = (1 + fib_t/100)^{1/52} - 1$. All data are obtained from Thomson Financial Datastream.

In order to get a better understanding of the segmentation of the European equity market, Figure 2.1 displays a hierarchical cluster dendrogram for the given sample of excess sector returns. The purpose of the cluster analysis is to divide the dataset into groups —
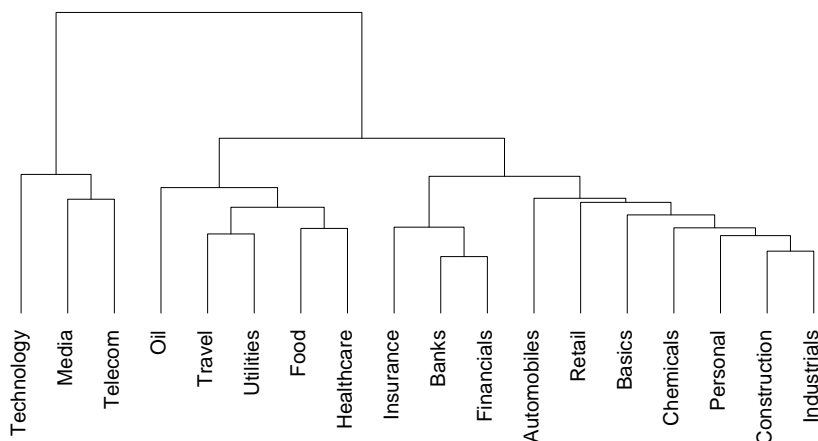
Figure 2.1: *Hierarchical cluster dendrogram for the set of excess sector returns.*

or clusters — of observations with similar characteristics. Hierarchical methods like the one employed here do not require the number of clusters to be predetermined.[2]

From top to bottom three major clusters can be identified. The outer left cluster consists of the three sectors Technology, Media and Telecommunications (TMT). The chart suggests that the TMT cluster is isolated from the rest of the market. This reflects the sectors' outstanding role during the *dotcom* bubble at the end of the 1990s. The center group of sectors ranges from Oil & Gas to Healthcare and represents the defensive end of the market. The remaining sectors constitute the cyclical part of the market. By moving the corresponding subtree one level further down, one can differentiate between financial and non-financial cyclicals.

2.2   Empirical properties

Figure 2.2 displays the time series of original prices, of excess returns as defined by (2.2) and of squared excess returns for the overall market and the two sectors Insurance and Food & Beverage. An informal first look at these graphs suggests that the original sector series in levels are trending. Formally, the existence of a unit root can be confirmed by an augmented Dickey-Fuller test or the Phillips-Perron test. As the existence of unit roots in financial price series is extensively discussed in the empirical literature, the corresponding results are not reported here; for a detailed outline of unit root tests, see, for example, Pagan (1996). In the following, only excess returns will be considered.

---

[2]The dendrogram has been generated using the `hclust()` function of the open-source statistical software package `R 2.1.1` (R Development Core Team 2005) which can be downloaded from `www.r-project.org`. For an introductory outline of cluster analysis and its implementation, see, for example, Struyf et al. (1996) and Kaufman and Rousseeuw (2005).
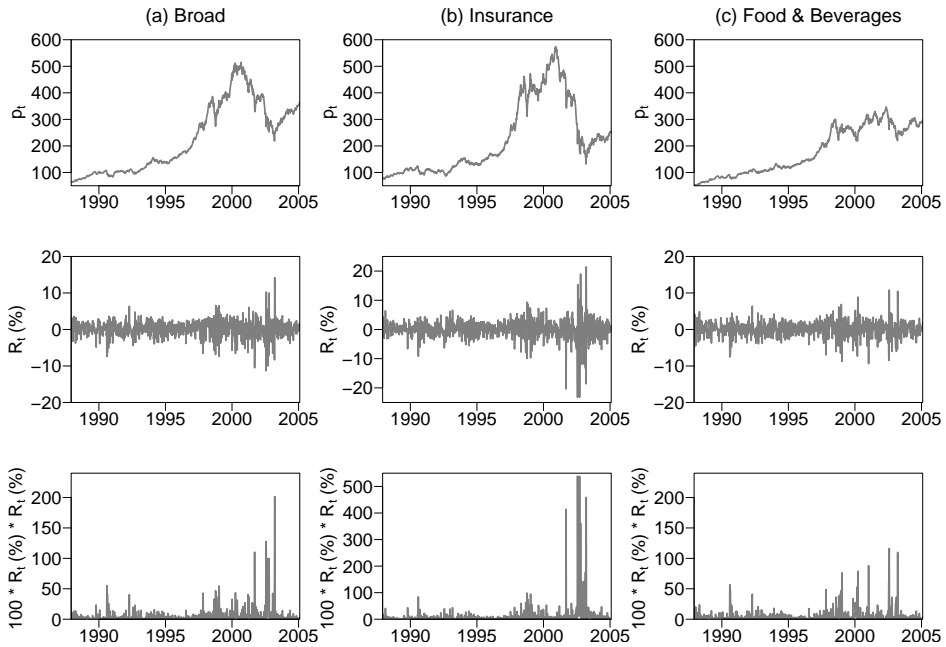
Figure 2.2: *Summaries of the weekly returns on the (i) broad market, (ii) the Insurance sector and (iii) Food & Beverages. Summaries from top to bottom are time series of original prices, excess returns and squared excess returns.*

Another obvious property illustrated by the returns and squared return series is that weeks of large absolute movements are followed by weeks with the same characteristics. This phenomenon, commonly referred to as *volatility clustering*, can be particularly observed in the second half of the sample, which includes the Asian crisis (1997), the Russian debt crisis (1998) and the boom and bust of the *dotcom* bubble (1998–2003).

Univariate descriptive statistics for the data and some standard test statistics, which are referred to in the following subsections, are provided in Table 2.2. It can be seen that over the entire sample, the highest average weekly excess returns are offered by Healthcare (0.17%), Oil & Gas (0.15%) and Utilities (0.15%). The lowest are realized for Automobiles & Parts (0.02%), Insurance (0.04%) and Retail (0.06%). The risk, as measured by the annualized standard deviation, ranges from 14.64% for the defensive Utilities to 30.45% for the high risk sector Technology.

2.2.1 Thick tails

The observed degree of kurtosis (ku) of market and sector returns reported in Table 2.2 generally exceeds the normal value of three. Compared to a normal distribution, the peaks are higher and the tails are heavier, which reflects that large outlying observations occur more often than can be expected under the assumption of normality. According

Table 2.2: *Summary statistics of weekly excess returns.*

Sector	T	μ^a	σ^b	sk^c	ku^d	JB^e	Qsq^f	lev^g	$\rho_{0i}{}^h$
Broad	*897*	*0.10*	*16.62*	*−0.30*	*6.83*	*560*	*332*	*−0.20*	*−*
Automobiles	897	0.02	23.77	−0.56	6.30	453	220	−0.15	0.80
Banks	897	0.14	19.47	−0.28	7.49	766	451	−0.18	0.91
Basics	897	0.12	20.47	−0.24	5.13	177	153	−0.13	0.73
Chemicals	897	0.09	18.52	−0.19	7.87	890	147	−0.14	0.81
Construction	897	0.08	17.69	−0.32	4.97	160	215	−0.14	0.83
Financials	897	0.07	18.69	−0.63	8.73	1287	378	−0.18	0.89
Food	897	0.10	15.29	−0.27	5.86	318	122	−0.14	0.70
Healthcare	897	0.17	18.23	0.18	5.52	243	146	−0.16	0.71
Industrials	897	0.07	17.86	−0.47	5.69	303	272	−0.19	0.91
Insurance	897	0.04	24.09	−0.85	13.97	4607	431	−0.18	0.88
Media	897	0.07	24.69	−0.62	9.89	1832	291	−0.09	0.82
Oil & Gas	897	0.15	19.22	−0.02	5.56	246	296	−0.16	0.66
Personal	683	0.09	18.57	−0.22	4.95	114	108	−0.22	0.86
Retail	683	0.06	21.50	−0.78	10.32	1594	49	−0.12	0.78
Technology	897	0.07	30.45	−0.55	6.68	553	603	0.01	0.81
Telecom	897	0.13	24.81	−0.18	5.36	213	325	−0.07	0.80
Travel	683	0.07	16.90	0.10	6.36	322	118	−0.19	0.80
Utilities	897	0.15	14.64	−0.45	5.15	203	171	−0.13	0.79

[a]The mean is expressed in percentage terms.

[b]The standard deviation is expressed in annualized percentage terms.

[c]sk denotes the skewness of the return series.

[d]ku denotes the kurtosis of the return series.

[e]JB is the Jarque-Bera statistic for testing normality. The test statistic is asymptotically χ^2 distributed with 2 degrees of freedom. The relevant critical value at the 99% level is 9.21.

[f]Qsq is the Ljung-Box portmanteau test for the null of no autocorrelation in the squared excess returns up to order 12. The test statistic is asymptotically χ^2 distributed with 12 degrees of freedom. The relevant critical value at the 99% level is 26.22.

[g]lev is the unconditional correlation between the squared excess return at date t and the excess return at date $t − 1$ for index i. Negative values for lev indicate that large volatility tends to follow upon negative returns.

[h]ρ_{0i} is the unconditional correlation between sector i and the overall market.

to the stated values of skewness (sk), the overall market as well as all sectors, except for Healthcare and Travel & Leisure, are negatively skewed. This might be an indication that large negative returns occur more often than large positive returns. The Jarque-Bera test statistics, as reported in column JB, confirm the departure from normality for all return series at the 1% significance level. Since the seminal works by Mandelbrot (1963) and Fama (1965) it is well known that many asset returns follow a leptokurtic distribution. The shortcomings related to the regularly made normality assumption are commonly addressed either by employing a fat-tailed distribution, such as Student-t, or by relying on a mixture of normals; for an overview of references on the subject, see Bollerslev (1987) or, more recently, Gettinby et al. (2004).

2.2.2   Volatility clustering

It appeared from Figure 2.2 that quiet periods, characterized by relatively small returns, alternate with relatively volatile periods, where price changes are rather large. This can be confirmed by looking at the first-order autocorrelation (AC) function of returns and squared returns in Figure 2.3. While the autocorrelations of the return series only show minor activity, the autocorrelation function of squared returns show significant correlations up to an extended lag length. The corresponding Box-Ljung statistics, reported in column Qsq of Table 2.2, confirm significant correlations for the squared returns at the 1% level for all series.

The phenomenon of persistently changing volatility over time has been first recognized by Mandelbrot (1963) and Fama (1965) who related volatility clustering to the observation of fat-tailed returns. The two most important concepts to deal with time-varying volatility are (i) the *autoregressive conditional heteroskedasticity* (ARCH) model by Engle (1982) together with the *generalized ARCH* (GARCH) model by Bollerslev (1986), and (ii) the *stochastic volatility* (SV) model by Taylor (1982, 1986). An alternative way to deal with conditional volatility is to employ a *Markov regime switching* model: the generation of excess returns under different volatility regimes is governed by different states, each driven by a first-order Markov chain. Leading papers on the subject include Hamilton and Susmel (1994) and Turner et al. (1989).

2.2.3   Leverage effects

Another common property of financial return series is the so-called *leverage effect*, where future volatility depends negatively on the sign of past returns (Black 1976). Column *lev* of Table 2.2 reports the estimated correlation coefficients between squared excess returns at date t and excess returns at date $t - 1$ for all sectors and the overall market. With the exception of Technology, all estimates are negative. This gives a first indication that negative sector returns are followed by a pick-up in volatility.

2.2.4   Volatility co-movements

From the squared excess returns shown in the bottom row of Figure 2.2, it can be seen that volatility peaks tend to occur at the same time for all three series (August 1990, January 1999, period between July 2002 and March 2003). This is a common observation as certain newsflow usually affects the volatility of different assets simultaneously. This phenomenon might lead to a failure of joint stationarity of two return series even though stationarity for the individual series holds. A possible solution to this problem is to employ a multivariate GARCH or a multivariate SV framework to deal with *volatility co-movements* by modeling conditional covariances. Important papers dealing with multivariate conditional heteroskedasticity include Diebold and Nerlove (1989), Harvey et al. (1992) and Harvey et al. (1994).
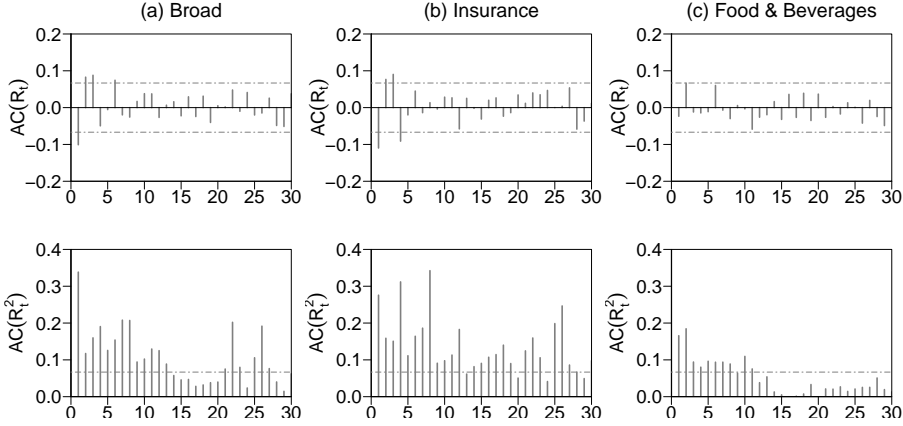
Figure 2.3: *Autocorrelation functions (first 30 lags) of the broad market, the Insurance sector and Food & Beverages. The top row displays correlograms of excess returns, the bottom row displays correlograms of the corresponding squares.*

2.3   Implications

The last column of Table 2.2 reports the unconditional correlation between the chosen market proxy and sector i for $i = 1, \ldots, 18$, denoted as ρ_{0i}. Correlations between an asset and the overall market play an important role in financial markets. They are widely used in portfolio and risk management applications. For the sample under consideration, all estimated values of ρ_{0i} are higher than 0.65. This indicates a strong linear association between the respective sectors and the market index over the entire sample. However, given the empirical properties described above, especially with regard to the observed time-variation of volatilities and covariances, it is reasonable to assume that the true degree of correlation is not constant but changing over time.

   This is illustrated by considering the following regression relationship between sector i and the market proxy:

$$R_{i,t} = \alpha_i + \beta_i R_{0,t} + \epsilon_{i,t}, \qquad t = 1, \ldots, T. \tag{2.3}$$

The ordinary least squares (OLS) estimator of the slope coefficient in this regression represents the usual estimator of an asset's *beta* in the context of the CAPM, which will be dealt with in more detail in the empirical part of this thesis. Under the assumption that the true value of beta is constant, it is defined as

$$\beta_i = \frac{Cov(R_0, R_i)}{Var(R_0)}, \tag{2.4}$$

where $Cov(R_0, R_i)$ is the unconditional covariance of the return of the market proxy with the return of sector i, and $Var(R_0)$ is the unconditional variance of market returns. Taking the observed stylized facts of volatility clustering and volatility co-movements
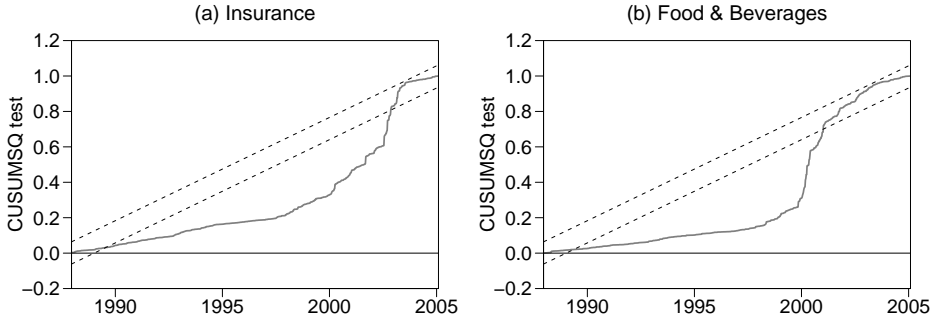
Figure 2.4: *CUSUMSQ tests with 5% confidence intervals for the excess return series of (a) Insurance and (b) Food & Beverages.*

into account, the numerator and denominator of (2.4) should be replaced by the conditional covariance and conditional variance, respectively, to allow the true beta to be time-varying.

Based on the cumulative sum of squares (CUSUMSQ) of recursive residuals, Figure 2.4 demonstrates the instability of beta at the example of the two sectors Insurance and Food & Beverages. The CUSUMSQ test, as proposed by Brown et al. (1975), detects instability in the regression coefficients caused by time-varying second moments if the cumulative sum of squares moves outside two critical lines. For both sectors, the signal line moves outside the lower critical bound. This indicates instability in the relationship between the respective sector and the overall market during the sample period. Similar observations leading to the same conclusion have been made for all sectors (not reported here). The test statistics and the significance lines have been computed using the object-oriented matrix programming language `Ox 3.30` by Doornik (2001) together with the package `SsfPack 2.3` by Koopman et al. (1999).

The findings above contradict the assumption of beta constancy and further motivate the scope of this thesis to deal with time-varying sensitivities in financial markets. Given previous evidence in the literature and economic arguments, which contradict the assumption of beta constancy, the focus throughout this thesis will be on the modeling of change, and not on testing the paradigm of beta stability. For an overview of alternative parameter stability tests in the context of beta, see, for example, Wells (1996, §2).

With regard to the modeling of conditional relationships in financial markets, the literature distinguishes between two different approaches: time-varying sensitivities can be modeled as linear functions of observable state variables as proposed, for example, by Shanken (1990). However, as it is not clear which instrumental variables should be included, any choice may exclude relevant conditioning information. An alternative procedure is to rely on advanced time series techniques to model time-varying betas as stochastic, possibly hidden processes. As demonstrated by Leusner et al. (1996) a stochastic process may capture the impact of the complete set of potential determinants of systematic risk, thus avoiding the omitted variables problem. Following this route,

throughout this thesis conditionality is dealt with by employing time series models. As it is typically not possible to model all distributional and temporal properties of return series simultaneously, different models are typically used to capture different empirical regularities (cf. Ghysels et al. 1996).

Time-varying relationships can be constructed either *directly* or *indirectly*. The direct approach will be implemented by employing a state space framework, where beta can be allowed to emerge either as a *continuous process* estimated via the Kalman filter, or as a *discrete process* in a Markov regime switching framework. Alternatively, indirect estimates of conditional sensitivities can be derived by capturing the underlying conditional variance and covariance components by a conditional heteroskedasticity model. Before applying these different concepts to analyze the time-varying relationship between macroeconomics, fundamentals and pan-European industry portfolios, the theoretical groundwork is made available in the subsequent theoretical part of this thesis.

Chapter 3

Linear Gaussian state space models and the Kalman filter

This chapter introduces the class of linear Gaussian state space models from the classical perspective of maximum likelihood estimation. The first section outlines the basic ideas behind state space modeling. Section 3.2 presents the general state space form of a dynamic system. Section 3.3 develops the Kalman filter and smoother, the basic tools to estimate models in state space form. Section 3.4 describes maximum likelihood estimation procedures for the unknown parameters. Section 3.5 extends the general state space model to allow for the incorporation of explanatory variables with constant as well as time-varying parameters. Model diagnostics and measures to assess the goodness of fit are described in Section 3.6. The chapter finishes with an illustration of how to specify state space models using the software package `SsfPack`.

3.1 Basic ideas of state space modeling

The state space form of a dynamic system with unobserved components is a very powerful and flexible instrument. A wide range of all linear and many nonlinear time series models can be handled, including regression models with changing coefficients, autoregressive integrated moving average (ARIMA) models and unobserved component models. A state space model consists of a *state equation* and an *observation equation*. While the state equation formulates the dynamics of the state variables, the observation equation relates the observed variables to the unobserved state vector. The state vector can contain trend, seasonal, cycle and regression components plus an error term. Models that relate the observations over time to different components, which are usually modeled as individual random walks, are referred to as *structural time series models*. The stochastic behavior of the state variable, its relationship to the data and the covariance structure of the errors depend on parameters that are also generally unknown. The state variable and the parameters have to be estimated from the data. Maximum likelihood estimates of the parameters can be obtained by applying the *Kalman filter*. Named after Kalman (1960, 1963) the Kalman filter is a recursive algorithm that computes estimates for the unobserved components at time t, based on the available information at the same date.

The Kalman filter has originally been applied by engineers and physicists to estimate the state of a noisy system. The classic Kalman filter application is the example of tracking an orbiting satellites whose exact position and speed, which are not directly measurable at any point of time, can be estimated using available data and well established physical laws. A discussion of engineering-type applications of the Kalman filter is provided by Anderson and Moore (1979). In economics and finance, we are regularly confronted with similar situations: either the exact value of the variable of interest is unobservable or the possibly time-varying relationship between two variables is unknown. Nevertheless, the propagation of the Kalman filter among econometricians and applied economists only really began with the introductory works of Harvey (1981) and Meinhold and Singpurwella (1983).

In contrast to the Box-Jenkins methodology, which still plays an important role in teaching and practicing time series analysis, the state space approach allows for a structural analysis of univariate as well as multivariate problems. The different components of a series, such as trend and seasonal terms, and the effects of explanatory variables are modeled explicitly. They do not have to be removed prior to the main analysis as is the case in the Box-Jenkins framework. Besides, state space models do not have to be assumed to be homogeneous, which results in a high degree of flexibility. This allows for time-varying regression coefficients, missing observations and calendar adjustments. Transparency is another important feature of structural models as they allow for a visual examination of the single components to check for derivations from expectations; see Durbin and Koopman (2001, §3.5) for a comparison of the state space framework and the Box-Jenkins approach.

Early applications of state space models and the Kalman filter to economics include Fama and Gibbons (1982) who model the unobserved ex-ante real interest rate as a state variable that follows an AR(1) process. Clark (1987) uses an unobserved-components model to decompose quarterly real GNP data into the two independent components of a stochastic trend component and a cyclical component. Another important contribution is the work of Stock and Watson (1991) who define an unobserved variable, which represents the state of the business cycle, to measure the common element of co-movements in various macroeconomic variables. Surveys on the applicability of the state space approach to economics and finance can be found in Hamilton (1994a) and Kim and Nelson (1999).

The state space approach offers attractive features with respect to their generality, flexibility and transparency. The lack of publicly available software to estimate these models has been the main reason why only relatively few economic and finance related problems have been analyzed in state space form so far. The subsequent sections aim at providing a presentation of the Gaussian state space model that is as compact and intuitive as possible, while being as comprehensive as necessary to render the employment of this versatile framework by applied researchers possible. More detailed treatments of state space models are given by Harvey (1989), Harvey and Shephard (1993) and Hamilton (1994a), among others. An outline with a focus on applications can be found at Kim and Nelson (1999). If not indicated otherwise, Durbin and Koopman (2001, §4–7) serve as standard reference for this chapter.

3.2   The state space form of a dynamic system

Let $\boldsymbol{y}_t$ denote a $N \times 1$ multivariate time series vector of observations whose development over time can be characterized in terms of an unobserved state vector $\boldsymbol{\xi}_t$ of dimension $m \times 1$, each at date t. A general linear Gaussian state space model that describes the dynamics of this system can be written as

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{T}_t \boldsymbol{\xi}_t + \boldsymbol{c}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t), \tag{3.1}$$

$$\boldsymbol{y}_t = \boldsymbol{Z}_t \boldsymbol{\xi}_t + \boldsymbol{d}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{H}_t), \tag{3.2}$$

for $t = 1, \ldots, T$. Equation (3.1) is known as the *state* or *transition equation* and (3.2) is referred to as the *observation* or *measurement equation*. The deterministic parameter matrices $\boldsymbol{T}_t$, $\boldsymbol{R}_t$ and $\boldsymbol{Z}_t$ are of dimension $m \times m$, $m \times r$ and $N \times m$, respectively, with $\boldsymbol{R}_t$ being simply the identity.[3] Unobserved structural components such as trend, seasonal and cycle may be modeled by an appropriate definition of $\boldsymbol{Z}_t$ and $\boldsymbol{\xi}_t$. The $m \times 1$ and $N \times 1$ vectors $\boldsymbol{c}_t$ and $\boldsymbol{d}_t$ can be used to represent known effects to the expected value of the observations and states by including dummy variables, or to incorporate explanatory variables with fixed coefficients. If not mentioned otherwise, $\boldsymbol{c}_t$ and $\boldsymbol{d}_t$ are set to zero in the following.

The $r \times 1$ and $N \times 1$ vectors $\boldsymbol{\eta}_t$ and $\boldsymbol{\epsilon}_t$ are serially uncorrelated, normally distributed error terms with mean zero and positive definite covariance matrices $\boldsymbol{Q}_t$ and $\boldsymbol{H}_t$ of dimensions $r \times r$ and $N \times N$, respectively:

$$E(\boldsymbol{\eta}_t \boldsymbol{\eta}_\tau') = \begin{cases} \boldsymbol{Q}_t & \text{for } t = \tau \\ \boldsymbol{0} & \text{otherwise,} \end{cases} \tag{3.3}$$

$$E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_\tau') = \begin{cases} \boldsymbol{H}_t & \text{for } t = \tau \\ \boldsymbol{0} & \text{otherwise.} \end{cases} \tag{3.4}$$

The state and observation disturbances are further assumed to be uncorrelated with each other at all lags:[4]

$$E(\boldsymbol{\eta}_\tau \boldsymbol{\epsilon}_t') = \boldsymbol{0}, \qquad \text{for all } \tau, t = 1, \ldots, T, \tag{3.5}$$

and to be independent from the initial state vector $\boldsymbol{\xi}_1$:

$$E(\boldsymbol{\eta}_t \boldsymbol{\xi}_1') = \boldsymbol{0}, \qquad E(\boldsymbol{\epsilon}_t \boldsymbol{\xi}_1') = \boldsymbol{0}, \qquad t = 1, \ldots, T. \tag{3.6}$$

The initial state vector, which is of dimension $m \times 1$, is assumed to be normally distributed with the $m \times 1$ mean vector $\boldsymbol{a}_1$ and the $m \times m$ covariance matrix $\boldsymbol{P}_1$:

$$\boldsymbol{\xi}_1 \sim N(\boldsymbol{a}_1, \boldsymbol{P}_1), \tag{3.7}$$

---

[3]In alternative specifications of the general state space model, the disturbances in the state equation are often defined as $\boldsymbol{\eta}_t^* = \boldsymbol{R}_t \boldsymbol{\eta}_t$ with the corresponding covariance matrix $\boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t'$. However, in case where $\boldsymbol{Q}_t$ is singular and $r < m$, it is useful to include $\boldsymbol{R}_t$ in front of the state disturbance term to work with nonsingular $\boldsymbol{\eta}_t$ rather than singular $\boldsymbol{\eta}_t^*$; see Durbin and Koopman (2001, §3.1) for details.

[4]Assumption (3.5) can be relaxed by allowing for correlated observation and state disturbances; see, for example, Anderson and Moore (1979, §5.3) or Harvey (1989, §3.2.4).

where $\boldsymbol{a}_1$ and $\boldsymbol{P}_1$ are assumed to be known for now. The initialization of the state vector will be discussed in §3.3.6.

The matrices $\boldsymbol{T}_t$, $\boldsymbol{c}_t$, $\boldsymbol{R}_t$, $\boldsymbol{Z}_t$, $\boldsymbol{d}_t$, $\boldsymbol{Q}_t$ and $\boldsymbol{H}_t$ are called *system matrices*. Unless stated otherwise, they are assumed to be non-stochastic, i.e. they change over time in a predetermined way. This leads to a system in which $\boldsymbol{y}_t$ can be expressed as a linear combination of $\boldsymbol{\xi}_1$ and past and present state and observation disturbances. A state space model with constant system matrices over time is referred to as being *time-homogeneous*.

Usually, at least some of the elements of the system matrices $\boldsymbol{T}_t$ and $\boldsymbol{R}_t$ in the state equation and $\boldsymbol{Z}_t$, $\boldsymbol{Q}_t$ and $\boldsymbol{H}_t$ in the observation equation depend on a vector $\boldsymbol{\psi}$ of unknown parameters, which are referred to as *hyperparameters*. While the hyperparameters control for the stochastic characteristics of the model, the parameters related to the vectors $\boldsymbol{c}_t$ and $\boldsymbol{d}_t$ have a deterministic impact on the observations and the expected state value. The values of the system matrices are typically unknown. Their estimation on the basis of the available observations will be the subject of §3.4. For now, the system matrices are assumed to be known.

Throughout this thesis, the software package `SsfPack 2.3` by Koopman et al. (1999) is employed to apply the Kalman filter to estimate models in state space form. `SsfPack` is a comprehensive collection of `C` routines for estimating models in state space form with general routines for smoothing, filtering, likelihood evaluation and forecasting linked to the computing environment `Ox 3.30` by Doornik (2001). With regard to the use of `SsfPack`, it will prove useful to represent the state space model introduced by (3.1)–(3.7) in compact form as

$$\left[\begin{array}{c} \boldsymbol{\xi}_{t+1} \\ \boldsymbol{y}_t \end{array}\right] = \boldsymbol{\Phi}_t \boldsymbol{\xi}_t + \boldsymbol{\delta}_t + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_t), \qquad t = 1, \ldots, T, \tag{3.8}$$

with

$$\boldsymbol{\Phi}_t = \left[\begin{array}{c} \boldsymbol{T}_t \\ \boldsymbol{Z}_t \end{array}\right], \quad \boldsymbol{\delta}_t = \left[\begin{array}{c} \boldsymbol{c}_t \\ \boldsymbol{d}_t \end{array}\right], \quad \boldsymbol{v}_t = \left[\begin{array}{c} \boldsymbol{R}_t \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_t \end{array}\right], \quad \boldsymbol{\Omega}_t = \left[\begin{array}{cc} \boldsymbol{Q}_t & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_t \end{array}\right], \tag{3.9}$$

where $\boldsymbol{\Phi}_t$ is of dimension $(m + N) \times m$, $\boldsymbol{\delta}_t$ is $(m + N) \times 1$, $\boldsymbol{v}_t$ is $(m + N) \times 1$ and $\boldsymbol{\Omega}_t$ is $(m + N) \times (m + N)$. The initial values are collected in the $(m + 1) \times m$ matrix $\boldsymbol{\Sigma}$ which is defined as

$$\boldsymbol{\Sigma} = \left[\begin{array}{c} \boldsymbol{P}_1 \\ \boldsymbol{a}_1' \end{array}\right]. \tag{3.10}$$

For a survey on using `SsfPack` for state space modeling in macroeconomics and finance, see Zivot et al. (2002).

3.3   The Kalman filter and smoother

Once a model is put into state space form, the Kalman filter can be employed to compute optimal forecasts of the mean and covariance matrix of the normally distributed state vector $\boldsymbol{\xi}_{t+1}$, based on the available information through time t. This section outlines how the Kalman filter can be used for estimation by *filtering* and *smoothing*: filtering bases an inference about the state vector only on the information up to time t; smoothing

incorporates the full set of information in the sample, where one distinguishes between *state smoothing* and *disturbance smoothing*. Subsection 3.3.4 shows how the Kalman filter deals with *missing observations*, which serves as a basis for the subsequent outline of using the Kalman filter for *forecasting* purposes. The *initialization* of the Kalman filter will be discussed in §3.3.6. The section closes with an interpretation of the Kalman filter when used in combination with non-Gaussian disturbances. The results in this section are given without proof and are primarily taken from Durbin and Koopman (2001, §4) where derivations can be found.

3.3.1 Filtering

The purpose of filtering is to update our knowledge of the state vector as soon as a new observation $\boldsymbol{y}_t$ becomes available. Reconsidering the state space model given by (3.1) through (3.7), the Kalman filter can be derived under the assumption that $\boldsymbol{a}_1$ and $\boldsymbol{P}_1$ are known. The objective is to get the conditional distribution of $\boldsymbol{\xi}_{t+1}$ for $t = 1, \ldots, T$, based on $\boldsymbol{Y}_t$, the set of observations up to time t, i.e. $\boldsymbol{Y}_t = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$.

3.3.1.1 The general form of the Kalman filter

Given the assumption of normality in all distributions of the system, all conditional distributions related to subsets of variables based on other subsets also have to be normally distributed. The required conditional distribution of $\boldsymbol{\xi}_{t+1}$ can be characterized by its mean $\boldsymbol{a}_{t+1}$ and covariance $\boldsymbol{P}_{t+1}$:

$$\boldsymbol{a}_{t+1} = E(\boldsymbol{\xi}_{t+1}|\boldsymbol{Y}_t), \tag{3.11}$$
$$\boldsymbol{P}_{t+1} = Var(\boldsymbol{\xi}_{t+1}|\boldsymbol{Y}_t), \tag{3.12}$$

where the mean of the conditional distribution of $\boldsymbol{\xi}_{t+1}$ represents an optimal estimator of the state vector at time $t + 1$: it minimizes the *mean squared error* (MSE) matrix, $E((\boldsymbol{\xi}_{t+1} - \boldsymbol{a}_{t+1})(\boldsymbol{\xi}_{t+1} - \boldsymbol{a}_{t+1})'|\boldsymbol{Y}_t)$, for all $\boldsymbol{\xi}_{t+1}$.

Under the assumption that $\boldsymbol{\xi}_t$ given $\boldsymbol{Y}_{t-1}$ is normally distributed with mean $\boldsymbol{a}_t$ and covariance $\boldsymbol{P}_t$, it can be shown that $\boldsymbol{a}_{t+1}$ and $\boldsymbol{P}_{t+1}$ can be calculated recursively from $\boldsymbol{a}_t$ and $\boldsymbol{P}_t$ as

$$\boldsymbol{a}_{t+1} = \boldsymbol{T}_t \boldsymbol{a}_t + \boldsymbol{K}_t \boldsymbol{v}_t, \tag{3.13}$$

and

$$\boldsymbol{P}_{t+1} = \boldsymbol{T}_t \boldsymbol{P}_t \boldsymbol{L}_t' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t', \tag{3.14}$$

with

$$\boldsymbol{v}_t = \boldsymbol{y}_t - E(\boldsymbol{y}_t|\boldsymbol{Y}_{t-1}) = \boldsymbol{y}_t - \boldsymbol{Z}_t \boldsymbol{a}_t, \tag{3.15}$$
$$\boldsymbol{F}_t = Var(\boldsymbol{v}_t) = \boldsymbol{Z}_t \boldsymbol{P}_t \boldsymbol{Z}_t' + \boldsymbol{H}_t, \tag{3.16}$$
$$\boldsymbol{K}_t = \boldsymbol{T}_t \boldsymbol{P}_t \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1}, \tag{3.17}$$
$$\boldsymbol{L}_t = \boldsymbol{T}_t - \boldsymbol{K}_t \boldsymbol{Z}_t, \tag{3.18}$$

for $t = 1, \ldots, T$. The $m \times N$ matrix $\boldsymbol{K}_t$ is referred to as the *Kalman gain* matrix. The $N \times 1$ vector $\boldsymbol{v}_t$ is the *one-step ahead prediction error* of $\boldsymbol{y}_t$ given $\boldsymbol{Y}_{t-1}$. It represents

the new information contained in the latest observation. This *innovation* term with $E(\boldsymbol{v}_t|\boldsymbol{Y}_{t-1}) = E(\boldsymbol{v}_t) = \boldsymbol{0}$ and $Cov(\boldsymbol{y}_\tau, \boldsymbol{v}_t) = \boldsymbol{0}$ for $\tau = 1, \ldots, t-1$, has a major impact on the updating process for the estimator of $\boldsymbol{\xi}_{t+1}$. The matrices $\boldsymbol{F}_t$ and $\boldsymbol{L}_t$ are of dimensions $N \times N$ and $m \times m$, respectively. $\boldsymbol{F}_t$ is assumed to be nonsingular.

The system of recursions (3.13) through (3.18) is called the *Kalman filter*. For more details, see Durbin and Koopman (2001, §4.2) who base the derivation of the filter on the application of standard properties of the multivariate normal distribution.[5]

3.3.1.2   The steady-state Kalman filter

For time-homogeneous state space models, it can be shown that the Kalman recursions for $\boldsymbol{P}_{t+1}$ and $\boldsymbol{K}_t$ converge to fixed matrices:

$$\lim_{t\to\infty} \boldsymbol{P}_{t+1} = \bar{\boldsymbol{P}}, \tag{3.19}$$

$$\lim_{t\to\infty} \boldsymbol{K}_t = \bar{\boldsymbol{K}}. \tag{3.20}$$

The Kalman filter solutions that are reached after convergence to $\bar{\boldsymbol{K}}$ and $\bar{\boldsymbol{P}}$ are called *steady-state solutions*. The computational costs related to filtering can be considerably reduced by replacing $\boldsymbol{P}_{t+1}$ and $\boldsymbol{K}_t$ by the steady-state solutions; see Hamilton (1994b, §13.5) for details.

3.3.2   State smoothing

State smoothing denotes the estimation of $\boldsymbol{\xi}_t$ based on $\boldsymbol{Y}_T$, the full set of observations. For convenience, $\boldsymbol{Y}_T$ can be represented as a stacked vector $\boldsymbol{y}$ with $\boldsymbol{y} = [\boldsymbol{y}_1' \cdots \boldsymbol{y}_T']'$. As the system's distributions are all normal, $\boldsymbol{\xi}_t$ given $\boldsymbol{y}$ is conditionally normal distributed with mean $\hat{\boldsymbol{\xi}}_t$ and covariance $\boldsymbol{V}_t$:

$$\hat{\boldsymbol{\xi}}_t = E(\boldsymbol{\xi}_t|\boldsymbol{y}), \tag{3.21}$$

$$\boldsymbol{V}_t = Var(\boldsymbol{\xi}_t|\boldsymbol{y}), \tag{3.22}$$

where $\hat{\boldsymbol{\xi}}_t$ is the *smoothed state* and $\boldsymbol{V}_t$ is the *smoothed state variance*. The notion of *state smoothing* refers to the process of computing $\hat{\boldsymbol{\xi}}_1, \ldots, \hat{\boldsymbol{\xi}}_T$. The objective is to compute $\hat{\boldsymbol{\xi}}_t$ and $\boldsymbol{V}_t$ by recursion under the assumption of (3.7) with $\boldsymbol{a}_1$ and $\boldsymbol{P}_1$ being assumed to be known. As the smoothed state estimator is based on more information than the filtered estimator, its MSE is generally smaller.

It can be shown that the smoothed state vector can be obtained by the following backwards recursion:

$$\hat{\boldsymbol{\xi}}_t = \boldsymbol{a}_t + \boldsymbol{P}_t \boldsymbol{r}_{t-1}, \tag{3.23}$$

with

$$\boldsymbol{r}_{t-1} = \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{v}_t + \boldsymbol{L}_t' \boldsymbol{r}_t, \tag{3.24}$$

---

[5]Alternative derivations have been proposed; see, for example, Kalman (1960) who uses orthogonal projections to derive the filter.

for $t = T, \ldots, 1$, where the $m \times 1$ vector $\boldsymbol{r}_{t-1}$ is constructed as a weighted sum of future innovations. As no innovations are available beyond time T, $\boldsymbol{r}_T$ equals zero. The smoothed state variance matrix $\boldsymbol{V}_t$ can be calculated recursively for $t = T, \ldots, 1$ as

$$\boldsymbol{V}_t = \boldsymbol{P}_t - \boldsymbol{P}_t \boldsymbol{N}_{t-1} \boldsymbol{P}_t, \tag{3.25}$$

with

$$\boldsymbol{N}_{t-1} = \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{Z}_t + \boldsymbol{L}_t' \boldsymbol{N}_t \boldsymbol{L}_t, \tag{3.26}$$

where $\boldsymbol{N}_t$ is the $m \times m$ variance matrix of $\boldsymbol{r}_t$.

Taken together, recursions (3.23) through (3.26), initialized by $\boldsymbol{r}_T = \boldsymbol{0}$ and $\boldsymbol{N}_T = \boldsymbol{0}$, provide an efficient algorithm for state smoothing, denoted as *state smoothing recursion*. It has been introduced to the econometrics literature by de Jong (1988a) and Kohn and Ansley (1989). While smoothing is performed by proceeding backwards through the observations using the state smoothing recursion, filtering is done by moving forwards applying the Kalman filter. During the filtering process $\boldsymbol{v}_t$, $\boldsymbol{F}_t$, $\boldsymbol{K}_t$, $\boldsymbol{a}_t$ and $\boldsymbol{P}_t$ have to be stored for $t = 1, \ldots, T$. Together, the Kalman filter and the state smoothing recursion are called *Kalman filter and smoother*.

3.3.3 Disturbance smoothing

Disturbance smoothing refers to the recursive estimation of the smoothed estimates of the state and observation disturbances, $E(\boldsymbol{\eta}_t | \boldsymbol{y})$ and $E(\boldsymbol{\epsilon}_t | \boldsymbol{y})$, given all the observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$. It represents an alternative and computationally more efficient way to calculate the path of the state vector. The smoothed estimates offer interesting features with respect to parameter estimation and diagnostic checking.

3.3.3.1 Disturbance smoothing recursion

Conditional on $\boldsymbol{y}$, the *smoothed disturbances* can be computed recursively as

$$\hat{\boldsymbol{\epsilon}}_t = E(\boldsymbol{\epsilon}_t | \boldsymbol{y}) = \boldsymbol{H}_t \boldsymbol{u}_t, \tag{3.27}$$
$$\hat{\boldsymbol{\eta}}_t = E(\boldsymbol{\eta}_t | \boldsymbol{y}) = \boldsymbol{Q}_t \boldsymbol{R}_t' \boldsymbol{r}_t, \tag{3.28}$$

for $t = T, \ldots, 1$. The $N \times 1$ *smoothing error* $\boldsymbol{u}_t$ is given by

$$\boldsymbol{u}_t = \boldsymbol{F}_t^{-1} \boldsymbol{v}_t - \boldsymbol{K}_t' \boldsymbol{r}_t, \tag{3.29}$$

where $\boldsymbol{r}_t$, as obtained from (3.24), can be reformulated as

$$\boldsymbol{r}_{t-1} = \boldsymbol{Z}_t' \boldsymbol{u}_t + \boldsymbol{T}_t' \boldsymbol{r}_t. \tag{3.30}$$

The corresponding *smoothed disturbance variance matrices* can be derived as

$$Var(\boldsymbol{\epsilon}_t | \boldsymbol{y}) = \boldsymbol{H}_t - \boldsymbol{H}_t \boldsymbol{D}_t \boldsymbol{H}_t, \tag{3.31}$$
$$Var(\boldsymbol{\eta}_t | \boldsymbol{y}) = \boldsymbol{Q}_t - \boldsymbol{Q}_t \boldsymbol{R}_t' \boldsymbol{N}_t \boldsymbol{R}_t \boldsymbol{Q}_t, \tag{3.32}$$

with the $N \times N$ matrix $\boldsymbol{D}_t$ defined as

$$\boldsymbol{D}_t = \boldsymbol{F}_t^{-1} + \boldsymbol{K}_t' \boldsymbol{N}_t \boldsymbol{K}_t. \tag{3.33}$$

The matrix N_{t-1} can be obtained from the backwards recursion given in (3.26) or alternatively as

$$N_{t-1} = Z_t' D_t Z_t + T_t' N_t T_t - Z_t' K_t' N_t T_t - T_t' N_t K_t Z_t, \qquad (3.34)$$

which is computationally less intense as it relies directly on the sparse system matrices T_t and Z_t. The system of (3.27) through (3.34), for $t = T, \ldots, 1$, initialized with $r_T = \mathbf{0}$ and $N_T = \mathbf{0}$, is denoted as *disturbance smoothing recursion*.

3.3.3.2 Fast state smoothing

The smoothing recursion for the state disturbance term can also be used for *fast state smoothing* as introduced by Koopman (1993). The idea is to calculate $\hat{\xi}_t$ for $t = 1, \ldots, T$, without the necessity to store a_t and P_t. This procedure results in lower computational costs.

Following the outline in Durbin and Koopman (2001, §4.4.2) the necessary recursion can be derived from the state equation in (3.1) which implies

$$\hat{\xi}_{t+1} = T_t \hat{\xi}_t + R_t \hat{\eta}_t. \qquad (3.35)$$

Substituting (3.28) into (3.35) yields

$$\hat{\xi}_{t+1} = T_t \hat{\xi}_t + R_t Q_t R_t' r_t, \qquad t = 1, \ldots, T, \qquad (3.36)$$

which is initialized for $t = 1$ by (3.23) together with (3.24). In contrast to the state smoothing recursion presented in §3.3.2, this recursion allows for the computation of smoothed states $\hat{\xi}_t$ without getting a_t and P_t involved.

The process of disturbance smoothing is comparable to that of state smoothing: while the Kalman filter is employed to proceed forwards, backwards proceeding through the data is done using the disturbance smoother. Due to its computational advantage, the disturbance smoother will be used whenever the only objective is to form an inference on the state vector and V_t is not required.

3.3.4 Missing observations

Within the state space framework, missing observations can be easily dealt with. Durbin and Koopman (2001, §4.8) show that for a missing set of observations, denoted as $y_\tau, \ldots, y_{\tau^*-1}$, the original filter and smoothing recursions can be used for all t. The one-step ahead forecast error v_t and the Kalman gain matrix K_t are simply set to zero for all missing data points. With $v_t = \mathbf{0}$ and $K_t = \mathbf{0}$ the filter recursions in (3.13) and (3.14) become

$$a_{t+1} = T_t a_t, \qquad (3.37)$$

$$P_{t+1} = T_t P_t T_t' + R_t Q_t R_t', \qquad t = \tau, \ldots, \tau^* - 1. \qquad (3.38)$$

The smoothing recursions in (3.24) and (3.26) can be written as

$$r_{t-1} = T_t' r_t, \qquad (3.39)$$

$$N_{t-1} = T_t' N_t T_t, \qquad t = \tau^* - 1, \ldots, \tau, \qquad (3.40)$$

while the other relevant smoothing steps remain unaffected.

3.3.5 Forecasting

The discussion related to missing observations serves as a basis to use the Kalman filter for forecasting purposes. The objective is to generate forecasts for $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$, denoted as $\boldsymbol{y}_{T+l}$, for *lead time* $l = 1, \ldots, L$. Let $\bar{\boldsymbol{y}}_{T+l}$ be the minimum MSE forecast of $\boldsymbol{y}_{T+l}$ given $\boldsymbol{y}$, i.e. $\bar{\boldsymbol{y}}_{T+l}$ has a minimum MSE matrix

$$\bar{\boldsymbol{F}}_{T+l} = E((\bar{\boldsymbol{y}}_{T+l} - \boldsymbol{y}_{T+l})(\bar{\boldsymbol{y}}_{T+l} - \boldsymbol{y}_{T+l})'|\boldsymbol{y}), \tag{3.41}$$

for all $\boldsymbol{y}_{T+l}$. Given the standard result that $E((X - \lambda)(X - \lambda)')$ is minimized for $\lambda = \mu$, where μ is the mean of a random variable X, it follows immediately that the conditional mean of $\boldsymbol{y}_{T+l}$ given $\boldsymbol{y}$ represents the minimum MSE forecast:

$$\bar{\boldsymbol{y}}_{T+l} = E(\boldsymbol{y}_{T+l}|\boldsymbol{y}). \tag{3.42}$$

From observation equation (3.2) we have $\boldsymbol{y}_{T+l} = \boldsymbol{Z}_{T+l}\boldsymbol{\xi}_{T+l} + \boldsymbol{\epsilon}_{T+l}$. Together with (3.42) and

$$\bar{\boldsymbol{a}}_{T+l} = E(\boldsymbol{\xi}_{T+l}|\boldsymbol{y}), \tag{3.43}$$
$$\bar{\boldsymbol{P}}_{T+l} = E((\bar{\boldsymbol{a}}_{T+l} - \boldsymbol{\xi}_{T+l})(\bar{\boldsymbol{a}}_{T+l} - \boldsymbol{\xi}_{T+l})'|\boldsymbol{y}), \tag{3.44}$$

the MSE forecast is given by

$$\bar{\boldsymbol{y}}_{T+l} = \boldsymbol{Z}_{T+l}E(\boldsymbol{\alpha}_{T+l}|\boldsymbol{y}) = \boldsymbol{Z}_{T+l}\bar{\boldsymbol{a}}_{T+l}, \tag{3.45}$$

with the MSE variance matrix

$$\bar{\boldsymbol{F}}_{T+l} = \boldsymbol{Z}_{T+l}\bar{\boldsymbol{P}}_{T+l}\boldsymbol{Z}'_{T+l} + \boldsymbol{H}_{T+l}. \tag{3.46}$$

Recursions for computing $\bar{\boldsymbol{a}}_{T+l}$ and $\bar{\boldsymbol{P}}_{T+l}$ for $l = 1, \ldots, L-1$ can be derived as

$$\bar{\boldsymbol{a}}_{T+l+1} = \boldsymbol{T}_{T+l}\bar{\boldsymbol{a}}_{T+l}, \tag{3.47}$$
$$\bar{\boldsymbol{P}}_{T+l+1} = \boldsymbol{T}_{T+l}\bar{\boldsymbol{P}}_{T+l}\boldsymbol{T}'_{T+l} + \boldsymbol{R}_{T+l}\boldsymbol{Q}_{T+l}\boldsymbol{R}'_{T+l}, \tag{3.48}$$

which are identical to the Kalman filter recursions for $\boldsymbol{a}_{T+l}$ and $\boldsymbol{P}_{T+l}$ in (3.13) and (3.14) with $\boldsymbol{v}_{T+l} = \boldsymbol{K}_{T+l} = \boldsymbol{0}$. As these are the same conditions that allow the Kalman filter to deal with missing observations in §3.3.4, an l-period-ahead forecast, denoted as $\bar{\boldsymbol{y}}_{T+l}$, can be calculated routinely by interpreting $\boldsymbol{y}_{T+1}, \ldots, \boldsymbol{y}_{T+L}$ as missing observations and by employing the Kalman filter beyond time $t = T$ with $\boldsymbol{v}_t$ and $\boldsymbol{K}_t$ set to zero (cf. Durbin and Koopman 2001, §4.9).

3.3.6 Initialization of filter and smoother

So far the distribution of the initial state vector $\boldsymbol{\xi}_1$ has been assumed to be $N(\boldsymbol{a}_1, \boldsymbol{P}_1)$ with known mean and covariance. In practice, however, at least some elements of the distribution of $\boldsymbol{\xi}_1$ are unknown. In the literature, three ways of initializing non-stationary state space models are discussed.

The first is to employ a *diffuse prior* that fixes $\boldsymbol{a}_1$ at an arbitrary value and allows the diagonal elements of $\boldsymbol{P}_1$ to go to infinity. Following Durbin and Koopman (2001, §5.1) a general specification for the initial state vector is given by

$$\boldsymbol{\xi}_1 = \boldsymbol{a} + \boldsymbol{A}\boldsymbol{\delta} + \boldsymbol{R}_0\boldsymbol{\eta}_0, \qquad \boldsymbol{\delta} \sim N(\boldsymbol{0}, \kappa \boldsymbol{I}_q), \qquad \boldsymbol{\eta}_0 \sim N(\boldsymbol{0}, \boldsymbol{Q}_0), \qquad (3.49)$$

where the $m \times 1$ vector $\boldsymbol{a}$ can be treated as a zero vector whenever none of the elements of of $\boldsymbol{\xi}_1$ are known constants. The $m \times q$ matrix $\boldsymbol{A}$ and the $m \times (m - q)$ matrix $\boldsymbol{R}_0$ are fixed and known selection matrices with $\boldsymbol{A}'\boldsymbol{R}_0 = \boldsymbol{0}$. The initial covariance matrix $\boldsymbol{Q}_0$ is assumed to be known and positive definite. The $q \times 1$ vector $\boldsymbol{\delta}$ is treated as a random variable with infinite variance and is called the *diffuse* vector as $\kappa \to \infty$. This leads to

$$\boldsymbol{P}_1 = \boldsymbol{P}_* + \kappa \boldsymbol{P}_\infty, \qquad (3.50)$$

with $\boldsymbol{P}_* = \boldsymbol{R}_0\boldsymbol{Q}_0\boldsymbol{R}_0'$ and $\boldsymbol{P}_\infty = \boldsymbol{A}\boldsymbol{A}'$. With some elements of $\boldsymbol{\xi}_1$ being diffuse, the initialization of the Kalman filter is referred to as *diffuse initialization*. However, in cases where $\boldsymbol{P}_\infty$ is a nonzero matrix, the standard Kalman filter cannot be employed as no real value can represent κ as $\kappa \to \infty$. It is necessary to find an approximation or to modify the Kalman filter in an appropriate way. The technique that will be used throughout this thesis is based on Harvey and Phillips (1979). The two authors propose to replace κ by a large but finite numerical value, which enables the use of the standard Kalman filter. They showed that this will yield a good approximation, where the remaining rounding errors are small.

As an alternative to the *large κ approximation*, exact initialization techniques have been developed for $\kappa \to \infty$. However, *exact treatments* turn out to be difficult to implement and cannot deal with the general multivariate linear Gaussian state space model directly. They will not find any consideration hereafter. For details on alternative exact treatments of the initial Kalman filter, see, for example, Ansley and Kohn (1985), de Jong (1988b, 1991) and Koopman (1997).

The assumption of an infinite variance might be regarded unnatural as all observed time series have a finite variance. This leads to the third way of initialization: Rosenberg (1973) considers $\boldsymbol{\xi}_1$ to be an unknown constant that can be estimated from the first observation $\boldsymbol{y}_1$ by maximum likelihood. It can be shown that by this procedure, the same initialization of the filter is obtained as by assuming that $\boldsymbol{\xi}_1$ is a random variable with infinite variance (cf. Durbin and Koopman 2001, §2.9).

In the following, unless genuine prior information on $\boldsymbol{a}_1$ and $\boldsymbol{P}_1$ is available, a diffuse prior with $\boldsymbol{a} = \boldsymbol{0}$, $\boldsymbol{P}_* = \boldsymbol{0}$ and $\boldsymbol{P}_\infty = \boldsymbol{I}$ will be used such that $\boldsymbol{\xi}_1 \sim N(0, \kappa\boldsymbol{I})$. Following the recommendation of Koopman et al. (1999), κ is first set to 10^6 and then multiplied by the maximum diagonal value of the residual covariances to adjust for scale:

$$\kappa = 10^6 \times \max\left\{1, \begin{bmatrix} \boldsymbol{Q}_t & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_t \end{bmatrix}\right\}. \qquad (3.51)$$

It should be noted that the initialization is not trivial for the multivariate case. It is possible that the part of $\boldsymbol{F}_t^{-1}$ that is linked to $\boldsymbol{P}_\infty$ is sometimes singular with different ranks. In these cases $\boldsymbol{F}_t^{-1}$ cannot simply be expanded in powers of κ^{-1} for the first few

elements of a multivariate series. As it is straightforward to deal with this problem for univariate series, Durbin and Koopman (2001, §5.1) propose to bring in the components of a multivariate series into the analysis one at a time. Thus, a multivariate series is converted into a univariate series. This leads to computational gains and significantly simplifies the process of diffuse initialization for multivariate series. For a detailed discussion of the univariate treatment of multivariate series, the reader is referred to Durbin and Koopman (2001, §6.4).

3.3.7 The Kalman filter with non-Gaussian errors

The derivation of the Kalman filter presented above and the related estimation procedures are based on the assumption of normally distributed disturbances. It was shown how the conditional distribution of the state vector $\boldsymbol{\xi}_t$ can be calculated recursively given the information set at time t, for all $t = 1, \ldots, T$. As these conditional distributions are also normal, they are fully specified by their first two moments. These can be computed by the Kalman filter. The conditional mean of the state vector represents an optimal estimator in the sense that is has minimum MSE matrix.

In case of non-Gaussian disturbances, the Kalman filter is no longer guaranteed to yield the conditional mean of the state vector. However, it nevertheless represents an optimal estimator in the sense that no other linear estimator has a smaller MSE. Therefore, the Kalman filter can still be employed when the normality assumption is dropped. The estimators obtained by maximizing the Gaussian likelihood function with observations that are not normally distributed are referred to as *quasi-maximum likelihood* (QML) estimators; cf. Hamilton (1994a).[6]

3.4 Maximum likelihood estimation

So far, the system matrices have been assumed to be known. In the more general case, they depend at least partially on $\boldsymbol{\psi}$, the vector of unknown parameters. This section explains how the vector of unknowns can be estimated by means of maximum likelihood (ML). After briefly summarizing the general idea behind the concept of ML, Subsection 3.4.1 introduces the loglikelihood function for the general Gaussian state space model. It will be demonstrated how the general model can be reparameterized in order to reduce the dimension of the vector of parameters by one. Subsections 3.4.2 and 3.4.3 briefly overview the two alternative concepts of maximizing the loglikelihood: direct numerical maximization and the EM algorithm. Finally, Subsection 3.4.4 shows how parameter restrictions can be implemented.

3.4.1 The loglikelihood function

In order to estimate a model by ML, it has to be parametric and fully specified through the joint probability density function; the parameter values have to contain all the necessary information for a simulation of the dependent variables. For the T sets of

---

[6]For an introduction to state space models that take non-Gaussianity explicitly into account, which is beyond the scope of this thesis, see, for example, Durbin and Koopman (2001, Part II).

observations, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$, which are assumed to be IID, the joint density is given by the product of the individual densities, denoted by $p(\cdot)$:

$$L(\boldsymbol{y}, \boldsymbol{\psi}) = p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T) = \prod_{t=1}^{T} p(\boldsymbol{y}_t), \tag{3.52}$$

where $L(\boldsymbol{y}, \boldsymbol{\psi})$ is the joint probability density function of the t-th set of observations. When the joint density is evaluated at a given data set, $L(\boldsymbol{y}, \boldsymbol{\psi})$ is referred to as a likelihood function of the model. To avoid computational difficulties caused by extreme numbers that may result from the product above, it is generally simpler to work with the natural logarithm of the likelihood function:

$$\log L(\boldsymbol{y}, \boldsymbol{\psi}) = \sum_{t=1}^{T} \log p(\boldsymbol{y}_t). \tag{3.53}$$

The likelihood function and its logarithm are often simply denoted as $L(\boldsymbol{y})$ and $\log L(\boldsymbol{y})$, respectively. If the vector of parameters is identifiable, the ML estimate of the parameters $(\hat{\boldsymbol{\psi}})$ is found by maximizing the likelihood with respect to $\boldsymbol{\psi}$.[7] For a general introduction into the methodology of ML, see, for example, Greene (2003, §17) or Davidson and MacKinnon (2004, §10).

3.4.1.1   Prediction error decomposition

As the observations for time series models are not generally independent, Equation (3.53) is replaced by a probability density function. The distribution of $\boldsymbol{y}_t$ is conditioned on $\boldsymbol{Y}_{t-1}$, the information set at time $t-1$:

$$\log L(\boldsymbol{y}) = \sum_{t=1}^{T} \log p(\boldsymbol{y}_t | \boldsymbol{Y}_{t-1}), \tag{3.54}$$

with $\boldsymbol{Y}_t = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$ and $p(\boldsymbol{y}_1 | \boldsymbol{Y}_0) := p(\boldsymbol{y}_1)$.

If the observation disturbances and the initial state vector in the general state space model (3.1)–(3.7) have a multivariate normal distribution, it can be shown that the conditional distribution of $\boldsymbol{y}_t$ itself is normal with conditional mean

$$E(\boldsymbol{y}_t | \boldsymbol{Y}_{t-1}) = \boldsymbol{Z}_t \boldsymbol{a}_t, \tag{3.55}$$

and conditional covariance

$$Var(\boldsymbol{y}_t | \boldsymbol{Y}_{t-1}) = \boldsymbol{F}_t. \tag{3.56}$$

The variance of the one-step ahead forecast error, $\boldsymbol{F}_t$, is defined as in (3.16). For Gaussian models, $\boldsymbol{y}_t$ is conditionally distributed as $N(\boldsymbol{Z}_t \boldsymbol{a}_t, \boldsymbol{F}_t)$ with conditional probability density function

$$p(\boldsymbol{y}_t | \boldsymbol{Y}_{t-1}) = \frac{1}{2\pi} |\boldsymbol{F}_t|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{v}_t' \boldsymbol{F}_t^{-1} \boldsymbol{v}_t\right). \tag{3.57}$$

---

[7] Identifiability means that the estimation yields a unique determination of the parameter estimates for a given set of data. For more details on the identifiability of structural time series models, see Harvey (1989, §4.4).

After substitution of (3.57) for the conditional density in (3.54), the loglikelihood function can be expressed as

$$\log L(\boldsymbol{y}) = -\frac{TN}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|\boldsymbol{F}_t| - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{v}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t, \qquad (3.58)$$

with $\boldsymbol{v}_t$ being defined as in (3.15). This representation of the loglikelihood traces back to Schweppe (1965). As $\boldsymbol{F}_t$ and $\boldsymbol{v}_t$ are part of the standard Kalman filter output for given values of $\boldsymbol{\psi}$, the loglikelihood can be obtained in routine fashion from the Kalman filter. As the conditional mean of $\boldsymbol{y}_t$ is equal to the minimum MSE of $\boldsymbol{y}_t$ (3.45), $\boldsymbol{v}_t$ can be interpreted as a vector of prediction errors. The loglikelihood (3.58) is referred to as *prediction error decomposition* (p.e.d); cf. Harvey (1989, §3.4).

When the Kalman filter is initialized using a diffuse prior with d diffuse elements in the state vector, the first d innovations and their corresponding variances are excluded from the prediction error decomposition. With summations starting at $t = d+1$ instead of $t = 1$, the joint density of $\boldsymbol{y}_{d+1}, \ldots, \boldsymbol{y}_T$, conditional on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_d$, is given by

$$\log L(\boldsymbol{y}) = -\frac{(T-d)N}{2}\log(2\pi) - \frac{1}{2}\sum_{t=d+1}^{T}\log|\boldsymbol{F}_t| - \frac{1}{2}\sum_{t=d+1}^{T}\boldsymbol{v}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t. \qquad (3.59)$$

Numerical maximization of the loglikelihood with respect to $\boldsymbol{\psi}$ yields the ML estimate for the vector of unknown parameters.

3.4.1.2 Concentrated loglikelihood

Maximum likelihood can be improved by concentrating one parameter out of the loglikelihood function. In the univariate case it is usually possible to reparameterize a model such that it depends on $\boldsymbol{\psi} = [\boldsymbol{\psi}_*' \quad \sigma_*^2]'$, where σ_*^2 is a positive scalar to which the variances of the disturbance terms are proportional. The vector $\boldsymbol{\psi}_*$ contains one parameter less than $\boldsymbol{\psi}$. With h_t and $\boldsymbol{Q}_t$ depending only on $\boldsymbol{\psi}_*$ and not on σ_*^2, the disturbance variances can be written as $Var(\epsilon_t) = \sigma_*^2 h_t$ and $Var(\boldsymbol{\eta}_t) = \sigma_*^2 \boldsymbol{Q}_t$. By setting h_t or one of the diagonal elements of $\boldsymbol{Q}_t$ equal to unity, σ_*^2 becomes equal to the variance of the corresponding disturbance term. In accordance with Harvey (1989, §3.4) the introduction of σ_*^2 gives the following prediction error decomposition for $N = 1$:

$$\log L(y) = -\frac{T-d}{2}\log(2\pi) - \frac{T-d}{2}\log\sigma_*^2 - \frac{1}{2}\sum_{t=d+1}^{T}\log f_t - \frac{1}{2\sigma_*^2}\sum_{t=d+1}^{T}\frac{v_t^2}{f_t}. \qquad (3.60)$$

For univariate series, the scalar $f_t = \boldsymbol{z}_t \boldsymbol{P}_t \boldsymbol{z}_t' + h_t$ replaces the $N \times N$ matrix $\boldsymbol{F}_t$ of (3.16). Differentiating (3.60) with respect to σ_*^2 and solving the first-order condition as a function of the data and the remaining parameters, yields the ML estimator of σ_*^2 for given values of $\boldsymbol{\psi}_*$:

$$\tilde{\sigma}_*^2(\boldsymbol{\psi}_*) = \frac{1}{T-d}\sum_{t=d+1}^{T}\frac{v_t^2}{f_t}. \qquad (3.61)$$

Substituting $\tilde{\sigma}_*^2$ back into (3.60) gives the *concentrated* or *profile loglikelihood* function that has to be maximized with respect to $\boldsymbol{\psi}_*$:

$$\log L_c(y) = -\frac{T-d}{2}\log(2\pi) - \frac{1}{2}\sum_{t=d+1}^{T}\log f_t - \frac{T-d}{2}(\log\tilde{\sigma}_*^2(\boldsymbol{\psi}_*)+1). \qquad (3.62)$$

When (3.62) is maximized instead of (3.59), the dimension of the vector of parameters to be estimated is reduced by one. In addition to the obtained gains in computational efficiency, the results are likely to be more reliable. As no exact gradients are available for the concentrated loglikelihood, it has to be maximized numerically.

If the Kalman filter is initialized by employing the large κ approximation, rounding errors can lead to numerical problems. In this thesis, a feasible possibility to overcome this problem in the univariate case is to calculate starting values from the first observations, as explained by Harvey (1989, §3.3.4). For a more general algorithm, see Ansley and Kohn (1985) who propose an analytical but complex and difficult to implement solution to the exact initialization problem. An alternative exact approach is discussed by de Jong (1991) who uses an augmentation technique to the exact Kalman filter. In a more recent work, Koopman (1997) proposes an exact analytical approach based on a trivial initialization and develops a diffuse loglikelihood, $\log L_d(\boldsymbol{y}|\boldsymbol{\psi})$. However, as it can be shown that the estimate $\hat{\boldsymbol{\psi}}$, obtained by maximizing $\log L(\boldsymbol{y}|\boldsymbol{\psi})$ for fixed κ, converges to the estimate obtained by maximizing the diffuse loglikelihood as $\kappa \to \infty$ (cf. Durbin and Koopman 2001, §7.3.1), the approach taken here can be considered a valid procedure for univariate models.

3.4.2 Numerical maximization

Given the sample observations, the loglikelihood can be maximized by means of direct numerical maximization. The basic idea behind this method is to find the value $\hat{\boldsymbol{\psi}}$ for which the loglikelihood is maximized. An algorithm is used to make different guesses for $\boldsymbol{\psi}$ and to compare the corresponding numerical values of the loglikelihood. To compute the ML estimates, the algorithm performs a series of steps, each time starting with a first guess for the unknown parameters. The algorithm then chooses the direction where to search, determines how far to move in the chosen direction, and computes and compares the value of the loglikelihood for the chosen values of $\boldsymbol{\psi}$. If $\boldsymbol{\psi}$ leads sufficiently close to a maximum of the loglikelihood, the algorithm stops, otherwise the search continues. Generally, numerical maximization methods differ with respect to the direction to search, the step size and the stopping rule (cf. Davidson and MacKinnon 2004, §6.4).

Many numerical maximization techniques are based on *Newton's method*: for a given starting value for $\boldsymbol{\psi}$, the direction of search is determined by the gradient or score vector, denoted as $\boldsymbol{g}(\boldsymbol{\psi})$; the step size is calculated from the Hessian matrix, denoted as $\boldsymbol{H}(\boldsymbol{\psi})$, which has a unique maximum only if it is negative definite for all $\boldsymbol{\psi}$. For a more detailed description of the different available algorithms, see, for example, Hamilton (1994b, §5.7) or Judge et al. (1985, §B.2). In practical applications, it is often impossible or computationally expensive to calculate the gradient and the Hessian analytically. However, it is usually feasible to compute $\boldsymbol{g}(\boldsymbol{\psi})$ numerically. For details on the calculation of the

score vector with respect to problems formulated in state space, the reader is referred to Durbin and Koopman (2001, §7.3.3). In order to avoid numerical or analytical computation, the Hessian can usually be approximated. In Ox, $\boldsymbol{H}(\boldsymbol{\psi})$ is approximated using the quasi-Newton method according to Broyden-Fletcher-Goldfarb-Shanno (BFGS), which ensures negative definiteness of the Hessian; details on this nonlinear optimization techniques are discussed by Greene (2003, §E.6.2).

3.4.3   The EM algorithm

The unknown elements of the system matrices can alternatively be estimated via the *expectation-maximization* (EM) algorithm by Dempster et al. (1977). The EM algorithm, named after its two steps of maximizing the expectation of the loglikelihood, is an iterative algorithm for maximum likelihood estimation. It has originally been developed for dealing with missing observations and can be employed for maximizing the loglikelihood of many state space models (cf. Shumway and Stoffer 1982; Watson and Engle 1983).

As in this thesis the likelihood will be generally maximized using numerical maximization procedures, both in the context of Kalman filter based state space models as well as the basic hidden Markov model to be introduced below, the reader is referred to the well-established literature for details on the EM algorithm. For an introductory outline, see, for example, Bulla (2006, Appendix A) and the references provided therein. A general demonstration of how to use the EM algorithm to compute the ML estimates in some common statistical applications is provided by McLachlan and Krishnan (1997); a Bayesian treatment is given by Tanner (1993).

3.4.4   Parameter restrictions

Sometimes the parameters of a model are not allowed to take any value in $\mathbb{R}$, the set of real numbers. In this case, it may become necessary to impose parameter constraints. For example, the elements of the covariance matrices $\boldsymbol{Q}_t$ and $\boldsymbol{H}_t$ in (3.1) and (3.2) are restricted to positive values. While it is theoretically possible to introduce such constraints into the numerical maximization procedure directly, it is not practically feasible. In order to employ Newton-type maximization routines implemented in standard statistical software packages, it is recommendable to perform any maximization with respect to unconstrained quantities. This can be achieved by transforming the parameters in an appropriate way. In case of a positive variance term σ^2, the constraint will be imposed by defining

$$\psi_\sigma = \log \sigma^2, \qquad -\infty < \psi_\sigma < \infty. \tag{3.63}$$

Once the loglikelihood is maximized using the transformed but unconstrained parameter ψ_σ, the constrained parameter can be calculated by back-transformation:

$$\hat{\sigma}^2 = \exp(\hat{\psi}_\sigma), \qquad \hat{\sigma}^2 > 0. \tag{3.64}$$

Similarly, whenever a parameter θ should be restricted to the range $[-a; a]$ with $a > 0$, it is possible to use the transformation

$$\psi_\theta = \frac{\theta}{\sqrt{a^2 - \theta^2}}, \qquad -\infty < \psi_\theta < \infty, \tag{3.65}$$

where the restricted parameter can be obtained as

$$\hat{\theta} = \frac{a\hat{\psi}_\theta}{\sqrt{1 + \hat{\psi}_\theta^2}}, \qquad -a < \hat{\theta} < a. \tag{3.66}$$

In case where a parameter θ is only allowed to take on values between zero and unity, the transformed parameter

$$\psi_\theta = \sqrt{\frac{\theta}{1 - \theta}}, \qquad -\infty < \psi_\theta < \infty, \tag{3.67}$$

will be used with θ being computed as

$$\hat{\theta} = \frac{\hat{\psi}_\theta^2}{1 + \hat{\psi}_\theta^2}, \qquad 0 \leq \hat{\theta} < 1. \tag{3.68}$$

With these transformations, the ML estimates are obtained in two steps. At first, the likelihood is maximized with respect to the unconstrained but transformed parameters. In a second step, the estimates for the transformed parameters are converted back to the constrained original parameters.

3.5   Introduction of explanatory variables

The general state space model can be extended to include explanatory variables that are able to partially explain the movements of the vector of observations. Subsection 3.5.1 introduces regression effects to the state space framework. Subsection 3.5.2 discusses the estimation of regression models with time-varying parameters and develops various dynamic processes for changing regression coefficients. The section closes with a brief discussion of how to set the initial values needed to initialize the Kalman filter in Subsection 3.5.3.

3.5.1   Incorporation of regression effects

A state space model can be allowed for regression effects by setting $\boldsymbol{d}_t$ in the observation equation (3.2) equal to the linear combination $\boldsymbol{x}_t\boldsymbol{\beta}_t$. Given an expanded set of information with k observable variables, which are assumed to be weakly exogenous[8] and to be linearly related to the univariate vector of observations y_t, the observation equation can be stated as

$$y_t = \boldsymbol{z}_t\boldsymbol{\xi}_t + \boldsymbol{x}_t\boldsymbol{\beta}_t + \epsilon_t, \qquad t = 1, \ldots, T, \tag{3.69}$$

---

[8]According to the definition provided by Harvey (1989, §7.1.4) explanatory variables are weakly exogenous if they can be dealt with as though they do not change in repeated samples. A formal treatment of the concept of exogeneity can be found in Engle et al. (1983).

where $\boldsymbol{x}_t$ is a $1 \times K$ vector of explanatory variables, which possibly contains unity as first element. The $k \times 1$ vector $\boldsymbol{\beta}_t$ contains the unknown regression coefficients. Weak exogeneity assures that it can be conditioned on $\boldsymbol{x}_t$ without losing any relevant information needed to efficiently estimate the unknown model parameters (cf. Harvey 1989, §7.1).

In case where $\boldsymbol{\beta}_t$ is known, the inclusion of explanatory variables does not affect the operation of the Kalman filter. In practical applications, however, the regression coefficients are usually unknown. In this case regression effects can be incorporated by including $\boldsymbol{\beta}_t$ in the state vector. This yields a state space model with an extended $(m+k) \times 1$ state vector $\boldsymbol{\xi}_t^\dagger = [\ \boldsymbol{\xi}_t'\ \ \boldsymbol{\beta}_t'\ ]'$ whose components can be estimated simultaneously for $t = 1, \ldots, T$:

$$\boldsymbol{\xi}_{t+1}^\dagger = \boldsymbol{T}_t \boldsymbol{\xi}_t^\dagger + \boldsymbol{R}_t \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t^\dagger), \tag{3.70}$$

$$y_t = \left[\ \boldsymbol{z}_t\ \ \boldsymbol{x}_t\ \right] \boldsymbol{\xi}_t^\dagger + \epsilon_t, \qquad \epsilon_t \sim N(0, h_t), \tag{3.71}$$

where $\boldsymbol{T}_t$, $\boldsymbol{R}_t$ and $\boldsymbol{\eta}_t$ are of dimensions $(m+k) \times (m+k)$, $(m+k) \times (r+k)$ and $(r+k) \times 1$, respectively. The $(r+k) \times (r+k)$ covariance matrix of the state errors is defined as

$$\boldsymbol{Q}_t^\dagger = \left[\begin{array}{cc} \boldsymbol{Q}_t & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_t^\beta \end{array}\right]. \tag{3.72}$$

The structure of $\boldsymbol{Q}_t^\beta$, a block matrix of dimension $k \times k$, determines the nature of the regression coefficients. For fixed coefficients, $\boldsymbol{Q}_t^\beta$ has to be set to a $k \times k$ zero matrix. For $\boldsymbol{Q}_t^\beta \neq \boldsymbol{0}$, the regression coefficients can be made time-varying. Details on the case with stochastic coefficients are provided in §3.5.2.

With $\boldsymbol{\beta}_1$ being treated as diffuse, the standard Kalman filter and smoother, as described in §3.3, can be employed to this extended state space model without further modification (cf. Durbin and Koopman 2001, §6.2). The regression coefficients can be estimated in the time-domain by employing the ML procedures introduced in §3.4. The Kalman filter can also be applied to a regression model with fixed coefficients that is formulated in state space form: in the absence of structural components, this yields the same results as can be obtained by recursive least squares estimation. This property can be used in connection with specification tests for structural breaks. The estimation yields two different types of residuals, which will be employed in §3.6 for specification testing and model diagnostic purposes. For details on the estimation of structural time series model with explanatory variables, see Harvey (1989, §7.3.2).

3.5.2   Time-varying parameter models

The coefficients of the explanatory variables can be made time-varying by incorporating them directly into the state vector, $\boldsymbol{\xi}_t$, where the variance of the corresponding state errors has to be nonzero. Since a regression with time-varying coefficients is a special case of the general state space introduced in §3.2, it can be handled routinely by the standard Kalman filter and smoother. The focus in this section is on the modeling of time-varying regression coefficients as both non-stationary and stationary processes.

Following Chow (1984) a general expression for the basic univariate time-varying coefficient regression model with homoskedastic error terms can be formulated as

$$y_t = \boldsymbol{z}_t \boldsymbol{\xi}_t + \epsilon_t, \qquad \epsilon_t \sim N(0, h), \tag{3.73}$$

where $\boldsymbol{z}_t$ denotes a $1 \times m$ row vector with $m = k$ fixed explanatory variables, of which unity might be the first element. The column vector $\boldsymbol{\xi}_t$ contains the regression coefficients whose behavior over time is represented by

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{T}\boldsymbol{\xi}_t + \boldsymbol{R}\boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(0, \boldsymbol{Q}), \tag{3.74}$$

where the system matrices $\boldsymbol{T}$ and $\boldsymbol{R}$ are assumed to be time-homogeneous. To allow for a meaningful economic interpretation, the state equation given in (3.74) can be rearranged by introducing the mean state vector $\bar{\boldsymbol{\xi}}$ and by replacing $\boldsymbol{\xi}_t$ by $\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}$. This leads to a representation in which $\bar{\boldsymbol{\xi}}$ can be interpreted as the mean coefficient over the entire sample. The matrix $\boldsymbol{T}$ is referred to as the *speed parameter*, which measures how fast the time-varying state vector returns to its mean:

$$\boldsymbol{\xi}_{t+1} - \bar{\boldsymbol{\xi}} = \boldsymbol{T}(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}) + \boldsymbol{R}\boldsymbol{\eta}_t. \tag{3.75}$$

The literature has not arrived yet at a consensus on how to introduce time-variation into the coefficients of explanatory variables. In the following, it will be demonstrated how different processes for the time-varying state vector can be derived depending on the chosen value for $\boldsymbol{T}$.

3.5.2.1   The random coefficient model

Setting $\boldsymbol{T}$ to an $m \times m$ zero matrix yields a model that describes the time path of a changing $\boldsymbol{\xi}_t$ as *random coefficients* (RC):

$$\boldsymbol{\xi}_{t+1} = \bar{\boldsymbol{\xi}} + \boldsymbol{R}\boldsymbol{\eta}_t. \tag{3.76}$$

The random coefficient model, originally introduced by Hildreth and Houck (1968) in a cross-sectional context, implies a long-run average coefficient around which the current estimate fluctuates randomly. The parameters to be estimated are $\bar{\boldsymbol{\xi}}$ and the variances of the error terms. As it is not possible to distinguish between a randomly behaving intercept and the observation disturbances, any intercept term has to be included as being fixed in the observation equation. This reduces the state vector's dimension by one (cf. Wells 1996, §5.3).

Overall, the practical relevance of this model is limited: as the stochastic properties of the underlying process are only reflected in the state disturbances, the average state always represents the best forecast. For a review of the random coefficient model, see, for example, Nicholls and Pagan (1985).

3.5.2.2   The random walk model

By setting the transition matrix to an m-dimensional identity matrix, one can derive an important modeling class according to which the behavior of the changing regression

coefficients over time can be described as a *random walk* (RW):

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{\xi}_t + \boldsymbol{R}\boldsymbol{\eta}_t. \tag{3.77}$$

The random walk is considered an essential scientific concept in the analysis of dynamics across many disciplines. It had already been widely established in the natural sciences when Samuelson (1965) introduced it to finance. Today, the random walk represents one of the most important ideas in financial economics.

An illustrative example of the applicability of random walks to economics is the time-varying parameter model for U.S. monetary growth presented by Kim and Nelson (1989). Excellent surveys of the random walk and its link to financial markets and security prices include Campbell et al. (1997, §2) and Lo and MacKinlay (1999).

3.5.2.3 The mean reverting model

Another important time-varying coefficient model is the *mean reverting* (MR) model, proposed by Rosenberg (1973). The regression coefficients are modeled as a stationary process:

$$\boldsymbol{\xi}_{t+1} - \bar{\boldsymbol{\xi}} = \boldsymbol{T}(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}) + \boldsymbol{R}\boldsymbol{\eta}_t, \tag{3.78}$$

where $\boldsymbol{\xi}_t$ is a vector of stochastic parameters with non-stochastic mean $\bar{\boldsymbol{\xi}}$ in stochastic equilibrium. For values of the roots of $\boldsymbol{T}$ inside the unit circle, stationarity of the stochastic process is guaranteed and the coefficients move around the constant mean $\bar{\boldsymbol{\xi}}$. The parameters to be estimated are $\bar{\boldsymbol{\xi}}$, $\boldsymbol{T}$, $\boldsymbol{Q}$ and h. Following Collins et al. (1987) it is assumed a priori that the major determinants of any time-varying regression coefficients change only gradually over time. To assure smoothly developing coefficients from period to period, the diagonal elements of $\boldsymbol{T}$ are restricted to values between zero and unity. For $\boldsymbol{T} = \boldsymbol{I}$, Equation (3.78) reduces to a random walk. For $\boldsymbol{T} = \boldsymbol{0}$, the model is equivalent to the random coefficient model.

The state equation of the mean reverting model differs from the general state equation in (3.74) by the incorporation of the parameter $\bar{\boldsymbol{\xi}}$. Substitution of (3.78) into (3.73) brings the mean reverting model into the same form as (3.70)–(3.71):

$$\left[\begin{array}{c} \boldsymbol{\xi}^*_{t+1} \\ \bar{\boldsymbol{\xi}} \end{array} \right] = \boldsymbol{T} \left[\begin{array}{c} \boldsymbol{\xi}^*_t \\ \bar{\boldsymbol{\xi}} \end{array} \right] + \boldsymbol{R}\boldsymbol{\eta}_t, \tag{3.79}$$

$$y_t = \boldsymbol{z}_t \left[\begin{array}{c} \boldsymbol{\xi}^*_t \\ \bar{\boldsymbol{\xi}} \end{array} \right] + \epsilon_t, \tag{3.80}$$

with $\boldsymbol{\xi}^*_t = \boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}$ for all t (cf. Chow 1984). The diagonal elements of $\boldsymbol{T}$ that correspond to $\bar{\boldsymbol{\xi}}$ are equal to unity. As $\bar{\boldsymbol{\xi}}$ is constant over time, the lower part of $\boldsymbol{\eta}_t$ is set to zero. In this representation, the state vector contains both time-varying as well as fixed regression coefficients and can be estimated by the standard Kalman filter.

In the literature, the mean reverting model has been applied to the testing and modeling of the time-variation of an asset's systematic risk; see, for example, Bos and Newbold (1984), Wells (1996, §5.3) and Brooks et al. (1998).

3.5.2.4   The moving mean reverting model

The most general of the models presented in this section is the *moving mean reverting* (MMR) model proposed by Wells (1994). It represents an extension to the mean reverting model. In the moving mean model the regression coefficients are mean reverting with the mean itself being modeled as a random walk:

$$\boldsymbol{\xi}_{t+1} - \bar{\boldsymbol{\xi}}_{t+1} = \boldsymbol{T}(\boldsymbol{\xi}_t - \bar{\boldsymbol{\xi}}_t) + \boldsymbol{\eta}_t, \tag{3.81}$$

with

$$\bar{\boldsymbol{\xi}}_{t+1} = \bar{\boldsymbol{\xi}}_t + \boldsymbol{\varsigma}_t, \tag{3.82}$$

where $\boldsymbol{R}$ is assumed to be the identity matrix. The dynamic process for the mean state can be brought into the state vector leading to a state space model similar to (3.79)–(3.80), only with $\bar{\boldsymbol{\xi}}$ being replaced by $\bar{\boldsymbol{\xi}}_t$. Note that in contrast to the MR model above, all elements of $\boldsymbol{\eta}_t$ have to be unequal to zero to allow for time-variation in the mean. A more detailed treatment of the moving mean model and its specification in `Ox` using `SsfPack` is given in §3.7.

3.5.3   Initial values

Two different sets of starting values are needed to initialize the Kalman filter for the various time-varying parameter regression models. With respect to the initial values of the state vector and its covariance, a diffuse prior as described in §3.3.6 is used for random coefficient and mean reverting models. As long as no numerical problems arise due to rounding errors, the same strategy is pursued for the random walk and the moving mean reverting model. In case of numerical difficulties, the initial state values of random walk and moving mean models are estimated via OLS using the first ten observations, as proposed by Wells (1996, §5.8) and Yao and Gao (2004).

The second type of initial values is concerned with the hyperparameters that have to be estimated by ML. The initial means are set to the OLS estimates over the entire sample. All diagonal elements of the transition matrix $\boldsymbol{T}$ that are neither zero nor unity, are set to 0.5. With the initial value for σ^2 set to 1.001, the variances of the state and observation disturbances are initialized using the parameter transformation introduced by (3.63).

3.6   Model diagnostics

Once a model has been implemented, it is necessary to check whether the assumptions underlying the model hold. By applying formal statistical test procedures, the quality of the estimated model can be evaluated. Subsection 3.6.1 introduces two different types of time series residuals that are commonly used for model diagnostic purposes. Subsection 3.6.2 summarizes the computation of basic measures of the goodness of fit for structural time series models. Subsection 3.6.3 refers to the employment of general diagnostic tests.

3.6.1 Residuals

Following the outline by Harvey (1989, §7.4.1) two sets of time series residuals can be computed for a state space model with an observation equation as defined in (3.69). While the generalized recursive residuals are calculated by standardization of the Kalman filter innovations, the generalized least squares residuals can be obtained by standardization of the Kalman smoother or disturbance smoother innovations.

3.6.1.1 Generalized recursive residuals

In accordance with the definition of an innovation term given in (3.15), $T - d - k$ one-step ahead prediction errors can be obtained by applying the Kalman filter to the state space model defined in (3.70)–(3.71):

$$v_t^\dagger = y_t - E(y_t | Y_{t-1}) = y_t - \boldsymbol{z}_t \boldsymbol{a}_t - \boldsymbol{x}_t \tilde{\boldsymbol{\beta}}_{t|t-1}, \qquad t = d + k + 1, \ldots, T, \qquad (3.83)$$

where $\tilde{\boldsymbol{\beta}}_{t|t-1}$ denotes the ML estimate of $\boldsymbol{\beta}_t$ based on the set of observations up to time $t - 1$. Standardizing the recursive residuals by dividing $v_t^\dagger$ by $\sqrt{f_t}$ as defined in (3.16) yields the generalized recursive residuals, denoted as $\tilde{v}_t^\dagger$:

$$\tilde{v}_t^\dagger = \frac{v_t^\dagger}{\sqrt{f_t}}. \qquad (3.84)$$

Without any stochastic model components, the Kalman filter becomes equivalent to standard OLS recursions such that $\tilde{v}_t^\dagger$ reduces to the recursive residuals. This property and the fact that recursive residuals are independent over time can be used in procedures to test for structural breaks.

3.6.1.2 Generalized least squares residuals

The application of the Kalman filter to the constructed series of observations $y_t - \boldsymbol{x}_t \tilde{\boldsymbol{\beta}}_{t|T}$, where $\tilde{\boldsymbol{\beta}}_{t|T}$ denotes the ML estimate of $\boldsymbol{\beta}_t$ given the complete sample up to time T, yields $T - d$ least squares residuals

$$v_t^+ = y_t - \boldsymbol{z}_t \boldsymbol{a}_t - \boldsymbol{x}_t \tilde{\boldsymbol{\beta}}_{t|T}, \qquad t = d + 1, \ldots, T. \qquad (3.85)$$

In case where the coefficient vector $\boldsymbol{\beta}_t$ is included in the state vector, overall two Kalman filters are needed: one to obtain $\tilde{\boldsymbol{\beta}}_{t|T}$ and another one that is applied to $y_t - \boldsymbol{x}_t \tilde{\boldsymbol{\beta}}_{t|T} = \boldsymbol{z}_t \boldsymbol{\xi}_t + \epsilon_t$ (cf. Durbin and Koopman 2001, §6.2.4). Dividing the least squares residuals by $\sqrt{f_t}$ yields the generalized least squares (GLS) residuals

$$\tilde{v}_t^+ = \frac{v_t^+}{\sqrt{f_t}}. \qquad (3.86)$$

In the absence of any stochastic components, these residuals resemble standard OLS residuals.

For a detailed description of applying the two types of residuals defined above to major test procedures for models in state space form, the reader is referred to Harvey (1989, §5 & §7.4).

3.6.2 Goodness of fit

Measures of the goodness of fit are usually computed on the basis of forecast errors. The basic goodness of fit measure for time series models is the prediction error variance, which also serves as an input to calculate the coefficient of determination. A comparison between models with different numbers of parameters is commonly made on the basis of information criteria.

3.6.2.1 Prediction error variance

In case of time-homogeneity, the prediction error variance (p.e.v.) is defined as

$$\sigma^2 = \sigma_*^2 \bar{f}, \tag{3.87}$$

where $\bar{f}$ represents the value to which f_t converges in steady-state. When a concentrated likelihood function is used with σ_*^2 being estimated by (3.61), σ^2 can be estimated as

$$\tilde{\sigma}^2 = \tilde{\sigma}_*^2 \bar{f}. \tag{3.88}$$

The incorporation of regression effects into a model slightly changes the definition of the estimator of σ_*^2. When dealing with time-varying regression models, the ML estimator of σ_*^2 for stochastic $\boldsymbol{\beta}_t$ is a function of the generalized recursive residuals:

$$s_*^2 = (T - d - k)^{-1} \sum_{t=d+k+1}^{T} \tilde{v}_t^{\dagger 2}, \tag{3.89}$$

where k is equal to the dimension of the vector of explanatory variables. For fixed $\boldsymbol{\beta}_t$, the ML estimator of σ_*^2 depends on the GLS residuals:

$$\tilde{\sigma}_*^{+2} = (T - d)^{-1} \sum_{t=d+1}^{T} \tilde{v}_t^{+2}, \tag{3.90}$$

where the sums of squares are identical in both cases. Taking (3.89) and (3.90) into account, the prediction error variance for models that contain explanatory variables can be estimated as

$$s^2 = s_*^2 \bar{f}, \tag{3.91}$$

or as

$$\tilde{\sigma}^{+2} = \tilde{\sigma}_*^{+2} \bar{f}. \tag{3.92}$$

According to Harvey (1989, §7.4.3) the prediction error variance can be approximated for large samples in terms of the unstandardized GLS residuals as defined in (3.85):

$$\tilde{\sigma}^2 = \frac{\bar{f}}{T - d} \sum_{t=d+1}^{T} \tilde{v}_t^{+2} = \frac{1}{T - d} \sum_{t=d+1}^{T} \frac{v_t^{+2}}{f_t} \bar{f} \approx \frac{1}{T - d} \sum_{t=d+1}^{T} v_t^{+2}. \tag{3.93}$$

3.6.2.2   Coefficient of determination

The coefficient of determination, R^2, is one of the most widely used measures of fit for time series models with stationary data. In the univariate case, the standard R^2 is computed by subtracting the ratio between the sum of squared errors (SSE) and the sum of squares of observations about their mean from unity:

$$R^2 = 1 - \frac{SSE}{\sum_{t=d+1}^{T} (y_t - \bar{y})^2},$$

(3.94)

where $\bar{y}$ denotes the mean of the dependent variable over the sample $t = d+1, \ldots, T$. The residual sum of squares is defined as

$$\text{SSE} = (T-d)\tilde{\sigma}^2 = (T-d-k)s^2.$$

(3.95)

An adjusted R^2 that penalizes models with a large number of explanatory variables, commonly denoted as $\bar{R}^2$, will be computed in standard fashion as

$$\bar{R}^2 = 1 - \frac{\left(1 - R^2\right)\left(T - d - 1\right)}{T - d - k}.$$

(3.96)

3.6.2.3   Information criteria

Another way of comparing different models is to use an information criterion that balances between the achieved goodness of fit and the degree of parsimony of a model. The two most widely used criteria are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), proposed by Akaike (1973) and Schwarz (1978), respectively. In the following, the AIC and BIC will be defined in terms of the value of the loglikelihood function for given $\hat{\psi}$:

$$\text{AIC} = [-2 \log L(y|\hat{\psi}) + 2w]/T,$$

(3.97)

$$\text{BIC} = [-2 \log L(y|\hat{\psi}) + w \log T]/T,$$

(3.98)

where w denotes the dimension of ψ. When using one of these two information criteria as a guide for model selection, the model with the smallest information criterion is to be preferred. Unless mentioned otherwise, the BIC will be preferred in the following as it imposes the higher penalty for each additional parameter (given that $T > 7$).

3.6.3   Diagnostics

Diagnostic tests are employed to check whether the disturbances of an estimated model are approximately random. Generally, both types of residuals introduced in §3.6.1 can be employed for model diagnostics, each revealing different aspects of the fitted model. While the recursive residuals yield the advantage of being serially uncorrelated, the least squares residuals are based on the information of the entire sample and allow for a direct comparison with OLS estimates. Unless stated otherwise, the basic diagnostics for testing on normality, heteroskedasticity and serial correlation will employ the GLS residuals defined in (3.86). Tests for detecting structural change are usually based on the

standardized one-step ahead prediction errors given in (3.84); definitions of the various test procedures and a discussion of their appropriateness for state space models can be found in Harvey (1989, §5.4 & §7.4).

3.7   Illustration: How to specify the MMR model for estimation using SsfPack

This section illustrates how to specify a state space model for estimation using SsfPack. Together with the technical paper by Koopman et al. (1999) the following outline allows for a hands-on application of the Kalman filter based methods discussed in this chapter. To demonstrate the application of the Kalman filter to the time-varying coefficient models presented in the previous section, we have a look at the univariate moving mean model with a single explanatory variable x_t, with regression coefficient β_t. The nonzero intercept term α_t is assumed to follow an AR(1) process. Following the logic of augmentation as in §3.5.1, this model is obtained by rewriting the observation equation as

$$
y_t = \begin{bmatrix} 1 & x_t & x_t \end{bmatrix} \begin{bmatrix} \alpha_t \\ \beta_t - \bar{\beta}_t \\ \bar{\beta}_t \end{bmatrix} + \epsilon_t, \qquad \epsilon_t \sim N\left(0, \sigma_\epsilon^2\right), \tag{3.99}
$$

with the corresponding state vector being formulated as

$$
\alpha_{t+1} = T_{11}\alpha_t + \nu_t, \qquad\qquad \nu_t \sim N\left(0, \sigma_\nu^2\right), \tag{3.100}
$$
$$
\beta_{t+1} - \bar{\beta}_{t+1} = T_{22}(\beta_t - \bar{\beta}_t) + \zeta_t, \qquad \zeta_t \sim N\left(0, \sigma_\zeta^2\right), \tag{3.101}
$$
$$
\bar{\beta}_{t+1} = \bar{\beta}_t + \varsigma_t, \qquad\qquad \varsigma_t \sim N\left(0, \sigma_\varsigma^2\right), \tag{3.102}
$$

where ν_t, ζ_t and ς_t are assumed to be normally distributed and mutually independent state disturbances.

In connection with SsfPack, it is useful to formulate the moving mean model in state space form by utilizing the compact representation of a state space model as presented in (3.8)–(3.10):

$$
\begin{bmatrix} \boldsymbol{\xi}_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{T} \\ \boldsymbol{z}_t \end{bmatrix} \boldsymbol{\xi}_t + \boldsymbol{v}_t \qquad \boldsymbol{v}_t \sim N\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & h \end{bmatrix}\right), \tag{3.103}
$$

with

$$
\boldsymbol{\xi}_t = \begin{bmatrix} \alpha_t \\ \beta_t - \bar{\beta}_t \\ \bar{\beta}_t \end{bmatrix}, \quad \boldsymbol{T} = \begin{bmatrix} T_{11} & 0 & 0 \\ 0 & T_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{z}_t = \begin{bmatrix} 1 & x_t & x_t \end{bmatrix},
$$

$$
\boldsymbol{v}_t = \begin{bmatrix} \boldsymbol{R} \begin{bmatrix} \nu_t \\ \zeta_t \\ \varsigma_t \end{bmatrix} \\ \epsilon_t \end{bmatrix}, \quad \boldsymbol{R} = \boldsymbol{I}_3, \quad \boldsymbol{Q} = \begin{bmatrix} \sigma_\nu^2 & 0 & 0 \\ 0 & \sigma_\zeta^2 & 0 \\ 0 & 0 & \sigma_\varsigma^2 \end{bmatrix}, \quad h = \sigma_\epsilon^2,
$$

$$\tag{3.104}$$

where the six parameters σ_ϵ^2, σ_ν^2, σ_ζ^2, σ_ς^2, T_{11} and T_{22} have to be estimated. For $T_{11} = 1$, α_t follows a random walk. If the variance of ν_t is equal to zero, the intercept term

becomes a constant. For values of σ_ς^2 close to zero, the model resembles a mean reverting model with a constant mean.

In `SsfPack`, the initial state values are collected in the matrix $\boldsymbol{\Sigma}$. In case of the moving mean model, it is defined as

$$
\boldsymbol{\Sigma} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \tag{3.105}
$$

where the value -1 indicates that the corresponding elements of the initial state vector are diffuse. If OLS estimates were used as initial values, the diagonal elements of the first three rows would be replaced by the corresponding elements of the estimated covariance matrix; the last row would contain the estimated OLS regression coefficients as initial values for the corresponding means of the state vector.

The lower part of the parameter matrix $\boldsymbol{\Phi}_t = [\boldsymbol{T} \quad \boldsymbol{z}_t]'$ contains the time-varying system element $\boldsymbol{z}_t$ with the explanatory variables. In `SsfPack`, an index matrix $\boldsymbol{J}_\Phi$, whose index refers to the explanatory data in $\boldsymbol{z}_t$, is required to specify the time-varying elements in $\boldsymbol{\Phi}_t$. The two matrices $\boldsymbol{\Phi}_t$ and $\boldsymbol{J}_\Phi$ must be of same dimension. With the exception of those elements for which time-variation in the corresponding elements of $\boldsymbol{\Phi}_t$ should be indicated, the elements of the index matrix take the value -1. An index value greater than -1 refers to the column of $\boldsymbol{z}_t$ with the time-varying values. For the moving mean model, the index matrix is given by

$$
\boldsymbol{J}_\Phi = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ 0 & 1 & 2 \end{bmatrix}. \tag{3.106}
$$

Notice that indexing in `Ox` generally starts at zero. Therefore, the (4,1) element of $\boldsymbol{J}_\Phi$ refers to the first column of $\boldsymbol{z}_t$.

As the corresponding elements in the last row of $\boldsymbol{J}_\Phi$ already indicate where the data for the respective columns of $\boldsymbol{z}_t$ come from, the values associated with $\boldsymbol{z}_t$ in the last row of $\boldsymbol{\Phi}_t$ are set to zero when dealing with time-variation. In the current example $\boldsymbol{\Phi}_t$ becomes

$$
\boldsymbol{\Phi}_t = \begin{bmatrix} T_{11} & 0 & 0 \\ 0 & T_{22} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{3.107}
$$

Generally, `SsfPack` allows to assign time-varying elements to any system matrix by using the index matrices $\boldsymbol{J}_\delta$, $\boldsymbol{J}_\Omega$ and $\boldsymbol{J}_\Phi$ to indicate the time-varying elements of $\boldsymbol{\delta}_t$, $\boldsymbol{\Omega}_t$ and $\boldsymbol{\Phi}_t$, respectively (cf. Zivot et al. 2002).

Chapter 4

Markov regime switching

Markov switching models of changing regimes are latent variable time series models. The observation-generating distribution depends on an unobserved, or *hidden*, state variable modeled as a Markov chain. Markov switching models, also commonly known as hidden Markov models (HMMs), offer a high degree of flexibility and can be employed for both univariate and multivariate series. A hidden Markov model represents a special class of dependent mixtures and consists of two processes: an unobservable m-state Markov chain that determines the state, or *regime*, and a state-dependent process of observations. The hidden Markov model is closely related to the general linear Gaussian state space model introduced in Chapter 3. Both models are state space models. The major difference is that HMMs have *discrete* states, while the Kalman filter based state space approach deals with unobserved *continuous* states.[9]

While HMMs have been employed by engineers in the context of signal-processing, over the last two decades extensive literature developed in automatic speech recognition, biosciences, image processing and climatology, among others. Important references include Baum and Petrie (1966), the tutorial by Rabiner (1989) and Ephraim and Merhav (2002). Economic and financial researchers are also frequently confronted with time series that experience changes in regime that are evoked by third factors. The shifts are not observed directly, and usually it is unknwown which regime currently prevails. However, it was not until the seminal works of Hamilton (1988, 1989, 1990) that economists and financial econometricians started to apply HMMs. Hamilton introduced a homogeneous Markov switching vector autoregressive model, in which the latent state process is independent from exogenous variables. Lahiri and Wang (1996) studied the comparative performance of various interest rate spreads as leading indicators for turning points of the U.S. business cycle. They assumed the U.S. economy to shift between two states with the shift between regimes being governed by a two-state Markov process. Fridman (1994) proposed a two-state CAPM where the states represent two market regimes of high and low volatility. The parameters are determined by an unobserved Markov chain. Another well-known application to finance is conducted by Rydén et al. (1998)

---

[9]For a review of the common properties of both concepts, see Roweis and Ghahramani (1999).

who employ a hidden Markov model to derive stylized facts of daily S&P 500 return series.

The standard regime switching model can be extended to provide an even higher degree of flexibility. Important specifications include Markov switching models with time-varying switching probabilities (Diebold et al. 1994; Filardo 1994) and hidden *semi*-Markov models, which allow for nonparametric state occupancy distributions, as proposed by Ferguson (1980). As these extensions are beyond the scope of this thesis, they will not find any consideration hereafter; for further reading, see, for example, Bulla (2006, §5) and the references given therein.

HMMs are employed in a wide spectrum of applications due to their overall versatility and mathematical tractability. Their attractiveness is grounded on the fact that the likelihood can be evaluated in a straightforward fashion, either by numerical maximization or by employing the EM algorithm. Moment properties can be easily derived, missing observations can be easily dealt with, and HMMs are often interpretable in a natural way. Besides, they are moderately parsimonious with a simple two-state model providing a good fit in many cases. Comprehensive treatments of HMMs are provided by Elliott et al. (1995), MacDonald and Zucchini (1997), Bhar and Hamori (2004) and Hamilton (1993), with the latter two focusing on applications in economics and finance.

The organization of this chapter is as follows. Section 4.1 briefly reviews independent mixture models and basic properties of Markov chains. Section 4.2 outlines the basic hidden Markov model. Section 4.3 briefly discusses parameter estimation based on the ML method. Section 4.4 looks at forecasting and decoding procedures, before Section 4.5 provides an overview of model selection and validation. To assure notational conformability, the outline in this chapter is, unless stated otherwise, based on Bulla (2006, §2–3) who also provided the code for conducting the computations related to HMMs in this thesis using the statistical software package R 2.1.1 (R Development Core Team 2005).

4.1   Basic concepts

Before introducing the theory of HMMs, it is instructive to look at two fundamental concepts as a basis to understand the basic structure of a hidden Markov model. Subsection 4.1.1 reviews the concept of mixture distributions. Subsection 4.1.2 provides a brief introduction to the basic properties of Markov chains, which are needed to construct HMMs.

4.1.1   Independent mixture distributions

Consider the weekly log-return series of the Technology sector. As discussed in §2.2.2, the series of all sectors show signs of positive autocorrelation in the squared returns. This characteristic is commonly referred to as volatility clustering, a common feature of financial returns data that usually induces excess kurtosis. Panel (a) of Figure 4.1 shows the weekly percentage log-returns on the DJ Technology sector. It can be seen that the sector's volatility is not constant over time: the level of volatility is clearly lower at the beginning and higher in the second half of the sample. As a consequence, a normal distribution is not capable of describing the return series adequately. Panel (b)
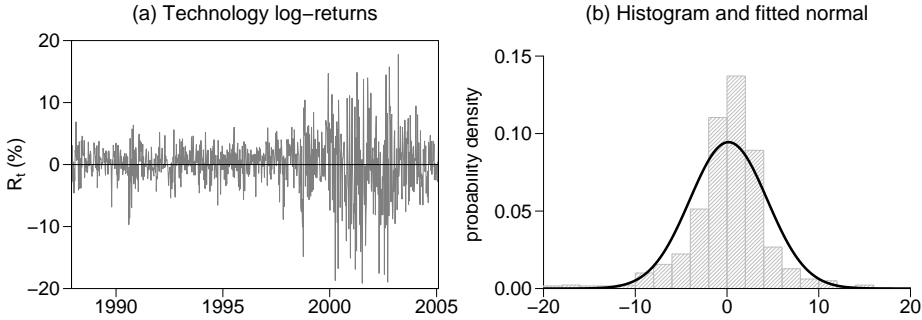
Figure 4.1: *(a) Weekly percentage log-return series of the Technology sector and (b) histogram with a fitted normal distribution.*

displays a histogram of the weekly returns together with a fitted normal distribution. It becomes obvious that the chosen normal underestimates the probability of both low returns around zero and extremely high absolute returns.

One possibility to overcome the shortcomings of the normal distribution is to employ a mixture of two normal distributions. Mixture distributions are useful in the context of overdispersed or multimodal data that may be caused by unobserved heterogeneity in the sample. An independent mixture distribution can either be modeled as a *discrete mixture*, which is characterized by a finite number of component distributions, or as a *continuous mixture*, which can be thought of as a discrete mixture consisting of infinitely many component distributions. Note that in both cases discrete or continuous distributions can be chosen as components. As only discrete mixtures are relevant for the HMMs considered in the context of this chapter and also for the applications in the empirical part of this thesis, the continuous case will not find any further consideration.[10]

In the general m-component case, the characteristics of the mixture distribution are determined by m random variables $X_1, \ldots, X_m$, and their probability or probability density functions, denoted as $p_i(x)$ or $f_i(x)$, respectively, for $i = 1, \ldots, m$. The mixture is performed using a discrete random variable S that determines from which random variable an observation is drawn. It can take values between 1 and m, each with a probability π_i:

$$S := \begin{cases} 1 \text{ with probability } \pi_1 \\ 2 \text{ with probability } \pi_2 \\ \vdots \\ m \text{ with probability } \pi_m, \end{cases} \tag{4.1}$$

where $\sum_{i=1}^{m} \pi_i = 1$ and $\pi_i \geq 0$ for $i = 1, \ldots, m$. With $\pi_1, \ldots, \pi_m$ representing the weights of the various components, the probability (density) function of the mixture distribution

---

[10]A survey of mixture distributions, is given, for example, by Titterington et al. (1985). For more details on continuous mixture distributions, see the references provided by Zucchini et al. (2006, §2.1).

based on X can be computed by a linear combination of the single components:

$$p(x) \quad = \quad \sum_{i=1}^{m} \pi_i p_i(x) \quad \text{(discrete case)}, \tag{4.2}$$

$$f(x) \quad = \quad \sum_{i=1}^{m} \pi_i f_i(x) \quad \text{(continuous case)}, \tag{4.3}$$

with the k-th moment of the mixture defined as a linear combination of the respective components moments

$$E(X^k) = \sum_{i=1}^{m} \pi_i E(X_i^k), \qquad k = 1, 2, \dots . \tag{4.4}$$

As it can be shown that the variance of a mixture model cannot be simply computed as a linear combination of the respective components variances, the standard equality

$$Var(X) = E(X^2) - (E(X))^2, \tag{4.5}$$

can be employed together with (4.4) to estimate the variance of a mixture model (cf. Zucchini et al. 2006, §2).

   The parameters of a mixture distribution are usually estimated by ML. For the example of the continuous case, the likelihood of an m-components mixture model can generally be stated as

$$L(\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_m, x_1, \dots, x_T) = \prod_{j=1}^{T} \sum_{i=1}^{m} \pi_i f_i(x_j, \boldsymbol{\psi}_i), \tag{4.6}$$

with observations $x_1, \dots, x_T$. The mixture weights $\pi_1, \dots, \pi_m$ and the parameter vectors of the component distributions are included in $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_m$. As the ML estimates $\hat{\boldsymbol{\psi}}_1, \dots, \hat{\boldsymbol{\psi}}_m$ represent a solution to a system of nonlinear equations, the likelihood can be maximized analytically only for rather trivial models. In most cases, the unknown parameters have to be estimated by employing direct numerical maximization procedures or by using the EM algorithm (cf. §3.4.3). Figure 4.2 shows the results of fitting mixtures of two (a) and three normals (b) to the weekly log-returns of the Technology sector. The components' weights have been obtained by a hidden Markov model for which more details will be provided in the section below. Compared to the case where the returns are fitted by a single normal distribution, the fit is clearly improved by both mixture models.

4.1.2 Markov chains

Let $\{S_t : t = 1, \dots, T\}$ be a stochastic process, i.e. a sequence of random variables that can assume an integer value in $S = \{1, \dots, m\}$, the state space. If for each date t, the probability that S_{t+1} is equal to a particular value $s_t \in S$ depends only on S_t, the current state of the process, such a process is called an m-state *Markov process*:

$$P(S_{t+1} = s_{t+1} | S_t = s_t, S_{t-1} = s_{t-1}, \dots, S_1 = s_1) = P(S_{t+1} = s_{t+1} | S_t = s_t). \tag{4.7}$$
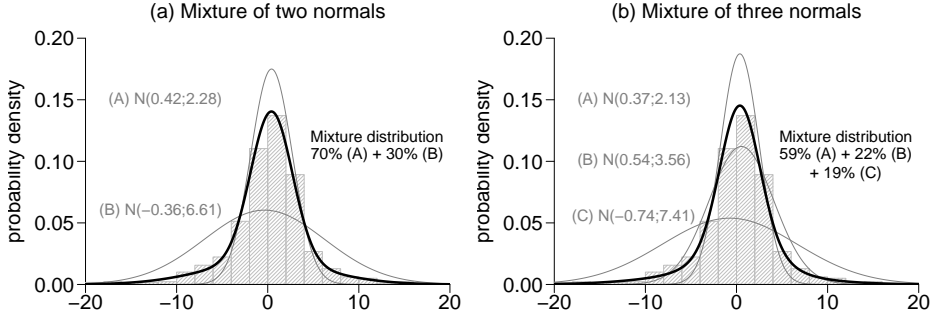
Figure 4.2: *Histogram of weekly log-returns of the Technology sector and fitted mixtures with (a) two and (b) three normal distributions.*

This is known as the *Markov property*. While both the state space and the time set can be either discrete or continuous, in the following only discrete-time Markov processes in discrete state space, referred to as *Markov chains*, are considered.

The probability of changing from state i to state j is called the *transition probability*. Even though the transition probability can be modeled as depending on date t in the general case, in this thesis only *homogeneous* Markov chains with constant transition probabilities over time will be considered, i.e., $\gamma_{ij} := P(S_{t+1} = j | S_t = i)$ for all t with $\sum_{j=1}^{m} \gamma_{ij} = 1$ and $1 \leq i \leq m$. The one-step ahead transition probabilities can be collected in an $m \times m$ matrix $\mathbf{\Gamma}$, which is referred to as *transition probability matrix*:

$$\mathbf{\Gamma} := \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm} \end{pmatrix}, \tag{4.8}$$

with each row summing up to unity. It can be shown that the matrix $\mathbf{\Gamma}(l)$ of l-period ahead transition probabilities $\gamma_{ij}(l) := P(S_{t+l} = j | S_t = i)$ can be computed by multiplying $\mathbf{\Gamma}$ by itself l times:

$$\mathbf{\Gamma}(l) := \begin{pmatrix} \gamma_{11}(l) & \gamma_{12}(l) & \cdots & \gamma_{1m}(l) \\ \gamma_{21}(l) & \gamma{22}(l) & \cdots & \gamma_{2m}(l) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1}(l) & \gamma_{m2}(l) & \cdots & \gamma_{mm}(l) \end{pmatrix} = \mathbf{\Gamma}^l. \tag{4.9}$$

In the following, any Markov chain will be assumed to be *irreducible*. This means that for all $i, j \in \{1, \ldots, m\}$ the probability of the chain changing from state i to state j, starting from state i, and vice versa, is positive.

While $\gamma_{ij}(l)$ describes the conditional probability of being in state j at time $t + l$, with the Markov chain starting from state i at time t, it does not provide the marginal probability of being in state i at a given time t. With the probability distribution of the initial state, $\boldsymbol{\pi}(1) := (\pi_1(1), \ldots, \pi_m(1)) = (P(S_1 = 1), \ldots, P(S_1 = m))$, the probability function of the state at time t is given by

$$\boldsymbol{\pi}(t) := (P(S_t = 1), \ldots, P(S_t = m)) = \boldsymbol{\pi}(t)\boldsymbol{\Gamma}^{l-1}. \tag{4.10}$$

For a homogeneous and irreducible Markov chain, $\boldsymbol{\pi}(t)$ can be shown to converge to a fixed vector $\boldsymbol{\pi}_s := (\pi_1, \ldots, \pi_m)$ for large t. This unique vector of dimension m satisfies

$$\boldsymbol{\pi}_s = \boldsymbol{\pi}_s\boldsymbol{\Gamma}, \tag{4.11}$$

and is called the vector of *stationary transition probabilities*. For existing $\boldsymbol{\pi}_s$, a Markov chain is referred to as being stationary if $\boldsymbol{\pi}_s$ describes the marginal distribution of the states for all $t = 1, \ldots, T$.

For more details on the well developed theory of Markov chains and further references, see, for example, Hamilton (1994b, §22).

4.2   The basic hidden Markov model

The sequence of observations and hidden states in an independent mixture model are independent by definition. Any potential correlation between the states cannot be captured by an independent mixture as it does not take into account the respective information. One method of modeling serially correlated time series is to use an unobserved Markov chain to select the parameters. This yields the hidden Markov model as a special dependent mixture model.

With $\{X_t\} = \{X_t, t = 1, \ldots, T\}$ denoting a sequence of observations and $\{S_t\} = \{S_t, t = 1, \ldots, T\}$ denoting a Markov chain in the state space $\{1, \ldots, m\}$, their respective histories up to time t can be written as

$$X^{(t)} := \{X_1, \ldots, X_t\}, \tag{4.12}$$
$$S^{(t)} := \{S_1, \ldots, S_t\}. \tag{4.13}$$

Consider a stochastic process that consists of two elements: (i) an underlying and unobserved parameter process $\{S_t\}$ for which the Markov property (4.7) holds, and (ii) a state-dependent observation process $\{X_t\}$, which fulfills the *conditional independence property*

$$P(X_t = x_t | X^{(t-1)} = x^{(t-1)}, S^{(t)} = s^{(t)}) = P(X_t = x_t \,|\, S_t = s_t). \tag{4.14}$$

This means that with known S_t, X_t only depends on S_t and not on the history of states or observations. The pair of stochastic processes $\{X_t\}$ and $\{S_t\}$ is referred to as an m-state hidden Markov model whose basic structure is illustrated in Figure 4.3, which is taken from Bulla (2006, §2).

Generally, different distributions are imposed for the various states. In this thesis, the Markov chain with transition probability matrix $\boldsymbol{\Gamma}$ will be assumed to be homogeneous
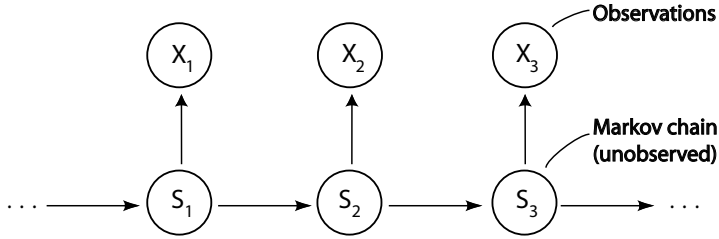
Figure 4.3: *Basic structure of a hidden Markov model (source: Bulla 2006, p. 17).*

and irreducible and to have a unique stationary distribution $\boldsymbol{\pi}_s$ (cf. §4.1.2). In practical applications, the underlying state process $\{S_t\}$ is hidden and only the state-dependent sequence of observations $\{X_t\}$ is known. Usually, the unobservable states can be interpreted in a natural way: for example, a hidden Markov model with two states could be employed in the context of the weekly return series of the Technology sector considered in §4.1.1: the observations in times of high volatility would be referred to by one state, while the other state would refer to low volatility markets.

For more details on the basic hidden Markov model, including a derivation of its moments and marginal distributions, see, for example, MacDonald and Zucchini (1997, §2).

4.3   Parameter estimation

The parameters of a hidden Markov model are generally estimated via the maximum likelihood principle. As the estimates represent a solution to a system of nonlinear equations, it is usually impossible to estimate the unknown parameters analytically. Instead, one has to refer either to direct numerical maximization procedures or to the EM algorithm, which was briefly mentioned in §3.4.3. In this thesis, the focus will be on parameter estimation by means of direct numerical maximization procedures. Provided that the initial values to be employed are adequately accurate, numerical maximization leads to faster convergence than the EM algorithm. Direct numerical maximization, which avoids the derivation of the required formula for the EM algorithm, can be used both for non-stationary and for stationary Markov chains; cf. Bulla and Berzel (2006) who compare the competing estimation procedures based on a simulation experiment. For details on the implementation of the EM algorithm with regard to HMMs, the reader is referred to Baum et al. (1970) and MacDonald and Zucchini (1997, §2).

The likelihood function of the basic hidden Markov model is presented in §4.3.1. Subsection 4.3.2 briefly overviews parameter estimation by numerically maximizing the likelihood.

4.3.1   The likelihood function

In order to estimate the parameters of a hidden Markov model by ML, a general likelihood function of a hidden Markov model in closed from can be derived as

$$L(\boldsymbol{\psi}) = P(\{X_1 = x_1, \ldots, X_T = x_T\}) = \boldsymbol{\pi}_s \boldsymbol{P}(x_1)\boldsymbol{\Gamma}\boldsymbol{P}(x_2)\boldsymbol{\Gamma}\ldots\boldsymbol{\Gamma}\boldsymbol{P}(x_T)\mathbf{1}', \qquad (4.15)$$

with parameter vector $\boldsymbol{\psi}$ and $\mathbf{1} := (1, \ldots, 1)$. The diagonal matrix $\boldsymbol{P}(x_t)$ contains the constant state-dependent probabilities of x_t for $t = 1, \ldots, T$; for a detailed derivation, see, for example, MacDonald and Zucchini (1997, §2.5).

Following Zucchini et al. (2006, §3.3) an appealing property of HMMs is that missing observations can be easily dealt with by replacing the respective diagonal matrices $\boldsymbol{P}(x_t)$ by the identity matrix $\boldsymbol{I}$. Besides, both stationary and non-stationary models can be handled. However, the evaluation of the likelihood usually involves a high number of matrix multiplications with elements between zero and unity. This frequently leads to problems of numerical underflow. Irrespective of whether the likelihood is maximized by employing methods of direct numerical maximization or the EM algorithm, this problem can be overcome by applying some form of rescaling as proposed, for example, by Rabiner (1989).

4.3.2   Direct numerical maximization

Following the outline of Zucchini et al. (2006, §4.3), direct numerical maximization of the likelihood in (4.15) has two major advantages over the EM algorithm: the flexibility to fit alternative and more complex specifications, and the ease with which missing observations can be dealt. On the other hand, two major problems are related to the implementation of parameter constraints and to numerical underflow.

Whenever the parameters of a hidden Markov model are not allowed to take all values in $\mathbb{R}$, the set of real numbers, it becomes necessary to consider parameter restrictions in the estimation procedure to be employed. As already outlined in the context of ML estimation of Kalman filter based Gaussian state space models, it is usually recommendable to reparameterize the model and to perform any maximization with respect to transformed but unconstrained quantities. The transformations and corresponding back-transformations introduced in §3.4.4 also apply to HMMs.

The problem of numerical underflow, where the likelihood for large t becomes very small for long series, is usually solved by considering the loglikelihood function. As for HMMs the likelihood represents a product of matrices, it is not trivial to maximize the loglikelihood. To allow for a recursive evaluation of the likelihood, a closed form of the loglikelihood is needed. Before the derivation is demonstrated in §4.3.2.2, the so-called *forward-backward probabilities* are introduced in §4.3.2.1.

4.3.2.1 Forward-backward probabilities

Using the notation for the history of observations and states introduced in (4.12) and (4.13), the likelihood function (4.15) can be reformulated as

$$L(\boldsymbol{\psi}) = \underbrace{\boldsymbol{\pi}_s \boldsymbol{P}(x_1)\boldsymbol{\Gamma}\ldots\boldsymbol{\Gamma}\boldsymbol{P}(x_t)}_{=:\boldsymbol{\alpha}_t}\underbrace{\boldsymbol{\Gamma}\boldsymbol{P}(x_{t+1})\boldsymbol{\Gamma}\ldots\boldsymbol{\Gamma}\boldsymbol{P}(x_T)\boldsymbol{1}'}_{=:\boldsymbol{\beta}_t'} = \boldsymbol{\alpha}_t\boldsymbol{\beta}_t', \qquad (4.16)$$

for $t = 1,\ldots,T$, with $\boldsymbol{\alpha}_0 = \boldsymbol{\pi}_s$ and $\boldsymbol{\beta}_T' = \boldsymbol{1}'$.

The introduced m-vector vector $\boldsymbol{\alpha}_t = (\alpha_t(1),\ldots,\alpha_t(m))$ contains the *forward probabilities*, defined as

$$\alpha_t(i) = P(X^{(t)} = x^{(t)}, S_t = i), \qquad (4.17)$$

for $i = 1,\ldots,m$. Each component $\alpha_t(i)$ can be interpreted as the probability of the hidden Markov model's history up to time t, with the Markov chain being at state i at the same date.

Similarly, $\boldsymbol{\beta}_t = (\beta_t(1),\ldots,\beta_t(m))$, consists of the so-called *backward probabilities* given as

$$\beta_t(i) = (P(X_{t+1} = x_{t+1},\ldots,X_T = x_T|S_t = i), \qquad (4.18)$$

which represent the conditional probability of future observations $x_{t+1},\ldots,x_T$, for state i of the Markov chain at time t; cf. Zucchini et al. (2006, §4.1) where a proof for these results is given.

4.3.2.2 Recursive evaluation of the loglikelihood

From the forward probabilities introduced above, for $L_T = \boldsymbol{\pi}_s \boldsymbol{B}_1\ldots\boldsymbol{B}_T\boldsymbol{1}'$ with $\boldsymbol{B}_t = \boldsymbol{\Gamma}\boldsymbol{P}(x_t)$, it follows that

$$L_T = \boldsymbol{\alpha}_T. \qquad (4.19)$$

We define $w_t := \boldsymbol{\alpha}_t\boldsymbol{1}'$, with the scalar w_t being equal to the sum of the forward probabilities, and $\boldsymbol{\phi}_t := \boldsymbol{\alpha}_t/w_t$, where $\boldsymbol{\phi}_t$ contains the rescaled forward probabilities. The loglikelihood L_T can be evaluated by the following recursion with starting equation

$$\boldsymbol{\phi}_0 = \frac{\boldsymbol{\alpha}_0}{w_0} = \frac{\boldsymbol{\pi}_s}{\boldsymbol{\pi}_s\boldsymbol{1}'} = \boldsymbol{\pi}_s, \qquad (4.20)$$

and updating equation

$$\boldsymbol{\phi}_t = \frac{\boldsymbol{\alpha}_t}{w_t} = \frac{\boldsymbol{\alpha}_{t-1}\boldsymbol{B}_t}{w_t} = \frac{w_{t-1}}{w_t}\boldsymbol{\phi}_{t-1}\boldsymbol{B}_t, \qquad (4.21)$$

for t=1,...,T, with $\boldsymbol{\alpha}_0 := \boldsymbol{\pi}_s$ and $\boldsymbol{\alpha}_t := \boldsymbol{\alpha}_{t-1}\boldsymbol{B}_t$. For $t = T$ the vector of rescaled forward probabilities can be derived as

$$\boldsymbol{\phi}_T = \frac{w_{T-1}}{w_T}\frac{w_{T-2}}{w_{T-1}}\ldots\frac{w_0}{w_1}\boldsymbol{\alpha}_T, \qquad (4.22)$$

which can be converted to yield the likelihood:

$$L_T = \prod_{t=1}^{T}\frac{w_t}{w_{t-1}}. \qquad (4.23)$$

Taking the logarithm of L_T gives the loglikelihood function as

$$\log L_T = \sum_{t=1}^{T} \log \left(\frac{w_t}{w_{t-1}} \right). \tag{4.24}$$

It can be shown that the ratio of the scaling factors can be obtained as $w_t/w_{t-1} = \phi_{t-1} \mathbf{B}_t \mathbf{1}'$. An efficient algorithm for recursively evaluating the loglikelihood based on the scaled forward probabilities can be derived with starting equations

$$\log L_0 = 0, \tag{4.25}$$

$$\phi_0 = \pi_s, \tag{4.26}$$

and updating equations

$$\mathbf{v}_t = \phi_{t-1} \mathbf{B}_t, \tag{4.27}$$

$$u_t = \mathbf{v}_t \mathbf{1}', \tag{4.28}$$

$$\log L_t = \log L_{t-1} + \log u_t, \tag{4.29}$$

$$\phi_t = \mathbf{v}_t u_t^{-1}. \tag{4.30}$$

To obtain the loglikelihood function, the loop is repeated for $t = 1, \ldots, T$ (cf. Zucchini et al. 2006, §4.3).

Together with a statistical software package that offers functions for numerical maximization (or minimization), this algorithm can be employed to estimate the unknown parameters of a hidden Markov model by means of direct numerical maximization. In this thesis, the ML parameters of HMMs are estimated using the R-functions `nlm()` and `optim()`, which can be employed to minimize a negative loglikelihood.

4.3.3   Standard errors of ML estimates

In the context of HMMs it is not trivial to determine the accuracy of the ML parameters estimated by one of the methods described above. As the elements of the vector of estimated parameters can be shown to be correlated, it is not possible to compute the standard error of the overall model directly. One possibility to obtain distributional properties of the parameter estimates is to apply parametric bootstrap methods. As these procedures are beyond the scope of this thesis, in the following the quality of a hidden Markov model will be evaluated based on its forecast performance rather than using classical statistical fit statistics. For details on how parametric bootstrap procedures can be employed to analyze the distributional properties of parameter estimates, the reader is referred to Zucchini et al. (2006, §4.4) and the references given therein.

4.4   Forecasting and decoding

Based on Zucchini et al. (2006, §5) this section briefly summarizes how HMMs can be used to produce forecasts (§4.4.1), and how information with regard to the unobservable states of the Markov chain can be obtained (§4.4.2). The inference regarding the hidden states is commonly referred to as *decoding*.

4.4.1 Forecast distributions

Before a *forecast distribution* for HMMs can be derived, it is necessary to compute the corresponding *conditional distribution*. Let X^{-u} denote the sequence of random variables X_t for $t = 1, \ldots, T$ with X_u being excluded, and let x^{-u} denote the observations x_t for $t = 1, \ldots, T$ with x_u being excluded:

$$X^{(-u)} = \{X_1, \ldots, X_{u-1}, X_{u+1}, \ldots, X_T\}, \tag{4.31}$$

$$x^{(-u)} = \{x_1, \ldots, x_{u-1}, x_{u+1}, \ldots, x_T\}. \tag{4.32}$$

Together with the likelihood function in (4.15) and the forward-backward probabilities defined in §4.3.2.1, the conditional distribution of X_u for $u = 1, \ldots, T$, given all other observations, can be derived as

$$P\left(X_u = x | X^{(-u)} = x^{(-u)}\right) = \frac{\boldsymbol{\alpha}_{u-1} \boldsymbol{\Gamma} \boldsymbol{P}(x) \boldsymbol{\beta}_u'}{\boldsymbol{\alpha}_{u-1} \boldsymbol{\Gamma} \boldsymbol{\beta}_u'}. \tag{4.33}$$

The numerator can be regarded as the likelihood of the observed series with x being substituted for x_u. The denominator can be interpreted as the likelihood of the series with x_u treated as being missing. Alternatively, the conditional probability can be interpreted as a mixture of state-dependent probability distributions.

The forecast distribution of a hidden Markov model, which is needed to compute the probability of an observation occurring l steps in the future, represents a special type of a conditional distribution. It can be derived as

$$P\left(X_{T+l} = x | X^{(T)} = x^{(T)}\right) = \frac{\boldsymbol{\alpha}_T \boldsymbol{\Gamma}^l \boldsymbol{P}(x) \mathbf{1}'}{\boldsymbol{\alpha}_T \mathbf{1}'} = \boldsymbol{\phi}_T \boldsymbol{\Gamma}^l \boldsymbol{P}(x) \mathbf{1}'. \tag{4.34}$$

The scaling algorithm introduced in §4.3.2.2 can be employed to avoid problems of numerical underflow. For increasing l the forecast distribution converges to $\boldsymbol{\pi}_s \boldsymbol{P}(x) \mathbf{1}'$, the stationary distribution.

4.4.2 Decoding

Decoding refers to the determination of the most probable states of the Markov chain for an estimated hidden Markov model. One distinguishes between *local* decoding and *global* decoding. The former refers to the derivation of the most likely state at date t, which can also be used to generate state predictions as will be shown in §4.4.2.1. Global decoding looks for the most probable sequence of states (§4.4.2.3).

4.4.2.1 Local decoding

With the forward-backward probabilities introduced above, the joint probability of the observations $X^{(T)} = x^{(t)}$ and the Markov chain S_t being in state i at date t, can be shown to be equal to

$$\alpha_t(i)\beta_t(i) = P\left(X^{(T)} = x^{(T)}, S_t = i\right), \tag{4.35}$$

such that the conditional distribution of the Markov chain given $X^{(T)} = x^{(T)}$ can be derived as

$$P\left(S_t = i | X^{(T)} = x^{(T)}\right) = \frac{\alpha_t(i)\beta_t(i)}{L_T}, \tag{4.36}$$

for each date $t = 1, \ldots, T$, and states $i = 1, \ldots, m$. The likelihood can be evaluated by employing the efficient scaling algorithm presented in §4.3.2.2. This conditional distribution represents a discrete probability function that is allowed to take values between 1 and m. For the given set of observations, the most likely state for each time t, denoted as i_t^*, is the one that maximizes (4.36):

$$i_t^* = \underset{i \in \{1,\ldots,m\}}{\operatorname{argmax}} \; P\left(S_t = i | X^{(T)} = x^{(T)}\right). \tag{4.37}$$

4.4.2.2 State predictions

Local decoding can be applied to derive the states of past as well as of future states. Following Zucchini et al. (2006, §5.3.3) the following expression shows how smoothed, filtered and predicted state estimates can be derived by local decoding for given observations x_t:

$$P\left(S_t = i | X^{(T)} = x^{(T)}\right) = \begin{cases} \dfrac{\alpha_t(i)\beta_t(i)}{L_T} & \text{for } 1 \leq t < T \quad \text{``smoothing''}, \\[2mm] \dfrac{\alpha_T(i)}{L_T} & \text{for } t = T \qquad \text{``filtering''}, \\[2mm] \dfrac{\boldsymbol{\alpha}_T(\boldsymbol{\Gamma}^{t-T})_{\bullet j}}{L_T} & \text{for } t > T \qquad \text{``predicting''}, \end{cases} \tag{4.38}$$

where $(\boldsymbol{\Gamma}^{t-T})_{\bullet j}$ refers to the j-th column of $\boldsymbol{\Gamma}^{t-T}$.

4.4.2.3 Global decoding

While local decoding determines the most likely state for a chosen date t, in practice it is often more important to derive the most probable sequence of unobserved states. Instead of considering the conditional probability of a single state, global decoding requires the maximization of the joint probability

$$(i_1^*, \ldots, i_T^*) = \underset{(i_1,\ldots,i_T) \in \{1,\ldots,m\}}{\operatorname{argmax}} P\left(S_1 = i_1, \ldots, S_T = i_T, X^{(T)} = x^{(T)}\right), \tag{4.39}$$

with all states being taken into account simultaneously. The most likely sequence of states is denoted by $(i_1^*, \ldots, i_T^*)$. This state sequence cannot be determined by using local decoding for each t separately as local decoding partially ignores the transition probability matrix $\boldsymbol{\Gamma}$. This might lead to a path with zero probability even though for each single t the most likely state is detected. Besides, evaluating all possible m^T joint probabilities would lead to a high computational burden. A more efficient way to find the most likely sequence of hidden states is needed. In the context of Markov switching

models, the *Viterbi algorithm* is employed for this purpose. The Viterbi algorithm represents an efficient dynamic programming algorithm. It was originally proposed by Viterbi (1967) as an error-correction procedure in noisy digital communication applications.

According to Rabiner (1989) the Viterbi algorithm can be summarized as follows. At first, it is necessary to define the quantity

$$\nu_t(i) := \max_{i_1,\ldots,i_{t-1}} P\left(S_1 = i_1, \ldots, S_{t-1} = i_{t-1}, S_t = i, X^{(t)} = x^{(t)}\right), \qquad (4.40)$$

which represents the highest likelihood along a single path at date t accounting for the first t observations $X^{(t)} = x^{(t)}$ and ending at state $S_t = i$. It can be shown that $\nu_t(i)$ satisfies

$$\nu_{t+1}(j) = [\max_i(\nu_t(i)\gamma_{ij})] \cdot p_j(x_{t+1}), \qquad (4.41)$$

from which a $T \times m$ matrix of values $\nu_t(i)$ for $t = 1, \ldots, T$, and $j = 1, \ldots, m$, can be derived. The required sequence of hidden states $(i_1^*, \ldots, i_T^*)$ is computed in four steps:

1. Initialization:

$$\nu_1(i) = \pi_i \cdot p_i(x_1), \qquad 1 \leq i \leq m, \qquad (4.42)$$

$$\kappa_1(i) = 0, \qquad (4.43)$$

where $\kappa_t(i)$ is another auxiliary variable whose role becomes clear in the second step.

2. Recursion:

$$\nu_t(j) = \max_{1 \leq i \leq m} [\nu_{t-1}(i)\gamma_{ij}] \cdot p_j(x_t), \qquad 1 \leq j \leq m,\ 2 \leq t \leq T, \qquad (4.44)$$

$$\kappa_t(j) = \operatorname*{argmax}_{1 \leq j \leq m}[\nu_{t-1}(i)\gamma_{ij}], \qquad 1 \leq j \leq m,\ 2 \leq t \leq T, \qquad (4.45)$$

where $\kappa_t(j)$ yields the state i at time $t-1$ which is most likely to lead to state j at time t.

3. Termination: The first recursion is terminated by maximizing $\nu_T(i)$ and storing the likelihood of the most probable path as a vector $\boldsymbol{p}^*$. The most probable state at time T is stored in i_T^*:

$$\boldsymbol{p}^* = \max_{1 \leq i \leq m} [\nu_T(i)], \qquad (4.46)$$

$$i_T^* = \operatorname*{argmax}_{1 \leq i \leq m}[\nu_T(i)]. \qquad (4.47)$$

4. Path backtracking: Starting with the most probable last state i_T^*, the most likely path can be computed by tracking back from $t = T - 1$ to $t = 1$ in a second recursion using κ_t:

$$i_t^* = \kappa_{t+1} i_{t+1}^*, \qquad t = T - 1, T - 2, \ldots, 1. \qquad (4.48)$$

The resulting optimal state sequence is referred to as *Viterbi path*. In order to avoid numerical underflow, the Viterbi algorithm can alternatively be stated in terms of logarithms.

4.5   Model selection and validation

A problem that naturally arises in connection with HMMs is the determination of the number of discrete states. A common approach to select an appropriate model is based on the information criteria presented in §3.6.2.3. According to the proposed AIC and BIC, the specification that yields the smallest relative information criterion is chosen.

While an information criterion allows for the selection of the best specification among various fitted models, it cannot guarantee appropriateness of the selected model. This can only be checked by assessing the model's fit. However, while it is straightforward to derive model diagnostics based on generalized recursive and generalized least squares residuals in the context of continuous state space models (cf. §3.6), assessing the fit of HMMs is more complex. As under a hidden Markov model the observations that are related to different states are produced by different distributions, the residuals should also be modeled by different distributions. The assumption of independently and identically distributed errors, which is commonly made in diagnostic testing procedures, does not even hold approximately for the residuals of a hidden Markov model (cf. Zucchini et al. 2006, §6.2). A possible solution is to employ so-called *pseudo residuals*, which allow for a comparison of observations induced by different distributions. As the concept of pseudo residuals is beyond the scope of this thesis, it will not find any consideration hereafter; for a comprehensive introduction to the subject, see, for example, Stadie (2002). Rather than relying on standard diagnostic tests, in the empirical part below the relative appropriateness of a hidden Markov model will be formally assessed based on its respective in- and out-of-sample forecast performance.

Chapter 5

Conditional heteroskedasticity models

It is a well established stylized fact of financial time series that the volatility of returns on financial assets changes persistently over time and across assets (cf. §2.2.2). The concept of conditional heteroskedasticity is key to many areas in finance and financial econometrics, whether it is in asset allocation, risk management, asset pricing or duration modeling (cf. Diebold 2004). In the context of this thesis, conditional heteroskedasticity models are mainly used to model time-varying relationships in an indirect way. As a simple regression coefficient is defined as a covariance-variance ratio, models of conditional heteroskedasticity can be used to obtain conditional betas based on conditional variance estimates.

The phenomenon of time-varying volatility, first recognized by Mandelbrot (1963) and Fama (1965), is commonly referred to as *volatility clustering*: quiet periods with small absolute returns are followed by volatile periods with large absolute returns, then quite periods again, and so on. This chapter presents the basic ideas of the two most important concepts of capturing time-varying volatility and excess kurtosis, which is usually induced by volatility clusters. The ARCH/ GARCH framework by Engle (1982) and Bollerslev (1986) represents practitioners' preferred tool to model and forecast volatility. An alternative way of modeling conditional heteroskedasticity is offered by the class of stochastic volatility (SV) models, introduced by Taylor (1982, 1986). In contrast to a GARCH model, the SV model adds an unobserved shock to the return variance. This leads to a higher degree of flexibility in characterizing the dynamics related to volatility.[11] The major difference between these two approaches is that ARCH models are observation-driven, while SV models are parameter-driven. In the context of an ARCH model, the conditional variance is a function of lagged observations, with the corresponding likelihood function being directly available. For SV models, the conditional variance depends on a latent component. As no analytic one-step ahead forecast densities are available, SV models can only be dealt with approximately or by employing numerically

---

[11]In addition to these two concepts of historical volatility, alternative measurements such as implied and realized volatility are available. As these procedures are based on option pricing data and high-frequency data for single stocks, respectively, they will not be considered in this thesis focusing on sectors. For details and an account on the available literature, see, for example, Andersen et al. (2002), Koopman et al. (2004) or Andersen et al. (2005).

intensive procedures. This is the major reason why practitioners prefer ARCH models to model time-varying volatilities in financial markets. However, the progress made with respect to the estimation of latent variable models over the past ten years increased the relative competitiveness of SV models. Yu (2002), for example, compares the ability of SV models to that of alternative ARCH-type models to predict stock price volatility and concludes that the SV model outperforms its competitors. Major advantages of the SV model, which have allowed the concept to grow up as a viable alternative to the modeling of conditional volatility, are:

- The ability to capture the stylized facts of excess kurtosis and leverage more naturally than a GARCH model.

- The provision of both filtered and smoothed estimates of conditional volatility.

- A treatment in continuous time, which is essential in mathematical finance and modern option pricing theory (not to be explored here).

Explicit comparisons of the basic ideas, estimation and inference issues related to ARCH and SV models are provided, for example, by Danielsson (1994), Jacquier et al. (1994), Pagan (1996), Shephard (1996) and Andersen et al. (2005).

This chapter is structured as follows. Section 5.1 introduces the basic GARCH framework. The concept of SV is the subject of Section 5.2. Alternative estimation methods with a focus on efficient Monte Carlo likelihood estimation are discussed. The chapter closes with a brief presentation of multivariate conditional heteroskedasticity models in Section 5.3.

5.1 Autoregressive conditional heteroskedasticity

The aim of this section is to provide a summary of the basic theory of GARCH models as a prerequisite for subsequent analyses. For more details and extensions of the standard GARCH model, the reader is referred to one of the excellent surveys that have been made available over the last fifteen years. For example, Bollerslev et al. (1992) give an overview of the numerous empirical applications related to finance; Bera and Higgins (1993) comprehensively treat many of the GARCH extensions; Bollerslev et al. (1994) evaluate the most important theoretical aspects; Palm (1996) gives a survey on GARCH modeling related to finance including multivariate models. More recent articles include Andersen and Bollerslev (1998), Engle (2001b) and Diebold (2004). A collection of the most influential papers on ARCH/ GARCH is presented by Engle (1995).

Subsection 5.1.1 introduces the basic univariate GARCH representation, summarizes the statistical properties of the model and shows how forecasts of conditional volatility can be produced. Subsection 5.1.2 looks at the two most influential nonlinear extensions of the GARCH model and demonstrates how to test for asymmetric effects. Subsection 5.1.3 discusses how to deal with non-Gaussianity. Subsection 5.1.4 closes with a summary of parameter estimation.

5.1.1 The GARCH(p, q) model

For an observed univariate time series y_t and a given information set Ω_{t-1}, which contains the relevant information through period $t - 1$, the functional form of y_t can be defined as

$$y_t = E(y_t|\Omega_{t-1}) + \epsilon_t, \tag{5.1}$$

with $E(\epsilon_t) = 0$ and $E(\epsilon_t \epsilon_s) = 0$ for all $t \neq s$. The conditional mean $E(y_t|\Omega_{t-1})$ represents the predictable part of y_t; the disturbance term ϵ_t represents the unpredictable part of y_t. In contrast to the standard regression model, in which ϵ_t is assumed to be unconditionally and conditionally homoskedastic, i.e., $E\left(\epsilon_t^2\right) = E(\epsilon_t^2|\Omega_{t-1}) = \sigma^2$ for all t, the conditional variance of ϵ_t is allowed to vary over time:

$$E(\epsilon_t^2|\Omega_{t-1}) = h_t, \tag{5.2}$$

with nonnegative function $h_t := h_t(\Omega_{t-1})$. A general expression for the conditional heteroskedasticity of ϵ_t is given by

$$\epsilon_t = z_t \sqrt{h_t}, \tag{5.3}$$

where z_t is an IID process with $E(z_t) = 0$ and $Var(z_t) = 1$. For now, z_t is assumed to be normally distributed. Taking (5.3) and the assumed properties of z_t together, it follows that ϵ_t is conditionally normal distributed with mean zero and variance h_t. With the unconditional variance of ϵ_t being assumed to be constant, it can be shown that the unconditional expectation of h_t is itself constant with $\sigma^2 := E\left(\epsilon_t^2\right) = E(E(\epsilon_t^2|\Omega_{t-1})) = E(h_t)$; cf. Franses and van Dijk (2000, §4) who, unless mentioned otherwise, serve as a basis for the outline on GARCH modeling in this section.

The model is completed by finding a specification for the behavior of h_t over time. In the basic ARCH(q) model by Engle (1982), the conditional variance of the error term at time t depends on the realized values of past squared disturbance terms:

$$h_t = \omega + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}^2, \tag{5.4}$$

Nonnegativeness and stationarity of h_t are guaranteed for $\omega > 0$, $\gamma_i \geq 0$ for $i = 1, \ldots, q$, and $\sum_{i=1}^{q} \gamma_i < 1$. With $\forall \gamma_i = 0$ the conditional variance reduces to a constant. For the case with $q = 1$, it becomes obvious that the ARCH model is able to capture volatility clustering: as h_t is a function of the previous squared shock, large shocks of either sign tend to be precedented by large shocks and vice versa. It can be shown that the ARCH(1) model captures the excess kurtosis usually inherent to financial time series.

While the ARCH(1) model captures the stylized facts of volatility clustering and excess kurtosis, it is unlikely that the model also accommodates for the characteristic features related to the autocorrelation function of squared disturbances ϵ_t^2. As illustrated by Figure 2.3, the first-order autocorrelation function of squared weekly returns is not very high, which implies a low value for γ_1. At the same time, the autocorrelation function is observed to be persistent, hence demanding a value of γ_1 that is close to unity. One possibility to cope with the persistent autocorrelation, is to add additional

lagged squared shocks to the function of the conditional variance. However, allowing q to become large results in many parameters to be estimated. This would be infeasible given the positiveness and stationarity conditions that have to be imposed. In many cases, an ARCH(q) process is not capable of capturing both the height and shape dimensions of the autocorrelation function; further generalization is required (cf. Pagan 1996).

By including a lagged conditional variance term to the conditional variance function, Bollerslev (1986) generalized the ARCH(q) model. The GARCH(p, q) class of models allows for a more flexible lag structure and for longer memory effects:

$$
\begin{aligned}
h_t &= \omega + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \delta_j h_{t-j} \\
&= \omega + \gamma(L)\epsilon_t^2 + \delta(L)h_t,
\end{aligned}
\tag{5.5}
$$

with $\omega > 0$ and L denoting the *lag operator*, where $\gamma(L) = \gamma_1 L + \cdots + \gamma_q L^q$ and $\delta(L) = \delta_1 L + \cdots + \delta_p L^p$. For $\delta_j = 0$ and $j = 1, \ldots, p$, the GARCH(p, q) process reduces to an ARCH(q) model. Following Bollerslev et al. (1994) the process is well-defined if all parameters in the infinite-order autoregressive representation

$$
h_t = (1 - \delta(L))^{-1}\gamma(L)\epsilon_t^2
\tag{5.6}
$$

are nonnegative. It is assumed that all the roots of $1 - \delta(L)$ lie outside the unit circle. The condition for covariance-stationary of the GARCH(p, q) process is $\sum_{i=1}^{q} \gamma_i + \sum_{j=1}^{p} \delta_j < 1$.

5.1.1.1   Statistical properties

A specification with $p = q = 1$ represents the simplest and most common GARCH model. As it is found to be sufficient for most empirical applications (cf. Bollerslev et al. 1992), only the GARCH(1,1) process will be considered in the following. The GARCH(1,1) model corresponds to an ARCH(∞) process, in which the conditional variance is guaranteed to be nonnegative with $\omega > 0$, $\gamma_1 > 0$ and $\delta_1 \geq 0$. Under the assumption of covariance-stationarity, which is assured for $\gamma_1 + \delta_1 < 1$, the unconditional variance of ϵ_t is defined. It can be calculated as

$$
\sigma^2 = \frac{\omega}{1 - \gamma_1 - \delta_1}.
\tag{5.7}
$$

If $(\gamma_1 + \delta_1)^2 + 2\gamma_1^2 < 1$, the unconditional fourth moment of ϵ_t exists. For normally distributed z_t, the kurtosis of ϵ_t is defined as

$$
K_\epsilon = \frac{3\left(1 - (\gamma_1 + \delta_1)^2\right)}{1 - (\gamma_1 + \delta_1)^2 - 2\gamma_1^2}.
\tag{5.8}
$$

The kurtosis of ϵ_t is always greater than three. It captures some of the excess kurtosis usually inherent to financial time series.

When the GARCH(1,1) model is applied to high-frequency data, a common finding is that current information has an impact on conditional variance forecasts for any horizon. This leads to a sum of γ_1 and δ_1 that is close or equal to unity. Engle and Bollerslev

(1986) refer to the class of models with $\gamma_1 + \delta_1 = 1$ as *integrated GARCH* (IGARCH). For IGARCH models the finite unconditional variance is not defined. Although the IGARCH model is not covariance-stationary, it can be shown that it is strictly stationary. Standard inference procedures remain asymptotically valid given that the sample size is large; cf. Bollerslev et al. (1992) where further references can be found.

To understand why the GARCH(1,1) process is qualified to model the stylized facts of a small first-order autocorrelation and a slow decay simultaneously, we have a look at the autocorrelations of ϵ_t^2:

$$\rho_1 = \gamma_1 + \frac{\gamma_1^2 \delta_1}{1 - 2\gamma_1\delta_1 - \delta_1^2}, \tag{5.9}$$

$$\rho_\tau = (\gamma_1 + \delta_1)^{\tau-1}\rho_1, \qquad \text{for } \tau = 2, 3, \dots . \tag{5.10}$$

With the decay factor of the exponentially declining autocorrelations being equal to $\gamma_1 + \delta_1$, the autocorrelations will decrease the more gradually the closer this sum is to unity (cf. Franses and van Dijk 2000, §4.1.1).

5.1.1.2 Forecasting

With respect to forecasts of the conditional variance, it follows from (5.5) that the optimal l-step ahead forecast can be calculated directly from h_{t+1}, which is part of the information set Ω_t:

$$\hat{h}_{t+l|t} = \omega + \gamma_1\hat{\epsilon}_{t+l-1|t}^2 + \delta_1\hat{h}_{t+l-1|t}^2. \tag{5.11}$$

For a covariance-stationary GARCH(1,1) process, it can be shown that this is equivalent to

$$\hat{h}_{t+l|t} = \sigma^2 + (\gamma_1 + \delta_1)^{l-1}\left(h_{t+1} - \sigma^2\right), \tag{5.12}$$

where the forecast reverts to σ^2 at an exponential rate (cf. Andersen et al. 2005).

5.1.2 Nonlinear extensions

The basic GARCH model is based on the assumption that positive and negative past shocks have the same effect on the conditional variance. However, many financial time series are *asymmetric*: negative shocks tend to have a bigger influence on future volatility than equally sized positive shocks. Asymmetric effects are often observed for aggregate equity indices (cf. Andersen et al. 2005), which will be in the focus of this thesis. This asymmetry, first observed by Black (1976), is commonly referred to as *leverage effect*. According to its definition, the fall of the value of equity amounts in an increased debt-to-equity ratio, the leverage. This implies an increased level of riskiness, which results in increased future volatility.[12] In the standard GARCH model, the conditional variance does not depend on the sign of the shocks such that asymmetries cannot be accommodated. Over the last fifteen years various nonlinear extensions of the GARCH model

---

[12]The *volatility-feedback hypothesis* by Campbell and Hentschel (1992) represents an alternative but less regarded explanation for volatility asymmetries, according to which positive volatility shocks lead to lower future returns.

have been proposed to capture asymmetric effects. Below, only the two most influential nonlinear extensions are discussed. For a more comprehensive overview of the broad spectrum of nonlinear GARCH specifications, see, for example, Hentschel (1995) or Franses and van Dijk (2000, §4.1.2).

5.1.2.1 Exponential GARCH

The first GARCH extension, in which the conditional volatility depends on both the size as well as the sign of lagged shocks, was proposed Nelson (1991). In its simplest specification with $p = q = 1$, the *exponential GARCH* (EGARCH) model can be written as

$$\log(h_t) = \omega + \gamma_1 z_{t-1} + \vartheta_1(|z_{t-1}| - E(|z_{t-1}|)) + \delta_1 \log(h_{t-1}). \qquad (5.13)$$

Due to the modeling of h_t in logarithms, no restrictions on the coefficients have to be imposed to ensure nonnegativity of the conditional volatility. Let the function $g(z_t)$ be defined as $g(z_t) := \gamma z_t + \vartheta_1(|z_t| - E(|z_t|))$ where the two summands relate to the sign and to the magnitude effect. By rewriting it as

$$g(z_t) = (\gamma_1 + \vartheta_1)z_t I(z_t > 0) + (\gamma_1 - \vartheta_1)z_t I(z_t < 0) - \vartheta_1(E|z_1|), \qquad (5.14)$$

with $I(\cdot)$ being an indicator function, it can be seen how asymmetric effects are incorporated: while the term $(\gamma_1 + \vartheta_1)$ is affected by positive shocks, negative shocks have an impact on $(\gamma_1 - \vartheta_1)$. Generally, positive shocks have a smaller effect on h_t than negative shocks of equally sized positive shocks (cf. Engle and Ng 1993). This becomes clear by having a look at the *news impact curve* (NIC)[13] for the EGARCH model, which is given by

$$NIC\left(\epsilon_t | h_t = \sigma^2\right) = \begin{cases} A \exp\left[\dfrac{(\gamma_1 + \vartheta_1)}{\sigma}\epsilon_t\right], & \text{for } \epsilon_t > 0, \\[2mm] A \exp\left[\dfrac{(\gamma_1 - \vartheta_1)}{\sigma}\epsilon_t\right], & \text{for } \epsilon_t < 0, \end{cases} \qquad (5.15)$$

with $A = \sigma^{2\delta_1} \exp(\omega - \vartheta_1\sqrt{2/\pi})$, for parameter constellation $\gamma_1 < 0$, $0 \leq \vartheta < 1$ and $\vartheta_1 + \delta_1 < 1$. As the EGARCH model is not differentiable with respect to z_{t-1} at zero, its estimation is more difficult than that of alternative asymmetric models. Another problem is related to forecasting. Usually, the researcher is interested in forecasting h_{t+l} and not $\log h_{t+l}$. This requires a transformation that depends on the complete l-step ahead forecast distribution, $f(y_{t+l}|\Omega_t)$, which is generally not available in closed-form.

5.1.2.2 GJR-GARCH

Glosten et al. (1993) and Zakoian (1994) independently introduced an alternative nonlinear extension to capture asymmetric effects: the *GJR-GARCH* or *threshold GARCH* model. According to Ling and McAleer (2002), the GJR-GARCH model represents the

---

[13]Engle and Ng (1993) introduced the NIC, defined as the functional relationship between the conditional variance and lagged shocks, as a measure of how the arrival of new information is reflected in volatility estimates. The NIC can be used to compare different GARCH models.

most widely used nonlinear GARCH specification. In the GJR-GARCH(1,1) model, the impact of lagged squared shocks is made dependent on the sign of ϵ_{t-1} by augmenting the GARCH(1,1) model with an additional ARCH term:

$$h_t = \omega + \gamma_1 \epsilon_{t-1}^2 + \vartheta_1 \epsilon_{t-1}^2 I(\epsilon_{t-1} < 0) + \delta_1 h_{t-1}. \qquad (5.16)$$

It can be seen immediately that negative lagged shocks have a bigger influence on h_t than positive shocks. For symmetrically distributed z_t, the process is covariance-stationary if $\gamma_1 + \delta_1 + \frac{1}{2}\vartheta_1 < 1$. The condition for the existence of the fourth moment is

$$\delta_1^2 + 2\delta_1\gamma_1 + K_z\gamma_1^2 + \delta_1\vartheta_1 + K_z\gamma_1\vartheta_1 + \frac{K_z}{2}\vartheta_1^2 < 1, \qquad (5.17)$$

where K_z denotes the kurtosis of z_t; cf. Ling and McAleer (2002) who derive conditions for the existence of moments and strict stationarity for the family of GARCH models. If (5.17) holds, the kurtosis of the unconditional distribution of ϵ_t can be calculated as

$$K_\epsilon = K_z \frac{1 - \left(\delta_1^2 + 2\gamma_1\delta_1 + \delta_1\vartheta_1 + \gamma_1\vartheta_1 + \gamma_1^2 + (1/4)\vartheta_1^2\right)}{1 - \left(\delta_1^2 + 2\gamma_1\delta_1 + \delta_1\vartheta_1 + K_z\gamma_1\vartheta_1 + K_z\gamma_1^2 + (K_z/4)\vartheta_1^2\right)}, \qquad (5.18)$$

cf. Verhoeven and McAleer (2003).

Assuming that the mean of z_t is zero, multi-period point forecasts of the conditional volatility can be implemented analogously to the standard GARCH model (5.12). The l-step ahead forecast is given by

$$\hat{h}_{t+l|t} = \sigma^2 + \left(\gamma_1 + \frac{1}{2}\vartheta_1 + \delta_1\right)^{l-1} \left(h_{t+1} - \sigma^2\right), \qquad (5.19)$$

where the unconditional variance is defined as $\sigma^2 = \omega(1 - \gamma_1 - \frac{1}{2}\vartheta_1 - \delta_1)^{-1}$ (cf. Andersen et al. 2005).

In a comparison of both nonlinear methods, Engle and Ng (1993) conclude that the GJR-GARCH model represents the better model. Although the corresponding NICs are quite similar, the variability of h_t as implied by the EGARCH model is too high. Taking this recommendation and the computational difficulties related to the EGARCH model into account, the GJR-GARCH method will be the preferred choice throughout this thesis whenever the explicit consideration of asymmetric GARCH effects seems to be appropriate.

5.1.2.3 Testing for asymmetric effects

Various diagnostic tests to check whether positive and negative shocks of the same magnitude have different effects on the conditional volatility have been proposed. For example, Sentana (1995) and Hagerud (1997) discuss test procedures based on the Lagrange multiplier principle. Alternatively, Engle and Ng (1993) present the *sign bias* (SB) test, the *negative sign bias* (NSB) test, the *positive sign bias* (PSB) test and a general test. All these tests yield the advantage of being directly applicable to the raw data series, without having to assume a specific volatility model.

The SB test checks whether the squared current residual depends on the sign of lagged residuals. It is defined as the t-statistic for the coefficient β_1 in the following regression:

$$\epsilon_t^{*2} = \beta_0 + \beta_1 S_{t-1}^- + \eta_t, \tag{5.20}$$

with $\epsilon_t := y_t - \mu$ and $\epsilon_t^* := \epsilon_t/\sigma$, where μ and σ refer to the unconditional first two moments of y_t, the series to be considered. S_{t-1}^- represents a dummy variable that takes the value of unity for negative ϵ_t and zero otherwise. Replacing S_{t-1}^- in (5.20) by $S_{t-1}^- \epsilon_{t-1}^*$, or by $S_{t-1}^+ \epsilon_{t-1}^*$ with $S_{t-1}^+ = 1 - S_{t-1}^-$, leads to the NSB test and PSB test, respectively. They test whether the conditional volatility depends on the size of past negative or positive shocks. As the corresponding test statistics are all t-ratios, they asymptotically follow the standard normal distribution.

A general test for nonlinear GARCH can be derived by conducting these tests jointly:

$$\epsilon_t^{*2} = \beta_0 + \beta_1 S_{t-1}^- + \beta_2 S_{t-1}^- \epsilon_{t-1}^* + \beta_3 S_{t-1}^+ \epsilon_{t-1}^* + \eta_t, \tag{5.21}$$

where the null hypothesis is defined as $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. The test-statistic is equal to T times the R^2 from this regression. It asymptotically follows a χ^2 distribution with three degrees of freedom.

5.1.3 Non-Gaussian conditional densities

Even though GARCH models have thicker than normal tails, it is still possible that the kurtosis induced by a GARCH model with conditional normal errors does not capture the leptokurtosis present in high-frequency financial time series completely (cf. Bollerslev et al. 1994). If not taken explicitly into account, this leads to a formally misspecified likelihood function. However, as long as $E(z_t|\Omega_{t-1}) = 0$ and $Var(z_t|\Omega_{t-1}) = 1$, the future conditional variance does not depend on the distribution of z_t. Under these circumstances, QML estimation still yields asymptotically valid volatility forecasts without having the distribution of z_t fully specified (cf. Andersen et al. 2005).

Alternatively, instead of relying on QML based procedures, fully efficient ML estimates can be obtained by considering alternative distributions with fatter than normal tails for z_t. The most common choice is to employ a standardized Student-t distribution with ν degrees of freedom as proposed by Bollerslev (1987). The t-distribution is symmetric around zero; the number of existing moments is restricted by ν. For the standard GARCH(1,1) model, the fourth moment of z_t exists for $\nu > 4$. It is given by

$$K_z = \frac{3(\nu - 2)}{\nu - 4}. \tag{5.22}$$

The value of K_z is greater than the normal value of three. This results in an unconditional kurtosis of ϵ_t that is also greater than in the normal case. The number of degrees of freedom is treated as an extra parameter and can be estimated together with the other model parameters. In case of the nonlinear GJR-GARCH(1,1) model with $z_t \sim t(\nu)$ and $\nu \geq 5$, and K_ϵ and K_z being defined as in (5.18) and (5.22), respectively, the fourth moment exists if (5.17) holds (cf. Ling and McAleer 2002).

To account for the observed leptokurtosis, several other parametric distributions have been suggested. Examples include the normal-Poisson mixture distribution (Jorion 1988)

or the generalized error distribution (Nelson 1991). As none of these will find consideration in the empirical part of this thesis, the reader is referred to Bollerslev et al. (1992) for a summary and an account of the relevant literature.

5.1.4   Parameter estimation

As a wide range of GARCH specifications can be estimated by standard econometric software packages, parameter estimation will only be briefly summarized. Throughout this thesis, all the computations related to GARCH models are carried out using `Ox 3.30` by Doornik (2001) together with the package `G@RCH 2.3` by Laurent and Peters (2002).

Following Bollerslev et al. (1994) GARCH models are usually estimated by ML, where the assumption of an IID distribution for z_t is made. It follows from (5.3) that $z_t(\boldsymbol{\theta}) := \epsilon(\boldsymbol{\theta}) h_t(\boldsymbol{\theta})^{(-1/2)}$, where the conditional mean and variance functions depend on the finite dimensional vector $\boldsymbol{\theta}$ with true value $\boldsymbol{\theta}_0$. Let $f(z_t; \boldsymbol{\nu})$ denote the conditional density function for the standardized innovations with mean zero, variance unity and nuisance parameters $\boldsymbol{\nu}$. Let $\boldsymbol{\psi}' := (\boldsymbol{\theta}', \boldsymbol{\nu}')$ be the combined vector of the parameters to be estimated. The conditional loglikelihood function for the t-th observation can be expressed as

$$l_t(y_t; \boldsymbol{\psi}) = \log f(z_t(\boldsymbol{\theta}); \boldsymbol{\nu}) - \frac{1}{2} \log h_t(\boldsymbol{\theta}), \qquad t = 1, \ldots, T. \tag{5.23}$$

The second term on the right hand side appears, because h_t depends on the unknown parameters in the conditional mean for y_t (cf. Franses and van Dijk 2000, §4.3.1). Once an explicit assumption for the conditional density in (5.23) has been made, the ML estimator for the true parameters $\boldsymbol{\psi}_0$, denoted as $\hat{\boldsymbol{\psi}}_{ML}$, can be calculated by maximizing the loglikelihood for the full sample:

$$\log L(y_T, y_{T-1}, \ldots, y_1; \boldsymbol{\psi}) = \sum_{t=1}^{T} l_t(y_t; \boldsymbol{\psi}), \tag{5.24}$$

where $y_T, y_{T-1}, \ldots, y_1$ refer to the sample realizations of the GARCH model.

As the first-order condition to be solved is nonlinear in the parameters, $\hat{\boldsymbol{\psi}}_{ML}$ is obtained by employing iterative optimization procedures. Following Bollerslev (1986) the most popular procedure to estimate GARCH models is the algorithm named after Berndt et al. (1974). Although convergence may fail in specifications with many parameters, usually no convergence problems arise in connection with univariate GARCH models and large data sets (cf. Alexander 2001, §4.3).

5.2   Stochastic volatility

This section introduces the concept of stochastic volatility (SV). In contrast to the class of GARCH models, the SV approach includes an unobservable shock to the return variance, which cannot be characterized explicitly based on observable past information. This raises the difficulty that no closed expression for the likelihood function exists. As the parameters of the SV model cannot be estimated by a direct application of standard maximum likelihood techniques, estimation is conducted by approximation or

via simulation-based techniques. In this regard, the SV model turns out to be strongly connected to the state space framework outlined in Chapter 3.

Subsection 5.2.1 introduces a representation for the univariate case, deals with the difficulties related to the estimation of SV models and summarizes the basic statistical properties of the model. Subection 5.2.2 gives an overview of the major procedures for statistical inference discussed in the literature. In Subection 5.2.3, the main ideas behind Monte Carlo likelihood, the estimation framework that will be used throughout this thesis for estimation purposes, will be highlighted. The final subsection contains a brief overview of possible extensions of the basic SV model.

5.2.1   The basic stochastic volatility model

Following Ghysels et al. (1996) the basic SV models in discrete time for a univariate time series y_t of asset log-returns can be written as:

$$y_t = \mu + \sigma_* \exp\left(\frac{1}{2}h_t\right)\epsilon_t, \qquad \epsilon_t \sim \text{IID}(0,1), \tag{5.25}$$

with

$$h_{t+1} = \phi h_t + \eta_t, \qquad \eta_t \sim \text{IID}\left(0,\sigma_\eta^2\right), \qquad h_1 \sim N\left(0,\sigma_\eta^2/(1-\phi^2)\right), \tag{5.26}$$

for $t = 1,\ldots,T$. The parameter μ denotes the unconditional expectation of the return process y_t. The scaling parameter σ_* is the average standard deviation with $\sigma_* > 0$. The unobserved log-volatility process is denoted by $h_t = \log\left(\sigma_t^2\right)$. As in the case of the EGARCH model (cf. §5.1.2) the use of logarithms ensures positivity of σ_t^2. The persistence parameter ϕ with $0 < \phi < 1$, plays a similar role as the sum $\gamma_1 + \delta_1$ for the GARCH(1,1) model. The parameter usually takes on values greater than 0.8. As most financial time series exhibit no pronounced serial dependence in the mean equation, the constant μ will be treated as fixed and set to zero in the following.[14] Although h_t can be modeled by any stationary autoregressive process, it is common to choose a first-order autoregressive process. The disturbances ϵ_t and η_t are Gaussian white noise. They are assumed to be uncorrelated, contemporaneously and at all lags. Franses and van Dijk (2000, §4.1.1) offer a useful interpretation of the two different shocks: ϵ_t represents the contents of new information (good or bad news); η_t reflects the shocks to the newsflow's intensity.

The SV model in (5.25) and (5.26) is commonly referred to as the *log-normal SV model*. It represents a state space model where the observation equation describes the relationship between the univariate vector of observations, $y = (y_1,\ldots,y_T)'$, and the state vector. The latent volatility process $\theta = (h_1,\ldots,h_T)'$ is specified in the state equation, which models the dynamic properties of h_t. As ϵ_t and h_t, which both enter the multiplication in the mean equation, are stochastic, the basic SV model is nonlinear. The Kalman filter as introduced in §3.3 cannot be applied directly for estimation purposes.

---

[14]Alternatively, some authors prefer to use mean-corrected returns $y_t^* := \log(P_t) - \log(P_{i,t-1}) - (1/T)\sum_{i=0}^{T}(\log(P_t) - \log(P_{i,t-1}))$, see, for example, Kim et al. (1998).

5.2.1.1 Linearized representation

Harvey et al. (1994) proposed to linearize the SV model by squaring the returns and taking logarithms:

$$\log y_t^2 = \log \sigma_*^2 + h_t + \log \epsilon_t^2, \tag{5.27}$$

$$h_{t+1} = \phi h_t + \eta_t, \tag{5.28}$$

where the disturbance terms in the transformed model are assumed to be uncorrelated, contemporaneously and at all lags. The standard normality of ϵ_t in the original mean equation (5.25) implies a $\log\left(\chi_1^2\right)$ distribution for $\log \epsilon_t^2$ with mean -1.27 and variance $\pi^2/2 = 4.93$. Taking logarithms of very small numbers leads to a heavily skewed distribution of $\log \epsilon_t^2$ with a long left-hand tail. Another important issue of practical relevance is the so-called *inlier problem*, which arises when taking logarithms of returns that are equal to zero. In case where zero returns go back to irregular observations, deletion of these observations avoids the inlier problem. For those inliers that cannot be assumed to result from data irregularities, Sandmann and Koopman (1998) describe various ways of dealing with them. They recommend to cut off any inliers by replacing zeros by the value 0.001.

Unfortunately, due to the non-Gaussianity of the observation disturbances, the prediction error decomposition in (3.58) does not yield the exact likelihood. Thus, a direct application of the Kalman filter using the QML method introduced in §3.3.7, only yields minimum mean square linear estimators. Even though the QML asymptotic theory is correct, it has been shown that due to the poor approximation of $\log \epsilon_t^2$ by a normal distribution, the corresponding QML estimator has poor small sample properties (Kim et al. 1998). Therefore, the QML method for estimating SV models will not find any consideration hereafter.

5.2.1.2 Statistical properties

According to Shephard (1996) the properties of the SV model as represented by (5.26) and (5.26) can be easily derived. For ϕ being restricted to be positive and smaller than unity, the standard Gaussian autoregression h_t will be strictly covariance stationary and follow a log-normal distribution with

$$E(h_t) = 0, \tag{5.29}$$

$$Var(h_t) = \frac{\sigma_\eta^2}{1 - \phi^2}. \tag{5.30}$$

Stationarity of y_t holds if and only if ϵ_t and h_t are stationary processes. Given the properties of the log-normal distribution, for stationary h_t all odd moments are zero. All even moments exist and are given by

$$\begin{aligned} E(y_t^r) &= E(\epsilon_t^r)E\left(\exp\left(\frac{r}{2}h_t\right)\right) \\ &= r! \exp\left(\frac{r^2 \sigma_h^2}{8}\right) / \left(2^{r/2}\left(\frac{r}{2}\right)!\right), \end{aligned} \tag{5.31}$$

where σ_h^2 denotes the variance of h_t, with r referring to the r-th moment. If the fourth moment of ϵ_t exists, it follows from (5.31) that the kurtosis of y_t can be computed as

$$K_y = \frac{E\left(y_t^4\right)}{E\left(y_t^2\right)^2} = 3\exp\left(\sigma_h^2\right), \tag{5.32}$$

which implies fatter than normal tails with $K_y > K_\epsilon$. As discussed in §5.1.3, a GARCH(1,1) model with conditional normal errors typically requires an alternative distribution with fatter tails to capture the excess kurtosis inherent to financial time series completely. The basic SV model does so naturally as $3\exp\left(\sigma_h^2\right)$ can take on any value.

The dynamic properties of y_t^2 can be characterized by the autocorrelations to be derived as

$$\rho_{y_t^2}(\tau) = \frac{\exp\left(\sigma_h^2\phi^\tau\right) - 1}{3\exp\left(\sigma_h^2\right) - 1} \approx \frac{\exp\left(\sigma_h^2\right) - 1}{3\exp\left(\sigma_h^2\right) - 1}\phi^\tau. \tag{5.33}$$

As in case of a GARCH(1,1) model, where the autocorrelation function is given by (5.9) and (5.10), the autocorrelations decay exponentially toward zero. For the SV model, the speed of this decay is determined by ϕ. In case of the linearized SV model as represented by (5.27) and (5.28), the autocorrelation function for $\log y_t^2$ is given by

$$\rho_{\log y_t^2}(\tau) = \frac{\phi^2}{1 + 4.93/\sigma_h^2}. \tag{5.34}$$

5.2.2 Alternative estimation procedures

The objective is the simultaneous estimation of the vector of unknown parameters, denoted as $\boldsymbol{\psi} = [\sigma_* \quad \phi \quad \sigma_\eta]'$, and the volatility vector h conditional on y. Various alternative methods have been proposed to estimate SV models. According to Shephard (2005, p. 13) estimation techniques for SV models can be categorized into (i) relatively simple but inefficient estimators, which are based on moments or approximations of the model; and (ii) computationally demanding procedures, which attempt to evaluate the full likelihood function.

5.2.2.1 Methods of moments and quasi maximum likelihood

In his introductory work on SV models, Taylor (1982) employed the *method of moments* for inference purposes. *Generalized methods of moments* (GMM) estimation techniques are based on the idea of utilizing the stationarity properties, which allow the sample moments to converge to their unconditional expectations. GMM was employed to estimate SV models, among others, by Melino and Turnbull (1990), Jacquier et al. (1994), Andersen and Sorensen (1996), Gallant et al. (1997) and van der Sluis (1997). While method of moments estimators are easy to implement, their efficiency is low. Besides, method of moments procedures do not yield an estimate of the conditional volatility directly, which makes an additional form of estimation necessary.

As mentioned above (cf. §5.2.1.1), an alternative estimation procedure, which is also convenient but quite inefficient with poor small sample properties, is the *quasi maximum*

likelihood (QML) approach as proposed by Harvey et al. (1994). It is based on the transformation of the nonlinear Gaussian SV model into a linear non-Gaussian state space model. In comparison to GMM procedures, QML has the attractive feature that it can be easily extended to non-stationary and multivariate settings. Estimates of the parameters as well as filtered and smoothed conditional volatility series can be readily obtained via the Kalman filter.

While the estimation procedures related to GMM and QML are based on finding solutions for analytical parameter functions, in many situations it is impossible to find such analytical expressions. This is where computationally more demanding simulation techniques come into play. They include the *simulated methods of moments* approach by Duffie and Singleton (1993), the *indirect inference* estimator by Gouriéroux et al. (1993) and the *moment matching* procedure by Gallant and Tauchen (1996).

5.2.2.2   Markov chain Monte Carlo

All estimation methods mentioned so far have in common that they are either based on moments or on an approximation of the exact likelihood. Alternatively, inference can be performed by computing the likelihood directly using computationally intensive methods. The most prominent approach to date is the application of *Markov chain Monte Carlo* (MCMC) techniques. Their origins trace back to the statistical physics literature. The first fully efficient parametric inference procedures for SV models based on MCMC have been developed by Shephard (1993) and Jacquier et al. (1994). Kim et al. (1998) discussed alternative simulation-based strategies for actually implementing MCMC to estimate SV models.

Following the outline of Andersen et al. (2005) MCMC can be used to deal with the high-dimensional inference problem inherent to the likelihood of the SV model from a Bayesian perspective. In the Bayesian MCMC approach, the model parameters are considered to be random variables. The entire latent volatility process is treated as an additional parameter. The focus is on the possibly high-dimensional joint distribution of this parameter vector, conditioned by the data, which is referred to as the *posterior distribution*. Although this density is analytically intractable, it can be characterized through a set of related conditional distributions. These allow a single parameter or a whole group of parameters to be expressed conditional on the remaining parameters. This feature is exploited by the MCMC procedure, which can be summarized as follows: after initialization of the vector of parameters conditioned on the observed data, the coefficients are drawn from the assumed prior distribution. The current draw for the parameters is combined with the dynamics of the SV model and the observations. A complete cycle through all conditional densities is called a *sweep* of the sampler. Depending on the form of these distributions, different samplers such as the *Metropolis-Hastings* algorithm or the *Gibbs* sampler or a combination of both can be used. Once the sample generated from the joint posterior distribution is long enough, inference on the parameters and the latent volatility process can be made. For a general introduction to MCMC procedures, see Geman and Geman (1984), West and Harrison (1997, §15) or Chib and Greenberg (1996). The idea of estimating SV models via MCMC procedures is comprehensively illustrated by Jacquier et al. (1994).

The enhancement of MCMC methods for the estimation of SV models and their extensions to cope with non-Gaussian errors and the leverage effect (cf. §5.1.2), is an active field with recent contributions, for example, by Jacquier et al. (2004), Omori et al. (2004) and Yu (2005). Overall, Bayesian estimators have been shown to outperform method of moments and QML approaches with respect to both, computing filtered volatility estimates and the estimation of parameters. On the other hand, they demand a large amount of computationally intensive simulations. Nontrivial modifications are required for certain extensions, such as the incorporation of explanatory variables.

5.2.2.3   Monte Carlo likelihood

These undesirable features of MCMC procedures have led to the second branch of inference that attempts to evaluate the full likelihood: *simulated maximum likelihood*, also referred to as *Monte Carlo likelihood* (MCL). According to Durbin and Koopman (2001, §8.3) MCL techniques are more transparent and computationally more convenient for the estimation of SV models than MCMC methods. As these two features are particularly valuable to practitioners who do not dispose of expert knowledge of simulation techniques, MCL will be preferred over MCMC in this thesis.

Danielsson (1994) was the first to apply MCL to the estimation of SV models. General contributions to the MCL literature were made by Shephard and Pitt (1997) and Durbin and Koopman (1997) who improved computational efficiency by employing *importance sampling* techniques. Sandmann and Koopman (1998) proposed an *efficient MCL* estimator, which was demonstrated to be a veritable alternative to MCMC procedures. They showed that efficient MCL, while offering comparable finite sample properties, is less computationally demanding than MCMC. As it is possible to approximate the likelihood arbitrarily close, inference can be performed by making use of Likelihood Ratio test statistics. Another comparative advantage of the proposed MCL technique is that only trivial modifications have to be imposed to extend the basic SV model to allow for heavy-tailed errors, leverage effects and explanatory variables; see, for example, Sandmann and Koopman (1998). Koopman and Hol-Uspensky (2001) proposed the *SV-in-Mean* model, in which the mean may also be influenced by changes in the conditional volatility. Liesenfeld and Richard (2003) generalized the importance sampling method employed by Danielsson (1994) by making use of the efficient importance sampling procedure proposed by Richard and Zhang (1996). Lee and Koopman (2004) compared the two different importance sampling techniques by considering a generalized SV model with Student-t distributed observation errors.

For a general introduction to SV models and a more detailed discussion of the various estimation techniques, see, for example, Shephard (1996), Ghysels et al. (1996) or Broto and Ruiz (2004). A collection of some of the most important papers on the topic is presented by Shephard (2005).

5.2.3   Efficient Monte Carlo likelihood estimation

The method of MCL will be used throughout this thesis for the estimation of SV models. In contrast to Sandmann and Koopman (1998), who employed MCL combined with the

linearized representation — thus dealing with a non-Gaussian state space model — the basic SV model without transformations will be considered in the following. As a comprehensive treatment of sampling methods would be beyond the scope of this thesis, only the main ideas behind the concepts of MCL and importance sampling will be outlined. For a discussion of importance sampling techniques, the reader is referred to Ripley (1987, §5). A comprehensive treatment of the handling of non-Gaussian and nonlinear state space models using simulation techniques, both from a classical and a Bayesian perspective, is given by Durbin and Koopman (2000) and Durbin and Koopman (2001, Part II). All inferential aspects will be considered from a classical standpoint.

5.2.3.1   The likelihood function

Following Durbin and Koopman (1997), the likelihood of the basic SV model in (5.25) and (5.26) is given by

$$L(\boldsymbol{\psi}) = p(y|\boldsymbol{\psi}) = \int p(y, \theta|\boldsymbol{\psi})d\theta = \int p(y|\theta, \boldsymbol{\psi})p(\theta|\boldsymbol{\psi})d\theta, \qquad (5.35)$$

where $p(y|\boldsymbol{\psi})$, $p(\theta|\boldsymbol{\psi})$, $p(y|\theta, \boldsymbol{\psi})$ and $p(y, \theta|\boldsymbol{\psi})$ denote the marginal densities of y and θ, the conditional density of y given θ for given $\boldsymbol{\psi}$ and the joint density of y and θ given $\boldsymbol{\psi}$, respectively. For the SV model, no analytical solution for this integral is available; direct numerical maximization is not feasible. The idea of employing MCL is to estimate the loglikelihood by simulation and find an estimate for $\boldsymbol{\psi}$ by numerical maximization. In principle, simulation could be used to estimate $L(\boldsymbol{\psi})$ by generating M independent draws $\theta^{(i)}$ from the density $p(\theta|\boldsymbol{\psi})$, for $i = 1, \ldots, M$. In a second step, the likelihood could be estimated by $M^{-1} \sum_{i=1}^{M} p(y|\theta^{(i)})$. However, as most replications $\theta^{(i)}$ would not resemble the process of θ under which y was observed, M would have to become large to gain some accuracy. As a consequence, this procedure turns out to be highly inefficient in practice (Lee and Koopman 2004).

5.2.3.2   Importance sampling

Computational efficiency can be improved by applying importance sampling techniques. A conditional Gaussian approximating density, denoted as $g(\theta|y, \boldsymbol{\psi})$, for which random draws can be obtained is chosen as *importance density*. The approximating density should be as close as possible to $p(\theta|y, \boldsymbol{\psi})$. Sampling from this density and estimating the original likelihood by appropriately adjusting the Gaussian likelihood of the approximating model using a correction factor, is referred to as importance sampling (cf. Durbin and Koopman 2001, §11.1). To simplify the notation, in the following a density's dependency on $\boldsymbol{\psi}$ will be suppressed as all densities depend on $\boldsymbol{\psi}$.

With the Gaussian importance density $g(\theta|y)$ the likelihood for the approximating Gaussian SV model can be rewritten as

$$L_G(\boldsymbol{\psi}) = g(y) = \frac{g(y, \theta)}{g(\theta|y)} = \frac{g(y|\theta)p(\theta)}{g(\theta|y)}. \qquad (5.36)$$

After substitution of (5.36) for $p(\theta)$ in (5.35), the likelihood of the original model can be stated as

$$L(\boldsymbol{\psi}) = L_G(\boldsymbol{\psi}) \int \frac{p(y|\theta)g(\theta|y)}{g(y|\theta)} d\theta$$
$$= L_G(\boldsymbol{\psi}) E_G \left(\frac{p(y|\theta)}{g(y|\theta)} \right), \tag{5.37}$$

where E_G denotes the expectation with respect to $g(\theta|y)$. The *simulation smoother*[15] by Durbin and Koopman (2002) is used to obtain M independent draws $\theta^{(i)}$ from $g(\theta|y)$. The Monte Carlo likelihood of the basic SV model can be calculated as the product of the Gaussian likelihood of the approximating model and a correction factor, which is obtained by simulation:

$$\hat{L}(\boldsymbol{\psi}) = L_G(\boldsymbol{\psi}) \left(\frac{1}{M} \sum_{i=1}^{M} \frac{p\left(y|\theta^{(i)}\right)}{g\left(y|\theta^{(i)}\right)} \right), \tag{5.38}$$

To minimize the required number of draws, $g(\theta|y)$ has to be chosen to be a good approximation to $p(\theta|y)$.

The importance sampling density can be obtained according to the procedure proposed by Shephard and Pitt (1997) and Durbin and Koopman (1997), which approximates the original model by a linear Gaussian state space model. In the approximating model, the state equation is still provided by (5.26), while the observation equation is represented by

$$y_t = h_t + a_t + b_t u_t. \tag{5.39}$$

The standard normally distributed errors are assumed to be uncorrelated with the state disturbances η_t. The location and scaling parameters a_t and b_t are responsible for a good match between the approximating and the original model. The conditional density $g(\theta|y)$ of the approximating model is taken as the importance density. For more details on the selection of the approximating model, see Durbin and Koopman (1997).

Based on the outline of Lee and Koopman (2004) the importance sampling procedure to calculate the likelihood of y given the vector of parameters $\boldsymbol{\psi}$ can be summarized as follows:

1. At the beginning, neither θ, a_t nor b_t are known. The algorithm that solves a_t and b_t analytically for given θ and y is started by choosing a trial value for θ.

2. Taking logarithms of the conditional densities of y given θ, both for the approximating and the original model, and equalizing the first two derivatives of the resulting log-density functions with respect to θ, yields some first estimates of a_t and b_t.

---

[15] According to Durbin and Koopman (2002, p. 603), "a simulation smoother in state space time series analysis is a procedure for drawing samples from the conditional distribution of state or disturbance vectors given the observations." These samples are useful for analyzing non-Gaussian and nonlinear state space models, both from a classical and from a Bayesian perspective.

3. An estimate for the unobserved process θ can be computed from the state space model given by (5.26) and (5.39) with (5.29) and (5.30) by making use of the Kalman smoother as introduced in §3.3.2.

4. Steps #2 and #3 are repeated using the smoothed estimates of θ until convergence of θ, a_t and b_t is achieved.

5. Finally, with the importance density $g\left(\theta^{(i)}|y\right)$ being evaluated using the Kalman smoother, the likelihood given by (5.38) can be estimated with the samples $\theta^{(i)}$ that are drawn from $g\left(\theta^{(i)}|y\right)$ via simulation smoothing.

In order to avoid numerical problems, this is usually done by working with logarithms of the likelihood. Let $w(\theta) = p(y|\theta)/g(y|\theta)$ and $w_i = w\left(\theta^{(i)}\right)$ for $i = 1, \ldots, M$. According to Durbin and Koopman (1997) an approximately unbiased estimator of $\log L(\boldsymbol{\psi})$ is given by

$$\log \hat{L}(\boldsymbol{\psi}) = \log L_G(\boldsymbol{\psi}) + \log \bar{w} + \frac{s_w^2}{2M\bar{w}^2}, \tag{5.40}$$

where $\bar{w} = M^{-1}\sum_{i=1}^{M} w_i$ and $s_w^2 = (M-1)^{-1}\sum_{i=1}^{M}(w_i - \bar{w})^2$ represent the sample mean and variance of w_i, respectively. The last term on the right hand side corrects for the bias that can be shown to arise from taking logarithms of (5.38). According to Durbin and Koopman (2000) the bias is usually so small that it can be neglected in practice. Numerical maximization of the simulated loglikelihood (5.40), using an iterative numerical optimization method as discussed in §3.4.2, yields the MCL estimate of $\boldsymbol{\psi}$. Starting values can be obtained from $\log L_G(y|\boldsymbol{\psi})$. The accuracy of the approximation of the likelihood function depends positively on M, the number of draws. Sandmann and Koopman (1998) showed that $N = 5$ is sufficient in empirical applications.

In this thesis, the SV model is estimated using `Ox 3.30` by Doornik (2001) together with the packages `SsfPack 2.3` and `SsfNonG` by Koopman et al. (1999). The relevant code for implementing the estimation of SV models in `Ox` has been downloaded from `www.feweb.vu.nl/koopman/sv/`.

5.2.3.3   Filtering, smoothing and forecasting

Once estimates for the unknown model parameters have been found as demonstrated above, another interest focuses on obtaining filtered, smoothed and predicted estimates of the conditional volatility process. In contrast to the GARCH model, where filtered estimates of the volatility process can be calculated recursively once the parameters are known, for an SV model the latent volatility has to be estimated. In addition to the filtered volatility also smoothed estimates are available via the Kalman filter and smoother (cf. §3.3); see Sandmann and Koopman (1998) for details.

According to Koopman et al. (2004) the one-step ahead forecast of conditional volatility can be calculated as

$$\hat{\sigma}_{T+1}^2 = \hat{\sigma}_*^2 \exp\left(\hat{h}_{T+1|T} + \frac{1}{2}p_{T+1|T}\right), \tag{5.41}$$

with $\hat{\sigma}_*^2$ denoting the ML estimate of σ_*^2. The estimator of h_{T+1} given all T observations and its MSE, denoted as $\hat{h}_{T+1|T}$ and $p_{T+1|T}$, are given by

$$\hat{h}_{T+1|T} = \frac{\sum_{i=1}^{M} w_i h_{T+1|T} \theta^{(i)}}{\sum_{i=1}^{M} w_i}, \tag{5.42}$$

$$p_{T+1|T} = \frac{\sum_{i=1}^{M} w_i p_{T+1|T} \theta^{(i)}}{\sum_{i=1}^{M} w_i}. \tag{5.43}$$

Based on the draw $\theta^{(i)}$, the Kalman filter is applied to the approximating model $g(\theta|y)$ to obtain $h_{T+1|T}\theta^{(i)}$ and $p_{T+1|T}\theta^{(i)}$.

5.2.4   Extensions

The SV model can be extended in various directions. As in the case of the GARCH framework, the most important extensions cope with alternative error distributions for the conditional mean innovations and leverage effects. An account of recently published papers related to extensions of the basic SV model, applying both MCMC and efficient MCL methods, is given in §5.2.2.

5.2.4.1   Heavy-tailed distributed errors

Empirical evidence of heavy-tailed ϵ_t in the context of SV models has been provided, for example, by Gallant et al. (1997). The efficient MCL method using importance sampling techniques presented in §5.2.3.2, can be adapted to consider Student-t distributed errors by replacing the Gaussian density $p(y|\theta)$ by a Student-t density with ν degrees of freedom. This leads to an altered pair of equations, which can be solved to obtain an approximating model through a_t and b_t. The importance density itself remains Gaussian; see Lee and Koopman (2004, Appendix B) for details.

Even though the basic SV model can be generalized by allowing for Student-t distributed mean errors, throughout this thesis only SV models with Gaussian errors will be considered. According to Ghysels et al. (1996) the SV model in its basic form is able to capture the excess kurtosis usually found in financial time series by considering y_t as a mixture of distributions, where the degree of mixing is governed by the parameter σ_η^2. As outlined in §5.2.1.2, the kurtosis of y_t is equal to $3\exp\left(\sigma_h^2\right)$ which can take any nonnegative value. Besides, the empirical analyses in this thesis are strictly based on weekly series, for which the evidence in favor of fat-tails is not as strong as for daily data (cf. Jacquier et al. 2004).

5.2.4.2   Asymmetric effects

As outlined in §5.1.2, asymmetric effects represent a well documented empirical stylized fact for many financial time series. It has been demonstrated in the previous section how to employ nonlinear GJR-GARCH and EGARCH extensions to specify the conditional volatility as a function of the sign and/ or size of past returns. For discrete time SV models, leverage effects can be implemented by allowing for contemporaneously negatively correlated observation and state disturbances. One of the earliest studies on

SV models with leverage was presented by Harvey and Shephard (1996) where inference is conducted via QML. More recently, asymmetric SV models are estimated using simulation-based methods; see, for example, Sandmann and Koopman (1998), Meyer and Yu (2000) or Jacquier et al. (2004).

Following the outline of Jungbacker and Koopman (2005b) the basic SV model in (5.25) and (5.26) can be extended to account for asymmetric effects by relaxing the assumption of uncorrelated ϵ_t and η_t:

$$\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \qquad |\rho| \leq 1, \tag{5.44}$$

for $t = 1, \ldots, T$ with $\rho := Corr(\epsilon_t, \eta_t)$. After reformulation of the state space model to include ρ, the parameters can be estimated by maximizing the Monte Carlo likelihood function in (5.40). The vector of unknown parameters ψ consists of σ_*, ϕ, σ_η and ρ.

Despite the efforts made to explicitly allow for asymmetric effects in the SV model, the basic SV model with uncorrelated errors has been found to have higher maximum likelihood values than any EGARCH model considered (cf. Danielsson 1994). At the same time, fewer parameters have to be estimated. Chan et al. (2005) note that leverage effects are mainly a feature of market-wide rather than industry-specific returns and volatility. Given these findings, in the subsequent analyses with a focus on industry portfolios, only the basic SV model will be employed.

5.3 Multivariate conditional heteroskedasticity

The classes of ARCH and SV models represent the two major techniques to model conditional volatility in financial markets. It can be assumed that certain news affect the volatility of different assets simultaneously. To make use of potential linkages, in many practical applications it is necessary to adapt the procedures that were discussed above in the context of univariate conditional heteroskedasticity models to a multivariate setting.

Covariances of various financial variables play an important role in financial economics: for example, they are used to estimate market betas, in the context of portfolio optimization, the pricing of derivatives and for hedging purposes. It is a common observation that joint stationarity does not hold even though the individual variables are stationary (cf. Alexander 2001, §1.4). Unconditional covariances between these variables do not exist. Multivariate GARCH and SV models can be employed to model the time-varying behavior of conditional covariances. In a multivariate setting two major obstacles have to be dealt with: (i) the lack of parsimony due the proliferation of the number of parameters in high-dimensional models, and (ii) the search for sufficient conditions to ensure positive definiteness. Multivariate models of conditional heteroskedasticity are a field of ongoing research. Engle (2001a, p. 54) notes in a survey article on the recent financial econometrics literature: "The most significant unsolved problem is the multivariate extension of many of these methods. Although various multivariate GARCH models have been proposed, there is no consensus on simple models that are satisfactory for big problems."

This section summarizes some of the most important multivariate conditional heteroskedasticity models. For those topics and the many extensions that are not covered here, the reader is referred to Bauwens et al. (2003) and Asai et al. (2006) who provide surveys of multivariate GARCH and multivariate SV models, respectively.

5.3.1   Multivariate GARCH

A general multivariate GARCH model for an N-dimensional process $\epsilon_t|\Omega_{t-1}$ is given by

$$\epsilon_t = z_t H_t^{1/2}, \tag{5.45}$$

where z_t is an N-dimensional IID process with zero mean and the identity matrix I_N as covariance matrix. These properties of z_t together with Equation (5.45) imply that $E(\epsilon_t|\Omega_{t-1}) = 0$ and $E(\epsilon_t\epsilon_t'|\Omega_{t-1}) = H_t$. For illustrative purposes, only the case with $N = 2$ will be considered in the following. In the bivariate case, the conditional covariance matrix is given by

$$H_t = \left[\begin{array}{cc} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{array} \right], \tag{5.46}$$

where H_t depends on lagged errors ϵ_{t-1} and on lagged conditional covariance matrices H_{t-1}. The most influential parameterizations of H_t can be summarized as follows.

5.3.1.1   The vech model

The most general representation of H_{t-1} is the *vech model* as proposed by Bollerslev et al. (1988). By employing the vech$(\cdot)$ operator, which vertically stacks the matrix elements on or below the principal diagonal and thus transforms an $N \times N$ matrix into an $N(N+1)/2 \times 1$ vector, all non-redundant elements of H_t are stacked into a column vector:

$$\text{vech}(H_t) = \omega^* + \Gamma^*\text{vech}(\epsilon_{t-1}\epsilon_{t-1}') + \Delta^*\text{vech}(H_{t-1}), \tag{5.47}$$

where $\omega^* = \text{vech}(\Omega)$ is a $N(N+1)/2 \times 1$ parameter vector and Γ^* and Δ^* are $N(N+1)/2 \times N(N+1)/2$ matrices. In the bivariate case, Equation (5.47) becomes

$$\left[\begin{array}{c} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{array} \right] = \left[\begin{array}{c} \omega_{11}^* \\ \omega_{21}^* \\ \omega_{22}^* \end{array} \right] + \left[\begin{array}{ccc} \gamma_{11}^* & \gamma_{12}^* & \gamma_{13}^* \\ \gamma_{21}^* & \gamma_{22}^* & \gamma_{23}^* \\ \gamma_{31}^* & \gamma_{32}^* & \gamma_{33}^* \end{array} \right] \left[\begin{array}{c} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}\epsilon_{1,t-1} \\ \epsilon_{2,t-1}^2 \end{array} \right]$$

$$+ \left[\begin{array}{ccc} \delta_{11}^* & \delta_{12}^* & \delta_{13}^* \\ \delta_{21}^* & \delta_{22}^* & \delta_{23}^* \\ \delta_{31}^* & \delta_{32}^* & \delta_{33}^* \end{array} \right] \left[\begin{array}{c} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{array} \right]. \tag{5.48}$$

Despite its flexibility, the vech model has two major drawbacks: to guarantee positive definiteness of H_t it is necessary to impose further constraints on Γ^* and Δ^*; see Engle and Kroner (1995) for a discussion. Besides, overall $N(N+1)/2 + N^2(N+1)^2/2$ parameters have to be estimated. As this number grows at a polynomial rate with increasing N, estimation of this general model may become quite cumbersome without further restrictions.

5.3.1.2   The diagonal vech model

The *diagonal vech model* is a first way to restrict Equation (5.47) and to reduce the number of parameters. Bollerslev et al. (1988) restrict the matrices $\mathbf{\Gamma}^*$ and $\mathbf{\Delta}^*$ to be diagonal such that the conditional covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ depends only on lagged cross-products of the residuals and its own lagged value. In this specification, each element of the conditional covariance matrix follows a univariate GARCH(1,1) model:

$$h_{ij,t} = \omega_{ij} + \gamma_{ij}\epsilon_{i,t-1}\epsilon_{j,t-1} + \delta_{ij}h_{ij,t-1}, \tag{5.49}$$

where ω_{ij}, γ_{ij} and δ_{ij} denote the ij-th element of the symmetric $N \times N$ matrices $\mathbf{\Omega}$, $\mathbf{\Gamma}$ and $\mathbf{\Delta}$, respectively. The latter two matrices are implicitly defined by $\mathbf{\Gamma}^* = \text{diag}(\text{vech}(\mathbf{\Gamma}))$ and $\mathbf{\Delta}^* = \text{diag}(\text{vech}(\mathbf{\Delta}))$.[16] Following Ding and Engle (2001) the diagonal vech model can be expressed in terms of the Hadamard product,[17] denoted by $\odot$:

$$\boldsymbol{H}_t = \mathbf{\Omega} + \mathbf{\Gamma} \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1} + \mathbf{\Delta} \odot \boldsymbol{H}_{t-1}, \tag{5.50}$$

The number of parameters is reduced to $3N(N + 1)/2$. Using this representation, $\boldsymbol{H}_t$ can be shown to be positive definite for positive definite $\mathbf{\Omega}$ and positive semi-definite $\mathbf{\Gamma}$ and $\mathbf{\Delta}$ (cf. Franses and van Dijk 2000, §4.7).

5.3.1.3   The BEKK model

A more general representation is the *BEKK model* of Engle and Kroner (1995). It includes all positive definite diagonal models and nearly all positive definite vech specifications. The model is named after an earlier version of the paper, which was based on the contributions of Baba, Engle, Kraft and Kroner. The BEKK model elegantly imposes the restrictions of positive definiteness of $\boldsymbol{H}_t$ by (i) decomposing the constant matrix into $\mathbf{\Omega}'\mathbf{\Omega}$, and (ii) by using quadratic forms of $\mathbf{\Gamma}$ and $\mathbf{\Delta}$ instead of imposing restrictions on these matrices. The model is given by

$$\boldsymbol{H}_t = \mathbf{\Omega}'\mathbf{\Omega} + \mathbf{\Gamma}'\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}\mathbf{\Gamma} + \mathbf{\Delta}'\boldsymbol{H}_{t-1}\mathbf{\Delta}, \tag{5.51}$$

where $\mathbf{\Omega}$, $\mathbf{\Gamma}$ and $\mathbf{\Delta}$ are symmetric $N \times N$ matrices with $\mathbf{\Omega}$ being upper triangular. In the bivariate case the BEKK model can be written as

$$\boldsymbol{H}_t = \mathbf{\Omega}'\mathbf{\Omega}' + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}' \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}\epsilon_{1,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$
$$+ \begin{bmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{bmatrix}' \boldsymbol{H}_{t-1} \begin{bmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{bmatrix}. \tag{5.52}$$

The number of parameters equals $N(5N + 1)/2$. For $N = 2$, two more unknowns than in the diagonal vech setting have to be estimated. The number of parameters can be

---

[16]The operation $\text{diag}(\boldsymbol{x})$ with $\boldsymbol{x} = [x_1 \cdots x_N]'$ denotes the $N \times N$ diagonal matrix with the diagonal elements given by $\boldsymbol{x}$.

[17]For two $N \times N$ matrices A and B, the Hadamard product $A \odot B$ is the $N \times N$ matrix that contains the element-by-element products $[a_{ij}b_{ij}]_{i,j=1,\ldots,N}$.

further reduced by imposing diagonality on $\boldsymbol{\Gamma}$ and $\boldsymbol{\delta}$, or by setting these matrices equal to scalars; both results in a loss of generality.

As the number of parameters to be estimated simultaneously still becomes large with increasing N, the resulting computational problems soon become insurmountable. Various approximations to compute multivariate GARCH conditional covariance matrices based on univariate GARCH models have been proposed. Two of the most influential specifications will be summarized below.[18]

5.3.1.4   The constant conditional correlation model

The *constant conditional correlation model* (CCC) by Bollerslev (1990) reduces the computational complexity of the general multivariate GARCH(1,1) model by assuming the conditional correlations between $\epsilon_{i,t}$ and $\epsilon_{j,t}$, denoted by ρ_{ij}, to be constant. The diagonal vech model with constant correlations can be defined as

$$h_{ii,t} = \omega_{ii} + \gamma_{ii}\epsilon_{i,t-1}^2 + \delta_{ii}h_{ii,t-1}, \tag{5.53}$$

$$h_{ij,t} = \rho_{ij}\sqrt{h_{ii,t}}\sqrt{h_{jj,t}}, \qquad\qquad \text{for all } i \neq j. \tag{5.54}$$

Equation (5.53) denotes the conditional variance of the returns of index i; Equation (5.54) is the conditional covariance between the returns of indices i and j. Alternatively, Bollerslev's model can be written as

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t, \tag{5.55}$$

where $\boldsymbol{D}_t$ is an $N \times N$ matrix with the conditional standard deviations $\sqrt{h_{ii,t}}$ on the diagonal; $\boldsymbol{R}$ denotes a symmetric positive definite $N \times N$ matrix that contains the unconditional correlations ρ_{ij}. In the bivariate case, the CCC model becomes

$$\boldsymbol{H}_t = \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{bmatrix}, \tag{5.56}$$

with only seven ($= N(N+5)/2$) parameters left to be estimated. The conditional covariance matrix is guaranteed to be positive definite, provided that $\boldsymbol{R}$ is positive definite and that the conditional variances $h_{ii,t}$ are positive (cf. Franses and van Dijk 2000, §4.7).

The consideration of different functional forms for the GARCH process, for example, allowing for asymmetric effects or non-Gaussian conditional densities, is straightforward: as the CCC model is exclusively based on univariate models, it is only necessary to respecify (5.53) in accordance with the outline presented in §5.1.2 or §5.1.3, respectively.

---

[18]Alternative simplifying approaches to estimate multivariate GARCH models, which will not find any consideration in this thesis, include the *factor multivariate GARCH* model by Diebold and Nerlove (1989) where the conditional covariance structure of the observed variables arises from a joint dependence on common factors; the *orthogonal GARCH* model by Alexander (2000) which is based on univariate GARCH volatilities of a limited number of uncorrelated risk factors obtained from a principal component analysis; and the *flexible multivariate GARCH* model by Ledoit et al. (2003) who propose to estimate univariate models and to transform the estimated parameters such that the conditional covariance matrix is positive definite.

5.3.1.5   The dynamic conditional correlation model

The *dynamic conditional correlation model* (DCC) by Engle (2000) generalizes the CCC model by allowing the conditional correlations to be time-varying:

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t, \tag{5.57}$$

The $N \times N$ matrix $\boldsymbol{R}_t$ contains the conditional correlations, which are estimated as follows: univariate GARCH models are estimated for each of the series in the N-dimensional vector of observations $\boldsymbol{Y}_t$. The resulting conditional standard deviations are used to build $\boldsymbol{D}_t$ and to define the $N \times 1$ vector of standardized residuals:

$$\tilde{\boldsymbol{\epsilon}}_t = \boldsymbol{D}_t^{-1} \boldsymbol{\epsilon}_t. \tag{5.58}$$

These are employed to calculate the $N \times N$ matrix of unconditional variances of $\tilde{\boldsymbol{\epsilon}}_t$:

$$\tilde{\boldsymbol{R}} = \frac{1}{T} \sum_{t=1}^{T} \tilde{\boldsymbol{\epsilon}}_t \tilde{\boldsymbol{\epsilon}}_t', \tag{5.59}$$

which is guaranteed to be positive definite. In the next step, the symmetric $N \times N$ covariance matrix of the standardized residuals, $\boldsymbol{Q}_t$, can be estimated as

$$\boldsymbol{Q}_t = (1 - \alpha - \beta) \tilde{\boldsymbol{R}} + \alpha (\tilde{\boldsymbol{\epsilon}}_{t-1} \tilde{\boldsymbol{\epsilon}}_{t-1}') + \beta \boldsymbol{Q}_{t-1}. \tag{5.60}$$

Finally, the matrix of conditional correlations is obtained by standardization of the elements in $\boldsymbol{Q}_t$:

$$\boldsymbol{R}_t = \text{diag} \left(q_{11,t}^{-1/2} \ldots q_{NN,t}^{-1/2} \right) \boldsymbol{Q}_t \, \text{diag} \left(q_{11,t}^{-1/2} \ldots q_{NN,t}^{-1/2} \right). \tag{5.61}$$

As long as the sum of the nonnegative scalars α and β is smaller than one, this process will be mean-reverting. For $\alpha + \beta = 1$, the process induced by $\boldsymbol{Q}_t$ will be integrated, and (5.60) reduces to an exponential smoothing process:

$$\boldsymbol{Q}_t = (1 - \lambda)(\tilde{\boldsymbol{\epsilon}}_{t-1} \tilde{\boldsymbol{\epsilon}}_{t-1}') + \lambda \boldsymbol{Q}_{t-1}, \tag{5.62}$$

with $0 < \lambda < 1$. According to Engle (2000) this leads to one of the simplest and most successful parameterizations of $\boldsymbol{R}_t$, in which the individual elements can alternatively be estimated as a geometrically weighted average of the standardized residuals:[19]

$$\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}} = \frac{\sum_{o=1}^{t-1} \lambda^o \tilde{\epsilon}_{i,t-o} \tilde{\epsilon}_{j,t-o}}{\sqrt{\sum_{o=1}^{t-1} \lambda^o \tilde{\epsilon}_{i,t-o}^2 \sum_{o=1}^{t-1} \lambda^o \tilde{\epsilon}_{j,t-o}^2}}. \tag{5.63}$$

With respect to forecasting dynamic conditional correlations, Engle (2000) developed the following forecasting expression for an l-step ahead forecast. It is based on an

---

[19]This specification is also used by RiskMetrics™, a well known risk management system developed by J.P. Morgan, with the smoothing parameter λ set to 0.94 for daily data (cf. Riskmetrics Group 1996).

approximation using a first-order Taylor series expansion of the correlation coefficient:

$$\rho_{ij,t+l} \approx \frac{\bar{q}_{ij}}{(\bar{q}_{ii}\bar{q}_{jj})^{1/2}} + \frac{q_{ij,t+l} - \bar{q}_{ij}}{(\bar{q}_{ii}\bar{q}_{jj})^{1/2}}$$
$$- \frac{1}{2}\frac{\bar{q}_{ij}}{(\bar{q}_{ii}\bar{q}_{jj})^{1/2}}\left(\frac{q_{ii,t+l} - \bar{q}_{ii}}{\bar{q}_{ii}} + \frac{q_{jj,t+l} - \bar{q}_{jj}}{\bar{q}_{jj}}\right), \tag{5.64}$$

with

$$E_t(q_{ij,t+l}) = \bar{\rho}(1 - \alpha - \beta) + \alpha E_t(\tilde{\epsilon}_{i,t+l-1}\tilde{\epsilon}_{j,t+l-1}) + \beta E_t(q_{ij,t+l-1}), \tag{5.65}$$

where $E_t(\tilde{\epsilon}_{i,t+l-1}\tilde{\epsilon}_{j,t+l-1}) = 1$ for $i = j$, and $E_t(\tilde{\epsilon}_{i,t+l-1}\tilde{\epsilon}_{j,t+l-1}) = E_t(\rho_{ij,t+l-1})$ for $i \neq j$.

The most important feature of the DCC model is that it can be estimated in two steps. This makes the method as feasible for the estimation of large variance-covariance matrices as the CCC model. Besides, the DCC model allows for the incorporation of nonlinear extensions and non-Gaussian conditional densities through the modeling of the univariate conditional variances, as introduced in §5.1.2 and §5.1.3, respectively. A drawback is that the dynamics of all conditional correlations are modeled to be the same as they all depend on the scalars α and β. For a review of the recent literature coping with this problem, see Bauwens et al. (2003, §2.3).

Even though the DCC model represents a more general approach to multivariate GARCH models, the assumption of constant correlations has been shown to be reasonable in many empirical applications; see, for example, Baillie and Bollerslev (1990) and Schwert and Seguin (1990). Hence, for the empirical analyses in this thesis the CCC model will be employed. As the multivariate SV model to be outlined below will also be based on the assumption of constant conditional correlations, the results obtained from the chosen multivariate conditional heteroskedasticity models can be compared in a straightforward fashion. In the absence of distractions caused by different ways of modeling conditional correlations, both chosen multivariate specifications purely reflect the respective volatility processes.

5.3.2   Multivariate stochastic volatility

In recent years, the basic univariate SV model has been extended to cope with multivariate $N \times 1$ time series vectors $\boldsymbol{y}_t = [y_{1,t} \cdots y_{N,t}]'$. Important works on the issue include, among others, Harvey et al. (1994), Pitt and Shephard (1999), Aguilar and West (2000), Chan et al. (2005), Jungbacker and Koopman (2005a) and Chib et al. (2006). Similar to the case of multivariate GARCH models, the literature on multivariate SV (MSV) focusses on (i) conditions that guarantee positive definiteness of the conditional covariance matrix, and (ii) restrictions that reduce the number of unknown parameters. According to Asai et al. (2006) the different approaches to address these two issues can be categorized as follows:

- A basic model with constant correlations.

- Asymmetric models.

- Factor models.

- Time-varying correlation models.

Asai et al. (2006, p. 171) note that "the MSV literature is still in its infancy" and that no paper has yet analyzed the relative performance of the various MSV models for predicting volatility. As these issues are beyond the scope of this thesis, only the basic MSV model will be employed in the empirical part of this thesis; for details on other MSV specifications, the reader is referred to the references cited above.

The basic MSV model has been proposed by Harvey et al. (1994). According to their representation, the univariate SV model in (5.25) and (5.26) can be generalized to multivariate series as follows:

$$y_{i,t} = \sigma_{*i} \exp\left(\frac{1}{2}\text{diag}(h_{i,t})\right)\epsilon_{i,t}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \qquad (5.66)$$

where $y_{i,t}$ denotes the t-th observation of series i, and $\boldsymbol{\epsilon_t} = [\epsilon_{1,t} \cdots \epsilon_{N,t}]'$ represents a multivariate normal vector with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}_\epsilon$. The elements of $\boldsymbol{\Sigma}_\epsilon$ on the principal diagonal are unity; the off-diagonal elements are denoted as ρ_{ij} with $\rho_{ii} = 1$ and $|\rho_{ij}| < 1$ for any $i \neq j$ and $i, j = 1, \ldots, N$. The state vector contains the unobserved multivariate volatility process:

$$h_{i,t+1} = \phi_i h_{i,t} + \eta_{i,t}, \qquad (5.67)$$

where $\boldsymbol{h}_t = [h_{1,t} \cdots h_{N,t}]'$ is the vector of log-volatilities; $\boldsymbol{\eta}_t = [\eta_{1,t} \cdots \eta_{N,t}]'$ is assumed to be multivariate normal with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}_\eta$ with elements $\sigma_{\eta,ij}$. The disturbance vectors $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_\tau$ are assumed to be uncorrelated with each other at all lags.

With the assumption that all off-diagonal elements of $\boldsymbol{\Sigma}_\eta$ are equal to zero, this MSV model is analogous to the constant conditional correlation model discussed in the context of multivariate GARCH models in §5.3.1.4: each conditional variance is specified in a univariate way, while the conditional covariance is computed as the product of a constant correlation coefficient and the corresponding conditional standard deviations. In reference to (5.55) the conditional covariance matrix can be written as

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t, \qquad (5.68)$$

where the $N \times N$ matrix $\boldsymbol{D}_t$ is now defined as $\boldsymbol{D}_t = \exp(\frac{1}{2}\text{diag}(\boldsymbol{h}_t))$; $\boldsymbol{R}$ is a symmetric positive definite $N \times N$ matrix with conditional correlation elements ρ_{ij}. With $\boldsymbol{h}_t$ representing log-volatilities, this specification guarantees positive definiteness of the conditional covariance matrix. At the same time, the number of the parameters to be estimated remains low.

Chapter 6

Time-varying market beta risk of pan-European sectors

Market beta represents one of the most widely used concepts in finance. It is used both by financial economists and practitioners. Important applications include the estimation of a security's sensitivity to the overall market, the identification of mispricings, cost of capital calculations and the evaluation of an asset manager's performance. In the context of the market model, in which beta is estimated via OLS, beta is assumed to be constant over time. However, as the real world is characterized by ongoing dynamics beta can be expected to depend on the available information at a given date and, hence, to vary over time (cf. Jagannathan and Wang 1996). Inspired by theoretical arguments that the systematic risk of an asset depends on microeconomic as well as macroeconomic factors, various studies over the last three decades have rejected the assumption of beta stability; see, for example, Fabozzi and Francis (1978), Bos and Newbold (1984) and Collins et al. (1987). While many papers have concentrated on testing the constancy of beta, only minor efforts have been made to explicitly model the time-varying behavior of systematic risk. In this chapter, a selection of modern time series techniques is employed to model and to analyze the time-varying behavior of sector-specific betas.[20]

From a practical perspective, betas prove to be especially useful in the context of sectors. Following Yao and Gao (2004) the analysis of industry effects and of the macroeconomic environment are two major determinants in the process of security analysis. Macroeconomics affect the security market through the impact on corporate earnings: under the assumption that a company's beta depends positively on the firm's sensitivity to the business cycle, sector betas can be employed to derive the implications of macroeconomic forecasts to specific industry portfolios. In spite of the empirical evidence that systematic risk on the sector level in the U.S. and in other regions of the world is time-variant, similar work in a pan-European context is still missing. This chapter aims at closing this gap. The stochastic behavior of beta for eighteen pan-European sector portfolios is modeled using time series techniques that are more sophisticated than OLS. These are based on the state space models and the conditional heteroskedasticity models outlined in the methodological part of this thesis. The data used in this chapter as

---

[20]This chapter is based on Mergner and Bulla (2008).

dependent variables are the weekly excess DJ Stoxx℠ sector return series introduced in §2.1; the DJ Stoxx℠ Broad index serves as market proxy.

The main purpose of the present chapter is to analyze the time-varying behavior of market betas for European sectors by applying elaborate modeling techniques. The respective ability of the various approaches to explain sector returns by movements of the overall market is compared. This chapter aims at contributing a comprehensive comparison of modeling techniques to the literature including non-standard procedures such as stochastic volatility, Markov switching and various Kalman filter based approaches. The results are expected to give an indication of how to proceed in the second part of empirical applications: the technique that turns out to model the systematic risk of European sectors most adequately will be employed in the subsequent chapter to analyze the time-varying impact of macroeconomic and fundamental determinants on industry portfolios.

The remainder of the chapter is organized as follows. Section 6.1 gives a brief overview of the CAPM and defines an unconditional beta. Section 6.2 reviews the literature on time-varying betas. The theoretical part of this thesis is linked to the empirical analyses to be conducted in this chapter by summarizing the competing modeling techniques together with their respective parameter estimates. Section 6.3 discusses the estimated conditional beta series and evaluates their in-sample and out-of-sample forecasting performances. Section 6.4 concludes.

6.1   The unconditional beta in the CAPM

The benchmark for time-varying market beta risk is the unconditional beta in the CAPM, introduced by Sharpe (1964) and Lintner (1965). The CAPM has played a dominant role in the field of asset pricing modeling over the last forty years. Assuming that investors can borrow and lend at a risk-free rate r^f, it linearly relates the expectation of the return on asset i, $E(r_i)$, to the systematic risk — or beta — of that security. The systematic risk is measured against the *market portfolio*, which consists of all securities on the market:

$$E(r_i) = r^f + \beta_i \left(E(r_0) - r^f \right), \tag{6.1}$$

with r_0 being the market return. Using excess returns as defined in (2.2) instead of real returns, yields a more compact version of the Sharpe-Lintner CAPM:

$$E(R_i) = \beta_i E(R_0), \tag{6.2}$$

where R_0 denotes the realized excess return of the market portfolio and R_i denotes the excess return to sector i for $i = 1, \ldots, N$. As the CAPM is designed as a single-period model a time dimension is missing.

The unconditional beta of an asset, β_i, is usually estimated by the OLS slope coefficient of the excess-return market model:

$$R_{i,t} = \alpha_i + \beta_i R_{0,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim \text{IID}(0, \sigma_i^2), \tag{6.3}$$

for period $t = 1, \ldots, T$. In the context of excess returns, α_i is expected to be zero. Beta is defined as

$$\hat{\beta}_i^{OLS} = \frac{Cov(R_0, R_i)}{Var(R_0)}. \tag{6.4}$$

The error terms $\epsilon_{i,t}$ are assumed to be normally and independently distributed over time with zero mean and constant variance σ_i^2; cf. Campbell et al. (1997, §5) who provide a review of the fundamental assumptions of the CAPM and their derivations.

Table 6.1 summarizes the OLS results of the excess market model for the eighteen DJ Stoxx℠ sector indices estimated over the full sample. As expected, the intercept is not different from zero at the 5% level for any sector. Therefore, if not mentioned otherwise, α_i will be assumed to be zero in the following. The estimated betas are all significant at the 1% level. Over the entire sample, the lowest beta was estimated for

Table 6.1: *OLS estimates of the excess market model.*

This table presents summary statistics for OLS estimation of the excess market model; *** means significance at the 1% level (**: 5%, *: 10%).

Sector	$\alpha \times 10^2$	β	R^2	JB^a	$Q(12)^b$	$LM(6)^c$
Automobiles	−0.09	1.15***	0.64	168.43***	11.75	53.38***
Banks	0.03	1.06***	0.82	2360.85***	27.67***	44.81***
Basics	0.03	0.90***	0.54	358.16***	20.31*	171.64***
Chemicals	0.00	0.91***	0.66	500.13***	19.48*	86.82***
Construction	−0.01	0.89***	0.69	369.89***	17.70	47.81***
Financials	−0.03	1.00***	0.79	599.60***	19.44*	79.92***
Food	0.04	0.65***	0.50	1485.44***	55.02***	184.76***
Healthcare	0.09	0.78***	0.50	324.62***	12.25	58.89***
Industrials	−0.03	0.98***	0.83	930.30***	13.41	58.00***
Insurance	−0.09	1.27***	0.77	1397.30***	22.58**	74.91***
Media	−0.05	1.22***	0.67	4508.10***	51.20***	74.82***
Oil & Gas	0.07	0.76***	0.43	279.22***	36.97***	114.59***
Personal	0.00	0.91***	0.74	1055.98***	21.01**	95.63***
Retail	−0.05	0.95***	0.61	964.95***	19.48*	7.26
Technology	−0.08	1.49***	0.66	1079.03***	10.21	98.73***
Telecom	0.01	1.19***	0.64	464.12***	24.50**	65.18***
Travel	−0.01	0.77***	0.65	598.81***	18.62*	38.80***
Utilities	0.08*	0.69***	0.62	38.38***	15.19	36.20***

$^a JB$ is the Jarque-Bera statistic for testing normality. The relevant critical values at the 95% (99%) level are 5.99 (9.21).

$^b Q(12)$ is the test statistic of the Ljung-Box portmanteau test for the null hypothesis of no autocorrelation in the errors up to order 12. In case of OLS, the Q-statistic is asymptotically χ^2 distributed with 12 degrees of freedom. The critical values at the 95% (99%) level are 21.03 (26.22).

$^c LM(6)$ is the LM-statistic of Engle's ARCH test for the null hypothesis of no ARCH effects up to order 6. The test statistic is asymptotically χ^2 distributed with 6 degrees of freedom. The relevant critical values at the 95% (99%) level are 12.59 (16.81).

Food & Beverages (0.65). The beta for Technology (1.49) was the highest confirming the sector's high-risk profile. The reported coefficients of determination suggest that between 43% (Oil & Gas) and 83% (Industrial Goods & Services) of the total return variation can be explained by movements of the overall market.

The last three columns provide test statistics to check whether the residuals are approximately normal and independent. As already indicated by the stylized facts discussed in §2.2, the Jarque-Bera and the ARCH-LM test statistics are highly significant for almost all sectors. The null hypothesis of normality can be rejected for every sector. The null of no heteroskedasticity can be rejected for all sectors with the exception of Retail. Regarding the Ljung-Box Q(l)-statistic to test for autocorrelation up to order l, the results are mixed: for seven sectors the null of no autocorrelation can be rejected at the 5% level. Overall, the standard regression models are not very well-specified as the assumptions underlying OLS do not generally hold.

6.2 Modeling conditional betas

The OLS estimator of beta is based on the assumption that the market model relationship is stable over the estimation period. Following the outline by Moonis and Shah (2002) sound economic arguments suggest that beta is not constant but varying over time. For example, beta can be shown to be linked to leverage. Thus changes in leverage, which may be induced by fluctuations of stock prices (cf. §5.1.2), cause changes in beta. Another reason for time-varying betas is related to information asymmetries in the market: whenever any news do not affect the returns of the overall market and a specific security equally, beta as a measure of the relative response of an asset changes. A macroeconomic argument for non-constant betas is the dependence of systematic risk on the level of the risk-free interest rate, which also changes over time. Aside from these economic arguments, time-variation of betas can be derived by the stylized fact of volatility clustering: time-variation in the second moments of returns imply time-varying conditional variances and covariances and, hence, time-variation in betas. Given these arguments in favor of considering systematic risk as being non-constant, various time series techniques can be employed to model conditional betas.

The first method is based on the multivariate GARCH model with constant conditional correlations (MGARCH) as introduced in §5.3.1.4. The conditional variance estimates derived by a GARCH(1,1) model are utilized to generate the series of conditional time-varying betas in an indirect way. This approach has been applied in previous studies: Giannopoulos (1995) uses weekly local stock market data over the period 1984 to 1993 to estimate time-varying country betas. Brooks et al. (1998) estimate time-dependent betas for Australian industry portfolios using monthly data covering the period from 1974 to 1996. Li (2003) studies the time-varying beta risk for sector portfolios in New Zealand by analyzing daily data from January 1997 to August 2002. Although the standard GARCH(1,1) model is able to describe the stylized facts of volatility clustering and excess kurtosis, asymmetric effects of positive and negative shocks on the conditional volatility are not captured (cf. discussion in §5.1.2). Therefore, nonlinear extensions of the basic GARCH model have been proposed and adopted to the modeling of time-

varying betas. For example, Braun et al. (1995) employ an EGARCH model to test for predictive asymmetries in beta; Faff et al. (2000) estimate time-varying systematic risk of UK industry indices by an EGARCH and a GJR-GARCH specification.

Event though GARCH is the most widely employed concept to model and forecast volatility, the class of SV models represents an attractive alternative. As shown in §5.2, by adding an additional contemporaneous shock to the return variance, SV models are more flexible in characterizing volatility dynamics than GARCH models. However, despite its theoretical appeal and its empirical superiority over GARCH models, the SV model is rarely used in practice for volatility forecasting or to model time-varying betas. This can be mainly attributed to the difficulties related to parameter estimation, which has been demonstrated above to be substantially more difficult for SV models. Nevertheless, the promising results by Li (2003), who uses an SV model to estimate non-constant betas for New Zealand industry portfolios, encourage further research in the applicability of SV models to the modeling of unconditional betas.

An alternative way of modeling the time-varying behavior of beta is based on the state space form of the CAPM using the Kalman filter. The Kalman filter allows to model and estimate time-varying betas directly by application of the time-varying parameter models introduced in §3.5.2. Different models for the dynamic process of conditional betas have been proposed. For U.S. data, Fabozzi and Francis (1978) and Collins et al. (1987) modeled beta as a random coefficient. The random coefficient model has also been applied by Wells (1994) for Swedish stocks and by Faff et al. (2000) for UK sectors. Two of the most prominent alternatives to model time-varying betas are the random walk model and the mean-reverting model. The former has been employed by Lie et al. (2000) for Australian financial stocks and by Li (2003) for New Zealand industry portfolios. The mean-reverting model has been used by Bos and Newbold (1984) for U.S. data and by Brooks et al. (1998) and Groenewold and Fraser (1999) for Australian industry portfolios. For their investigation of the systematic risk of Australian industrial stock returns, Yao and Gao (2004) also considered a moving mean reverting model, in which the mean beta is allowed to vary over time as proposed by Wells (1994).

The last approach to be considered in this chapter uses a Markov switching framework, which belongs to the large class of Markov switching models introduced in Chapter 4. Although Markov switching regression models have been applied in many different settings, the literature dealing with time-varying betas is relatively thin. Fridman (1994) considered monthly data from 1980 to 1991 to analyze the excess returns of three oil companies. The fitting of a two-state regression model led to an improved assessment of systematic risk associated with each security. Two interesting effects were observed:

- Beta increases whenever the process is in the more volatile state.

- The state associated with higher volatility tends to be less persistent than the state associated with lower volatility.

Huang (2000) also considered a Markov switching model with one high-risk and one low-risk state. Using monthly return data from 1986 to 1993, he performed several test to check the consistency of different states with the CAPM. He rejected the hypothesis that the data were from the same state.

6.2.1   GARCH conditional betas

While in the traditional CAPM returns are assumed to be IID, it is well established in the empirical finance literature that this is not the case for returns in many financial markets. Signs of autocorrelation and regularly observed volatility clusters contradict the assumption of an identical and independent return distribution over time. In this case, the variance-covariance matrix of sector and market returns is time-dependent. A non-constant beta can be defined by extending (6.4) to

$$\hat{\beta}_{i,t}^{GARCH} = \frac{Cov(R_{i,t}, R_{0,t})}{Var(R_{0,t})}. \tag{6.5}$$

The conditional beta is obtained indirectly as the ratio between the time-varying conditional covariance between a sector and the overall market, and the time-varying conditional market variance.

A bivariate version of the MGARCH model with constant conditional correlations can be used to compute time-varying betas. Let $\boldsymbol{R}_t = [R_{i,t} \quad R_{0,t}]'$ be a 2×1 time-series vector, where $R_{i,t}$ represents the return series of sector i for $i = 1, \dots, 18$, and $R_{0,t}$ denotes the return series of the broad market. Consider a system of $N = 2$ conditional mean equations:

$$\boldsymbol{R}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \tag{6.6}$$

where $\boldsymbol{\mu} = [\mu_1 \quad \mu_2]'$ is a 2×1 vector of constants and $\boldsymbol{\epsilon}_t = [\epsilon_{1,t} \quad \epsilon_{2,t}]'$ denotes a 2×1 time series vector of residuals, conditioned by the complete information set Ω_{t-1}. In accordance with (5.45) a general bivariate GARCH model for the two-dimensional process $\boldsymbol{\epsilon}_t | \Omega_{t-1}$ is represented by

$$\boldsymbol{\epsilon}_t = \boldsymbol{z}_t \boldsymbol{H}_t^{1/2}, \tag{6.7}$$

where $\boldsymbol{z}_t$ is a two-dimensional IID process with mean zero and covariance matrix $\boldsymbol{I}_2$. This implies $E(\boldsymbol{\epsilon}_t | \Omega_{t-1}) = \boldsymbol{0}$ and $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' | \Omega_{t-1}) = \boldsymbol{H}_t$. The most influential parameterizations for $\boldsymbol{H}_t$ have been summarized in §5.3.1, where it has also been discussed why the constant correlation model is to be preferred in the following.

Referring to (5.53) and (5.54) the assumption of constant conditional correlations leads to a bivariate GARCH model of the form

$$h_{ii,t} = \psi_{ii} + \gamma_{ii}\epsilon_{i,t-1}^2 + \delta_{ii}h_{ii,t-1}, \tag{6.8}$$

$$h_{ij,t} = \rho_{ij}\sqrt{h_{ii,t}}\sqrt{h_{jj,t}}, \qquad \text{for all } i \neq j, \tag{6.9}$$

which can be substituted into (6.5) to estimate a time-varying beta of sector i:

$$\hat{\beta}_{i,t}^{GARCH} = \frac{Cov(R_{0,t}, R_{i,t})}{Var(R_{0,t})} = \rho_{0i}\frac{\sqrt{h_{ii,t}}}{\sqrt{h_{00,t}}}, \tag{6.10}$$

where ρ_{0i} denotes the unconditional correlation between the overall market and sector i.

The last step before actually estimating ARCH-based time-varying betas is to find an adequate functional form for the univariate GARCH(1,1) models for market and sector excess returns in (6.8). Two issues have to be dealt with: excess kurtosis and potential

Table 6.2: *Testing for asymmetric ARCH effects.*

This table presents the test statistics together with the corresponding p-values for the Sign Bias (SB), Negative Sign Bias (NSB), Positive Sign Bias (PSB) test and the general test for asymmetric volatility effects.

Sector	SB test		NSB test		PSB test		General test	
	Test	p-value	Test	p-value	Test	p-value	Test	p-value
Broad	*2.65*	*0.004*	*−10.08*	*0.000*	*0.52*	*0.303*	*116.50*	*0.000*
Automobiles	1.91	0.028	−7.75	0.000	0.73	0.234	72.22	0.000
Banks	2.25	0.012	−9.81	0.000	1.56	0.060	115.21	0.000
Basic	1.94	0.026	−5.59	0.000	−0.75	0.226	33.27	0.000
Chemicals	2.22	0.013	−7.36	0.000	0.66	0.254	64.19	0.000
Construction	2.49	0.006	−6.56	0.000	−0.44	0.330	47.27	0.000
Financials	2.85	0.002	−9.78	0.000	1.74	0.041	112.64	0.000
Food	1.32	0.093	−5.60	0.000	−0.99	0.161	37.66	0.000
Healthcare	2.31	0.011	−7.87	0.000	−0.20	0.419	70.92	0.000
Industrials	2.59	0.005	−8.38	0.000	−0.76	0.224	74.77	0.000
Insurance	1.67	0.047	−9.19	0.000	1.03	0.152	93.70	0.000
Media	1.16	0.123	−6.12	0.000	2.44	0.007	57.07	0.000
Oil & Gas	1.59	0.056	−8.95	0.000	1.12	0.131	100.34	0.000
Personal	3.42	0.000	−8.61	0.000	−1.07	0.141	74.99	0.000
Retail	2.36	0.009	−4.47	0.000	−0.49	0.310	20.79	0.000
Technology	−0.29	0.384	−3.05	0.001	3.96	0.000	35.96	0.000
Telecom	1.71	0.043	−4.25	0.000	1.18	0.118	26.49	0.000
Travel	2.72	0.003	−7.83	0.000	−0.78	0.219	64.52	0.000
Utilities	1.85	0.032	−6.72	0.000	0.73	0.232	56.20	0.000

asymmetric effects. It has been shown in §2.2.1 that the sector returns in the selected sample are highly leptokurtic. Following the outline in §5.1.3, a standardized Student-t distribution can be considered for the innovation terms z_t in (6.7) to avoid a reliance on QML based procedures. With respect to the second issue, it has been outlined in §5.1.2 that aggregate equity indices are frequently observed to exhibit asymmetric effects. The test results presented in Table 6.2 confirm that this feature also holds true for the pan-European sectors to be considered here. The application of the SB, NSB, PSB tests and the general test for asymmetric effects as described in §5.1.2.3, gives substantial evidence of asymmetric conditional volatility effects. With the exception of Media and Technology, both the sign and negative size effects appear to be of greater importance than positive size effects.

To decide on the functional form to be utilized, various univariate GARCH(1,1) models are fitted to all sectors and the broad market. The different specifications allow for Gaussian versus t-distributed errors, for symmetric versus asymmetric effects and for a constant versus no constant in the mean equation (6.6). The different models are compared using the estimated BIC as defined in §3.6.2.3. The results in Table 6.3 reveal that, other things being equal, on average specifications with t-distributed errors offer lower information criteria than in case with normally distributed innovations z_t.

Table 6.3: *Comparison of different GARCH(1,1) specifications.*

To decide whether (i) the innovations z_t in (6.7) should be modeled by a normal or a standardized Student-t distribution, (ii) asymmetric effects should be explicitly taken into account, and (iii) a nonzero constant should be included in the mean equation in (6.6), eight different GARCH(1,1) specifications have been analyzed. This table reports the estimated values of the corresponding log-likelihood functions and the Bayesian information criteria.

| Sector | symmetric GARCH | | | | | | | | asymmetric GARCH | | | | | | | |
| | $[z_t \sim N, \mu \neq 0]$ | | $[z_t \sim N, \mu = 0]$ | | $[z_t \sim t, \mu \neq 0]$ | | $[z_t \sim t, \mu = 0]$ | | $[z_t \sim N, \mu \neq 0]$ | | $[z_t \sim N, \mu = 0]$ | | $[z_t \sim t, \mu \neq 0]$ | | $[z_t \sim t, \mu = 0]$ | |
	$\log L$	BIC	$\log L$	BIC	$\log L$	BIC	$\log L$	BIC	$\log L$	BIC	$\log L$	BIC	$\log L$	BIC	$\log L$	BIC
Broad	*2215.9*	*−4.91*	*2209.6*	*−4.90*	*2228.5*	*−4.93*	*2220.1*	*−4.92*	*2226.2*	*−4.93*	*2223.5*	*−4.93*	*2234.6*	*−4.94*	*2229.5*	*−4.93*
Automobiles	1871.9	−4.14	1870.7	−4.15	1892.5	−4.18	1890.6	−4.19	1879.9	−4.15	1879.7	−4.16	1895.0	−4.18	1893.9	−4.18
Banks	2112.8	−4.68	2108.0	−4.68	2132.4	−4.72	2124.1	−4.71	2119.3	−4.69	2116.7	−4.69	2136.0	−4.72	2129.9	−4.71
Basics	1988.2	−4.40	1985.5	−4.40	2013.8	−4.45	2009.6	−4.45	1995.7	−4.41	1993.8	−4.42	2017.7	−4.45	2014.6	−4.45
Chemicals	2110.1	−4.67	2106.3	−4.67	2117.5	−4.68	2112.2	−4.68	2120.7	−4.69	2119.3	−4.69	2125.5	−4.69	2122.8	−4.70
Construction	2112.9	−4.68	2109.8	−4.68	2129.7	−4.71	2124.3	−4.71	2126.0	−4.70	2124.6	−4.71	2137.6	−4.72	2134.5	−4.72
Financials	2144.4	−4.75	2137.4	−4.74	2175.7	−4.81	2168.1	−4.80	2156.2	−4.77	2152.0	−4.77	2181.6	−4.82	2175.3	−4.81
Food	2240.5	−4.97	2238.1	−4.97	2257.3	−5.00	2252.1	−4.99	2247.5	−4.97	2246.3	−4.98	2261.8	−5.00	2258.0	−5.00
Healthcare	2083.3	−4.61	2076.7	−4.61	2090.4	−4.62	2083.8	−4.62	2088.4	−4.62	2084.6	−4.62	2093.1	−4.62	2088.4	−4.62
Industrials	2144.2	−4.75	2137.8	−4.74	2158.9	−4.78	2150.9	−4.77	2154.8	−4.77	2151.3	−4.77	2164.3	−4.78	2158.9	−4.78
Insurance	1973.8	−4.37	1971.3	−4.37	2006.3	−4.44	2003.1	−4.44	1981.3	−4.38	1980.4	−4.39	2008.6	−4.43	2006.2	−4.44
Media	1927.4	−4.27	1925.1	−4.27	1941.8	−4.29	1937.5	−4.29	1930.7	−4.27	1929.8	−4.27	1943.5	−4.29	1940.7	−4.29
Oil & Gas	2048.7	−4.54	2044.0	−4.53	2056.8	−4.55	2052.6	−4.55	2061.0	−4.56	2059.4	−4.56	2066.7	−4.56	2064.9	−4.57
Personal	1570.8	−4.56	1567.7	−4.56	1577.9	−4.57	1574.2	−4.57	1582.9	−4.59	1582.3	−4.60	1587.4	−4.59	1586.2	−4.60
Retail	1466.1	−4.25	1464.6	−4.26	1521.2	−4.41	1517.6	−4.41	1478.8	−4.28	1478.5	−4.29	1526.2	−4.41	1524.1	−4.42
Technology	1780.0	−3.94	1776.9	−3.94	1793.5	−3.96	1788.3	−3.96	1781.8	−3.93	1779.5	−3.94	1794.3	−3.96	1790.2	−3.95
Telecom	1863.2	−4.12	1858.5	−4.12	1866.1	−4.12	1861.0	−4.12	1864.9	−4.12	1861.3	−4.12	1867.4	−4.12	1863.3	−4.12
Travel	1640.8	−4.77	1637.4	−4.77	1652.8	−4.79	1649.7	−4.79	1653.5	−4.79	1652.9	−4.80	FTC[a]	FTC	FTC	FTC
Utilities	2270.9	−5.03	2265.7	−5.03	2284.4	−5.06	2276.1	−5.04	2273.4	−5.03	2269.7	−5.03	2285.6	−5.05	2278.6	−5.04
Mean BIC		−4.55		−4.55		−4.58		−4.58		−4.56		−4.56		−4.57		−4.57

[a]Failed to converge.

Regarding asymmetric effects, symmetric GARCH models with t-distributed innovations are superior to the alternative GJR-counterparts. These results are in line with Bollerslev et al. (1992, §3.3). They argue that leverage effects are often attributable to a few outliers, which can be better captured by employing a fat-tailed distribution than by explicitly allowing for asymmetric effects. The inclusion of a constant term generally improves the values of the corresponding log-likelihood without leading to higher BIC. These results suggest to model ARCH-based conditional betas by a bivariate GARCH framework, in which univariate t-GARCH(1,1) models with nonzero constants in the mean equation are fitted to the excess returns of each sector and the overall market. The estimation results are summarized in Table 6.4. The corresponding beta series are denoted as $\beta_{i,t}^{tG}$.

Table 6.4: *Parameter estimates for t-GARCH(1,1) models.*

This table reports the estimated parameters for the t-GARCH(1,1) models for the eighteen DJ STOXX℠ sectors and the DJ Stoxx℠ Broad as market index; *** means significance at the 1% level (**: 5%, *: 10%).

Sector	$\mu \times 10^2$	$\omega \times 10^4$	γ	δ	DF^a	JB^b	$Q(12)^c$
Broad	*0.25****	*0.16***	*0.13****	*0.84****	*9.29****	*127.12****	*16.96*
Automobiles	0.17*	0.25**	0.10***	0.88***	7.03***	223.82***	17.46
Banks	0.26***	0.12**	0.12***	0.87***	7.88***	245.82***	13.14
Basics	0.22***	0.05	0.06***	0.94***	6.63***	580.94***	22.99**
Chemicals	0.22***	0.27**	0.14***	0.82***	9.48***	44.56***	9.70
Construction	0.23***	0.24**	0.11***	0.85***	8.26***	279.77***	18.71*
Financials	0.25***	0.16*	0.14***	0.84***	6.27***	485.41***	37.67***
Food	0.20***	0.23**	0.11***	0.84***	7.79***	144.74***	12.76
Healthcare	0.27***	0.23*	0.10***	0.86***	10.54***	27.956***	13.70
Industrials	0.26***	0.16**	0.14***	0.84***	8.26***	139.28***	34.59***
Insurance	0.18**	0.14**	0.09***	0.90***	6.01***	648.31***	14.16
Media	0.23***	0.18**	0.10***	0.88***	7.76***	88.42***	10.92
Oil & Gas	0.22***	0.20	0.08***	0.89***	9.43***	29.39***	9.38
Personal	0.23***	0.51**	0.16***	0.76***	12.10***	69.55***	18.74*
Retail	0.24***	0.67**	0.14***	0.78***	6.17***	14986.00***	10.97
Technology	0.29***	0.14**	0.11***	0.89***	8.70***	137.26***	17.75
Telecom	0.29***	0.14*	0.09***	0.90***	15.42***	8.94**	19.20*
Travel	0.17**	0.54**	0.15***	0.75***	8.43***	90.72***	6.26
Utilities	0.24***	0.38**	0.15***	0.76***	7.76***	90.24***	19.19*

$^a DF$ denotes the number of degrees of freedom of the Student-t distribution, which has been estimated along with the other parameters of the t-GARCH(1,1) models.

$^b JB$ is the Jarque-Bera statistic for testing normality. The relevant critical values at the 95% (99%) level are 5.99 (9.21).

$^c Q(12)$ is the test statistic of the Ljung-Box portmanteau test for the null hypothesis of no autocorrelation in the errors up to order 12. The critical values at the 95% (99%) level are 21.03 (26.22).

While the constant term in the mean equation, μ_i, is statistically significant at the 1% level for 17 sectors, ω_i is different from zero at the 5% level only for 13 sectors. The coefficients for the ARCH and GARCH terms, γ_i and δ_i, are always significantly different from zero at the 1% level. They are all positive and sum up to less than unity, such that positive definiteness and stationarity are guaranteed. The highest degree of persistence can be obsreved for Basic Resources, Technology and Telecommunications. On the other hand, the models for Travel & Leisure, Utilities and Retail are the least persistent. With regard to the reported diagnostics, the Jarque-Bera statistics indicate that the residuals are generally not normally distributed. With the exception of Basic Resources, Financial Services and Industrials, where the null of white noise can be rejected at the 5% or higher level, the residuals show no signs of autocorrelation.

Following (6.10) the correlation coefficient ρ_{0i} between a sector and the overall market is the other factor that is needed to calculate GARCH conditional betas. The unconditional correlations, estimated in the context of stylized facts over the entire sample, have been reported in the last column of Table 2.2.

6.2.2 Stochastic volatility conditional betas

By including an additional contemporaneous shock to the return variance, SV models allow the conditional mean and the conditional volatility of return series to be characterized by different shocks. As discussed in §5.2, SV models offer a higher degree of flexibility and imply excess kurtosis. This qualifies them to be more appropriate in describing financial time series than GARCH-type approaches. Hence, SV models represent an alternative to the modeling of time-varying betas.

According to (5.25) and (5.26) a general SV model can be represented by its first two moments. The mean equation is given by

$$R_{i,t} = \sigma_{i,t}\epsilon_{i,t}, \qquad \epsilon_{i,t} \sim \text{IID}(0,1), \qquad t = 1, \ldots, T, \tag{6.11}$$

where $R_{i,t}$ is the return series of index i. The disturbances are assumed to be identically and independently normally distributed with zero mean and unit variance. The variance equation is given by

$$\sigma_{i,t}^2 = \sigma_{*i}^2 \exp(h_{i,t}), \tag{6.12}$$

where the actual volatility $\sigma_{i,t}^2$ is the product of the positive scaling factor σ_{*i}^2 and the exponential of the stochastic process $h_{i,t}$. The log-volatility $h_{i,t}$ is modeled as a first-order autoregressive process:

$$h_{i,t+1} = \phi_i h_{i,t} + \sigma_{\eta i}\eta_{i,t}, \qquad \eta_{i,t} \sim \text{IID}(0,1), \qquad h_{i,1} \sim N\left(0, \frac{\sigma_{\eta i}^2}{1 - \phi_i^2}\right). \tag{6.13}$$

The persistence parameter ϕ_i is restricted to be positive and smaller than one to ensure stationarity of $h_{i,t}$. The disturbances $\epsilon_{i,t}$ and $\eta_{i,t}$ are assumed to be uncorrelated, contemporaneously and at all lags. The inclusion of an unobservable shock to the return variance makes the variance a latent process. As this process cannot be characterized explicitly with respect to observable past information, the parameters of the SV model

Table 6.5: *Parameter estimates for stochastic volatility models.*

This table reports the estimated parameters for the SV models for the eighteen DJ Stoxx℠ sectors and the DJ Stoxx℠ Broad as market index. Figures in parentheses denote the lower and upper bounds of the asymmetric asymptotic 95% confidence intervals. For the *JB*- and *Q*-test statistics, *** means significance at the 1% level (**: 5%, *: 10%).

Sector	ϕ	σ_η^2	σ_*^2	JB-statistics[a]	$Q(12)$[b]
Broad	*0.966*	*0.039*	*3.677*	*13.57******	*16.07*****
	(0.924;0.986)	*(0.018;0.083)*	*(2.498;5.413)*		
Automobiles	0.964	0.041	7.651	8.30**	10.77
	(0.920;0.984)	(0.020;0.087)	(5.255;11.139)		
Banks	0.977	0.035	4.362	10.37***	14.89*
	(0.947;0.990)	(0.018;0.068)	(2.585;7.360)		
Basics	0.973	0.029	6.156	1.71	13.93*
	(0.925;0.991)	(0.010;0.085)	(4.066;9.320)		
Chemicals	0.958	0.047	4.737	9.28**	8.42
	(0.906;0.982)	(0.021;0.108)	(3.346;6.706)		
Construction	0.955	0.039	4.707	9.20*	12.42
	(0.903;0.980)	(0.018;0.086)	(3.482;6.363)		
Financials	0.958	0.067	4.013	7.85**	31.58***
	(0.911;0.981)	(0.033;0.139)	(2.662;6.049)		
Food	0.941	0.053	3.446	9.55***	11.30
	(0.878;0.972)	(0.025;0.114)	(2.630;4.516)		
Healthcare	0.953	0.038	5.031	2.47	13.10
	(0.882;0.982)	(0.014;0.103)	(3.783;6.689)		
Industrials	0.965	0.043	4.314	11.47***	26.75***
	(0.938;0.981)	(0.023;0.083)	(2.831;6.574)		
Insurance	0.979	0.035	5.833	6.28**	15.40*
	(0.952;0.991)	(0.018;0.067)	(3.267;10.414)		
Media	0.978	0.037	6.829	5.73*	7.71
	(0.947;0.991)	(0.018;0.074)	(3.950;11.808)		
Oil & Gas	0.963	0.031	5.518	2.71	10.87
	(0.894;0.988)	(0.010;0.098)	(4.004;7.605)		
Personal	0.947	0.042	5.219	7.69**	19.61**
	(0.864;0.980)	(0.015;0.117)	(3.842;7.088)		
Retail	0.917	0.093	5.975	6.35**	14.18*
	(0.830;0.962)	(0.042;0.205)	(4.442;8.035)		
Technology	0.989	0.021	9.470	7.88**	14.98*
	(0.970;0.996)	(0.010;0.046)	(4.138;21.672)		
Telecom	0.989	0.013	7.590	0.81	17.80**
	(0.966;0.997)	(0.006;0.030)	(3.895;14.789)		
Travel	0.905	0.089	4.074	4.33	8.27
	(0.807;0.956)	(0.039;0.206)	(3.138;5.289)		
Utilities	0.908	0.084	3.215	14.62***	19.99**
	(0.837;0.950)	(0.044;0.159)	(2.561;4.038)		

[a]The relevant critical values for the Jarque-Bera test at the 95% (99%) level are 5.99 (9.21).

[b]In a structural model, the *Q*-statistic of the Ljung-Box portmanteau test for the null hypothesis of no autocorrelation in the errors up to order l is asymptotically χ^2 distributed with $l - w - 1$ degrees of freedom, where w denotes the total number of estimated parameters (Harvey 1989, p. 259). The relevant critical values at the 95% (99%) level are 15.51 (20.09).

cannot be estimated by a direct application of standard maximum likelihood techniques. As outlined in §5.2.2, several procedures for estimating SV models have been proposed. These include method of moments and QML estimators, Bayesian approaches based on MCMC techniques, the MCL estimator and the efficient MCL procedure. A consensus on how to estimate SV models is still missing.

As outlined in §5.2.3, in this thesis SV models are estimated via efficient MCL using importance sampling techniques. Recalling the MSV model in (5.68) time-varying sector betas can be constructed as

$$\hat{\beta}_{i,t}^{SV} = \rho_{0i} \frac{\sqrt{h_{ii,t}}}{\sqrt{h_{00,t}}}. \tag{6.14}$$

In contrast to the GARCH based conditional betas defined in (6.10) smoothed rather than filtered estimates of the conditional variance series of market and sector returns, $h_{00,t}$ and $h_{ii,t}$, are employed. All the available information up to and including date T can be relied upon for in-sample analysis purposes.

A summary of the estimation results of the considered univariate SV models for European sectors over the full sample period is given in Table 6.5. The asymmetric 95% confidence intervals for the persistence parameter ϕ_i are generally narrow. This indicates a high level of significance. The degree of volatility persistence ranges from a low for Travel & Leisure, to the highest level for Technology and Telecommunications. This compares well to the GARCH results. For the two other parameters, $\sigma_{\eta i}^2$ and σ_{*i}^2, both the asymmetric confidence intervals as well as the range of parameter estimates across sectors, are wider. For the sectors Retail, Travel & Leisure and Utilities, the combination of a low persistence parameter and a high value for $\sigma_{\eta i}^2$, which measures the variation of the volatility process, implies that the process of volatility is less predictable for these three sectors. The highest levels of volatility, as indicated by a high scaling parameter σ_{*i}^2, are found for Automobiles & Parts and the three sectors Telecommunications, Media and Technology (TMT). This finding broadly corresponds to the calculated standard deviations of weekly returns in Table 2.2. Notably, the null of normality cannot be rejected at the 5% level for seven sectors. Compared to the GARCH models estimated in the previous subsection, the estimated Jarque-Bera test statistics are significantly lower for all sectors and also for the overall market. The reported Q-statistics indicate some degree of autocorrelation at the 5% level for five sectors as well as the market index.

Figure 6.1 allows for a comparison of conditional volatility estimates based on a t-GARCH(1,1) and a SV model. The graph displays the filtered and smoothed conditional volatility estimates for the Telecommunications sector. Overall, the filtered conditional volatility series show a very similar pattern. The range of the GARCH based series is greater than in case of its SV counterpart. The smoothed estimate, which — per definition — takes the full set of observations into account, reveals that both filtered series tend to overstate the level of volatility.

6.2.3 Kalman filter based approaches

In contrast to volatility-based techniques, where the conditional beta series can only be constructed indirectly after the conditional variances of the market and sector i have
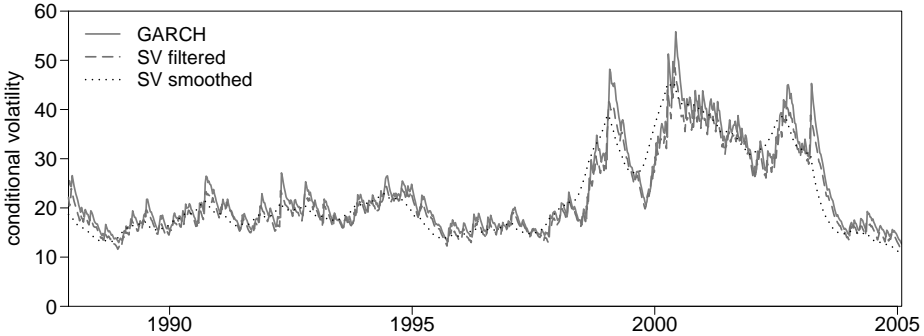
Figure 6.1: *Conditional volatility estimates for the Telecommunications sector.*

been obtained first, the state space approach allows to model and estimate the time-varying structure of beta directly. As outlined in Chapter 3, state-space models can be estimated through a recursive algorithm known as the Kalman filter.

In state space form, the excess-return market model in (6.3) is modified to become an observation equation:

$$R_{i,t} = \beta_{i,t} R_{0,t} + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim N(0, \sigma_{\epsilon i}^2). \tag{6.15}$$

In accordance with (3.75) the dynamic process of the unobserved time-varying state vector $\beta_{i,t}$ can be generally defined by the state equation

$$\beta_{i,t+1} - \bar{\beta}_i = \phi_i(\beta_{i,t} - \bar{\beta}_i) + \eta_{i,t}, \qquad \eta_{i,t} \sim N(0, \sigma_{\eta i}^2), \tag{6.16}$$

where ϕ_i denotes the constant transition (or speed) parameter. The system of equations (6.15)–(6.16) represents a special case of the general state space model introduced in §3.2. The constant variances $\sigma_{\epsilon i}^2$ and $\sigma_{\eta i}^2$ together with the transition parameter ϕ_i are the hyperparameters of the system. As discussed in §3.5.2, alternative specifications for the stochastic process of $\beta_{i,t}$ are derived by formulating different assumptions on ϕ_i and $\bar{\beta}_i$.

6.2.3.1   The random walk model

The RW model represents the first state space specification of the evolution of the time-varying beta to be employed in the following. By setting ϕ_i to unity, the beta coefficient develops as a random walk:

$$\hat{\beta}_{i,t+1}^{RW} = \beta_{i,t} + \eta_{i,t}, \tag{6.17}$$

where the two hyperparameters $\sigma_{\epsilon i}^2$ and $\sigma_{\eta i}^2$ have to be estimated.

The parameter estimates and diagnostic test statistics for the RW model are reported in Table 6.6, which summarizes the estimation results for all considered Kalman filter models. The estimated hyperparameters are all significant at the 1% level. While the

null hypothesis of normality can be generally rejected, signs of autocorrelation in the residuals are observed at the 5% level for twelve out of eighteen sectors. Even though the ARCH-test is highly significant for all sectors, for eleven sectors the corresponding LM-statistics are lower than in case of OLS. It can be concluded that the RW specification is able to capture the heteroskedasticity in the resulting residuals at least partially. With regard to the goodness of fit, the R^2 is always significantly higher than in case of OLS.

6.2.3.2   The mean reverting model

As an alternative to the random walk, the dynamic process of beta can be modeled as being mean reverting. In the MR model, an autoregressive process of order one, AR(1), with a constant mean is used for the evolution of beta:

$$\hat{\beta}_{i,t+1}^{MR} - \bar{\beta}_i = \phi_i(\beta_{i,t} - \bar{\beta}_i) + \eta_{i,t}. \tag{6.18}$$

In order to use the Kalman filter for estimating the MR model, the mean beta coefficient is augmented into the state vector. This leads to the following state space model:

$$\begin{bmatrix} \beta_{i,t+1} - \bar{\beta}_i \\ \bar{\beta}_i \end{bmatrix} = \begin{bmatrix} \phi_i & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_{i,t} - \bar{\beta}_i \\ \bar{\beta}_i \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{i,t} \\ 0 \end{bmatrix}, \tag{6.19}$$

$$y_t = \begin{bmatrix} x_t & x_t \end{bmatrix} \begin{bmatrix} \beta_{i,t} - \bar{\beta}_i \\ \bar{\beta}_i \end{bmatrix} + \epsilon_t, \tag{6.20}$$

where overall four parameters ($\sigma_{\epsilon i}^2$, $\sigma_{\eta i}^2$, $\bar{\beta}_i$, ϕ_i) have to be estimated. As outlined in §3.5.2.3, the AR(1) parameter ϕ_i should be restricted to values between zero and unity. This constraint is implemented in the estimation procedure by application of the parameter transformation (3.67).

The estimated variances of observation and state disturbances are significant at the 1% level for every sector. Compared to the RW model above, the estimated variance for the observation equation, σ_i^2, is generally lower in case of the MR parameterization; the opposite holds with respect to $\sigma_{\eta i}^2$, the variance of the dynamic process of the time-varying beta. For the MR model, two additional parameters have been estimated. The value for $\bar{\beta}_i$, which compares well to the estimated OLS betas as reported in Table 6.1, is always significant at the 1% level. The estimates for the second extra parameter can be clustered across all sectors into three groups. In the first group, ϕ_i is close to unity. The closer to unity the transition parameter gets, the more the conditional beta resembles its RW counterpart. In case of Food & Beverages, Healthcare and Personal & Household Goods, the MR betas literally follow a random walk. In the second group with values for ϕ_i around 0.5, the conditional betas return faster to their individual means, which implies more noisy series of conditional betas. In the third group, where ϕ_i is close to zero, the resulting beta series follow a random coefficient model as defined in §3.5.2.1. Note that the estimated speed parameters in the last group (Industrials and Travel & Leisure) are not statistically significant.

According to the fit statistics, the residuals are generally non-normal. With the exception of Healthcare and Retail, the null of no autocorrelation can be rejected for all sectors. According to the conducted ARCH-test, the null of no heteroskedasticity cannot

Table 6.6: *Parameter estimates for Kalman filter models.*

This table reports the estimated parameters for the Kalman filter based models for the eighteen DJ Stoxx℠ sectors. All estimates for σ^2, σ^2_η, σ^2_ζ and $\bar{\beta}$ are significant at the 1% level. For ϕ and the test statistics that refer to the observation errors, *** means significance at the 1% level (**: 5%, *: 10%).

Sector	Model	$\sigma^2 \times 10^4$	$\sigma^2_\eta \times 10^2$	$\sigma^2_\zeta \times 10^2$	ϕ	$\bar{\beta}$	$\log L$	BIC	R^2	JB^a	$Q(12)^b$	$LM(6)^c$
Automobiles	RW	3.20	0.80				2289.9	−5.09	0.72	39.65***	13.63	23.15***
	MR	2.61	16.61		0.55***	1.15	2304.5	−5.11	0.81	84.93***	13.79*	13.02**
	MMR	2.50	19.96	0.22	0.01		2309.5	−5.12	0.82	115.11***	14.26**	10.03
	GWR	0.02	0.25				4668.2	−10.39	0.69	9.60***	14.01	4.30
Banks	RW	1.06	0.19				2792.9	−6.21	0.86	368.83***	13.31	97.17***
	MR	0.75	9.38		0.41***	1.03	2808.8	−6.23	0.92	1186.10***	25.16***	15.65**
	MMR	0.75	7.44	0.10	0.00		2825.1	−6.27	0.92	1052.70***	21.14***	13.50**
	GWR	0.02	0.07				4664.6	−10.39	0.85	186.35***	17.28**	1.10
Basics	RW	3.22	0.47				2297.7	−5.11	0.62	417.28***	25.31***	116.97***
	MR	3.07	2.08		0.92***	0.96	2303.6	−5.11	0.65	525.94***	22.91***	112.63***
	MMR	2.77	11.51	0.23	0.00		2308.5	−5.12	0.71	759.67***	27.08***	80.00***
	GWR	0.02	0.13				4665.4	−10.39	0.70	20.71***	23.04***	4.18
Chemicals	RW	1.85	0.23				2547.9	−5.67	0.73	287.48***	16.34*	70.59***
	MR	1.78	0.89		0.94***	0.91	2551.5	−5.66	0.75	360.87***	15.81**	63.37***
	MMR	1.62	5.40	0.10	0.40*		2555.4	−5.67	0.79	578.40***	20.31***	52.85***
	GWR	0.02	0.05				4653.5	−10.36	0.73	28.35***	19.46**	8.76

Continued on the next page...

[a] JB is the Jarque-Bera statistic for testing normality. The relevant critical values at the 95% (99%) level are 5.99 (9.21).

[b] $Q(l)$ is the test statistic of the Ljung-Box portmanteau test for the null hypothesis of no autocorrelation in the errors up to order l. In a structural model, the Q-statistic is asymptotically χ^2 distributed with $l - w - 1$ degrees of freedom, where w denotes the total number of estimated parameters (Harvey 1989, p. 259).

[c] $LM(l)$ is the LM-statistic of Engle's ARCH test for the null hypothesis of no ARCH effects up to order l. The test statistic is asymptotically χ^2 distributed with l degrees of freedom. The relevant critical values at the 95% (99%) level are 12.59 (16.81).

Table 6.6 — continued

Sector	Model	$\sigma^2 \times 10^4$	$\sigma_\eta^2 \times 10^2$	$\sigma_\zeta^2 \times 10^2$	ϕ	$\bar{\beta}$	$\log L$	BIC	R^2	JB	$Q(12)$	$LM(6)$
Construction	RW	1.56	0.06				2638.4	−5.87	0.74	200.45***	21.75***	58.17***
	MR	1.20	11.34		0.36***	0.94	2629.8	−5.83	0.84	354.47***	29.36***	32.31***
	MMR	1.22	7.46	0.04	0.00		2659.5	−5.90	0.83	260.43***	24.24***	33.40***
	GWR	0.02	0.04				4693.4	−10.45	0.75	11.52***	28.33***	7.73
Financials	RW	1.26	0.16				2721.2	−6.05	0.82	696.12***	16.87*	37.17***
	MR	0.88	9.98		0.20***	0.94	2756.7	−6.12	0.90	1175.70***	32.93***	2.70
	MMR	0.88	9.73	0.00	0.16**		2757.9	−6.12	0.90	1190.30***	32.55***	2.80
	GWR	0.02	0.04				4657.2	−10.37	0.79	76.45***	18.01**	12.36*
Food	RW	1.68	0.20				2593.0	−5.77	0.64	3049.40***	22.14***	95.44***
	MR	1.67	0.23		0.99**	0.68	2593.7	−5.75	0.64	3121.40***	22.07***	94.78***
	MMR	1.35	7.91	0.12	0.00		2609.3	−5.79	0.75	4750.80***	19.23***	69.40***
	GWR	0.02	0.05				4669.3	−10.40	0.74	145.55***	8.14*	5.90
Healthcare	RW	2.84	0.18				2364.9	−5.26	0.57	184.90***	12.15	43.46***
	MR	2.82	0.31		0.98***	0.81	2366.2	−5.25	0.57	195.24***	12.71*	42.82***
	MMR	2.56	6.31	0.13	0.00		2370.7	−5.26	0.64	298.90***	12.86*	29.77***
	GWR	0.02	0.09				4663.0	−10.38	0.59	74.59***	11.97	2.67
Industrials	RW	0.98	0.06				2843.1	−6.32	0.84	923.81***	22.72***	59.00***
	MR	0.59	11.04		0.00	1.00	2881.8	−6.40	0.93	862.32***	22.16***	12.08*
	MMR	0.59	9.49	0.04	0.00		2888.0	−6.41	0.93	881.82***	26.64***	15.03**
	GWR	0.02	0.02				4675.5	−10.41	0.85	54.04***	26.17***	1.97
Insurance	RW	1.83	0.30				2548.5	−5.67	0.84	215.62***	20.19**	42.17***
	MR	1.26	13.67		0.63***	1.15	2570.5	−5.70	0.92	716.45***	15.38**	8.05
	MMR	1.25	11.80	0.16	0.04		2600.7	−5.77	0.91	550.26***	17.39**	8.89
	GWR	0.02	0.15				4699.6	−10.46	0.80	14.01***	26.69***	4.02
Media	RW	3.00	0.51				2326.5	−5.17	0.75	2047.20***	37.53***	78.38***
	MR	1.92	37.69		0.26***	1.18	2338.5	−5.18	0.89	1185.30***	22.59***	7.92
	MMR	2.00	25.25	0.22	0.00		2363.2	−5.24	0.87	1271.90***	25.13***	10.04
	GWR	0.02	0.17				4660.8	−10.38	0.72	84.28***	21.59**	3.67

Table 6.6 — continued

Sector	Model	$\sigma^2 \times 10^4$	$\sigma_\eta^2 \times 10^2$	$\sigma_\zeta^2 \times 10^2$	ϕ	$\bar{\beta}$	$\log L$	BIC	R^2	JB	Q(12)	LM(6)
Oil & Gas	RW	3.78	0.14				2242.6	−4.99	0.48	314.84***	26.11***	101.05***
	MR	3.24	12.76		0.44***	0.75	2249.5	−4.99	0.61	794.60***	22.84***	56.55***
	MMR	3.24	11.63	0.05	0.35***		2252.0	−4.99	0.60	802.55***	22.17***	56.92***
	GWR	0.02	0.09				4648.9	−10.35	0.52	10.94***	13.06	0.99
Personal	RW	1.53	0.08				2012.1	−5.87	0.77	758.40***	26.17***	100.02***
	MR	1.52	0.13		0.98**	0.95	2012.5	−5.85	0.78	732.66***	26.14***	101.70***
	MMR	1.36	3.52	0.07	0.04		2016.4	−5.87	0.82	427.06***	19.92***	90.20***
	GWR	0.02	0.04				3540.5	−10.35	0.78	54.93***	17.48**	23.13***
Retail	RW	2.98	0.78				1765.7	−5.15	0.68	85.79***	20.45***	15.24**
	MR	2.40	16.47		0.33**	0.90	1782.3	−5.18	0.78	72.60***	14.63**	3.08
	MMR	2.39	15.28	0.11	0.04		1785.2	−5.19	0.78	81.97***	15.98**	2.78
	GWR	0.02	0.13				3561.9	−10.41	0.63	71.40***	16.73*	4.78
Technology	RW	4.43	2.04				2128.8	−4.73	0.77	1426.00***	22.32***	46.33***
	MR	4.14	6.24		0.92***	1.48	2137.5	−4.74	0.79	1653.00***	22.24***	39.01***
	MMR	3.36	31.22	0.35	0.04		2162.2	−4.79	0.85	2917.00***	25.84***	21.99***
	GWR	0.02	0.26				4676.0	−10.41	0.74	69.99***	8.42	1.97
Telecom	RW	3.51	0.52				2258.2	−5.02	0.71	167.57***	19.67**	54.57***
	MR	3.40	1.55		0.95***	1.25	2262.6	−5.01	0.73	167.95***	21.36***	49.59***
	MMR	2.92	14.93	0.20	0.03		2274.9	−5.04	0.79	393.25***	18.00**	24.05***
	GWR	0.02	0.14				4662.5	−10.38	0.67	50.42***	5.39	9.96
Travel	RW	1.80	0.05				1960.8	−5.72	0.68	441.48***	16.13*	55.50***
	MR	1.36	10.91		0.04	0.75	1971.9	−5.74	0.80	416.05***	11.88	22.53***
	MMR	1.33	9.39	0.03	0.01		1982.3	−5.77	0.81	397.43***	13.07*	23.61***
	GWR	0.02	0.04				3537.5	−10.34	0.66	440.77***	24.13***	6.16
Utilities	RW	1.38	0.14				2682.9	−5.97	0.68	74.91***	20.21**	15.59***
	MR	1.35	0.37		0.97***	0.74	2685.5	−5.96	0.69	95.78***	21.89***	14.21**
	MMR	1.23	3.57	0.07	0.00		2691.4	−5.97	0.73	171.70***	23.10***	10.42
	GWR	0.02	0.07				4655.2	−10.36	0.66	35.79***	17.48**	4.48

be rejected at the 5% in five cases. A comparison of the resulting LM-statistics reveals that, with the exception of Personal & Household Goods, the test statistic usually takes on smaller values for the MR than for the RW model. With respect to the estimated goodness of fit, the coefficient of determination for the MR models is usually higher than for the RW models. However, the better fit comes at the cost of a less parsimonious model. For seven sectors, the Bayesian information criterion of the RW model is lower.

6.2.3.3 The moving mean reverting model

The moving mean reverting model is the third Kalman filter based specification to be considered. The MMR model has been demonstrated in §3.5.2.4 to generalize the MR by allowing the mean beta to vary over time. The state vector is given by

$$
\left[\begin{array}{c} \beta_{i,t+1} - \bar{\beta}_{i,t+1} \\ \bar{\beta}_{i,t+1} \end{array} \right] = \left[\begin{array}{cc} \phi_i & 0 \\ 0 & 1 \end{array} \right] \left[\begin{array}{c} \beta_{i,t} - \bar{\beta}_{i,t} \\ \bar{\beta}_{i,t} \end{array} \right] + \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right] \left[\begin{array}{c} \eta_{i,t} \\ \zeta_{i,t} \end{array} \right], \tag{6.21}
$$

with

$$
\left[\begin{array}{c} \eta_{i,t} \\ \zeta_{i,t} \end{array} \right] \sim N \left(\left[\begin{array}{c} 0 \\ 0 \end{array} \right], \left[\begin{array}{cc} \sigma_{\eta i}^2 & 0 \\ 0 & \sigma_{\zeta i}^2 \end{array} \right] \right), \tag{6.22}
$$

where estimates for the parameters $\sigma_{\epsilon i}^2$, $\sigma_{\eta i}^2$, $\sigma_{\zeta i}^2$ and ϕ_i have to be found. The various initial values needed to implement the Kalman filter for the different models are set as discussed in §3.5.3.

It can be seen from Table 6.6 that all estimated variance terms are significant at the 1% level. On average, the observation errors are less dispersed than in case of the MR model above. This comes at the cost of a higher variance in the state errors. For seven sectors, $\sigma_{\zeta i}^2$ is smaller than 0.001. In these cases the MMR model effectively comes close to being a MR model. With respect to the transition parameter ϕ_i, it is interesting to note that, with the exception of Chemicals, Financial Services and Oil & Gas, the estimates are not statistically significant at the 10% level. Taking this result together with the low values for $\sigma_{\zeta i}^2$, the corresponding MMR models de facto behave like random coefficient models.

As for the RW and MR models above, the JB-statistic is highly significant for all sectors. With the exception of Retail and Healthcare, the error terms are significantly autocorrelated. The LM-tests are comparable to those computed for the MR models. For six sectors, the null of no autocorrelation cannot be rejected. The reported values for R^2 signal a comparable fit of MMR and MR models. According to the information criteria, the MMR model is to be preferred for all sectors with the exception of Financial Services and Oil & Gas.

6.2.3.4 The generalized random walk model

The Kalman filter models considered so far are based on the assumption of normally distributed errors. As discussed earlier, the assumption of normality is incompatible with the sector series at hand. When the Kalman filter is employed in case of non-normal errors, the Kalman filter still leads to consistent QML estimators as outlined in §3.3.7. However, the obtained estimators are not efficient. As discussed in Chapter 5,

the two most important sources of non-Gaussianity are volatility clustering and outliers. As neither of the Kalman filter based approaches above is capable of fully coping with these issues, it might be sensible to remove these influences altogether. For this purpose, a three-stage estimation procedure following Ghysels et al. (1996) is developed:

1. Correct the observations for heteroskedasticity.

2. Cut down the remaining outliers.

3. Apply the Kalman filter based random walk model to the transformed observations.

In order to account for ARCH effects, less weight should be placed on data points with high conditional volatility. In the OLS context, this approach is referred to as *generalized least squares* (GLS). Following the outline by Davidson and MacKinnon (2004, §7), who give a general introduction to GLS and related topics, the concept of GLS can be summarized as follows: in a standard linear regression model of the form

$$y = X\beta + \epsilon, \qquad \epsilon \sim \text{IID}(\mathbf{0}, \boldsymbol{\Omega}), \tag{6.23}$$

the parameter vector $\boldsymbol{\beta}$ can only be estimated efficiently by least squares if the residuals are uncorrelated and homoskedastic. Whenever these assumptions are violated, an efficient GLS estimator of $\boldsymbol{\beta}$ can be found by appropriately transforming the regression so that the Gauss-Markov conditions are satisfied. The corresponding transformation depends on $\boldsymbol{\Psi}$, a quadratic matrix that is usually triangular with

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{\Psi}\boldsymbol{\Psi}'. \tag{6.24}$$

Premultiplication of (6.23) by $\boldsymbol{\Psi}'$ yields the transformed regression model that can be estimated by OLS to obtain efficient estimates:

$$\boldsymbol{\Psi}'y = \boldsymbol{\Psi}'X\beta + \boldsymbol{\Psi}'\epsilon. \tag{6.25}$$

The GLS estimator for $\boldsymbol{\beta}$ is given as

$$\hat{\boldsymbol{\beta}}^{GLS} = (X'\boldsymbol{\Psi}\boldsymbol{\Psi}'X)^{-1}X'\boldsymbol{\Psi}\boldsymbol{\Psi}'y = (X'\boldsymbol{\Omega}^{-1}X)^{-1}X'\boldsymbol{\Omega}^{-1}y. \tag{6.26}$$

In case of heteroskedastic but uncorrelated errors, the covariance matrix is diagonal and a GLS estimator can be obtained by means of *weighted least squares* (WLS). Each observation is weighted proportionally to the inverse of the nonconstant diagonal elements of $\boldsymbol{\Omega}$. With w_t^2 denoting the t-th element of $\boldsymbol{\Omega}$ and w_t^{-1} denoting the t-th element of $\boldsymbol{\Psi}$, for a typical observation at time t, the transformed regression model in (6.25) can be written as

$$w_t^{-1}y_t = w_t^{-1}X_t\beta + w_t^{-1}\epsilon_t. \tag{6.27}$$

The dependent and independent variables are simply multiplied by w_t^{-1}, where the weight observations depend negatively on the variance of the disturbance term. It can be shown that the variance of the disturbances is equal to unity.

As the precise form of the covariance matrix is usually unknown in practice, a consistent estimate of $\boldsymbol{\Omega}$ can be employed to get *feasible GLS* estimators. A common way to

correct for heteroskedasticity is to base w_t on a filtered or smoothed estimator of the conditional volatility, calculated by one of the methods discussed in Chapter 5. Even though this does not necessarily give the most efficient estimator in small samples, the feasible GLS estimator is usually more efficient than OLS in the presence of heteroskedasticity (cf. Ghysels et al. 1996).

The concept of WLS can be easily adapted to the random walk model discussed above. For the original state space model of the form

$$R_{i,t} = \beta_{i,t} R_{0,t} + \epsilon_{i,t}, \tag{6.28}$$

$$\beta_{i,t+1} = \beta_{i,t} + \eta_{i,t}, \tag{6.29}$$

the weighted transformation is

$$R_{i,t}/w_t = \beta_{i,t}^{WLS}(R_{0,t}/w_t) + \epsilon_{i,t}/w_t, \tag{6.30}$$

$$\beta_{i,t+1}^{WLS} = \beta_{i,t}^{WLS} + \eta_{i,t}^{WLS}, \tag{6.31}$$

with regressand $R_{i,t}/w_t$ and regressor $R_{0,t}/w_t$. Summary statistics are reported in terms of the transformed regressand to ensure orthogonality of the corresponding residuals.

In order to estimate the transformed model in (6.30) and (6.31) consistent estimates of the weighting factor are needed. In the following, w_t is set equal to the filtered conditional standard deviation estimated from an auxiliary heteroskedastic regression model. The behavior of $h_{i,t}$ is modeled by a t-GARCH(1,1) model:

$$R_{i,t} = \beta_i^* R_{0,t} + \epsilon_{i,t}^*, \tag{6.32}$$

$$\epsilon_{i,t}^* = z_{i,t}\sqrt{h_{i,t}}, \tag{6.33}$$

$$h_{i,t} = \omega_i + \gamma_i \epsilon_{i,t-1}^{*2} + \delta_i h_{i,t-1}, \tag{6.34}$$

where $z_{i,t}$ is assumed to be t-distributed. The weighting factor $\hat{w}_{i,t}$ can now be computed in annualized percentages as

$$\hat{w}_{i,t} = 100 * \sqrt{Ah_{i,t}}, \tag{6.35}$$

where the annualizing factor $A = 52$ is equal to the number of weekly returns per year (cf. Alexander 2001, §1.1). For all $t = 1, \ldots, T$, $\hat{w}_t$ depends on the information set Ω_{t-1}, which only contains relevant information through date $t - 1$. Therefore, the transformation based on filtered volatility estimated over the full sample is valid both for in-sample and out-of-sample purposes.

The second issue addressed above is related to outliers. For many estimation methods outliers can be captured straightforwardly by employing a t-distribution for the disturbance terms. The introduction of t-distributed errors to a Kalman filter model is not trivial as the resulting likelihood cannot be evaluated analytically. One way of handling non-Gaussian state space models is to employ importance sampling techniques as introduced in the context of SV models in §5.2.3; see Durbin and Koopman (2000, §11) for details. However, in order to limit the level of complexity, in the following outliers should be treated without relying on simulation-based techniques. A common way to remove the influence of outliers is to truncate the variables that enter the model. According to Granger et al. (2000) outlying observations can be identified as those data points that
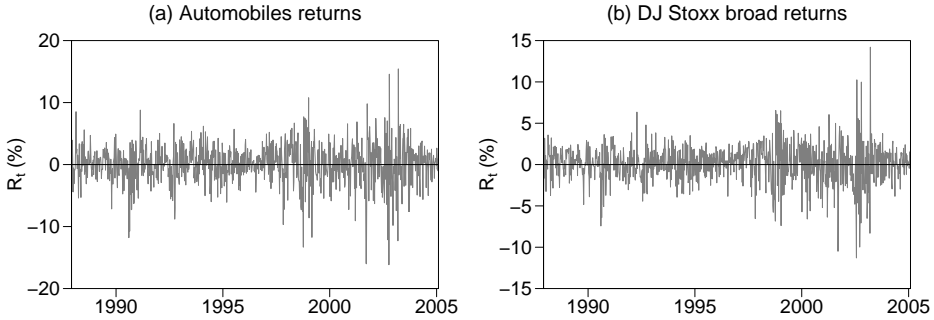
Figure 6.2: *Weekly excess log-return series of (a) Automobiles and (b) the broad market.*

are at least four standard deviations away from the mean of regressand and regressors, respectively. They propose to employ outlier-reduced series where any data point above or below $\hat{\mu} \pm 4\hat{\sigma}$ is replaced by $\hat{\mu} \pm 4\hat{\sigma}$; $\hat{\mu}$ and $\hat{\sigma}$ are estimated from the original data. In this thesis, more rigid limit lines at three standard deviations are imposed. Any outlying observation in (6.30) is simply set equal to the mean of y_t/w_t or x_t/w_t plus — or minus for negative outliers — three standard deviations. Obviously, unless outliers and the standard deviations are determined simultaneously, outliers will distort the standard deviations to be estimated. This leads to inflated sigmas and possibly masks influential observations. However, the objective of the "three sigma"-rule is not to fundamentally remove all outliers, but to approximately remove the biggest distractive effects from the data. Therefore, the chosen procedure can be considered being appropriate in the following.

The resulting model will be referred to as the *generalized random walk* (GRW) model. The estimates of conditional beta series are denoted as $\hat{\beta}_{i,t}^{GRW}$. Of course, other methods for dealing with non-normality are available. For example, in order to take conditional heteroskedasticity explicitly into account, Harvey et al. (1992) proposed a modified Kalman filter estimated by QML, and Kim (1993) developed a state space models with Markov switching heteroskedasticity. An alternative approach to deal with outliers is discussed by Judge et al. (1985, §20) who, instead of truncating the independent and dependent variables directly, truncate the residuals from a robust regression. While those techniques may be more sophisticated, in the following the methodology should be kept as simple and as relevant for practical implementation purposes as possible. Hence, the preference will be on the methodology described above. Even though this approach is simple, its relevance can be tested in a straightforward fashion: if the forecasting accuracy is superior in comparison to the standard RW model, then the proposed modifications can be regarded as being justified. Note that all reported diagnostics refer to the trimmed generalized input variables, while the error measures to be used in the next section to evaluate the forecast performances will be based on the original return series. This allows for a fair comparison of the different Kalman filter based models.
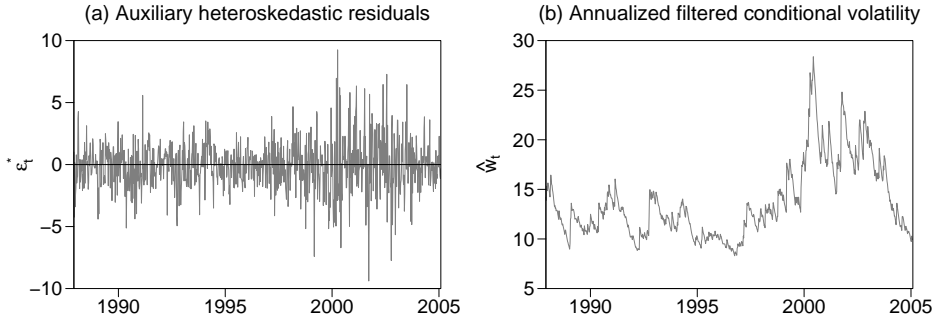
Figure 6.3: *(a) Residuals from the auxiliary heteroskedastic regression model and (b) GLS weighting factor for Automobiles and the overall market.*

The described procedure to deal with volatility clusters and outliers shall be illustrated at the example of the Automobiles sector. Figure 6.2 shows the weekly excess log-returns on the sector and the DJ Stoxx℠ Broad index. It is obvious that the volatility of both series is not constant over time. It can be seen that some periods are characterized by small absolute returns and others by large absolute returns. Especially in the second half of the period, which is effected by the Asian crisis, the Russian crises and the boom and bust of the new economy, some outlying absolute returns exceed the value of 10%. As outlined above, the influence of heteroskedasticity can be removed by means of weighted least squares. The weighting factor w_t can be derived by an auxiliary heteroskedastic regression model. Figure 6.3 displays the auxiliary residuals in Panel (a); Panel (b) shows the weighting factor computed as the annualized conditional t-GARCH volatility estimate.

The estimated conditional volatility confirms the impression of volatility clustering. It is used to transform the return series according to (6.30) to yield the weighted return series, which is plotted in Figure 6.4. As expected, the WLS transformation removed the heteroskedasticity in the series. What remains are a few outliers defined as those observations outside $\hat{\mu} \pm 3\hat{\sigma}$ that can now be capped (floored) according to the "three sigma"-rule. The trimmed generalized series can now be utilized as dependent and independent variables to estimate time-varying GRW beta series. Figure 6.5 illustrates the difference between the RW and GRW beta series for the Automobiles sector: the proposed procedure to deal with heteroskedasticity and outliers leads to a smoother conditional beta series whose major pattern remains intact.

The estimation results for all GRW models are summarized in Table 6.6. The estimated variance of observation and state disturbances are significant at the 1% level for all sectors. Although the null of normality can be rejected without exception, the reported JB-statistics are all significantly lower than for the Kalman filter based models considered above. The null of no autocorrelation can be rejected at the 5% level for eleven sectors. According to the reported LM-tests, the weighted transformation removed the volatility clusters for all sectors except for Personal & Household Goods.
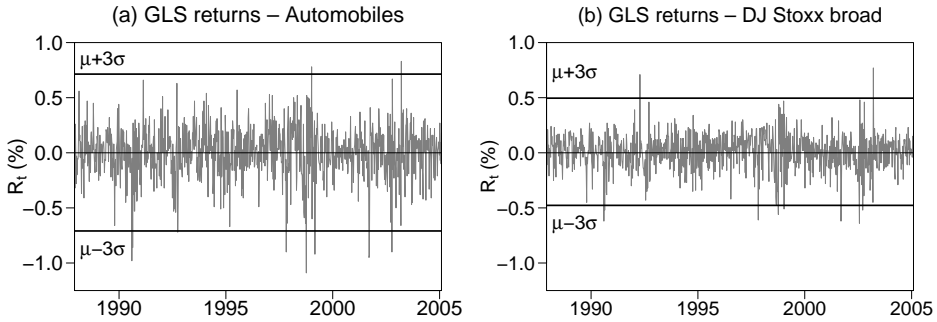
Figure 6.4: *Weighted weekly excess log-return series of (a) Automobiles and (b) the broad market.*

The reported values for R^2 and BIC should be interpreted with caution as they only refer to the second step of the proposed GRW approach. This is the major reason for not relying on fit statistics to compare and evaluate the in-sample performances of the various modeling techniques below.

6.2.4 Markov switching based approaches

The Markov regime switching approach outlined in Chapter 4 represents an alternative way to model time-varying betas. In contrast to the Kalman filter based specifications above, conditional betas switch between discrete states. The implicit assumption of regime switching models is that the observed data result from a process that undergoes abrupt changes, induced, for example, by political or environmental events.

In the Markov switching framework, the systematic risk of an asset is determined by the different regimes of beta, which are driven by an unobserved Markov chain. The switching behavior of beta is governed by the transition probability matrix $\boldsymbol{\Gamma}$. Under the assumption of a model with two states, $\boldsymbol{\Gamma}$ is of the form

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}. \tag{6.36}$$

The entries of each line describe the interaction of the two regimes from which beta is drawn: γ_{11} is the probability of staying in the first state from period t to period $t+1$, γ_{12} is the probability of switching from the first to the second state. The second row of the matrix $\boldsymbol{\Gamma}$ can be interpreted analogously.

In this thesis, two different Markov switching models are employed. The first approach is a simple *Markov switching* (MS) regression model. Let $\{s_1, \ldots, s_T\}$ denote the state sequence that represents the different regimes. Driven by the transition probability matrix of a stationary Markov chain, the states take values in $\{1, \ldots, m\}$. Following Huang (2000) the regime-switching CAPM is specified by

$$R_{i,t} = \alpha_{is_t} + \beta_{is_t} R_{0,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_{is_t}^2), \tag{6.37}$$
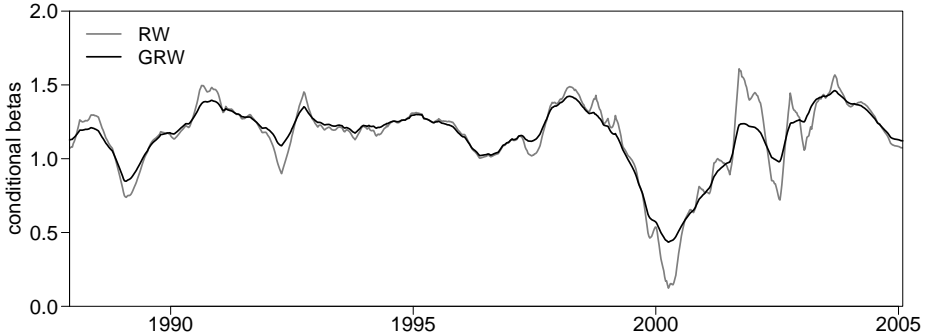
Figure 6.5: *Conditional random walk and generalized random walk beta estimates for the Automobiles sector.*

where the regression coefficients α_{is_t} and β_{is_t} depend on state s_t. Note that the model is designed to accommodate both the correlations across return series and the serial correlation of the individual series.

The second approach synchronizes the switching times of beta according to different market conditions by making additional assumptions on the market returns. This model will be referred to as *Markov switching market* (MSM) model. Rydén et al. (1998) showed that the temporal and distributional properties of daily return series can be well governed by a hidden Markov model with conditional normal distributions. Following their approach, the dynamics of the assets' returns also follow the regime-switching regression in (6.37). At the same time, the distribution of the market return is given by:

$$R_{0,t} = \mu_{s_t} + \eta_{s_t}, \quad \eta_{s_t} \sim N(0, \sigma^2_{0s_t}). \tag{6.38}$$

This means that in the MSM model the regime of the market changes together with the regime of the regression setup. Modeling both regimes as depending on the same state sequence allows for direct conclusions from the market condition on the asset's risk as represented by beta.

The estimation procedures for these two Markov switching models are based on the maximum likelihood method for hidden Markov models. The respective likelihood of both models is available in an explicit form such that the parameters of the models can be estimated directly by numerical maximization of the loglikelihood (cf. §4.3.2). The estimates for the model parameters include the state-dependent betas for each asset i and state j denoted by $\hat{\beta}^{MS}_{ij}$ or $\hat{\beta}^{MSM}_{ij}$. All computations were carried out using the statistical software package R, version 2.1.1 (R Development Core Team 2005); see Mergner and Bulla (2008) for details.

As mentioned above the state sequence is not directly observable. Information about the state-distribution at time t has to be derived in order to obtain in-sample estimates as well as out-of-sample forecasts of conditional betas. The desired probabilities of a sojourn in state j at time t can be computed by smoothing, filtering and state prediction

algorithms (cf. §4.4.2.2). Given the state-distribution at time t, estimates for the time-varying betas can be calculated by weighting the state-dependent $\hat{\beta}_{ij}^{MS}$ ($\hat{\beta}_{ij}^{MSM}$) with the probability of a sojourn in the corresponding state:

$$\hat{\beta}_{i,t}^{MS(MSM)} = \sum_{j=1}^{m} \Big[\beta_{ij} \cdot P(S_t = j | R_{0,1}, \ldots, R_{0,T}, R_{i,1}, \ldots, R_{i,T}) \Big], \qquad (6.39)$$

with

$$P(S_t = j | R_{0,1}, \ldots, R_{0,T}, R_{1,1}, \ldots, R_{1,T}) = \begin{cases} \dfrac{\alpha_t(j)\beta_t(j)}{L} & \text{for } 1 \le t \le T \\[2ex] \dfrac{\alpha_t(j)(\mathbf{\Gamma}^{t-T})_{\bullet j}}{L} & \text{for } T < t, \end{cases} \qquad (6.40)$$

where $\alpha_t(j)$ and $\beta_t(j)$ are the forward-backward probabilities introduced in §4.3.2.1, and $(\mathbf{\Gamma}^{t-T})_{\bullet j}$ denotes the j-th column of the matrix $\mathbf{\Gamma}^{t-T}$.

   The estimation results are summarized in Table 6.7. As expected, all alphas are very close to zero. For almost all sectors the high- and the low-risk states are well identifiable. However, the two state-dependent betas are quite close together in case of the MS model for the sectors Industrials and Retail, and in case of the MSM model for Industrials. Generally, the MSM model is characterized by a weaker separation of the two regimes: the state-dependent betas lie closer together than the betas of the corresponding MS model. This phenomenon can be explained by the lack of flexibility of the former model due to the enforced synchronous switching with the market regimes. The estimates for the expected market returns μ_1 and μ_2 of the MSM model are very close to zero. This supports Rydén et al. (1998) who proposed means equal to zero for daily return series. The estimates for γ_{11} and γ_{22} with values between 95% and 99% for most sectors show a high persistence both for the high- and the low-risk state. The observation of Fridman (1994) who reports lower persistence of the high-risk state cannot be confirmed.

Table 6.7: *Parameter estimates for Markov switching models.*

This table reports the estimated parameters for the MS/MSM models for the eighteen DJ Stoxx$^{\text{SM}}$ sectors.

Sector	Model	$\alpha_1; \alpha_2(\times10^4)$	$\beta_1; \beta_2$	$\sigma_{i1}^2; \sigma_{i2}^2(\times10^2)$	γ_{11}	γ_{22}	$\mu_1; \mu_2(\times10^3)$	$\sigma_{01}^2; \sigma_{02}^2(\times10^2)$
Automobiles	MS	$-13.4; -4.87$	1.26; 1.03	1.51; 2.89	0.993	0.980		
	MSM	$-3.05; -35.5$	1.22; 1.07	1.47; 2.92	0.976	0.926	2.40; -2.97	1.56; 3.70
Banks	MS	1.57; 8.70	1.01; 1.11	0.70; 1.77	0.988	0.970		
	MSM	2.02; 1.07	0.97; 1.10	0.67; 1.75	0.962	0.904	2.30; -2.02	1.58; 3.36
Basics	MS	$-3.26; 8.62$	1.05; 0.84	1.04; 2.64	0.998	0.998		
	MSM	$-3.90; 9.56$	1.02; 0.82	1.08; 2.95	0.972	0.917	$-2.58; -2.28$	1.61; 3.31
Chemicals	MS	1.70; -4.37	1.00; 0.87	0.93; 2.05	0.998	0.996		
	MSM	1.04; -12.6	0.98; 0.85	1.05; 2.38	0.978	0.901	2.65; -4.16	1.63; 3.70
Construction	MS	$-7.04; 0.22$	1.14; 0.77	0.92; 1.75	0.997	0.993		
	MSM	$-4.38; -4.11$	1.07; 0.79	0.94; 1.89	0.982	0.959	2.55; -3.26	1.55; 3.42
Financials	MS	$-0.61; -3.31$	0.91; 1.08	0.84; 1.82	0.997	0.990		
	MSM	$-0.05; 3.33$	0.91; 1.05	0.81; 1.89	0.969	0.883	2.52; -3.25	1.60; 3.62
Food	MS	$-0.95; 1.96$	0.91; 0.49	0.76; 2.23	0.993	0.984		
	MSM	$-1.48; -1.47$	0.89; 0.52	0.78; 2.34	0.985	0.928	2.59; -2.51	1.60; 3.33
Healthcare	MS	14.5; -0.88	0.91; 0.68	1.13; 2.39	0.981	0.971		
	MSM	10.0; 2.72	0.90; 0.73	1.16; 2.40	0.969	0.952	2.62; -1.31	1.51; 3.11
Industrials	MS	$-0.59; -7.62$	1.02; 0.96	0.62; 1.45	0.998	0.996		
	MSM	1.52; -1.41	1.00; 0.97	0.64; 1.57	0.975	0.951	2.24; -1.45	1.59; 3.38

Continued on the next page. . .

Table 6.7 — continued

Sector	Model	$\alpha_1; \alpha_2 (\times 10^4)$	$\beta_1; \beta_2$	$\sigma_{i1}^2; \sigma_{i2}^2 (\times 10^2)$	γ_{11}	γ_{22}	$\mu_1; \mu_2 (\times 10^3)$	$\sigma_{01}^2; \sigma_{02}^2 (\times 10^2)$
Insurance	MS	$-5.81; -8.25$	1.12; 1.37	1.01; 2.48	0.996	0.989		
	MSM	$-6.14; 1.58$	1.11; 1.39	1.05; 2.64	0.971	0.901	2.44; -2.93	1.64; 3.72
Media	MS	$-0.81; -12.4$	1.05; 1.41	1.20; 3.40	0.989	0.960		
	MSM	$-2.57; -1.51$	1.04; 1.32	1.21; 3.34	0.991	0.935	2.37; -4.21	1.65; 3.60
Oil & Gas	MS	6.02; 3.57	0.96; 0.58	1.37; 2.86	0.992	0.983		
	MSM	6.12; 1.92	0.91; 0.71	1.38; 2.94	0.987	0.926	2.20; -1.68	1.62; 3.45
Personal	MS	$-2.33; 6.06$	0.95; 0.80	1.03; 1.87	0.995	0.980		
	MSM	$-1.76; -5.45$	0.99; 0.85	1.72; 1.84	0.986	0.962	2.79; -4.19	1.72; 3.88
Retail	MS	2.29; -11.9	0.90; 0.97	1.12; 2.48	0.909	0.891		
	MSM	2.50; -19.6	0.87; 0.99	1.45; 2.83	0.967	0.872	1.68; -5.45	1.68; 4.04
Technology	MS	2.45; -1.61	1.18; 1.71	1.25; 3.85	0.996	0.990		
	MSM	3.67; -7.46	1.17; 1.60	1.25; 3.80	0.993	0.947	2.48; -2.35	1.62; 3.38
Telecom	MS	6.73; -8.51	1.09; 1.31	1.49; 3.24	0.997	0.989		
	MSM	3.29; -3.24	1.12; 1.21	1.49; 3.22	0.992	0.974	2.76; -4.10	1.69; 3.60
Travel	MS	0.05; -0.20	0.82; 0.50	1.09; 2.71	0.967	0.733		
	MSM	$-1.36; -8.04$	0.84; 0.73	1.05; 2.13	0.976	0.922	3.04; -5.01	1.67; 3.94
Utilities	MS	6.97; -8.05	0.82; 0.56	1.03; 1.63	0.994	0.977		
	MSM	7.07; -3.89	0.81; 0.63	1.03; 1.67	0.989	0.960	2.90; -5.67	1.59; 3.69

6.3   Analysis of empirical results

Following the methodological outline and the presentation of estimation results above, this section discusses the estimated conditional betas derived by the selected modeling approaches. The resulting beta series are compared in §6.3.1. The in-sample forecasting performance of the different modeling approaches is evaluated in §6.3.2. The respective out-of-sample predictive power is looked at in §6.3.3.

6.3.1   Comparison of conditional beta estimates

According to the discussed estimation results for the various modeling techniques, time-varying betas have been calculated for the eighteen pan-European DJ Stoxx℠ sector indices. All conditional beta series can be summarized by their respective mean and range. To avoid an unfair bias against the Kalman filter, which is likely to produce large outliers in the first stages of estimation, in the subsequent analyses the first fifty conditional beta estimates for any of the chosen modeling techniques will be excluded. Figure 6.6 utilizes boxplots to characterize the conditional betas for the example of the Insurance sector; for a complete summary covering all sectors, see Table C.1 in the appendix.
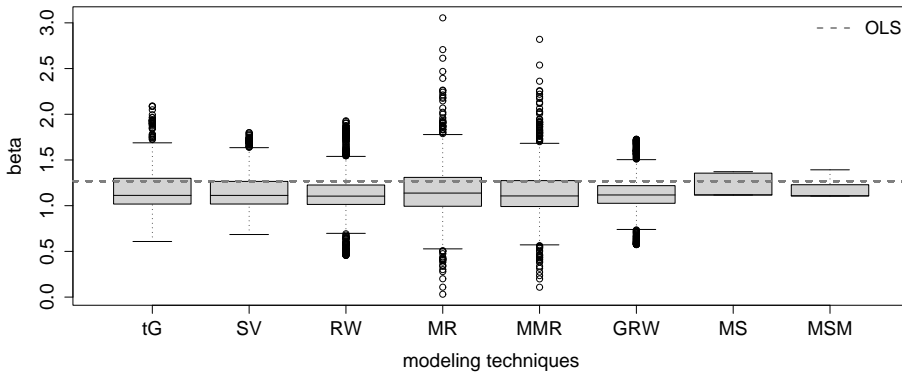
Figure 6.6: *Boxplots of the conditional beta series for the Insurance sector.*

Even though the mean conditional betas are usually close to their OLS peers, a wide range of mean betas can be observed for every individual sector. The widest beta range across sectors is observed for the two Kalman filter based MR and MMR models, followed by the t-GARCH and RW approaches. On the other end of the spectrum, the minimum and maximum of conditional betas estimated by the two Markov switching approaches do not deviate far from their respective mean. The evolution of conditional Insurance betas over time is displayed in Figures 6.7–6.9. They illustrate general similarities and differences between the alternative conditional beta series.
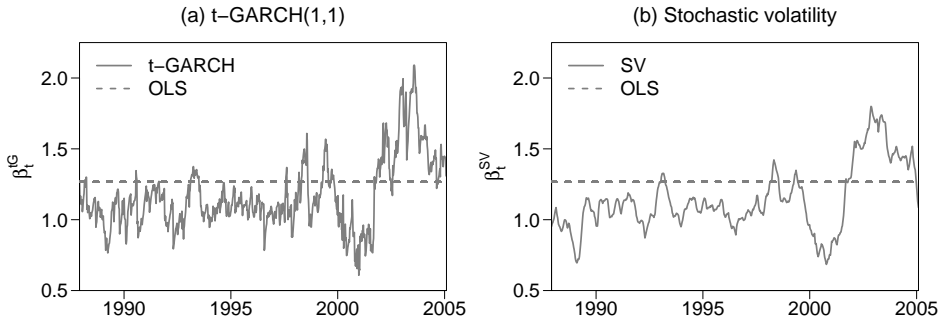
Figure 6.7: *t-GARCH and stochastic volatility based betas for the Insurance sector.*

As already indicated by the range of conditional betas, the Kalman filter and GARCH based techniques display the greatest variation. The time series of systematic risk exhibit the greatest amplitude when modeled either by the MMR or the MR model. They seem to be the techniques that are most flexible in capturing changes in a sector's sensitivity to the overall market over time.
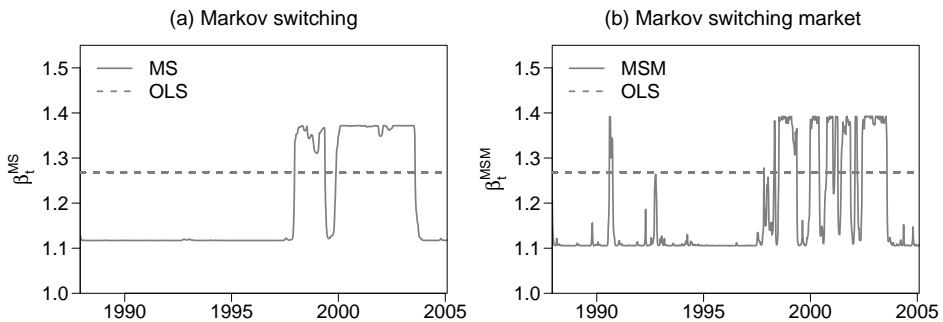
Figure 6.8: *Markov switching betas for the Insurance sector.*

With the exception of the Markov switching framework, the evolution of the Insurance beta during the stock market bubble at the end of the 1990s and its aftermath is described in a similar way by all techniques. During the twelve months period before the market peaked, a sharp fall of the Insurance beta below unity is indicated. In the subsequent two years the sector's beta more than doubles. The highest beta values are reached when either modeled by one of the two mean reverting approaches or by the GARCH framework. On the other hand, the Markov switching models are not able to reflect the developments and dramatic shifts in terms of exposure to market risk during the bubble period. The MSM model switches back and force between the different states and fails to give a clear direction of the sector's sensitivity to the overall market. Irrespective of the
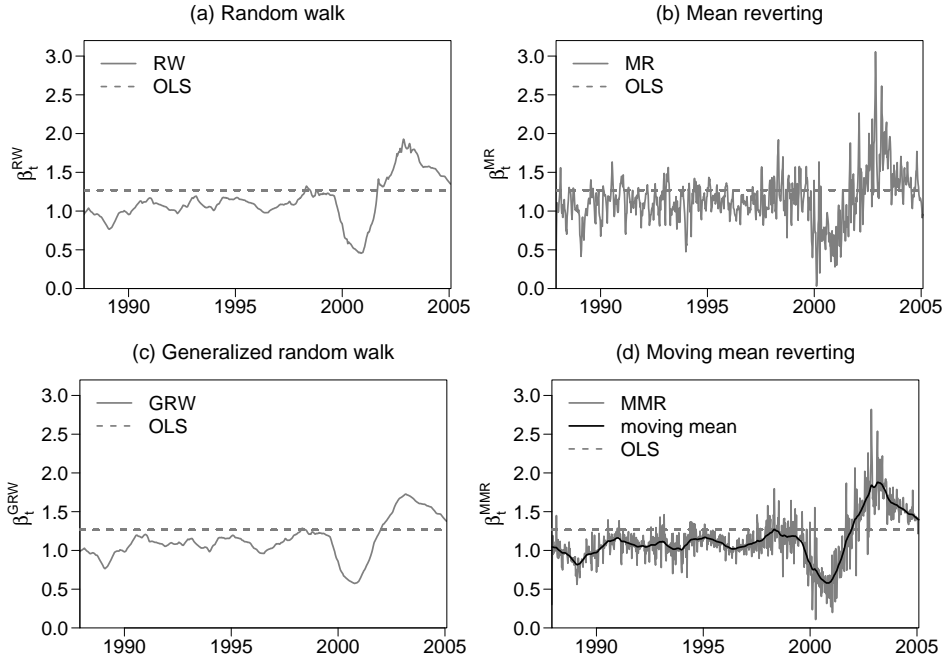
Figure 6.9: *Kalman filter betas for the Insurance sector.*

chosen modeling technique, the conditional beta of the Insurance sector tends to return to pre-bubble levels at the end of the sample period. Overall, similar patterns can also be observed for most other sectors as illustrated by Figures B.1–B.4 in the appendix.

The comparison of the different Kalman filter based betas for the Insurance sector illustrates that the RW and GRW series on the one hand side, and the MR and MMR series on the other hand side, show a very similar shape. As discussed in §6.2.3.4, the GRW beta can be thought of as a smoothed RW beta. For Kalman filter betas, the characteristics of the stochastic process of systematic risk depend strongly on the estimated values for the transition parameter: while ϕ_i is highly significant and close to 0.5 for the MR model, it is insignificant and close to zero for the MMR model. At the same time, the variance term σ_ζ is significantly different from zero. Hence, the MMR model turns out to behave like a random coefficient model that fluctuates randomly around a moving mean.

6.3.2   In-sample forecasting accuracy

The results above strongly indicate that systematic risk is not stationary and that the nature of the time-varying behavior of beta depends on the chosen modeling technique. In order to determine the relatively best measure of time-varying systematic risk, the quality of estimated conditional betas could be evaluated based on the goodness of fit

and diagnostic statistics reported in §6.2. However, in the context of this thesis at least two problems arise in connection with such an approach. The first issue is related to the residuals that are employed to calculate fit statistics. With the exception of the GARCH model, the fit statistics to be considered could either be calculated based on recursive residuals or generalized residuals (cf. §3.6.1). For a comparison with OLS, the generalized residuals should be used as these are also based on the full sample. However, for GARCH based models only recursive residuals are available. As a consequence, all derived fit statistics would employ less information than in case of the alternative modeling techniques. The GARCH based approach to conditional betas would be at a structural disadvantage. The second problem refers to the way time-varying betas are derived. While GARCH based betas are constructed indirectly, conditional betas derived by a state space model are calculated directly. As the resulting test statistics refer to different aspects of the respective models, they should only be employed for comparative purposes within the same modeling class.

To avoid these problems, in the following the different techniques are formally ranked based on their in-sample forecast performance. Following previous studies, the two main criteria used to evaluate and compare the respective in-sample forecasts are the *mean absolute error* (MAE) and the *mean squared error* (MSE):

$$MAE_i = \frac{1}{T} \sum_{t=1}^{T} |\hat{R}_{i,t} - R_{i,t}|, \tag{6.41}$$

$$MSE_i = \frac{1}{T} \sum_{t=1}^{T} (\hat{R}_{i,t} - R_{i,t})^2, \tag{6.42}$$

where T is the number of forecast observations; $\hat{R}_{i,t} = \hat{\beta}_{i,t} R_{0,t}$ denotes the series of return forecasts for sector i, calculated as the product of the conditional beta series estimated over the entire sample and the series of market returns. The latter is assumed to be known in advance, which is a commonly made assumption in the context of forecast evaluation. The forecast quality is inversely related to the size of these two error measures.

Figure 6.10 displays the average MAE and MSE measures across all sectors for the different modeling techniques on the left hand side. Panel (b) shows the respective average ranks of mean absolute and mean squared errors for each approach under consideration. For a more detailed sectoral breakdown, see Tables C.2 and C.3 in the appendix. A comparison of the different modeling techniques confirms the conjecture that the forecast performance of standard OLS is worse than for any time-varying technique. However, compared to the GARCH based techniques and the two Markov switching approaches, the degree of OLS' inferiority is remarkably low.[21] The MSE of OLS equals 2.89, which is only slightly higher than the MSE for the Markov switching model (2.78).

For the investigated sample, the two mean reverting techniques clearly outperform their competitors. The average mean squared error of 1.56 is nearly 50% lower than in case of OLS. With respect to both error measures, the MMR model ranks first in ten

---

[21]When interpreting the in-sample results, it should be remembered that, in contrast to all other techniques, the GARCH based betas are based on filtered instead of smoothed estimates.
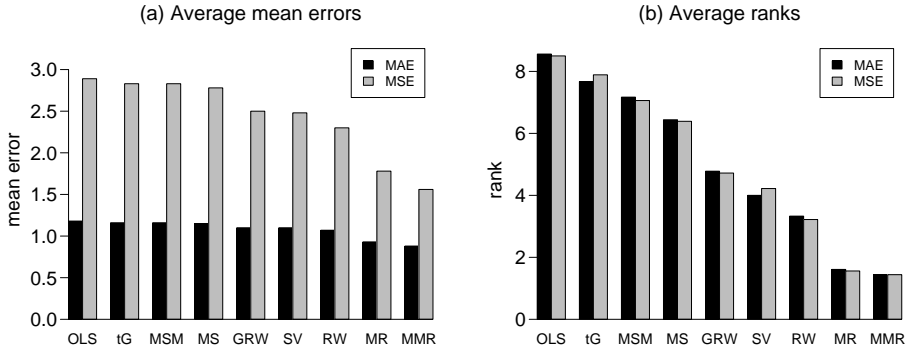
Figure 6.10: *In-sample forecasting evaluation: (a) average MAE and MSE across sectors and (b) average ranks across sectors.*

occasions and second in eight cases. With the exception of the MAE for Personal & Household Goods, each time the MMR only ranks second, the MR takes the top spot. The average rank of the mean absolute and mean squared errors for the RW model is equal to 3.3 and 3.2, respectively. Whenever the RW model does not rank third, it is usually outperformed by the SV model. The GRW model ranks behind the SV model. As the proposed generalization was intended not to capture every spike and to yield smoother conditional betas than the standard RW model, the comparably weak in-sample performance is not surprising. Within the class of volatility models, the SV approach seems to be better qualified to model the time-varying behavior of systematic risk than the well established GARCH model. On average, the MAE (MSE) for the SV model is 6% (13%) lower than the error measures for the GARCH based models, and 25% (59%) higher compared to the overall best model. Within the Markov switching framework, the MS betas lead to lower average errors than the MSM in case of fourteen sectors.

While the mean error criteria can be used to evaluate the average forecast performance over a specified period of time for each model and each sector individually, they do not allow for an analysis of forecast performances across sectors. From a practical perspective, it is interesting to see how close the rank order of forecasted sector returns corresponds to the order of realized sector returns at any time. *Spearman's rank correlation* coefficient, ρ_t^S, represents a non-parametric measure of correlation that can be used for ordinal variables in a cross-sectional context. It is introduced as the third evaluation criteria: after ranking the predicted and observed sector returns separately for each date, where the sector with the highest return ranks first, ρ_t^S can be computed as

$$\rho_t^S = 1 - \frac{6 \sum_{i=1}^{N_t} D_{i,t}^2}{N_t(N_t^2 - 1)}, \tag{6.43}$$

with $D_{i,t}$ being the difference between the corresponding ranks for each sector, and N_t being the number of pairs of sector ranks, each at time t. Figure 6.11 plots histograms
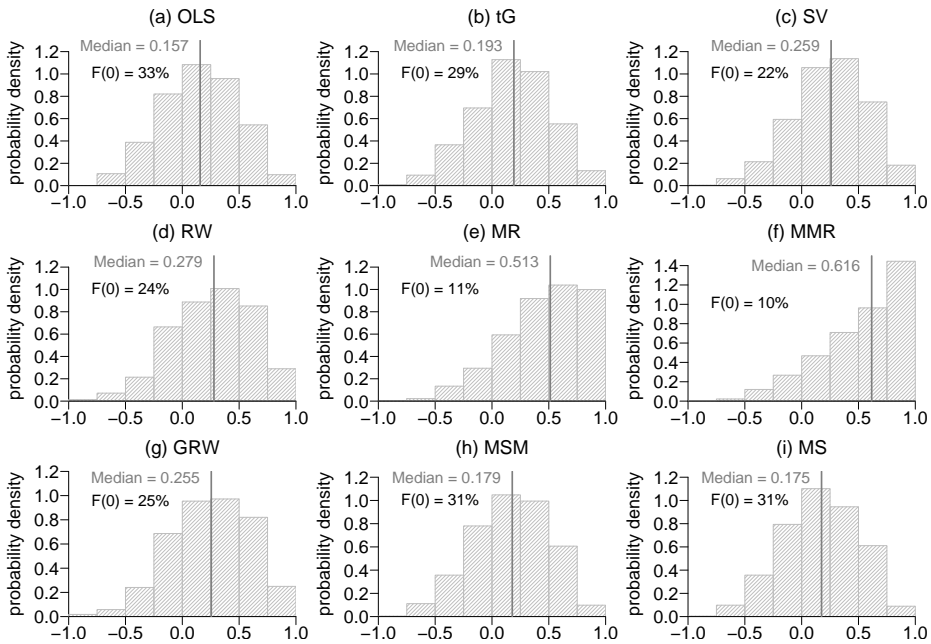
Figure 6.11: *Histograms of Spearman's in-sample rank correlations.*

of the in-sample rank correlations for the different modeling techniques together with their respective medians. The reported value of $F(0)$ denotes the proportion of rank correlations that are smaller than zero.

The highest medians of in-sample rank correlations are observed for the MMR and the MR models, where 50% of the computed rank correlations exceed the value of 0.616 and 0.513, respectively. For both models the distribution of rank correlations is negatively skewed, with the share of negative rank correlations being smaller than 10% (11%) for the MMR (MR) model. This confirms the finding gained from the analysis of mean errors above that these two models provide the best in-sample measures of time-varying betas. The next best results are observed for the RW, the SV and the GRW model. In contrast to the two mean reverting models, about a quarter of computed rank correlations is negative. The GARCH and Markov switching models do only slightly better than OLS.

Overall, the analysis of in-sample estimates suggests that time-varying European sector betas as modeled by one of the proposed Kalman filter approaches are superior to the considered alternatives. This is in line with previous findings presented by Brooks et al. (1998) and Faff et al. (2000) for industry portfolios in Australia and the UK.

6.3.3   Out-of-sample forecasting accuracy

The in-sample analysis is useful to assess the various techniques' ability to fit the data. However, as Wells (1996, p. 101) notes, "we should [...] be aware that because a

model gives a satisfactory description of historical phenomena does not [...] mean that it will perform as well on the future." Therefore, the indispensable extension to the model comparisons conducted above is to evaluate the respective out-of-sample forecast performances. In a first step, 100 beta and return forecasts based on 100 samples of 520 weekly observations are estimated for all proposed time-varying techniques. Within this rolling window forecasting procedure, the sample is rolled forward by one week, while the sample size is kept constant at 520. The first sample starts 24 March 1993 and ends 5 March 2003. It is used to calculate the out-of-sample conditional beta forecasts for 12 March 2003 based on the chosen modeling technique. The 100-th beta forecast for 2 February 2005 is then generated based on the last sample starting 15 February 1995 and ending 26 January 2005. The reason why the out-of-sample period is chosen to be limited to one hundred observations is to ensure proper convergence of the conditional volatility models. As will be seen from this analysis, the three Kalman filter based models RW, MMR and GRW prove to be superior in an out-of-sample context.

In a second step, an out-of-sample period of ten years based on 520 samples of 150 weekly observations is considered for these three models. As the Kalman filter does not require as many observations as the GARCH or stochastic volatility based models, an analysis based on an extended out-of-sample period is added. An estimation period of 150 weeks is chosen to allow for an out-of-sample period of 10 years for all eighteen sectors, including the three newly formed sectors for which only 683 instead of 897 weekly observations are available. An estimation period of 150 data points is somehow arbitrary and represents a compromise. According to Lo and MacKinlay (1997, §5.1) the most common choice to estimate the market model is to employ five years of monthly data. The current chapter deals with the forecastability of systematic risk at a more short-term horizon, which is reflected by the use of weekly data. Therefore, an estimation period roughly corresponding to three years of data is justified: while it contains enough data to generate stable parameter estimates, it is short enough to reflect current market conditions. Remembering that the Kalman filter naturally puts more weight on the most recent observations, the out-of-sample results for the Kalman filter models do not critically depend on the length of the estimation period.

The goal of this analysis is to identify the Kalman filter model that generates the best out-of-sample forecasts of the time-varying exposure to systematic risk over the more representative time horizon of ten years. The to be identified overall best approach will be chosen in the next chapter to model and forecast the time-varying relationship between macroeconomics and sector allocation.

6.3.3.1 Step I: Out-of-sample period of 100 weeks to compare all conditional modeling techniques

To compare the out-of-sample performances of all conditional betas, the average MAE and MSE measures across sectors and the respective average ranks are shown in Figure 6.12; the complete sectoral breakdown of out-of-sample error measures is provided in the appendix in Tables C.4 and C.5. Note that the MR model failed to converge for the Food & Beverages sector such that the out-of-sample mean error measures as well
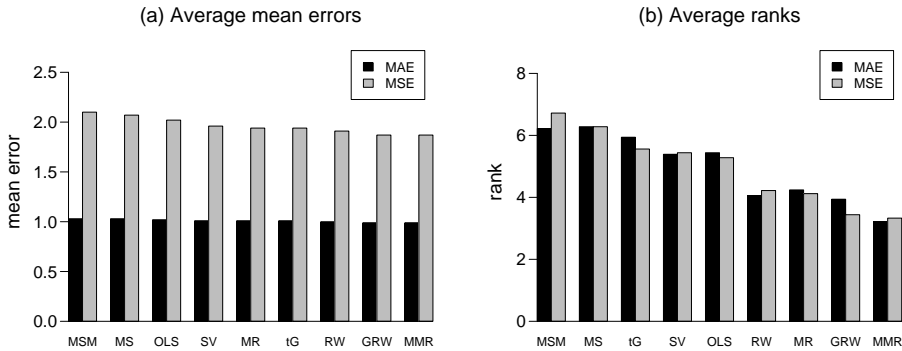
Figure 6.12: *Out-of-sample forecasting evaluation (100 samples): (a) average MAE and MSE across sectors and (b) average ranks across sectors.*

as the cross-sectional analysis for this modeling technique are based only on seventeen sectors.

According to the chosen error measures, the best out-of-sample forecast performances are offered by the GRW and the MMR models, for which the lowest average mean errors are observed. The MR model, which does very well in-sample, only yields the fifth lowest MSE. Within the class of conditional volatility models, no clear winner can be proclaimed as the GARCH and SV models approximately produce the same forecast errors. Disappointing forecast performances are observed for the two Markov switching models, which do even worse than standard OLS. While the average errors related to OLS are higher than for the volatility based techniques, the average relative ranks are even lower, being only inferior to the Kalman filter models. Generally, the average mean errors lie more closely together than for the in-sample results. The highest average MAE (MSE) as observed for the MSM model, the approach that offers the worst out-of-sample performance, is 4.5% (12.7%) higher than for the overall best model. Overall, the Kalman filter remains superior, even though it is not as dominant as it is in-sample: for both error measures, the average rank of the overall best model drops to around three, which compares to an average rank near one for the best technique in-sample. Only in five (six) occasions the MMR (GRW) model yields the lowest or second-lowest MSE.

The Kalman filter's superiority is also broadly confirmed in a cross-sectional setting utilizing Spearman's rank correlation coefficient. The histograms in Figure 6.13 show that all estimated medians are positive. The highest medians are observed for the RW and the MMR model, followed by the SV and the MR. The lowest medians are observed for the standard OLS and the two Markov switching models. Compared to the in-sample analysis, the difference between the median of the best model (RW: 0.298) and the overall worst model (MSM: 0.244) is smaller. The fraction of adverse signals is also broadly comparable, with values of F(0) around 25% for most of the different techniques. Outstanding in this context is the GRW model where only 14% of the computed rank
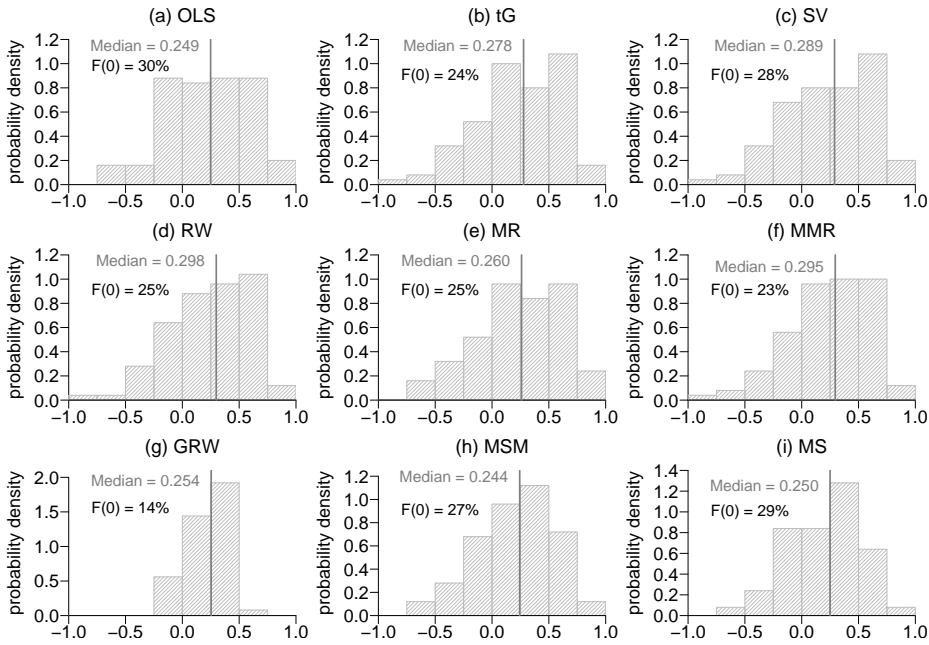
Figure 6.13: *Histograms of Spearman's out-of-sample rank correlations (100 samples).*

correlations are smaller than zero. This means that the risk of generating a misleading signal is reduced.

When evaluating a model, a common way to take the risk related to a forecast explicitly into account is to calculate an *information ratio* (IR). Alexander (2001, p. 445) defines an information ratio as "the mean prediction error divided by the standard deviation of the prediction error". In the context of the cross-sectional analysis using rank correlations, an information ratio for a given modeling techniques n can be defined as

$$IR_n = \frac{E(\rho_n^S)}{\sqrt{Var(\rho_n^S)}}, \tag{6.44}$$

where $\rho_n^S = \{\rho_{n,1}^S, \rho_{n,2}^S, \ldots, \rho_{n,T}^S\}$. The computed information ratios are reported in Table 6.8. For the chosen out-of-sample period, the GRW model is confirmed to offer the best risk-adjusted forecasting performance, followed by the MMR and the RW model. To check which of these three models yields the best forecasting performance over a more representative period of time, an out-of-sample period of ten years based on 520 samples

Table 6.8: *Information criteria of out-of-sample rank correlations.*

OLS	tG	SV	RW	MR	MMR	GRW	MSM	MS
0.63	0.65	0.64	0.70	0.64	0.72	1.42	0.62	0.65

of 150 weekly observations is considered in the second step of the proposed evaluation procedure.

6.3.3.2 Step II: Out-of-sample period of ten years to identify the overall best modeling approach

Table 6.9 summarizes the calculated mean absolute and squared errors for the RW, GRW and MMR model, respectively, by reporting their averages across all sectors; for a comprehensive sectoral breakdown, see Table C.6 in the appendix. According to the considered average mean errors, the out-of-sample forecasting performance of all three models over the last ten years of the given sample is very similar. Even though the average MAE is slightly lower for the MMR model and the average MSE is lowest for the RW model, based on these results neither approach significantly stands out. This is also confirmed by looking at the relative ranks, which indicate that the different modeling approaches yield an average rank of around two.

Table 6.9: *Average out-of-sample MAE and MSE across sectors (520 samples).*

	Mean absolute error ($\times 10^2$)			Mean squared error ($\times 10^4$)		
	RW	*GRW*	*MMR*	*RW*	*GRW*	*MMR*
Average error	1.311	1.310	1.309	3.584	3.600	3.589
Average rank	2.33	2.11	1.94	1.94	2.22	2.17

When Spearman's rank correlation coefficient is employed to evaluate the forecasting performance in a cross-sectional context, the RW model has a small advantage over its competitors. While Figure 6.14 illustrates that the realized rank correlations are similarly distributed for all three models, the median rank correlation as well as the information criteria of the RW model are slightly higher.

Overall, neither the less parsimonious moving mean reverting nor the generalized random walk model, which are both motivated by their respective capability to capture volatility clusters and outliers, yield a forecasting advantage over the random walk specification. This result suggests that heteroskedasticity and outliers can be considered as being "third-order" problems in the context of applying the Kalman filter to model the time-varying behavior of systematic risk for pan-European sector indices.

6.4   Concluding remarks

Despite the considerable empirical evidence that systematic risk is not constant over time, only few studies deal with the explicit modeling of the time-varying behavior of betas. Previous studies with a focus on Australia, India, New Zealand, the U.S. and the UK primarily employed Kalman filter and GARCH based techniques. The empirical analysis presented in this chapter contributes an investigation of time-varying betas for pan-European industry portfolios. The spectrum of modeling techniques is extended by (i) incorporating two Markov switching approaches, whose capabilities to model time-
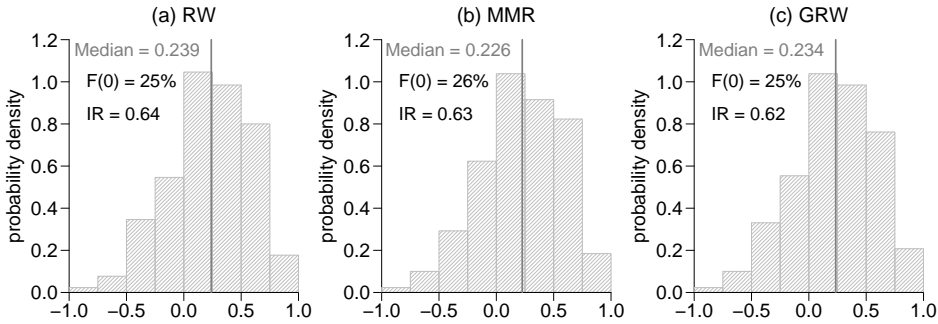
Figure 6.14: *Histograms of Spearman's out-of-sample rank correlations for the random walk, the moving mean reverting and the generalized random walk model (520 samples).*

varying betas have not been compared to the proposed alternatives in previous studies; (ii) by incorporating the stochastic volatility model, which so far has only been used by Li (2003) to model time-varying betas; and (iii) by proposing the generalized random walk model to deal with the common phenomena of heteroskedasticity and non-normality.

The in-sample forecast performances of the analyzed modeling approaches suggest that the extent to which sector returns can be explained by movements of the overall market is always higher in combination with time-varying betas. This confirms previous findings that sector betas are not stable over time. Based on the employed evaluation criteria, the in- and out-of-sample forecast performance of the various techniques has been compared. The results indicate that time-varying sector betas are best described by a random walk process estimated via the Kalman filter. The in-sample results generally support Kalman filter based approaches. In an out-of-sample context, where the relative advantage is less pronounced, their superiority is maintained only partially. The findings of Li (2003), according to which stochastic volatility based conditional betas outperform the other techniques, cannot be confirmed. The two proposed Markov switching models deliver an unsatisfactory out-of-sample predictive performance, which is even inferior to that of standard OLS.

It has been shown that the forecasting ability of the random walk model is not negatively affected by the influence of volatility clusters and outliers. The random walk specification outperforms all of the alternative Kalman filter based approaches, including the suggested generalized random walk model. The obtained quasi-maximum likelihood estimators are not only consistent, they also generate the best out-of-sample forecasts of conditional betas. As a consequence, the violation of the homoskedasticity and normality assumption can be neglected in the following. For the remainder of this thesis, time-varying factor sensitivities will be generally modeled as random walks estimated via the Kalman filter.

The methodology used in this study can be extended in a couple of directions. It would be of interest to see how the forecasting accuracy of the various models depends on the chosen length of the forecasting period. The performance of the Kalman filter might

be further improved by adding exogenous factors to explain the time-varying behavior of systematic risk. Some first steps into this direction have been made by Abell and Krueger (1989) and Andersen et al. (2005) who link betas to macroeconomics; Liodakis et al. (2003) use company fundamentals, momentum and liquidity data as determinants of time-varying betas. Another interesting path to continue research on conditional betas is to analyze the performance of employing multivariate conditional heteroskedasticity models with truly dynamic conditional correlations.

Chapter 7

A Kalman filter based conditional multifactor pricing model

Factor pricing models relate the risk of a portfolio to a possibly multidimensional set of common factors. They are widely employed by academics and investment professionals to analyze risk or to derive return predictions. The quality of a factor model depends on the selection of risk factors and on the estimated factor sensitivities. Using a Kalman filter based conditional multifactor pricing model, this chapter aims at analyzing the time-varying impact of fundamentals and macroeconomics on pan-European industry portfolios. Commercial portfolio risk models, such as BARRA (`www.barra.com`) or Citigroup's macro-based European Risk Attribute Model (Brennan et al. 2000), are usually based on single stock returns. They account for potential time-variation in factor sensitivities by assigning more weight to the most recent observations. The conditional modeling approach employed in this chapter is different in both aspects: (i) sectors instead of single stocks are used as dependent variables, and (ii) time-varying factor sensitivities are modeled explicitly as individual stochastic processes. A major advantage of building a factor model on sectors is the reduction of a potential errors-in-variables problem, which typically arises whenever independent variables, such as factor sensitivities in a cross-sectional setting, are measured with error. The proposed conditional multifactor pricing model can be applied to analyze current portfolio risks at the sector-level, to construct sector portfolios and to evaluate the past performance of an implemented sector strategy.

Given the results of the available literature on multifactor pricing models and the demonstrated significant relationship between macroeconomics, fundamentals and equity returns, this chapter focuses on the time-variation in macroeconomic and fundamental factor loadings of pan-European sectors. In the previous chapter, the ability of various modeling techniques to model the time-varying behavior of betas have been compared. As the results suggest that the stochastic process of systematic risk can be best described and predicted by a random walk process estimated via the Kalman filter, the same modeling technique is employed here. This is in contrast to the literature where conditional sensitivities are commonly captured either by allowing for an interaction term, by selecting different factors at different times, or by employing a discrete regime-

switching framework. In this thesis, time-varying sensitivities will be modeled directly as continuous stochastic random walk processes based on the state space framework introduced in Chapter 3.

The research goal of this chapter is to study the practical relevance of time-varying factor loadings in a multifactor pricing framework. Based on the conditional multiple beta series to be estimated via the Kalman filter, it will be analyzed whether an explicit consideration of the time-varying impact of macroeconomics and fundamentals on European industry portfolios can be exploited in a profitable way, either from a risk management or from a portfolio management perspective. The present chapter contributes to the literature a synthesis of the classical cross-sectional regression approach by Fama and MacBeth (1973) and conditional factor loadings.

The chapter is organized as follows. After briefly summarizing the concept of factor modeling, Section 7.1 reviews the anomalies literature with a focus on macroeconomic and fundamental factors. Section 7.2 outlines the conditional multifactor methodology. It will be differentiated between time series and cross-sectional regressions. Section 7.3 introduces the set of common risk factors. Section 7.4 discusses the empirical results and evaluates the relative out-of-sample forecasting ability of the proposed conditional multifactor specification. It will be analyzed whether the statistical superiority of the Kalman filter based model can be exploited in practice. Section 7.5 concludes.

7.1 Factor modeling

As outlined in §6.1 the most widely used factor model is the CAPM, in which the excess return of the overall market is chosen as the single common factor. The measure of systematic risk is referred to as market beta. Despite the CAPM's popularity, various studies over the last three decades have suggested that a single beta model, while describing a large portion of the common variation in returns, may not be sufficient to explain the cross-section of returns. The biggest challenge to the CAPM includes the empirical evidence that macroeconomic sources of risk and company-specific characteristics are priced beyond market risk. In accordance with these well-documented findings, the empirical deficiencies of the CAPM are most commonly explained by missing risk factors or by a misapproximation of the total wealth portfolio.[22] This leads to multifactor pricing models as motivated by the Intertemporal Capital Asset Pricing Model (ICAPM), introduced by Merton (1973), or the Arbitrage Pricing Theory (APT), developed by Ross (1976). The basic idea of a multifactor pricing model is that the common variation in asset returns can be accounted for by multiple common components, or risk factors.[23] Without explicitly differentiating between the ICAPM and the APT in the following,

---

[22]Alternatively, instead of explaining the violations of the CAPM using risk-based arguments, nonrisk-based explanations have been proposed. They include, among others, data-snooping biases, the existence of market frictions, transaction costs and liquidity effects. As nonrisk-based explanations will not find any consideration in this chapter, the reader is referred to MacKinlay (1995) for more details.

[23]An introduction to the basic conception of asset pricing theory is provided in Appendix A. As the theory on multifactor models is well-established, it is not intended to derive the underlying assumptions of the various multifactor pricing models in this thesis; for any details and

unconditional multifactor models, which describe the return generating process for asset i by linking returns to common sources of risk, can be specified in the general form of a *beta pricing model*:

$$r_{i,t} = \alpha_i + \sum_{k=1}^{K} \beta_{ik} f_{k,t} + \epsilon_{i,t}, \tag{7.1}$$

where β_{ik} is referred to as *factor loading* or *exposure*, i.e. the sensitivity of the i-th asset on the k-th risk factor, denoted by $f_{k,t}$ for $k = 1, \ldots, K$. The asset-specific intercept is given by α_i. The idiosyncratic components of the return series, captured by the disturbances $\epsilon_{i,t}$, are assumed to be uncorrelated with zero mean for $i = 1, \ldots, N$, and to be uncorrelated with the set of risk factors, each for all t. Whenever the set of common factors does not have mean zero, it is usually convenient to construct zero mean factors as

$$\tilde{\boldsymbol{f}}_t := \boldsymbol{f}_t - E(\boldsymbol{f}), \tag{7.2}$$

where $\boldsymbol{f}_t$ is the $K \times 1$ vector of risk factors (cf. Cochrane 2005, §9.4). In a world with existing risk-free assets, the constant term is usually assumed to be equal to the risk-free interest rate and (7.1) can be written in terms of excess returns, $R_{i,t}$, as defined in (2.2). With α_i set to zero, this can be represented in matrix notation as

$$R_{i,t} = \boldsymbol{\beta}_i' \boldsymbol{f}_t + \epsilon_{i,t}, \tag{7.3}$$

where $\boldsymbol{\beta}_i$ is a $K \times 1$ vector of factor loadings of asset i to the set of common factors.

A central derivation of the APT theorem is that in the absence of arbitrage, expected excess returns conditioned by the information set at date $t - 1$, Ω_{t-1}, depend approximately linear on the factor loadings:

$$E(\boldsymbol{R}_t | \Omega_{t-1}) \approx \boldsymbol{B} \boldsymbol{\lambda}_t, \tag{7.4}$$

where $\boldsymbol{R}_t$ is an $N \times 1$ vector of excess returns, $\boldsymbol{B}$ is an $N \times K$ matrix of factor loadings and $\boldsymbol{\lambda}_t$ is a $K \times 1$ vector of *factor risk premia*. The risk premium can be interpreted as the compensation for owning one unit of the k-th factor risk. In the following, we align ourselves with the majority of the literature and assume that (7.4) holds exactly:

$$E(\boldsymbol{R}_t | \Omega_{t-1}) = \boldsymbol{B} \boldsymbol{\lambda}_t. \tag{7.5}$$

By testing whether risk premia are significantly different from zero, one can determine which risk factors are priced by the market. As indicated by the chosen notation, $\boldsymbol{R}_t$ and $\boldsymbol{\lambda}_t$ are usually considered for a given date t, while $\boldsymbol{B}$ is assumed to remain constant over time.

Despite its intuitive appeal, the theory on multifactor models does not prescribe which and how many factors should be included. Answers to these questions can be derived from the so-called *anomalies literature*. It has emerged in the course of empirical works

---

further references, see one of the many excellent texts on the subject, for example, Fama and French (1996), Campbell et al. (1997, §6) or Cochrane (2005, §9).

testing the validity of the CAPM. Various empirical findings have demonstrated the existence of factors with explanatory power for the cross-section of returns beyond market beta risk. This section provides a brief literature review on the different factor types that can be categorized into *macroeconomic*, *fundamental*, *momentum* and *statistical* factors.

7.1.1   Factor taxonomy

Macroeconomic variables constitute the first set of potential risk factors that can be assumed to be rewarded by the market. They are intended to capture the state of the economy or to forecast future economic conditions. Commonly employed macroeconomic factors include interest rates, production growth, consumer confidence, credit spreads, steepness of the yield curve and shifts in energy prices. Macroeconomic factor models with multiple betas, where each beta relates an asset to a particular economic risk, allow fund managers to gain top-down insights into how their portfolios are affected by different economic scenarios. Alternatively, fundamental factor models assume that sensitivities to firm characteristics such as the price-earnings ratio (PE), leverage or size are capable of explaining the cross-section of returns. Although it is not yet clear which systematic risks are approximated by fundamental factors, this second type of model has been very successful empirically. Employing the third type of factors, momentum models are based on the empirical finding that past return patterns may offer an indication of future returns. In contrast to those factor categories, statistically derived factors are not observable and have to be inferred from the return data using statistical factor selection procedures.

7.1.1.1   Macroeconomic factors

The possibility that macroeconomic factors may successfully predict security returns has spawned a remarkable bulk of literature that analyzes whether stock and bond returns can be predicted using macroeconomic variables. One of the best known studies is that of Chen et al. (1986). In an APT framework, the authors implement expected and unexpected inflation, industrial production, the spread between short- and long-term interest rates and the default premium, defined as the yield spread between high and low rated bonds. The chosen risk factors are found to be significantly priced. The predictive power of the default premium has been confirmed, among others, by Fama and French (1989) and Keim and Stambaugh (1986). The results presented by Campbell (1987) imply that excess stock returns are predicted by the state of the term structure of interest rates. A further indication of the importance of interest rates as well as their volatility is provided by Shanken (1990). In an analysis of the source of predictability of monthly stock and bond returns, Ferson and Harvey (1991) look at a set of state variable proxies: the value-weighted New York Stock Exchange index return less the 1-month Treasury-bill return, per capita growth of personal consumption expenditures, unanticipated inflation, the yield spread between Baa-rated corporate bonds and a long-term government bond, the change in the slope of the yield curve, the real 1-month Treasury-bill return and the dividend yield on the S&P 500. In a later study on risk and

predictability of international equity returns (Ferson and Harvey 1993) the authors add a dollar and an oil price factor. Jones and Kaul (1996) explicitly document the impact of changes in the oil price on the stock market. In another application, Chan et al. (1998) use the growth rate of monthly industrial production, the default premium, the real interest rate defined as the difference between the return on one-month Treasury-bills and the relative change in the monthly consumer price inflation, the slope of the yield curve, the change in the monthly expected inflation and the maturity premium, defined as return difference between long-term government bonds and the one-month Treasury-bill rate, as macroeconomic variables. The authors conclude that only the default and the maturity premium are significantly related to stock returns. In a work on maximized predictability in stock and bond markets in the US, Lo and MacKinlay (1997) rely on the dividend yield, the default spread, the maturity spread, the return on the S&P 500 and an interest-rate trend, calculated as the change of average yields on a long-term government bond. More recently, Lettau and Ludvigson (2001) successfully employ the log consumption-wealth ratio as a conditioning factor.

7.1.1.2 Fundamental factors

The second category of factors is related to firm-specific attributes. Various empirical studies have illustrated that it is possible to earn risk-adjusted returns by constructing portfolios in accordance with fundamental factors. Basu (1977) finds the *PE effect*: firms with low PEs have higher sample returns and firms with high PEs have lower sample returns than can be expected in the context of a mean-variance efficient market portfolio. Banz (1981) documents the *size effect* with higher than expected returns for firms with a small market capitalization. Bhandari (1988) documents a positive relationship between average returns and leverage. Rosenberg et al. (1985) report the so-called *value premium*, where the average returns are positively related to the book-to-market equity ratio, which is defined as a company's book value (BV) to its market value (MV). At the beginning of the 1990s, Chan et al. (1991) confirms the value premium also for Japanese equities. Subsequent studies, see, for example, Fama and French (1993, 1995, 1998), Lakonishok et al. (1994) and Daniel and Titman (1997), gave further confirmation of the book-to-market anomaly and tried to find different explanations for the value premium. In today's portfolio management industry, the most important investment style is based on the value premium: a *value investor* invests in firms with the highest book-to-market ratios, which means investing in the relatively cheapest value companies.

Fama and French (1992, 1996) developed a more comprehensive framework. Instead of conducting individual analyses for the various anomalies, they take the interdependencies between the different variables explicitly into account. They analyze the empirical relationships between the expected return of a stock, its beta and other fundamentals such as size, book-to-market equity, leverage and earnings-price ratios. Their work is considered a milestone as they interpret the combination of different variables as a multidimensional measure for risk. The most widely used fundamental multifactor model, which dominates today's empirical research, is the *three-factor model* by Fama and French (1993). It explains the cross-section of expected returns by three factors: a market proxy, size and the book-to-market ratio. Even though a solid economic theory

explaining which non-diversifiable risks are proxied by size and book-to-market still has to be developed, these two factors explain average returns better than the theoretically easier to justify macroeconomic factors (cf. Cochrane 2005, §20).

7.1.1.3   Momentum and reversal

Research on momentum and reversal strategies has started with the works of DeBondt and Thaler (1985, 1987) who report *price reversals* over the long-term (three to five years) where stock prices overreact and eventually mean-revert. The authors demonstrate that a portfolio that is long past losers and short past long-term winners, yields a better performance than a portfolio constructed in the opposite way. Jegadeesh (1990) finds that price reversals also occur over the very short term (one week to one month). Over the medium term (three to twelve months) Jegadeesh and Titman (1993, 2001) and Chan et al. (1996) document *price momentum* effects: past winners continue to outperform past losers.

Long-term price reversals are consistent with the Fama-French three-factor model, because stocks that do poorly over a long time horizon build up a value premium. By contrast, short-term reversals and price momentum cannot be explained by the Fama-French model. Some authors relate momentum profits to microstructure explanations such as calendar or illiquidity effects; see Cochrane (2005, §20) for an overview. However, once transaction costs are taken into account, momentum and reversal strategies, which both require frequent trading, are not exploitable as shown by Carhart (1997). Therefore, given these non-supportive arguments together with the findings of Grundy and Martin (2001), who conclude that momentum strategies are even less attractive when applied to sectors instead of single stocks, momentum and reversal effects will not find any consideration in the following.

7.1.1.4   Statistical factors

An alternative to employ macroeconomic or fundamental variables is to derive the factors statistically. Instead of using observable real world variables, the factors are inferred from the return data by applying statistical factor selection procedures. The two primary approaches are a two-step factor analysis and principal components; see Campbell et al. (1997, §6) for a summary. Statistical factor models yield the advantage that no external explanatory variables are required and that they provide an answer to the number of unknown factors. Moreover, one does not have to deal with problems related to multicollinearity as the statistically derived factors are usually orthogonal. However, despite their attractive in-sample features, their out-of-sample performance is usually poor (cf. Chan et al. 1998). Besides, purely statistical factors do not offer an economic interpretation. In this thesis statistically derived factors will not be considered as risk factors.

7.1.2   Number of factors

The multifactor model proposed in (7.3) does not yet specify K, the number of required factors. In order to avoid overfitting and to arrive at a specification of maximum practical

usefulness, K should generally be as small as possible. The employment of fewer factors increases the chance that the results hold in- and also out-of-sample (cf. Chan et al. 1998). In empirical work, the lack of specification of K is usually handled by repeated modeling with different numbers of factors, where the sensitivity of any results to increasing K is observed. Alternatively, the adequacy of opting for K explanatory variables can be tested explicitly by employing a likelihood ratio test, for example. Overall, various studies have suggested that the adequate number of factors should be between four and six, which serves as a guideline in this chapter; cf. Campbell et al. (1997, §6.4.2) for an overview and further references.

7.1.3   Time-varying factor loadings

The central notion throughout this thesis is that in an inherently dynamic world, where expected returns and risk premia vary over the business cycle, as demonstrated, among others, by Keim and Stambaugh (1986), Fama and French (1989) and Ferson and Harvey (1991), factor sensitivities should also be time-varying. This implies conditional versions of (7.3) and (7.4), respectively.

According to Shanken (1990) many authors have traditionally modeled time-varying multiple sensitivities as linear functions of observable, possibly macroeconomic state variables. For example, Lo and MacKinlay (1997) use an interaction term, constructed as the product of the dividend yield and the return on a market proxy, to capture time-variation in market betas. An alternative route is taken by Pesaran and Timmermann (1995) who find that, due to the conditional nature of return volatility, the predictive ability of particular macroeconomic factors is changing over time. The authors propose a recursive forecasting model with a base set of explanatory variables from which a model with possibly different factors is specified at each date. A third way of addressing sensitivities that evolve through time is to permit for a nonlinear relation between risk factors and returns. This is illustrated by González-Rivera (1997) who integrate the conditional APT with betas modeled as random coefficients. In a more recent work on the predictability of U.S. sectors, Johnson and Sakoulis (2003) employ a Kalman filter with Bayesian parameter estimation to model the time-varying link between returns and lagged macroeconomic information as following a random walk.

As demonstrated in the previous chapter on time-varying market beta risk, a wide spectrum of econometric techniques to model conditional sensitivities is available. While each approach has its own merits, the empirical results presented above suggest that time-varying systematic risk for pan-European industry portfolios can be best modeled and predicted by employing a Kalman filter based random walk model. In addition to the superior in- and out-of-sample forecasting performance of the random walk specification, the computational burden related to Kalman filter estimation is substantially lower than in case of stochastic volatility or hidden Markov models. This is true for the estimation of parameters and also for the generation of out-of-sample forecasts of time-varying sensitivities: within a state space framework one-step ahead predictions are readily available. The Kalman filter's computational advantage represents a major determinant of practical usefulness and implementability. Considering these properties together with the superior forecasting performance of the random walk specification, the time-varying

coefficients of the multifactor model in this chapter will be modeled as individual random walks estimated via the Kalman filter.

7.2   Specification of a conditional multifactor risk model

Following the review of the standard unconditional multifactor pricing framework, this subsection derives a conditional multifactor risk model, on which the subsequent empirical analysis will be based. Subsection 7.2.1 outlines the conditional time series representation. Subsection 7.2.2 summarizes how to estimate factor risk premia in a cross-sectional setting and how to conduct inferences about them.

7.2.1   Time series representation

The starting point for modeling the time-varying impact of macroeconomics and fundamentals on pan-European sector allocation is the general multifactor beta pricing model as given by (7.3). A multifactor model for the realized excess returns $R_{i,t}$ with time-varying betas can be stated in state space form with observation equation

$$R_{i,t} = \boldsymbol{\beta}'_{i,t}\boldsymbol{f}_t + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim N(0, \sigma_i^2), \tag{7.6}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$; $\boldsymbol{f}_t$ and $\boldsymbol{\beta}_{i,t}$ are the $K \times 1$ vectors of risk factors and corresponding factor loadings, respectively; $\epsilon_{i,t}$ is the vector of normally distributed disturbances with unconditional variance σ_i^2. The factor realizations are assumed to be stationary with unconditional moments

$$E(\boldsymbol{f}_t) = \mathbf{0}, \tag{7.7}$$
$$Cov(\boldsymbol{f}_t) = \boldsymbol{\Omega}_f, \tag{7.8}$$

and to be uncorrelated with the error terms:

$$Cov(f_{k,t}, \epsilon_{i,t}) = 0, \tag{7.9}$$

for all i, k and t.

In accordance with (3.77) the evolution of the k factor loadings $\beta_{ik,t}$ can be modeled as individual random walks by the following set of K state equations:

$$\begin{aligned}
\beta_{i1,t+1} &= \beta_{i1,t} + \eta_{i1,t}, & \eta_{i1,t} &\sim N(0, \sigma_{\eta i1}^2), \\
\beta_{i2,t+1} &= \beta_{i2,t} + \eta_{i2,t}, & \eta_{i2,t} &\sim N(0, \sigma_{\eta i2}^2), \\
&\vdots \\
\beta_{iK,t+1} &= \beta_{iK,t} + \eta_{iK,t}, & \eta_{iK,t} &\sim N(0, \sigma_{\eta iK}^2).
\end{aligned} \tag{7.10}$$

The system of equations (7.6)–(7.10) is a special case of the general state space framework presented in §3.2. The assumptions made in (3.3)–(3.7) apply. The constant variances σ_i^2 and $\sigma_{\eta ik}^2$ represent the $K + 1$ unknown hyperparameters of the system, which can be estimated by means of maximum likelihood as discussed in §3.4. The path of time-varying factor sensitivities can be tracked and predicted by the Kalman filter and smoother as outlined in §3.3.

7.2.2 Cross-sectional regressions

The ability of the proposed time-varying multifactor specification to predict the cross-sectional variation of pan-European industry portfolios is tested by employing the cross-sectional regression approach proposed by Fama and MacBeth (1973). Instead of using time series averages to conduct a single cross-sectional regression, the basic idea of the Fama-MacBeth approach is to regress the cross-section of returns for each time period on a set of predetermined factor loadings. In a second step, inference about the estimated risk premia is conducted to analyze the explanatory power of the corresponding risk factors. For regressors that do not vary over time, the Fama-MacBeth procedure can be shown to be equivalent (i) to a single cross-sectional regression that is based on sample averages, and (ii) to a pooled approach with stacked time series and cross-sections. In contrast to these alternative procedures, the Fama-MacBeth methodology can easily handle time-varying right-hand variables (cf. Cochrane 2005, §12.3). This qualifies the approach as the preferred cross-sectional regression procedure in the following. As a novelty, the procedure outlined below combines the traditional Fama-MacBeth approach with the paradigm of changing betas.

7.2.2.1 The Fama-MacBeth approach

In accordance with Ferson and Harvey (1991) the Fama-MacBeth approach is implemented in two steps. At first, time series instruments for the betas are estimated for each asset. While the instruments are usually derived by means of standard OLS, in this chapter OLS betas are replaced by the Kalman filter based random walk specification as proposed in (7.6)–(7.10). This leads to a Fama-MacBeth approach with conditional betas. The set of estimated time-varying factor sensitivities is taken as an estimator of the matrix of factor loadings $\boldsymbol{B}$ in (7.5), denoted as $\hat{\boldsymbol{B}}$. In the second step, the cross-section of ex-post excess returns is regressed for each date t on the matrix of ex-ante instruments to obtain overall T estimates of the $K \times 1$ vector of risk premia, $\boldsymbol{\lambda}_t$:

$$\boldsymbol{R}_t = \hat{\boldsymbol{B}}_{t-1}\boldsymbol{\lambda}_t + \boldsymbol{\nu}_t, \tag{7.11}$$

where $\boldsymbol{\nu}_t$ is the $N \times 1$ vector of cross-sectional disturbances. Due to the use of excess returns, the above equation does not contain an intercept term. The employed dating convention indicates that the betas are time-varying and based only on information available at date $t-1$. The regression in (7.11) allows for a decomposition of excess returns for each t. The term $\hat{\boldsymbol{B}}_{t-1}\boldsymbol{\lambda}_t$ refers to the component that traces back to the cross-sectional risk structure as measured by the conditional betas. This is the part of the returns that is predictable. The remainder term $\boldsymbol{\nu}_t$ is not correlated with the measures of risk and should thus be unpredictable.

A risk factor is priced if the expected value of the corresponding ex-post risk premium is significantly different from zero. Following common practice, the significance of the involved risk factors is tested by defining $\lambda_k = E(\lambda_{k,t})$ for $k = 1, \ldots, K$, and employing a usual t-test with test statistic

$$w(\hat{\lambda}_k) = \frac{\hat{\lambda}_k}{\hat{\sigma}_{\lambda_k}}, \tag{7.12}$$

with mean

$$\hat{\lambda}_k = \frac{1}{T} \sum_{t=1}^{T} \hat{\lambda}_{k,t}, \tag{7.13}$$

and standard error

$$\hat{\sigma}_{\lambda_k}^2 = \frac{1}{T(T-1)} \sum_{t=1}^{T} (\hat{\lambda}_{k,t} - \hat{\lambda}_k)^2, \tag{7.14}$$

where $w(\hat{\lambda}_k)$ is Student-t distributed with $T-1$ degrees of freedom (cf. Campbell et al. 1997, §5.8).

7.2.2.2   Econometric issues

Several problems related to the methodology may arise. The conditional betas in (7.11) are unknown and have to be estimated in a first step based on a time series regression. This induces an errors-in-variables problem. Secondly, even if the true betas were known, the errors of the proposed cross-sectional regression would be likely to be heteroskedastic and correlated, because the stock returns that are used as dependent variables are usually correlated for a given date t. As a consequence, the estimated standard errors might be unreliable (cf. Ferson and Harvey 1991).

The empirical evidence presented by Ferson and Harvey (1999) suggests that the main results are robust to errors-in-variables. Nevertheless, different ways to explicitly cope with the issue have been proposed. Shanken (1992) presents a general procedure that corrects for errors in the estimates of conditional betas by using adjusted standard errors. Alternatively, better estimates of the conditional betas can be produced by using returns of portfolios rather than of single stocks (cf. Chen et al. 1986). As this thesis exclusively deals with industry portfolios, the impact of a potential errors-in-variables problem can be assumed to be of limited nature. The second problem of heteroskedastic and correlated errors is especially alarming in small samples where the t-ratios might be biased. One way to address this is to employ an efficient generalized least squares estimator where the T estimates are weighted according their reciprocal variances; see Ferson and Harvey (1999) for details. In the following, the analysis will be conducted based on ten years of weekly data. The sample is sufficiently large to assume that the corresponding t-ratios will be unbiased. Therefore, throughout this chapter, the widely used unadjusted standard errors of the Fama-MacBeth approach will be relied upon.

7.3   The risk factors

The weekly excess returns on the eighteen DJ Stoxx℠ sector indices introduced in §2.1 constitute the set of dependent variables to be studied. This section describes the risk factors to be used in the context of the proposed conditional multifactor model.

According to the literature review provided in §7.1, various fundamental factors, in particular size and valuation, as well as different proxies for macroeconomic risks, which can be thought of having an impact on equity returns, have been previously analyzed. As this thesis focuses on weekly data, the spectrum of available macroeconomic series is limited. Many factors that are typically employed in empirical research, such as growth

Table 7.1: *Glossary and definition of risk factors.*

Symbol	Definition	Mnemonic[a]
TS	Change in the euro term structure defined as the difference between the yield of a 10-year German government bond and a money market rate (assembled series: until 18.11.92 German 3-month rate; thereafter Euro 3-month rate)	BDBRYLD; GERMDRQ; EMINT3M
OIL	Log-return of the crude oil-brent 1-month forward contract	OILBRNI
FX	Log-return of the synthetic US-dollar to euro exchange rate	USEURWD
VGS	Difference between log-returns of the chosen value and growth indices (assembled series: until 2.7.97 Frank Russell 1000 Value and Growth indices; thereafter DJ Stoxx℠ TMI Value and Growth indices)	FRUS1VA; FRUS1GR; DJTSVAE; DJTSGRE
SIZ	Difference between log-returns of the DJ Stoxx℠ TMI Large 200 index and the DJ Stoxx℠ TMI Small 200 index	DJSLARG; DJSSMAL
BMR	Residual excess benchmark return calculated from a time-varying regression of the log-return of the DJ Stoxx℠ Broad return index in excess of the risk-free rate, calculated from the 3-month FIBOR, on the set of macroeconomic and fundamental factors	DJSTOXX; FIBOR3M

[a]This column provides the codes used to access the corresponding data series via Thomson Financial Datastream.

in gross domestic product, unexpected inflation or consumer and industrial confidence indicators, are only surveyed monthly or quarterly. Table 7.1 lists the variables to be employed as risk factors in this study. They are all assumed to capture underlying exogenous influences that have an impact on all sectors. The focus in this chapter is on the modeling of risk using conditional factor loadings and not on factor selection procedures. Hence the factors have been chosen from widely tested variables that can be justified theoretically and that have been successfully employed in the literature in a similar context, as reviewed in §7.1.1. Following Ferson and Harvey (1991) who used a similar set of risk factors to study predictable components of stock and bond returns, it cannot be claimed that these factors uniquely pick up all relevant fundamental and macroeconomic risks. These variables could well represent a joint approximation for a set of unobserved variables that influence asset returns. However, the chosen factors are theoretically appealing. They each have been previously documented to capture different aspects of systematic risk.

In order to work with a set of independent variables of the same length, the beginning of the sample period is set to the earliest date for which historical pricing data for all eighteen DJ Stoxx℠ sectors is available (as outlined in §2.1, Stoxx Ltd. redefined its

sector classification in 2004, which led to the introduction of three new sectors whose histories are shorter than for the remaining market segments). The sample spans a period of 683 weekly excess return observations from 8 January 1992 to 2 February 2005. All sector returns, as well as the proposed macroeconomic and fundamental factor series, are obtained from Thomson Financial Datastream.

7.3.1 Macroeconomic risk variables

The choice of macroeconomic factors is based on previous empirical findings, which demonstrate that the corresponding risk premia are significantly different from zero. The theoretical reasons behind the selection of macroeconomic factors can be summarized as follows.

7.3.1.1 European term structure

The term structure represents one of the most widely used variables to reflect systematic changes in the state of the economy. It is usually defined as the weekly change in the difference between the yield of a 10-year Treasury bond and a 3-month Treasury bill. This factor is chosen to pick up risks that are reflected by a changing shape of the yield curve, as proposed, among others, by Chen et al. (1986), Ferson and Harvey (1991) or Johnson and Sakoulis (2003). Changes in the slope of the yield curve capture both changes in the set of investment opportunities and changes in inflation expectations, which are partly driven by growth expectations. With a focus on pan-European industry portfolios in this thesis, the term structure variable (TS) is derived by European interests rates. These are approximated by the yield of a 10-year German government bond and a European money market rate. Until 18 November 1992 the latter is approximated by the German money market rate, thereafter by the 3-month Euro rate.

The term structure variable is highly correlated both with changes in German bond yields ($\rho = 0.60$) and changes in European money market rates ($\rho = -0.64$). To avoid problems related to multicollinearity, TS is the only interest rate variable that is included in the set of employed risk factors.

7.3.1.2 Oil price

Given the global economy's dependence on oil as a major energy resource, fluctuations in the price of oil can be expected to have an effect on security prices, as shown, for example, by Jones and Kaul (1996). Hence, the oil price constitutes another factor that is typically included in the set of systematic risk factors (see, for example, Chen et al. 1986; Hamao 1988). The OIL series used here is defined as weekly differences in the logarithm of the crude oil-brent 1-month forward contract.

7.3.1.3 Dollar

Exchange rate risks have been successfully documented to be priced in international security markets. The results of Dumas and Solnik (1995), among others, support the existence of a premium for foreign exchange risks. Even though the euro was not

introduced as a common European currency until 1999, markets started to price the effect of a common exchange rate much earlier. This anticipation was mainly driven by the economic convergence across Europe (cf. Brennan et al. 2000). Therefore, a synthetic dollar to euro time series is chosen as common risk factor, where FX denotes the series of corresponding log-returns.

7.3.2   Fundamental risk variables

The selection of fundamental factors is motivated by the Fama and French (1993) three-factor model. Based on the concept of factor mimicking portfolios, the Fama-French model incorporates, in addition to a market factor, a size factor as well as a factor that measures the spread between value and growth. The idea behind mimicking portfolios is to construct portfolios in a way such that their returns mimic a specific factor. For example, a mimicking portfolio for the size factor can be obtained by sorting all eligible stocks by their market capitalization in a first step. Based on its respective rank each stock is assigned to one of five portfolios, where the smallest and largest stocks form portfolios 1 and 5, respectively. The return spread between the highest and lowest ranked portfolio represents the return of the size mimicking portfolio, which can be employed as a proxy for a common size factor (cf. Chan et al. 1998). Instead of calculating proprietary factor mimicking portfolios, return differences between corresponding indices, calculated by professional index providers such as Stoxx Ltd. or MSCI, for example, can be relied upon as respective proxies.

7.3.2.1   Size

The size factor (SIZ) to be used in the following, is defined as the difference between the log-returns of the DJ Stoxx℠ Total Market Index (TMI) Large 200 index and the DJ Stoxx℠ TMI Small 200 index. Stoxx Ltd. publishes three so-called *size indices*. The Large 200 index covers the 200 largest stocks as ranked by free-float market capitalization of the DJ Stoxx℠ TMI universe. The index covers 95% of the total market capitalization of all stocks that are traded on the main exchanges in the UK, Switzerland, Denmark, Iceland, Norway, Sweden and the Eurozone ex Slovenia. The Medium 200 index covers the next-largest 200 companies, and the Small 200 index again the next-largest 200 companies of the DJ Stoxx℠ TMI universe. All indices are weighted by the free-float market capitalization of the index constituents. The index composition as well as the weights are reviewed quarterly; for more details, see Stoxx Ltd. (2007).

7.3.2.2   Value-growth-spread

Analogously to the size factor, the value-growth risk factor (VGS) is derived as the difference between the log-returns of the DJ Stoxx℠ TMI Value and the DJ Stoxx℠ TMI Growth index. Stoxx Ltd. offers these two so-called *style indices*, which are also based on the DJ Stoxx℠ TMI universe. The style characteristics of a stock depend on six factors: projected and trailing price to earnings ratio, projected and trailing earnings growth, price to book ratio and dividend yield. The index composition is reviewed semi-

annually with the weightings being reviewed quarterly; see Stoxx Ltd. (2005) for more details.

As the history of these two indices only goes back to 30 June 1997, VGS is an assembled series: prior to inception of the two DJ Stoxx℠ style indices, the value-growth-spread is proxied by the difference between the log-returns of the Frank Russell 1000 Value and the Frank Russell 1000 Growth index. Comparable to the DJ Stoxx℠ style indices for European stock markets, the objective of the two Frank Russell indices is to give a comprehensive picture of value and growth stocks for a universe of the 1000 U.S. common stocks with the highest market capitalization. The style index membership is determined by ranking each stock in the Russell 1000 universe by their respective price to book ratio and their projected long-term growth. The indices are reconstituted on a yearly basis; more details on the methodology are provided by Russell Investment Group (2007).

Following this procedure, the value and growth indices, from which the time series of the VGS factor is derived, refer to different regions. However, during the period for which weekly style returns are available for both European and U.S. markets, the correlation between the respective value (growth) returns is as high as 0.76 (0.73). Therefore, under the assumption that style performance in different regions — particularly in the U.S. and Europe — is highly correlated, the chosen procedure can be considered feasible.

7.3.3 The market factor

Given the importance of the market portfolio as a measure of aggregate wealth as indicated by the CAPM (cf. §6), a market factor is typically included as a central risk factor in the context of a multifactor pricing model. Even though the goal of this chapter is to analyze the time-varying impact of macroeconomic and fundamental factors on European industry portfolios, the consideration of a non-macroeconomic, non-fundamental equity market factor is economically plausible: as stock prices respond very quickly to new information, aggregate market returns can be expected to reflect additional information that is not yet captured by the factors above (cf. Chen et al. 1986).

As illustrated in Table 7.2, the excess log-returns of the DJ Stoxx℠ Broad index, which serve as a proxy for the overall market throughout this thesis, are highly correlated with some of the chosen macroeconomic and fundamental risk factors. A common approach to assure that the set of considered risk factors is reasonably independent is to define a so-called *residual market factor*. This factor measures market returns after eliminating the effects of the macroeconomic and fundamental factors. Brennan et al. (2000), for example, regress the series of market returns on the set of the remaining risk factors by using standard OLS. The obtained disturbance terms are then used as the residual market factor.

As a means to avoid any look-ahead bias, in this chapter the residual benchmark factor, BMR, is derived as the series of one-step ahead prediction errors. These are obtained from an auxiliary regression with time-varying coefficients modeled as individual random walks (cf. §3.6.1.1). In order to use the Kalman filter for estimation purposes, the

auxiliary regression is stated in state space form with observation equation

$$
\begin{aligned}
R_{0,t} &= \beta_{01,t}VGS_t + \beta_{02,t}SIZ_t + \beta_{03,t}TS_t + \beta_{04,t}OIL_t + \beta_{05,t}FX_t + \epsilon_{0,t}, \\
&\quad \epsilon_{0,t} \sim N(0,\sigma_0^2),
\end{aligned} \tag{7.15}
$$

where $R_{0,t}$ denotes excess returns on the DJ Stoxx℠ Broad benchmark index. The set of five state equations is given by

$$
\beta_{0k,t+1} = \beta_{0k,t} + \eta_{0k,t}, \qquad \eta_{0k,t} \sim N(0,\sigma_{\eta 0k}^2), \tag{7.16}
$$

for $k = 1,\ldots,5$. In accordance with the definition of one-step ahead prediction errors, $v_t^\dagger$, given in (3.83), the residual benchmark series is defined as

$$
BMR_t := v_t^\dagger, \tag{7.17}
$$

for $t = d + k + 1,\ldots,T$, where d refers to the number of diffuse state elements. In order to receive a series BMR that is of length T, an extended history of 735 observations with starting date 9 January 1991 is employed for estimating the system of equations (7.15)–(7.16). That way, with $t = -51,\ldots,683$, the series of one-step-ahead prediction errors obtained in (7.17) contains $735 - k - d$ elements of which the last 683 are taken as residual market factor series BMR. The estimation results are summarized in Table 7.3. As illustrated by the estimated R^2, only 57% of the variation of the excess benchmark returns are explained by the chosen macroeconomic and fundamental factors. The remaining unexplained information is captured by the generated BMR series.

7.3.4 Summary statistics

Summary statistics for the set of chosen risk factors are given in Table 7.4. As illustrated by the second column, with the exception of the size factor, all reported mean values are not significantly different zero. For the size factor, the null hypothesis of a zero mean can be rejected at the 5% significance level. For reasons of parsimony, without having to construct zero mean factors as defined in (7.2), the first moment assumption made in (7.7) is considered to hold at least approximately. According to the estimated annualized standard deviations, the macroeconomic factors are on average more volatile than the residual benchmark and fundamental factors. In columns 4–9, the risk factors' autocorrelations for different lags are displayed; the last column shows the test statistic for the Ljung-Box test of no autocorrelation. While the BMR factor displays elevated autocorrelations for the first two lags, the two fundamental factors show elevated autocorrelations at the first three lags. For all three equity market related factors, the Q-test

Table 7.2: *Correlations between the DJ Stoxx℠ Broad index and the chosen macroeconomic and fundamental risk factors (8.1.1992–2.2.2005).*

	VGS	SIZ	TS	OIL	FX
DJ Stoxx℠ Broad	-0.27	0.56	0.10	0.01	0.40

Table 7.3: *Parameter estimates for auxiliary regression (9.1.1991–2.2.2005).*

This table reports the estimated parameters for the auxiliary Kalman filter regression used to generate the series of the residual market factor BMR. All parameters and test statistics are significant at the 1% level.

$\sigma_0^2 \times 10^4$	$\sigma_{\eta 01}^2 \times 10^4$	$\sigma_{\eta 02}^2 \times 10^4$	$\sigma_{\eta 03}^2 \times 10^4$	$\sigma_{\eta 04}^2 \times 10^4$	$\sigma_{\eta 05}^2 \times 10^4$
2.72	12.23	32.84	1.07	0.15	10.84
$\log L$	BIC	R^2	JB^a	$Q(12)^b$	$LM(6)^c$
1906.1	-5.13	0.57	252.81	36.25	19.74

$^a JB$-statistics is the Jarque-Bera statistic for testing normality. The relevant critical values at the 95% (99%) level are 5.99 (9.21).

$^b Q(l)$ is the test statistic of the Ljung-Box portmanteau test for the null hypothesis of no autocorrelation in the errors up to order l. In a structural model, the Q-statistic is asymptotically χ^2 distributed with $l - w - 1$ degrees of freedom where w denotes the total number of estimated parameters (Harvey 1989, p. 259).

$^c LM(l)$ is the LM-statistic of Engle's ARCH test for the null hypothesis of no ARCH effects up to order l. The test statistic is asymptotically χ^2 distributed with l degrees of freedom. The relevant critical values at the 95% (99%) level are 12.59 (16.81).

indicates significant autocorrelation up to order 12. In contrast, neither the oil price nor the exchange rate factor exhibit signs of serial autocorrelation. The result for the term structure factor is mixed: while no autocorrelation is detected at low lag orders, the Q(12)-statistic indicates significant autocorrelation at the 5% level. Overall, these results are in line with the levels of autocorrelation found in similar sets of risk factors, as documented, for example, by Chen et al. (1986) or Ferson and Harvey (1993).

An important aspect in selecting the explanatory variables is to ensure that they do not systematically move together. The absence of multicollinearity is a prerequisite to determine the individual contributions of the various factors to overall risk, and to estimate the unknown parameters with the required degree of precision. A common rule of thumb for ruling out the presence of severe multicollinearity is to check whether the R^2 from the estimated model is greater than the individual correlations between the set of regressors (cf. Alexander 2001, §A.4.1). Table 7.5 displays contemporaneous correlations between the risk factors for different time periods. Panel A covers to full sample from 8 January 1992 to 2 February 2005. The remaining panels refer to four subperiods. The first break takes place 15 February 1995. This date represents the end of the formation period and the beginning of the out-of-sample period, which is chosen to be of length 520 observations or ten years. The out-of-sample period itself is broken into three subsamples according to different market regimes: Panel C reaches to the end of the *dotcom* bubble, which took off at the end of the 1990s; Panel D covers the subsequent bear market whose end is marked by the low point in the DJ Stoxx℠ Broad as observed on 12 March 2003; and Panel E refers to the market recovery that began thereafter.

The risk factors are only mildly correlated both over the entire sample and across subperiods. For most factors, the pairwise correlation is around 0.1 or lower. For

Table 7.4: *Summary statistics for the set of risk factors (8.1.1992–2.2.2005).*

	μ^a	σ^b	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_{12}	$Q(12)^c$
BMR	0.03 (0.68)	13.14	0.14	0.13	0.02	0.02	0.03	-0.08	33.45 (0.01)
VGS	0.03 (0.52)	9.73	0.03	0.10	0.13	0.03	0.02	0.04	81.54 (0.00)
SIZ	0.11 (0.04)	9.80	-0.11	0.11	0.15	0.02	-0.01	0.06	46.70 (0.00)
TS	0.71 (0.14)	90.29	0.00	0.03	0.11	0.00	0.02	0.02	25.09 (0.01)
OIL	0.14 (0.39)	29.51	-0.02	0.04	0.03	-0.03	0.03	0.03	12.98 (0.37)
FX	-0.01 (0.82)	10.06	0.01	-0.02	0.00	0.05	0.08	-0.08	14.39 (0.28)

   [a]The mean is expressed in percentage terms. Figures in parentheses denote p-values for a simple t-test of the null hypothesis of a zero mean.
   [b]The standard deviation is expressed in annualized percentage terms.
   [c]Q(l) is the test statistic of the Ljung-Box portmanteau test for the null hypothesis of no autocorrelation in the errors up to order l. The Q-statistic is asymptotically χ^2 distributed with l degrees of freedom. Figures in parentheses denote p-values.

the full sample, the highest correlations occur between SIZ and FX (-0.204) and between SIZ and VGS (-0.227). This indicates that the outperformance of small caps is associated (i) with a falling dollar, and (ii) with value outperforming growth. The highest correlation coefficients are measured for Panel E, which is based on fewer observations than the other subperiods. During this time, the exchange rate factor's correlation with the value-growth-spread (-0.370) and the term structure (-0.385) is elevated. Overall, the observed pairwise correlations indicate that the selected risk factors are far from being excessively collinear, which suggests that none is dispensable.

7.4   Empirical results

Employing the risk factors defined above together with the conditional multifactor model given by (7.6)–(7.10) implies the following data-generating process for the set of excess sector returns, $R_{i,t}$. The observation equation can be written as

$$
\begin{aligned}
R_{i,t} &= \beta_{i1,t}BMR_t + \beta_{i2,t}VGS_t + \beta_{i3,t}SIZ_t + \beta_{i4,t}TS_t + \beta_{i5,t}OIL_t \\
&\quad + \beta_{i6,t}FX_t + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim N(0, \sigma_i^2),
\end{aligned} \tag{7.18}
$$

for $i = 1, \ldots, 18$. The state equation is given as

$$
\boldsymbol{\beta}_{i,t+1} = \boldsymbol{\beta}_{i,t} + \boldsymbol{\eta}_{i,t}, \qquad \boldsymbol{\eta}_{i,t} \sim N(\mathbf{0}, \boldsymbol{Q}), \tag{7.19}
$$

where $\boldsymbol{\beta}_{i,t+1}$ and $\boldsymbol{\eta}_{i,t}$ are 6×1 vectors of states and state disturbances, respectively; $\boldsymbol{Q} = \text{diag}(\sigma_{\eta i1}^2, \ldots, \sigma_{\eta i6}^2)$ is the variance term of state errors. With reference to (7.11) we are ultimately interested in the factor loadings $\boldsymbol{\beta}_{i,t-1}$, which are based only on information available at date $t-1$. Hence, filtered state estimates as defined in (3.13) serve as time series estimates of time-varying factor loadings.

   In order to assess the relative superiority of the proposed dynamic specification with conditional factor loadings modeled as individual random walks, three alternative models

Table 7.5: *Risk factor cross-correlations.*

This table reports the estimated contemporaneous correlations between the risk factors for different time periods. Panel A covers to full sample from 8 January 1992 to 2 February 2005. The remaining panels refer to four subperiods: the first break takes place at 15 February 1995 representing the end of the formation period and the beginning of the out-of-sample period (520 obsrevations). The out-of-sample period itself is broken into three subsamples according to different market regimes: Panel C reaches to end of the TMT bubble, Panel D covers the subsequent bear market, and Panel E refers to the market recovery that began thereafter.

Symbol	BMR	VGS	SIZ	TS	OIL	FX
A. 8 January 1992–2 February 2005 ($t = 1, \ldots, 683$)						
BMR	1.000					
VGS	−0.024	1.000				
SIZ	−0.023	−0.227	1.000			
TS	−0.011	−0.020	−0.047	1.000		
OIL	0.035	0.020	−0.009	−0.007	1.000	
FX	0.007	0.036	−0.204	−0.104	−0.001	1.000
B. 8 January 1992–15 February 1995 ($t = 1, \ldots, 163$)						
BMR	1.000					
VGS	0.024	1.000				
SIZ	0.093	−0.235	1.000			
TS	−0.011	−0.097	−0.222	1.000		
OIL	0.117	0.105	0.009	−0.012	1.000	
FX	−0.067	0.022	−0.240	0.037	−0.017	1.000
C. 22 February 1995–8 March 2000 ($t = 164, \ldots, 427$)						
BMR	1.000					
VGS	−0.027	1.000				
SIZ	−0.010	−0.255	1.000			
TS	−0.007	0.104	−0.196	1.000		
OIL	0.036	0.011	0.028	0.000	1.000	
FX	0.009	0.055	−0.206	−0.135	0.005	1.000
D. 15 March 2000–12 March 2003 ($t = 428, \ldots, 584$)						
BMR	1.000					
VGS	0.000	1.000				
SIZ	−0.010	−0.173	1.000			
TS	0.039	−0.121	0.286	1.000		
OIL	0.021	−0.040	0.072	0.122	1.000	
FX	0.006	0.053	−0.155	−0.141	−0.085	1.000
E. 19 March 2003–2 February 2005 ($t = 585, \ldots, 683$)						
BMR	1.000					
VGS	−0.156	1.000				
SIZ	−0.225	0.110	1.000			
TS	−0.169	0.006	−0.002	1.000		
OIL	−0.149	0.233	0.018	−0.099	1.000	
FX	0.112	−0.370	−0.041	−0.385	0.020	1.000

are considered. Each of the alternative specifications is based on the simple factor model

$$
\begin{aligned}
R_i \;=\;& \beta_{i1}BMR_t + \beta_{i2}VGS_t + \beta_{i3}SIZ_t + \beta_{i4}TS_t + \beta_{i5}OIL_t \\
& + \beta_{i6}FX_t + \epsilon_{i,t}, \qquad \epsilon_{i,t} \sim N(0, \sigma_i^2).
\end{aligned}
\tag{7.20}
$$

The first alternative model employs recursive least squares (RLS). The factor model in (7.20) is estimated repeatedly using expanding subsets of the sample. As outlined in §7.3.4, the first 163 observations represent the formation period, on which the first estimate of the vector of factor loadings, $\hat{\boldsymbol{\beta}}_{i,163}$, is based. Adding the subsequent observation to the subset of data leads to an estimate of $\boldsymbol{\beta}_{i,164}$, adding the next observation yields an estimate of $\boldsymbol{\beta}_{i,165}$, and so on. This procedure is repeated until the number of observations in the subset grows to $T = 682$. This gives overall 520 estimates of ex-ante factor loadings; the last one is computed on 26 January 2005 based only on information up to this date.

The other two alternative models are based on rolling regressions with estimation periods of one and five years, respectively. In the literature, factor models are most commonly estimated using five years of data (cf. Campbell et al. 1997, §5.1). In this chapter, a one-year rolling period is chosen so that short-term effects can be captured by simple OLS. In case of the one-year rolling regression (RR1) specification, (7.20) is estimated using a sample spanning an estimation period from $t - 51$ to t. The sample is rolled forward week-by-week for $t = 163, \ldots, 682$ such that the 520-th loading is obtained using data from 4 February 2004 to 26 January 2005. Analogously, for the five-years rolling regression (RR5) setup the sample size is targeted at 260 observations. As before, 520 ex-ante estimates of the vector of factor loadings are generated. As the complete data set contains only 683 weeks, the first 97 ex-ante instruments are generated based on expanding samples consisting of less than 260 data points. The first estimate is available for 15 February 1995 and is based on 163 observations. After 18 December 1996 all factor loadings are based on individual samples that contain the respective previous five years of weekly data.

Following the procedures described above, four different estimates of the 18×6 matrix $\boldsymbol{B}_t$ of ex-ante factor loadings, denoted as $\hat{\boldsymbol{B}}_t^{KF}$, $\hat{\boldsymbol{B}}_t^{RLS}$, $\hat{\boldsymbol{B}}_t^{RR1}$ and $\hat{\boldsymbol{B}}_t^{RR5}$, are obtained for $t = 163, \ldots, 682$. In a second step, these estimates will be employed as instruments to estimate the corresponding risk premia in a cross-sectional setting according to (7.11). Unless mentioned otherwise, all estimation procedures in this chapter have been carried out using `Ox 3.30` by Doornik (2001). For all Kalman filter based computations, `Ox` has been used together with the package `SsfPack 2.3` by Koopman et al. (1999).

The remainder of this section presents the empirical results and is organized as follows: Subsection 7.4.1 summarizes the estimation results and compares the various factor loading series. Subsection 7.4.2 compares the out-of-sample forecasting accuracy of the Kalman filter based sensitivities with that of the proposed alternatives. Subsection 7.4.3 analyzes whether the superior forecasting performance of the Kalman filter based specification can be exploited either from a risk management or from a portfolio management perspective.

7.4.1 Estimation of factor loadings

The estimation results for the fully specified Kalman filter based model in (7.18)–(7.19) are provided in Table 7.6. The estimated hyperparameters are broadly significant. While the market factor is significant for all sectors except for Travel & Leisure, both fundamental factors are significant at the 1% level for all sectors. With regard to the set of macroeconomic factors it can be observed that not all factors are relevant for all sectors. The term structure factor, which is supposed to pick up systematic risks that are related to a changing slope of the yield curve, turns out to be significant for thirteen out of eighteen sectors. The estimated sensitivities to changes in the oil price are significantly different from zero for twelve sectors. Surprisingly, the oil price does not represent a systematic risk factor for the energy-related Chemicals, Utilities and Travel & Leisure sectors. This leaves the latter as the only sector with only four significant factors. The overall most important macroeconomic factor in explaining the sector return series over time is the exchange rate. With the exception of Healthcare, Personal & Household Goods and Telecommunications, the dollar is significant for all sectors. As measured by the adjusted coefficient of determination, the proposed specification with the selected set of explanatory variables explains between 65% and 93% of the variance of sectors. The lowest values of $\bar{R}^2$ are observed for Travel & Leisure and the defensive sectors Food & Beverages, Healthcare, Oil & Gas and Utilities. The model is found to be more appropriate for financials and non-financial cyclicals than for defensives and TMT.

Regarding the estimation results of the alternative least squares based multiple factor specifications given by (7.20), the finding that the market and fundamental factors are more significant than the macroeconomic factors is only partially maintained. For the RLS specification, both the market and the size factor are significant for all eighteen sectors. With the exception of Construction and Healthcare, the value-growth spread is significant at least at the 10% level for all sectors. In contrast to the observation made above, the term structure factor also represents a systematic risk factor for sixteen sectors. The exchange rate factor is only significantly different from zero for Automobiles. Changes in the oil price represent the second most important macroeconomic factor. As expected the explanatory power is lower than in case of the more flexible Kalman filter based model: as measured by the adjusted R^2 between 49% (Food & Beverages) and 78% (Banks) of the sector variance can be explained by movements of the chosen systematic risk factors. The estimation results for the rolling RR1 and RR5 specifications are summarized by reporting the respective ranges of estimated coefficients. It can be seen that all factors are significant for each sector at some stage. This supports the hypothesis of time-varying factor sensitivities. For a detailed summary of the estimated parameters of the alternative least squares based multifactor models, see Table C.7 in the appendix.

Examples of the paths of Kalman filter based factor loadings are illustrated in the following set of figures. All charts show filtered state estimates so that large outliers may occur at the beginning of the respective series of factor loadings. The two panels in Figure 7.1 display the respective time-varying exposure of Technology and Food & Beverages to the value-growth spread. Technology's value-growth beta became clearly negative during the bubble and its subsequent aftermath, which reflects the growth

Table 7.6: *Parameter estimates for multifactor Kalman filter models.*

This table reports the estimated parameters for the Kalman filter based multifactor specification for the eighteen DJ Stoxx^{SM} sectors. Columns 3–8 report the estimated state disturbance variance terms $\sigma_{\eta k}^2$ for $k = 1, \ldots, 6$, with k referring to the respective regressor (1: BMR, 2: VGS, 3: SIZ, 4: TS, 5: OIL, 6: FX); *** means significance at the 1% level (**: 5%, *: 10%). The last column shows adjusted coefficients of determination.

Sector	$\sigma^2 \times 10^4$	$\sigma_{\eta 1}^2 \times 10^4$	$\sigma_{\eta 2}^2 \times 10^4$	$\sigma_{\eta 3}^2 \times 10^4$	$\sigma_{\eta 4}^2 \times 10^4$	$\sigma_{\eta 5}^2 \times 10^4$	$\sigma_{\eta 6}^2 \times 10^4$	$\log L$	BIC	$\bar{R}^2$
Automobiles	3.14***	8.80***	43.37***	19.61***	3.03***	0.11***	42.13***	1692.1	−4.89	0.77
Banks	0.78***	1.68***	29.82***	96.09***	2.63***	0.00	3.32***	2105.7	−6.10	0.93
Basics	3.03***	6.07***	68.37***	8.24***	1.66***	0.10***	25.96***	1716.0	−4.96	0.70
Chemicals	1.92***	2.52***	16.46***	36.45***	0.90***	0.00	60.08***	1861.0	−5.38	0.77
Construction	1.32***	2.31***	18.37***	24.25***	1.06***	0.13***	27.40***	1979.0	−5.73	0.81
Financials	1.04***	1.11***	11.78***	59.83***	2.36***	1.12***	13.26***	2025.7	−5.86	0.89
Food	1.84***	5.66***	25.15***	18.80***	0.22***	0.01*	6.08***	1900.2	−5.50	0.65
Healthcare	2.57***	0.79***	61.77***	39.41***	0.00	0.03***	0.00	1802.7	−5.21	0.65
Industrials	0.79***	2.59***	43.62***	29.01***	1.52***	0.00**	24.84***	2113.8	−6.12	0.91
Insurance	1.79***	31.85***	0.00***	85.91***	3.87***	0.00	8.54***	1855.6	−5.37	0.89
Media	2.42***	22.02***	64.54***	30.86***	1.64***	0.56***	5.13***	1775.1	−5.13	0.85
Oil & Gas	2.86***	0.40***	12.48***	48.60***	0.00	0.91***	31.06***	1758.9	−5.08	0.67
Personal	1.53***	6.73***	46.08***	19.53***	0.71***	0.00	0.00	1946.7	−5.63	0.79
Retail	1.69***	2.62***	9.74***	31.96***	0.22***	0.20***	16.79***	1923.5	−5.57	0.72
Technology	3.04***	36.69***	165.38***	28.17***	4.10***	0.12***	11.59***	1680.2	−4.85	0.88
Telecom	3.21***	2.99***	22.99***	20.19***	0.00	0.03***	0.00	1738.8	−5.02	0.78
Travel	3.14***	0.00	1.56***	47.01***	0.00	0.00	43.06***	1737.6	−5.02	0.66
Utilities	1.52***	4.65***	21.57***	29.40***	0.00	0.00	4.10***	1970.7	−5.70	0.69
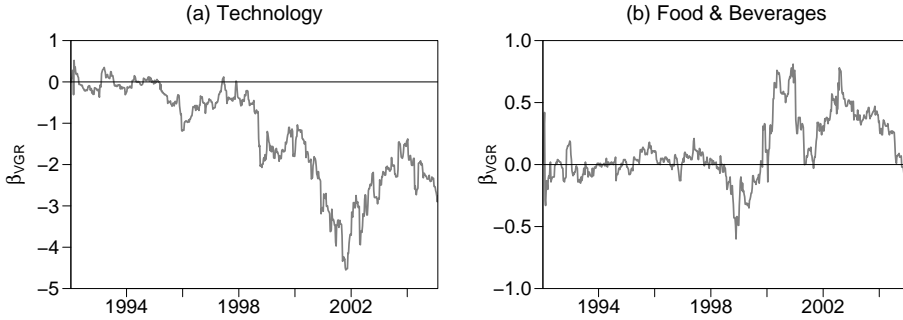
Figure 7.1: *VGR factor loadings for Technology and Food & Beverages.*

characteristics of the sector. As can be seen from Panel (b), the market perception of whether the Food & Beverages sector has more growth or more value attributes changes over time. Four distinct regimes can be identified. Until October 1997, the VGR beta of the sector moves around zero. Over the subsequent two years, when the whole market rose and growth outperformed value, the beta becomes slightly negative. After the market peaked in March 2000, the VGR beta of Food & Beverages turns positive. This is intuitive given the sector's defensive characteristics: market participants exited the overcrowded and expensive growth segments of the market and looked for cheaper and safer alternatives such as Food & Beverages. This led to a rising VGR beta. At the end of the sample, the estimated sensitivity returns toward zero, which means that the value-growth spread no longer represents a systematic risk factor of the sector. The observation that the VGR beta is only significantly different from zero around the TMT bubble holds for other sectors as well.

Figure 7.2 displays the time-varying size exposure for Industrials and Healthcare. While the Industrials sector index is comprised of many small caps, the Healthcare sector
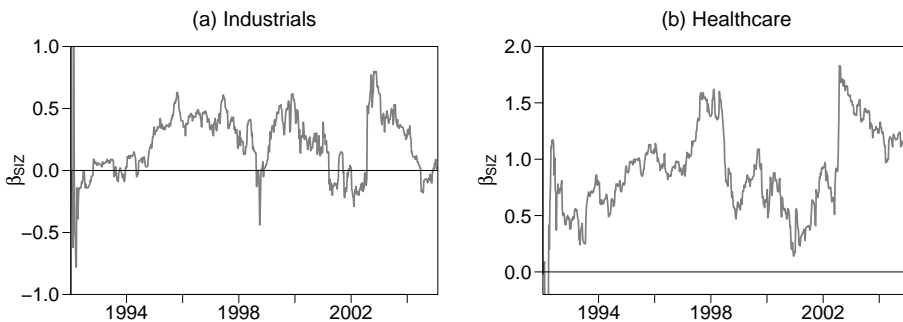
Figure 7.2: *SIZ factor loadings for Industrials and Healthcare.*
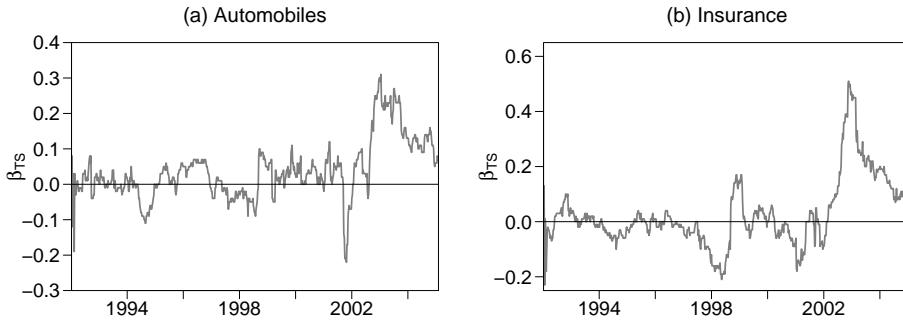
Figure 7.3: *TS factor loadings for Automobiles and Insurance.*

is dominated by large cap pharmaceutical stocks, such as GlaxoSmithKline, Novartis and Sanofi-Aventis. Therefore, size beta of Healthcare can be expected to be higher. At the first glance, both series mostly move in positive territory, which is representative for all sectors. This means that over the entire estimation period returns are usually positively related to an outperformance of large caps over small caps. As can be seen from Panel (a), the size beta of the Industrials sector occasionally becomes negative and rarely takes on values greater than 0.5. This is in contrast to the average size beta of Healthcare, which moves around unity as shown in Panel (b).

Changes in the slope of the yield curve are generally assumed to capture changes in the state of the economy and would naturally be expected to have a systematic impact on cyclical sectors. The two panels in Figure 7.3 illustrate the paths of TS loadings for the non-financial cyclical sector Automobiles and the financial cyclical Insurance sector, respectively. For both sectors, the sensitivities alternate between positive and negative territory: at some stages of the cycle the corresponding sectors rise when the yield curve
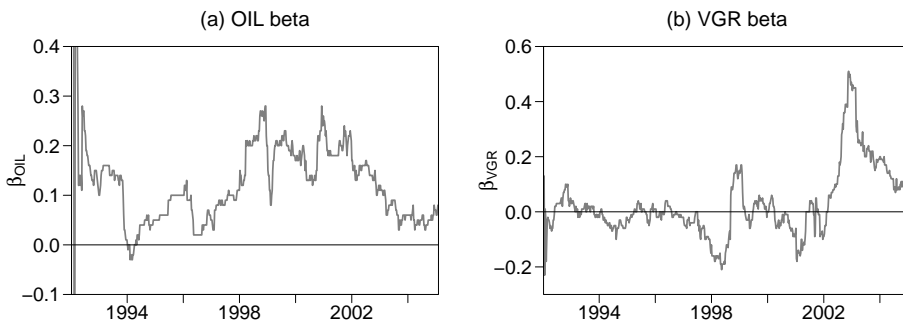
Figure 7.4: *OIL and VGR betas for the Oil & Gas sector.*
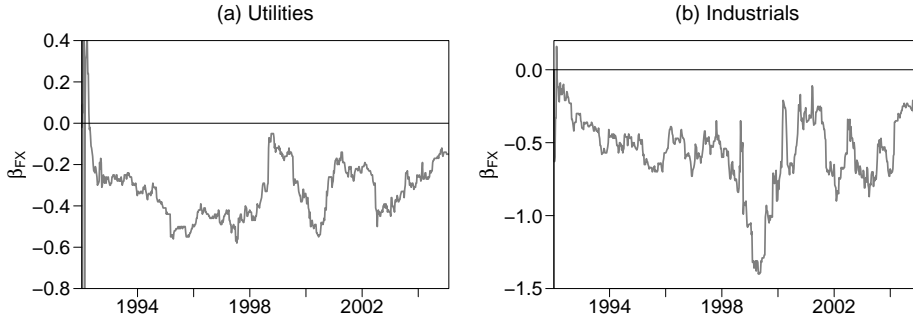
Figure 7.5: *FX factor loadings for Utilities and Industrials.*

steepens, while at other times the returns are positively related to a flattening of the curve.

For the Oil & Gas sector, the loadings on the oil price are positive over the complete out-of-sample period (t=163,...,683). This could be expected as the sales and earnings of oil companies are directly linked to changes in the price of oil. This does not mean that the systematic importance of the oil price factor is not changing over time. As illustrated by the left panel of Figure 7.4, the OIL beta of the sector even decreases toward zero after July 2001. At the same time, the sector's VGR beta starts to pick up. Between 2002 and 2003 the value-growth spread becomes relatively more important in explaining the series of Oil & Gas returns than changes in the price of oil.

As indicated above, the US-dollar is found to be the most important systematic macroeconomic risk factor. All sectors tend to be negatively related to a rising euro. This observation is especially true for export-oriented industrial cyclical sectors such as Industrials and Automobiles; it only holds to a lower degree for Utilities as shown by Figure 7.5.

7.4.2   Out-of-sample forecasting performance

The estimation results above indicate that the sensitivities to the chosen multiple systematic risk factors are varying over time. As demonstrated in Chapter 6, the out-of-sample forecasting accuracy of the estimated conditional betas can be evaluated and compared to the proposed least squares based alternatives using the *mean absolute error* (MAE) and *mean squared error* (MSE) criteria. Their definition is repeated here for convenience:

$$MAE_i = \frac{1}{T}\sum_{t=1}^{T} |\hat{R}_{i,t} - R_{i,t}|, \tag{7.21}$$

$$MSE_i = \frac{1}{T}\sum_{t=1}^{T} (\hat{R}_{i,t} - R_{i,t})^2, \tag{7.22}$$

Table 7.7: *Average out-of-sample errors across sectors for multiple factor models.*

	Mean absolute error ($\times 10^2$)				Mean squared error ($\times 10^4$)			
	KF	RLS	$RR5$	$RR1$	KF	RLS	$RR5$	$RR1$
Average error	1.270	1.401	1.371	1.345	3.143	4.051	3.767	3.505
Average rank	1.00	3.67	3.00	2.33	1.00	3.89	3.06	2.06

for $t = 164, \ldots, 683$ with $T = 520$ where $\hat{R}_{i,t}$ represents the series of return forecasts for sector i defined as

$$\hat{R}_{i,t} = \hat{\boldsymbol{\beta}}'_{i,t|t-1} \boldsymbol{f}_t, \qquad (7.23)$$

The betas employed to generate the return forecasts for date t are based only on information available at time $t - 1$. For all specifications, betas are predicted naïvely with the respective latest estimate of beta being used as the corresponding one-step ahead prediction. The series of factor realizations, $\boldsymbol{f}_t$, are assumed to be known with perfect foresight, which represents a common *modus operandi* in empirical research to isolate the impact of conditionality.

The resulting mean error measures are summarized in Table 7.7 which reports the respective averages together with the corresponding average ranks each across all sectors for all four employed specifications; for a detailed sectoral breakdown, see Table C.8 in the appendix. For the sample under consideration, the Kalman filter based betas yield the relatively best out-of-sample forecasts. For both error measures, the Kalman filter based specification ranks first for all eighteen sectors. Compared to the least squares alternatives, the mean absolute error is between 5.9% and 10.4% lower. When using the mean squared error as evaluation criterion, the superiority of the Kalman filter based betas becomes even more obvious: compared to the RLS, $RR5$ and $RR1$ alternatives, the average MSE is 28.9%, 19.9% and 11.5% lower, respectively.

7.4.3 Practical relevance of time-variation in factor loadings

The results above indicate that the proposed dynamic specification, where conditional betas are modeled as individual random walks estimated via the Kalman filter, produces the relatively most accurate estimates of time-varying factor loadings. This subsection examines whether the statistical superiority of the proposed methodology can be exploited in practice, either in the pricing of risk or from a portfolio management perspective.

7.4.3.1 Risk pricing

As outlined in Subsection 7.2.2, a common procedure to analyze how well the chosen set of risk factors explains the cross-section of assets is to employ the Fama-MacBeth cross-sectional regression approach. Based on the proposed set of explanatory variables, (7.11) implies that the cross-section of sector returns follows a linear factor model of the

following form:

$$
\begin{aligned}
\boldsymbol{R}_t \ = \ & \lambda_{1,t}\hat{\boldsymbol{\beta}}_{1,t-1} + \lambda_{2,t}\hat{\boldsymbol{\beta}}_{2,t-1} + \lambda_{3,t}\hat{\boldsymbol{\beta}}_{3,t-1} + \lambda_{4,t}\hat{\boldsymbol{\beta}}_{4,t-1} + \lambda_{5,t}\hat{\boldsymbol{\beta}}_{5,t-1} \\
& \lambda_{6,t}\hat{\boldsymbol{\beta}}_{6,t-1} + \boldsymbol{\nu}_t, \qquad \text{for} \hspace{4.5cm} (7.24)
\end{aligned}
$$

for $t = 164, \ldots, 683$, with $\boldsymbol{R}_t$ and $\hat{\boldsymbol{\beta}}_{1,t-1}, \ldots, \hat{\boldsymbol{\beta}}_{6,t-1}$ denoting 18×1 vectors of excess sector returns and factor loadings, respectively. The factor loadings used as independent variables have been estimated from one of the alternative time series regressions (KF, RLS, RR5 or RR1) of excess sector returns on the set of explanatory variables given by (7.18)–(7.20). Estimation of (7.24) for each date t of the out-of-sample period yields a time series of associated risk premia $\lambda_{k,t}$, $k = 1, \ldots, 6$, each of length 520.[24] In accordance with (7.12)–(7.14) the significance of the chosen risk factors is tested by employing a t-test to the time series means of estimated risk premia.

The cross-sectional regression results are summarized in Table 7.8, where the coefficient averages together with the corresponding Fama-MacBeth t-statistics are reported. To ascertain whether the proposed Kalman filter based conditional factor loadings are relatively better in explaining sector pricing, the cross-sectional regressions have also been conducted for the betas estimated by recursive least squares and by rolling OLS. Table 7.8 is divided into four panels: Panel A covers the complete out-of-sample period from 22 February 1995 to 2 February 2005; Panels B to D refer to the three subperiods of observed bull market, bear market and market recovery regimes. Overall, there is only little evidence that the multifactor model is better able to describe the cross-section of returns when conditional Kalman filter based factor loadings are used as independent variables. Over the entire estimation period, the average $\bar{R}^2$ is only slightly higher than in case of the three alternative least squares specifications. The four different models explain between 28.9% (KF) and 28.0% ($RR1$) of the total cross-sectional variation of excess sector returns. The only risk premium that is significantly different from zero over the complete sample is estimated for the Kalman filter based TS factor.

Dividing the sample under consideration into three market-regime dependent subperiods reveals that both the average ability to explain the cross-sectional variability and the significance of estimated risk premia differ across subperiods. For the bull market period (Panel B) the risk premium of the SIZ factor is significantly positive for all models. The sign on SIZ reflects the outperformance of large caps over small caps during the *dotcom* bubble. Besides, for the KF and $RR1$ based factor loadings the risk premium of the TS factor is significantly negative at the 10% level. This means that during

---

[24]For each cross-sectional regression at a given date t, six parameters have to be estimated with the dependent variable consisting of 18 data points only. Similar settings with comparable degrees of freedom are commonly employed in the literature, see, for example, Chen et al. (1986) who use 20 portfolios as dependent variables to estimate up to seven parameters in the cross-section. Nevertheless, in order to check whether the conclusion to be drawn in this subsection is sensitive to the dimension of the vector of dependent variables, (7.24) has also been estimated using 37 sectors instead of the 18 supersectors employed throughout this thesis. As the employment of a finer market segmentation leads to the same conclusion with respect to the practical relevance of time-varying factor loadings, only the case with 18 supersectors is discussed here; results of the Fama-MacBeth procedure using 37 sectors are available on request.

Table 7.8: *Fama-MacBeth regression results I.*

This table presents the estimation results for the Fama-MacBeth regressions where the regressors are the out-of-sample factor loadings on the proposed set of explanatory variables. The factor loadings have been estimated from one of the time series regressions of excess sector returns on the set of explanatory variables as presented in §7.4.1. Average coefficients are expressed as percent per week $\times 10$. The number of cross-sectional regressions is 520 for each model; t-statistics are reported in parentheses with the relevant critical value at the 95% (90%, 99%) level being -1.96 (-1.65, -2.59). Significant average coefficients are printed in **boldface**. The last column reports average adjusted coefficients of determination.

Model	BMR	VGS	SIZ	TS	OIL	FX	$\bar{R}^2$
A. Complete out-of-sample period (22 February 1995–2 February 2005)							
KF	0.214	0.114	0.993	**−12.041**	0.321	−0.284	0.289
	(0.186)	(0.128)	(1.268)	(−1.694)	(0.062)	(−0.237)	
RLS	1.030	0.152	1.099	19.764	3.356	0.492	0.284
	(0.543)	(0.099)	(1.059)	(0.519)	(0.381)	(0.180)	
RR5	1.214	0.993	1.427	16.663	1.700	0.980	0.282
	(0.755)	(0.754)	(1.448)	(0.781)	(0.223)	(0.475)	
RR1	−0.675	0.524	0.912	−7.340	−1.292	−0.316	0.280
	(−0.553)	(0.610)	(1.152)	(−0.813)	(−0.331)	(−0.273)	
B. Bull market subperiod (22 February 1995–8 March 2000)							
KF	0.730	−2.000	**2.706**	**−18.968**	0.401	−1.132	0.209
	(0.490)	(−1.464)	(2.346)	(−1.829)	(0.052)	(−0.624)	
RLS	3.113	−3.797	**3.755**	58.609	21.553	3.575	0.196
	(1.254)	(−1.503)	(2.680)	(1.229)	(1.488)	(0.857)	
RR5	1.754	−1.375	**4.358**	48.533	12.872	2.164	0.215
	(0.769)	(−0.625)	(3.205)	(1.320)	(0.985)	(0.635)	
RR1	−0.483	−1.538	**3.056**	**−23.658**	−1.853	−1.515	0.189
	(−0.324)	(−1.234)	(2.632)	(−1.756)	(−0.323)	(−0.930)	
C. Bear market subperiod (15 March 2000–12 March 2003)							
KF	−2.923	**3.322**	−2.012	−6.410	−1.153	1.093	0.420
	(−1.127)	(1.907)	(−1.360)	(−0.434)	(−0.126)	(0.489)	
RLS	−6.521	**6.264**	−2.647	4.862	**−25.850**	−5.049	0.441
	(−1.534)	(2.399)	(−1.160)	(0.055)	(−1.723)	(−0.953)	
RR5	−3.093	**5.020**	−2.569	−12.368	−14.200	−0.227	0.419
	(−0.938)	(2.256)	(−1.210)	(−0.406)	(−1.217)	(−0.067)	
RR1	−3.658	**3.948**	**−2.742**	6.247	0.181	−0.970	0.406
	(−1.290)	(2.137)	(−1.774)	(0.365)	(0.026)	(−0.446)	
D. Market recovery subperiod (19 March 2003–2 February 2005)							
KF	**3.813**	0.661	1.187	−2.502	2.444	−0.205	0.296
	(2.026)	(0.897)	(0.891)	(−0.276)	(0.239)	(−0.104)	
RLS	**7.449**	0.989	−0.044	−60.188	1.151	1.057	0.268
	(2.479)	(0.992)	(−0.029)	(−0.971)	(0.141)	(0.306)	
RR5	**6.603**	0.919	−0.053	−22.285	−2.879	−0.264	0.244
	(2.563)	(1.116)	(−0.039)	(−0.921)	(−0.411)	(−0.104)	
RR1	3.546	0.591	0.989	14.628	−2.134	3.920	0.319
	(1.613)	(0.838)	(0.841)	(0.998)	(−0.273)	(1.600)	

this subperiod sectors that performed well whenever the yield curve flattened have been preferred. While the model explains only around a fifth of the total variation during the bull market, it captures between 40.6% (RR1) and 44.1% (RLS) of total variation during the subsequent period of market decline (Panel C). In this market environment the VGS factor is positively priced for all models, which reflects the outperformance of value over growth between March 2000 and March 2003. For the RLS and $RR1$ based betas, both the OIL and the SIZ factor are also found to be significant. After the end of the bear market the residual market factor BMR dominates. In accordance with Panel D the associated risk premium is significantly positive for all models except for the $RR1$ specification. Following previous market phases, where either large caps or value assets have been en vogue, market participants started to focus on classical market beta exposure to benefit from the market recovery. For this subperiod, the chosen systematic risk factors account for between 24.4% (RR5) and 31.9% (RR1) of total cross-sectional variation. Notably, the risk premium for the FX factor is not significant in any subperiod. The OIL factor is only significantly priced by the RLS model in a single subperiod (bear market). Re-estimation of the Fama-MacBeth regressions without these two factors leaves the major findings unaltered; a detailed summary of the corresponding estimation results is given in Table C.9 in the appendix.

The findings above indicate that the statistical superiority of the Kalman filter based factor loadings does not lead to a significant improvement in the pricing of risk. When factor loadings are estimated by one of the alternative least squares specifications, similar results are obtained with regard to both the identification of significant risk premia, and the degree to which the cross-sectional variation of excess sector returns can be explained.

7.4.3.2 Portfolio management perspective

This subsection analyzes whether portfolios built on Kalman filter based return forecasts would have been more profitable on average than alternatively constructed portfolios. The respective hypothetical performances are compared using a *backtesting* procedure. Backtests are widely established in the financial community to test models and investment strategies prior to implementation (cf. Campbell et al. 2007). All backtests presented here have been performed using R (R Development Core Team 2005) in combination with the `backtest` package, version 0.1-1, by Campbell et al. (2006b).

In a first step, so-called *rank scores* are assigned to the assets used to construct the portfolios. These scores are based on $\hat{R}_{i,t}$, the series of return forecasts computed in §7.4.2 for the different modeling techniques according to (7.23) for $t = 164, \ldots, 683$. For each forecasting period, scores from 1 to 18 are assigned to each sector, where the sector with the respective lowest predicted return ranks first. Following the outline by Campbell et al. (2006a) these scores are used to split the return forecasts into six quintiles, for which the average real returns are computed for each modeling approach. The spread between the observed average future returns for the highest and the lowest quintile represents a simple measure for the accuracy of the different sector rankings. Table 7.9 summarizes the results for all four models.

For the return predictions based on Kalman filter derived factor loadings, the average return for the lowest quintile, i.e. the mean weekly forward return for sectors with a

Table 7.9: *Summary results of simple backtests.*

Model	low	Q2	Q3	Q4	Q5	high	spread
KF	−0.0113	−0.0031	0.0001	0.0026	0.0050	0.0133	0.0246
RLS	−0.0106	−0.0027	−0.0005	0.0027	0.0053	0.0124	0.0230
RR5	−0.0107	−0.0031	0.0000	0.0022	0.0052	0.0130	0.0237
RR1	−0.0112	−0.0025	0.0006	0.0020	0.0050	0.0128	0.0240

rank score in the lowest quintile, is −1.13%. Together with an average return of 1.33% for sectors in the highest quintile, this results in an average weekly spread of 2.46%. This suggests that on average the high-ranked sectors perform significantly better than the sectors in the lowest quintile. Positive spreads, albeit of smaller magnitude, are also observed for the least squares based alternatives. Compared to the conditional return predictions, the spread is between six (*RR1*) and sixteen (*RLS*) basis points lower per week. The results of this simple backtest indicate that portfolios built on return forecasts that utilize conditional factor loadings would have performed better than portfolios alternatively built on least squares based return predictions.

While this procedure yields a simple measure for the average weekly performance over the entire sample period, it does not take into account how the different portfolios develop over time. For this purpose, a backtesting procedure referred to as *natural backtest* is employed. Following Campbell et al. (2006a) a natural backtest is implemented as follows: in the starting period an equally weighted *long-short portfolio* is constructed by investing in the top three sectors and shorting the three sectors in the lowest quintile. Subsequently, the portfolio is reviewed weekly and rebalanced according to the latest quintile changes, i.e. a long (short) position in a sector that has left the top (bottom) quintile is closed and new positions are entered. According to this procedure, for each modeling approach 520 portfolios are formed over the out-of-sample period between 22 February 1995 and 2 February 2005.

Figure 7.6 displays the evolution of the Kalman filter based quintile portfolios and the corresponding *spread return*, which is defined as the difference between the top and the bottom quintile portfolio. The spread return represents a central output of a natural backtest. As illustrated in Panel (a) the cumulative performance of the higher quintile portfolios moves on a higher path than the cumulative performance of the lower quintile portfolios. In the first three years of the out-of-sample period, the six quintile portfolios move closely together. After 1998, when world equity markets entered the final stages of the bull market, the individual quintile spreads began to widen. This observation particularly holds for the spread between the highest and the lowest quintile. In the period between the market peak in 2000 and the end of the bear market in 2003, the widening accelerates even further. Panel (b) shows the weekly spread return. With a few exceptions, the spread return is positive. Three volatility regimes can be observed: following a period of relatively low spread dispersion, volatility increases and peaks in 2000. Since then the observed spread dispersion consistently decreases back toward pre-1998 levels. Similar patterns can be observed for the portfolios alternatively built on least squares based factor loadings; for an illustration, see Figure B.5 in the appendix.
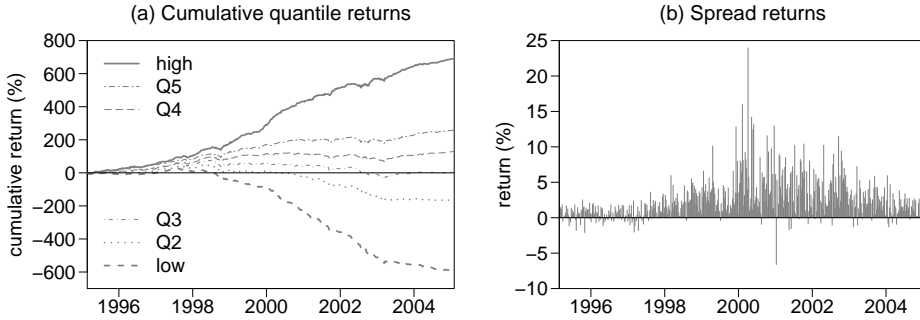
Figure 7.6: *(a) Cumulative quintile and (b) spread returns for the KF portfolios.*

Summary results for all conducted natural backtests are presented in Table 7.10. The mean spread returns in the second column correspond to the reported spreads for the simple backtests above (cf. Table 7.9). The first piece of extra information that is received by conducting a natural backtest is the standard deviation of spread returns, which is reported in the third column. Compared to the alternative specifications, the higher average spread return offered by utilizing Kalman filter based betas does not come at the cost of a significant pick-up in volatility. Column 4 reports non-annualized information ratios, defined as the ratio between the mean spread return and its standard deviation, as a measure for risk-adjusted returns. With an information ratio of 0.8542, the average risk-adjusted performance of the Kalman filter based portfolio is slightly higher than the performance of any alternative portfolio. With perfect foresight of the systematic risk factors, the portfolio constructed using conditional random walk betas would have led to an cumulative return of 1281% over the entire sample period. As transaction costs are ignored and risk factors can only be predicted with a limited degree of precision in the real world, this is only a hypothetical number. Nevertheless, the reported total returns in column 5 can be employed to compare the relative performance of the various modeling techniques. Other things being equal, the utilization of

Table 7.10: *Summary results of natural backtests.*

Model	$\mu(\text{spread})^a$	$\sigma(\text{spread})^b$	Information ratio[c]	total return[d]
KF	0.0246	0.0288	0.8542	12.81
RLS	0.0230	0.0275	0.8371	11.97
$RR5$	0.0237	0.0279	0.8501	12.34
$RR1$	0.0240	0.0289	0.8300	12.48

[a]$\mu(\text{spread})$ is the mean spread return.
[b]$\sigma(\text{spread})$ is the standard deviation of spread returns.
[c]The non-annualized information ratio is defined as $\mu(\text{spread})/\sigma(\text{spread})$.
[d]Total return is defined as the cumulative spread return over the entire sample.
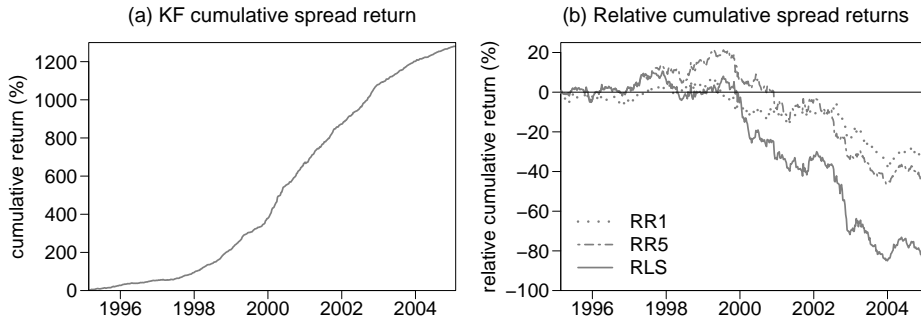
Figure 7.7: *(a) Cumulative KF spread return and (b) cumulative spread returns of the alternative portfolios relative to the KF based portfolios.*

conditional factor loadings to generate return forecasts results in long-short portfolios that are characterized by both higher absolute and higher risk-adjusted returns.

In order to analyze how the performance of the Kalman filter based strategy relative to that of the alternative least squares based strategies evolves over time, we look at *relative cumulative spread returns*. The relative cumulative spread return of an alternative least squares based long-short portfolio at date t for $t = 164, \ldots, 683$, is computed as the difference between its cumulative spread return and the KF cumulative spread return at the same date. The left hand side of Figure 7.7 shows the KF cumulative spread return, which is steadily growing over time. Panel (b) presents the relative cumulative spread returns for the RLS, $RR5$ and $RR1$ based long-short portfolios, respectively. It can be seen that during the low volatily regime at the beginning of the sample, all relative cumulative spread return series move sideways around the zero line. In 1999, the $RR1$ series reaches 20%, which means that this strategy has been temporally able to outperform the KF strategy. At the end of 1999, the KF long-short portfolio started to consistently outperform all alternative strategies. The sharpest relative gains are observed for the last few months before the market peaked, and also in 2002 when global equity markets collapsed.

The results in this section demonstrate the practical relevance of employing conditional factor loadings from a portfolio management perspective. The analysis of long-short portfolios reveals that the use of Kalman filter based factor loadings, which take the time-varying nature of fundamental and macroeconomic systematic risks explicitly into account, results in better return forecasts of pan-European sectors. The relative performance of Kalman filter based long-short portfolios is analyzed over time by looking at relative cumulative spread returns. It is found that the superior performance traces back to the increased flexibility of conditional factor loadings, which particularly pays off during more volatile market regimes.

7.5   Concluding remarks

This chapter has explored the conditional nature of systematic influences on pan-European industry portfolios. While factor models with multiple risk factors play a dominant role in the modeling of risk and return, only relatively few approaches take potential time-variation of factor sensitivities explicitly into account. The empirical analysis presented in this chapter models the loadings to multiple fundamental and macroeconomic common factors as individual stochastic random walk processes estimated via the Kalman filter. The factors are chosen a priori in accordance with empirical findings of the anomalies literature. In order to estimate factor risk premia in a cross-sectional setting, the methodology proposed in this thesis combines the classical Fama-MacBeth approach with the paradigm of time-varying betas.

In the first part of the empirical analysis, the out-of-sample forecasting accuracy of the proposed conditional modeling approach is compared to that of various least squares based alternatives. According to the selected error measures, the Kalman filter based specification offers the best out-of-sample predictions of conditional betas for all sectors. This confirms time-variation in the relationship between sector returns and the selected systematic risk factors. In line with previous findings, the market risk factor and the fundamental factors size and valuation turn out to have a bigger impact on pan-European industry portfolios than common macroeconomic risk factors.

The second part of the empirical analysis addresses the practical relevance of explicitly considering the conditionality of factor sensitivities. For the sample under consideration the conclusion is mixed. From a risk management perspective, the findings derived by a Fama-MacBeth regression procedure with conditional betas as instruments suggest that the statistical superiority of Kalman filter based betas does not result in an improved pricing of risk. On the other hand, from a portfolio management perspective the use of conditional factor loadings is found to be of practical relevance. A backtesting procedure that relies on the construction of long-short portfolios is employed to derive the empirical results. Using out-of-sample estimates of conditional betas to compute return forecasts over ten years, it is demonstrated that Kalman filter based conditional betas lead to improved return forecasts of pan-European industry portfolios.

Overall, this chapter adds to the literature an empirical analysis that studies the practical relevance of using time-varying factor loadings in a multifactor pricing model. Methodologically, a Fama-MacBeth approach with conditional instruments is proposed. The findings of this chapter lead to the conclusion that time-varying sensitivities carry implications for the time series prediction of returns, and less so for the analysis of risk. To further clarify the practical importance of time-varying betas, future research could examine the modified Fama-MacBeth procedure in more detail. In order to increase the dimension of the vector of dependent variables, single stock instead of sector returns could be employed. Another interesting aspect would be a multivariate approach to the introduced Kalman filter based multifactor model, where the equations for all sectors are estimated simultaneously to make use of potential interrelationships between the series under consideration.

Chapter 8

Conclusion and outlook

This thesis deals with the modeling of change in the context of widely established concepts in finance. Advanced time series models are applied to analyze the time-varying relationship between systematic risks and pan-European industry portfolios. Three major research objectives are addressed: (i) to provide a notationally conformable introduction to the econometric toolbox needed for an elaborate modeling of changing relationships over time, (ii) to compare the different modeling techniques' ability to characterize and predict the time-varying nature of systematic beta risk as a stochastic process, and (iii) to analyze the practical relevance of taking conditionality of factor sensitivities explicitly into account. This final chapter reviews the main results and points to possible future directions to continue research connected to the findings presented in this thesis.

A sound theoretical foundation of how advanced time series techniques can be employed to model changing relationships over time constitutes a prerequisite to carry out empirical research on the modeling of change. As we can choose from a wide spectrum of available modeling techniques, the first step in conducting research is to decide on how a problem should be approached. A general guide toward the selection and specification of a model is to look at the empirical stylized facts of the series of interest. Chapter 2 introduces the set of weekly sector return series used throughout this thesis. A review of the series' major empirical properties illustrates why the paradigm of stable sensitivities has to be questioned: the distributions of the return series at hand are found to be highly leptokurtic, the squared series are significantly autocorrelated, most series exhibit a leverage effect and volatilities of different assets are linked to each other. These empirical properties, especially with regard to the observed volatility clustering and volatility co-movements, imply that the true degree of the linear association between sector returns and systematic influences is not constant but changing over time. This can be confirmed for all sectors by a formal testing procedure for structural breaks.

The stylized facts not only motivate the modeling of change, they also serve as a guide to which modeling techniques might be relied upon. As a single model usually cannot capture all identified distributional properties simultaneously, different models are employed to capture different properties of a series. The finding of volatility clustering points toward the employment of conditional heteroskedasticity modeling, a key concept in many areas of finance and financial econometrics that can be used to model

conditional regression coefficients in an indirect way. Within the versatile class of het-
eroskedasticity models, many different specifications that explicitly allow for particular
empirical properties, such as fat-tailed distributions, leverage effects or volatility co-
movements, are considered. As an alternative to the indirect approach, time-varying
relationships can be modeled and predicted directly: in the context of a state space
framework, the path of beta either emerges as a hidden continuous process estimated
via the Kalman filter, or as a hidden discrete process in a Markov regime switching
framework.

In the theoretical part of this thesis, the methodology to analyze the time-varying
relationship between common systematic risk factors and sector returns is introduced.
The econometric toolbox provided in Chapters 3–5 equips the reader with a range of
contemporary time series techniques of different, and often non-economic, origin: Gaus-
sian state space models and the Kalman filter, Markov regime switching models, and
linear and nonlinear GARCH models, and stochastic volatility models. Although choices
had to be made with respect to both breadth and depth in deciding which econometric
techniques should be employed in the course of this thesis, many features, reaching from
basic representations, estimation and path-tracking procedures to selected extensions,
are discussed. Linkages between the different concepts are emphasized: the introduc-
tion of linear Gaussian state space model and the Kalman filter lays the ground for
the strongly connected Markov regime switching and stochastic volatility models. The
presentation of the various concepts in a unified notational framework hopefully encour-
ages empirical researchers with a focus on finance-related disciplines — practitioners
and academics alike — to utilize these advanced time series concepts to conduct fruitful
empirical research in a field that is characterized by daily dynamics and ongoing change.

The empirical investigation in Chapter 6 adds to the literature a comprehensive analy-
sis of the ability of different elaborate time series concepts to model conditional beta risk
in a pan-European context. Compared to earlier studies for other regions of the world,
the spectrum of modeling techniques employed here is expanded to also include two
Markov regime switching models, and a bivariate stochastic volatility model that is es-
timated via simulation-based efficient Monte Carlo likelihood and importance sampling
techniques. A generalized random walk model is proposed to deal with heteroskedasticity
and non-normality in the context of a three-stage estimation procedure. The in-sample
results suggest that sector returns can be better explained by movements of the over-
all market when betas are allowed to vary over time. Compared to standard OLS,
each time-varying approach delivers superior in-sample forecasting results. This implies
confirmation of the conditionality assumption. An evaluation of the various modeling
techniques' ability to produce out-of-sample forecasts of conditional sector betas reveals
that the Kalman filter based random walk model offers the relatively best predictive
performance. The random walk specification is closely followed by the moving mean re-
verting and the proposed generalized random walk model. Unsatisfactory out-of-sample
results are obtained by the two Markov regime switching models. It is concluded that
time-varying market beta for pan-European sectors can be best described as a stochas-
tic random walk process. The finding that the generalized random walk model, which
has been introduced to take volatility clustering and outliers explicitly into account,

cannot outperform the random walk model, suggests that the corresponding quasi maximum likelihood estimator is well applicable in the presence of heteroskedasticity and non-Gaussianity.

While answering the question which of the selected modeling approaches is most suitable to predict the path of time-varying betas for the sample under consideration, many interesting challenges are open for future research. It would be of interest how the ranking of relative forecasting performances is affected both by the length of the forecasting period and by the frequency of return data. The performance of the Kalman filter could be further improved by adding exogenous top-down or bottom-up factors, or a combination of both, to explain the path of beta. Besides, depending on the eventual progress in the development of multivariate conditional heteroskedasticity models, it might be worth to consider multivariate ARCH and stochastic volatility models with truly dynamic conditional correlations.

Targeting the third research objective of this thesis, Chapter 7 contributes an empirical evaluation of the practical relevance of explicitly considering time-variation in factor sensitivities. The task is approached in the context of a conditional multifactor pricing framework for pan-European industry portfolios with macroeconomic and fundamental risk factors. Motivated by the findings on time-varying betas, the conditional factor sensitivities are modeled as individual stochastic random walk processes estimated via the Kalman filter. Methodologically, a synthesis of the classical Fama-MacBeth cross-sectional regression procedure with time-varying betas as instruments is introduced. Following a review of the anomalies literature, the following variables are selected as risk factors: the European term structure, the oil price, the dollar-euro exchange rate, the value-growth spread and size. All factors can be justified theoretically and have been successfully employed in the literature. A so-called residual benchmark factor, which is derived as the series of one-step ahead prediction errors obtained from an auxiliary regression with time-varying coefficients, is introduced and employed as additional risk factor. Compared to least squares based alternatives, the proposed Kalman filter based specification is found to yield the relatively best ex-ante forecasts. Fundamental risk factors are found to be of greater importance than macroeconomic factors.

The findings with regard to the practical relevance of explicitly considering the conditionality of factor loadings are mixed. On the one hand, the results obtained by the proposed Fama-MacBeth cross-sectional analysis with conditional instruments suggest that the documented statistical superiority of Kalman filter based betas is not associated with an improved pricing of risk. On the other hand, from a portfolio management perspective with a focus on producing return forecasts, the results are encouraging: the employed backtests indicate that, everything else being equal, improved return predictions can be obtained by replacing stable factor sensitivities by conditional loadings to take the time-varying nature of fundamental and macroeconomic risks into account. The superior performance of Kalman filter based long-short portfolios over time is grounded on the higher degree of flexibility offered by Kalman filter based factor loadings, which particularly pays off during more volatile market regimes.

As the question, whether the paradigm of time-varying betas is of practical relevance, cannot be answered unambiguously, more effort is needed for clarification. Avenues to

conduct future research connected to the analysis presented here, include further treatment of the proposed Fama-MacBeth procedure with conditional betas as instruments. Another interesting path would be a multivariate approach. Simultaneous estimation of the conditional Kalman filter based multifactor models could lead to better estimates of time-varying sensitivities and an improved forecasting performance.

Appendix A

Brief review of asset pricing theory

The concept of mean-variance efficient portfolios introduced in the seminal work of Markowitz (1959) represents a cornerstone of modern finance theory. The risk-return relationship of a portfolio is usually characterized by an asset pricing model, i.e. a linear factor model that decomposes the return on an asset into a possibly multidimensional set of common risk factors and an asset-specific component. Linear pricing models are commonly employed to predict returns, to identify risk sensitivities and to estimate abnormal returns. Following Cochrane (2005, §1–2) this section briefly summarizes the basic ideas of asset pricing theory.

A.1 The discount factor view of asset pricing

The cornerstone of asset pricing theory is that the value of an asset is equal to the expected discounted payoff. The risk related to the asset's payment is explicitly taken into account. One distinguishes between *relative* and *absolute* asset pricing. The former refers to the pricing of an asset given the prices of other assets. In absolute pricing, which is at the heart of finance and of this thesis, each asset is valued according to its exposure to fundamental macroeconomic risks.

By employing the discount factor view of asset pricing as proposed, among others, by Rubinstein (1976) or Hansen and Jagannathan (1991), asset pricing can be summarized by the following two equations:

$$p_t = E_t(m_{t+1}x_{t+1}), \tag{A.1}$$

$$m_{t+1} = f(data, parameters), \tag{A.2}$$

where p_t is the asset's price at date t, x_{t+1} is the asset payoff at time $t+1$ and m_{t+1} denotes the stochastic discount factor. This simple and universal approach allows for a separation of (i) determining the empirical representation in (A.1), and (ii) specifying the model assumptions in (A.2). By making different choices for the function $f(\cdot)$, different asset pricing models can be derived.

A.2   The consumption-based model

The most basic specification of the pricing equation presented in (A.1) can be derived from the first-order condition of an investor's marginal utility of consumption. The marginal loss of utility of consuming less today to buy more assets, should be equal to a higher degree of marginal utility gained from consuming more of the assets' future payoff. Hence, the value of an asset today should correspond to the expected discounted value of its payoff. A stream of uncertain future cash flows can be valued by determining the worth of this payoff to a typical investor using his or her utility function of consumption:

$$U(c_t, c_{t+1}) = u(c_t) + \delta E_t \left(u(c_{t+1}) \right), \tag{A.3}$$

where c_t is consumption at time t, c_{t+1} denotes random consumption at time $t+1$, and $u(\cdot)$ represents a concave and increasing utility function that reflects a declining marginal value of more consumption. Investors' typical preference of a steady stream of consumption and their impatience for future cash flows are captured by the curvature of the utility function and by δ, the *subjective discount factor*. Under the assumption that the investor is able to freely buy and sell any amount of the payoff x_{t+1} at a price p_t, the basic pricing equation can be derived as the first-order condition for optimal portfolio formation and consumption:

$$p_t = E_t \left(\delta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right). \tag{A.4}$$

The *stochastic discount factor* or *marginal rate of substitution*, which describes an investor's willingness to intertemporal substitution of consumption, is defined as

$$m_{t+1} := \delta \frac{u'(c_{t+1})}{u'(c_t)}. \tag{A.5}$$

According to Cochrane (2005, p. 6) Equation (A.4) represents "the central pricing equation". Most of the asset pricing theory simply consists of rearrangements and specializations of this formula. The expectation is generally conditioned on information up to time t; the price and the payoff are always considered at times t and $t+1$, respectively. In the following the subscripts will be suppressed and (A.1) is simply stated as $p = E(mx)$.

A.3   Alternative asset pricing models

Despite its theoretical appeal as a universal way to value any uncertain cash flow and security, the consumption-based model has not worked well in applied work. Possible explanations for poor model performance relate, among others, to the use of wrong utility functions and measurement errors in consumption data. This motivates alternative asset pricing models to avoid the empirical shortcomings of the consumption-based approach. Different functions for m_{t+1} have been proposed. The alternative approaches either employ different utility functions or link asset prices not to consumption data but to other factors or macroeconomic aggregates. The direct modeling of marginal utility based on alternative variables leads to *factor pricing models*, on which the analysis in the empirical part of this thesis is grounded.

Factor pricing models specify the stochastic discount factor as a linear function of the form

$$m = a + \boldsymbol{b}'\boldsymbol{f}, \tag{A.6}$$

with free parameters a and $\boldsymbol{b}$. The $K \times 1$ vector of factors $\boldsymbol{f}$ is chosen as a proxy for an investor's marginal utility growth. As demonstrated by Cochrane (2005, §6), (A.6) is equivalent to a multiple-beta model of expected returns:

$$E(R_i) = \gamma + \boldsymbol{\lambda}'\boldsymbol{\beta}_i, \tag{A.7}$$

with the $K \times 1$ vector $\boldsymbol{\beta}_i$ containing the multiple regression coefficients of returns R_i on $\boldsymbol{f}$ for assets $i = 1, \ldots, N$. This specification, usually referred to as a *beta pricing model*, states that each expected return is proportional to the asset specific $\boldsymbol{\beta}_i$, which is also known as the *quantity of risk*. The $k \times 1$ vector of free parameters $\boldsymbol{\lambda}$, which is the same for all assets i, can be interpreted as the *price of risk*. In a world with existing risk-free assets, i.e. a zero-beta portfolio, the constant γ is usually assumed to be equal to the risk-free interest rate, denoted by r^f; the economic model in (A.7) can be written in terms of returns in excess of the risk-free rate with γ being set to zero.

In order to identify the factors $\boldsymbol{f}$ that can serve as appropriate proxies for marginal utility growth, one looks for variables for which (A.6) approximately holds. As consumption is economically linked to the state of the economy, macroeconomic variables, such as interest rates, GDP growth and broad-based portfolios, constitute the first set of factors. Consumption can be assumed to also depend on current newsflow that signals future income and consumption changes. Variables that either indicate changes in consumption and/or other macroeconomic indicators, or predict asset returns directly, also qualify as potential factors; important variables include dividend yields, stock returns or the term premium (Cochrane 2005, §9).

The most important factor pricing models include the single-factor Capital Asset Pricing Model (CAPM), the Intertemporal CAPM (ICAPM) and the Arbitrage Pricing Theory (APT). The latter two allow for multiple sources of systematic risk. They all represent specializations of the consumption-based model, in which extra assumptions allow for the use of other variables to proxy for marginal utility growth. They can be summarized as follows:

- The CAPM, developed by Sharpe (1964) and Lintner (1965), is the first and still most widely used factor pricing model. It linearly relates the expected return of an asset to the return's covariance with the return on the *wealth portfolio*. The return on total wealth is usually approximated by the return on a broad-market stock portfolio.

- The ICAPM by Merton (1973) is grounded on equilibrium arguments where an investor tries to hedge uncertainty about future returns by demanding assets that do well on bad news. In equilibrium, expected returns depend on the covariation with current market returns and on the covariation with news that predict changes in the investment opportunity set of an investor. The ICAPM can be represented by (A.6) where each state variable that forecasts future market returns can be a factor.

- The APT, introduced by Ross (1976) as an alternative to the CAPM, is based on arbitrage arguments. The starting point of the APT is a statistical characterization of the return covariance matrix. The idea is that the common variation in returns can be related to common components, or *risk factors*, that describe the covariance matrix of returns. In this setting, idiosyncratic movements in returns are not priced as they can be completely diversified away. In contrast to the ICAPM, the factors f in (A.6) are assumed to provide a description of the return covariance matrix and to be IID and orthogonal (cf. Cochrane 2005, §9).

Overall, the ICAPM and the APT are more general than the CAPM as they allow for multiple risk factors. They both offer the advantage of not necessarily requiring the identification of the wealth portfolio. On the other hand, the nature and the number of factors are not specified by the underlying models. For more details on multifactor asset pricing models, their underlying assumptions and the various factor selection procedures, see, for example, Fama and French (1996), Campbell et al. (1997, §6) or Cochrane (2005, §9).

Appendix B

Figures
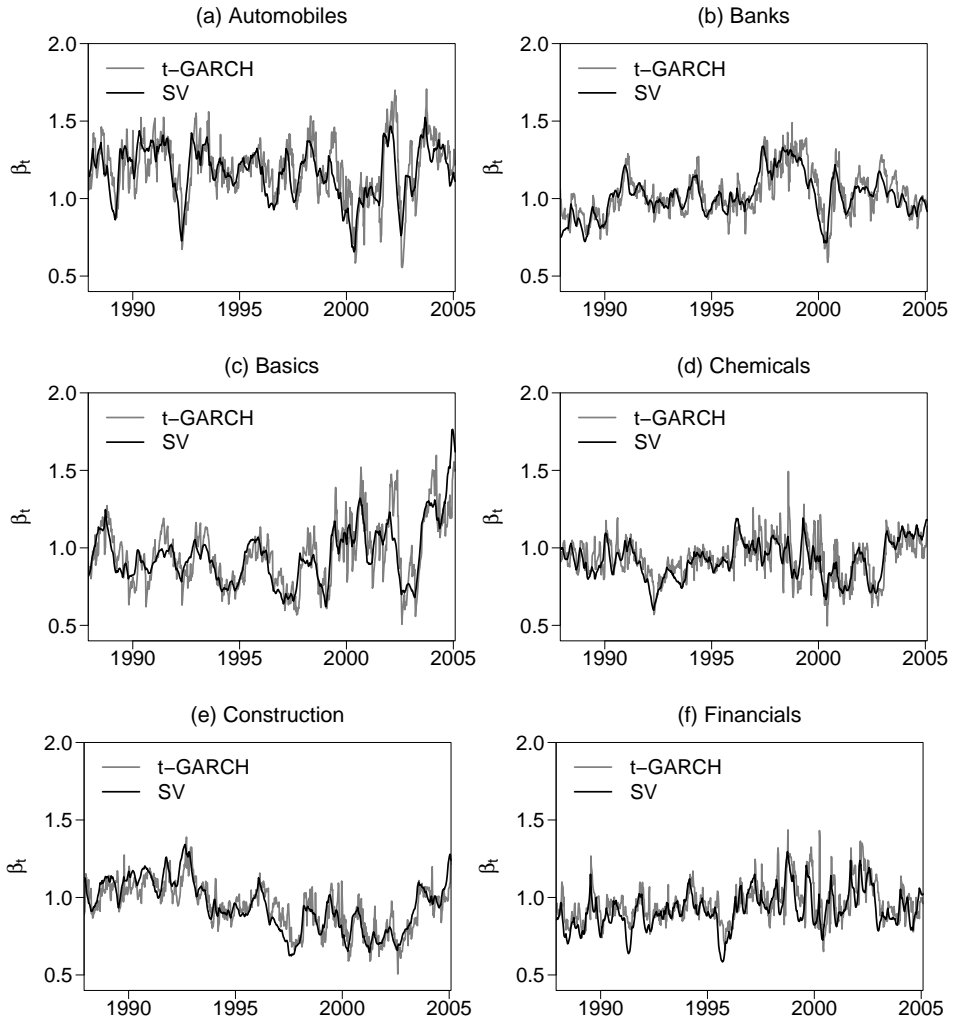
Figure B.1: *t-GARCH and stochastic volatility conditional betas.*
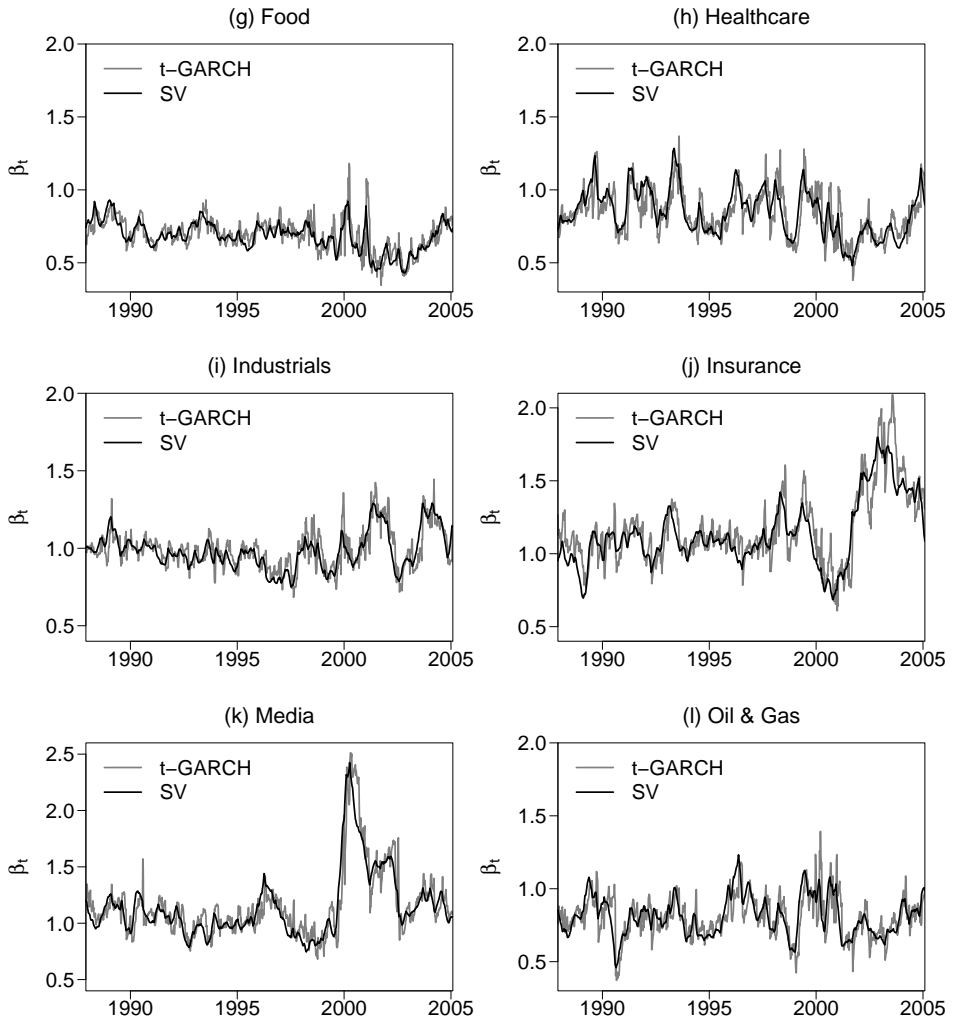
Figure B.1 — continued
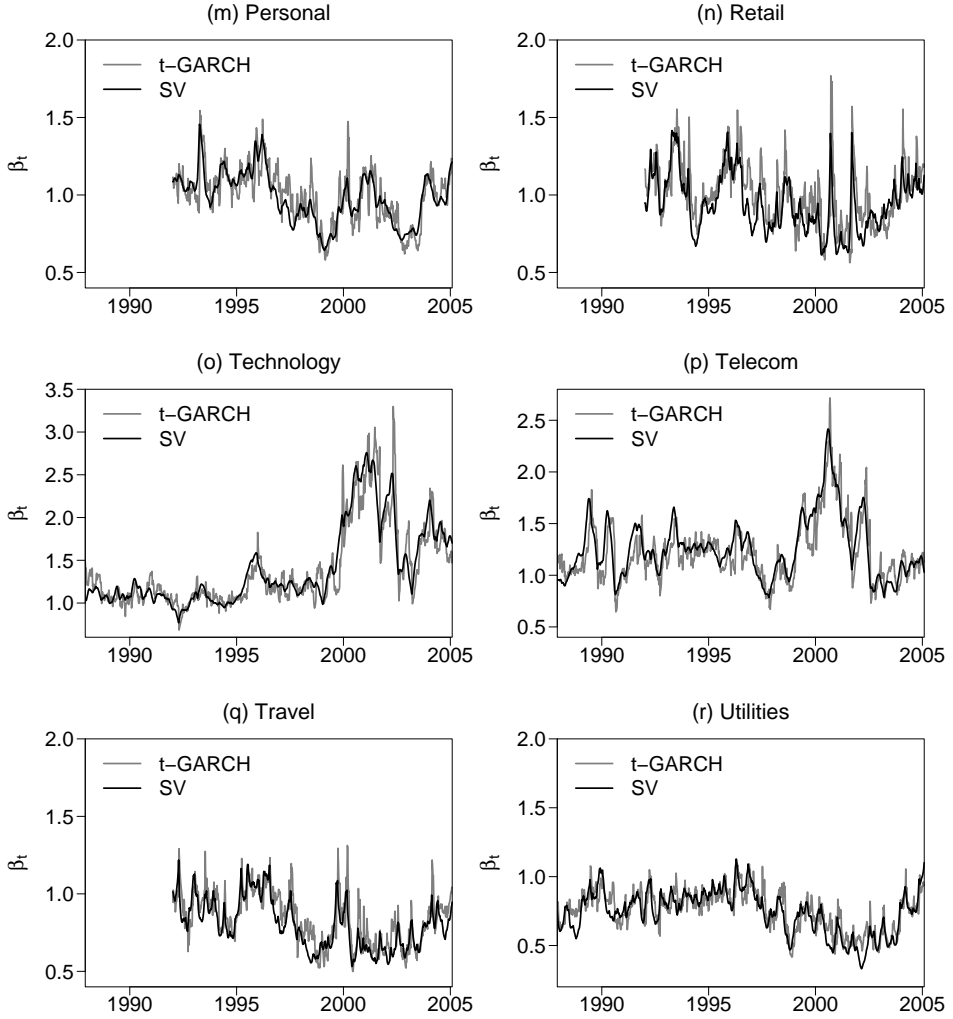
Figure B.1 — continued
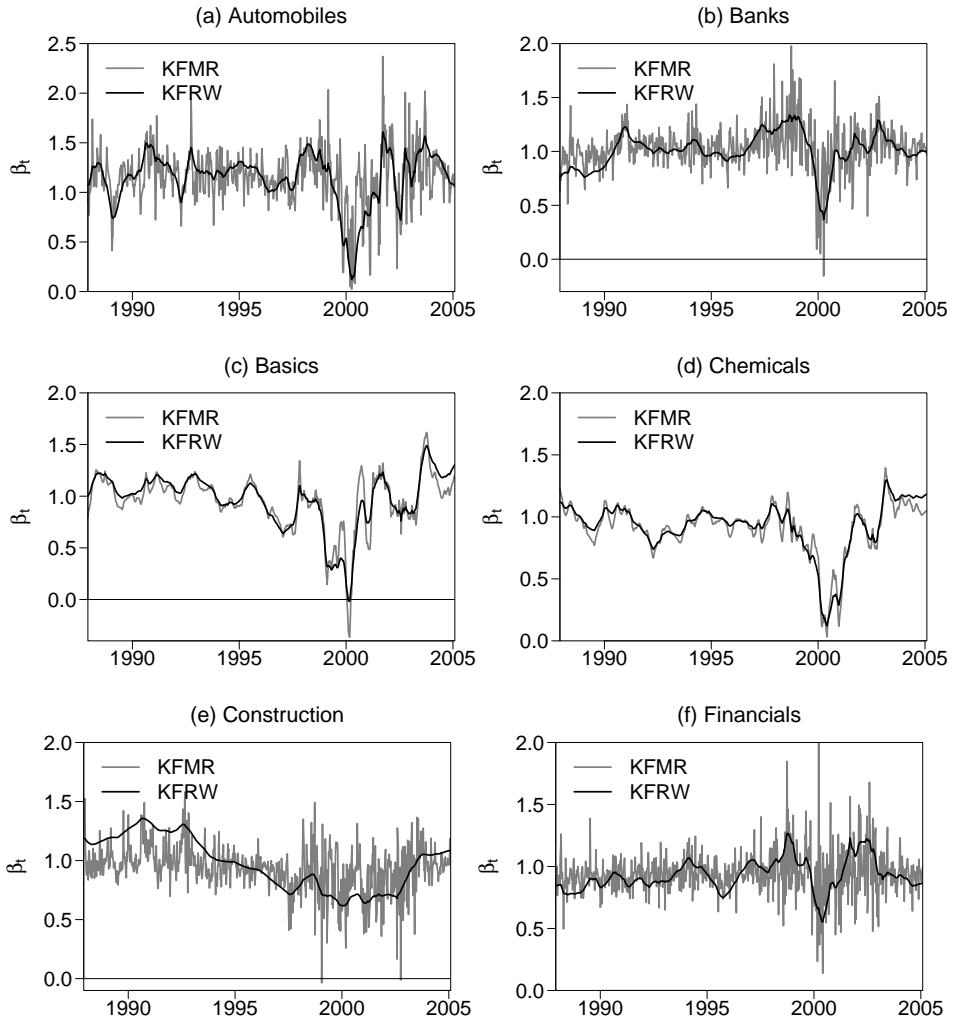
Figure B.2: *Random walk and mean reverting conditional betas.*
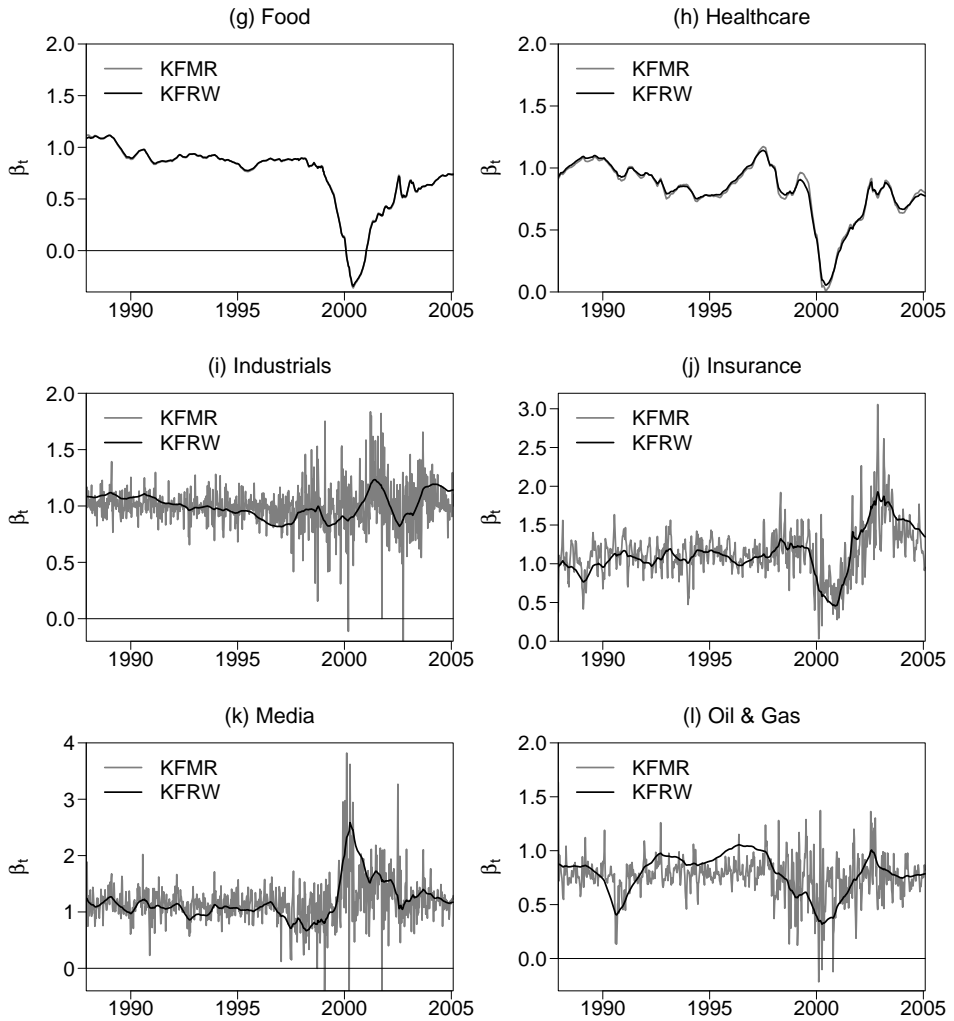
Figure B.2 — continued
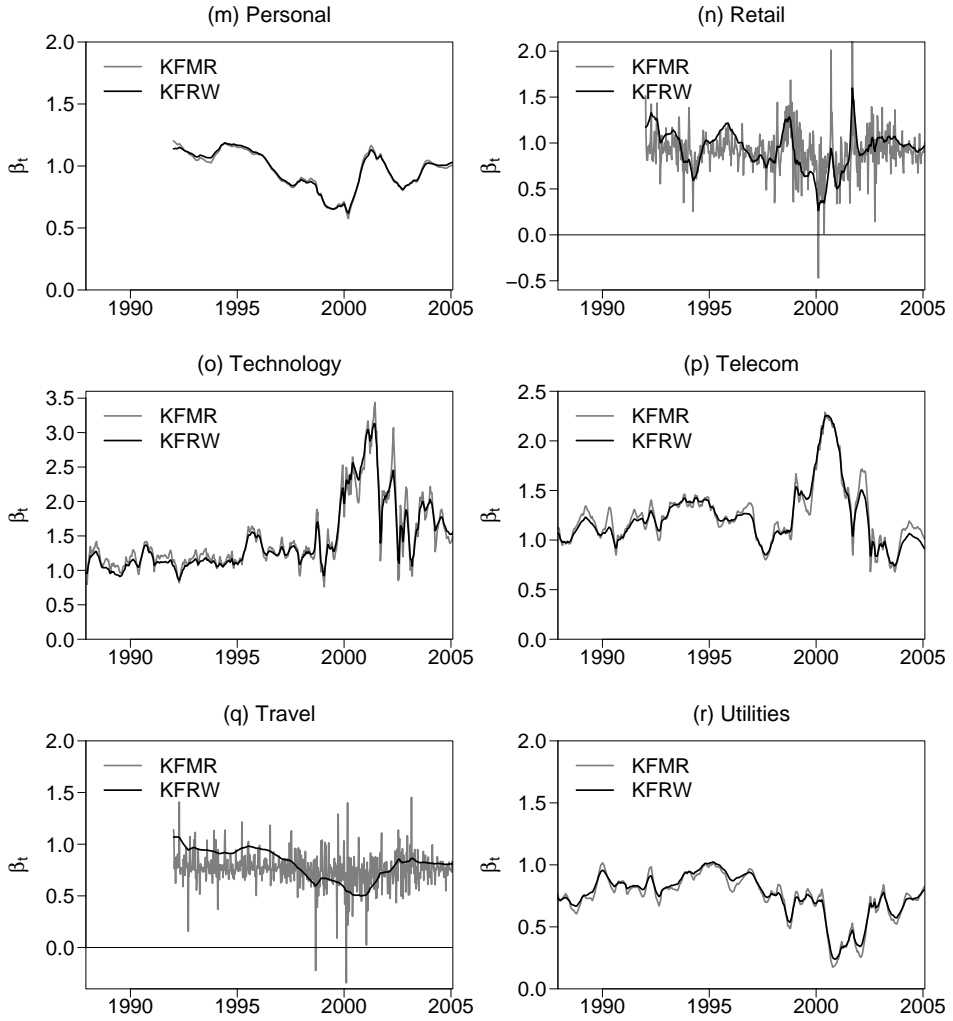
Figure B.2 — continued
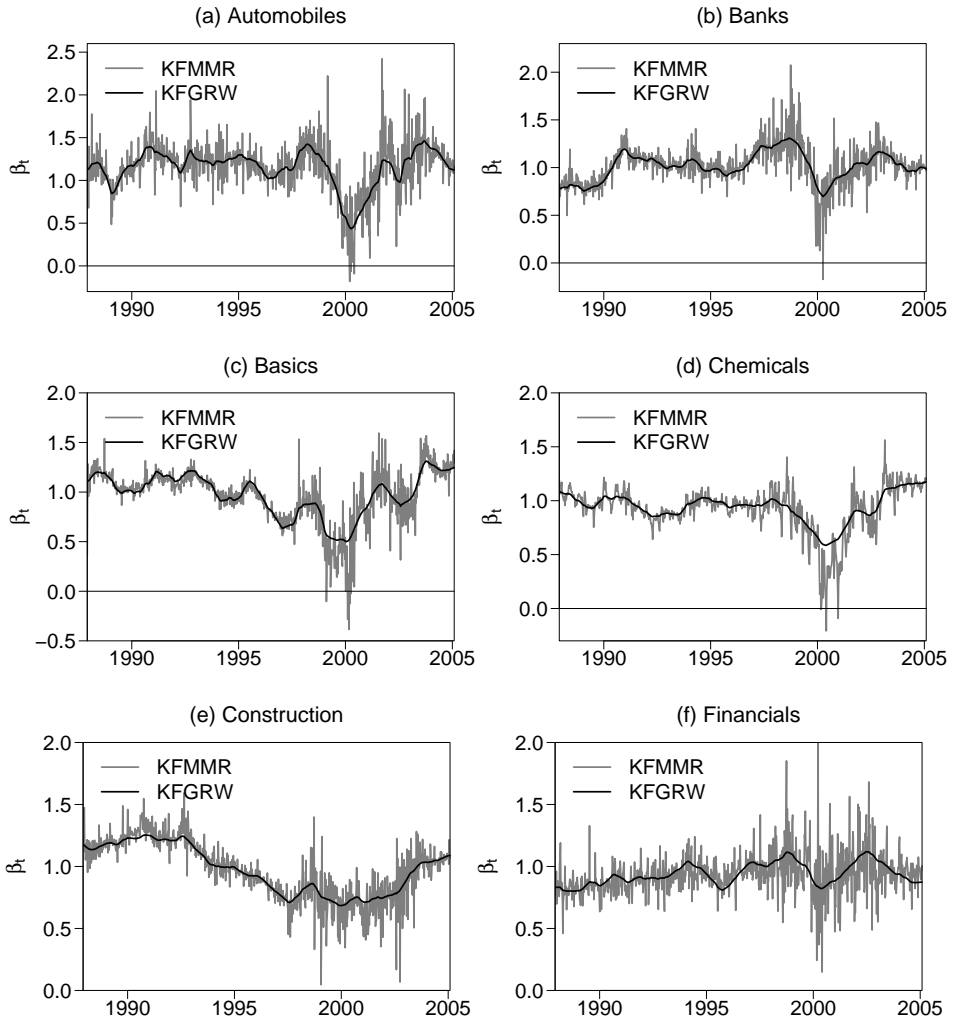
Figure B.3: *Moving mean reverting and generalized random walk conditional betas.*
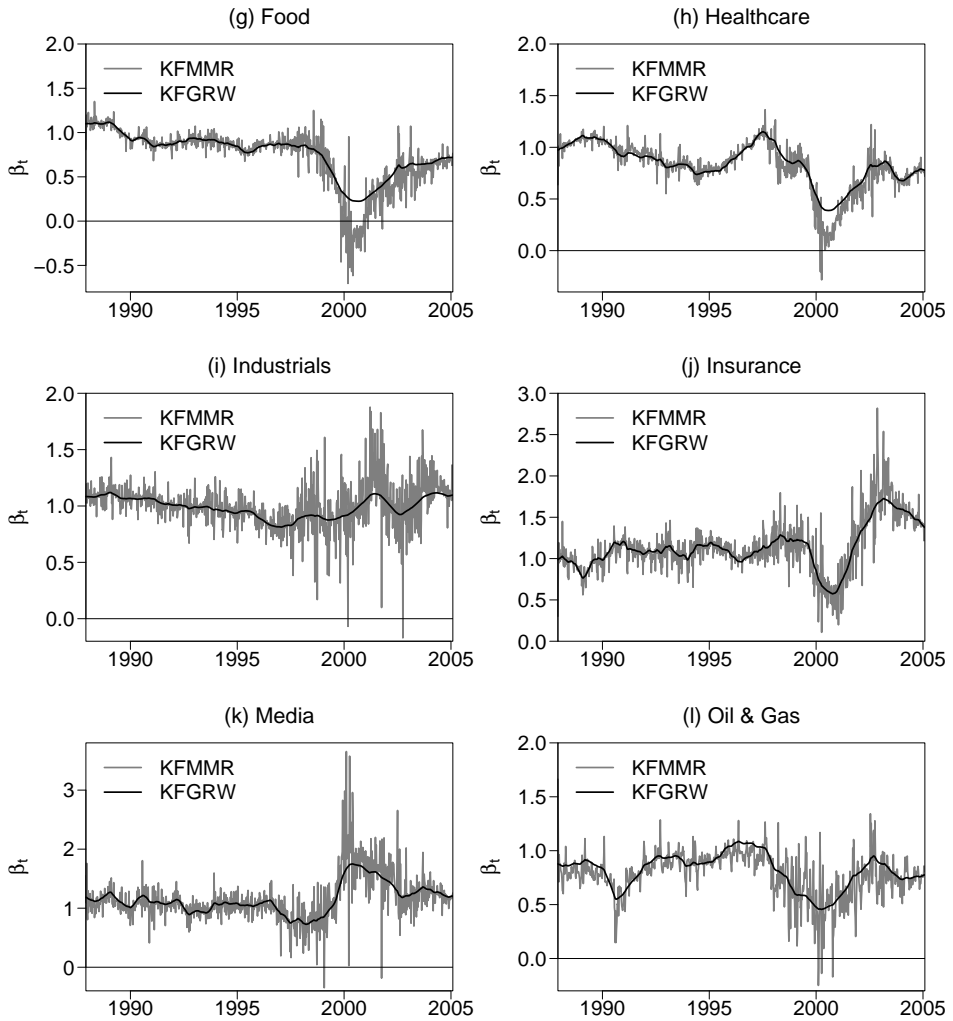
Figure B.3 — continued
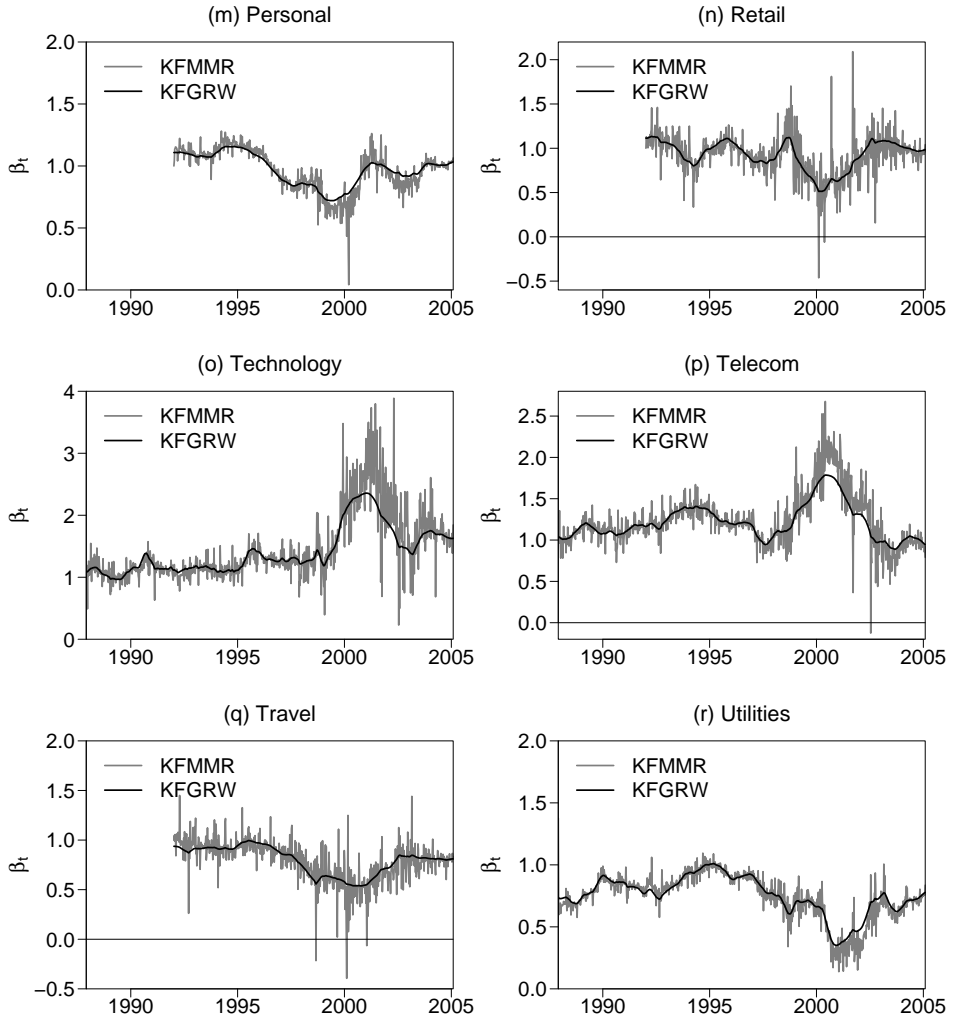
Figure B.3 — continued
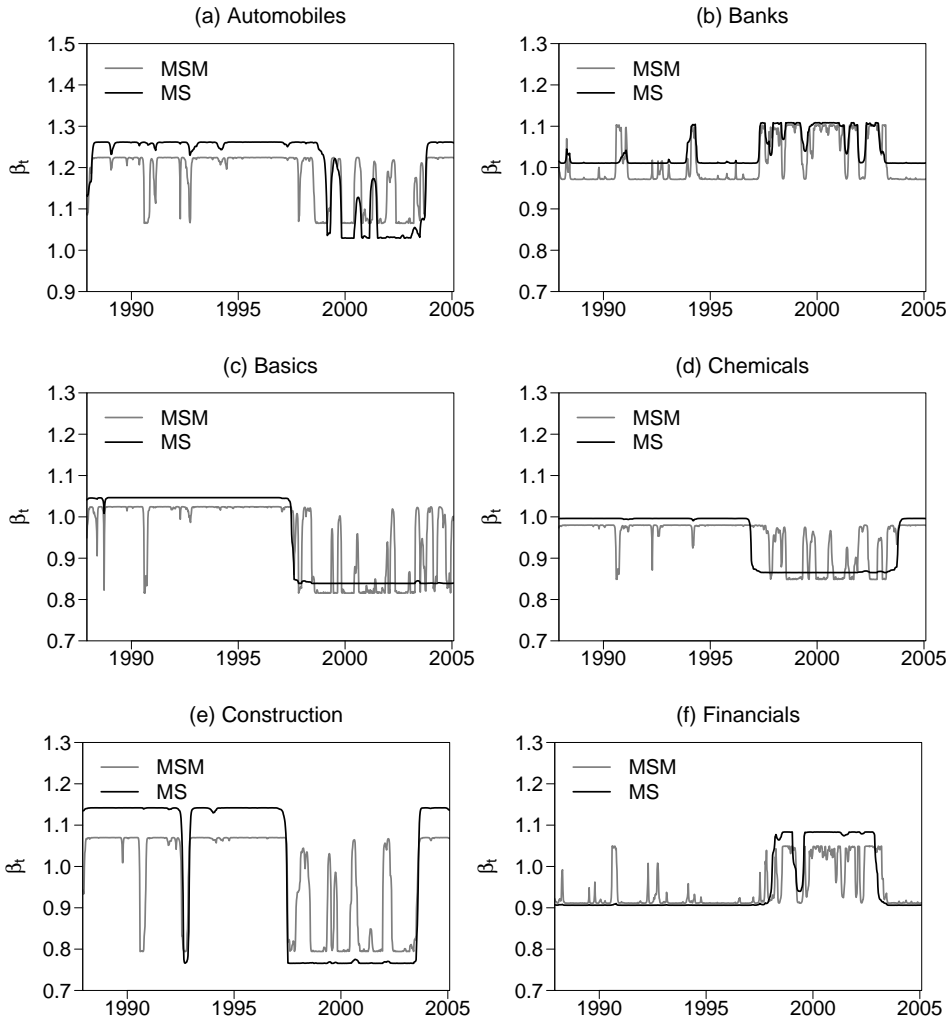
Figure B.4: *Markov switching and Markov switching market conditional betas.*
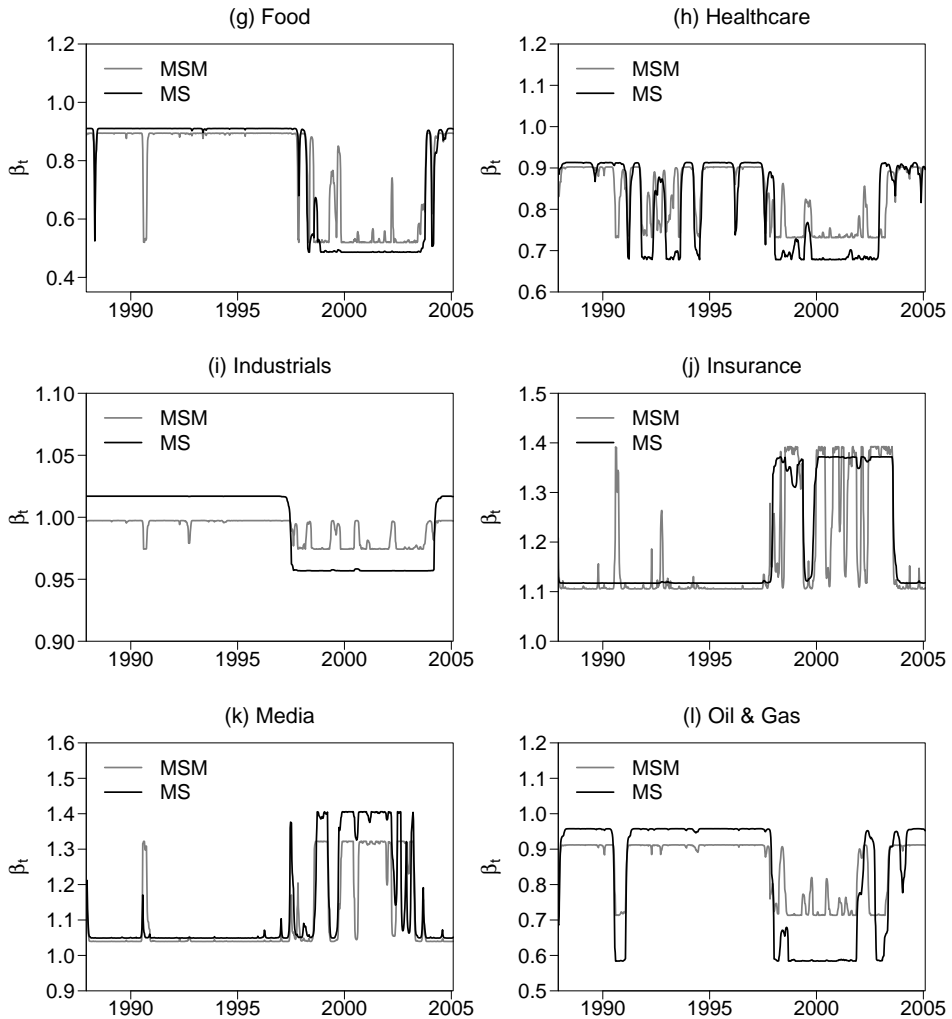
Figure B.4 — continued
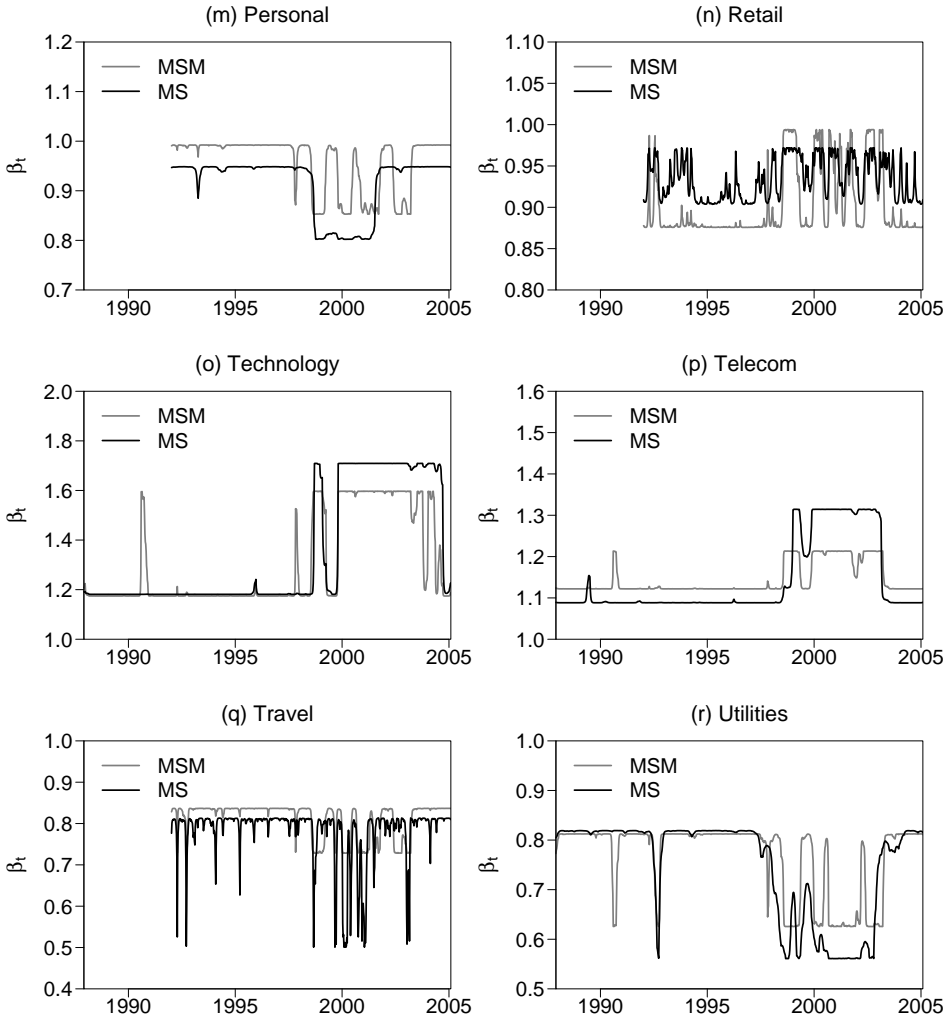
Figure B.4 — continued
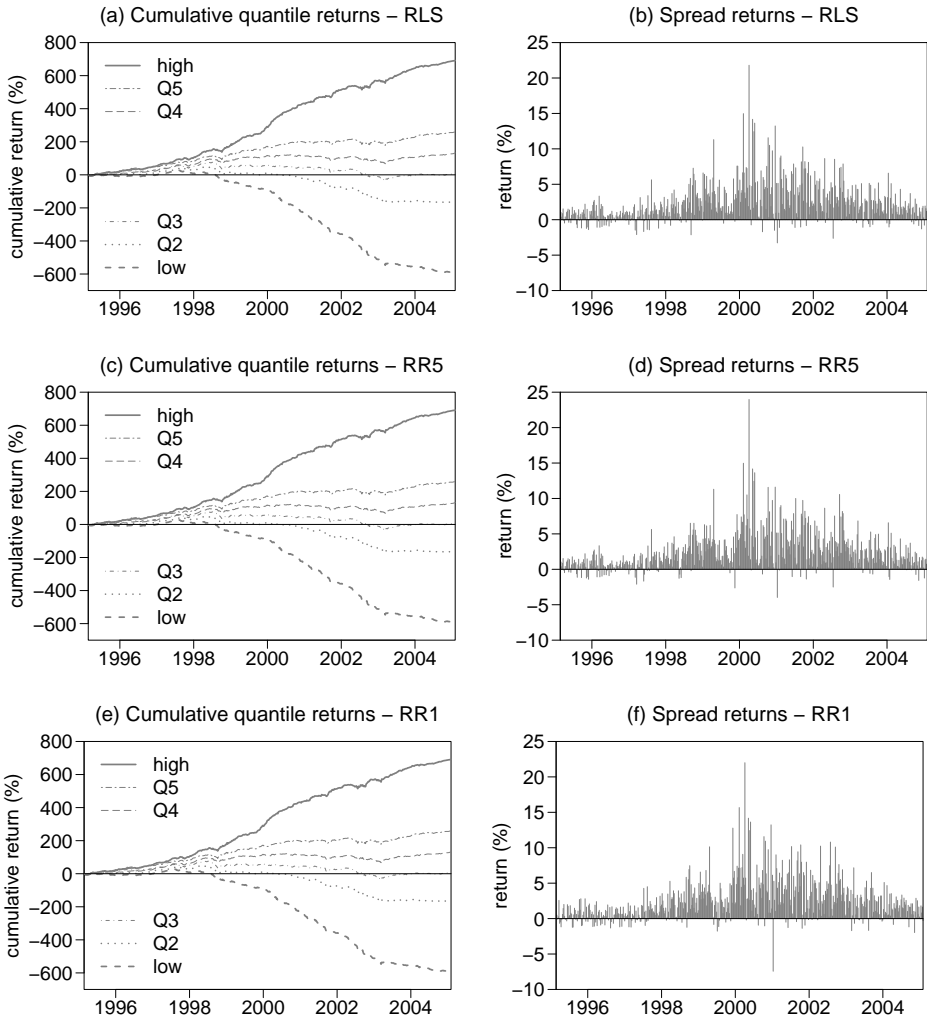
Figure B.5: *Cumulative quintile returns and spread returns for the RLS, RR5 and RR1 portfolios.*

Appendix C

Tables

Table C.1: *Comparison of OLS betas and various conditional beta series.*

This table characterizes the various conditional beta series by reporting the mean betas and their range (in parentheses). As the Kalman filter is likely to produce large outliers in the first stages of estimation, the first fifty conditional beta estimates for any of the chosen techniques to model conditional betas are not included in the subsequent analyses to avoid an unfair bias against the Kalman filter.

Sector	β^{OLS}	β^{tG}	β^{SV}	β^{RW}	β^{MR}	β^{MMR}	β^{GRW}	β^{MSM}	β^{MS}
Automobiles	1.15	1.19	1.17	1.15	1.15	1.15	1.15	1.18	1.20
		(0.56;1.71)	(0.66;1.52)	(0.12;1.61)	(0.02;2.37)	(−0.18;2.42)	(0.43;1.46)	(1.07;1.22)	(1.03;1.26)
Banks	1.06	1.03	1.03	1.02	1.03	1.03	1.03	1.01	1.04
		(0.59;1.49)	(0.71;1.34)	(0.37;1.34)	(−0.16;1.98)	(−0.17;2.07)	(0.70;1.31)	(0.97;1.10)	(1.01;1.11)
Basics	0.90	0.96	0.95	0.96	0.94	0.95	0.96	0.95	0.95
		(0.51;1.60)	(0.62;1.76)	(−0.02;1.49)	(−0.36;1.62)	(−0.38;1.60)	(0.50;1.31)	(0.82;1.02)	(0.84;1.05)
Chemicals	0.91	0.93	0.92	0.91	0.90	0.91	0.94	0.95	0.94
		(0.50;1.49)	(0.60;1.19)	(0.12;1.30)	(0.03;1.40)	(−0.21;1.56)	(0.59;1.17)	(0.85;0.98)	(0.87;1.00)
Construction	0.89	0.94	0.94	0.96	0.93	0.96	0.96	0.98	0.99
		(0.51;1.39)	(0.62;1.34)	(0.62;1.36)	(−0.04;1.58)	(0.05;1.59)	(0.68;1.25)	(0.79;1.07)	(0.77;1.14)
Financials	1.00	0.97	0.93	0.94	0.95	0.94	0.95	0.95	0.96
		(0.65;1.44)	(0.58;1.30)	(0.55;1.27)	(0.14;2.08)	(0.15;2.08)	(0.80;1.12)	(0.91;1.05)	(0.91;1.08)
Food	0.65	0.69	0.68	0.71	0.71	0.71	0.74	0.77	0.76
		(0.35;1.18)	(0.43;0.93)	(−0.35;1.12)	(−0.36;1.12)	(−0.70;1.25)	(0.22;1.12)	(0.52;0.89)	(0.49;0.91)
Healthcare	0.78	0.84	0.85	0.81	0.81	0.81	0.83	0.83	0.81
		(0.38;1.37)	(0.48;1.28)	(0.06;1.14)	(0.01;1.17)	(−0.28;1.36)	(0.39;1.15)	(0.73;0.90)	(0.68;0.91)
Industrials	0.98	0.99	0.98	0.99	1.00	1.00	0.98	0.99	0.99
		(0.68;1.45)	(0.75;1.29)	(0.82;1.23)	(−0.20;1.84)	(−0.17;1.88)	(0.81;1.12)	(0.97;1.00)	(0.96;1.02)

Continued on the next page...

Table C.1 — continued

Sector	β^{OLS}	β^{tG}	β^{SV}	β^{RW}	β^{MR}	β^{MMR}	β^{GRW}	β^{MSM}	β^{MS}
Insurance	1.27	1.17	1.15	1.14	1.15	1.15	1.14	1.18	1.20
		(0.61;2.09)	(0.68;1.80)	(0.46;1.93)	(0.03;3.05)	(0.11;2.82)	(0.57;1.73)	(1.11;1.39)	(1.12;1.37)
Media	1.21	1.17	1.15	1.18	1.18	1.16	1.15	1.11	1.13
		(0.68;2.51)	(0.75;2.43)	(0.67;2.59)	(−0.54;3.82)	(−0.35;3.65)	(0.73;1.75)	(1.04;1.32)	(1.05;1.41)
Oil & Gas	0.76	0.81	0.81	0.78	0.75	0.78	0.81	0.85	0.83
		(0.37;1.39)	(0.46;1.23)	(0.32;1.06)	(−0.22;1.37)	(−0.25;1.34)	(0.46;1.09)	(0.71;0.91)	(0.58;0.96)
Personal	0.91	0.99	0.98	0.96	0.95	0.96	0.96	0.96	0.91
		(0.58;1.54)	(0.64;1.46)	(0.62;1.19)	(0.58;1.19)	(0.04;1.28)	(0.72;1.16)	(0.85;0.99)	(0.80;0.95)
Retail	0.95	1.00	0.94	0.91	0.90	0.90	0.91	0.90	0.93
		(0.56;1.77)	(0.61;1.42)	(0.26;1.60)	(−0.47;2.11)	(−0.46;2.09)	(0.51;1.12)	(0.88;0.99)	(0.90;0.97)
Technology	1.49	1.40	1.42	1.46	1.49	1.47	1.42	1.31	1.36
		(0.68;3.30)	(0.77;2.76)	(0.85;3.13)	(0.76;3.44)	(0.23;3.89)	(0.96;2.36)	(1.17;1.60)	(1.18;1.71)
Telecom	1.19	1.23	1.27	1.25	1.27	1.25	1.22	1.15	1.15
		(0.65;2.72)	(0.78;2.42)	(0.74;2.26)	(0.68;2.29)	(−0.13;2.68)	(0.89;1.79)	(1.12;1.21)	(1.09;1.31)
Travel	0.77	0.84	0.79	0.79	0.75	0.78	0.79	0.81	0.78
		(0.50;1.31)	(0.53;1.19)	(0.50;0.98)	(−0.34;1.45)	(−0.39;1.44)	(0.54;1.00)	(0.73;0.84)	(0.50;0.81)
Utilities	0.69	0.76	0.75	0.75	0.74	0.75	0.76	0.76	0.74
		(0.42;1.10)	(0.33;1.13)	(0.24;1.02)	(0.17;1.02)	(0.14;1.09)	(0.35;1.01)	(0.63;0.81)	(0.56;0.82)

Table C.2: *In-sample mean absolute errors.*

This table reports the estimated in-sample MAE ($\times 10^2$) for the eighteen DJ Stoxx℠ sectors. For each sector i, figures in parentheses denote the relative rank of a model's MAE, where the model with the smallest MAE ranks first.

Sector	β^{OLS}	β^{tG}	β^{SV}	β^{RW}	β^{MR}	β^{MMR}	β^{GRW}	β^{MSM}	β^{MS}
Automobiles	1.469	1.484	1.386	1.350	1.120	1.081	1.399	1.468	1.456
	(8)	(9)	(4)	(3)	(2)	(1)	(5)	(7)	(6)
Banks	0.796	0.795	0.742	0.736	0.536	0.550	0.752	0.786	0.791
	(9)	(8)	(4)	(3)	(1)	(2)	(5)	(6)	(7)
Basics	1.365	1.383	1.339	1.240	1.191	1.078	1.286	1.355	1.353
	(8)	(9)	(5)	(3)	(2)	(1)	(4)	(7)	(6)
Chemicals	1.095	1.090	1.036	1.007	0.973	0.886	1.044	1.090	1.089
	(9)	(8)	(4)	(3)	(2)	(1)	(5)	(7)	(6)
Construction	1.015	0.982	0.926	0.932	0.724	0.759	0.938	0.988	0.979
	(9)	(7)	(3)	(4)	(1)	(2)	(5)	(8)	(6)
Financials	0.869	0.878	0.790	0.810	0.600	0.600	0.834	0.856	0.854
	(8)	(9)	(3)	(4)	(1)	(2)	(5)	(7)	(6)
Food	1.042	1.039	1.000	0.884	0.881	0.723	0.917	0.978	0.956
	(9)	(8)	(7)	(3)	(2)	(1)	(4)	(6)	(5)
Healthcare	1.327	1.344	1.283	1.250	1.242	1.136	1.269	1.315	1.304
	(8)	(9)	(5)	(3)	(2)	(1)	(4)	(7)	(6)
Industrials	0.735	0.726	0.676	0.692	0.455	0.458	0.704	0.734	0.732
	(9)	(6)	(3)	(4)	(1)	(2)	(5)	(8)	(7)
Insurance	1.146	1.076	0.997	0.988	0.727	0.737	1.004	1.130	1.132
	(9)	(6)	(4)	(3)	(1)	(2)	(5)	(7)	(8)
Media	1.362	1.276	1.204	1.184	0.812	0.858	1.230	1.340	1.324
	(9)	(6)	(4)	(3)	(1)	(2)	(5)	(8)	(7)
Oil & Gas	1.486	1.498	1.462	1.433	1.231	1.238	1.443	1.473	1.451
	(8)	(9)	(6)	(3)	(1)	(2)	(4)	(7)	(5)
Personal	0.971	0.963	0.901	0.918	0.912	0.824	0.933	0.966	0.967
	(9)	(6)	(2)	(4)	(3)	(1)	(5)	(7)	(8)
Retail	1.350	1.369	1.264	1.246	1.026	1.018	1.291	1.347	1.350
	(8)	(9)	(4)	(3)	(2)	(1)	(5)	(6)	(7)
Technology	1.693	1.550	1.477	1.402	1.328	1.102	1.492	1.616	1.595
	(9)	(6)	(4)	(3)	(2)	(1)	(5)	(8)	(7)
Telecom	1.515	1.469	1.412	1.397	1.357	1.178	1.436	1.514	1.499
	(9)	(6)	(4)	(3)	(2)	(1)	(5)	(8)	(7)
Travel	0.993	1.015	0.936	0.959	0.747	0.748	0.962	0.995	0.970
	(7)	(9)	(3)	(4)	(1)	(2)	(5)	(8)	(6)
Utilities	0.980	0.965	0.901	0.903	0.881	0.816	0.920	0.964	0.949
	(9)	(8)	(3)	(4)	(2)	(1)	(5)	(7)	(6)
Average MAE	1.178	1.161	1.096	1.074	0.930	0.877	1.103	1.162	1.153
Average Rank	8.56	7.67	4.00	3.33	1.61	1.44	4.78	7.17	6.44

Table C.3: *In-sample mean squared errors.*

This table reports the estimated in-sample MSE $(\times 10^4)$ for the eighteen DJ Stoxx$^{\text{SM}}$ sectors. For each sector i, figures in parentheses denote the relative rank of a model's MSE, where the model with the smallest MSE ranks first.

Sector	β^{OLS}	β^{tG}	β^{SV}	β^{RW}	β^{MR}	β^{MMR}	β^{GRW}	β^{MSM}	β^{MS}
Automobiles	3.935	3.978	3.350	3.069	2.129	1.988	3.374	3.925	3.855
	(8)	(9)	(4)	(3)	(2)	(1)	(5)	(7)	(6)
Banks	1.336	1.328	1.106	1.043	0.571	0.590	1.131	1.314	1.323
	(9)	(8)	(4)	(3)	(1)	(2)	(5)	(6)	(7)
Basics	3.821	4.022	3.723	3.158	2.894	2.409	3.410	3.781	3.778
	(8)	(9)	(5)	(3)	(2)	(1)	(4)	(7)	(6)
Chemicals	2.308	2.304	2.042	1.852	1.727	1.457	2.033	2.289	2.291
	(9)	(8)	(5)	(3)	(2)	(1)	(4)	(6)	(7)
Construction	1.889	1.741	1.565	1.568	0.958	1.029	1.603	1.800	1.724
	(9)	(7)	(3)	(4)	(1)	(2)	(5)	(8)	(6)
Financials	1.472	1.502	1.197	1.240	0.682	0.684	1.353	1.447	1.428
	(8)	(9)	(3)	(4)	(1)	(2)	(5)	(7)	(6)
Food	2.333	2.442	2.240	1.680	1.670	1.145	1.857	2.170	2.099
	(8)	(9)	(7)	(3)	(2)	(1)	(4)	(6)	(5)
Healthcare	3.324	3.509	3.177	2.884	2.842	2.392	2.991	3.288	3.238
	(8)	(9)	(5)	(3)	(2)	(1)	(4)	(7)	(6)
Industrials	1.116	1.105	0.960	0.994	0.421	0.430	1.042	1.115	1.112
	(9)	(6)	(3)	(4)	(1)	(2)	(5)	(8)	(7)
Insurance	2.713	2.325	1.899	1.811	0.968	0.995	1.955	2.608	2.624
	(9)	(6)	(4)	(3)	(1)	(2)	(5)	(7)	(8)
Media	4.035	3.443	3.036	2.966	1.344	1.509	3.291	3.909	3.836
	(9)	(6)	(4)	(3)	(1)	(2)	(5)	(8)	(7)
Oil & Gas	4.173	4.366	4.101	3.810	2.879	2.902	3.870	4.106	3.937
	(8)	(9)	(6)	(3)	(1)	(2)	(4)	(7)	(5)
Personal	1.741	1.816	1.577	1.535	1.515	1.234	1.602	1.718	1.713
	(8)	(9)	(4)	(3)	(2)	(1)	(5)	(7)	(6)
Retail	3.559	3.766	3.000	2.855	1.931	1.936	3.325	3.538	3.551
	(8)	(9)	(4)	(3)	(1)	(2)	(5)	(6)	(7)
Technology	6.282	5.652	4.806	4.282	3.803	2.685	5.108	6.002	5.900
	(9)	(6)	(4)	(3)	(2)	(1)	(5)	(8)	(7)
Telecom	4.412	4.091	3.644	3.495	3.295	2.527	3.790	4.406	4.340
	(9)	(6)	(4)	(3)	(2)	(1)	(5)	(8)	(7)
Travel	1.928	1.997	1.757	1.765	1.089	1.080	1.776	1.912	1.798
	(8)	(9)	(3)	(4)	(2)	(1)	(5)	(7)	(6)
Utilities	1.613	1.571	1.374	1.365	1.308	1.118	1.416	1.556	1.498
	(9)	(8)	(4)	(3)	(2)	(1)	(5)	(7)	(6)
Average MSE	2.888	2.831	2.475	2.298	1.779	1.562	2.496	2.827	2.780
Average Rank	8.50	7.89	4.22	3.22	1.56	1.44	4.72	7.06	6.39

Table C.4: *Out-of-sample mean absolute errors (100 samples).*

This table reports the estimated out-of-sample MAE $(\times 10^2)$ for the eighteen DJ Stoxx℠ sectors. For each sector i, figures in parentheses denote the relative rank of a model's MAE, where the model with the smallest MAE ranks first.

Sector	β^{OLS}	β^{tG}	β^{SV}	β^{RW}	β^{MR}	β^{MMR}	β^{GRW}	β^{MSM}	β^{MS}
Automobiles	1.309	1.287	1.294	1.289	1.336	1.257	1.254	1.279	1.299
	(8)	(4)	(6)	(5)	(9)	(2)	(1)	(3)	(7)
Banks	0.456	0.489	0.482	0.478	0.447	0.480	0.488	0.459	0.454
	(3)	(9)	(7)	(5)	(1)	(6)	(8)	(4)	(2)
Basics	1.584	1.496	1.517	1.485	1.495	1.493	1.489	1.568	1.583
	(9)	(5)	(6)	(1)	(4)	(3)	(2)	(7)	(8)
Chemicals	1.106	0.991	1.026	0.931	0.959	0.942	0.948	1.083	1.090
	(9)	(5)	(6)	(1)	(4)	(2)	(3)	(7)	(8)
Construction	0.938	0.872	0.863	0.870	0.926	0.882	0.881	0.909	0.916
	(9)	(3)	(1)	(2)	(8)	(5)	(4)	(6)	(7)
Financials	0.679	0.710	0.699	0.694	0.666	0.678	0.699	0.673	0.673
	(5)	(9)	(8)	(6)	(1)	(4)	(7)	(3)	(2)
Food	0.941	0.959	0.957	0.931	FTC[a]	0.929	0.935	0.933	0.951
	(5)	(8)	(7)	(2)	(−)	(1)	(4)	(3)	(6)
Healthcare	1.032	1.096	1.095	1.024	1.022	1.028	1.021	1.046	1.011
	(6)	(9)	(8)	(4)	(3)	(5)	(2)	(7)	(1)
Industrials	0.794	0.738	0.728	0.740	0.788	0.741	0.745	0.806	0.799
	(7)	(2)	(1)	(3)	(6)	(4)	(5)	(9)	(8)
Insurance	1.098	0.991	1.011	1.017	1.080	1.000	1.004	1.127	1.097
	(8)	(1)	(4)	(5)	(6)	(2)	(3)	(9)	(7)
Media	0.941	1.012	1.028	0.964	0.956	0.973	0.964	0.975	1.032
	(1)	(7)	(8)	(3)	(2)	(5)	(4)	(6)	(9)
Oil & Gas	1.232	1.261	1.251	1.244	1.239	1.245	1.255	1.259	1.261
	(1)	(8)	(5)	(3)	(2)	(4)	(6)	(7)	(9)
Personal	0.665	0.666	0.667	0.662	0.655	0.659	0.655	0.662	0.668
	(6)	(7)	(8)	(5)	(2)	(3)	(1)	(4)	(9)
Retail	1.072	1.141	1.123	1.122	1.108	1.095	1.092	1.107	1.099
	(1)	(9)	(8)	(7)	(6)	(3)	(2)	(5)	(4)
Technology	1.818	1.778	1.752	1.820	1.813	1.762	1.781	1.900	1.885
	(6)	(3)	(1)	(7)	(5)	(2)	(4)	(9)	(8)
Telecom	1.161	1.089	1.078	1.057	1.053	1.044	1.065	1.147	1.129
	(9)	(6)	(5)	(3)	(2)	(1)	(4)	(8)	(7)
Travel	0.738	0.757	0.753	0.739	0.741	0.736	0.739	0.748	0.761
	(2)	(8)	(7)	(4)	(5)	(1)	(3)	(6)	(9)
Utilities	0.833	0.838	0.824	0.846	0.843	0.842	0.849	0.902	0.831
	(3)	(4)	(1)	(7)	(6)	(5)	(8)	(9)	(2)
Average MAE	1.022	1.009	1.008	0.995	1.007	0.988	0.992	1.032	1.030
Average Rank	5.44	5.94	5.39	4.06	4.24	3.22	3.94	6.22	6.28

[a]Failed to converge.

Table C.5: *Out-of-sample mean squared errors (100 samples).*

This table reports the estimated out-of-sample MSE ($\times 10^4$) for the eighteen DJ Stoxx℠ sectors. For each sector i, figures in parentheses denote the relative rank of a model's MSE, where the model with the smallest MSE ranks first.

Sector	β^{OLS}	β^{tG}	β^{SV}	β^{RW}	β^{MR}	β^{MMR}	β^{GRW}	β^{MSM}	β^{MS}
Automobiles	3.146	3.101	3.205	3.176	3.223	2.999	2.993	3.027	3.128
	(6)	(4)	(8)	(7)	(9)	(2)	(1)	(3)	(5)
Banks	0.349	0.373	0.362	0.364	0.328	0.373	0.387	0.344	0.344
	(4)	(8)	(5)	(6)	(1)	(7)	(7)	(3)	(2)
Basics	4.062	3.566	3.744	3.538	3.606	3.570	3.558	3.962	4.057
	(9)	(3)	(6)	(1)	(5)	(4)	(2)	(7)	(8)
Chemicals	2.177	1.670	1.789	1.424	1.443	1.448	1.451	2.198	2.184
	(7)	(5)	(6)	(1)	(2)	(3)	(4)	(9)	(8)
Construction	1.500	1.317	1.297	1.312	1.462	1.355	1.352	1.443	1.441
	(9)	(3)	(1)	(2)	(8)	(5)	(4)	(7)	(6)
Financials	0.878	0.959	0.886	0.894	0.817	0.855	0.916	0.879	0.826
	(4)	(9)	(6)	(7)	(1)	(3)	(6)	(5)	(2)
Food	1.384	1.471	1.457	1.374	FTC[a]	1.362	1.388	1.952	1.476
	(3)	(6)	(5)	(2)	(−)	(1)	(4)	(8)	(7)
Healthcare	1.864	2.179	2.151	1.838	1.815	1.840	1.805	1.952	1.780
	(6)	(9)	(8)	(4)	(3)	(5)	(1)	(7)	(1)
Industrials	1.095	1.026	0.992	1.027	1.082	1.003	1.013	1.124	1.114
	(7)	(4)	(1)	(5)	(6)	(2)	(3)	(9)	(8)
Insurance	2.695	1.996	2.134	2.495	2.362	2.164	2.084	2.737	2.622
	(8)	(1)	(3)	(6)	(5)	(4)	(2)	(9)	(7)
Media	1.499	1.740	1.788	1.564	1.535	1.590	1.553	1.626	1.845
	(1)	(7)	(8)	(4)	(2)	(5)	(3)	(6)	(9)
Oil & Gas	2.510	2.665	2.646	2.528	2.549	2.550	2.553	2.613	2.641
	(1)	(9)	(8)	(2)	(3)	(4)	(5)	(6)	(7)
Personal	0.715	0.737	0.745	0.696	0.686	0.693	0.687	0.702	0.722
	(6)	(8)	(9)	(4)	(1)	(3)	(2)	(5)	(7)
Retail	2.237	2.623	2.527	2.356	2.360	2.278	2.272	2.329	2.369
	(1)	(9)	(8)	(5)	(6)	(3)	(2)	(4)	(7)
Technology	5.790	5.409	5.428	5.822	5.735	5.569	5.599	6.448	6.322
	(6)	(1)	(2)	(7)	(5)	(3)	(4)	(9)	(8)
Telecom	2.316	2.045	1.993	1.905	1.895	1.869	1.949	2.274	2.213
	(9)	(6)	(5)	(3)	(2)	(1)	(4)	(8)	(7)
Travel	0.944	0.976	0.986	0.911	0.938	0.910	0.912	0.977	1.014
	(5)	(6)	(8)	(2)	(4)	(1)	(3)	(7)	(9)
Utilities	1.127	1.103	1.087	1.194	1.192	1.148	1.169	1.269	1.150
	(3)	(2)	(1)	(8)	(7)	(4)	(5)	(9)	(5)
Average MSE	2.016	1.942	1.957	1.912	1.943	1.865	1.869	2.103	2.069
Average Rank	5.28	5.56	5.44	4.22	4.12	3.33	3.44	6.72	6.28

[a]Failed to converge.

Table C.6: *Out-of-sample mean absolute and squared errors (520 samples).*

This table reports the estimated out-of-sample mean errors for an out-of-sample period of ten years. For each sector i, figures in parentheses denote the relative rank of a model's mean absolute and squared error, respectively, where the model with the smallest error ranks first.

Sector	MAE ($\times 10^2$)			MSE ($\times 10^4$)		
	β^{RW}	β^{GRW}	β^{MMR}	β^{RW}	β^{GRW}	β^{MMR}
Automobiles	1.607	1.596	1.598	4.534	4.568	4.480
	(3)	(1)	(2)	(2)	(3)	(1)
Banks	0.905	0.907	0.905	1.740	1.777	1.722
	(2)	(3)	(1)	(2)	(3)	(1)
Basics	1.705	1.703	1.710	5.495	5.490	5.546
	(2)	(1)	(3)	(2)	(1)	(3)
Chemicals	1.286	1.294	1.286	2.909	2.944	2.916
	(2)	(3)	(1)	(1)	(3)	(2)
Construction	1.077	1.071	1.075	2.127	2.117	2.137
	(3)	(1)	(2)	(2)	(1)	(3)
Financials	1.019	1.013	1.012	1.999	1.983	2.008
	(3)	(2)	(1)	(2)	(1)	(3)
Food	1.167	1.177	1.172	2.777	2.828	2.795
	(1)	(3)	(2)	(1)	(3)	(2)
Healthcare	1.405	1.397	1.397	3.730	3.664	3.697
	(3)	(2)	(1)	(3)	(1)	(2)
Industrials	0.884	0.880	0.884	1.563	1.562	1.560
	(2)	(1)	(3)	(3)	(2)	(1)
Insurance	1.215	1.199	1.200	3.190	3.043	3.035
	(3)	(1)	(2)	(3)	(2)	(1)
Media	1.512	1.516	1.535	5.034	5.133	5.193
	(1)	(2)	(3)	(1)	(2)	(3)
Oil & Gas	1.697	1.712	1.712	5.313	5.357	5.357
	(1)	(4)	(3)	(1)	(2)	(3)
Personal	1.003	1.000	1.005	1.828	1.821	1.834
	(2)	(1)	(3)	(2)	(1)	(3)
Retail	1.390	1.389	1.370	4.063	4.109	3.997
	(4)	(3)	(1)	(3)	(4)	(2)
Technology	2.018	2.016	2.001	9.047	9.192	9.171
	(3)	(2)	(1)	(1)	(3)	(2)
Telecom	1.624	1.625	1.623	5.181	5.236	5.194
	(2)	(3)	(1)	(1)	(3)	(2)
Travel	1.046	1.042	1.034	2.145	2.126	2.113
	(4)	(3)	(2)	(4)	(3)	(2)
Utilities	1.040	1.043	1.045	1.832	1.842	1.844
	(1)	(2)	(3)	(1)	(2)	(3)
Average error	1.311	1.310	1.309	3.584	3.600	3.589
Average rank	2.33	2.11	1.94	1.94	2.22	2.17

Table C.7: *Parameter estimates for least squares based multifactor models.*

This table reports the estimated parameters for the alternative least squares based multifactor models. Columns 3–8 report the estimated regression coefficients β_k for $k = 1, \ldots, 6$, with k referring to the respective regressor (1: BMR, 2: VGS, 3: SIZ, 4: TS, 5: OIL, 6: FX). For the RLS specification, the final OLS estimates are reported where *** means significance at the 1% level (**: 5%, *: 10%). The estimation results of the two rolling OLS specifications, RR5 and RR1, are summarized by reporting the respective ranges of estimated coefficients. Figures in parentheses denote the lowest and highest p-values as observed over the entire estimation period. The last column shows adjusted coefficients of determination.

Sector	Model	β_1	β_2	β_3	β_4	β_5	β_6	$\bar{R}^2$
Automobiles	RLS	1.16***	0.12*	0.85***	0.03***	−0.02	−0.64***	0.59
	RR5	1.10;1.35	−0.59;0.35	0.12;1.18	0.00;0.08	−0.07;0.05	−0.95;−0.47	0.57;0.77
		(0.00;0.00)	(0.00;0.74)	(0.00;0.39)	(0.00;0.99)	(0.04;1.00)	(0.00;0.00)	
	RR1	0.63;1.57	−1.28;0.93	0.15;1.42	−0.09;0.26	−0.14;0.22	−1.46;0.21	0.33;0.88
		(0.00;0.00)	(0.00;0.99)	(0.00;0.58)	(0.00;0.99)	(0.02;1.00)	(0.00;0.98)	
Banks	RLS	1.06***	0.22***	1.14***	0.02***	−0.03**	−0.43	0.78
	RR5	0.90;1.22	−0.12;0.31	0.50;1.35	−0.02;0.06	−0.04;0.00	−0.61;−0.34	0.73;0.85
		(0.00;0.00)	(0.00;0.94)	(0.00;0.00)	(0.00;0.79)	(0.04;0.98)	(0.00;0.00)	
	RR1	0.83;1.41	−0.29;1.00	0.46;1.69	−0.09;0.21	−0.14;0.08	−0.99;−0.13	0.67;0.97
		(0.00;0.00)	(0.00;0.96)	(0.00;0.02)	(0.00;0.98)	(0.01;1.00)	(0.00;0.49)	
Basics	RLS	1.02***	0.22***	0.36***	0.03***	−0.03	−0.58	0.51
	RR5	0.96;1.15	−0.26;0.41	−0.13;0.74	−0.01;0.10	−0.08;0.01	−0.81;−0.42	0.45;0.78
		(0.00;0.00)	(0.00;0.71)	(0.00;1.00)	(0.00;0.98)	(0.00;0.98)	(0.00;0.00)	
	RR1	0.59;1.52	−0.56;1.13	−0.36;0.78	−0.10;0.28	−0.22;0.15	−1.29;0.16	0.23;0.89
		(0.00;0.03)	(0.00;0.99)	(0.00;0.99)	(0.00;0.96)	(0.03;1.00)	(0.00;0.94)	
Chemicals	RLS	0.90***	0.19***	0.76***	0.02***	−0.03*	−0.43	0.58
	RR5	0.81;1.00	−0.19;0.31	0.33;1.07	0.00;0.06	−0.06;0.03	−0.86;−0.22	0.48;0.69
		(0.00;0.00)	(0.00;0.98)	(0.00;0.00)	(0.00;1.00)	(0.03;1.00)	(0.00;0.01)	
	RR1	0.33;1.30	−0.81;0.64	0.06;1.68	−0.08;0.15	−0.13;0.17	−1.21;0.27	0.33;0.90
		(0.00;0.17)	(0.00;1.00)	(0.00;0.82)	(0.00;0.99)	(0.00;0.99)	(0.00;0.98)	

Table C.7 — continued

Sector	Model	β_1	β_2	β_3	β_4	β_5	β_6	$\bar{R}^2$
Construction	RLS	0.95***	−0.02	0.38***	0.02***	−0.01	−0.48	0.64
	RR5	0.87;1.15	−0.23;0.02	−0.10;0.63	−0.01;0.07	−0.06;0.03	−0.72;−0.28	0.62;0.77
		(0.00;0.00)	(0.01;1.00)	(0.00;1.00)	(0.00;0.94)	(0.04;1.00)	(0.00;0.00)	
	RR1	0.74;1.14	−0.77;0.81	−0.13;1.04	−0.04;0.19	−0.08;0.11	−1.00;0.11	0.54;0.87
		(0.00;0.00)	(0.00;1.00)	(0.00;1.00)	(0.00;0.98)	(0.01;1.00)	(0.00;0.98)	
Financials	RLS	1.04***	0.10**	0.91***	0.02***	−0.01	−0.41	0.72
	RR5	0.84;1.16	−0.04;0.16	0.11;1.24	−0.02;0.05	−0.04;0.01	−0.64;−0.33	0.66;0.82
		(0.00;0.00)	(0.00;1.00)	(0.00;0.20)	(0.00;0.98)	(0.06;1.00)	(0.00;0.00)	
	RR1	0.63;1.29	−0.40;0.60	−0.32;1.56	−0.09;0.19	−0.10;0.22	−1.17;0.03	0.56;0.94
		(0.00;0.00)	(0.00;0.99)	(0.00;0.99)	(0.00;1.00)	(0.00;1.00)	(0.00;0.98)	
Food	RLS	0.57***	0.32***	0.70***	0.01**	−0.05***	−0.36	0.49
	RR5	0.37;0.83	−0.14;0.51	0.41;0.85	−0.01;0.05	−0.06;0.00	−0.65;−0.19	0.35;0.79
		(0.00;0.00)	(0.00;1.00)	(0.00;0.00)	(0.00;0.96)	(0.01;1.00)	(0.00;0.03)	
	RR1	0.06;0.91	−0.55;0.88	−0.11;1.18	−0.06;0.12	−0.15;0.13	−0.93;0.09	0.04;0.87
		(0.00;0.78)	(0.00;1.00)	(0.00;0.87)	(0.00;1.00)	(0.00;1.00)	(0.00;0.98)	
Healthcare	RLS	0.50***	0.01	1.10***	0.01	−0.07***	−0.44	0.54
	RR5	0.39;0.66	−0.54;0.25	0.65;1.30	−0.01;0.01	−0.09;0.00	−0.63;−0.38	0.44;0.66
		(0.00;0.00)	(0.00;0.96)	(0.00;0.00)	(0.22;1.00)	(0.00;0.96)	(0.00;0.00)	
	RR1	0.08;0.99	−1.28;0.50	0.19;1.82	−0.11;0.08	−0.20;0.09	−1.00;−0.02	0.15;0.90
		(0.00;0.69)	(0.00;1.00)	(0.00;0.55)	(0.00;1.00)	(0.00;1.00)	(0.00;0.89)	
Industrials	RLS	1.03***	−0.26***	0.46***	0.02***	−0.01	−0.57	0.73
	RR5	0.90;1.13	−0.38;0.00	0.07;0.63	−0.01;0.06	−0.05;0.03	−0.77;−0.38	0.73;0.86
		(0.00;0.00)	(0.00;0.96)	(0.00;0.26)	(0.00;0.77)	(0.00;1.00)	(0.00;0.00)	
	RR1	0.70;1.24	−1.23;0.37	−0.19;0.83	−0.05;0.20	−0.08;0.13	−1.40;−0.21	0.54;0.93
		(0.00;0.00)	(0.00;1.00)	(0.00;0.99)	(0.00;0.99)	(0.01;1.00)	(0.00;0.05)	

Table C.7 — continued

Sector	Model	β_1	β_2	β_3	β_4	β_5	β_6	$\bar{R}^2$
Insurance	RLS	1.18***	0.11*	1.39***	0.03***	−0.04**	−0.42	0.66
	RR5	0.95;1.39	−0.22;0.27	0.37;1.91	−0.02;0.10	−0.06;0.01	−0.63;−0.29	0.68;0.80
		(0.00;0.00)	(0.01;1.00)	(0.00;0.00)	(0.00;1.00)	(0.08;0.99)	(0.00;0.01)	
	RR1	0.52;2.09	−0.70;0.51	0.23;2.61	−0.16;0.39	−0.22;0.22	−1.07;0.08	0.36;0.93
		(0.00;0.00)	(0.00;1.00)	(0.00;0.27)	(0.00;1.00)	(0.01;1.00)	(0.00;0.98)	
Media	RLS	1.07***	−1.06***	0.64***	0.03***	0.01	−0.70	0.64
	RR5	0.65;1.37	−1.47;0.11	0.23;0.80	−0.01;0.10	−0.09;0.10	−0.97;−0.36	0.58;0.79
		(0.00;0.00)	(0.00;1.00)	(0.00;0.02)	(0.00;0.95)	(0.00;1.00)	(0.00;0.00)	
	RR1	0.33;1.67	−1.85;0.38	−0.32;1.34	−0.04;0.27	−0.17;0.14	−1.30;−0.23	0.38;0.93
		(0.00;0.22)	(0.00;0.99)	(0.00;0.99)	(0.00;1.00)	(0.01;1.00)	(0.00;0.34)	
Oil & Gas	RLS	0.64***	0.48***	0.92***	0.01**	0.12***	−0.64	0.54
	RR5	0.54;0.87	−0.01;0.65	0.45;1.35	0.00;0.03	0.05;0.20	−0.97;−0.55	0.50;0.65
		(0.00;0.00)	(0.00;1.00)	(0.00;0.00)	(0.06;1.00)	(0.00;0.17)	(0.00;0.00)	
	RR1	0.41;1.23	−0.13;1.57	0.18;1.86	−0.12;0.11	−0.02;0.32	−1.67;0.13	0.43;0.90
		(0.00;0.06)	(0.00;0.99)	(0.00;0.35)	(0.00;1.00)	(0.00;0.96)	(0.00;1.00)	
Personal	RLS	0.89***	−0.14***	0.64***	0.01**	−0.02	−0.57	0.66
	RR5	0.80;1.03	−0.26;−0.01	0.37;0.85	−0.02;0.05	−0.04;0.01	−0.71;−0.47	0.58;0.73
		(0.00;0.00)	(0.00;0.88)	(0.00;0.00)	(0.00;0.99)	(0.04;0.98)	(0.00;0.00)	
	RR1	0.59;1.37	−0.98;0.41	0.05;1.07	−0.07;0.14	−0.10;0.12	−1.03;−0.11	0.28;0.94
		(0.00;0.00)	(0.00;1.00)	(0.00;0.78)	(0.00;1.00)	(0.00;1.00)	(0.00;0.30)	
Retail	RLS	0.75***	0.09**	0.74***	0.01***	−0.02	−0.39	0.59
	RR5	0.60;0.94	−0.18;0.17	0.46;0.88	−0.01;0.05	−0.06;0.02	−0.60;−0.25	0.47;0.68
		(0.00;0.00)	(0.01;1.00)	(0.00;0.00)	(0.00;1.00)	(0.04;1.00)	(0.00;0.00)	
	RR1	0.45;0.97	−0.57;0.75	−0.22;1.19	−0.03;0.12	−0.15;0.12	−0.94;0.10	0.21;0.90
		(0.00;0.01)	(0.00;1.00)	(0.00;0.98)	(0.00;1.00)	(0.00;1.00)	(0.00;0.99)	
				Continued on the next page...				

Table C.7 — continued

Sector	Model	β_1	β_2	β_3	β_4	β_5	β_6	$\bar{R}^2$
Technology	RLS	1.33***	−1.52***	0.85***	0.03***	0.04*	−0.73	0.66
	RR5	1.02;1.54	−2.04;−0.01	0.32;1.35	−0.02;0.12	−0.05;0.10	−1.08;−0.38	0.65;0.82
		(0.00;0.00)	(0.00;0.91)	(0.00;0.00)	(0.00;0.98)	(0.01;0.99)	(0.00;0.00)	
	RR1	0.83;2.08	−3.91;0.42	−0.29;1.61	−0.05;0.49	−0.16;0.30	−1.90;−0.10	0.54;0.92
		(0.00;0.00)	(0.00;0.98)	(0.00;0.94)	(0.00;1.00)	(0.00;1.00)	(0.00;0.56)	
Telecom	RLS	0.99***	−1.09***	1.10***	0.03***	−0.01	−0.49	0.70
	RR5	0.89;1.29	−1.48;−0.12	0.56;1.31	0.00;0.06	−0.06;0.01	−0.65;−0.40	0.62;0.82
		(0.00;0.00)	(0.00;0.38)	(0.00;0.00)	(0.00;0.99)	(0.01;1.00)	(0.00;0.00)	
	RR1	0.65;1.68	−1.95;0.42	0.22;1.90	−0.03;0.20	−0.11;0.10	−0.92;−0.06	0.53;0.90
		(0.00;0.00)	(0.00;0.98)	(0.00;0.22)	(0.00;0.99)	(0.04;1.00)	(0.00;0.69)	
Travel	RLS	1.04***	0.18***	0.63***	0.02***	−0.03*	−0.58	0.58
	RR5	0.89;1.11	−0.22;0.32	0.13;0.93	−0.01;0.06	−0.07;0.02	−0.85;−0.35	0.47;0.65
		(0.00;0.00)	(0.00;1.00)	(0.00;0.35)	(0.00;0.89)	(0.06;1.00)	(0.00;0.00)	
	RR1	0.59;1.38	−0.39;1.11	−0.56;1.20	−0.09;0.21	−0.12;0.12	−1.48;−0.08	0.29;0.88
		(0.00;0.02)	(0.00;1.00)	(0.00;0.95)	(0.00;1.00)	(0.03;1.00)	(0.00;0.68)	
Utilities	RLS	0.64***	−0.10**	0.68***	0.01	−0.02*	−0.30	0.56
	RR5	0.50;0.89	−0.34;0.06	0.42;0.75	−0.01;0.02	−0.08;0.01	−0.46;−0.24	0.50;0.65
		(0.00;0.00)	(0.00;1.00)	(0.00;0.00)	(0.00;1.00)	(0.01;0.99)	(0.00;0.00)	
	RR1	0.25;1.12	−0.75;0.75	−0.28;1.20	−0.09;0.08	−0.13;0.08	−0.75;0.22	0.21;0.87
		(0.00;0.04)	(0.00;0.98)	(0.00;1.00)	(0.00;0.99)	(0.01;0.99)	(0.00;0.99)	

Table C.8: *Out-of-sample errors for multiple factor models.*

This table reports the estimated out-of-sample mean errors for the four multiple factor specifications. For each sector i, figures in parentheses denote the relative rank of a model's mean absolute and squared error, respectively, where the model with the smallest error ranks first.

Sector	MAE ($\times 10^2$)				MSE ($\times 10^4$)			
	KF	RLS	$RR5$	$RR1$	KF	RLS	$RR5$	$RR1$
Automobiles	1.617	1.720	1.703	1.670	4.706	5.655	5.458	5.109
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Banks	0.868	1.007	0.982	1.002	1.560	2.248	2.069	1.988
	(1)	(4)	(2)	(3)	(1)	(4)	(3)	(2)
Basics	1.610	1.713	1.671	1.663	5.002	5.553	5.313	5.227
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Chemicals	1.336	1.448	1.439	1.432	3.084	3.743	3.707	3.532
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Construction	1.039	1.133	1.120	1.115	1.990	2.531	2.351	2.201
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Financials	1.001	1.146	1.115	1.110	1.892	2.639	2.385	2.303
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Food	1.162	1.249	1.247	1.206	2.731	3.089	3.053	2.872
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Healthcare	1.283	1.357	1.368	1.348	3.181	3.547	3.538	3.488
	(1)	(3)	(4)	(2)	(1)	(4)	(3)	(2)
Industrials	0.907	1.054	1.011	1.002	1.617	2.277	2.037	1.941
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Insurance	1.280	1.509	1.466	1.392	3.279	5.460	4.878	4.183
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Media	1.470	1.721	1.643	1.502	3.989	6.319	5.447	4.197
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Oil & Gas	1.487	1.567	1.575	1.510	3.949	4.490	4.546	4.160
	(1)	(3)	(4)	(2)	(1)	(3)	(4)	(2)
Personal	1.085	1.168	1.156	1.159	2.164	2.541	2.455	2.347
	(1)	(4)	(2)	(3)	(1)	(4)	(3)	(2)
Retail	1.132	1.140	1.148	1.177	2.383	2.552	2.532	2.519
	(1)	(2)	(3)	(4)	(1)	(4)	(3)	(2)
Technology	1.642	2.096	1.939	1.862	5.588	9.273	7.738	6.797
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Telecom	1.424	1.551	1.469	1.445	3.789	4.589	4.059	3.949
	(1)	(4)	(3)	(2)	(1)	(4)	(3)	(2)
Travel	1.443	1.495	1.476	1.499	3.798	4.238	4.071	4.172
	(1)	(3)	(2)	(4)	(1)	(4)	(2)	(3)
Utilities	1.065	1.151	1.155	1.110	1.864	2.171	2.175	2.104
	(1)	(3)	(4)	(2)	(1)	(3)	(4)	(2)
Average error	1.270	1.401	1.371	1.345	3.143	4.051	3.767	3.505
Average Rank	1.00	3.67	3.00	2.33	1.00	3.89	3.06	2.06

Table C.9: *Fama-MacBeth regression results II.*

This table presents the estimation results for the Fama-MacBeth regressions with the regressors being the out-of-sample loadings on the BMR, VGS, SIZ and TS factors only. The factor loadings have been estimated from one of the time series regressions of excess sector returns on the set of explanatory variables as presented in §7.4.1. Average coefficients are expressed as percent per week $\times 10$. The number of cross-sectional regressions is 520 for each model. t-statistics are reported in parentheses with the relevant critical value at the 95% (90%, 99%) level being -1.96 (-1.65, -2.59). Significant average coefficients are printed in bold numbers. The last column reports average adjusted coefficients of determination.

Model	BMR	VGS	SIZ	TS	$\bar{R}^2$
A. 22 February 1995–2 February 2005					
KF	1.128	0.532	0.346	**−13.225**	0.229
	(1.073)	(0.650)	(0.463)	(−1.895)	
RLS	0.841	0.546	1.145	11.757	0.216
	(0.681)	(0.381)	(1.130)	(0.315)	
RR5	0.717	0.680	1.329	14.079	0.213
	(0.585)	(0.555)	(1.400)	(0.688)	
RR1	0.312	0.647	0.316	−7.041	0.218
	(0.291)	(0.780)	(0.416)	(−0.827)	
B. 22 February 1995–8 March 2000					
KF	2.155	−1.375	**1.856**	**−23.141**	0.140
	(1.637)	(−1.132)	(1.690)	(−2.262)	
RLS	1.518	−3.263	**3.549**	26.942	0.116
	(1.123)	(−1.450)	(2.605)	(0.587)	
RR5	1.071	−2.175	**3.853**	33.279	0.130
	(0.788)	(−1.104)	(2.905)	(0.942)	
RR1	1.408	−1.498	**2.031**	−18.327	0.127
	(1.043)	(−1.264)	(1.820)	(−1.461)	
C. 15 March 2000–12 March 2003					
KF	−2.727	**3.362**	**−2.466**	−5.080	0.371
	(−1.159)	(1.973)	(−1.708)	(−0.353)	
RLS	−3.974	**6.663**	−2.134	26.108	0.377
	(−1.359)	(2.443)	(−0.956)	(0.296)	
RR5	−3.363	**5.329**	−2.130	0.556	0.362
	(−1.113)	(2.370)	(−1.058)	(0.019)	
RR1	−3.081	**3.908**	**−2.729**	−2.148	0.362
	(−1.295)	(2.155)	(−1.855)	(−0.128)	
D. 19 March 2003–2 February 2005					
KF	**4.505**	1.128	0.780	0.301	0.241
	(2.243)	(1.564)	(0.652)	(0.035)	
RLS	**6.670**	1.001	−0.068	−51.495	0.228
	(2.523)	(1.033)	(−0.044)	(−0.815)	
RR5	**6.246**	0.918	0.087	−15.673	0.199
	(2.778)	(1.211)	(0.064)	(−0.669)	
RR1	2.774	**1.194**	0.569	15.294	0.231
	(1.327)	(1.784)	(0.472)	(1.180)	

References

Abell, J. D. and T. M. Krueger (1989). Macroeconomic influences on beta. *Journal of Economics and Business 41*(2), 185–193.

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics 18*(3), 338–357.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov and F. Csake (Eds.), *Proceedings of the Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, Hungary.

Alexander, C. (2000). Orthogonal methods for generating large positive semi-definite covariance matrices. ISMA Centre Discussion Papers in Finance 2000-06, ISMA Centre, University of Reading, UK. Available at `http://www.ismacentre.rdg.ac.uk` (9.5.2006).

Alexander, C. (2001). *Market Models. A Guide to Financial Data Analysis.* West Sussex, UK: John Wiley & Sons.

Andersen, T. and B. Sorensen (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics 14*(3), 392–452.

Andersen, T. G. and T. Bollerslev (1998). ARCH and GARCH models. In S. Kotz, C. B. Read, and D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences*, Volume II, Chapter 49, pp. 6–16. New York, NY, USA: John Wiley & Sons.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2005). Volatility forecasting. Working Paper 11188, National Bureau of Economic Research, Cambridge, MA, USA. Available at `http://www.nber.org/papers/w11188` (6.4.2005).

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2002). Parametric and nonparametric volatility measurement. NBER Technical Working Paper 0279, National Bureau of Economic Research, Inc. Available at `http://ideas.repec.org/p/nbr/nberte/0279.html` (6.2.2006).

Andersen, T. G., T. Bollerslev, F. X. Diebold, and J. Wu (2005). A framework for exploring the macroeconomic determinants of systematic risk. *American Economic Review Papers and Proceedings*. Available at `http://www.aeaweb.org/annual_mtg_papers/2005/0107_1015_0101.pdf` (17.1.2005).

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering.* Englewood Cliffs, NJ, USA: Prentice Hall.

Ansley, C. F. and R. Kohn (1985). Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics 13*(4), 1286–1316.

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: A review. *Econometric Reviews 25*(2–3), 145–175.

Baillie, R. T. and T. Bollerslev (1990). A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets. *Journal of International Money and Finance 9*(3), 309–324.

Banz, R. W. (1981). The relationship between return and the market value of common stocks. *Journal of Financial Economics 9*(1), 3–18.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance 32*(3), 663–682.

Baum, L. E. and T. Petrie (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics 37*, 1554–1563.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics 41*, 164–171.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2003). Multivariate GARCH models: A survey. Discussion Paper 2003/31, CORE (Center for operations research and econometrics), Louvain-la-Neuve, France. Available at `http://ssrn.com/abstract=411062` (29.12.2004).

Bera, A. K. and M. L. Higgins (1993). ARCH models: Properties, estimation and testing. *Journal of Economic Surveys 7*(4), 305–366.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974). Estimation inference in nonlinear structured models. *Annals of Economic and Social Measurement 3*(4), 653–665.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance 43*(2), 507–528.

Bhar, R. and S. Hamori (2004). *Hidden Markov Models: Applications to Financial Economics.* Advanced Studies in Theoretical and Applied Econometrics. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the American Statistical Association Annual Meeting, Business and Economic Statistic Section*, Washington, DC, USA, pp. 177–181.

Blanchard, O. J. and S. Fischer (1989). *Lectures on Macroeconomics.* Cambridge, MA, USA: The MIT Press.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics 69*(3), 542–547.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics 72*(3), 498–505.

Bollerslev, T., Y. Chou, R. and F. Kroner, K. (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics 52*(1–2), 5–59.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 49, pp. 2959–3038. Amsterdam, The Netherlands: Elsevier Science B.V.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*(1), 116–131.

Bos, T. and P. Newbold (1984). An empirical investigation of the possibility of stochastic systematic risk in the market model. *Journal of Business 57*(1), 35–41.

Braun, P. A., D. B. Nelson, and A. M. Sunier (1995). Good news, bad news, volatility and betas. *Journal of Finance 50*(5), 1575–1603.

Brennan, N., S. Hartmann, S. Fagg, M. Dowle, A. Luck, G. Brar, and M. Turner (2000). A new European risk attribute model (ERAM). Technical report, Schroder Salomon Smith Barney, Equity Research Europe, Quantitative Strategy.

Brooks, R. D., R. W. Faff, and M. D. McKenzie (1998). Time–varying beta risk of Australian industry portfolios: A comparison of modelling techniques. *Australian Journal of Management 23*(1), 1–22.

Broto, C. and E. Ruiz (2004). Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys 18*(5), 613–649.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B 37*(2), 149–192.

Bulla, J. (2006). *Application of Hidden Markov Models and Hidden Semi-Markov Models to Financial Time Series*. Ph. D. thesis, Institute for Statistics and Econometrics, Georg-August-University, Göttingen, Germany.

Bulla, J. and A. Berzel (2006). Computational issues in parameter estimation for stationary hidden Markov models. *Computational Statistics (to appear)*.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics 18*(2), 373–399.

Campbell, J. Y. and L. Hentschel (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics 31*(3), 281–318.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton, NJ, USA: Princeton University Press.

Campbell, K., J. Enos, D. Gerlanc, and D. Kane (2006a). *Backtests: using the `backtest` package*. Available at `http://www.r-project.org` (8.2.2007).

Campbell, K., J. Enos, D. Gerlanc, and D. Kane (2006b). *The `backtest` package, version 0.1-1*. Available at `http://www.r-project.org` (8.2.2007).

Campbell, K., J. Enos, and D. Kane (2007). *Performing equity investment simulations with the `portfolioSim` package*. Available at `http://www.r-project.org` (8.2.2007).

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Chan, D. X., R. Kohn, and C. Kirby (2005). Multivariate stochastic volatility models with correlated errors. SSRN Working Paper Series, Social Science Research Network. Available at `http://ssrn.com/abstract=876348` (9.5.2006).

Chan, L. K., Y. Hamao, and J. Lakonishok (1991). Fundamentals and stock returns in japan. *Journal of Finance 46*(5), 1739–1764.

Chan, L. K., N. Jegadeesh, and J. Lakonishok (1996). Momentum strategies. *Journal of Finance 51*(5), 1681–1713.

Chan, L. K., J. Karceski, and J. Lakonishok (1998). The risk and return from factors. *Journal of Financial and Quantitative Analysis 33*(3), 159–188.

Chen, N.-F., R. Roll, and S. A. Ross (1986). Economic forces and the stock market. *Journal of Business 59*(3), 383–403.

Chib, S. and E. Greenberg (1996). Markov Chain Monte Carlo simulation methods in econometrics. *Econometric Theory 12*(3), 409–431.

Chib, S., F. Nardari, and N. Shephard (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics 134*(2), 341–371.

Chow, G. C. (1984). Random and changing coefficient models. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, pp. 1213–1245. Amsterdam, The Netherlands: Elsevier Science Publishers B.V.

Clark, P. K. (1987). The cyclical component of U.S. economic activity. *Quarterly Journal of Economics 102*(4), 797–814.

Cochrane, J. H. (2005). *Asset Pricing* (Revised ed.). Princeton, NJ, USA: Princeton University Press.

Collins, D. W., J. Ledolter, and J. Rayburn (1987). Some further evidence on the stochastic properties of systematic risk. *Journal of Business 60*(3), 425–448.

Daniel, K. and S. Titman (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance 52*(1), 1–33.

Danielsson, J. (1994). Stochastic volatility in asset prices, estimation with simulated maximum likelihood. *Journal of Econometrics 64*(1–2), 375–400.

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford, UK: Oxford University Press.

de Jong, P. (1988a). A cross–validation filter for time series models. *Biometrika 75*(3), 594–600.

de Jong, P. (1988b). The likelihood for a state space model. *Biometrika 75*(1), 165–169.

de Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics 19*(2), 1073–1083.

DeBondt, W. F. M. and R. H. Thaler (1985). Does the stock market overreact? *Journal of Finance 40*(3), 793–805.

DeBondt, W. F. M. and R. H. Thaler (1987). Further evidence on investor overreaction and stock market seasonality. *Journal of Finance 42*(3), 557–581.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B 39*(1), 1–38.

Diebold, F. X. (2004). The Nobel Memorial Prize for Robert F. Engle. *Scandinavian Journal of Economics 106*(2), 165–185.

Diebold, F. X., J.-H. Lee, and G. C. Weinbach (1994). Regime switching with time varying transition probabilities. In C. Hargreaves (Ed.), *Non-stationary Time Series Analyses and Cointegration*, pp. 283–302. Oxford, UK: Oxford University Press.

Diebold, F. X. and M. Nerlove (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics 4*(1), 1–21.

Ding, Z. and R. F. Engle (2001). Large scale conditional covariance matrix modeling, estimation and testing. *Academia Economic Papers 29*(2), 157–184.

Doornik, J. A. (2001). *Ox: An Object-Oriented Matrix Programming Language* (4 ed.). London, UK: Timberlake Consultants Press.

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica 61*(4), 929–952.

Dumas, B. and B. Solnik (1995). The world price of foreign exchange risk. *Journal of Finance 50*(2), 445–479.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non–Gaussian state space models. *Biometrika 84*(3), 669–684.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society. Series B 62*, 3–56.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series. Oxford, UK: Oxford University Press.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–615.

Elliott, R. J., L. Aggoun, and J. B. Moore (1995). *Hidden Markov models, Estimation and Control*, Volume 29 of *Applications of Mathematics*. New York, NY, USA: Springer.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F. (Ed.) (1995). *ARCH: Selected Readings*. Oxford, UK: Oxford University Press.

Engle, R. F. (2000). Dynamic conditional correlation: A simple class of multivariate GARCH models. Available at `http://weber.ucsd.edu/~mbacci/ engle/391.pdf` (17.1.2005).

Engle, R. F. (2001a). Financial econometrics—a new discipline with new methods. *Journal of Econometrics 100*(1), 53–56.

Engle, R. F. (2001b). GARCH 101: The use of ARCH/ GARCH models in applied econometrics. *Journal of Economic Perspectives 15*(4), 157–168.

Engle, R. F. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews 5*(1), 1–50.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*(2), 277–304.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*(1), 122–150.

Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance 48*(5), 1749–1778.

Ephraim, Y. and N. Merhav (2002). Hidden Markov processes. *IEEE Transactions on Information Theory 48*(6), 1518–1569.

Fabozzi, F. J. and J. C. Francis (1978). Beta as a random coefficient. *Journal of Financial and Quantitative Analysis 13*(1), 101–116.

Faff, R. W., D. Hillier, and J. Hillier (2000). Time varying beta risk: An analysis of alternative modelling techniques. *Journal of Business Finance and Accounting 27*(5), 523–554.

Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business 38*(1), 34–105.

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*(1), 23–49.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance 50*(1), 131–155.

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*(1), 55–84.

Fama, E. F. and K. R. French (1998). Value versus growth: The international evidence. *Journal of Finance 53*(6), 1975–1999.

Fama, E. F. and M. R. Gibbons (1982). Inflation, real returns, and capital investment. *Journal of Monetary Economics 9*(3), 297–323.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Ferguson, J. D. (1980). Variable duration models for speech. *Proceedings of the Symposium on the Applications of Hidden Markov Models to Text and Speech*, 143–179. Princeton, New Jersey, USA.

Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *Journal of Political Economy 99*(2), 385–415.

Ferson, W. E. and C. R. Harvey (1993). The risk and predictability of international equity returns. *Review of Financial Studies 6*(3), 527–566.

Ferson, W. E. and C. R. Harvey (1999). Conditioning variables and the cross section of stock returns. *Journal of Finance 54*(4), 1325–1360.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics 12*(3), 299–308.

Franses, P. H. and D. van Dijk (2000). *Non–linear Time Series Models in Empirical Finance*. Cambridge, UK: Cambridge University Press.

Fridman, M. (1994). A two state capital asset pricing model. IMA Preprint Series, Institute of Mathematics and its Applications, University of Minnesota, Minneapolis, MN, USA. Available at `http://www.ima.umn.edu/ preprints/March1994/1221.pdf` (14.3.2005).

Gallant, R. A., D. Hsieh, and G. Tauchen (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics 81*(1), 159–192.

Gallant, R. A. and G. Tauchen (1996). Which moments to match. *Econometric Theory 12*(4), 657–681.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*(6), 721–741.

Gettinby, G. D., C. D. Sinclair, D. M. Power, and R. A. Brown (2004). An analysis of the distribution of extreme share returns in the UK from 1975 to 2000. *Journal of Business Finance and Accounting 31*(5–6), 607–646.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics*, Volume 14, pp. 128–198. Amsterdam, The Netherlands: North-Holland.

Giannopoulos, K. (1995). Estimating the time varying components of international stock markets' risk. *European Journal of Finance 1*, 129–164.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess returns on stocks. *Journal of Finance 48*(5), 1779–1801.

González-Rivera, G. (1997). The pricing of time-varying beta. *Empirical Economics 22*(3), 345–363.

Gouriéroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8 (Supplement: Special issue on econometric inference using simulation techniques)*, S85–S118.

Granger, C. W. J., S. A. Spear, and Z. Ding (2000). Stylized facts on the temporal and distributional properties of absolute returns: An update. In W. S. Chan, W. K. Li, and H. Tong (Eds.), *Statistics and Finance: An Interface*, pp. 97–120. London, UK: Imperial College Press.

Greene, W. H. (2003). *Econometric Analysis* (5 ed.). Upper Saddle River, NJ, USA: Prentice Hall.

Groenewold, N. and P. Fraser (1999). Time-varying estimates of CAPM betas. *Mathematics and Computers in Simulations 48*, 531–539.

Grundy, B. and J. S. Martin (2001). Understanding the nature of the risks and the source of the rewards to momentum investing. *Review of Financial Studies 14*(1), 29–78.

Hagerud, G. E. (1997). *A new non-linear GARCH model*. Ph.D. thesis, IFE, School of Economics, Stockholm, Sweden.

Hamao, Y. (1988). An empirical examination of arbitrage pricing theory: Using Japanese data. *Japan and the World Economy 1*(1), 45–61.

Hamilton, J. D. (1988). Rational expectations econometric analysis of changes in regime. an investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control 12*(2–3), 385–423.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics 45*(1–2), 39–70.

Hamilton, J. D. (1993). Estimation, inference and forecasting of time series subject to changes in regime. In G. S. Maddala, C. R. Rao, and H. D. Vinod (Eds.), *Handbook of Statistics*, Volume 11, pp. 231–260. Amsterdam, The Netherlands: Elsevier Science Publishers B. V.

Hamilton, J. D. (1994a). State-space models. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 50, pp. 3039–3080. Amsterdam, The Netherlands: Elsevier Science B. V.

Hamilton, J. D. (1994b). *Time Series Analysis*. Princeton, NJ, USA: Princeton University Press.

Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics 64*(1–2), 307–333.

Hansen, L. P. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy 99*(2), 225–262.

Harvey, A. C. (1981). *Time Series Models*. Oxford, UK: Philip Allan and Humanities Press.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

Harvey, A. C. and G. D. A. Phillips (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika 66*(1), 49–58.

Harvey, A. C., E. Ruiz, and E. Sentana (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics 52*(1-2), 129–157.

Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*(2).

Harvey, A. C. and N. Shephard (1993). Structural time series models. In G. S. Maddala, C. R. Rao, and H. D. Vinod (Eds.), *Handbook of Statistics*, Volume 11, pp. 261–302. Amsterdam, The Netherlands: Elsevier Science Publishers B. V.

Harvey, A. C. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics 14*(4), 429–434.

Hentschel, L. (1995). All in the family: Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics 39*(1), 71–104.

Hildreth, C. and J. Houck (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association 63*(322), 584–595.

Huang, H.-C. (2000). Tests of regimes-switching CAPM. *Applied Financial Economics 10*, 573–578.

Isakov, D. and F. Sonney (2004). Are practitioners right? On the relative importance of industrial factors in international stock returns. *Swiss Journal of Economics and Statistics 140*(3), 355–379.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics 12*(4), 371–389.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*(1), 185–212.

Jagannathan, R. and Z. Wang (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance 51*(1), 3–53.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jegadeesh, N. and S. Titman (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance 56*(2), 699–720.

Johnson, L. D. and G. Sakoulis (2003). Maximizing equity market sector predictability in a Bayesian time varying parameter model. Draft version. Available at `http://ssrn.com/abstract=396642` (10.11.2004).

Jones, C. M. and G. Kaul (1996). Oil and the stock markets. *Journal of Finance 51*(2), 463–491.

Jorion, P. (1988). On jump processes in the foreign exchange and stock markets. *Review of Financial Studies 1*(4), 427–445.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T.-C. Lee (1985). *The Theory and Practice of Econometrics* (2 ed.). New York, NY, USA: John Wiley & Sons.

Jungbacker, B. and S. J. Koopman (2005a). Monte carlo likelihood estimation for three multivariate stochastic volatility models. Working Paper, Department of Econometrics, Free University Amsterdam, Amsterdam, The Netherlands. Available at `http://staff.feweb.vu.nl/koopman/publications/msv.pdf` (10.4.2006).

Jungbacker, B. and S. J. Koopman (2005b). On importance sampling for state space models. Tinbergen Institute Discussion Papers 05-117/4, Tinbergen Institute. Available at `http://ideas.repec.org/p/dgr/uvatin/20050117.html` (7.2.2006).

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering—Transactions of the ASME. Series D 82*, 35–45.

Kalman, R. E. (1963). New methods in Wiener filtering theory. In J. L. Bogdanoff and F. kozin (Eds.), *Proceedings of the First Symposium on Engineering Application of Random Function Theory and Probability*. John Wiley & Sons.

Kaufman, L. and P. J. Rousseeuw (2005). *Finding Groups in Data: An Introduction to Cluster Analysis* (2 ed.). Hoboken, NJ, USA: Wiley-Interscience.

Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics 17*(2), 357–390.

Kim, C.-J. (1993). Unobserved-component time series models with Markov-switching heteroskedasticity: Changes in regime and the link between inflation rates and inflation uncertainty. *Journal of Business and Economic Statistics 11*(3), 341–349.

Kim, C.-J. and C. R. Nelson (1989). The time-varying parameter model for modeling changing conditional variance: The case of the lucas hypothesis. *Journal of Business and Economic Statistics 7*(4), 433–440.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching*. Cambridge, MA, USA: The MIT Press.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*(3), 361–393.

Kohn, R. and C. F. Ansley (1989). A fast algorithm for signal extraction, influence and cross-validation in state space models. *Biometrika 76*(1), 65–79.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika 80*(1), 117–126.

Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association 92*(440), 1630–1638.

Koopman, S. J. and E. Hol-Uspensky (2001). The stochastic volatility in mean model: Empirical evidence from international stock markets. Working Paper, Department of Econometrics, Free University Amsterdam, Amsterdam, The Netherlands. Available at `http://staff.feweb.vu.nl/koopman/papers/svm300701.pdf` (18.3.2005).

Koopman, S. J., B. Jungbacker, and E. Hol (2004). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. Project, Department of Econometrics, Free University Amsterdam, Amsterdam, The Netherlands. Available at `http://staff.feweb.vu.nl/koopman/projects/2004forc.pdf` (21.3.2005).

Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal 2*(1), 113–166. Available at `http://www.ssfpack.com` (19.3.2005).

Lahiri, K. and J. G. Wang (1996). Interest rate spreads as predictors of business cycles. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics*, Volume 14, Chapter 10, pp. 297–315. Amsterdam, The Netherlands: Elsevier Science B. V.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance 49*(5), 1541–1578.

Laurent, S. and J.-P. Peters (2002). G@RCH 2.2, an Ox package for estimating and forecasting various ARCH models. *Journal of Economic Surveys 16*, 447–484.

Ledoit, O., P. Santa-Clara, and M. Wolf (2003). Flexible multivariate GARCH modeling with an application to international stock markets. *The Review of Economics and Statistics 85*(1), 735–747.

Lee, K. M. and S. J. Koopman (2004). Estimating stochastic volatility models: A comparison of two importance samplers. *Studies in Nonlinear Dynamics and Econometrics 8*(2), Article 5.

Lettau, M. and S. Ludvigson (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy 109*(6), 1238–1287.

Leusner, J., J. D. Akhavein, and J. Swamy (1996). Solving an empirical puzzle in the capital asset pricing model. Finance and Economics Discussion Series 1996-14, The Federal Reserve Board, Washington, DC, USA. Available at `http://www.federalreserve.gov/pubs/feds/1996/199614/199614pap.pdf` (18.6.2006).

Li, X. (2003). On unstable beta risk and its modelling techniques for New Zealand industry portfolios. Working Paper 03.01, Massey University Commerce, Auckland, New Zealand. Available at `http://ssrn.com/abstract=503722` (29.12.2004).

Lie, F., R. Brooks, and R. Faff (2000). Modelling the equity beta risk of Australian financial sector companies. *Australian Economic Papers 39*, 301–311.

Liesenfeld, R. and J.-F. Richard (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance 10*(4), 505–531.

Ling, S. and M. McAleer (2002). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics 106*(1), 109–117.

Lintner, J. (1965). The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*(1), 13–37.

Liodakis, M., G. Brar, R. Crenian, and M. Dowle (2003). Estimating betas using fundamentals. Technical report, Citigroup Smith Barney, Equity Research Europe, Quantitative Strategy.

Lo, A. W. and A. C. MacKinlay (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies 1*(1), 41–66.

Lo, A. W. and A. C. MacKinlay (1997). Maximizing predictability in the stock and bond markets. *Macroeconomic Dynamics 1*, 102–134.

Lo, A. W. and A. C. MacKinlay (1999). *A Non-Random Walk Down Wall Street*. Princeton, NJ, USA: Princeton University Press.

MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*, Volume 70 of *Monographs on Statistics and Applied Probability*. London, UK: Chapman & Hall.

MacKinlay, A. C. (1995). Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics 38*(1), 3–28.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business 36*(4), 394–419.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York, NY, USA: John Wiley & Sons.

McLachlan, G. J. and T. Krishnan (1997). *The EM Algorithm and Extensions*. New York, NY, USA: John Wiley & Sons.

Meinhold, R. J. and N. D. Singpurwella (1983). Understanding the Kalman filter. *American Statistician 37*(2), 123–127.

Melino, A. and S. M. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics 45*(1–2), 239–265.

Mergner, S. and J. Bulla (2008). Time-varying beta risk of pan-European industry portfolios: A comparison of alternative modeling techniques. *European Journal of Finance* (forthcoming).

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica 41*(5), 867–887.

Meyer, R. and J. Yu (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal 3*, 198–215.

Moonis, S. A. and A. Shah (2002). Testing for time variation in beta in India. Available at `http://www.mayin.org/ajayshah/pdfdocs` (3.1.2005).

Moskowitz, T. J. and M. Grinblatt (1999). Do industries explain momentum? *Journal of Finance 54*(4), 1249–1290.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

Nicholls, D. F. and A. R. Pagan (1985). Varying coefficient regression. In E. J. Hannan, P. R. Krishnaiah, and M. M. Rao (Eds.), *Handbook of Statistics*, Volume 5, pp. 413–449. Amsterdam, The Netherlands: Elsevier Science Publishers B. V.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2004). Stochastic volatility with leverage: fast likelihood inference. Economics Papers 2004-W19, Economics Group, Nuffield College, University of Oxford. Available at `http://ideas.repec.org/p/nuf/econwp/0419.html` (7.5.2005).

Pagan, A. R. (1996). The econometrics of financial markets. *Journal of Empirical Finance 3*(1), 15–102.

Palm, F. C. (1996). GARCH models of volatility. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics*, Volume 14, Chapter 7, pp. 209–240. Amsterdam, The Netherlands: Elsevier Science B. V.

Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance 50*(4), 1201–1228.

Pitt, M. K. and N. Shephard (1999). Time varying covariances: A factor stochastic volatility approach. In J. M. Bernardo, J. O. Berger, A. P. David, and A. F. M. Smith (Eds.), *Bayesian Statistics 6*, pp. 547–570. Oxford, UK: Oxford University Press.

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*(2), 257–286.

Richard, J. F. and W. Zhang (1996). Econometric modeling of UK house prices using accelerated importance sampling. *Oxford Bulletin of Economics and Statistics 58*(4), 601–613.

Ripley, B. D. (1987). *Stochastic Simulation.* New York, NY, USA: John Wiley & Sons.

Riskmetrics Group (1996). *Riskmetrics technical document* (4 ed.).

Rosenberg, B. (1973). Random coefficients models: the analysis of a cross-section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement 2*(4), 399–428.

Rosenberg, B., K. Reid, and R. Lanstein (1985). Persuasive evidence of market inefficiency. *Journal of Portfolio Management 11*(3), 9–17.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Roweis, S. T. and Z. Ghahramani (1999). A unifying review of linear Gaussian models. *Neural Computation 11*(2), 305–345.

Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics 7*(2), 407–425.

Russell Investment Group (2007). *Russell Indexes Construction and Methodology report.* Russell Investment Group. Available at `http://www.russell.com/ Indexes/PDF/Methodology.pdf` (26.1.2007).

Rydén, T., T. Terasvirta, and S. Asbrink (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics 13*(3), 217–244.

Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review 6*, 41–49.

Sandmann, G. and S. J. Koopman (1998). Estimation of stochastic volatility models through Monte Carlo Maximum Likelihood. *Journal of Econometrics 87*(2), 271–301.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistic 6*(2), 461–464.

Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory 11*, 61–70.

Schwert, G. W. and P. J. Seguin (1990). Heteroskedasticity in stock returns. *Journal of Finance 45*(4), 1129–1155.

Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies 62*(4), 639–661.

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics 45*(1–2), 99–120.

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies 5*(1), 1–33.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*, 425–442.

Shephard, N. (1993). Fitting nonlinear time-series models with applications to stochastic variance models. *Journal of Applied Econometrics 8* (Supplement: Special Issue on Econometric Inference Using Simulation Techniques), S135–S152.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility models. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London, UK: Chapman & Hall.

Shephard, N. (Ed.) (2005). *Stochastic Volatility: Selected Readings*. Oxford, UK: Oxford University Press.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84* (3), 653–667.

Shumway, R. H. and D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis 3* (4), 253–264.

Stadie, A. (2002). *Überprüfung stochastischer Modelle mit Pseudo-Residuen*. Ph. D. thesis, Institute for Statistics and Econometrics, Georg-August-University, Göttingen, Germany.

Stock, J. H. and M. W. Watson (1991). A probability model of the coincident economic indicators. In K. Lahiri and G. H. Moore (Eds.), *Leading Economic Indicators: New Approaches and Forecasting Records*, pp. 63–89. Cambridge, UK: Cambridge University Press.

Stoxx Ltd. (2004). *Stoxx Ltd.'s shift to Industry Classification Benchmark (ICB)*. Stoxx Ltd. Available at `http://www.stoxx.com/indexes/stoxx_shifto_icb.pdf` (7.3.2005).

Stoxx Ltd. (2005). *Methodology Overview: Style Indices*. Stoxx Ltd. Available at `http://www.stoxx.com/download/indices/methodology/djs_tmistyle_me.pdf` (26.1.2007).

Stoxx Ltd. (2007). *Methodology Overview: Size Indices*. Stoxx Ltd. Available at `http://www.stoxx.com/download/indices/methodology/sxxp_me.pdf` (26.1.2007).

Struyf, A., M. Hubert, and P. J. Rousseeuw (1996). Clustering in an object-oriented environment. *Journal of Statistical Software 1* (4), 1–30. Available at `http://www.jstatsoft.org` (29.9.2004).

Tanner, M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions* (2 ed.). New York, NY, USA: Springer-Verlag.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes—a study of daily sugar prices 1961–79. In O. D. Anderson (Ed.), *Time Series Analysis: Theory and Practice*, Volume 1, pp. 203–226. Amsterdam, The Netherlands: North-Holland.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester, UK: John Wiley & Sons.

Titterington, D. M., A. F. M. Smith, and U. E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester, UK: John Wiley & Sons.

Turner, C. M., R. Startz, and C. R. Nelson (1989). A Markov model of heteroskedasticity, risk, and learning in the stock market. NBER Working Papers 2818, National Bureau of Economic Research, Inc. Available at `http://ideas.repec.org/p/nbr/nberwo/2818.html` (1.10.2006).

van der Sluis, P. J. (1997). EmmPack 1.01: C/C++ code for use with Ox for estimation of univariate stochastic volatility models with the efficient method of moments. *Studies in Nonlinear Dynamics and Econometrics 2* (3), 77–94.

Verhoeven, P. and M. McAleer (2003). Fat tails and asymmetry in financial volatility models. CIRJE F-Series 211, CIRJE, Faculty of Economics, University of Tokyo. Available at `http://ideas.repec.org/p/tky/fseres/2003cf211.html` (15.5.2006).

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory 13* (2), 260–269.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, mimic, and varying coefficient regression models. *Journal of Econometrics 23* (3), 385–400.

Wells, C. (1994). Variable betas on the stockholm exchange 1971-1989. *Applied Economics 4*, 75–92.

Wells, C. (1996). *The Kalman Filter in Finance*. Advanced Studies in Theoretical and Applied Econometrics. Dordrecht, The Netherlands: Kluwer Academic Publishers.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (2 ed.). New York, NY, USA: Springer-Verlag.

Yao, J. and J. Gao (2004). Computer-intensive time-varying model approach to the systematic risk of Australian industrial stock returns. *Australian Journal of Management 29*(1), 121–146.

Yu, J. (2002). Forecasting volatility in the New Zealand stock market. *Applied Financial Economics 12*, 193–202.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics 127*(2), 165–178.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control 18*(5), 931–955.

Zivot, E., J. Wang, and S. J. Koopman (2002). State space modeling in macroeconomics and finance using SsfPack in S+FinMetrics. Technical report. Available at `http://faculty.washington.edu/ezivot/statespacesurvey.pdf` (13.4.2005).

Zucchini, W., A. Berzel, and J. Bulla (2006). Hidden Markov models. Lecture notes, Institute for Statistics and Econometrics, Georg-August-University, Göttingen, Germany.

State space models play a key role in the estimation of time-varying sensiti-vities in financial markets. The objective of this book is to analyze the rela-tive merits of modern time series techniques, such as Markov regime switching and the Kalman filter, to model structural changes in the context of widely used concepts in finance.

The presented material will be useful for financial economists and practitio-ners who are interested in taking time-variation in the relationship between financial assets and key economic factors explicitly into account. The empirical part illustrates the application of the various methods under consideration. As a distinctive feature, it includes a comprehensive analysis of the ability of time-varying coefficient models to estimate and predict the conditional nature of systematic risks for European industry portfolios.

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Universitätsverlag Göttingen