

Multifractal Models: Estimation, Forecasting and Option Pricing

Inaugural-Dissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Andrés Esteban Leövey, M.Sc.
aus Buenos Aires, Argentina

Hagen, 2015

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan:
Prof. Dr. Achim Walter

Erstberichterstattender:
Prof. Dr. Thomas Lux

Zweitberichterstattender:
Prof. em. Dr. Friedrich Wagner

Tag der Abgabe der Arbeit:
5. August 2014

Tag der mündlichen Prüfung:
31. Januar 2015

To my wife and family, for all their unconditional support.

Acknowledgements

First of all, I would like to express my sincere gratitude to Prof. Dr. Thomas Lux. The work presented in this thesis has gained greatly from his advise and support. I wish also to express my appreciation to Dr. Reiner Franke and to Prof. Dr. Roman Liesenfeld for all the valuable suggestions that have helped improve this thesis. I am grateful to Prof. em. Dr. Friedrich Wagner for agreeing to serve as a referee. I would like in addition to thank my colleagues at the Chair and at the PhD Program for all the long talks we have had and the liters of coffee we have enjoyed together.

I'm also grateful for the financial support received from the Deutsche Forschungsgemeinschaft (DFG) and from the Landesstipendium zur Förderung des wissenschaftlichen und künstlerischen Nachwuchses der CAU Kiel. This thesis would not have been possible without their contribution.

Finally, I would kindly thank my parents and family for the all the moral support during this period. Last but not least, to my beautiful wife, who has unconditionally stayed by my side along the way. To them I dedicate this thesis.

Executive Summary

This thesis consists of studies in the field of multifractal models applied to financial time series. The first chapter reviews the empirical regularities found in financial time series and provides a short introduction to the multifractal literature, preparing the terrain for the following chapters. Chapter 2 investigates new estimation and forecasting tools for the well-known multiplicative Lognormal cascade process. The chapter demonstrates that despite the combinatorial nature of the model similar estimation methods can be developed as for its causal counterpart in financial econometrics, without giving up the time-honored mechanisms of statistical physics for turbulent flows. Chapter 3 presents a study on the identification and estimation of the Markov Switching Multifractal (MSM) model of Calvet and Fisher^[1] with finitely many levels in continuous time. The chapter proposes a two-step mixed SMM approach for the estimation of the process and evaluates the resulting performance of this new estimator via both a Monte Carlo and an empirical study. Chapter 4 introduces a new way to obtain parameter estimates for a discrete-time MSM process through calibration to option price data. It focuses thereafter on the generation and evaluation of forecasts based on the historical as well as on the option-price time series. Appendix A provides additional results for the estimation procedures presented in Chapter 2. Appendix B goes into the details of the suggested future line of work with MSM models. A recollection of the main findings from this thesis is presented in the Conclusion.

Contents

Introduction	11
1 The Multifractal Construction as a Stochastic Volatility Model	15
1.1 Preliminaries	15
1.2 Theoretical Background	27
2 Lognormal Cascades. Parameter Estimation and Forecasting	54
2.1 Introduction	54
2.2 The Process	58
2.3 Estimation Methodology	61
2.4 Uncertainty of the number of cascade components	67
2.5 Forecasting Methodology	69
2.6 Empirical Evidence	72
2.7 Conclusion	75
2.A Appendix: Analytical Moments	86
2.B Appendix: Forecasting Formulae	93
3 Continuous-Time MSM Model. Simulation-Based Estimation	94
3.1 Introduction	94
3.2 Preliminaries	98
3.3 The continuous-time MSM model	101
3.4 The Moments	103
3.5 Estimation Methodology	107
3.6 Simulation Study	114
3.7 Empirical Results	122

3.8 Conclusion	127
3.A Appendix: Formulation of the Moments	137
4 Implied MSM Volatility Forecasting	142
4.1 Introduction	142
4.2 Model Description	146
4.3 Model Implementation	154
4.4 The Data	163
4.5 Empirical Evidence	170
4.6 Conclusion	182
Conclusions	191
A Additional Estimation Results for Chapter 2	198
B The Asymmetric MSM model	205
Bibliography	236

List of Tables

2.1	Lognormal Cascades - Monte Carlo results for GMM estimator . . .	78
2.2	Lognormal Cascades - Monte Carlo results for $\widehat{\chi}^2$, $\widehat{\lambda}_q^2$, and \widehat{F}_2 estimators	79
2.3	Lognormal Cascades - Monte Carlo results for GMM with cascade- level uncertainty added	80
2.4	Lognormal Cascades - Monte Carlo assessment of best linear fore- casts based on GMM and $\widehat{\lambda}_q^2$ estimation	81
2.5	Lognormal Cascades - Monte Carlo assessment of best linear fore- casts with cascade-level uncertainty	82
2.6	Lognormal Cascades - Empirical parameter estimates	83
2.7	Lognormal Cascades - Empirical forecast: MSE	84
2.8	Lognormal Cascades - Empirical forecast: MAE	85
3.1	Continuous-Time MSM - Monte Carlo results for GMM estimation. Bernoulli factors ($\gamma_{\bar{k}} = 3$)	130
3.2	Continuous-Time MSM - Monte Carlo results for GMM estimation. Bernoulli factors ($\gamma_{\bar{k}} = 0.69315$)	131
3.3	Continuous-Time MSM - Monte Carlo results for SMM estimation under alternative starting weighting matrices	132
3.4	Continuous-Time MSM - Monte Carlo results for SMM estimation. Bernoulli factors ($\gamma_{\bar{k}} = 3$)	132
3.5	Continuous-Time MSM - Monte Carlo results for SMM estimation. Bernoulli factors ($\gamma_{\bar{k}} = 0.69315$)	133

LIST OF TABLES

3.6	Continuous-Time MSM - Monte Carlo results for SMM estimation. Bernoulli factors ($k = 10, 15, 20$)	133
3.7	Continuous-Time MSM - Monte Carlo results for SMM estimation. Lognormal factors ($\gamma_{\bar{k}} = 0.69315$)	134
3.8	Continuous-Time MSM - Monte Carlo results for SMM estimation. QV moments ($\gamma_{\bar{k}} = 0.69315$)	134
3.9	Continuous-Time MSM - Empirical parameter estimates. Foreign exchange daily asset returns	135
3.10	Continuous-Time MSM - Empirical parameter estimates. Stock index returns	136
4.1	Implied MSM Volatility Forecasting - Simulated calibration exercise	186
4.2	Implied MSM Volatility Forecasting - Monte Carlo results for cal- ibration of parameters	186
4.3	Implied MSM Volatility Forecasting - In-sample Analysis: Empir- ical estimates and Goodness-of-fit	187
4.4	Implied MSM Volatility Forecasting - In-sample Error Analysis for Calibrated Option Prices	188
4.5	Implied MSM Volatility Forecasting - Out-of-sample Forecasting Results. Error Statistics	189
4.6	Implied MSM Volatility Forecasting - Out-of-sample Forecasting Results. Regression Analysis	190
A.1	Lognormal Cascades - Monte Carlo results for the GMM estimator	200
A.2	Lognormal Cascades - Monte Carlo results for $\hat{\lambda}_q^2$ estimator	201
A.3	Lognormal Cascades - Monte Carlo results for different $\hat{\lambda}_q^2$ estimators	202
A.4	Lognormal Cascades - Monte Carlo results for different $\hat{\chi}^2$ esti- mators (bins of equal probability mass)	203
A.5	Lognormal Cascades - Monte Carlo results for different $\hat{\chi}^2$ esti- mators (different number of bins)	204
B.1	Asymmetric MSM - Transition probabilities for $M_{k,t}$	217

List of Figures

1.1	Multifractal Volatility - Monthly- and daily-return time series of the S&P500 index (March 1950 - May 2012)	52
1.2	Multifractal Volatility - A <i>conservative</i> multifractal measure and a squared-returns sample series of the YEN (January 1979 - May 1995)	53
2.1	Lognormal Cascades - The multiplicative Lognormal cascade process.	77
3.1	Continuous-Time MSM - The SMM objective function	129
4.1	Implied MSM Volatility Forecasting - The Simulated (S)-NLS objective function	185
4.2	Implied MSM Volatility Forecasting - Integrated RV time series of the FDAX (July 2007 - February 2010)	185
B.1	Asymmetric MSM - The asymmetric MSM model for $\rho = 0$, $\rho = 1$, and $\rho = 3$	216
B.2	Asymmetric MSM - The S&P500 and a asymmetric MSM daily return time series.	217

Introduction

In this thesis I study the development of multifractal models, from their estimation and forecasting in the traditional grid-bound setting to the application of continuous-time estimation, option pricing, and volatility forecasting techniques in their causal counterparts. I conclude with a summary of the main findings and a proposal for future research avenues.

Chapter 1 starts with a description of the key research problems faced in financial econometrics followed by a short summary of the models commonly employed nowadays. After a brief description of their performance and the reasons for their potential shortcomings, it turns to the introduction of the multifractal literature, preparing the terrain for the following sections. The remaining of this thesis is organized in self-contained chapters each focusing on a specific issue involving multifractal models. These chapters are the result of my work during the last four years.

In Chapter 2, the well known multiplicative Lognormal cascade process is studied, in which the multiplication of Gaussian and Lognormally distributed random variables yields time series with intermittent bursts of activity. A previous version of this chapter can be found under the title “Parameter Estimation and Forecasting for Multiplicative Lognormal Cascades” in the Kiel Working Papers

series (cf. Leövey and Lux^[2]). A slightly different version has been published in April 2012 in *Physical Review E* (cf. Leövey and Lux^[3]).

Parameters of the multiplicative Lognormal cascade process have traditionally been estimated by fitting the numerical approximation of the associated non-Gaussian probability density function (PDF) to empirical data (cf. Castaing *et al.*^[4]). The preference for this robust, albeit oversimplified, method has resided in the practical difficulties arising from the non-stationarity of the process and the combinatorial nature of its formalism. More recently, alternative estimators based upon the kurtosis of the series (cf. Beck^[5]) or upon the q th order absolute moments (cf. Kiyono *et al.*^[6]) have also been introduced. Chapter 2 pursues this latter moment-based approach further and develops a more rigorous Generalized Method of Moments (GMM) estimation procedure to cope with the documented difficulties of previous methodologies.

Chapter 2 shows, in addition, that the estimated parameters can be used for forecasting the evolution of the turbulent flow by employing the Levinson-Durbin algorithm for best linear forecasts. Results from this implementation are compared via Monte Carlo simulations. Finally, an overall test of the approach is pursued by estimation and forecasting of volatility for a sample of financial data from stock and foreign exchange markets.

Chapter 3 focuses on the identification and estimation of the Markov Switching Multifractal (MSM) model of Calvet and Fisher^[1] with finitely many levels in continuous time. A version of this chapter can be found as a stand-alone manuscript under the title “Simulation-Based Estimation of the Continuous-Time Markov Switching Multifractal Model”.

The discrete-time MSM model has only recently been estimated by Calvet and

Fisher^[7] and Lux^[8]. The estimation of its continuous-time formulation presents, on the other hand, additional challenges. These arise from the evolution of the latent random volatility process between observations and the lack of a closed-form solution for the probabilistic law.

The chapter proceeds first with a Monte Carlo evaluation of the loss in accuracy incurred when employing the misspecified GMM estimator of Lux^[8] to the incomplete record generated by the (sampled) continuous-time model. As a remedy to the inaccuracies found in the previous step, a Simulated Method of Moments (SMM) methodology is suggested, where one can generate estimates consistent with the presumed continuous-time fluctuations. The methodology is applicable to returns as well as to realized-volatility (RV) time series arising from either a Bernoulli or a Lognormal volatility process. The algorithm provides also potentially large improvements upon the traditional two-step SMM approach in terms of computational cost. The reason for this resides in the iterative scheme applied, initialized with the results obtained by the misspecified GMM approach. A thorough analysis is pursued thereafter adopting different distributional specifications. In the next step, the chapter investigates empirically whether with the continuous-time model there are significant improvements in the in-sample fitting upon that of the discrete-time counterpart. To ease the comparison, the analysis is pursued with stock-index time series previously employed in Lux *et al.*^[9] as well as with foreign-exchange time series examined in Chapter 2.

Chapter 4 follows a different venue by introducing a new way to obtain parameter estimates for a discrete-time MSM Bernoulli process through calibration to option price data. A version of this chapter can be found as a stand-alone manuscript under the title “Implied Multifractal Volatility Forecasting: The Complementary

Benefits of Information from Derivative Contracts and Historical Data”.

The chapter introduces first as underlying a dividend-discounted stock index, for which the time evolution in both the ‘statistical’ setting as well as the equivalent ‘risk-neutral’ counterpart is modeled. While the process in the former setting is estimated by the well known Maximum Likelihood (ML) estimation procedure, the ‘risk-neutral’ version is calibrated via a Simulated Non-Linear Least Squares (SNLS) optimization procedure.

The empirical analysis is undertaken by selecting a highly liquid stock index, like the DAX. Chapter 4 uses the information collected from the ML estimation and SNLS methodologies to analyze the goodness of fit in-sample as well as the ability of the ‘risk-neutral’ MSM process to capture information regarding future volatility contained in option prices data. Hence, Chapter 4 complements previous studies on MSM processes based solely on backward-looking data like daily-return time series. By Bayesian updating, the usual forecasting methodology with MSM, quadratic variation (QV) forecasts are performed for both parameter sources, and compared against the actual RV for the same period, constructed on the base of 5-minute interval transaction prices of the DAX.

In the conclusion, I summarize the major findings in this thesis regarding the key research problems. A number of avenues for further research are also suggested, with special interest put in the extension of the traditional MSM model of Calvet and Fisher^[1,7]. Appendix A will be dedicated for complementary Monte Carlo estimation results accompanying the analysis of Chapter 2. Appendix B provides some details on the extension proposed for future work on the traditional MSM model.

Chapter 1

The Multifractal Construction as a Stochastic Volatility Model

1.1 Preliminaries

As one of most important approaches in financial econometrics, time series modeling focuses on capturing the main characteristics of asset return distributions with the goal of understanding and subsequently replicating the nontrivial statistical properties of these assets. From this perspective, the seemingly random variations of asset prices share some properties that are common across a wide range of instruments, markets and time periods. The collection of such properties is better known as the *stylized facts* and is the result of taking a common denominator among the most commonly observed phenomena in instruments and markets over the last half of century or more (cf. Cont^[10]).

The traditional approach in time series analysis has been to focus on discrete-time returns calculated from various sampling intervals. Since the early contributions

of Mandelbrot^[11] and Fama^[12,13], the depiction of the (possibly non-Gaussian) underlying distribution follows from the analysis of the sample moments of the series. According to this strategy, the most well-known properties recollected in the literature can be outlined as follows (cf. Pagan^[14], Campbell *et al.*^[15], Cont^[10], and Lux^[16]):

- Absence of return autocorrelations. Asset returns are time-dependent though the autocorrelations are often very low, except for intraday high-frequency data in which microstructure effects may arise.
- Leptokurtosis. Return distributions present excess kurtosis with respect to the Gaussian distribution and may eventually follow a cubic law of their tails. The excess kurtosis of the tails can in most cases be well described by a power-law or Pareto-like function, with a tail index which is finite, higher than two and less than five.
- Skewness. While there is a certain asymmetry between the probability of positive and negative returns, this fact depends highly on the asset class considered. Foreign exchange rates and certain index stock returns display high symmetry in up and down movements. Single and index stocks that do not capitalize dividends present, on the other hand, large downward jumps but not equally large upward shocks.
- Aggregational Normality. With larger scales of time aggregation, asset return moments become closer and closer to those of the Gaussian distribution. At particular shorter time horizons though, return fluctuations seem to obey more leptokurtic distribution laws.

- Volatility peaks. A high degree of variability can be observed in returns at any time scale. This is depicted by the presence of spikes in the squared return series and quantified by irregular bursts in the conditional volatility measures.
- Positive volatility autocorrelation. Volatility events seem to cluster in time. Low fluctuation periods are followed by low fluctuation periods, while large fluctuation periods are followed by further large ones.
- Conditional leptokurtosis. Even after correcting daily returns by traditional measures of volatility clustering (e.g. via ARCH-type models), the residual time series exhibits leptokurtosis.
- Slow decay of absolute return autocorrelations. When considering various powers of absolute returns, much higher and longer lasting autocorrelations are typically observed.
- Return-volatility dependence. Most measures of return variation show a negative correlation to the return series innovations. This correlation seems to vary in time and be relative to the magnitude of the return innovations, not only its sign.
- Volatility-trading volume correlation. Trading volume has a high contemporaneous correlation with volatility.

From examination of these *stylized facts*, it is clear that the conditional and unconditional second moments play a central role in the depiction of the probability distribution associated to the return series. Note, on the other hand, that for decades the Gaussian probability law has been employed as an approximation

to empirical return distributions. Famous contributions are found in studies of portfolio evaluation, market-efficiency theory, and even in option pricing. Given the time-honored use of this distribution in the literature, the importance of the volatility behavior resides not only in the depiction of the second moments but also in its role played in the higher-order statistics, among them, the various types of autocorrelations and kurtosis measures. As a result, a great deal of effort in the financial literature has been linked to the enhancement of the properties of the Gaussian distribution by introducing more realistic volatility features.

Initially introduced by Engle^[17] and Bollerslev^[18], ARCH-type models have been the first successful attempt to replicate some of the aforementioned *stylized facts*. The approach consists in modeling the time dependence of volatility as a deterministic function of the realized past innovations. Due to the sequential nature of the link built between past innovations and contemporaneous volatility though, this framework does not offer an integral view of return fluctuations at different time scales (cf. Jondeau *et al.*^[19]).¹ Note, moreover, that by definition the approach does not contemplate the use of an exogenous fluctuation source for the volatility dynamics. This fact, together with the limited range of variation arising from the approach, and the necessity to specify parametrically each additional feature of the *stylized facts* represent well documented drawbacks of this framework.

A successful alternative to ARCH-type models was first proposed by Taylor^[21,22], and it is denoted nowadays as the stochastic volatility (SV) model. The approach consists in introducing a latent state variable for the evolution of volatility. In

¹ Drost and Nijman^[20] have more recently introduced a new category of ARCH-type processes that may allow for a congruence of return distributions at different time scales.

the traditional setting, the former is depicted by an autoregressive drift and a Lognormally distributed innovation. The intuition behind this framework is that asset returns may be well approximated by a mixture distribution, where the mixing factor is linked to the second moment of the first distribution (cf. Shephard and Andersen^[23]). The usual construction builds upon the Mixture of Distributions Hypothesis (MDH) of Clark^[24], who depicted continuously compounded log-prices as being driven by the subordination of a Brownian motion to a non-decreasing process acting as a mixture. The latter process resulted from the sum of independent and identically distributed (i.i.d.) positive random variables reflecting information arrivals. More generally, and as long as the mixing process remains independent of the Brownian motion, Clark's^[24] MDH may depict a return fluctuation with serially uncorrelated, albeit dependent, increments, which in discrete time accommodates Taylor's model (cf. Shephard and Andersen^[23]). By definition, a second source of fluctuations makes SV models more flexible than the ARCH-type frameworks. The use of this artifact also embodies the change in focus away from the adjustment of past filtered observations as a tool to represent volatility. A classical example of the change in paradigms can be found when contemplating the correlation between past returns and contemporaneous volatility in empirical stock times series. While ARCH-type models recur to functions of past return innovations and, as such, to actual realizations of random variables to contemplate such effects, asymmetric volatility responses are incorporated in SV models by introducing a parametric dependence between the return innovation and the volatility process distributions. One should note on the other hand that due to the general lack of closed-form solutions for these mixed probability laws, implementation of SV models require alternative ways of handling the inherently

latent nature of the state variables. In short, these procedures tend to be more complex and computationally more expensive than the ones used for ARCH-type models.²

It wasn't though until the analytical and technical developments in the areas of option pricing and, separately, of high-frequency time series that SV models were brought to the center of attention (cf. Andersen *et al.*^[32], Shephard^[33], and Shephard and Andersen^[23]). In this sense, option-price and high-frequency data sources have presented alternative ways to unveil the essence of these state variables. The advent of high-frequency data has prompted authors, on the one hand, to the study of 'smoother' estimators of volatility based upon the increments of quadratic variation (QV) (cf. Shephard^[33]). In this direction, the realized volatility (RV) metric, has led to the notion that return fluctuations can be handled as an 'observable' variable, otherwise inherently latent when considering daily-return fluctuations (cf. Andersen *et al.*^[34] and Andersen *et al.*^[32]). The distinctive nature of derivative markets relies, on the other hand, on the fact that agents must engage in forming *expectations* about future non-linear payoffs of the underlying asset. Derivative prices cluster therefore market 'beliefs' about the possible evolution of the underlying asset fluctuations over the life-span of the derivative contract. I propose at this stage a short detour to review the empirical regularities found in these two fields of study that will contribute to a better

² Nowadays standard approaches for these models recur to simulation techniques. Well documented applications for inference include, for instance, the Markov chain Monte Carlo (MCMC) (cf. Jones^[25], Eraker^[26], and Roberts and Stramer^[27]) and the Simulated Maximum Likelihood (SML) (cf. Elerian *et al.*^[28], Brandt and Santa-Clara^[29], Durham and Gallant^[30], and Durham^[31]) methods. The common feature in these methodologies resides in the sample infilling of the return time series with artificially generated data belonging to the latent variable. This is done to be able to depict the potential evolution of the volatility process in the empirical scenario (cf. Shephard and Andersen^[23]).

understanding of the success as well as of the shortcomings of SV models.

To start with, recall that the use of high-frequency data allows for a deeper analysis of continuous-time return fluctuations, usually inaccessible in the presence of monthly-, weekly-, or even daily-sampled data. Consider, for instance, the study of the return fluctuations of the S&P500 index, based upon monthly and daily return series solely.³ Figure 1.1 at the end of this chapter depicts both sample series. The monthly series (top panel) exhibits only one positive and some negative outliers, but it looks otherwise rather *normal*. One could in fact derive the set of reliable statistics summarizing the overall behavior of the series at this frequency and come to the conclusion that its data generating process (DGP) is Gaussian-like. A derivation from this series of the *dynamic* properties of the DGP that are applicable even at very short time intervals would be nonetheless moot. The shift from the monthly to the daily frequency (bottom panel) allows for a better visualization of the return fluctuation, which by the size and frequency of the large shocks exhibits not only a negative skewness but also a high degree of leptokurtosis. In attempting to describe the DGP, note that kurtosis is closely connected to the range of values the volatility process may take as it considers both tails of a distribution alike. Skewness, on the other hand, focuses on the asymmetry of the distribution of the DGP with respect to the mean, and it may be analyzed separately once the nature of the volatility behavior is accounted for. The arising body of literature focusing on volatility shows that ARCH-type and traditional SV models may eventually be incapable of generating the sudden spikes in return fluctuation levels. This is true despite the volatility pulsations

³ The S&P500 index has been one of the most widely used assets for the evaluation of extensions to the traditional SV model (cf. Bates^[35], Chernov and Ghysels^[36], Benzoni^[37], Chernov *et al.*^[38], Pan^[39], Eraker *et al.*^[40], Eraker^[41]).

generated by the traditional SV model being considerably higher than that of the ARCH-type models. Hence, the matter at this frequency becomes whether the features captured in the sample (conditional) second and fourth moments (in this case of the S&P500 index) could have been solely generated by a sequence of unlikely draws of the latent state variable in the volatility process under the traditional SV. If this were not the case, or even if it were but at the cost of employing a calibration that reverts the situation and imposes the process to wildly evolve with only a few transitory periods of low activity, the introduction of additional state variables for the improvement of the flexibility of the model would be plausible. Some authors have suggested, for instance, the introduction in the DGP of the S&P500 index of a jumping process with a low rate of innovation arrivals. This additional process may not only lead to an increase in the level of volatility and kurtosis but also contribute to the partial replication of the skewness observed in the bottom panel of Figure 1.1 (cf. Jondeau *et al.*^[19]). Note that in this case, one would still face the difficulty of pinpointing the impact of the *dynamic* volatility component to the overall return fluctuation, which may not be uniquely determined when employing daily observations only. Consequently, the study of latent fluctuation sources still remains a challenge at this frequency. Given the various arrival rates the latent state variables may have though, data sampled at increasingly shorter time intervals may allow for a better allocation of the different return fluctuation sources (cf. Aït-Sahalia^[42]).

The development of SV models in the context of high-frequency fluctuations has started with concurrent papers from Andersen and Bollerslev^[43] and Barndorff-Nielsen and Shephard^[44]. In their work on foreign exchange returns, Andersen *et al.*^[34,45] have been able to establish some additional *stylized facts* of volatility

self. These are:

- Lognormality of the daily volatility. When taking logs of the square root of RV, the latter become well approximated by a Gaussian distribution.
- Positive autocorrelation of daily volatility. The square root of RV is persistent with highly significant autocorrelations beyond the 20th order.
- Slow volatility decay. Hyperbolic decay in the autocorrelations of logarithmic square root of RV giving signs of a long-range dependence.

The authors notice moreover that once foreign exchange daily returns have been standardized by the *ex post* RV, these are well approximated by a Gaussian distribution without further evidence of volatility clustering nor leptokurtosis. These results strengthen the notion of the existence of a second underlying factor, possibly Lognormally distributed, driving the trend of volatility. They limit in addition the universality of ‘conditional leptokurtosis’ of returns as an *stylized fact*, which may be in part the result of ARCH-type models not being able to generate enough excess kurtosis (cf. Andersen *et al.*^[34]). One should note on the other hand that the presence of large positive autocorrelations in volatility motivates the use of more complicated versions of the SV model. The long memory of volatility may lead, for instance, to the introduction of heterogenous components of various durations. This, together with the fact that traditional SV models are occasionally unable to generate sufficient leptokurtosis, may be perceived as non-negligible drawbacks of the standard framework.⁴

⁴ At this point, it would be fair to mention that the *stylized facts* of the RV series may alone be insufficient to prove the existence of a DGP with stochastic volatility. Cont and Tankov^[46] have made the appealing argument that a return series generated by a Student-t random walk (SRW) with tail index equal to 3.8 is able to generate similar RV features as the ones depicted here, despite the SRW having i.i.d. increments and constant (conditional and

The interest in SV models in the context of option pricing emerged when market participants first noticed the presence of smiles and smirks in implied volatility surfaces.⁵ While many attempts to explain this matter were made, it was the essay of Hull and White^[48] that caught the general attention by indicating that these features could be replicated by SV models. Further publications of Stein and Stein^[49] and Heston^[50] deriving quasi-analytical solutions for a computationally efficient calculation of option prices have led to the preponderance of SV models in this field of study.

Derivative prices cluster *market expectations* about the possible price evolution of the underlying asset. Due to the wide range of beliefs that may be at any time present, absence of arbitrage can only be assured in these markets when an investor with risk-neutral preferences exists. *Stylized facts* refer in this context to the risk-neutral distribution invoked in the pricing of derivative contracts (cf. Garcia *et al*^[51]). Following the conventions in this kind of studies, I refer to these *stylized facts* in terms of the implied volatility surface:

- Leptokurtosis. The implied volatility surface exhibits a convexity for strikes other than at-the-money.

unconditional) second moments. Hence, the mere presence of heavy tails in the DGP may lead to falsely associate these empirical regularities to heteroscedasticity and stochastic volatility (cf. Cont and Tankov^[46]). Note, nonetheless, that because processes with i.i.d. increments are unable to generate volatility clustering, a slow decay of absolute return autocorrelations, nor a dependence between volatility and past returns, one may interpret this contention as a further case against traditional SV models with short-lived volatility memory and moderate kurtosis levels.

⁵ The implied volatility surface is a three dimensional plot with axes representing the maturity, the moneyness, and the implied volatility of the option. The latter is calculated by finding the volatility parameter in the Black-Scholes formula by which the predicted value of the derivative contract matches its current market value (cf. Black and Scholes^[47]). A correct price prediction by the Black-Scholes formula for all degrees of moneyness would lead to a flat surface.

- Negative skewness. The convexity of the implied volatility curve is more pronounced at one side of the at-the-money strikes than at the other.⁶
- Time-varying distribution. The asymmetry and curvature shape of the implied curves change across maturities. Skewness and curvature seem to be strong in the short maturities and fades-out in the longer ones.
- Slow volatility decay. Implied volatility shapes tend to have a non-negligible curvature still at very long maturities.

Additionally to these findings, it has been noted that the RV of the underlying asset over the life-span of the option differs substantially from the implied volatility extracted from at-the-money contracts (cf. Garcia *et al*^[51]). This divergence between *ex ante* implied volatility and *ex post* RV has led to reinforce the notion of a time-varying and possibly stochastic volatility. It is also noteworthy that the level of leptokurtosis and skewness depend on the different maturities of the options. This may indicate that additional state variables with varying arrival times may be required. It has also been reported that SV models alone are unable to generate enough sudden spikes in leptokurtosis at short maturities, which is needed to contemplate the chance of immediate extreme events. Clearly, the magnitude of the required convexity, the possibly heterogeneous durations, and the presence of a slow decay of volatility may indicate that the source of the return fluctuations may lie outside the traditional SV frame (cf. Garcia *et al*^[51]).

The amount of research accumulated over the last two decades has in summary

⁶ There is a general agreement among studies that risk-neutral distributions are more negatively skewed than return distributions from historical time series. The reason for this resides in the nature of the alternative information sources, that is, *ex-ante* expectations on the one hand and *ex-post* outcomes on the other. Note, nonetheless, that the asymmetry in the distribution depends heavily on whether dividends are included in the study.

shown that more complicated designs are needed to replicate the dynamic of the return fluctuations. Current extensions to the traditional SV model already conceive, among others, the introduction of jumps into the volatility process (cf. Barndorff-Nielsen and Shephard^[44] and Eraker *et al.*^[40]) or, alternatively, to model the volatility process as a function of a number of separate stochastic processes or factors (cf. Barndorff-Nielsen and Shephard^[44] and Chernov *et al.*^[52]). The problem with the growing complexity of these models though is the arising difficulty in identifying and efficiently estimating the increasingly larger parameter space, which together with the general lack of closed-form analytical solutions is what has mainly characterized this strand of research (cf. Shephard^[33]).

Multifractal models arise in the financial literature as an important alternative to the approaches presented so far. This new generation of models distinguishes itself by its parsimony, the focus on temporal multiscaling, and the capacity to replicate sudden increase in fluctuation spikes without recurring to boundary parameter values. The multifactor nature of the volatility process may also allow in certain settings to replicate nearly every *stylized fact* of the return time series. In its causal form, the model fits into the literature of regime-switching and Itô diffusions, and as such, it can clearly be included amidst the extensive SV literature. The basic principles for the construction of these models are, on the other hand, very different from those used in the models reviewed so far. The next section is dedicated to an introduction to this framework.

1.2 Theoretical Background

We begin with a formal introduction to the multifractal literature and other related concepts. A general introduction on stochastic processes can directly be taken from Billingsley^[53], Karatzas and Shreve^[54], and Cont and Tankov^[46], among many others. Here, we will focus on those topics that will help us evaluate the subject matter of the following chapters. The concepts and definitions regarding multifractal phenomena are addressed in Mandelbrot^[55,56,57,58], Mandelbrot *et al.*^[59], and Calvet and Fisher^[1,7,60]. For brevity, we will outline the common properties among the most well-known multifractal constructions from these authors only, while the specific details regarding their construction will be picked up upon later on.⁷ Finally, no claims of authorship are made for the definitions and propositions presented in this section.

Definition (Stochastic process) The *stochastic process* $X(t)$ (or alternatively X_t in the discrete-time case) is a *collection of random variables* X on the *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. The *index* t admits the interpretation of *time* in the *bounded interval* $[0, T]$, with $T < \infty$.⁸

Every stochastic process in our first definition is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, being Ω the sample space, \mathcal{F} the sigma algebra of Borel sets, $\mathcal{F}(t)$ (or

⁷ Alternative designs based upon the Multifractal Random Walk (MRW) (cf. Bacry *et al.*^[61], Pochart and Bouchaud^[62], Saichev and Sornette^[63], Saichev and Filimonov^[64], Bacry *et al.*^[65]) or on the Self-Excited Multifractal (SEMF) (cf. Filimonov and Sornette^[66]) models exist, and the reader should refer to these works in case of interest. These models have been left aside in this thesis though, due to the comparatively lower impact they have made in the financial literature so far.

⁸ Note that a discrete time process X_t is only valued between the uniform time-increments Δt , taken at points $t_1, \dots, t_k \geq 0$. A continuous-time process $X(t)$ arises by letting $\Delta t \rightarrow 0$. Eventually, we will also consider the $\delta_{\Delta t}$ -increment of the price process X_t , or $X(t)$, respectively.

alternatively \mathcal{F}_t) the filtration of sub-sigma algebras of \mathcal{F} , and \mathbb{P} the probability measure assigning the probability of occurrence of each event in (Ω, \mathcal{F}) . Along this thesis, we will fix our attention on the real-valued random variables X only, with $\mathbb{E}[|X|] < \infty$; conditional expectations with respect to $\mathcal{F}(t)$ (or \mathcal{F}_t) are assumed to exist and are denoted by $\mathbb{E}_t[\cdot]$.

Recall that the probability measure \mathbb{P} need not be unique. In particular, another probability measure $\tilde{\mathbb{P}}$ can assign a different probability of occurrence to an event in (Ω, \mathcal{F}) . A natural way to compare these two probability measures is to look at their ratio $\frac{\tilde{\mathbb{P}}(\mathcal{A})}{\mathbb{P}(\mathcal{A})}$ for various measurable sets $\mathcal{A} \in \mathcal{F}$. Because this ratio would be undefined for a denominator equal to zero, this comparison makes only sense if $\tilde{\mathbb{P}}(\mathcal{A}) = 0$ every time $\mathbb{P}(\mathcal{A}) = 0$ (cf. Cont and Tankov^[46]). In case the reverse relationship is also true, one encounters the following definitions.

Definition (Equivalent probability measures) The *probability measures* \mathbb{P} and $\tilde{\mathbb{P}}$ in (Ω, \mathcal{F}) are *equivalent* if they agree on which *sets* in \mathcal{F} have *probability zero*.

Definition (Radon-Nikodym variable) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a *probability space*, $\tilde{\mathbb{P}}$ the *equivalent probability measure* of \mathbb{P} in (Ω, \mathcal{F}) , and Λ a *nonnegative random variable* with $\mathbb{E}[\Lambda] = 1$. For $\mathcal{A} \in \mathcal{F}$, Λ is called the *Radon-Nikodym variable* of $\tilde{\mathbb{P}}$ with respect to \mathbb{P} , when it relates the two by

$$\tilde{\mathbb{P}}(\mathcal{A}) = \int_{\mathcal{A}} \Lambda \, d\mathbb{P}.$$

To illustrate this definition, consider two random variables $X \stackrel{d}{=} N(\mu, \sigma^2)$ and $\tilde{X} \stackrel{d}{=} N(0, \sigma^2)$, and the set $\mathcal{A} = (-\infty, a]$.⁹ As absolutely continuous random

⁹ The application of this definition in the Gaussian case only works for the mean parameter.

1.2 Theoretical Background

variables, X and \tilde{X} have the density function $f(\cdot)$, with $f(\cdot) = f_\varepsilon(y)/\sigma$ and where $f_\varepsilon(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$ is the standard Gaussian distribution, with $y = \frac{x - \mu}{\sigma}, \frac{\tilde{x}}{\sigma}$. Our goal is to evaluate how the probability allocation of X under \mathbb{P} is connected to that of \tilde{X} under $\tilde{\mathbb{P}}$ through the Radon-Nikodym variable. It follows from the above definition that

$$\tilde{\mathbb{P}}(\tilde{X} \leq a) = \tilde{\mathbb{E}}[\mathbb{I}_{\mathcal{A}}] = \mathbb{E}[\Lambda \cdot \mathbb{I}_{\mathcal{A}}] = \int_{-\infty}^a \Lambda \cdot f(x) dx,$$

where $\mathbb{I}_{\mathcal{A}}$ is a variable assigning the value of $\mathbf{1}$ if $\tilde{x} \in \mathcal{A}$ and $\mathbf{0}$ otherwise.¹⁰ For a constant $\theta \equiv \frac{\mu}{\sigma}$, the random variable $\Lambda \equiv \exp\{-\theta Y - \frac{1}{2}\theta^2\}$ leads to the Gaussian density $f(\tilde{x}) = \Lambda \cdot f(x)$, with $Y = \frac{X - \mu}{\sigma}$.

Another well-known application of the Radon-Nikodym variable is found with the exponentially distributed random variable Z , having density $f(z) = \phi e^{-\phi z}$ and a positive constant ϕ . In this case, the Λ that connects Z under \mathbb{P} and \tilde{Z} under $\tilde{\mathbb{P}}$ looks like $\Lambda \equiv \frac{\phi'}{\phi} \exp\{-(\phi' - \phi)Z\}$, with ϕ' being another positive constant.

Our last example will help us illustrate the application of these concepts in the finite state space.¹¹ Consider the sample space

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

resulting of tossing a coin three times independently, where $\omega_i = H$ represents the coin showing ‘head’ and $\omega_i = T$ showing ‘tail’ in the toss i . The event of obtaining for instance three heads is represented by $\mathcal{B} = \{HHH\}$. The event of

The variance parameter should remain the same on the other hand.

¹⁰ Note that $\int_{\mathcal{A}} \Lambda d\mathbb{P}$ is a Lebesgue integral, where in case a density $f(x)$ exists, the calculation can be performed by the usual Riemann integral.

¹¹ This example can be found in Shreve^[67]. We refer the reader to this source for a treatment of equivalent probability measures and Radon-Nikodym variables in the Binomial Asset Pricing model.

1.2 Theoretical Background

two heads in the first two tosses is denoted also by $\mathcal{C} = \{HHH, HHT\}$. Note that we may retain in our construction the concept of time when each toss is made at a subsequent instant.¹²

We assume also that due to the nature of the coin, the probability of a head is $p = \frac{2}{3}$ and of a tail is $q = 1 - p = \frac{1}{3}$. The probability measure \mathbb{P} is

$$\begin{aligned} \mathbb{P}(HHH) &= \frac{8}{27}, & \mathbb{P}(HHT) &= \frac{4}{27}, & \mathbb{P}(HTH) &= \frac{4}{27}, & \mathbb{P}(HTT) &= \frac{2}{27}, \\ \mathbb{P}(THH) &= \frac{4}{27}, & \mathbb{P}(THT) &= \frac{2}{27}, & \mathbb{P}(TTH) &= \frac{2}{27}, & \mathbb{P}(TTT) &= \frac{1}{27}. \end{aligned}$$

Assume now that another coin exists with probabilities $\tilde{p} = \frac{1}{2}$ for a head and $\tilde{q} = \frac{1}{2}$ for a tail. The corresponding probability measure is

$$\begin{aligned} \tilde{\mathbb{P}}(HHH) &= \frac{1}{8}, & \tilde{\mathbb{P}}(HHT) &= \frac{1}{8}, & \tilde{\mathbb{P}}(HTH) &= \frac{1}{8}, & \tilde{\mathbb{P}}(HTT) &= \frac{1}{8}, \\ \tilde{\mathbb{P}}(THH) &= \frac{1}{8}, & \tilde{\mathbb{P}}(THT) &= \frac{1}{8}, & \tilde{\mathbb{P}}(TTH) &= \frac{1}{8}, & \tilde{\mathbb{P}}(TTT) &= \frac{1}{8}. \end{aligned}$$

For $\mathcal{A} = \omega_1\omega_2\omega_3 \in \mathcal{F}$, the *Radon-Nikodym variable* is defined as the ratio

$$\Lambda \equiv \frac{\tilde{\mathbb{P}}(\mathcal{A})}{\mathbb{P}(\mathcal{A})} = \left(\frac{\tilde{p}}{p}\right)^{\#H(\mathcal{A})} \cdot \left(\frac{\tilde{q}}{q}\right)^{\#T(\mathcal{A})},$$

where $\#H(\mathcal{A})$ denotes the number of heads in the sequence $\omega_1\omega_2\omega_3$, and $\#T(\mathcal{A})$ the number of tails. As a result, the value of Λ for each event arises to

$$\begin{aligned} \Lambda(HHH) &= \frac{27}{64}, & \Lambda(HHT) &= \frac{27}{32}, & \Lambda(HTH) &= \frac{27}{32}, & \Lambda(HTT) &= \frac{27}{16}, \\ \Lambda(THH) &= \frac{27}{32}, & \Lambda(THT) &= \frac{27}{16}, & \Lambda(TTH) &= \frac{27}{16}, & \Lambda(TTT) &= \frac{27}{8}. \end{aligned}$$

The application of the *Radon-Nikodym variable* and the notion of *equivalent probability measures* will become handy in Chapter 4 when dealing with the evaluation of option prices. We continue now with some properties of stochastic processes necessary for the rest of this thesis. The depiction of a time course with potential

¹² In this case, we should also need to specify the collection of subsets \mathcal{F}_t , for $t = 0, 1, 2, 3$.

accelerating and decelerating periods is of special interest:

Definition (Time deformation) The *process* $\theta(t)$ is a *time-deformation process* in $[0, T]$ if $\theta(0) = 0$, and $\theta(t)$ is a *nondecreasing, right-continuous function* of the *clock time* t in $[0, T]$, with $\mathbb{E}[\theta(t)] < \infty$.

While the process $\theta(t)$ in our definition may be either deterministic or stochastic, the latter choice presents more interesting features. Consider, for instance, the case of the stochastic process $\sigma(t)$, traditionally used to depict the time-evolution of volatility. One could define $\theta(t) \equiv \int_0^t \sigma^2(s) ds$ and regard $\sigma(t)$ as the intensity change of information arrivals in Clark's^[24] MDH. $\theta(t)$ is referred to as a stochastic 'trading time' in this case.

The stochastic process $\theta(t)$ could alternatively be seen as a counting (jump) process, being the Poisson process with independent information arrivals its simplest form. $\theta(t)$ could also be understood amidst the random-measure framework of Probability Theory, which is where the multifractal literature finds its origins (cf. Mandelbrot^[55]). Note, however, that the distinction between counting processes and random measures is rather theoretical, and the application of the latter does not present major advantages in a practical context. Taking this into account, the theory of multifractality is seen to have evolved towards the analysis of stochastic processes. We will present a detailed overview of this course of events in the following pages. Before this, two other important classes of processes need to be introduced:

Definition (Markov process) Let $X(t)$ be an *adapted process*, with $s \leq t$ in the *bounded interval* $[0, T]$. Assume that for every *nonnegative, Borel-measurable function* h , there is another *Borel-measurable function* g such that

$$\mathbb{E}_s[h[X(t)]] = g[X(s)].$$

$X(t)$ is then called a *Markov process*.

The Markov property is satisfied when the evolution of a process after time t is independent of its past before that instant. This is commonly known as the 'memoryless' property of Markov processes. Some authors emphasize this feature by writing $\mathbb{P}[X(t) | \mathcal{F}(s)] = \mathbb{P}[X(t) | X(s)]$, for every random variable X depending on $\mathcal{F}(s)$, $s < t$.

Definition (Martingale process) The *process* $X(t)$ is said to be a *submartingale* (*supermartingale*) if, for every $s \leq t$ in the *bounded interval* $[0, T]$, one observes $\mathbb{E}_s[X(t)] \geq X(s)$ ($\mathbb{E}_s[X(t)] \leq X(s)$). $X(t)$ is a *martingale* if it is both a *submartingale* and a *supermartingale*.

Well-known continuous-time martingales are the Brownian motion and the compensated Poisson process. More generally, consider the Itô diffusion $I(t) \equiv \int_0^t \sigma(s) dB(s)$ in $[0, T]$, with $\sigma(t)$ representing an adapted process such that $\mathbb{E} \left[\int_0^T \sigma^2(s) ds \right] < \infty$, and $B(t)$ a Brownian motion. As a $B(t)$ -based mixing distribution process, $I(t)$ obeys likewise the definition of a martingale.

A standard example of a submartingale is, on the other hand, the Poisson process. Note also that by Jensen's inequality a convex function $\psi(\cdot)$ of a martingale $X(t)$ is a submartingale. In particular, $X(t)^2$ is a submartingale.

Submartingales are widely employed in the financial literature. Note, for instance, that under the Asset Pricing Theory securities exposed to systematic risk must compensate their holders with an excess of return. As a result, several security types do not behave like martingales but rather like submartingales. This leads us to the following theorem:

Theorem (Doob-Meyer decomposition) Let $B(t)$ be a *right-continuous martingale* and $\theta(t)$ an *increasing process*. The *right-continuous submartingale* $X(t)$ can be expressed as $X(t) = B(t) + \theta(t)$.¹³

Consider the log-price $X(t)$ composed of a Brownian motion $B(t)$ and a finite variation process $\theta(t)$ constituting the expected mean return. This setting allows us to replicate the statistical properties of a security as required under the Asset Pricing Theory by employing $B(t)$ to depict unpredictable price changes and $\theta(t)$ investors' expectations about compensation for systematic risk. This approach will become helpful in the implementation of multifractal processes in the following chapters.

Another standard example of the application of the theorem consists in the Poisson process $Y(t)$ with intensity λ . The Doob-Meyer decomposition of $Y(t)$ reads $Y(t) = Z(t) + \theta(t)$, where $Z(t)$ is a compensated Poisson process and $\theta(t) = \lambda t$. More generally, consider a process $U(t)$ with independent increments, such that $\mathbb{E}[U(t)] < \infty$. The resulting process $U(t) - \mathbb{E}[U(t)]$ is a martingale with independent increments regardless of whether $U(t)$ is a sub- or a supermartingale.

The previous examples focus on the application of $\theta(t)$ as a simple drift in the log-price $X(t)$. The Doob-Meyer decomposition presents also an additional application in the study of sample-path variation of a process, where $\theta(t)$ may show more complex features.

Definition (Quadratic variation process) Let $X(t)$ be a *square-integrable process*, such that $X(0) = 0$ and $\mathbb{E}[X(t)^2] < \infty$. The *quadratic variation (QV) process* $[X](t)$, with $[X](0) = 0$, is the unique *adapted process* for which $X(t)^2 -$

¹³ We refer the interested reader to Karatzas and Shreve^[54] for a more detailed version of this theorem as well as for its proof.

$[X](t)$ is a *martingale*. In particular, $[X](t)$ is the *increasing process* in the *Doob-Meyer decomposition* of $X(t)^2$.

A QV process admits the alternative representation as a limiting process: for a sequence of n uniform partitions of size $\Delta t = \frac{t}{n}$ in $[0, t]$, with $\Delta t \rightarrow 0$ as $n \rightarrow \infty$, let $[X](t) \equiv \text{p lim } \sum_{j=1}^n [X(j\Delta t) - X((j-1)\Delta t)]^2$. The QV process is defined as a limit of positive sums, capturing the variation of the process along the sample path. Note also that it does not involve the use of moments, meaning that the process is well defined even for those processes with infinite variance.¹⁴

To illustrate, consider a Brownian motion $B(t)$, with $[B](t) = t$. It follows that $M(t) \equiv B(t)^2 - t$ is a martingale, with $M(t) = 2 \int_0^t B(s) dB(s)$ (cf. Shreve^[68]). Note moreover that $[B](t)$ does not depend on the path of $B(t)$. This is a special feature of Brownian motions. For other martingales like the Itô diffusion $X(t) = \int_0^t \sigma(s) dB(s)$, $[X](t)$ can depend on the path of $X(t)$ provided $\sigma(s)$ can depend on the path. Furthermore, $[X](t)$ may presents jumps at those instants $X(t)$ does. If $X(t)$ is otherwise continuous and has sample paths of finite variation, it follows that $[X](t) = 0$. This introduces a difference between the variability of the drift processes presented in the previous definition and the *intermittency* of processes with nonzero QV. The latter are said to be inherently ‘more random’ (cf. Shreve^[68], Cont and Tankov^[46]).

Note, finally, that $[X](t)$ can present sudden jumps and path-dependent behavior even when $X(t)$ follows a mixing distribution process where volatility is not necessarily the mixing factor. In turn, the process $X(t)$ may be devised by linking its components over the time scale. This leads us to the following definition:

¹⁴The concept of QV, together with its realized volatility (RV) estimator, will accompany us along Chapter 3 and Chapter 4.

Definition (Compound process) Let $B(t)$ be a *stochastic process*, and $\theta(t)$ a *nondecreasing function of clock time t* . The process $X(t) \equiv B[\theta(t)]$ is called a *compound process*.

The process $X(t)$ is the result of subordinating the process $B(t)$ by a (possibly) stochastic increasing process $\theta(t)$. The usual choice for $B(t)$ consists of a martingale processes with independent increments, like the Brownian motion. Here, the independence of $B(t)$ and $\theta(t)$ leads to a zero-mean and symmetric process $X(t)$, with

$$\text{VAR}[X(t)] = \sigma^2 \mathbb{E}[\theta(t)], \quad \mathbb{KUR}[X(t)] = \frac{3 \text{VAR}[\theta(t)]}{\mathbb{E}[\theta(t)]^2},$$

and $\mathbb{KUR}[X(t)]$ referring to kurtosis as a scaled version of the fourth moment of $X(t)$.

Note that one could still make use of this definition in the SV case by letting $\theta(t) \equiv \int_0^t \sigma^2(s) ds$, where $\sigma(s)$ is the stochastic volatility process in $[0, t]$. In general, any random $\theta(t)$ leads to a leptokurtic $X(t)$ through its second moments. The following theorem facilitates the study of martingales by means of compound processes:

Theorem (Dambis-Dubins-Schwartz) Let $X(t)$ be a *continuous-time martingale* with associated QV process $[X](t)$ satisfying $\lim_{t \rightarrow \infty} [X](t) = \infty$. Define, for each s in $[0, T]$, the *time function* $\theta(s) = \inf\{t \geq 0; [X](t) > s\}$. The *compound process*

$$B(s) \equiv X[\theta(s)], \quad \mathcal{G}(s) \equiv \mathcal{F}[\theta(s)],$$

is a *Brownian Motion* with filtration $\mathcal{G}(s)$. Specifically, the filtration $\mathcal{G}(s)$ is

1.2 Theoretical Background

right-continuous, and $\mathcal{G}(0)$ contains all the negligible events of \mathcal{F} . It follows that $X(t) = B[[X](t)]$, for t in $[0, T]$.¹⁵

In simple terms, the continuous-time martingale $X(t)$ can be rewritten as a Brownian-based compound process with time-change equal to $[X](t)$. Many of the relevant statistical properties of a martingale can therefore be read from the even moments of $B[[X](t)]$. This feature will become handy in Chapter 3 when addressing the estimation of multifractal processes in continuous time.

So far, we have mainly revisited processes in their continuous-time formulation. A significant body of literature concentrates also on their discrete-time counterparts. Essential for the coherence between the two views is the distribution the price process has, as the time-step Δt approaches zero. In this regard, Mandelbrot^[11] has been the first to claim that the time scale selected to analyze a time series should not modify the statistical properties of the process itself. This notion can be captured in the following definition:

Definition (Self-similar process) A random process $X(t)$ that satisfies

$$\{X(ct_1), \dots, X(ct_k)\} \stackrel{d}{=} \{c^H X(t_1), \dots, c^H X(t_k)\}$$

for some $H > 0$ and all $c, k, t_1, \dots, t_k \geq 0$, is called *self-similar* or *self-affine*. The number H is the *self-similarity index*, or *scaling exponent*, of the process $X(t)$.

The Brownian motion is self-similar with $H = \frac{1}{2}$. Besides the Brownian motion, the fractional Brownian motion and the Lévy stable process are the most common examples of self-similar processes in the financial literature (cf. Calvet and

¹⁵ We refer the reader to Karatzas and Shreve^[54] for a proof of the theorem.

Fisher^[60]).

Self-similarity is a intuitively appealing property when studying asset return distributions, as it leads to consider an universal probability law applicable to all time scales. In practical terms, it implies that a process $X(t)$ in the interval $[0, T]$ should exhibit the same properties as in the stretched interval $[0, 1]$, say. As it turned out though, financial times series, like several other time series arising from physics, astronomy, biology and medicine, are not really self-similar but resemble the aggregational normality feature; they have thinner tails and become less peaked in the bells when the sampling interval decreases (cf. Campbell *et al.*^[15], Sorriso-Valvo *et al.*^[69], Kiyono *et al.*^[70,71]). It is therefore clear that a more flexible relationship is necessary to account for such a phenomenon.

The multifractal literature meets its point of departure in the natural sciences when Mandelbrot^[55] first proposed as a generalization of the self-similarity paradigm a probabilistic approach for the distribution of energy dissipation in intermittent turbulence. Multifractality procures a solution to the aforementioned empirical violations by replacing c^H in the definition of self-similarity by a positive (random) factor M_c . Early implementations of multifractality, on the other hand, have been based on the concepts of mass conservation and random measures in the \mathbb{R}^d space (cf. Mandelbrot^[55]). These have been extended later on to the notion of multifractal processes (cf. Mandelbrot^[56]), with successful applications in a large number of fields (cf. Sorriso *et al.*^[69], Carius and Ingelman^[72], Kiyono *et al.*^[70,71], Bullmore *et al.*^[73], Sinha-Ray *et al.*^[74], and Baptista *et al.*^[75] among many others).¹⁶ A short digression on these concepts will bring some light into the benefits of these and subsequent constructions.

¹⁶ Cf. Mandelbrot^[57] for a classic exposition of fractals and their early applications.

Definition (Multifractal measure) *A random measure μ defined on $[0,1]$ is called multifractal if it satisfies for all $q \in \mathcal{Q}$:*

$$\mathbb{E} [\mu[t, t + \Delta t]^q] \sim c(q)(\Delta t)^{\tau(q)+1} \text{ as } \Delta t \rightarrow 0,$$

where \mathcal{Q} is an interval containing $[0,1]$, $\tau(q)$ and $c(q)$ are deterministic functions defined on \mathcal{Q} , and the operator \sim implies that if $h(\Delta t) \sim g(\Delta t)$, the two functions h and g satisfy $h(\Delta t)/g(\Delta t) \rightarrow 1$.

The first generation of models known for satisfying this definition are hierarchical ‘cascades’ of combinatorial nature. The construction of these cascades starts by assigning a random measure μ_0 to a bounded interval $[0, 1]$, say, and consequently employing an iterative transformation to it. In the first step, μ_0 is divided into two subintervals receiving positive constants m_0 and m_1 , respectively. In this simple version, the constants may be chosen to obey $m_1 = 1 - m_0$, with $0 \leq m_0 \leq 1$. The resulting measure is commonly referred to as μ_1 , which no longer is uniform in $[0, 1]$ but rather has a step-function shape; the left interval has a height of m_0 , while the right one has a height of $1 - m_0$.¹⁷ In the next cascade step, the two intervals of μ_1 are split up again into two subintervals receiving factors m_0 and $1 - m_0$, assigned from left to right. This leads to the measure μ_2 consisting of four intervals, each with its own probability mass: $(m_0)^2$, $(m_0)(1 - m_0)$, $(1 - m_0)(m_0)$, $(1 - m_0)^2$. The procedure can be repeated *ad infinitum* leading μ_∞ to weakly converges to the measure μ .

One should note that these transformations never alter the total mass of μ_0 , they

¹⁷ In the simplest case, both intervals have the same length, that is, μ_0 is divided into $[0, 1/2]$ and $[1/2, 1]$, though other alternatives are possible as well.

1.2 Theoretical Background

only spread it off by the factors m_0 and $1 - m_0$ along the original support. One speaks in this case of a *conservative* measure, given that the original mass is preserved at each iteration step. It follows that in any interval of size $\Delta t = 2^{-k}$, the probability mass amounts to

$$\mu_k[t, t + \Delta t] = m_0^{k v_0} (1 - m_0)^{k v_1},$$

where $t = \sum_{i=1}^k \eta_i 2^{-i}$, $\eta_1, \dots, \eta_k \in \{0, 1\}$, and v_0 and v_1 denote the relative frequency of 0s and 1s in the series (η_1, \dots, η_k) (cf. Mandelbrot *et al.*^[76]). Also, the nature of the factor assignment to each interval of size 2^{-k} has led μ to be referred to as the Binomial Multifractal measure.

Many variations of this hierarchical procedure exist. One could think of generating $b \geq 2$ subintervals per iteration, for instance. Subintervals indexed from left to right by $\beta \in \{0, \dots, b - 1\}$ receive mass m_0, m_1, \dots, m_{b-1} and conserve mass by requiring $\sum_{\beta=0}^{b-1} m_\beta = 1$. Instead of always assigning the fraction m_0 to the left descendent, one could alternatively randomize this transformation by making a draw from a random variable M_β that takes values m_0, m_1, \dots, m_{b-1} with probabilities p_0, p_1, \dots, p_{b-1} , or for that matter, from a more general random variable $M \geq 0$. A popular example of this generalization is the Lognormal Multifractal model, where M obeys a Lognormal distribution (cf. Mandelbrot^[56], Mandelbrot *et al.*^[59], and Calvet and Fisher^[77] among others). Note that these factors remain independent across the cascade steps. Mass conservation within a cascade step, on the other hand, is fulfilled by setting a constraint in the joint distribution of $\{M_0, \dots, M_{b-1}\}$ so that $\sum_{\beta=0}^{b-1} M_\beta = 1$.

Figure 1.2 illustrates the generation of a *conservative* measure with $b = 2$ and

1.2 Theoretical Background

random variables $M \in \{m_0, 1 - m_0\}$ with $m_0 = 0.7$. Three iteration levels are shown: μ_1 , μ_2 , and μ_{12} , together with a sample of daily squared returns used as a proxy for the volatility series of the Japanese Yen (YEN). One can perceive that the resemblance of μ_{12} to the empirical time series of volatility is perplexing.

A further extension in the construction conceives the independence of $\{M_0, M_1, \dots, M_{b-1}\}$ within each cascade step. M_β may be allowed in this case to have the same distribution law as the random variable $M \in \{m_0, 2 - m_0\}$, or again, more generally, of $M \geq 0$. Note that when the independence of $\{M_0, M_1, \dots, M_{b-1}\}$ is considered, mass concentration can only be assured on average, that is, provided $M_\beta \stackrel{d}{=} M \quad \forall \beta: \mathbb{E} \left[\sum_{\beta=0}^{b-1} M_\beta \right] = 1 \Leftrightarrow \mathbb{E}[M] = 1/b$. One refers in this case to a *canonical* measure.

Let us verify now how the last two constructions satisfy the definition of multifractality. We note that at each starting point $t = \sum_{i=1}^k \eta_i b^{-i}$, with $\eta_1, \dots, \eta_k \in \{0, \dots, b-1\}$, the *conservative* measure μ_k in the interval $\Delta t = b^{-k}$ is $\mu(\Delta t) = M_{\eta_1} \cdot M_{\eta_1, \eta_2} \cdots M_{\eta_1, \eta_2, \dots, \eta_k}$. The exact conservation of mass and the independence of multipliers across cascade steps lead furthermore to $\mathbb{E}[\mu(\Delta t)^q] = [\mathbb{E}(M^q)]^k$, or respectively to

$$\mathbb{E}[\mu(\Delta t)^q] = (\Delta t)^{\tau(q)+1},$$

with $\tau(q) = -\log_b \mathbb{E}(M^q) - 1$, and where $k \rightarrow \infty$ implies that $\Delta t \rightarrow 0$ (cf. Mandelbrot *et al.*^[59], Calvet and Fisher^[77]).

In the *canonical* case, that is, when mass conservation is satisfied only on average, it is convenient to start from $\mu[0, 1] = \Omega \geq 0$ being a random variable. This leads to $\mu(\Delta t) = M_{\eta_1} \cdot M_{\eta_1, \eta_2} \cdots M_{\eta_1, \eta_2, \dots, \eta_k} \cdot \Omega_{\eta_1, \eta_2, \dots, \eta_k}$, where $\Omega_{\eta_1, \eta_2, \dots, \eta_k}$ has the same distribution as Ω . Applying this into the definition of multifractal measures leads

to

$$\mathbb{E} [\mu(\Delta t)^q] = \mathbb{E} [\Omega^q] (\Delta t)^{\tau(q)+1},$$

with $c(q) = \mathbb{E} [\Omega^q]$ being a deterministic function of q and $\tau(q)$ set as before.

The notion of multifractality can be extended to multifractal processes:

Definition (Multifractal process) *A stochastic process $X(t)$ is called multifractal if it has stationary increments $\delta_{\Delta t} X(t) \equiv X(t + \Delta t) - X(t)$ that satisfy the moment scaling rule*

$$\mathbb{E} [|\delta_{\Delta t} X(t)|^q] \sim c_X(q) (\Delta t)^{\tau_X(q)+1} \text{ as } \Delta t \rightarrow 0.$$

The requisite of stationary increments implies that the probability law of $\delta_{\Delta t} X(t)$ does not depend on t . Regarding the scaling function $\tau_X(q)$, we note that by setting $q = 0$, all $\tau_X(q)$ have the same intercept $\tau_X(0) = -1$. Furthermore, a self-similar process has a linear $\tau_X(q)$. This can be verified from the invariance condition $X(t) \stackrel{d}{=} t^H X(1)$ in the definition of self-similarity, which implies that $\mathbb{E} [|X(t)|^q] = t^{Hq} \mathbb{E} [|X(1)|^q]$. By plugging-in this into the latter definition, one obtains $\tau_X(q) = H \cdot q - 1$, a linear function of q .

Like multifractal measures, multifractal processes are characterized by a non-linear scaling function $\tau_X(q)$.¹⁸ Due to its capacity to explain various physical phenomena, ‘multiscaling’ (or the non-linear functional relationship between the probability laws of price changes $\delta_{\Delta t} X(t)$ at different time scales Δt) has been the main property of multifractal processes studied in the natural sciences. While successful there, initial multifractal constructions were not widely embraced in

¹⁸ Multifractal models have in fact a strictly concave $\tau_X(q)$.

the financial literature. The main reason for this lies on the inherited focus on random measures for the generation of the process, while its combinatorial nature presented difficulties at the time of incorporating these concepts into the available financial literature. A number of events contributed to turn the situation around. On the one hand, Mandelbrot *et al.*^[59] introduced the Multifractal Model of Asset Returns (MMAR). Though still of combinatorial nature, the MMAR represented a bridge between the multifractal perspective of fluctuations and the financial concept of trading time, which helped shifting the focus from the analysis of multiscaling to the study of volatility (cf. Calvet *et al.*^[60], and Clark^[24]). On the other hand, a number of articles arising from the econophysics field have helped to raise acceptance of this new framework by pointing out the similarities between the distribution of asset returns and of the fluid turbulence (cf. Demos *et al.*^[78], Ghashghaie *et al.*^[79], Arneodo *et al.*^[80], Muzy *et al.*^[81], Kozuki and Fuchikami^[82], Ausloos and Ivanova^[83], Kiyono *et al.*^[6,84], and Nawroth *et al.*^[85] among others). Due to the importance of the analogy between the volatility of asset returns and the velocity differences of two points in a turbulent flow, I dedicate Chapter 2 of this thesis to its analysis. In the chapter, I further show that despite of its combinatorial nature, the hierarchical cascade construction suggested there (and alternative to the MMAR) represents a valid method for the modeling of asset return variations.

Turning back to the MMAR, the construction is based upon the more general proposition (cf. Calvet and Fisher^[60]):

Proposition (Multifractality of a log-price process) Given conditions 1-3 below, the log-price $X(t) \equiv \ln[Y(t)] - \ln[Y(0)]$ is a *multifractal process* with

scaling function $\tau_X(q) \equiv \tau_\theta(q/2)$.

Condition 1 $X(t)$ is a compound process, where $B(t)$ is a Brownian motion, and $\theta(t)$ is a stochastic time deformation.

Condition 2 The time deformation $\theta(t)$ is the Cumulative Distribution Function (CDF) of a multifractal measure μ defined on $[0, T]$.¹⁹

Condition 3 The processes $B(t)$ and $\theta(t)$ are independent.

Clearly, one can verify that under these conditions

$$\mathbb{E}[|X(t)|^q] = \mathbb{E}[\theta(t)^{q/2}] \mathbb{E}[|B(1)|^q].$$

Being a CDF function of the multifractal measure $\mu[0, T]$, the trading time θ satisfies the relation $\mathbb{E}[\theta(t)^q] \sim c_\theta(q) t^{\tau_\theta(q)+1}$. The moments of the process $X(t)$ therefore scale as a power function of the frequency of observation:

$$\mathbb{E}[|X(t)|^q] \sim c_X(q) t^{\tau_X(q)+1} \text{ as } t \rightarrow 0,$$

where

$$\tau_X(q) = \tau_\theta(q/2) \text{ and } c_X(q) = c_\theta(q/2) \mathbb{E}[|B(1)|^q].$$

As a parallel to the early implementations in the financial literature of Clark's^[24] MDH, this proposition implies that the increment $\delta_{\Delta t}\theta(t)$ works as a volatility factor of the log-prices differences $\delta_{\Delta t}X(t)$, while $\delta_{\Delta t}B(t)$ determines the direction of the price movement. In short, the MMAR achieves the goal of translating the intermittent burst of activity generated by the multifractal construction into the notion of volatility.

¹⁹ The *canonical* mass conservation rule translates here into $\mathbb{E}[\theta(T)] = T$.

1.2 Theoretical Background

Several empirical studies have focused on verifying the goodness of this new approach. Among the early applications can be listed, for instance, Calvet and Fisher^[77,86,87]. Further studies can be found in Fillool^[88], Yalamova^[89], Eisler and J. Kertész^[90], Pantanella and Pianese^[91], Chen *et al.*^[92], among others.

The main drawback of the MMAR resides in the measure μ being defined on a fixed grid-bound interval (cf. Lux^[8]). The resulting non-stationarity of the model still presents serious obstacles for the application of estimation and forecasting tools. As a remedy to this deficiency, Calvet and Fisher^[1] proposed the Poisson Multifractal Model (PMM), based upon a grid-free multifractal measure μ' . This second generation class of model overcomes the major weaknesses of the MMAR by randomizing the construction of subintervals at each step k with a sequence of Poisson arrival times $\{T_{k,i}\}_{i=1}^n$, while preserving the hierarchical definition of mass subintervals in the cascade. The PMM exhibits also strictly stationary increments, which in connection with the Poisson arrival times greatly facilitates the implementation of estimation and forecasting tools. Given the almost contemporaneous rise of the Markov Switching Multifractal (MSM) model of Calvet and Fisher^[7] though, little work has been done with the PMM.

The MSM model builds upon the PMM, by relaxing the hierarchical definition of subintervals $T_{k,i}$. Consequently, the sequence of arrival times $\{T_{k+1,i}\}_{i=1}^n$ at level $k+1$ now extends in parallel to the sequences at other levels. As pointed out by Calvet and Fisher^[7], this independence of the switches of the factors M at level $k+1$ from those occurring at other levels provides the MSM model with better empirical features than the PMM, while preserving the moment scaling properties.²⁰ All in all, the MSM model is able to replicate the majority of the

²⁰ The MSM model preserves actually the scaling properties over a finite horizon only (cf.

stylized facts, that is, absence of autocorrelations, aggregational normality, zero excess skewness and leptokurtosis of raw returns, volatility clustering, jumps and long memory of volatility, as well as a slow decay of autocorrelation in absolute returns (cf. Zhong and Zhao^[93]). Provided though that the model does not obey the traditional definition of a hierarchical cascade structure used for the elaboration of multifractal measures, one may turn to a broader specification of **Condition 2** above to include this and other potential constructions within the framework:

Condition 2' *The time deformation $\theta(t)$ is any increasing function satisfying the definition of multiscaling.*

Calvet and Fisher^[7] present the proof of how the PMM and the MSM model satisfy this new condition. We skip therefore this step here and turn to the discussion of estimation and forecasting applications directly. We proceed with the following definition:

Definition (Multifractal asset returns) Given a *multifractal process* $X(t) \equiv \ln [Y(t)] - \ln [Y(0)]$ with *increments*

$$\delta_{\Delta t} X(t) \equiv X(t + \Delta t) - X(t) = B[\theta(t + \Delta t)] - B[\theta(t)],$$

the *returns* $x_{t+\Delta t}$ are *distributed* as

$$x_{t+\Delta t} \stackrel{d}{=} [\delta_{\Delta t} \theta(t)]^{1/2} \cdot \delta_{\Delta t} B(t).$$

Calvet and Fisher^[7].

1.2 Theoretical Background

In a nutshell, the return $x_{t+\Delta t}$ of a multifractal process $X(t)$ can be modeled as the product of a Gaussian random variable $\delta_{\Delta t} B(t) \equiv \varepsilon_{t+\Delta t} \stackrel{d}{=} N(0, \sqrt{\Delta t})$ and the ‘volatility’ of the process in that time interval, measured as the square root of the marginal increase in the subordinator, or $\sigma(M_{t+\Delta t}) \equiv [\delta_{\Delta t} \theta(t)]^{1/2}$. In this case, $\sigma(\cdot)$ denotes a function of $M_{t+\Delta t}$, the vector of positive random variables M of dimension equal to the number of cascade steps.²¹

Note also that when each volatility component M is Bernoulli distributed, i.e. when $M \in \{m_0, m_1\}$, only a finite number of volatility states emerge. This fact has led Calvet and Fisher^[7] to introduce in the context of the MSM models a Maximum Likelihood (ML) estimation procedure based upon Hamilton’s^[94,95] version of the expectation maximization (EM) algorithm. Forecasting of the return fluctuations becomes available, in this case, by a Bayesian update of the volatility states. The authors further show that this model outperforms others like GARCH, FIGARCH and MS-GARCH, in- and out-of-sample when analyzing foreign-exchange returns. Lux^[8] complements the previous work on daily return series by focusing on those scenarios where either due to the large number of cascade steps or due to the continuous-space nature of the underlying distribution of M , a Markov chain update is no longer possible. His methodology is based upon the Generalized Method of Moments (GMM) and allows, among other things, the choice between Bernoulli- and Lognormally-distributed volatility components M . By means of a best-linear forecasting (BLF) algorithm, the flexibility reflected in the estimation approach is also present in the prediction of future fluctuations. One should notice nonetheless that BLF is suboptimal with respect to predictions

²¹ Mass conservation can be translated here to a point-wise metric $\mathbb{E}[\sigma^2(M_{t+\Delta t})] = 1$, around which $\sigma(M_{t+\Delta t})$ oscillates. This phenomenon is usually referred to as the volatility clustering property of multifractal models.

1.2 Theoretical Background

based on Bayesian updating. Empirical evaluations confirm that this prediction-loss is minimal and that the advantage of MSM over the GARCH, FIGARCH and MS-GARCH models is not restricted to the Bernoulli case. Further studies based upon the moments of univariate multifractal time series include Lux and Kaizoji^[96], Lux and Morales-Arias^[97,98], and Lux *et al.*^[9].

Most of the applications of MSM models depicted so far involve daily-return time series only. Provided though that the access to high-frequency financial asset prices allows nowadays the ‘observation’ of a continuous-time process with increasing precision, the building-blocks of the continuous-time MSM model may resemble reality more accurately. The early evaluation of this premise has prompted, for instance, empirical studies on the scaling properties of the continuous-time MSM process (cf. Calvet and Fisher^[60,77]).²² When studying intraday fluctuations as a proxy for continuous-time intermittency, however, there exists a lower bound on the sampling interval that can be instrumentally used for computation. This results from limited finite sampling or from the existence of market microstructure noise (cf. Andersen *et al.*^[32]). Chapter 3 of this thesis tackles these issues when addressing the estimation of the MSM model in continuous time. The chapter proposes a Simulated Method of Moments (SMM) approach in which the moments derived from the continuous-time process are approximated to obtain an asymptotically efficient estimator as $\Delta t \rightarrow 0$. The approach allows the use of a sample series of raw returns or, alternatively, of the QV process. For this, two alternative sets of moments are employed depending on the information entering the estimation procedure. The first set focuses on the evolution of the continuous-

²² Cf. Di Matteo *et al.*^[99] for an analysis of the scaling behavior of the discrete-time MSM model in terms of daily returns, comparable to that of Calvet and Fisher^[77] for the continuous-time MSM process.

1.2 Theoretical Background

time returns and is applicable when only a sample of daily prices is available to the researcher or, alternatively, when a measurement of daily returns is generated via record aggregation. The second set targets more directly the latent volatility process. These moments consist of functions of the QV associated to the return process and are applicable when a high-frequency metric like RV can be obtained from a sample of intra-daily returns. The latter direction is based on the previous work of Lux *et al.*^[9] where a modification of the log-moments of Lux^[8] allows to handle return fluctuations directly as an observable variable.

The study of MSM models could alternatively focus on the sample of option-price time series or on the combined set of both historical return and option-price time series. A closer look at the *stylized facts* of daily return and option-price time series highlights the potential benefits of employing these two datasets *simultaneously*. Calvet *et al.*^[100] have made use of this strategy for the in-sample analysis of their most recent version of the MSM model. The authors obtain one set of estimates by updating the volatility states within the combined sample of returns and option prices for the construction of their joint density. Concretely, the approach consist in combining a fast numerical method for the calculation of option prices together with a particle filter for the update of the state variables M_t in the historical returns. The estimation is performed by ML, providing of the usual asymptotic properties. Results show that the approach of Calvet *et al.*^[100] delivers significantly higher in-sample likelihood than the benchmarks consisting of ARCH-type, SV, and pure-jump models.

When using standalone option prices on the other hand, one issue has received a considerable amount of attention since early studies of volatility. This refers to the degree of significance the term structure and the shape of the implied volatil-

1.2 Theoretical Background

ity surface possess for the forecasting of return fluctuation.²³ In this case, the advantage of complementing option-price time series with historical financial data resides in the possibility to combine forecasts generated from each independent source of information. Chapter 4 of this thesis commits to these issues in the context of MSM models. The chapter departs from a discrete-time model including risk preferences of an idealized investor and suggest an equivalent ‘risk-neutral’ formulation under which the calibration to option prices is performed. For the empirical analysis, a highly liquid stock index like the German DAX is selected, for which the necessary amount of historical return and option-price time series is available. The information collected from the calibration methodology is first used to assess the ability of the ‘risk-neutral’ MSM process to capture information regarding future volatility contained in option price data. Later on, QV forecasts are generated for both the statistical and the risk-neutral processes and compared to the observed RV, derived from transaction prices of the DAX over 5-minute intervals.

Current extensions to the MSM model seek to increase the applicability of the framework. Calvet *et al.*^[107], Liu and Lux^[108], and Idier^[109], for instance, have posed extensions to the multivariate case. Baruník *et al.*^[110] and Chen *et al.*^[111], on the other hand, employ the MSM model for the analysis of intertrade durations, providing the first application in the market microstructure literature. Lux and Morales-Arias^[97] have tested the impact of Student-t distributed returns innovations on volatility forecasting. Further extensions have sought the replication of the few remaining *stylized facts* not contemplated by the original construction,

²³ Some strategies can be found in Jorion^[101], Bakshi *et al.*^[102], Christensen and Prabhala^[103], Christensen and Hansen^[104], Fleming^[105], and Blair *et al.*^[106], among others.

1.2 Theoretical Background

which due to the complexity of the model may require arduous reformulations. Note, for instance, that the symmetry of returns, as well as the symmetry of volatility responses to return innovations, is a feature of MSM models by construction. This arises as a consequence of **Condition 3** introduced in the last proposition. It seems also to be shared by most types of multifractal models (cf. Zhong and Zhao^[93]), signaling how deep-rooted symmetry is in the construction. A modification of this property for the inclusion of ‘leverage’ effects represents so a great challenge that cannot be easily undertaken without altering other highly appreciated properties of the model. Because of this, Calvet and Fisher^[112] suggest to directly avoid the issue of this connection to turn, alternatively, to a contemporaneous link between volatility and prices by using equilibrium valuation methods. Their approach, however, is intimately connected to dividend-paying stocks and requires the acceptance of equilibrium-based valuation relationships that are otherwise difficult to extrapolate to other type of assets without further adjustments. Calvet and Fisher^[113] develops the previous idea further for the conception of an option pricing model in continuous time. In this framework, market pricing generates a contemporaneous negative correlation between volatility changes and price jumps leading to a negative skewness of returns. Calvet *et al.*^[100] introduce an alternative version of the previous approach, together with a pricing routine for its efficient estimation.

Appendix B at the end of this thesis suggests an alternative framework for the contemplation of volatility responses to past innovations within the context of discrete time series modeling. The approach consist in relaxing the assumption about the marginal distribution of the volatility shocks when new arrivals occur. Concretely, probability mass from this distribution conditionally on the Gaussian

1.2 Theoretical Background

shock of the previous period is shifted so that past negative Gaussian shocks may increase the contemporaneous probability mass of drawing high volatility shocks. Because the dependence is introduced in the conditional distribution of the volatility shocks only, the variables self remain independent of the contemporaneous information arrivals and Gaussian shocks. As a result, most of the time-honored properties of MSM remain unaltered while the new model gains heavily in flexibility. Appendix B suggests at last an estimation methodology that will be the central focus of future work.

The MSM model is all in all a complex but flexible framework that has been extended along many directions in recent years. The multifractal literature stands nonetheless in its early stage, and as a research field, it provides considerable potential for further developments.

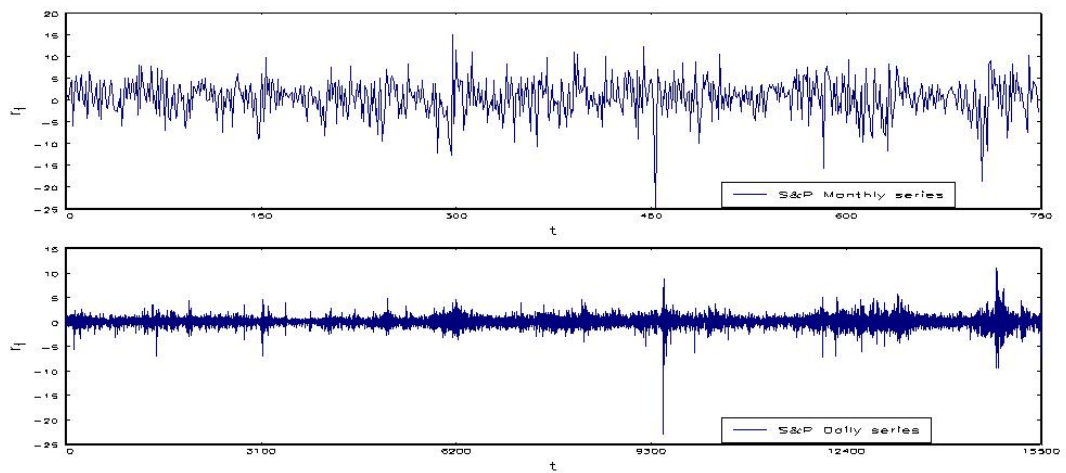


Figure 1.1: Monthly (top panel) and daily (bottom panel) sample return series of the S&P500 index. The series runs from 01/03/1950 until 01/05/2012. In the top panel, the series exhibits a Gaussian-like fluctuations despite the three large negative, and one positive, outliers. The first negative outlier represents the market crash of 1987, while the positive and the second negative ones are not really large single shocks but the accumulation of a series of smaller returns. Though insufficient for the identification of all variability sources, the shift from the monthly to the daily frequency already allows for a better visualization of the fluctuations.

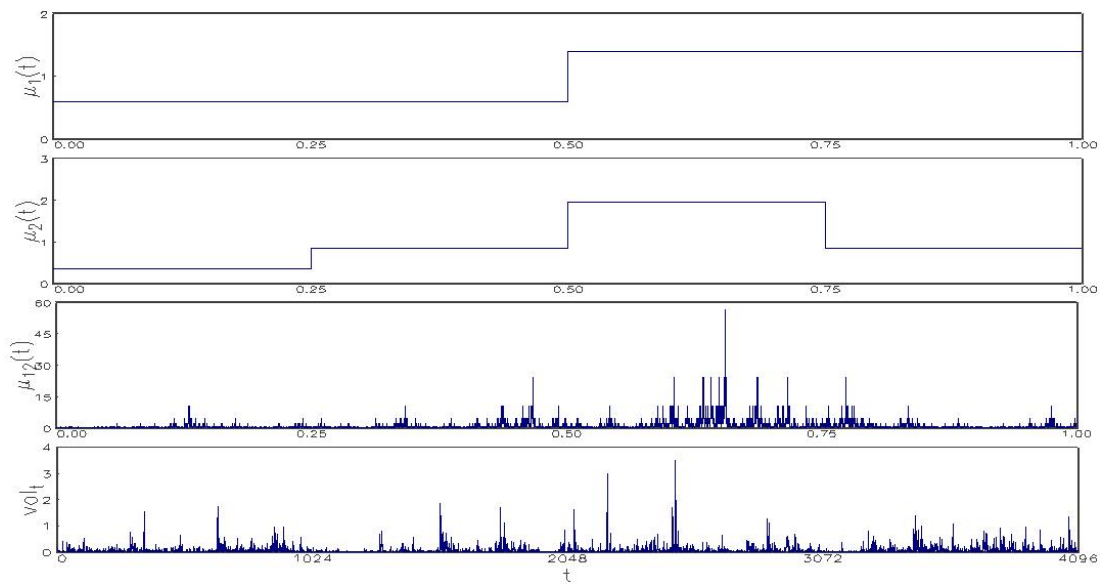


Figure 1.2: Example of the construction of a *conservative* measure with random variables $M \in \{m_0, 1 - m_0\}$ and $m_0 = 0.7$. From top to bottom: a draw of the first level μ_1 , a draw of the second level μ_2 , a draw of the 12th level μ_{12} , and an empirical sample series of squared returns of the Japanese Yen (YEN) as a proxy for volatility. The latter series consists of ≈ 16.25 years of data starting on the 2nd of January of 1979. For better visualization, μ_k was multiplied at each cascade level k by 2^k .

Chapter 2

Lognormal Cascades. Parameter Estimation and Forecasting

2.1 Introduction

The multifractal literature meets its point of departure in the natural sciences as Mandelbrot^[55] first proposed a probabilistic approach for the distribution of energy dissipation in intermittent turbulence. A remarkable property of this phenomenon is the inhomogeneity of its variance, resulting in non-Gaussian probability density functions (PDFs) (cf. Kiyono^[114]). Due to the potential complexity of this setting, the usual approach to describe such non-Gaussian PDFs is by superposition of Gaussian distributions through the diffusion parameter (cf. Castaing *et al.*^[4], and Beck and Cohen^[115]).

In their seminal study, Castaing *et al.*^[4] have suggested the following equation for the characterization of the PDF of velocity differences in fully developed turbulent flows:

$$P_{\lambda, \sigma_0}(x) = \int_0^\infty \frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{\ln^2(\sigma/\sigma_0)}{2\lambda^2}\right) P_\sigma\left(\frac{x}{\sigma}\right) \frac{d\sigma}{\sigma^2}, \quad (2.1)$$

where λ and σ_0 are positive parameters characterizing the PDF of the variable σ , and P_σ is the PDF of a stationary and zero mean random variable x . Both λ and σ_0 determine not only the second moment associated with $P_{\lambda, \sigma_0}(x)$ but also the kurtosis. When $\sigma_0 = 1$ and $\lambda > 0$, $P_{\lambda, \sigma_0}(x)$ represents a mixture of distributions with a variance greater than one and excess kurtosis. In the limit $\lambda \rightarrow 0$, on the other hand, we observe $\sigma \rightarrow \sigma_0$ and $P_{\lambda, \sigma_0}(x)$ becomes a standard, mesokurtic Gaussian distribution. Equation (2.1), therefore, covers a whole spectrum of processes that can be used to describe more complex fluctuations than those originating from a Gaussian source. Stochastic processes corresponding to the PDF in eq. (2.1) could be of the form:

$$x_i = \exp(\xi_i) \varepsilon_i, \quad (2.2)$$

where ε_i and ξ_i are mutually independent and follow Gaussian distributions $\varepsilon_i \stackrel{d}{=} N(0, \sigma_1^2)$ and $\xi_i \stackrel{d}{=} N(\ln \sigma_0, \lambda^2)$.¹ The resulting intermittency generated from processes of the type of eq. (2.2) has been found to approximate quite well the fluctuations observed in data from various fields, such as from hadron collision (cf. Carus and Ingelman^[72]), solar wind (cf. Sorriso-Valvo *et al.*^[69]), as well as human heartbeat fluctuations (cf. Kiyono *et al.*^[70,71]), high-resolution satellite images (cf. Arneodo *et al.*^[116]) and, finally, in data of stock index (cf. Kiyono *et al.*^[84]) and foreign exchange rate fluctuations (cf. Ghashghaie *et al.*^[79], and Calvet and Fisher^[86], among others).

¹ Though not expressed explicitly, σ_1^2 can clearly be accommodated in eq. (2.1) via P_σ .

It is also well-known that the phenomenological approach by Castaing *et al.*^[4] allows for non-linear scaling of absolute moments or multifractality of the underlying data-generating process. Considering a continuous-time process $X(t)$ with increments between times t and $t + \Delta t$: $\delta_{\Delta t} X(t) = X(t + \Delta t) - X(t)$, self-similarity of the associated PDF amounts to:

$$P(\delta_{\Delta t} X) \stackrel{d}{=} s^H P(s^H \delta_{s\Delta t} X), \quad (2.3)$$

with H the pertinent (Hurst) exponent for the renormalization of the PDF under changes of the scale s ($s > 0$). In order to account for multi-scaling in a series, a unique scaling exponent H is not appropriate so that one has to extend the previous approach. As originally suggested by Mandelbrot^[58,117], by replacing the constant factor s^H in eq. (2.3) by a random factor M_s depending on the scale, we obtain:

$$\delta_{s\Delta t} X(st) \stackrel{d}{=} M_s \delta_{\Delta t} X(t). \quad (2.4)$$

It can be shown that such a scale-dependent multiplicative random modulation of $P(\delta_{\Delta t} X)$ leads to a non-linear scaling of absolute moments. The stochastic process of eq. (2.2) is an example of a process characterized by such non-linear scaling and, consequently, the PDF of eq. (2.1) is a potential outcome of such a stochastic extension of the notion of a self-similar process. Considering a cascade scale Δt and a finer scale $s\Delta t$ ($s < 1$) in eq. (2.4), the PDF of eq. (2.1) indeed characterizes their relationship, with the random factor M_s being represented by the Lognormally distributed random variable $\exp(\xi_s)$.

In the tradition of Castaing *et al.*^[4], practical implementations of eq. (2.1) have

mostly resorted to numerical approximations of the shape of the PDF minimizing the χ^2 statistics with respect to the empirical PDF to obtain parameter estimates. To avoid certain problems related to this method, Kiyono *et al.*^[6] suggest an alternative procedure based on $\mathbb{E}[|x|^q]$, the q th order absolute moments. Another moment-based estimator has been proposed by Beck^[5], who uses the standardized fourth moment (or ‘flatness’ of the PDF). In this chapter, we introduce a new alternative estimation procedure based on a Generalized Method of Moments (GMM) framework and demonstrate its superior performance. Our approach is motivated by a similar estimator that has been proposed in Lux^[8] for the causal Markov-Switching Multifractal (MSM) model of Calvet and Fisher^[1]. While our methodology is also based on moment matching, it differs from the approaches of Beck^[5] and Kiyono *et al.*^[6] in two important aspects: First, our moments are computed with respect to the joint distribution of x_i at different points of the cascade and, as such, they are exact moments of the underlying process. In contrast, the moments proposed by Beck and Kiyono *et al.* are computed from the marginal distribution $P_{\lambda, \sigma_0}(x)$ of eq. (2.1). These moments are exact for the multiplicative Lognormal model of eq. (2.2) with independent draws ξ_i but not for a model with added cascade-like structure. Secondly, by using a GMM approach, we use more than one moment condition and systematically exploit therefore the degree of uncertainty in various moments.

For the practical use of parameter estimates, we develop a forecasting scheme based on the best linear forecast algorithm that dispenses with the necessity to work with an approximation to the PDF of the coarse scale Δt process. We finally test its out-of-sample accuracy via Monte Carlo simulations and provide an empirical application. The remainder of this chapter is structured as follows.

Section 2.2 introduces a detailed description of the process. Section 2.3 details the estimation methodology and compares our GMM estimates with previous approaches via Monte Carlo simulations. Section 2.4 shows how the estimator behaves under misspecification concerning the number of cascade levels. Section 2.5 introduces the best linear forecast algorithm, and section 2.6 presents empirical results for both parameter estimation and forecasting for a sample of financial data. Section 2.7 concludes and the appendices collect explicit formulas for the particular moments used in our GMM and best linear forecasting approaches.

2.2 The Process

To illustrate our procedure, we will first concretize the hypothesized data generating process (DGP). Although several ways to simulate intermittent fluctuations exist, we follow here the algorithm of Kiyono *et al.*^[6] for the generation of a cascade with n levels, and consider a fixed grid of 2^n points defining a sequence of uniform time intervals. In the first cascade step, we take the whole discrete set $[1, 2^n]$ and divide it into two sets of the same length. To each subset $[1, 2^{n-1}]$ and $[2^{n-1} + 1, 2^n]$ we uniformly assign a random weight $M_1(k) = \exp[\omega_1(k)]$ ($k = 0, 1$). In the next step, we further divide $[1, 2^{n-1}]$ and $[2^{n-1} + 1, 2^n]$ into two new sets each, and assign in the same fashion the random weights $M_2(k) = \exp[\omega_2(k)]$ ($k = 0, 1, 2, 3$). This procedure is repeated for $j = 1, \dots, n$ leading to the final sequence of products of weights $\prod_{j=1}^n M_j(k)$ attached to the data points $\{1, \dots, 2^n\}$. We obtain the Lognormal cascade as a compound process on the bounded interval $[1, 2^n]$ by multiplying the sum of the

Lognormal weights with a Normally distributed random variable ε :

$$x_i \equiv \left[\prod_{j=1}^n M_j \left(\lfloor \frac{i-1}{2^{n-j}} \rfloor \right) \right] \varepsilon_i = \exp \left[\sum_{j=1}^n \omega_j \left(\lfloor \frac{i-1}{2^{n-j}} \rfloor \right) \right] \varepsilon_i, \quad (2.5)$$

where $\lfloor \cdot \rfloor$ represents the floor function and $\varepsilon_i \stackrel{d}{=} N(0, \sigma_1^2)$. It is common to select $\omega_j(\cdot) \stackrel{d}{=} N(\tilde{\mu}, \tilde{\sigma}^2)$ so that the sum of $\omega_j(\cdot)$ is $N(n\tilde{\mu}, n\tilde{\sigma}^2)$ distributed. Hence, $\sum_{j=1}^n \omega_j(\cdot)$ in eq. (2.5) corresponds to ξ_i in eq. (2.2), and x_i fits into the framework of eq. (2.2) with $\sigma_0 = \exp(n\tilde{\mu})$ and $\lambda^2 = n\tilde{\sigma}^2$. Note, however, that here the ξ_i are not independent draws but are correlated via the cascade structure. In the presentation of their estimator, Kiyono *et al.* assume that $\sigma_1 = 1$ and that $\sigma_0 = \exp(-\lambda^2)$, which in our context would be equivalent to require that $\omega_j(\cdot) \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$ with $\lambda_0^2 = \frac{\lambda^2}{n} = -\tilde{\mu} = \tilde{\sigma}^2$. Figure 4.2 shows an illustration of a $n = 12$ -level cascade with standardized factors $\omega_j(\cdot) \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$. In the top three panels we exhibit draws at the first level $M_1(k) = \exp[\omega_1(k)]$ ($k = 0, 1$), the second level $M_2(k) = \exp[\omega_2(k)]$ ($k = 0, 1, 2, 3$) and the 10th level $M_{10}(k) = \exp[\omega_{10}(k)]$ ($k = 0, \dots, 9$), respectively, while in the fourth panel an outcome of the corresponding ‘time series’ $\{x_i\}_{i=1}^{2^n}$ is displayed. To overcome the statistical difficulties that may arise from such a non-stationary construction, we go one step further and allow for an infinite sequence of independent cascades following the same generative principle, concatenating these series of sequences one after the other. This assumption leads to a sequence of data points $\{\dots, m2^n + 1, m2^n + 2, \dots, (m+1)2^n, (m+1)2^n + 1, (m+1)2^n + 2, \dots\}$, with $m = 0, 1, \dots$ an infinite sequence of repetitions of the same process of generation of a stochastic cascade of length 2^n . Our time series of measurements of

the multiplicative Lognormal cascade process is consequently given by

$$x_t \equiv \exp \left[\sum_{j=1}^n \omega_j^{(m)} \left(\left\lfloor \frac{t-2^n(m-1)-1}{2^{n-j}} \right\rfloor \right) \right] \varepsilon_t, \quad (2.6)$$

where again $\varepsilon_t \stackrel{d}{=} N(0, \sigma_1^2)$.² The ‘multipliers’ $\omega_j^{(m)}(\cdot) = \ln M_j^{(m)}(\cdot)$ are assumed to be new draws for each newly started cascade, so that the process $\{x_t\}_{t=1}^\infty$ does not exhibit any obvious periodic structure, which distinguishes our algorithm from so-called cyclo-stationary processes (e.g. weather signals) that have clearly defined deterministic (e.g. sinusoidal) components (cf. Gardner *et al.*^[118]).³

Note that, the time series in eq. (2.6) can also be described by eq. (2.2), taking into account the particular structure of the conditional distribution for the draws $\omega_j^{(m)}(\cdot)$ (or equivalently, ξ_i) as imposed by the cascade structure. There are two ways to look at our infinite cascade process: First, under knowledge of the actual position, the joint distribution of observations at some time points $\{t_1, \dots, t_k\}$ and $\{t_1 + z, \dots, t_k + z\}$ would clearly be different. This holds independently of whether any sequence would extend beyond the boundary of a single cascade or not. However, under ignorance of the current position, both sequences could be considered to be draws from a stationary process and would, thus, be characterized by the same joint distribution (and, of course, by the same moments). We adopt this second perspective and consider data samples being drawn from this infinite repetition of independent random cascade processes at arbitrary starting points.⁴ The bottom panel of Figure 4.2 shows a sample of 7,500 observations

² As the identification of the repetition number of the cascade is irrelevant for the variable ε , this sequence can simply be indexed by time t .

³ In our case the independent draws of the ‘multipliers’ have an effect that would be similar to reshuffling of the seasons in annual data.

⁴ For this reason, we will drop from this point on the notation of m in $\omega_j^{(m)}(\cdot)$ and we will

of the $\{x_t\}_{t=1}^{\infty}$ process as a result of concatenating three $n = 12$ -level bounded cascades.

Despite the non-standard nature of the $\{x_t\}_{t=1}^{\infty}$ process (i.e. the application of a combinatorial construction in a time series context), our process is stationary under the second perspective (which corresponds to the limited information available to the empirical researcher), and many standard procedures for statistical inference become now available. On the contrary, when considering the original process from eq. (2.5) over a bounded interval only, the non-stationarity of the process would have followed trivially. As a consequence, standard ‘regularity conditions’ (cf. Harris and Mátyás^[119]) for many standard methods of statistical inference would have been violated. As we will see in the following, our approach allows us to compute exact conditional and unconditional moments for our GMM estimation procedure that universally apply to any set of observations arising from the process $\{x_t\}_{t=1}^{\infty}$ of eq. (2.6). Due to the analytical structure of these moments (cf. Appendix 2.A), standard regularity conditions such as differentiability and boundedness of the moments are now clearly satisfied.

2.3 Estimation Methodology

GMM is a very general statistical approach for estimation of the parameters of a model. Given a set of analytical moments, the vector of parameter estimates, say φ , is obtained as the result of the minimization of an objective function of identify the ‘multipliers’ simply by $\omega_j(t)$.

2.3 Estimation Methodology

the following form:

$$\widehat{\varphi}_T = \arg \min_{\varphi \in \Phi} f_T(\varphi)' \Omega_T f_T(\varphi), \quad (7)$$

with Φ being the parameter space, $f_T(\varphi)$ the vector of differences between a set of sample and analytical moments, Ω_T a positive definite and possibly random weighting matrix, and $\widehat{\varphi}_T$ is the solution to this optimization problem, i.e. the argument (arg) that minimizes the objective function $f_T(\varphi)' \Omega_T f_T(\varphi)$, the weighted difference between sampled and analytical moments (cf. Hansen^[120]). Using log-absolute moments in the implementation of $f_T(\varphi)$, Lux^[8] has applied this estimation method to the iterative MSM model, demonstrating that it provides reliable parameter estimates even for small sample sizes. In the following, we will apply a similar approach in our analysis of multiplicative Lognormal cascades.

Let us consider the log-absolute difference $\zeta_{t,\ell} \equiv \ln |x_t| - \ln |x_{t-\ell}|$, with ℓ representing the lag at which the difference is taken. In order to exploit the scaling properties of the cascade process, we select as in Lux^[8] autocovariances of the overlapping log differences $\zeta_{t,\ell}, \zeta_{t+1,\ell}, \dots$. A closer look at these yields:

$$\begin{aligned} \zeta_{t,\ell} &\equiv \ln |x_t| - \ln |x_{t-\ell}| \\ &= \sum_{j=1}^{\ell} [\omega_j(t) - \omega_j(t - \ell)] + \ln |\varepsilon_t| - \ln |\varepsilon_{t-\ell}|. \end{aligned} \quad (2.8)$$

As one can see, these log-absolute differences remain unaffected by σ_1 , the scale factor in eq. (2.6) that is typically needed to match the order of magnitude of the data under scrutiny. Our moment conditions will consist, for $p = 1, 2$, of the following set of moments:

$$\mathfrak{Mom}(\ell, p) = \mathbb{E}[\zeta_{t+\ell,\ell}^p \zeta_{t,\ell}^p], \quad (2.9)$$

together with a raw moment like $\mathbb{E}[x_t^2] = \sigma_1^2$ for the identification of σ_1 (note that we have standardized the distribution of $\omega_j(\cdot)$ in a way to guarantee that the second moment of the first term on the right-hand side of eq. (2.6) is equal to unity; cf. also Appendix 2.B). With this device, the resulting estimates of σ_1 from GMM are identical to the sample standard deviation of the $\{x_t\}_{t=1}^\infty$ process and the covariance matrix between both sets of parameters would be block diagonal. Appendix 2.A contains the explicit derivations for the moments introduced in eq. (2.9).

The estimator $\widehat{\lambda}_q^2$ of Kiyono *et al.*^[6] eq.(5) is derived from $\mathbb{E}[|x|^q]$, the absolute moment of power q for the marginal PDF of eq. (2.1):

$$\widehat{\lambda}_q^2 = \frac{2}{q(q-2)} \left[\ln \left(\frac{\sqrt{\pi} \mathbb{E}[|x|^q]}{2^{q/2}} \right) - \ln \Gamma \left(\frac{q+1}{2} \right) \right] \quad (10)$$

(where $q \neq 0, 2$) after standardizing the mentioned PDF by setting $\sigma_0 = \exp(-\lambda^2)$ in eq. (2.2). Note that eq. (2.1) is not the PDF of the ensemble of observations from a cascade process as it applies strictly only for independent draws of ξ_i in eq. (2.2). Given the stage of the cascade, it, however, characterizes the marginal PDF of the process at any position t . Since $\widehat{\lambda}_q^2$ is not derived from the exact PDF of the cascade process, it will in all likelihood be an inconsistent estimator for such a model. The same applies to the traditional $\widehat{\chi}^2$ estimator of Castaing *et al.*^[4] and the flatness estimator \widehat{F}_2 of Beck^[5]. As we will see, this conjecture is confirmed by our Monte Carlo simulations below. In our cascade-setting, the mentioned standardization implies, on the other side, that $\omega_j(t) \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$ in eq. (2.6), and so as stated before, $\widehat{\lambda}_q^2$ captures the overall intermittency $\lambda^2 = n \lambda_0^2$. In practice, $\mathbb{E}[|x|^q]$ is calculated from a zero-mean unit-variance series so that be-

fore being able to compute this moment, the series $\{x_t\}_{t=1}^T$ must be detrended and consequently standardized by the ad-hoc sample standard deviation estimator $\widehat{\sigma}_1$. The value of q is arbitrary *a priori*, but as the authors suggest, one can numerically compare the root mean squared errors (RMSE) of $\widehat{\lambda}_q^2$ under different q and select the optimal one.

We proceed by reporting results of several Monte Carlo studies designed to explore the applicability of our GMM estimator and its performance in comparison to the aforementioned estimators. To this end, we first apply Kiyono *et al.*'s^[6] standardization for the generation of the data. In the following, we choose $q = 0.5$ for the $\widehat{\lambda}_q^2$ estimator and consider 20 bins with equal probability mass for the $\widehat{\chi}^2$ estimator; that is, the PDF of eq. (2.1) is binned using a varying split of the support $[x_{j-1}, x_j]$, for $j = 1, \dots, 20$, so that each bin contains a probability mass of $CDF_{\lambda, \sigma_0}(x_j) - CDF_{\lambda, \sigma_0}(x_{j-1}) \approx 0.05$.⁵

We apply both estimators for various sample lengths T_i , namely $T_1 = 2, 500$, $T_2 = 5, 000$, and $T_3 = 10, 000$. The GMM procedure aims at exploiting the intermittency at different cascade levels, and therefore, the moments in eq. (2.9) depend on the choice of the number and values of lags ℓ . After many trials, for which results are not presented here, we found that using three lags leads to a good compromise between computational speed and quality of the estimates. In short, the values $\ell = 1, 14, 64$ are chosen to capture the intermittency generated by the last seven cascade levels.⁶ We use the iterative GMM version instead of the simple two-step GMM scheme, where a new weighting matrix Ω_T is computed and the whole estimation process is repeated until convergence of both the parameter

⁵ Results for alternative choices other than the ones presented in the tables are available in Appendix A at the end of this thesis.

⁶ The value $\ell = 14$ is as good as any in (8, 16).

2.3 Estimation Methodology

estimates and the weighting matrix is obtained (cf. Hansen *et al.*^[121]).

Table 2.1 shows the results from our GMM approach, while Table 2.2 presents the outcomes from the older methods, where we have normalized the results in Table 2.2 by the total number of cascade levels n for better comparability. As we can infer from both Tables, all estimators start out very similarly with a slight advantage of the moment-based estimators of Kiyono *et al.* and Beck in terms of RMSE at relatively small parameter values λ_0^2 or low cascade levels n . This appears plausible as these scenarios are closest to the case of independent ξ_i for which the latter would be a consistent estimator. For a fixed number of cascade levels n , however, the bias of the raw moment-based estimators increases considerably the higher λ_0^2 gets. Interestingly, the inconsistent $\widehat{\chi}^2$ estimator still performs surprisingly well and even marginally better than our GMM estimator for $\lambda_0^2 = 0.15$ and $n = 8$. This indicates that with a limited number of cascade steps, the resulting PDF is still not too different from that of eq. (2.1). However, the limitations of this approach become apparent when looking at the case $n = 16$, which shows sizable biases as compared to those from the GMM approach. Note that the relatively good performance of the $\widehat{\chi}^2$ estimator is in contrast to results displayed in Kiyono *et al.*^[6]. Experiments with different settings indicate that the relatively favorable results of the $\widehat{\chi}^2$ estimator in Table (2.2) are due to our use of bins with equal probability mass, while other bin structures would typically give worse outcomes.

Equivalently, for a fixed λ_0^2 , the bias of the moment-based estimators increases with a higher number of cascade levels n . GMM, on the other hand, shows only very slight increases of RMSE when either λ_0^2 or n increases so that its advantage becomes more and more pronounced for high levels of intermittency and high

number of cascade levels. The decrease of biases and sampling variability for the moment-based and $\widehat{\chi}^2$ estimators appears also much slower with increasing sample sizes than for the GMM estimator at high levels n . While the latter seems to nicely satisfy squared-root consistency when doubling the sample size from T_1 to T_2 and from T_2 to T_3 , the former almost never do so. For high λ_0^2 or n RMSEs of the traditional methods appear almost constant across sample sizes.

Table 2.1 and 2.2 also present the results for σ_1 . In Table 2.2, only one sample standard deviation for the moment-based estimators is presented given that these only provide an estimate of λ_0^2 . One observes that the GMM estimator in Table 2.1 agrees closely with the sample standard deviation in Table 2.2, which actually is to be expected under the block-diagonal structure of the covariance matrix of the moment conditions. In all cases, the estimator seems to be more biased the higher both λ_0^2 and n , the number of cascade levels. As in Lux^[8], this might be due to the fact that a higher λ_0^2 and n generate enhanced fluctuations of the product of volatilities, which might interfere with the estimation of the constant scale factor σ_1 . Somewhat surprisingly, Table 2.1 shows that in the case $\lambda_0^2 = 0.15$ and $n = 16$, an apparent violation of the square-root consistency can be perceived for σ_1 when doubling the sample size T . Additional analysis undertaken with larger sample sets, left aside in this chapter for brevity, indicate that this behavior is only restricted to sample sizes T_1 , T_2 , and T_3 , and we recover a ‘nice’ behavior for increasingly larger samples. The reason seems to be that: initially, for small sample sizes (T_1) in relation to the sample size of a bounded cascade with $n = 16$, the probability of encountering a major node at which many switches occur will be low so that the behavior of moments is quite regular. However, with medium sample sizes like T_3 , the probability of meeting a node

2.4 Uncertainty of the number of cascade components

with many switches becomes much larger and the remaining data points in that same sample may not be enough to compensate for this disruption. Thus, if n is very large, preasymptotic fluctuations of the quality of estimated parameters cannot be excluded even for data sets in the range of 10,000 observations.⁷ We note, however, that this apparently only happens for both very large n together with a high intermittency parameter λ_0^2 .

Nevertheless, with this particular caveat notwithstanding, the complete set of our simulations indicate that the GMM estimates are generally as well-behaved as they are expected to be.

2.4 Uncertainty of the number of cascade components

Our initial study (Tables 2.1 and 2.2) on GMM performance has been based on the assumption that we have exact knowledge about the relevant number of cascade steps. The lack of such knowledge introduces an additional source of uncertainty. To investigate the effect of such uncertainty we extend our previous analysis and generate samples of size $T = 10,000$ for a cascade of $n = 11$ levels with different values of λ_0^2 . We apply then our GMM estimator for a range of hypothesized cascade levels from 8 to 14 and contemplate the change in the estimated value as well as in the objective function. Results of pertinent Monte Carlo simulations are presented in Table 2.3.

⁷In principle, it is quite plausible that the range of preasymptotic volatility of estimates scales with the cascade level n .

2.4 Uncertainty of the number of cascade components

As it turns out, the additional uncertainty does not impede the correct estimation of the intermittency parameter even if the cascade generating the data has a higher or lower number of components than the one used for estimation. As it can be seen from Table 2.3, the absolute percentage difference (APD) between the estimates is at most three percent, which occurs with low λ_0^2 . In addition, the difference of the objective function compared to that of $n = 11$ increases with the difference of the assumed cascade steps from the true $n = 11$, with more change happening for lower than for larger values. The difference is more pronounced the higher λ_0^2 is. However, a large deviation between the minimized objective functions does not directly carry over to APDs, which appear to be smaller throughout the range of n considered. The reason for this is that our moment conditions focus on capturing the fluctuation generated at a cascade of size 2^ℓ so that for any higher cascade level, the number of anticipated switches decreases proportionally and, eventually, when the length of the cascade level is larger than the sample size T the number of added switches is at most one per level. As such, higher cascade levels add very little to the analytical moments, whereas the estimate of σ_1 absorbs higher level cascade components to some extent and, therefore, shows a bias that increases with n (the same observation has been made for the MSM model in Lux^[8]). In conclusion, our GMM procedure seems to provide reliable estimates of the intermittency generating parameter λ_0^2 even with uncertainty regarding the number of cascade steps.

2.5 Forecasting Methodology

Lux^[8] has introduced best linear forecasts to predict out-of-sample fluctuations of realizations of the causal Markov-switching multifractal process of Calvet and Fisher^[1,7]. Given a zero-mean weakly stationary process $\{Z_t\}$, the standard approach for construction of best linear h -step forecasts amounts to predicting the realization of the process at time horizons h by

$$\widehat{Z}_{t+h} = \sum_{i=1}^t \phi_{ti}^{(h)} Z_{t+1-i} = \Phi_t^{(h)} \mathbf{Z}_t, \quad (2.11)$$

with the vector of weights $\Phi_t^{(h)} = (\phi_{t1}^{(h)}, \phi_{t2}^{(h)}, \dots, \phi_{tt}^{(h)})'$ being any solution to the system $\Gamma_t \Phi_t^{(h)} = \gamma_t^{(h)}$, where $\Gamma_t = [\gamma(i-j)]_{i,j=1,\dots,t}$ is the variance-covariance matrix, and $\gamma_t^{(h)} = (\gamma(h), \gamma(h+1), \dots, \gamma(t+h-1))'$ denotes the vector of t elements of lag h auto-covariances and beyond (cf. Brockwell and Davis^[122]).

One consequence of the periodicity of size 2^n introduced to the series by the concatenation of cascades is that the long memory of the process is bounded by the length of that period. As such, its autocovariances would rapidly drop to zero after lag 2^n so that the inclusion of all available data, as one might consider when dealing with long-memory processes, should have no practical influence on the resulting forecasts beyond the maximum lag.

In the implementation of the procedure involving eq. (2.11), we use the iterative algorithm developed by Brockwell and Dahlhaus^[123] algorithm 5. For the implementation, one needs the autocovariances of the quantity one wishes to predict. In our case, our aim is to predict squared returns, x_t^2 , as a proxy of volatility which requires analytical solutions for $\mathbb{E}[x_{t+\ell}^2 x_t^2]$. With this in mind, we define a

series of zero-mean squared fluctuations:

$$Z_t \equiv x_t^2 - \mathbb{E}[x_t^2] = x_t^2 - \hat{\sigma}_1^2, \quad (2.12)$$

where $\hat{\sigma}$ is the estimate of the scale factor σ_1 in eq. (2.6). Also, $\hat{\sigma}_1^2$ appears only in the mean value of eq. (2.12), but it drops from the coefficients $\phi_{t,i}^{(h)}$. Appendix 2.B presents the pertinent formulae for the variance and auto-covariances of the intermittency generating part of a series of length T .

We explore again the performance of our proposed methodology via Monte Carlo simulations, assuming that one knows the exact number of cascade levels in the DGP. We restrict ourselves to one sample of size $T = 7500$, where we use the first 5,000 entries for in-sample parameter estimation and the remainder for an assessment of the out-of-sample forecasting performance in terms of mean squared error (MSE) and mean absolute error (MAE). Both MSE and MAE are standardized relative to the MSE and MAE of the most naive forecast, that is, the sample variance or squared "historical volatility" of a random-walk (RW) during the in-sample period, for the same sample, so that values below one indicate an improvement against the constant variance forecast based on a RW. We confine ourselves to a comparison of the quality of forecasts based on the GMM estimator, on the one hand, and on the Kiyono *et al.*'s estimator, on the other. A closer look at Table 2.4 shows that forecasts based on the $\hat{\lambda}_q^2$ estimate plus the sample standard deviation are fairly similar to those based on GMM. Indeed, both forecasts outperform the naive forecast at similar rates. This advantage of the model-based forecasts over the naive predictor initially increases with the degree of intermittency of the time series, i.e. λ_0^2 , but declines at the upper end

of the spectrum of values used in our Monte Carlo study. It seems worthwhile emphasizing that we have kept the in-sample period constant at $T = 5,000$ at all times. This means that the information used to estimate the parameters has not been updated over the out-of-sample period. The increase in biases in the estimate of σ_1 with increasing n does not appear to constitute a major obstacle for the prediction of future fluctuations.

The U-shape of the prediction accuracy with varying λ_0^2 is reminiscent of similar observations in Lux^[8]. Apparently, there are two opposite forces at work here: with small λ_0^2 , an increase of this parameter leads to a better forecasting performance simply because, then, the fluctuations become more pronounced, and the series shows more of a deviation from a random walk, while at very high values ($\lambda_0^2 = 0.15$) these fluctuations become more intermittent and less predictable given the sample size T available for estimation. This may also explain why the inconsistent estimator of Kiyono *et al.*^[6] is even marginally better than the GMM estimator at higher values of λ_0^2 : its strong downward bias (cf. Table 2.2) leads to smoother forecasts which on average might lead to somewhat smaller errors than forecasts based on a more accurate estimate.⁸

Next, we consider the case with added uncertainty on the number of cascade levels n of the series. Table 2.5 presents the forecasting results of a series generated by $n = 11$ and $T = 5,000$ splitted into two subsamples of 2,500 for estimation and forecasting. The process has been analyzed for a sequence of cascade levels ranging from 8 to 50, for which GMM estimation and subsequent out-of-sample

⁸ While the lack of an advantage of the GMM estimates compared to the inconsistent moment estimator of Kiyono *et al.* might appear disappointing, we should note that the use of the later for forecasting already implies quite some effort in computing exact moments (to implement eq. (2.12)). Hence, at this point a reliance on the moment estimator would be moot anyway.

forecasting exercises have been conducted. As we can see the MSEs and MAEs stabilize at the ‘true’ n : while using too low a number of cascade steps leads to suboptimal performance in forecasting, using even unboundedly high levels of n is almost completely harmless (except for a slightly higher variability of MSEs and MAEs around their means as indicated by their standard errors). Since, in practice, n will be typically unknown, these results speak in favor of using deliberately large hypothesized values of n in empirical research.

2.6 Empirical Evidence

Starting with Ghashghaie *et al.*^[79], a fair amount of effort has been spent particularly by physicists on the analogy between turbulence and financial markets (cf. Kiyono *et al.*^[6,84], Muzy *et al.*^[81], Arneodo *et al.*^[80], Ausloos and Ivanova^[83], and Kozuki and Fuchikami^[82]). In this strand of research, the main goal has been to retrieve the functional form of the relationship among PDFs of price changes at different scales. Instead, we focus in this chapter on forecasting turbulence (volatility) on the base of a cascade model of intermittent fluctuations. Recently, multiscale descriptions have also been used for the prediction of financial time series, cf. Nawroth *et al.*^[85].

Our analysis is based on data from seven different foreign exchange markets: the Canadian Dollar (CND), the Japanese Yen (YEN), the Swedish Krona (SEK), the Swiss Franc (CHF), the Australian Dollar (AUD), the Deutsche Mark- extended by the EURO since 1999- (DEM/EUR), and the British Pound (UKP), all against the U.S. Dollar. Further, we have analyzed the price of gold in U.S.

Dollars. All time series start on the 2nd of January of 1979 and extend until the 2nd of July of 2010.

Due to the slight variation in the number of active trading days among markets, we use the first ≈ 21.8 years of data for in-sample estimation and leave the remaining years for out-of-sample evaluation of volatility forecast. This gives exactly 5,500 in-sample observations for each asset and never less than 2000 observations for the out-of-sample analysis.

Though not compulsory for our GMM methodology, we employ the mentioned standardization for ω_j so that $\omega_j(t) \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$.⁹ Table 2.6 reports in-sample parameter estimates for the intermittency parameter λ_0^2 and for σ_1 , together with their standard errors and the corresponding probability of Hansen's test statistics $J_T = f_T(\hat{\varphi})' \Omega_T f_T(\hat{\varphi})$, where the estimation procedure has been repeated for $n = 8, \dots, 20$. At a significance level of 0.05, the J -test statistics would allow to reject the multifractal cascade as the DGP only for the Swiss Francs (CHF), on the base of our chosen moment functions. Further, we can see that while the number of cascade levels at which the lowest objective function was obtained varies from asset to asset, the Maximum APD among different n remains always below 2.5 %, except for the case of the Swiss Franc (CHF).

Finally, the forecasting procedure has also been applied for all model specifications $n = 8, \dots, 20$. The forecasting results under the MSE and MAE criteria for the highest cascade level are presented in Table 2.7 and 2.8, respectively.

Also presented are the forecasting results of a fitted GARCH(1,1) model for each

⁹The set of suggested moments in our GMM procedure allows the alternative specification $\sigma_1 = \exp(-n \tilde{\mu})$, together with an additional estimator for $\tilde{\sigma}^2$, the variance of each ω_j in eq. (2.6). In this case, however, the covariance matrix of the parameters would no longer be block diagonal.

asset.¹⁰ Though we abstain from presenting all details here, we find that the differences in forecast ability for different n have been marginal, with the forecasts for $n = 20$ being very close to that of all other n . This agrees with results reported in Lux^[8] and Calvet and Fisher^[7] who arrive at a similar conclusion regarding the saturation of forecasting performance beyond a certain threshold. As one can observe in Table 2.7 and 2.8, our procedure performs quite well for most of the series, particularly for the MSEs where results are almost always statistically significantly better than RW forecasts at the 99% level of the test statistic for nested models of Clark and West^[124,125].¹¹ In some cases, statistical significance was found even when the reported MSE is slightly higher than one. The reason is that if the two forecasts are highly correlated the series with a lower variance of squared-forecasting errors should be preferred even if its mean is slightly worse. As concerns comparison between the $n = 20$ Lognormal cascade and GARCH(1,1), most often GARCH(1,1) has a slight advantage for the smaller lags, while the cascade model provides better forecasts for larger horizons. Given that the GARCH(1,1) model has only short-term dependence (and two estimated parameters for the exact structure of this short-term dependence), while the cascade model is designated to capture dependence over larger horizons, these results very much coincide with our expectations. However, differences in both directions are mostly non-significant under the modified Diebold and Mariano^[126] test statistic (Harvey *et al.*^[127]) at the 95 % level. It is noteworthy nonetheless that the cascade forecasts typically dominate over the longer horizons which shows

¹⁰ Estimation results for the GARCH(1,1) are not presented here but are available upon request.

¹¹ Best-linear and GARCH forecasts can be considered to nest the naive forecasts from a random-walk (RW). Best-linear multifractal forecasts and GARCH forecasts are, however, not nested so that the non-adjusted version of the test applies.

the added-value of the long-term dependence for the multiplicative structure of volatility.

The MAE results in Table 2.8 display more significant results, with the $n = 20$ Lognormal cascade forecasts performing almost always better than GARCH(1,1), with a statistical significance of 95%. Note that under the MAE criterion, the cascade models mostly also dominates over GARCH(1,1) at short horizons, and even significantly so. There is also a difference in the significance of results against the RW-based forecast in this table with a somewhat smaller number of improvement for both the cascade and the GARCH(1,1) models. This, however, may be based on the nature of the modified Diebold and Mariano test employed here, as the MAE logic of averaging L_1 distances precludes us from employing the Clark and West^[124,125] adjustments for nested models. Overall, while the results for MAEs are not entirely homogeneous they appear quite encouraging particularly as concerns potential improvements against the GARCH(1,1) benchmark.

2.7 Conclusion

We have proposed in this chapter a GMM approach for estimation of Lognormal cascade processes, which compares favorably with previously proposed $\hat{\chi}^2$ and raw moment-based estimators. Our numerical analysis suggests that the GMM estimator is indeed consistent and asymptotical Normally distributed. Further, our methodology allows us to retrieve the cascade parameter value with high accuracy even when the number of levels of the cascade is unknown.

To apply the estimates obtained for forecasting the future evolution of a cas-

cade, we have developed a forecast methodology based on the Levinson-Durbin algorithm for best linear forecasts. Our methodology circumvents the statistical problems related to the definition of a cascade process on a bounded interval by allowing for a new initialization of the process each time the endpoint of the cascade is reached. We also show that the size of the interval, i.e. the number of cascade steps, has virtually no influence on the estimated intermittency parameter. The predictive power of forecasts based on past realizations is similarly relatively insensitive to the number of steps beyond some threshold. Somewhat surprisingly, using more precise GMM estimates yields virtually the same forecasting performance as combining the inconsistent Kiyono *et al.*'s estimator with best linear forecasts based on accurate moment conditions (a feature probably due to a lucky interplay between the bias of the estimator by Kiyono *et al.* and the volatility of forecasts for different parameters and cascade levels).

The applicability of our procedure is confirmed by an extensive simulation analysis. Our empirical application consists in the estimation of the intermittency parameter and the forecasting of volatility for various foreign exchange markets and the gold market. Our results suggest that cascade models, even with their grid-bound nature of volatility components, capture a non-trivial part of the variability of price fluctuations. This supports previous findings for the causal Markov-Switching Multifractal model (Calvet and Fisher^[1]). However, our use of the combinatorial structure of models of turbulence in physics demonstrates that similar results can be obtained without giving up the time-honored generating mechanisms for turbulent flows in statistical physics. Our approach might, therefore, be valuable for observations of turbulent processes in other areas that are not easily cast into a causal time series framework.

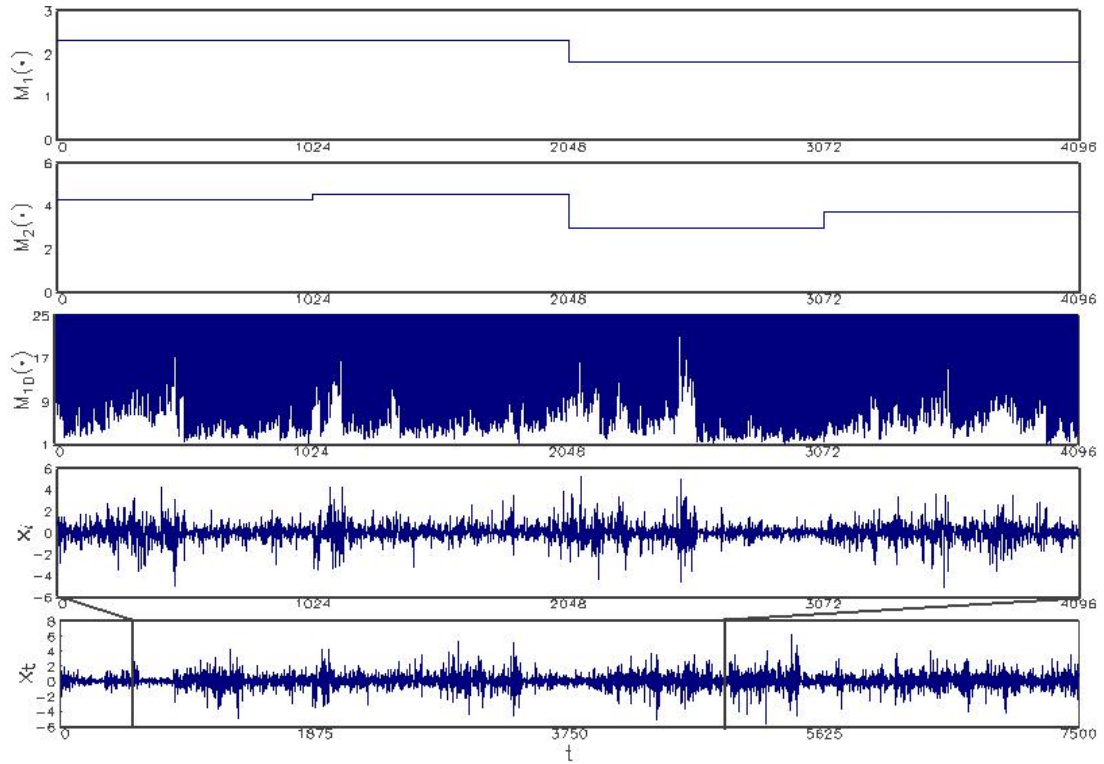


Figure 2.1: Illustration of one sample of the $\{x_t\}_{t=1}^{\infty}$ process of eq. (2.6) with $n = 12$. From top to bottom: The first level of draws of Lognormal random variables, the second level, the 10th level, the corresponding bounded $\{x_i\}_{i=1}^{2^n}$ process according to eq. (2.5), and a sample of 7500 points of the stationary $\{x_t\}_{t=1}^{\infty}$ process of eq. (2.6). The standardization of the PDF in eq. (2.1) suggested by Kiyono *et al.* for the construction of the cascades has been followed, with parameter value $\lambda_0^2 = 0.035$. For better visualization of the samples in the last two panels, $\sigma_1 = 2^{-n}$ was chosen to scale the overall magnitude of intermittency.

Table 2.1: Monte Carlo results for GMM estimator.

n	$\lambda_0^2 = 0.01$			$\lambda_0^2 = 0.05$			$\lambda_0^2 = 0.15$			
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	
8	$\bar{\lambda}_0^2$	0.010	0.011	0.011	0.047	0.049	0.050	0.142	0.146	0.148
	RMSE	.009	.008	.005	.017	.011	.007	.024	.016	.011
8	$\bar{\sigma}_1$	1.000	1.002	1.000	1.000	0.995	0.998	0.969	0.967	0.993
	RMSE	.039	.027	.019	.094	.066	.048	.216	.161	.139
16	$\bar{\lambda}_0^2$	0.010	0.011	0.011	0.047	0.049	0.050	0.142	0.146	0.148
	RMSE	.010	.007	.005	.016	.011	.007	.024	.017	.012
16	$\bar{\sigma}_1$	0.976	0.979	0.993	0.916	0.900	0.917	0.611	0.661	0.740
	RMSE	.224	.207	.182	.521	.433	.372	.676	.761	.840

NOTE: All simulations are based on a process with $\varepsilon \stackrel{d}{=} N(0,1)$, $\omega_j \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$, and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. RMSE denotes the root mean squared error. GMM was executed using lags $\ell = 1, 14, 64$. For each case, 400 Monte Carlo runs have been carried out.

Table 2.2: Monte Carlo results for $\hat{\chi}^2$, $\hat{\lambda}_q^2$, and \hat{F}_2 estimators.

E_i	n	$\lambda_0^2 = 0.01$						$\lambda_0^2 = 0.05$						$\lambda_0^2 = 0.15$					
		T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3			
$\hat{\chi}^2$	8	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.149	0.151	0.152			
		RMSE	.004	.003	.002	.008	.006	.004	.008	.006	.004	.008	.006	.019	.013	.009			
	16	$\bar{\sigma}_1$	1.004	1.004	1.001	1.024	1.019	1.024	1.024	1.019	1.024	1.024	1.050	1.050	1.050	1.062			
		RMSE	.049	.036	.024	.112	.076	.057	.112	.076	.057	.112	.076	.214	.162	.127			
$\hat{\lambda}_q^2$	8	$\bar{\lambda}_0^2$	0.007	0.007	0.008	0.034	0.036	0.040	0.034	0.036	0.040	0.040	0.119	0.125	0.129				
		RMSE	.005	.004	.003	.020	.018	.015	.020	.018	.015	.020	.018	.057	.048	.040			
	16	$\bar{\sigma}_1$	0.983	0.981	0.990	0.857	0.919	0.910	0.857	0.919	0.910	0.910	0.778	0.856	0.952				
		RMSE	.226	.195	.181	.417	.411	.351	.417	.411	.351	.417	.411	.709	.655	.753			
\hat{F}_2	8	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.049	0.049	0.050	0.049	0.049	0.050	0.050	0.142	0.144	0.147				
		RMSE	.002	.002	.001	.007	.005	.004	.007	.005	.004	.007	.005	.025	.019	.017			
	16	$\bar{\lambda}_0^2$	0.007	0.007	0.008	0.033	0.036	0.040	0.033	0.036	0.040	0.040	0.090	0.100	0.111				
		RMSE	.004	.003	.003	.019	.016	.013	.019	.016	.013	.019	.016	.064	.055	.044			
$\hat{\sigma}_1$	8	$\bar{\lambda}_0^2$	0.009	0.010	0.010	0.044	0.047	0.047	0.044	0.047	0.047	0.099	0.103	0.110					
		RMSE	.002	.002	.001	.013	.013	.010	.013	.013	.010	.013	.013	.057	.052	.046			
	16	$\bar{\lambda}_0^2$	0.006	0.007	0.008	0.026	0.030	0.033	0.026	0.030	0.033	0.052	0.059	0.066					
		RMSE	.004	.004	.003	.025	.022	.019	.025	.022	.019	.025	.022	.099	.092	.085			
$\hat{\sigma}_1$	8	$\bar{\sigma}_1$	1.000	1.001	1.000	0.997	0.992	0.996	0.997	0.992	0.996	0.973	0.970	0.993					
		RMSE	.037	.026	.019	.084	.063	.046	.084	.063	.046	.223	.156	.141					
	16	$\bar{\sigma}_1$	0.973	0.979	0.992	0.879	0.899	0.932	0.879	0.899	0.932	0.637	0.680	0.758					
		RMSE	.228	.206	.184	.463	.429	.397	.463	.429	.397	.910	.901	.701					

NOTE: All simulations are based on a process with $\varepsilon \stackrel{d}{=} N(0,1)$, $\omega_j \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$, and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters, while RMSE denotes the root mean squared error. E_i denotes the estimation method: $\hat{\chi}^2$, Kiyono *et al.*'s estimator $\hat{\lambda}_q^2$, Beck's flatness or fourth moment estimator \hat{F}_2 , and the simple sample standard deviation $\hat{\sigma}_1$. The $\hat{\chi}^2$ estimator was calculated using 20 bins with equal probability mass, that is, with a varying support-size $[x_{j-1}, x_j]$, for $j = 1, \dots, 20$, so that in each bin a probability mass of $CDF_{\lambda, \sigma_0}(x_j) - CDF_{\lambda, \sigma_0}(x_{j-1}) \approx 0.05$ is obtained. $\hat{\lambda}_q^2$ was calculated using $q = 0.5$ after the series was filtered by the sample estimate $\hat{\sigma}_1$. The same filtered series was employed for the \hat{F}_2 estimator, too. All entries referring to the parameter λ_0^2 were obtained by normalizing the resulting estimates by n . For each case, 400 Monte Carlo runs have been carried out.

Table 2.3: Monte Carlo results for GMM with cascade-level uncertainty added.

λ_0^2	n						
	8	9	10	11	12	13	14
0.01	0.011	0.011	0.011	0.011	0.011	0.011	0.011
RMSE	.005	.005	.005	.005	.005	.005	.005
APD	0.031	0.030	0.024	0.000	0.023	0.025	0.026
Qmin.	0.994	0.997	0.999	1.000	1.000	1.001	1.001
0.05	0.050	0.050	0.050	0.050	0.050	0.050	0.050
RMSE	.007	.007	.007	.007	.007	.007	.007
APD	0.006	0.003	0.001	0.000	0.001	0.001	0.001
Qmin.	0.991	0.990	0.996	1.000	1.003	1.004	1.005
0.15	0.151	0.149	0.149	0.149	0.148	0.148	0.148
RMSE	.012	.012	.012	.012	.012	.012	.012
APD	0.014	0.006	0.002	0.000	0.001	0.002	0.002
Qmin.	1.156	1.025	1.002	1.000	1.002	1.003	1.004

NOTE: All simulations are based on a process with $n = 11$ cascade levels, $\varepsilon \stackrel{d}{=} N(0, 1)$, $\omega_j \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$, and $\sigma_1 = 1$. The sample length is $T = 10,000$. GMM was executed for each number of cascade levels in the table using lags $\ell = 1, 14, 64$. For each case, 400 Monte Carlo runs have been carried out. λ_0 is the corresponding mean of the estimated parameters, while RMSE denotes the root mean squared error. For each iteration, the absolute difference between an estimate with cascade level n and the estimate with $n = 11$ as a proportion to the latter has been calculated. The Absolute Percentage Difference (APD) denotes the mean of such series. Qmin is the average of the objective value at the optimum for each n divided by the one with $n = 11$.

Table 2.4: Monte Carlo assessment of best linear forecasts based on GMM and $\widehat{\lambda}_q^2$ estimation.

T	$ \lambda_0^2$	$n = 10$		$n = 16$									
		$\widehat{\lambda}_q^2$	GMM	$\widehat{\lambda}_q^2$	GMM								
MSE	1	.954 (.023)	.911 (.039)	.936 (.056)	.959 (.025)	.911 (.039)	.938 (.054)	.912 (.075)	.835 (.166)	.828 (.228)	.919 (.073)	.835 (.165)	.850 (.195)
	10	.973 (.020)	.964 (.030)	.982 (.030)	.975 (.020)	.964 (.031)	.983 (.028)	.931 (.075)	.883 (.170)	.880 (.211)	.935 (.072)	.883 (.169)	.908 (.161)
	20	.980 (.018)	.975 (.026)	.987 (.024)	.981 (.018)	.975 (.026)	.989 (.022)	.938 (.074)	.895 (.171)	.890 (.201)	.942 (.071)	.895 (.169)	.920 (.149)
	100	.994 (.010)	.993 (.014)	.996 (.010)	.994 (.009)	.993 (.015)	.998 (.011)	.956 (.074)	.916 (.165)	.918 (.164)	.958 (.070)	.917 (.161)	.946 (.114)
MAE	1	.965 (.052)	.897 (.112)	.901 (.127)	.968 (.049)	.897 (.113)	.897 (.126)	.936 (.140)	.806 (.278)	.833 (.325)	.941 (.134)	.807 (.273)	.855 (.232)
	10	.977 (.045)	.952 (.096)	.970 (.088)	.979 (.042)	.952 (.097)	.964 (.091)	.948 (.137)	.857 (.278)	.908 (.297)	.952 (.131)	.859 (.268)	.924 (.189)
	20	.983 (.040)	.967 (.084)	.981 (.072)	.984 (.038)	.967 (.085)	.975 (.080)	.954 (.135)	.875 (.272)	.925 (.280)	.957 (.128)	.877 (.262)	.938 (.173)
	100	.994 (.021)	.992 (.042)	.997 (.030)	.995 (.020)	.993 (.046)	.992 (.060)	.968 (.122)	.913 (.241)	.960 (.221)	.970 (.115)	.915 (.227)	.963 (.133)

NOTE: The table shows mean squared errors (MSE) and mean absolute errors (MAE) for two different cascade sizes n . MSE and MAE are given in % of the pertinent MSEs and MAEs of a naive forecast from a RW using the in-sample variance. All entries are averaged over 400 Monte Carlo runs (with standard errors given in parenthesis). In each run, an overall sample of 7500 entries from a random starting point has been drawn. From that starting point on, an in-sample period of 5,000 entries for parameter estimation and an adjacent out-of-sample period of 2,500 entries for evaluation of forecasting performance were selected. Parameter values have been estimated with Kiyono's $\widehat{\lambda}_q^2$ and the GMM estimator respectively. GMM was executed using lags $\ell = 1, 14, 64$. The column T represents the forecast horizons, whereas the row λ_0^2 describes the selected intermittency values.

Table 2.5: Monte Carlo assessment of best linear forecasts with cascade-level uncertainty.

T	n										
	8	9	10	11	12	13	14	...	20	...	50
MSE	1	.896 (.073)	.893 (.078)	.891 (.081)	.890 (.083)	.890 (.084)	.890 (.085)	.890 (.085)	.890 (.086)	.890 (.086)	.890 (.086)
	10	.954 (.051)	.947 (.064)	.843 (.072)	.941 (.078)	.940 (.081)	.939 (.083)	.939 (.084)	.939 (.085)	.939 (.085)	.940 (.085)
	20	.970 (.038)	.961 (.054)	.956 (.066)	.953 (.074)	.951 (.078)	.951 (.081)	.951 (.082)	.951 (.084)	.951 (.084)	.951 (.084)
	100	1.000 (.013)	.992 (.018)	.983 (.036)	.977 (.052)	.974 (.063)	.972 (.070)	.972 (.073)	.971 (.077)	.971 (.077)	.971 (.077)
MAE	1	.890 (.126)	.884 (.146)	.881 (.163)	.880 (.175)	.878 (.185)	.878 (.190)	.878 (.193)	.878 (.198)	.878 (.198)	.878 (.199)
	10	.948 (.082)	.940 (.111)	.934 (.136)	.931 (.157)	.930 (.172)	.929 (.182)	.929 (.187)	.928 (.195)	.928 (.195)	.928 (.197)
	20	.967 (.061)	.957 (.090)	.950 (.119)	.947 (.143)	.945 (.162)	.944 (.174)	.944 (.180)	.943 (.190)	.943 (.190)	.943 (.192)
	100	.999 (.038)	.991 (.040)	.984 (.063)	.979 (.094)	.976 (.122)	.975 (.141)	.975 (.152)	.975 (.168)	.975 (.168)	.975 (.172)

NOTE: All simulations are based on a process with $n = 11$ cascade levels, $\varepsilon \stackrel{d}{=} N(0, 1)$, $\omega_j \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$, and $\sigma_1 = 1$. MSE and MAE are given in % of the pertinent MSEs and MAEs of a naive forecast from a RW using the in-sample variance. All entries are averages over 400 Monte Carlo runs (with standard errors given in parenthesis). In each run, an overall sample of 5,000 entries from a random starting point with $\lambda_0^2 = 0.05$ have been drawn. From that starting point on, an in-sample period of 2,500 entries for parameter estimation and an adjacent out-of-sample period of 2,500 entries for evaluation of forecasting performance were selected. Employed parameter values were estimated with GMM, using lags $\ell = 1, 14, 64$. The column T represents the forecast horizons.

Table 2.6: Empirical parameter estimates.

	Asset							
	CND	YEN	SEK	CHF	AUD	DM	UKP	Gold
$\widehat{\lambda}_0^2$.0154	.0357	.0167	.0119	.0357	.0238	.0191	.0339
(SE)	(.006)	(.007)	(.006)	(.005)	(.007)	(.006)	(.006)	(.007)
Max APD	.021	.006	.003	.165	.017	.011	.005	.003
$\widehat{\sigma}_1$.2781	.6945	.6556	.7533	.6208	.7144	.6546	1.3499
(SE)	(.009)	(.021)	(.038)	(.017)	(.032)	(.018)	(.020)	(.107)
J_{prob}	.243	.843	.718	.005	.500	.695	.550	.260
n_{min}	8	8	19	8	8	20	20	12

NOTE: Empirical estimates for standardized cascade shocks $\omega_j(t) \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$ and overall order of fluctuation magnitude σ_1 , obtained via GMM from a sample of 5500 entries for each asset. Each column show the estimate corresponding to the lowest objective function obtained for the range of cascade sizes $n = 8, \dots, 20$ for each asset. GMM was executed using lags $\ell = 1, 14, 64$. SE are the standard errors of the pertinent estimates and the entry J_{prob} gives the probability of the corresponding J statistics. n_{min} is the number of cascade levels at which the lowest objective function was obtained. The Maximum Absolute Percentage Difference (M-APD) is taken between the $\widehat{\lambda}_0^2$ with the lowest and the highest objective function from the employed range.

Table 2.7: Empirical forecast: MSE.

T	Asset									
	CND	YEN	SEK	CHF	AUD	DM	UKP	Gold		
GMM20	1	.801**	.956**	.852**	.943**	.786	.925**	.890**	.911**	
	5	.808**	.970**	.840**	.951**	.851**	.900**	.877**	.894**	
	20	.855**	.989**†	.867**	.980**	.927**	.914**	.917**	.917**	
	50	.913**†	.996**	.919**	.986**	.968**	.952**	.962**	.964**	
	100	.923**	.999**	.970**	1.015**	.973**	.988**	1.008**	.984**	
GARCH	1	.792**	.958**	.836**†	.942**	.788	.917**	.874**†	.903**	
	5	.794**	.971**	.827**†	.953**	.902*	.900**	.877**	.899**	
	20	.847**	.996**	.851**	.985**	1.056**	.913**	.908**	.917**	
	50	.933**	1.006**	.899**	.994**	1.203**	.961**	.948**	1.009**	
	100	.952**	1.011**	.948**	1.013**	1.252**	1.000	.997**	1.050**	

NOTE: Multifractal and GARCH(1,1) mean squared errors (MSE) in % of the pertinent MSEs of a 'naive' forecast from a RW using the historical variance. GMM20 denotes multifractal forecasts that were employed using a cascade level $n = 20$. Multifractal parameter values were estimated with the GMM estimator executed using lags $\ell = 1, 14, 64$.

* denotes an improvement against RW model which is significant at the 95% level.

** denotes an improvement against RW model which is significant at the 99% level.

† denotes an improvement significant at the 95% level (GMM20 against GARCH(1,1) and *vice versa*).

Comparisons against RW are based on the test statistic for nested models of Clark and West^[124,125]. Comparisons against GARCH(1,1) are based on the modified Diebold and Mariano^[126] test statistic by Harvey *et al.*^[127].

Table 2.8: Empirical forecast: MAE.

T	Asset	Asset									
		CND	YEN	SEK	CHF	AUD	DM	UKP	Gold		
GMM20	1	1.094	.949**	1.051†	.923**†	1.090†	.885**†	.911**†	.838**†		
	5	1.085	.949**†	1.045†	.916**†	1.093†	.872**†	.905**†	.838**†		
	20	1.080	.958*†	1.035†	.925**†	1.107†	.882**†	.910**†	.860**†		
	50	1.070	.964	1.044†	.932*†	1.120†	.900**†	.918*†	.887**†		
	100	1.061	.969	1.060†	.942	1.119†	.915*†	.942†	.904†		
GARCH	1	1.103	.950**	1.060	.939**	1.172	.908**	.920**	.849**		
	5	1.088	.957**	1.061	.939**	1.219	.907**	.921**	.857**		
	20	1.073	.974	1.068	.961**	1.352	.950**	.937*	.900*		
	50	1.037 †	.988*	1.088	.983*	1.641	1.016	.955	.991		
	100	1.013 †	.998	1.119	.998	2.082	1.064	.994	1.091		

NOTE: Multifractal and GARCH(1,1) mean squared errors (MSE) in % of the pertinent MSEs of a 'naive' forecast from a RW using the historical variance. GMM20 denotes multifractal forecasts that were employed using a cascade level $n = 20$. Multifractal parameter values were estimated with the GMM estimator executed using lags $\ell = 1, 14, 64$.

* denotes an improvement against RW model which is significant at the 95% level.

** denotes an improvement against RW model which is significant at the 99% level.

† denotes an improvement significant at the 95% level (GMM20 against GARCH(1,1) and *vice versa*).

Comparisons against RW and GARCH(1,1) are based on the modified Diebold and Mariano [126] test statistic by Harvey *et al.* [127].

2.A Appendix: Analytical Moments

We first consider the following definition:

$$\eta_{t,\ell} \equiv \sum_{j=1}^n [\omega_j(t) - \omega_j(t - \ell)], \quad (2.A.1)$$

where obviously it holds that $\mathbb{E}[\eta_{t,\ell}] = 0$. Lux^[8] shows that the moments of $\zeta_{t,\ell}$, defined as in eq. (2.8), can be written as:

$$\mathbb{E}[\zeta_{t+\ell,\ell} \zeta_{t,\ell}] = \mathbb{E}[\eta_{t+\ell,\ell} \eta_{t,\ell}] + \mathbb{E}^2[\ln |\varepsilon_i|] - \mathbb{E}[(\ln |\varepsilon_t|)^2] \quad (2.A.2)$$

and

$$\begin{aligned} \mathbb{E}[\zeta_{t+\ell,\ell}^2 \zeta_{t,\ell}^2] &= \mathbb{E}[\eta_{t+\ell,\ell}^2 \eta_{t,\ell}^2] - 4 \{ \mathbb{E}[\eta_{t,\ell}^2] - \mathbb{E}[\eta_{t+\ell,\ell} \eta_{t,\ell}] \} \{ \mathbb{E}^2[\ln |\varepsilon_i|] - \mathbb{E}[(\ln |\varepsilon_t|)^2] \} \\ &\quad + 3 \mathbb{E}^2[(\ln |\varepsilon_t|)^2] - 4 \mathbb{E}[\ln |\varepsilon_t|] \mathbb{E}[(\ln |\varepsilon_t|)^3] + \mathbb{E}[(\ln |\varepsilon_t|)^4]. \end{aligned} \quad (2.A.3)$$

As can be seen in eq. (2.A.2) and (2.A.3), the moments in eq. (2.9) require to compute $\mathbb{E}[\eta_{t,\ell}^2]$, $\mathbb{E}[\eta_{t+\ell,\ell} \eta_{t,\ell}]$, and $\mathbb{E}[\eta_{t+\ell,\ell}^2 \eta_{t,\ell}^2]$. These moments require two inputs, the probabilities for the renewal of the ‘multipliers’ ω_j , and the distribution of ω_j itself.

In the first case, $\eta_{t,\ell}^2$ elements would be different from zero if the pertinent ω_j has different realizations in $t + \ell$ and t . Therefore, the expected value of $\eta_{t,\ell}^2$ is

2.A Appendix: Analytical Moments

computed as

$$\begin{aligned}
 \mathbb{E}[\eta_{t+\ell,\ell}^2] &= \left(\frac{2}{2^n}\right) \sum_{j=1}^n \{\mathbb{E}[(\omega_j)^2] - \mathbb{E}^2[(\omega_j)]\} \{\mathbb{I}(2^{n-j} > \ell) (2^j \ell) + \mathbb{I}(2^{n-j} \leq \ell) 2^n\} \\
 &= \left(\frac{2 \tilde{\sigma}^2}{2^n}\right) \sum_{j=1}^n \{\mathbb{I}(2^{n-j} > \ell) (2^j \ell) + \mathbb{I}(2^{n-j} \leq \ell) 2^n\},
 \end{aligned} \tag{2.A.4}$$

where $\mathbb{I}(\cdot)$ represents the indicator functions, ω_j is distributed $N(\tilde{\mu}, \tilde{\sigma}^2)$, and 2^{n-j} accounts for the length of each subinterval at cascade step j . To understand the computation of eq. (2.A.4) note that 2^j is the number of different Lognormal draws at cascade level j while 2^{n-j} is the number of successive elements with equal contributions at level j within a single sequence of a cascade with 2^n time-ordered observations. In eq. (2.A.4), we distinguished between the cases $2^{n-j} > \ell$ and $2^{n-j} \leq \ell$. In the first case, we have to account for the sequence of the ℓ -first consecutive equal contributions at level j . From these ℓ of the 2^{n-j} numbers, a difference of ℓ will reach into the next box and, hence, have a nonzero value. If $2^{n-j} \leq \ell$ all ℓ -differences at level j will lead out of the individual box so that nonzero values will be estimated for all 2^n admissible starting points.

Calculations become slightly more involved for the autocovariances of $\eta_{t,\ell}$. First of all, we know that

$$\eta_{t+\ell,\ell} \eta_{t,\ell} = \left(\sum_{j=1}^n [\omega_j(t+\ell) - \omega_j(t)] \right) \left(\sum_{s=1}^n [\omega_s(t) - \omega_s(t-\ell)] \right). \tag{2.A.5}$$

Because of independence of realizations of any pair of volatility components j and s , only summands with $j = s$ give nonzero contributions. As such, eq. (2.A.5)

2.A Appendix: Analytical Moments

becomes

$$\eta_{t+\ell,\ell} \eta_{t,\ell} = \sum_{j=1}^n [\omega_j(t+\ell) \omega_j(t) - \omega_j^2(t) - \omega_j(t+\ell) \omega_j(t-\ell) + \omega_j(t) \omega_j(t-\ell)].$$

Further, different realizations in $t+\ell$ and t , and in t and $t-\ell$, that is, $\omega_j(t+\ell) \neq \omega_j(t) \neq \omega_j(t-\ell)$ at each cascade step j may now exist depending on the relationship between 2^{n-j} and 2ℓ , and 2^{n-j} and ℓ . We find, then, the autocovariances of $\eta_{t,\ell}$ using the first and second moments $\mathbb{E}[(\omega_j)^2] - \mathbb{E}^2[(\omega_j)]$:

$$\begin{aligned} \mathbb{E}[\eta_{t+\ell,\ell} \eta_{t,\ell}] = & - \left(\frac{\tilde{\sigma}^2}{2^n} \right) \sum_{j=1}^n \left\{ \mathbb{I}(2^{n-j} \leq \ell) 2^n + \mathbb{I}(2^{n-j} > \ell) \right. \\ & \left. \left[\mathbb{I}(2^{n-j} < 2\ell) \left[\mathbb{I}(j > 1) 2^j (2\ell - 2^{n-j}) \right] \right] \right\}. \end{aligned} \tag{2.A.6}$$

Eq. (2.A.6) can be understood following the sequence of different cases we distinguish: First, if $2^{n-j} \leq \ell$, any ℓ -difference at level j involves two different random numbers, and so, all admissible values, 2^n , make a nonzero contribution. If $2^{n-j} > \ell$, at least $2^{n-j} < 2\ell$ and $j > 1$ must hold to have any nonzero entries. Their number can then be determined by the following considerations: The term 2^j is the number of boxes of size 2^{n-j} on the bounded interval while $2\ell - 2^{n-j}$ determines the number of possible starting points of nonzero double differences of two times the size ℓ .

Calculations for the autocovariances of $\eta_{t,\ell}^2$ are more complex. We can arrive at the closed-form solutions for the total sample counterpart by identifying the

2.A Appendix: Analytical Moments

nonzero entries in the cascade with respect to

$$\eta_{t+\ell,\ell}^2 \eta_{t,\ell}^2 = \left(\sum_{j=1}^n [\omega_j(t+\ell) - \omega_j(t)] \right)^2 \left(\sum_{s=1}^n [\omega_s(t) - \omega_s(t-\ell)] \right)^2, \quad (2.A.7)$$

which requires to identify three different cases:

- (1) $j = s$ and $\omega_j(t+\ell) \neq \omega_j(t) \neq \omega_j(t-\ell)$ leading to entries of the form

$$(\omega_j(t+\ell) - \omega_j(t))^2 (\omega_j(t) - \omega_j(t-\ell))^2, \quad (2.A.8)$$

We count here the same number of entries as in the case of $\eta_{t+\ell,\ell} \eta_{t,\ell}$. Using the identities $\mathbb{E}[(\omega_j)^3] = 3 \tilde{\mu} \tilde{\sigma}^2 + \tilde{\mu}^3$ and $\mathbb{E}[(\omega_j)^4] = 3 \tilde{\sigma}^4 + 6 \tilde{\mu}^2 \tilde{\sigma}^2 + \tilde{\mu}^4$, the expectation of eq. (2.A.8), once the number of cases is counted, is

$$\mathbb{E}[(\omega_j)^4] + 3 \mathbb{E}^2[(\omega_j)^2] - 4 \mathbb{E}[(\omega_j)^3] \mathbb{E}[(\omega_j)] = 6 \tilde{\sigma}^4.$$

Together, we obtain

$$\begin{aligned} \kappa_1 = 6 \tilde{\sigma}^4 \sum_{j=1}^n \left\{ \mathbb{I}(2^{n-j} \leq \ell) 2^n + \mathbb{I}(2^{n-j} > \ell) \right. \\ \left. \left[\mathbb{I}(2^{n-j} < 2\ell) \left[\mathbb{I}(j > 1) 2^j (2\ell - 2^{n-j}) \right] \right] \right\} \end{aligned} \quad (2.A.9)$$

- (2) $j \neq s$ and $\omega_j(t+\ell) \neq \omega_j(t)$ and $\omega_s(t) \neq \omega_s(t-\ell)$ for the case of entries like

$$(\omega_j(t+\ell) - \omega_j(t))^2 (\omega_s(t) - \omega_s(t-\ell))^2, \quad (2.A.10)$$

In this case, we can simplify our computations by considering only contri-

2.A Appendix: Analytical Moments

butions for which the second index is lower than the first. The reason is that if a term is non-zero for the higher index, it is the lower index that determines whether the complete expression ((2.A.10)) is vanishing or not.¹² Therefore, we define for $h > l$

$$\begin{aligned} \psi(h) \equiv & \mathbb{I}(2^{n-h} \leq \ell) [\mathbb{I}(2^{n-l} > \ell) 2^l \ell + \mathbb{I}(2^{n-l} \leq \ell) 2^n] + \\ & + \mathbb{I}(2^{n-h} > \ell) \mathbb{I}(2^{n-h} < 2\ell) 2^l (2\ell - 2^{n-h}). \end{aligned} \quad (2.A.11)$$

In eq. (2.A.11) we encounter the following cases: First, if $2^{n-h} \leq \ell$ and $2^{n-l} > \ell$, a number of 2^l boxes has each ℓ nonzero contributions for the double differences. Further, if both $2^{n-h} \leq \ell$ and $2^{n-l} \leq \ell$, all countable elements, i.e. 2^n cases, make a nonzero contribution. If, finally, $\ell < 2^{n-h} < 2\ell$ we have 2^l boxes with $2\ell - 2^{n-h}$ nonzero elements each.

Furthermore, for non-zero entries, the expectation of eq. (2.A.10) is defined as

$$4 \mathbb{E}^2[(\omega_j)^2] - 8 \mathbb{E}[(\omega_j)^2] \mathbb{E}^2[(\omega_j)] + 4 \mathbb{E}^4[(\omega_j)] = 4 \tilde{\sigma}^4$$

Together with the number of cases, we obtain

$$\kappa_2 = 4 \tilde{\sigma}^4 \sum_{j=1}^n \sum_{\substack{s=1 \\ s \neq j}}^n [\mathbb{I}(j > s) \psi(j) + \mathbb{I}(j < s) \psi(s)], \quad (2.A.12)$$

(3) $j \neq s$ and $\omega_m(t + \ell) \neq \omega_m(t) \neq \omega_m(t - \ell)$ for $m = j, s$ leads to entries of the

¹² Please remember that a lower index means a position in a longer subinterval, so if the term is non-zero there, it must also be non-zero at a higher index level or shorter subinterval respectively.

format

$$\begin{aligned}
 & (\omega_j(t + \ell) - \omega_j(t)) (\omega_j(t) - \omega_j(t - \ell)) \\
 & (\omega_s(t + \ell) - \omega_s(t)) (\omega_s(t) - \omega_s(t - \ell)).
 \end{aligned} \tag{2.A.13}$$

As we can see, we have here a double term at cascade step j and another double term at a higher or lower cascade step s . Since both pairs of terms must not disappear simultaneously, once we are in, let's say, $\mathbb{I}(j < s)$, the total number of double terms is fully determined by the number of double terms with the index j . In summary, we have the same number of double terms for index j as in $\eta_{t+\ell, \ell} \eta_{t, \ell}$, so for $h > l$ we define the counting formula as

$$\begin{aligned}
 \varphi(l) = \sum_{l=1}^n \left\{ \mathbb{I}(2^{n-l} \leq \ell) 2^n + \mathbb{I}(2^{n-l} > \ell) \right. \\
 \left. [\mathbb{I}(2^{n-l} < 2\ell) [\mathbb{I}(l \geq 1) 2^l (2\ell - 2^{n-l})]] \right\}.
 \end{aligned} \tag{2.A.14}$$

Given the previous explanations, the components of eq. (2.A.14) are easily explained.

Using these results and the fact that the expectation of eq. (2.A.13) is $2\tilde{\sigma}^4$ we find

$$\kappa_3 = 2\tilde{\sigma}^4 \sum_{j=1}^n \sum_{\substack{s=1 \\ s \neq j}}^n [\mathbb{I}(j > s) \varphi(s) + \mathbb{I}(j < s) \varphi(j)]. \tag{2.A.15}$$

2.A Appendix: Analytical Moments

Putting (2.A.9), (2.A.12), and (2.A.15) together we finally obtain

$$\mathbb{E}[\eta_{i+\ell,\ell}^2 \eta_{i,\ell}^2] = \left(-\frac{1}{2^n}\right) (\kappa_1 + \kappa_2 + \kappa_3). \quad (2.A.16)$$

As the last element we need, the log-absolute moments of the standard Gaussian variates ε_i in eq. (2.A.3) can be easily obtained by using the Gamma function and its derivatives.

2.B Appendix: Forecasting Formulae

For computing linear forecasts of $x_t^2 - \widehat{\sigma}_1^2$ in eq. (2.12), we need the second moment and the autocovariances of the x_t . Let us define

$$\vartheta_t \equiv \exp \left[2 \sum_{j=1}^n \omega_j^{(m)} \left(\left\lfloor \frac{t-2^n(m-1)-1}{2^{n-j}} \right\rfloor \right) \right] = \exp \left[2 \sum_{j=1}^n \omega_j(t) \right] = \left[\prod_{j=1}^n M_j(t) \right]^2, \quad (2.B.1)$$

for $m = 0, 1, \dots$, and $\exp[\omega_j(t)] = M_j(t)$. Due to the standardization of the PDF in eq. (2.1) by $\sigma_0 = \exp(-\lambda^2)$ in eq. (2.2), it follows that $\omega_j(t) \stackrel{d}{=} N(-\lambda_0^2, \lambda_0^2)$, which conveys that $\mathbb{E}[\vartheta_i] = 1$ and the cascade level conserves mass on average. The second moment of the volatility process is

$$\mathbb{E}[\vartheta_t^2] = \mathbb{E} \left[\exp \left[4 \sum_{j=1}^n \omega_j(t) \right] \right] = \mathbb{E} \left[\left(\prod_{j=1}^n M_j(t) \right)^4 \right] = \exp(4 n \lambda_0^2), \quad (2.B.2)$$

while the autocovariance of the volatility process is given by:

$$\begin{aligned} \mathbb{E}[\vartheta_{t+\ell} \vartheta_t] &= \mathbb{E} \left[\prod_{j=1}^n [M_j^2(t) M_j^2(t+\ell) \mathbb{I}(M_j(t) \neq M_j(t+\ell)) + \right. \\ &\quad \left. + M_j^4(t) \mathbb{I}(M_j(t) = M_j(t+\ell))] \right] \\ &= \prod_{j=1}^n \left[\mathbb{E} [M_j^2(t) M_j^2(t+\ell) \mathbb{I}(M_j(t) \neq M_j(t+\ell))] + \right. \\ &\quad \left. + \mathbb{E} [M_j^4(t) \mathbb{I}(M_j(t) = M_j(t+\ell))] \right] \\ &= \prod_{j=1}^n \left[\mathbf{1} \mathbb{I}(2^{n-j} \leq \ell) + \left[\exp(4 \lambda_0^2) \frac{2^{n-2^j} \ell}{2^n} + \mathbf{1} \frac{2^j \ell}{2^n} \right] \mathbb{I}(2^{n-j} > \ell) \right]. \end{aligned}$$

Chapter 3

Continuous-Time MSM Model. Simulation-Based Estimation

3.1 Introduction

The inherent time-varying nature of volatility is one of the main areas of interest in financial economics and mathematical finance. In this strand of research, stochastic volatility (SV) models represent the most versatile instrument for capturing this phenomenon, with noteworthy developments in the areas of option pricing, portfolio selection, risk management as well as in the analysis of high-frequency time series (cf. Andersen *et al.*^[32], Shephard^[33], and Shephard and Andersen^[23] for a review of the history and applications of SV models).

The usual criterion for a construction to be classified as a SV model is the presence of a second source of uncertainty in the diffusion element of the price dynamics. This ad-hoc process acts as a mixing factor, providing a standard model of the desired time-varying features (cf. Chapter 1). One should note also that the

increase in complexity derived from a latent source of fluctuation rises not only benefits but also challenges, e.g. in the identification of the model. A great deal of effort has been devoted accordingly to the study of inference methodologies, with several approaches suggested for discrete-time SV models (cf. Shephard and Andersen^[23] for a detailed description).

On the other hand, the estimation of continuous-time models has been found rather difficult. This is not surprising, considering the impossibility to completely record continuous-time fluctuations and the general lack of closed-form solutions of the ‘aggregated’ returns. Simulation-based methods have proven here to deliver promising results, albeit being the application in each case heavily dependent on the specification of the SV model at hand. Among extant approaches, two clear strategies can be distinguished. On the one hand, methods like the Markov chain Monte Carlo (MCMC) (cf. Jones^[25], Eraker^[26], and Roberts and Stramer^[27]) and the simulated Maximum Likelihood (SML) (cf. Elerian *et al.*^[28], Brandt and Santa-Clara^[29], Durham and Gallant^[30], and Durham^[31]) attempt to amend a partial recording of the continuous-time fluctuation via data augmentation.¹ In the second group, we encounter methodologies that generate moment-condition proxies for the true continuous-time data generating process (DGP). The calculation of these moments is based on the simulation of the ‘complete path’ of an underlying continuous-time model. The main advantage of these approaches is their broader applicability. Indirect inference methods (II) (cf. Gouriéroux *et al.*^[128]) and the efficient method of moments (EMM) (cf. Gallant and Tauchen^[129]) are procedures that recur to a set of moment conditions that is endogenously defined by means of an auxiliary model. In contrast, the Simulated Method of Moments

¹ Cf. Shephard and Andersen^[23].

(SMM) (cf. Duffie and Singleton^[130]) uses sets of moment conditions of the ‘true’ continuous-time DGP.²

The Multifractal Model of Asset Returns (MMAR) of Mandelbrot *et al.*^[59] has been the first attempt at formulating a multifractal diffusion process in continuous time to capture the dynamics of financial price fluctuations. Due to its combinatorial nature, however, early research on the base of this formulation has been limited to tests based on moment scaling. The Markov Switching Multifractal model (MSM) of Calvet & Fisher^[1] generalizes the MMAR by randomizing news arrival times, guaranteeing a strictly stationary stochastic process that also is characterized by a causal rather than combinatorial development of volatility components. The continuous-time MSM provides a parsimonious diffusion with a very versatile structure due to its hierarchical Markov switching over multiple frequencies. When the number of frequencies is finite, the model fits naturally into the literature on regime-switching and Itô diffusions and, as such, can be classified as a particular variant of a SV model.

The estimation of MSM and MMAR have traditionally shared the same fate SV continuous-time models have had in that the standard econometric techniques have been applicable only after a discretization of the process on a finite grid. The ML estimation methodology from Calvet & Fisher^[7] or the GMM scheme from Lux^[8] rest on the existence of a discrete-time process generating the set of observations. In Chapter 2 we apply likewise a GMM methodology based upon Lux^[8] for the estimation and forecasting of discrete-time Lognormal cascades similar to those deriving from MMAR. As it turns out, the MSM model is at

² A thorough overview of the SMM as well as the other aforementioned methods is provided in Carrasco and Florens^[131].

least equivalent and even has a competitive edge against the more time-honored ARCH-type models in the forecasting of volatility.

One may wonder what the benefit of focusing on a continuous-time model may be provided the discrete-time counterpart already provides satisfactory results. One important reason for considering continuous-time models is that with the ongoing development of financial markets the period at which electronic exchanges remain closed has significantly been reduced. This allows information arrivals to be processed at practically any second of the day. A continuous-time framework may thus resemble reality more accurately than a market construction receiving the information flow only once a day at closing time. Today's availability of high-frequency financial asset prices has additionally allowed to approach a continuous-time limit quite closely. Then again, there is a definite lower-bound horizon that can be used in practice for the estimation of continuous-time models, due to either limited finite sampling or market microstructure noise (cf. Andersen *et al.*^[32]). Aggregation of returns from the underlying high-frequency asset prices has thus become a necessary step to reduce the effects from measurement error when conducting estimations (cf. Zhang *et al.*^[132]).

This chapter is the first known attempt of application of a parameter estimation algorithm for the continuous-time MSM model. To pursue this, we adopt a special version of the Simulated Method of Moments approach. Our strategy consists in a two-step procedure, where the GMM scheme of Lux^[8] is first applied to the tightly discretized version of the MSM model. Despite misspecification, the daily return moments of the GMM scheme provides us with valuable initial estimates at a very low computational cost. The second step consists in an iterative SMM approach initialized at the values and weighting matrix estimates derived from

the first step. As in Lux^[8], our procedure accommodates different distributions of the volatility process, among which the Bernoulli and the Lognormal distributions are the most commonly used. In our implementation we employ two alternative sets of moments depending on the information used in the estimation procedure. The first set focuses on the evolution of continuous-time returns. It is applicable when only a sample of daily prices is available to the researcher or, alternatively, when a measurement of daily returns is generated via record aggregation. The second set targets more directly the latent volatility process. These moments consist of functions of the quadratic variation (QV) associated to the return process and are applicable when a high-frequency metric like realized volatility (RV) can be obtained from a sample of intra-daily returns.

The rest of this chapter is organized as follows. In section 3.2 we start with a general description of SV models, and we proceed in section 3.3 with a specification of the MSM continuous-time process. We continue in section 3.4 with a description of the moments applied. In section 3.5 we describe the methodology implemented while in section 3.6 we provide details of Monte Carlo simulations. In section 3.7 we assess the performance of our methodology in an empirical implementation with foreign-exchange and stock-index time series. We conclude this chapter with our final remarks in section 3.8.

3.2 Preliminaries

In the context of time series analysis, a continuous-time asset pricing model typically rests on one or more stationary processes, depending on the nature of the

fluctuations intended. It is well known that under assumptions of absence of arbitrage and finiteness in the mean-variation the log-price process can be considered as a semi-martingale. Following Andersen *et al.*^[34], a continuous log-price process $P(t) \equiv \ln [Y(t)]$ admits a canonical decomposition on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, namely

$$P(t) = P(0) + A(t) + X(t),$$

where $A(t)$ represents the finite variation predictable components, $X(t)$ is the local martingale part and, by definition, $A(0) \equiv X(0) \equiv 0$.³ The standard SV model consist then in the fluctuating martingale

$$X(t) = \int_0^t \sigma(s) dB(s) \tag{3.1}$$

composed of two processes, the Brownian motion $B(t)$ and the non-negative spot volatility process $\sigma(t)$ that is assumed to have càdlàg sample paths.⁴ If for the moment we disregard the predictable components, $X(t)$ depicts the change in log-prices $P(t) - P(0)$ and has continuous sample paths even if $\sigma(t)$ does not. A necessary and sufficient condition for $X(t)$ to be a martingale is that

$$\mathbb{E} \left[\left(\int_0^t \sigma^2(s) ds \right)^{1/2} \right] < \infty.$$

Note that the assumption of independence between $B(t)$ and $\sigma(t)$ allows us to express $X(t)$ in eq. (3.1) as the compound process $X(t) = B[\theta(t)]$, where $\theta(t)$ is

³ We refer the reader to Andersen *et al.*^[34] for a detailed description of the cases covered by this specification of $A(t)$.

⁴ In French “continue à droite, limite à gauche” denotes a right-continuous function with left limits.

the integrated variance (IV) process

$$\theta(t) = \int_0^t \sigma^2(s) ds.$$

Alternatively, if we start from the assumption that $X(t)$ is modeled as a continuous-time martingale, $X(t)$ can be represented by the Dambis-Dubins-Schwartz Theorem as a time-changed Brownian motion, where the time-change is depicted by the quadratic variation (QV) process

$$[X](t) = \text{p lim} \sum_{j=1}^n [X(j\Delta t) - X((j-1)\Delta t)]^2, \quad (3.2)$$

for a sequence of n uniform partitions of size $\Delta t = \frac{t}{n}$ over the interval $[0, t]$ with $\Delta t \rightarrow 0$ as $n \rightarrow \infty$ (cf. Shephard and Andersen^[23] and Chapter 1 of this thesis for more details).

As in Andersen *et al.*^[34], let us finally denote

$$R(t) \equiv X(t) - X(t-1) = \int_{t-1}^t \sigma(s) dB(s) \quad (3.3)$$

as the continuously compounded return over the interval $[t-1, t]$ and

$$[R](t) \equiv [X](t) - [X](t-1) \quad (3.4)$$

as the increment in QV over $[t-1, t]$. Note that while preserving the main properties of $X(t)$ and $[X](t)$ respectively, $R(t)$ and $[R](t)$ allow us to focus on predefined time intervals facilitating the application of time series estimation techniques.

3.3 The continuous-time MSM model

We turn now to the specification of the process $\sigma(t)$. For this, we consider the continuous-time MSM process of Calvet & Fisher^[1,60,113], whose IV process leads to the Poisson multifractal measure with non-hierarchical time subintervals (cf. Chapter 1 of this thesis for more details). This continuous-time diffusion is based upon the Markov state vector $M(t)$ with \bar{k} components:

$$M(t) \equiv (M_1(t), M_2(t), \dots, M_{\bar{k}}(t)) \in \mathbb{R}_+^{\bar{k}}, \quad t \in [0, \infty),$$

which are mutually independent across the \bar{k} ‘cascade levels’. The duration of each component $M_k(t)$, $k \in \{1, \dots, \bar{k}\}$ is heterogenous as the frequencies of arrivals are assumed to progress geometrically with k , i.e. at stage k the intensity parameter γ_k is given by $\gamma_k = b \gamma_{k-1}$, for $b \in (1, \infty)$ and a strictly positive intensity γ_{k-1} . The change in any $M_k(t)$ over the infinitesimal interval dt is drawn from the same marginal distribution M when triggered by a Poisson arrival with pertinent intensity γ_k . When no trigger event occurs, the variable $M_k(t)$ remains constant. In short, the dynamics of the $M_k(t)$ between t and $t + dt$ are given by

$$\begin{aligned} M_k(t + dt) \text{ drawn from distribution } M & \quad \text{with probability } \approx \gamma_k dt \\ M_k(t + dt) = M_k(t) & \quad \text{with probability } \approx 1 - \gamma_k dt, \end{aligned}$$

where the switching events and new draws from M are assumed to be mutually independent across k and dt . M is required to be a distribution with positive support and mean equal to unity, i.e. $M \geq 0$ and $\mathbb{E}(M) = 1$. As in Lux^[8], we consider M to be either a Bernoulli distribution with states m_0 and $2 - m_0$

3.3 The continuous-time MSM model

($1 \leq m_0 < 2$) with equal probability, or a Lognormal distribution $\text{LN}(\tilde{\mu}, \tilde{\sigma}^2)$. In the latter case, $M(t)$ consists of \bar{k} different realizations of Lognormal random draws at every given instant t . Condition $\mathbb{E}(M) = 1$ implies that

$$\exp\left(\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2\right) = 1,$$

and the specification of the Lognormal distribution can be reduced to one parameter, say λ_0 , with $\tilde{\mu} = -\lambda_0$ and $\tilde{\sigma} = \sqrt{2\lambda_0}$.

The geometric progression of the Poisson arrival intensities $\gamma \equiv (\gamma_1, \gamma_2, \dots, \gamma_{\bar{k}})$ may be reformulated as

$$\gamma_k \equiv \gamma_{\bar{k}} b^{k-\bar{k}}$$

where $\gamma_{\bar{k}} > 0$ determines the time horizon of the fastest arrival, and $b \in (1, \infty)$ the progression among arrival rates.

Taking all components together, the volatility process is given by

$$\sigma(t) = \sigma(M(t)) \equiv \sigma_1 \left(\prod_{k=1}^{\bar{k}} M_k(t) \right)^{1/2} \quad (3.5)$$

where σ_1 is a positive constant that captures the unconditional level of fluctuations. As an Itô diffusion, the continuous-time MSM process $X(t)$ is required to fulfill $\int_0^t \sigma^2(M(s)) ds < \infty$, which is satisfied as $\mathbb{E} \left[\left(\int_0^t \sigma^2(M(s)) ds \right)^{1/2} \right] = \sigma_1 \sqrt{t} < \infty$. The process $X(t)$ has strictly stationary increments by construction, which are governed by the distribution M and the parameter space $(\sigma_1, b, \gamma_{\bar{k}})$. $X(t)$ is furthermore a martingale, that produces a bounded and absolutely continuous QV process.

3.4 The Moments

We proceed by describing the moments employed in the SMM methodology of the next section. From this point on, we make the stronger assumption that $A(t)$ is a predetermined function at the filtration set $\mathcal{F}(t)$. This allows us to separate the estimation of the drift, which could be executed beforehand, from the SV part, which is the focus of this chapter.⁵

A common situation faced in a study like ours is the impossibility to determine the analytical form of the density of $X(t)$ on the path $X_t \in \{X_{0+\Delta t}, \dots, X_1, \dots, X_T\}$ when observations p_i consist only of discrete measurements at integer instants $i = 0, \dots, T$. The existence of discrete observations p_i results commonly from infrequent recording times, namely, when only a sequence of daily log-prices $\mathcal{P} = \{p_0, p_1, \dots, p_T\}$ is recovered despite the existence of intraday activity. In this case, the modeling assumptions imply that each of these recorded prices depicts the compounding of the process along the sample path provided by the previous data point, that is, the observations p_t and p_{t-1} are the only recorded points of all the sampled series between $P(t)$ and $P(t-1)$ in eq. (3.3) and so $r_t \equiv p_t - p_{t-1}$. On the other hand, when $\mathcal{P}_{\Delta t} = \{p_0, p_{0+\Delta t}, \dots, p_1, p_{1+\Delta t}, \dots, p_{T-1}, p_{T-1+\Delta t}, \dots, p_T\}$ is available, the main reason for record aggregation of high-frequency data is the presence of pronounced and systematic intraday patterns in return volatility which may render the direct employment of these intraday observations for estimation troublesome. Record aggregation may be used in this sense to avoid

⁵ This assumption is not as restrictive as it first may seem given the fact that in daily observations a constant mean and a very slight, if any, autocorrelation at the first lag are typically found.

contamination with microstructure noise by creating measures of daily returns

$$r_t \equiv \sum_{j=1}^{1/\Delta t} p_{t-1+j\Delta t} - p_{t-1+(j-1)\Delta t} \quad (3.6)$$

which may disregard, for instance, the noisy overnight returns. Under any of these two circumstances, it seems suitable for the estimation of the continuous-time MSM process to use moments based on the returns $R(t)$ of eq. (3.3), where the time included on the interval $[t-1, t]$ represents one day.

Alternatively, record aggregation may be used for constructing a realized volatility (RV) estimator

$$RV_{0,t} \equiv \sum_{j=1}^{t/\Delta t} (p_{j\Delta t} - p_{(j-1)\Delta t})^2$$

that converges to $RV_{0,t} \xrightarrow{p} [X](t)$ of eq. (3.2) as $\Delta t = \frac{t}{n} \rightarrow 0$ for $n \rightarrow \infty$, in the absence of observation errors (cf. Andersen *et al.*^[34]). In practice, however, a sample of squared intra-daily returns taken at a fixed Δt is commonly aggregated into a time series of daily increments $RV_{t-1,t} \equiv RV_t$, annihilating the idiosyncratic noise over the interval $[t-1, t]$ and leading to a ‘smoother’ estimator of $[R](t)$ in eq. (3.4) (cf. Andersen *et al.*^[32]).⁶ Provided that a reliable measurement like RV_t can be computed, we will employ as an alternative in our SMM procedure moments based solely on $[R](t)$.

Before turning to the definitions of the actual functions, note that the strict stationarity property of the continuous-time MSM process implies that $X(t)$ in eq. (3.1) (from where $R(t)$ and $[R](t)$ inherit their properties, respectively) may be

⁶ We assume that no jumping terms are present in the time series at hand. Empirical work in the last section will show that this assumption may not be fully justified for all assets.

approximated with good accuracy by sufficiently small discrete increments. In this chapter, we use the Euler discretization procedure first introduced by Calvet and Fisher^[7,133], and detailed in Appendix 3.A, for simulating synthetic samples of the MSM process and extracting their finite sample moments. The estimation problem of the continuous-time MSM model consists therefore in finding parameter values for which the empirical sample moments match as closely as possible those from simulated data of the hypothesized DGP. SMM extends in this way the applicability of the GMM estimator to a larger class of situations for which the moment conditions of interest do not have analytic representations (cf. Duffie and Singleton^[130]).

The functions based upon the continuously compounded returns $R(t)$ consist of the logarithmic differences

$$\zeta(t, \ell) \equiv \ln |R(t)| - \ln |R(t - \ell)|,$$

where ℓ represents the lag in the set of integer instants $1, \dots, T - 1$ at which the difference is taken. Moments of this type have already been employed for the discrete-time analogue of the MSM model in Lux^[8].

Alternatively, the functions based upon the daily increments $[R](t)$ consist of

$$\xi(t, \ell) \equiv \ln [R]^{1/2}(t) - \ln [R]^{1/2}(t - \ell). \tag{3.7}$$

An application of the discrete-time version of these functions was first proposed by Lux *et al.*^[9] for the study of daily RV. However, the focus there was the forecastability of QV increments on the base of a discrete-time MSM model. We will

concentrate here on the estimation of the continuous-time MSM process instead, and consider RV_t merely as a smoother function of the latent continuous-time fluctuations. The application of eq. (3.7) thus resembles the work of Bollerslev and Zhou^[134], where a method of moments approach for the estimation of a traditional SV model was employed. In their approach, the authors obtain analytical moments for the latent IV process, the realization of which was approximated by a RV series. In our case, however, the lack of closed-form solutions will lead us to the simulation of the underlying DGP for the subsequent computation of moments (cf. Corradi and Distaso^[135]).

In order to exploit the scaling properties of the cascade process, we select autocovariances of the overlapping logarithmic differences $\zeta(t, \ell), \zeta(t + 1, \ell), \dots$ as in Lux^[8] in the first case, and of $\xi(t, \ell), \xi(t + 1, \ell), \dots$ as in Lux *et al.*^[9] in the second. The transformed variables $\zeta(t, \ell)$ and $\xi(t, \ell)$ are additive measures of the continuous-time fluctuation and their moments will depend on the parameters of the volatility process, as in the discrete-time MSM model. In Appendix 3.A, we also show that the unconditional standard deviation σ_1 drops out in any of these two formulations. For this reason, the moment conditions in the first case will consist of

$$\mathfrak{Mom}(\ell, q) = \mathbb{E} [\zeta^q(t + \ell, \ell) \cdot \zeta^q(t, \ell)], \quad (3.8)$$

for $q = 1, 2$ and $\ell = 1, 5, 10, 50$, together with a raw moment like $\mathbb{E}[R(t)^2]$ for the identification of σ_1 .⁷ In the second case, we will make use of

$$\mathfrak{Mom}(\ell, q) = \mathbb{E} [\xi^q(t + \ell, \ell) \cdot \xi^q(t, \ell)] \quad (3.9)$$

⁷ The value of the largest ℓ may vary according to the sample size at hand without sizeable changes in the results.

together with $\mathbb{E}[[R](t)]$, also for $q = 1, 2$ and $\ell = 1, 5, 10, 50$. Appendix 3.A contains the explicit statements involved in the moments introduced in eq. (3.8) and in eq. (3.9), as well as the simulation procedure used for their numerical approximation.

For the rest of this chapter we will fix $b = 2$ and consider a Poisson arrival with prespecified intensity $\gamma_{\bar{k}}$ as in Lux^[8], Lux and Morales-Arias^[97], and Lux *et al.*^[9]. Our parameter space φ will consequently consist of σ_1 and either m_0 or λ_0 according to the distribution chosen for the \bar{k} factors $M_k(t)$.

3.5 Estimation Methodology

GMM is a very general statistical approach for estimation of the parameters of a model (cf. Hansen^[120]). Given a set of analytical moments $k(\varphi)$, the vector of parameter estimates of the true φ_0 is obtained as the result of the minimization of an objective function of the following form:

$$\widehat{\varphi}_T(\Omega) = \arg \min_{\varphi \in \Phi} \Psi_T(\varphi) = \arg \min_{\varphi \in \Phi} f_T(\varphi)' \Omega f_T(\varphi), \quad (3.10)$$

with Φ representing the parameter space; $f_T(\varphi)$ the vector of differences between a set of sample and analytical moments $\frac{1}{T} \sum_{t=1}^T [K(r_t) - k(\varphi)]$, with $K(\cdot)$ being a function of the observation r_t so that $\mathbb{E}[K(r_t)] = k(\varphi_0)$; and Ω a positive definite and possibly random weighting matrix. When suitable sets of ‘regularity conditions’ (also referred to as Hansen’s^[120] regularity conditions) are fulfilled (cf. Harris and Mátyás^[119]), $\widehat{\varphi}_T$ is consistent and asymptotically Gaussian. The

estimator $\widehat{\varphi}_T$ then converges to

$$\sqrt{T} (\widehat{\varphi}_T(\Omega) - \varphi_0) \xrightarrow{d} \mathcal{N} (0, \Xi_1^{-1} \Xi_2 \Xi_1^{-1}),$$

where the factors $\Xi_1 = F_0' \Omega F_0$ and $\Xi_2 = F_0' \Omega V[f_0(\varphi_0)] \Omega F_0$ are composed of constant limiting matrices $F_0 \equiv \mathbb{E}[\partial k(\varphi_0)/\partial \varphi]$ and $V[f_0(\varphi_0)] \equiv \lim_{T \rightarrow \infty} T \cdot \mathbb{V}\mathbb{A}\mathbb{R}[f_T(\varphi_0)]$, to which $F_T = \partial f_T(\widehat{\varphi}_T)/\partial \varphi$ and $V[f_T(\widehat{\varphi}_T)] = T \cdot \mathbb{V}\mathbb{A}\mathbb{R}[f_T(\widehat{\varphi}_T)]$ converge.

Standard asymptotic results show that any Ω satisfying positive definiteness will lead to a consistent and asymptotically Gaussian GMM estimator, with pertinent asymptotic variance of $\widehat{\varphi}_T$. It can be shown that taking an efficient estimator of the inverse of the asymptotic variance-covariance matrix of the moments as the weighting matrix Ω is optimal in that the resulting $\widehat{\varphi}_T$ will be the most efficient estimator in the class of all asymptotically Gaussian estimators. In practice, however, the latter choice relies on the unknown true parameter vector φ_0 and on the set of analytical moments employed. In Lux^[8], for instance, the selection of these moments has been mostly focused on exploiting systematically the degree of uncertainty generated by the ‘long memory’ property of the MSM process. Admittedly, MSM models are characterized by only an ‘apparent’ long memory, with an asymptotic hyperbolic decline of the autocorrelation of absolute powers over horizons $1 \ll \tau \ll b^{\bar{k}}$ and exponential decline thereafter (cf. Calvet and Fisher^[7]). This proximity to a power-law structure may nonetheless cause practical concerns when working with finite samples. For $b = 2$ and $k > 12$, for instance, the extent of the power-law scaling might exceed the size of most available data for daily financial prices (cf. Lux^[8]). To circumvent these potential problems, log-moments

are chosen as the log transformation guarantees that autocorrelations are equal to zero beyond a certain time horizon.

In his application, Lux^[8] also uses an iterative GMM scheme for the estimation of the discrete-time MSM. After a solution to eq. (3.10) is found, the procedure carries on with the computation of a new weighting matrix $\Omega_T = [V[f_T(\hat{\varphi}_T)]]^{-1}$ that functions as an input in the next evaluation of eq. (3.10). The estimation procedure continues until convergence of both the parameter estimates and the weighting matrix is obtained (cf. Hansen *et al.*^[121]). Preliminary Monte Carlo results on the discrete-time MSM show a consistent slight advantage in terms of smaller biases and dispersion of the estimator with the iterative GMM in comparison to the alternative scheme, the two-step GMM. In empirical work, the former methodology may also be profitably adapted to SMM estimators as it may lead to smaller asymptotic confidence intervals. We will turn back again to this point in the next section and will proceed here with the general properties of the current procedure.

The SMM estimator is the counterpart of GMM replacing the analytical moment-based function $f_T(\varphi)$ by

$$f_{ST}(\varphi) = \frac{1}{T} \sum_{t=1}^T \left[K(r_t) - \frac{1}{ST} \sum_{s=1}^S \sum_{t=1}^T \tilde{k}(u_{t,\Delta t}^s; \varphi) \right], \quad (3.11)$$

with $u_{t,\Delta t}^s$ being a draw from the distribution function used in the simulation, $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ the simulator such that $\lim_{T \rightarrow \infty} f_{ST}(\varphi) = 0$ if and only if $\varphi = \varphi_0$, and \mathcal{S} the number of simulated samples of size T (cf. Appendix 3.A of this chapter and Gouriéroux and Monfort^[136]).

The SMM approach is not new to the literature of multifractal models. In an

earlier version of Calvet and Fisher^[7], the authors have implemented an SMM approach for the estimation of the discrete-time MSM based on a variety of moment-scaling properties, parameters of log – log regressions of the sample autocovariogram, slope parameters from log – periodogram regressions, high-frequency autocovariances, and tail index estimates. Calvet and Fisher^[77] devised an SMM adaptation of the ‘scaling estimator’ from Calvet *et al.*^[59], but they focused mostly on the identification of the process and later discarded the procedure in favor of Maximum Likelihood (ML) techniques. In this chapter, we will focus on the log moments $\mathfrak{M}\mathfrak{om}(\ell, q)$ of eq. (3.8) and $\mathfrak{M}\mathfrak{om}(\ell, q)$ of eq. (3.9), originally devised by Lux^[8] and Lux *et al.*^[9] in the GMM estimation of the discrete-time MSM model. Note, also, that the set of analytical moments to be replaced by the simulator $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ in eq. (3.11) refers, strictly speaking, to the moments derived from the discretized version of $X(t)$ in eq. (3.1) (cf. for the calculation of $\zeta(t, \ell)$ and $\xi(t, \ell)$, the use of $R(t)$ and $[R](t)$ with that of their counterparts R_t of eq. (3.A.3) and $[R]_t$ of eq. (3.A.4) in Appendix 3.A, respectively). Provided though that the approximation error of these moments to the true moments of the continuous-time MSM model vanishes for an increasingly smaller Δt , we make reference only to the latter in this chapter.

As in eq. (3.10), we refer to the vector $\widehat{\varphi}_{sT}$ as the solution to

$$\widehat{\varphi}_{sT}(\Omega) = \arg \min_{\varphi \in \Phi} \Psi_{sT}(\varphi) = \arg \min_{\varphi \in \Phi} f_{sT}(\varphi)' \Omega f_{sT}(\varphi), \quad (3.12)$$

with Φ being the parameter space, and Ω the weighting matrix. A set of assumptions under which the SMM estimator can achieve consistency and asymptotic normality has been depicted by Duffie and Singleton^[130]. We discuss now the

applicability of these conditions together with the resulting properties of $\widehat{\varphi}_{ST}$. For consistency, the authors consider first that a convergent weighting matrix Ω_T is available, such that $\Omega_T \rightarrow [V^s[f_0(\varphi_0)]]^{-1}$, where $V^s[f_0(\varphi_0)]$ is a nonsingular matrix containing the limiting second moments of $f_{ST}(\varphi_0)$.⁸ The remaining conditions identify those cases under which the SMM estimator is asymptotically valid provided an arbitrary initial state of the simulation process. Note, in our case, that the MSM process $X(t)$ of eq. (3.1) has strictly stationary increments, that $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ in eq. (3.11) is Lipschitz, uniformly in probability, and that its latent Markov state vector $M(t)$ obeys the ergodicity property (cf. Calvet and Fisher^[60]). Consequently, we may avoid those cases that hinder consistency by setting the proper initial conditions on the first place, namely, by adopting the ergodic distribution of $M(t)$ as its initial distribution at $t = 0$.

With respect to Ω_T , Duffie and Singleton^[130] suggest to estimate $V^s[f_0(\varphi_0)]$ based upon empirical sample moments only. These moments can be obtained by the Newey-West^[137] heteroscedasticity- and autocorrelation-consistent (HAC) estimator. Note though that the proximity to long-memory of the MSM model, the slow convergence rate of spectral estimators like the one proposed here, and the potentially partial record of the DGP in the empirical sample make the estimation of $V^s[f_0(\varphi_0)]$ based on empirical data simply a first working approach. One may alternatively use simulations for the computation of Ω_T . This strategy may prove advantageous provided it allows to control the sample size through \mathcal{S} (cf. Duffie and Singleton^[130]). Recall also our argument for consistency: the simulation process closely mimics the DGP for $\Delta t \rightarrow 0$ and the ergodic distribution of $M(t)$

⁸ The matrix $V^s[f_0(\varphi_0)]$ would be the equivalent of $V[f_0(\varphi_0)]$ for the GMM case, had an analytical expression for the moments $\mathfrak{M}\text{om}(\ell, q)$ of eq. (3.8) or $\mathfrak{M}\text{om}(\ell, q)$ of eq. (3.9) existed.

is used as the initial distribution for the generation of $\tilde{k}(u_{t,\Delta t}^s; \varphi)$. As a result, we may depict the evolution of the process in better detail than when using the empirical sample only. All this makes the estimation of $V^{\mathcal{S}}[f_0(\varphi_0)]$ based upon the Newey-West^[137] HAC estimator of $V[f_{sT}(\hat{\varphi}_{sT})]$ a more reasonable strategy in our case. Note that in this case an initial consistent estimate $\hat{\varphi}_{sT}$ is necessary, obtained possibly in a prior run of the SMM estimator.⁹

With the assumptions for consistency satisfied, asymptotic normality basically follows by the continuous differentiability of $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ with respect to φ , and the ‘geometric’ ergodicity of $X(t)$ (cf. Duffie and Singleton^[130]). As $T \rightarrow \infty$, the estimator $\hat{\varphi}_{sT}$ of eq. (3.12) behaves asymptotically

$$\sqrt{T}(\hat{\varphi}_{sT} - \varphi_0) \xrightarrow{d} N\left(0, G_0^{-1} V^{\mathcal{S}}[f_0(\varphi_0)] (G_0^{-1})'\right),$$

where $V^{\mathcal{S}}[f_0(\varphi_0)]$ and $G_0 \equiv \mathbb{E}[\partial \tilde{k}(\cdot; \varphi_0) / \partial \varphi]$ are constant limiting matrices, and $\tilde{k}(\cdot; \varphi_0)$ is taken from the ergodic distribution. Note that $V^{\mathcal{S}}[f_0(\varphi_0)] \equiv \lim_{T \rightarrow \infty} T \cdot \mathbb{V}\text{AR}[f_{sT}(\varphi_0)]$, to which $V[f_{sT}(\hat{\varphi}_{sT})] = T \cdot \mathbb{V}\text{AR}[f_{sT}(\hat{\varphi}_{sT})]$ converges.¹⁰

G_0 is approximated in our implementation by finite (central) differences of $f_{sT}(\hat{\varphi}_{sT})$, where the accuracy of this approximation will depend on the magnitude of the step-size ϵ_T used to take the differences. Note that ϵ_T depends on T . The resulting G_T is consistent for $\epsilon_T \rightarrow 0$ and $\epsilon_T^{-1} = O(\sqrt{T})$ as $T \rightarrow \infty$ (cf. Newey and

⁹ We have also considered the use of non-parametric bootstrap methods to improve the finite sample properties of Ω_T . In our context though, their implementation would not be without problems given the large block-size required in the resampling algorithm to capture the long temporal dependence of the process (cf. Brown and Newey^[138], and Inoue and Shintani^[139] for some studies on bootstrapping for GMM, and Winker *et al.*^[140], and Franke and Westerhoff^[141] for applications of bootstrap-based SMM to time series).

¹⁰ As in Duffie and Singleton^[130], $\mathbb{V}\text{AR}[f_{sT}(\hat{\varphi}_{sT})] = (1 + \frac{1}{S}) \mathbb{V}\text{AR}[f_T(\hat{\varphi}_T)]$, which arise from the fact that simulations instead of analytical expressions are used for the moments.

McFadden^[142]).

We turn at last to the optimization algorithm. For this, note the main criterium for the selection of our moments is their capacity to capture in great detail the fluctuations generated by $M(t)$ in its course along the hidden states. This arises, in particular, from the application in the moments $\mathfrak{Mom}(\ell, q)$ of the overlapping log-differences $\zeta(t, \ell)$, or $\xi(t, \ell)$ respectively. Note furthermore that the parameters involved in the generation of scenarios are $(\varphi, b, \gamma_{\bar{k}})$, the cascade levels \bar{k} , and the control variable \mathcal{S} . Figure 3.1 illustrates the sampled $\Psi_{\mathcal{S}T}(\varphi)$ of eq. (3.12) for a given Δt . One can see that for certain constellations of these parameters, $\Psi_{\mathcal{S}T}(\varphi)$ may appear to be non-smooth with respect to m_0 . This may contradict at first sight the aforementioned statement regarding the continuous differentiability of $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ with respect to φ . To show this is not the case, consider in Figure 3.1 the top and bottom panels on the left. The bottom panel depicts a much smoother surface $\Psi_{\mathcal{S}T}(\varphi)$ despite presenting a larger number of cascade levels \bar{k} . The reason for this is that any additional component $M_k(t)$ enters with a proportionally smaller probability of switching in a finite sample. As such, the behavior of $\Psi_{\mathcal{S}T}(\varphi)$ is governed by the random draws based on the switching probabilities of $M_k(t)$ rather than on the parameter governing the distribution M (cf. Appendix 3.A). Noteworthy is finally that when moving from the left to the right panels, some of the jumpy behavior of $\Psi_{\mathcal{S}T}(\varphi)$ is reduced with a sufficiently large \mathcal{S} .¹¹ The actual size of the reduction will depend in the end on the parameter constellation present, making a sufficiently large size of \mathcal{S} difficult to specify beforehand.

¹¹ The requirement of a large \mathcal{S} for the reduction of the sampling variability of our procedure may also contribute to improve the finite sample properties of Ω_T .

The non-smoothness of $\Psi_{sT}(\varphi)$ in our optimization problem requires for the solution an algorithm that seeks an interior extremum without recurring to gradients. Under these circumstances, researchers often recur to slow but robust algorithms like the Nelder-Mead simplex method. As shown in Figure 3.1, the non-smoothness of $\Psi_{sT}(\varphi)$ may be sufficiently reduced for a sufficiently large \mathcal{S} , which inclines us to use methods like Powell's algorithm (cf. Press *et al.*^[143]). Powell's method is a conjugate-direction set approach that requires a one-dimensional minimization sub-algorithm such as Brent's method. For a problem with a small number of parameters, this method may usually prove to converge faster when some 'smoothness' in the objective function is present. An implementation of the Powell's algorithm for the software package Gauss can be found on Mark's^[144] website. We have extended in our case the mentioned implementation to allow for multithreading in a multi-core processor with the software package GAUSS v.11.

3.6 Simulation Study

We proceed in this section with a Monte Carlo simulation analysis of the suggested estimation procedure. Admittedly, results will refer to the discretized version of the continuous-time model with discretization step Δt , but will be asymptotically valid for $\Delta t \rightarrow 0$.

Our starting point consists in the generation of artificial data via the simulation of the MSM process under the discretization scheme of Appendix 3.A. Note that it would also be convenient at this stage to emulate different sampling frequencies

encountered in empirical data. In practice, one observes that governmental and other official institutions recollect for instance data once a day, while in almost all financial asset markets data recollection is available at nearly millisecond intervals.

In a preliminary set of Monte Carlo runs, we explore how the GMM estimator based on the misspecified discrete MSM model performs, when applied to discretely sampled data from the continuous-time process. In this analysis, we will set $k = 8$, $\sigma_1 = 1$, and consider Bernoulli factor values $m_0 = 1.3, 1.4$, and 1.5 . The artificial data will be generated according to $\Delta t = 1, 0.1$, and 0.01 . A step-size $\Delta t = 1$ recovers then the discrete-time MSM fluctuation as a relatively crude approximation to the continuous-time MSM. For $\Delta t < 1$, high-frequency returns from the sampled ‘continuous’ process have been capitalized to obtain a ‘daily’ return measurement like r_t in eq. (3.6) (cf. Andersen *et al.*^[34]).¹²

Table 3.1 and 3.2 presents the GMM results based on the moments of the discrete MSM model of Lux^[8], for a sampled record of the continuous-time process generated according to $\gamma_{\bar{k}} = 3$ and $\gamma_{\bar{k}} = 0.69315$, respectively.¹³ These moments are the equivalent to $\mathfrak{M}\mathfrak{om}(\ell, q)$ in eq. (3.8) (together with the raw moment $\mathbb{E}[R(t)^2]$), where $R(t)$ is replaced by the daily discrete-time return r_t (cf. eq. (3.A.1) in Appendix 3.A for a detailed description of these moments in continuous time). GMM was executed using lags $\ell = 1, 5, 10, 50$ for the observation sample lengths $T_1 = 2, 500$, $T_2 = 5, 000$, and $T_3 = 10, 000$. For each case, 400 Monte Carlo runs

¹² Note also that from the property of log-returns and the structure of the Euler scheme of eq. (3.A.5), the sampled daily return obtained from the difference of daily prices $r_t \equiv p_t - p_{t-1}$ will coincide in our Monte Carlo procedure with the return r_t of eq. (3.6) obtained from aggregation of the high-frequency increments when $\Delta t < 1$.

¹³ For $\Delta t = 1$, $\gamma_{\bar{k}} = 3$ leads to a switching probability of 0.95 in the discrete-time MSM (cf. Calvet and Fisher^[7]). $\gamma_{\bar{k}} = 0.69315$ implies, on the other hand, a switching probability of 0.5 over the same time-step (Lux^[8], Lux and Morales-Arias^[97] and Lux *et al.*^[9]).

have been carried out. Table 3.1 and 3.2 present the corresponding means of the estimated parameters as well as the finite sample standard error (FSSE) and root mean squared error (RMSE), respectively.

Results in Table 3.1 and 3.2 for $\Delta t = 1$ are in harmony with previous studies of discrete-time MSM (cf. Calvet and Fisher^[7] and Lux^[8]) in that the finite sample bias for the correctly specified discrete model ($\Delta t = 1$) uniformly vanishes at a $T^{1/2}$ rate when doubling the sample size from T_1 to T_2 , and from T_2 to T_3 . As the length of Δt is reduced, however, we are attempting to estimate a continuous-time MSM model using the misspecified moments of its discrete-time analogue. The bias obtained in the parameter estimates \hat{m}_0 is now the product of misspecification as well as of finite sampling, and as such, it can only diminish to a certain degree. This is captured for $\Delta t = 0.1$ and 0.01 in Table 3.1, for instance, where one can observe that the mean estimates \bar{m}_0 under T_3 are even more biased than \bar{m}_0 under T_1 for $\Delta t = 1$. Though not homogeneously, violations to the square-root consistency rule are also perceivable in these settings. In Table 3.2, a deterioration is also perceivable in the estimator \bar{m}_0 when comparing the results for $\Delta t = 0.1$ and 0.01 to those for $\Delta t = 1$ with $m_0 = 1.5$, and to a lesser degree with $m_0 = 1.4$. In this case, however, the effect is more moderate due to the lower degree of fluctuations generated by a process with $\gamma_{\bar{k}} = 0.69315$. Noteworthy is finally that the deterioration of the average estimate does not necessarily increase from $\Delta t = 0.1$ to $\Delta t = 0.01$.

Table 3.1 and 3.2 also exhibit the results for $\bar{\sigma}_1$. As one can see, GMM recovers a very precise estimate of σ_1 regardless of the Δt employed in the generation of the data. FSSE and RMSE values for $\Delta t = 0.1$ and 0.01 are also very close to those from $\Delta t = 1$. The fact that σ_1 is almost exactly recovered from the discrete model

suggests that it could be useful to exploit some of the information from estimating the simple discrete-time model. Note also that the fact that the parameter estimates do not vary much between $\Delta t = 0.1$ and 0.01 may provide some indirect evidence that we can achieve a good approximation to the continuous-time dynamics under $\Delta t = 0.1$. To enhance computational speed, we will then select for all future simulation exercises $\Delta t = 0.1$ for the generation of the artificial data as well as for the computation of the simulated moments.¹⁴

Before turning to the general evaluation of the SMM procedure, we provide some insights into its practical implementation. When adopting as a weighting matrix $\Omega_T = [V[f_{sT}(\widehat{\varphi}_{sT})]]^{-1}$ and employing the Newey-West^[137] HAC estimator to compute $V[f_{sT}(\widehat{\varphi}_{sT})]$, the traditional SMM method consists in a two-step algorithm. In a first step, one may solve $\Psi_{sT}(\varphi)$ of eq. (3.12) with $\Omega = \mathbb{I}$, where \mathbb{I} is the identity matrix. $\Psi_{sT}(\varphi)$ in the second step is then recomputed using Ω_T obtained from the first step. Preliminary work with simulated data on this two-step approach has shown that if the procedure is allowed to run further, the initialization of the algorithm with the identity matrix leads to an estimate that must be updated on average three and a half times until convergence of both parameters and Ω_T is obtained. Note that while this is not necessary for consistency, it may indicate that there is some room for improvements in finite samples. Additional experiments have shown that using as an initial weighting matrix the inverse of the variance-covariance matrix of the (misspecified) GMM estimator from the discrete-time MSM model of daily returns provides a valuable improvement in computational time. Indeed, only 2.25 iterations on average are required in this

¹⁴ In his study of the traditional SV model, Benzoni^[37] also applies a $\Delta t = 0.1$ when simulating.

case for convergence of both parameters and weighting matrix. Table 3.3 exhibits the results of this comparison, with SMM_1 denoting a procedure initialized using the weighting matrix resulting from the prior application of GMM and SMM_2 one initialized using the identity matrix. The data was generated for a process with $k = 8$, $(m_0, \sigma_1, b, \gamma_k) = (1.4, 1, 2, 3)$, and a sample length of $T = 2,500$. For the generation of the moments, $\mathcal{S} = 150$ simulated samples have been employed. For each case, a total of 400 Monte Carlo runs has again been carried out. Apart from the lower number of iterations needed, SMM_1 results for parameter m_0 are on average much less biased than those of SMM_2 . Furthermore, the initialization with a weighting matrix from the prior application of GMM leads to a reduction in the dispersion of the estimates of m_0 . On the contrary, \bar{m}_0 , FSSE, RMSE from SMM_2 are not better than those from (misspecified) GMM in Table 3.1. Provided that the time required for the initial GMM estimator of the auxiliary discrete-time MSM model of daily returns is negligible in comparison to a full round of SMM iteration, a good initial ‘hint’ from GMM for the starting values of SMM proves very valuable. In summary, the mixed iterative SMM, initialized with a weighting matrix resulting from the prior application of GMM, delivers better results in a more efficient way behaving on average almost as a two-step algorithm.

Additional experiments have been conducted for further reduction of computational time and of the sample variability of the simulations employed in $f_{\mathcal{S}T}(\varphi)$ in eq. (3.11). When applying the resulting estimate $\hat{\sigma}_1$ from GMM as a starting value for the SMM optimization, for instance, faster convergence is achieved. We will thus employ this estimate as a starting value in all our further analyses. On the other hand, we have found no consistent decreases in sample variation of the

simulator when a simple variance reduction technique like antithetic variates was employed.

Table 3.4 and 3.5 present the SMM results for a process with $k = 8$, $\sigma_1 = 1$, and Bernoulli factors $m_0 = 1.3, 1.4$, and 1.5 . Once again, we employ the moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.8) together with $\mathbb{E}[R(t)^2]$, whereas $\mathcal{S} = 150$ simulated samples have been employed for their generation. As one can observe, \bar{m}_0 improves in terms of variance reduction and RMSE when going from T_1 to T_3 in both tables. Global results are in general encouraging, showing that a large \mathcal{T} effectively leads to low FSSE and RMSE figures. Under $\gamma_{\bar{k}} = 3$, we obtain slightly smaller FSSE and RMSE figures. The reason is that there is a larger probability of switches at all levels, which may allow for a better detection of changes in the vector $M(t)$. Further inspection of individual runs show that in only a few cases SMM ends up in boundary solutions, in contrast to GMM. This is most noticeable for $m_0 = 1.3$, where GMM for the discrete model has been reported to be biased downwards (cf. Lux^[8]). Furthermore, only a very small number of runs have finished in error (and their results consequently been dropped out), due to nonconvergence or breakdown of the estimation, corroborating the convenient behavior of the SMM procedure at the given size of \mathcal{S} . We note finally that between both tables results for $\bar{\sigma}_1$ coincide for $m_0 = 1.3$, and 1.4 , whereas for $m_0 = 1.5$ the lower number of switches under $\gamma_{\bar{k}} = 0.69315$ results in a slightly negative effect on the accuracy of this estimator. This source of inaccuracy though, slowly disappears as the sample size increases.

For better comparability with previous studies (cf. Lux^[8], Lux and Morales-Arias^[97], and Lux *et al.*^[9]), we will focus from now on on a process with $\gamma_{\bar{k}} = 0.69315$ only. Table 3.6 presents the results for an increasing number of cascade

components $k = 10, 15,$ and 20 . To preserve space, we restrict ourselves here to only one sample size T_2 . Comparison to the entries from Table 3.5 shows that the estimator is relatively insensitive to the addition of components despite the increased sample variability in these simulation. This may be the result of very infrequent switches of higher cascade steps producing little contributions to the log differences. As in Lux^[8] though, the addition of nearly constant cascade entries makes the estimation of σ_1 more cumbersome as it becomes difficult to distinguish between unconditional variability, σ_1 , and long-lived high-level components.

We consider now the estimation of the model with Lognormal cascade components $M_k(t)$. For a better comparison with the results obtained so far, we start with the set of moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.8) and $\mathbb{E}[R(t)^2]$. In a separate estimation exercise, we focus on the ‘direct’ estimation of the latent volatility process $\sigma(M(t))$ of eq. (3.5) based on the alternative set of moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.9) together with $\mathbb{E}[[R](t)]$ (cf. Appendix 3.A for more details). To this end, we employ in our experiments high-frequency simulations based on $\Delta t = 0.1$ and compute ‘daily’ returns r_t (cf. Table 3.7) as well as the aggregation of intra-daily squared returns into a measure of RV (cf. Table 3.8). Note that we have used the same set of random numbers to generate both samples of artificial return and RV series in each Monte Carlo run. As a result, the differences in the performance of the SMM estimator between Tables 3.7 and 3.8 are entirely due to the estimation approach. The other variables used in the simulation of the process are the number of cascade levels $k = 8$, and parameter values $\sigma_1 = 1$ and $\lambda_0 = 0.05, 0.1,$ and 0.15 .

The results in Table 3.7 for the estimation of the return process are encouraging,

with small fluctuations of $\bar{\lambda}_0$ around the true λ_0 for $\lambda_0 = 0.05$, when moving from T_1 to T_3 . FSSE and RMSE figures are, as expected, slightly higher than those published by Lux^[8] for the discrete-time MSM, due to the higher variability arising from the simulation of a continuous-time Lognormal cascade in eq. (3.11) and the lower precision of SMM compared to GMM. As in the Bernoulli case $m_0 = 1.5$ of Table 3.5, we observe also a slight bias in $\bar{\sigma}_1$ for $\lambda_0 = 0.15$ at T_1 . The latter, however, is of no serious concern as it smoothly disappears as the sample size increases.

We turn finally to the results of Table 3.8. The calculation of the moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.9) (together with $\mathbb{E}[[R](t)]$) follows from the daily increments $[R](t)$ of eq. (3.4). As with all previous estimations, we have initialized the SMM algorithm with the outcomes from the prior execution of the (misspecified) GMM estimator, this time using the moments of Lux *et al.*^[9]. One can observe when comparing the figures in Tables 3.7 and 3.8 that using high-frequency data for computing RV measures for unit-time intervals leads to huge gains in efficiency of the resulting estimates. Compared to estimates in Table 3.7 based on simple aggregation of high-frequency returns at unit-time intervals, estimates based upon RV have lower biases (particularly for high λ_0) and FSSE as well as RMSE that are only about one fifth of the size of those for the estimation of the return process. Note that this gain in efficiency is reduced without having to incur higher computational costs. This would be different if intra-daily observations would be used *directly* in the estimation procedure. While this would be perfectly sensible and computationally straightforward (as we could extract discrete observations from our continuous-time MSM at any available time horizon), the inflation of the sample size would make the SMM estimation more costly both in Monte

Carlo runs and empirical applications. Table 3.8, thus, demonstrates that using RV provides a very convenient avenue to improve parameter estimates at small additional computational costs.

3.7 Empirical Results

We consider in this section the analysis of empirical time series for two different data sets. Application of our mixed iterative SMM procedure proceeds as in the previous section. This time, several starting values are used for the intermittency parameters m_0 and λ_0 in the SMM stage to avoid settling down at local minima. We note though that results have tended to be rather insensitive to these initial guesses in the end.

We first analyze daily observations from foreign exchange markets. This data set covers only one observation per asset per day consisting of the latest transaction price. Secondly, we investigate as in Lux *et al.*^[9] aggregated intradaily return and RV series from stock indexes. These aggregated figures were calculated according to Andersen *et al.*^[34], that is, with a 30-min interval for the calculation of intraday returns.¹⁵ In all cases, we use a $\Delta t = 0.05$ for the simulation of the underlying DGP.

We turn first to our analysis based on data from seven foreign exchange rates and the price of gold. The currencies are: the Canadian Dollar (CND), the Japanese Yen (YEN), the Swedish Krona (SEK), and the Swiss Franc (CHF), the Australian Dollar (AUD), the Deutsche Mark- extended by the EURO since 1999-

¹⁵ Up to this point, we only possess the aggregated figures of daily return and RV series. Access to the raw data may allow us in the future to consider alternative intervals for data aggregation.

(DEM/EUR), and the British Pound (UKP), all against the U.S. Dollar. Finally, the price series of gold that is also quoted in U.S. Dollars. All time series consists of 5,500 daily observations starting on the 2nd of January of 1979 (cf. Chapter 2).

Table 3.9 reports estimates for the intermittency parameters m_0 and λ_0 , together with their standard errors and the value of the test for overidentifying restrictions $J_T = T \frac{S}{S+1} \Psi_{sT}(\hat{\varphi}_{sT})$, where the estimation procedure has been repeated for $k = 5, 10, 15, 20$ (cf. Carrasco and Florens^[131]). Results correspond to the set of moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.8) together with $\mathbb{E}[R(t)^2]$. First of all, we notice that for a given series, estimates fluctuate somewhat for different k . This contrasts with the behavior of the GMM estimator of the discrete-time MSM, for which results practically do not change at all beyond $k = 10$ (cf. Lux^[8]). The reason for this resides in that the increasing number of cascade steps introduces additional sampling error via $\frac{1}{sT} \sum_{s=1}^S \sum_{t=1}^T \tilde{k}(u_{t,\Delta t}^s; \varphi)$. Underlined entries of the J -test statistics indicate at a significance level of 0.05 those cases for which the null-hypothesis of the continuous-time MSM being the DGP cannot be rejected. We find that on the base of our moment functions and time-step Δt , the continuous-time MSM model can only be rejected for the Canadian Dollar (CND) for two out of four specifications, and in no case for the other series.

It is particularly interesting that Table 3.9 shows no rejection on the continuous-time MSM as the DGP of the Swiss Franc (CHF), which somehow differs from previous results obtained in Chapter 2 with the discrete-time Lognormal cascade or in other studies with the discrete-time MSM (cf. Lux^[8]). Application of GMM under the assumption of $\Delta t = 1$ (1st-step) shows, however, that the discrete-time MSM cannot be rejected at any k either (not shown here). The difference in

the theoretical construction of the model with respect to the Lognormal cascade (cf. Chapter 2), and the inclusion of 900 more observations in the data set in comparison with the number of observations employed in the prior study of the discrete-time MSM (cf. Lux^[8]) seem therefore to be the sources of the perceived inconsistencies of results. Note, finally, the very close agreement of goodness-of-fit for the Bernoulli and Lognormal model.

We turn now to the aggregates of intradaily returns and volatility measures from stock indexes as in Lux *et al.*^[9]. This data set comprises three European indexes, namely, the DAX, the FTSE100, and the CAC40, and two US indexes, the NYSE Composite and the S&P500. The DAX and the FTSE100 include the most highly capitalized companies in Germany and Great Britain, respectively. The French CAC40 index is composed of the 40 largest companies listed in Euronext Paris in terms of order book volume and market capitalization. The NYSE Composite tracks all companies listed on the New York Stock Exchange, whereas the S&P500 includes the largest capitalized actively traded companies in the NYSE and the NASDAQ markets. The available sample sizes of daily observations are $T=2614$ for the DAX (for the period 04/01/1999-30/04/2009), $T=2445$ for the CAC40 (for the period 13/06/2000-09/02/2010), $T=1637$ for the FTSE100 (or the period 01/07/2003-04/01/2010), $T=5682$ for the NYSE Composite (for the period 02/01/1987-09/02/2010), and $T=6692$ for the S&P500 (for the period 01/02/1983-04/01/2010). The relatively small size of the FTSE100 sample leads us to consider the following lags $\ell = 1, 5, 10, 20$ for the moments. We refer the reader to Lux *et al.*^[9] for a detailed description of data sources, composition, and other particulars.

From now on, we will consider the Lognormal MSM model only. The choice is

based upon the descriptive statistics for the stock index data (cf. Lux *et al.*^[9]) and, in particular, on the fact that the distribution of the logarithmic RV_t can be well approximated by a Gaussian distribution (cf. Andersen *et al.*^[34]). Table 3.10 reports parameter estimates for the intermittency parameter λ_0 , their standard errors, and the value of the test for overidentifying restrictions J_T . The estimation procedures have been repeated for $k = 5, 10, 15, 20$ for better comparison.

We first direct our attention to the results on the right-hand side of Table 3.10, arising from the analysis of aggregate intradaily return series. We recall that these results correspond to the set of moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.8) together with $\mathbb{E}[R(t)^2]$. As in the case of foreign exchange series, the estimation outcomes seem to fluctuate quite a bit as k increases. Furthermore, at a significance level of 0.05 the underlined entries for the DAX, CAC40, and the FT100 indicate that the null-hypothesis of the continuous-time MSM being the DGP cannot be rejected. In contrast, the two US indexes reject the continuous-time MSM model even at the 0.01 significance level. These results could be somewhat expected provided that a strong negative skewness is a well-known feature characterizing the latter two indexes (cf. Lux *et al.*^[9]). Note that this attribute of the US indexes have prompted many authors to include additional sources of fluctuation in the return process like asymmetric shocks or unpredictable negative jumps (cf. Chernov *et al.*^[38] and Garcia *et al.*^[51]).

We consider now the results on the left-hand side of Table 3.10, based on the RV series. These results correspond to the set of moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.9) together with $\mathbb{E}[[R](t)]$. We perceive for all assets strong rejections of the J_T statistic, significant at the 0.01 level. In the case of the two US indexes, this may have been expected. Note that as a traditional source of pronounced neg-

ative skewness unpredictable jumps generate a nonzero. Remarkable is, on the other hand, the rejection for the European indexes. The SMM estimator applying the moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.9) hence appears more sensitive to differences between the volatility dynamics of the MSM model and the ‘observed’ return variation in the empirical samples.¹⁶ That is, the more efficient use of intradaily information when considering daily RV rather than raw returns may be better suited to detect differences between the data and the hypothesized DGP. Admittedly, the rejections on the left-hand side of Table 3.10 may just be an indication that, like any model, the present one would be rejected when a sufficient amount of data is available. These rejections may alternatively be the product of misspecification arising from unmodeled sources of return fluctuation found in the empirical data. Recall that the MSM model in this chapter has a zero mean and sets all autocorrelations of the return process to zero. This leads us to adjust in our implementation the sample daily return series by filtering out the empirical mean and return autocorrelations. The RV sample on the other hand allows no such filtering on the daily basis. Hence, the effect caught in intradaily raw returns from a potential drift in the DGP is added up to the daily frequency in the former case but is later adjusted –at least partially. The same drift, however small in size, is first squared before being summed up to RV_t , without posterior potential adjustment. When the sample series include severe market crisis periods, as the samples in our study do, the outcome of squaring this drift may no longer be negligible anymore. Further analysis of the high-frequency observations in this data set should shed light on this issue in the future.

¹⁶ Note that this does not necessarily imply that an MSM model based upon RV would be inferior in a task like forecasting to the MSM based upon squared returns (cf. Lux *et al.*^[9]).

3.8 Conclusion

We have proposed in this chapter a SMM approach with two alternative sets of moments for the estimation of the continuous-time MSM model with finitely many levels. The model fits into the broad class of SV models, but conceives volatility as a hierarchical process with a ‘cascade’ of different layers. As it lacks a closed-form solution for its transient density, a discretization scheme has to be applied for estimation and forecasting.

We show that the performance of the SMM estimator can be improved upon by using a two-step mixed SMM approach for the estimation of the continuous-time MSM. The first step consists in the application of Lux’s^[8] GMM procedure. Results from this step are incorporated as starting values in an iterative SMM algorithm, which is left running until convergence of both the parameter estimates and the weighting matrix is obtained. The iterative nature of the algorithm may appear computationally costly. We find, however, that due to the essentially costless introduction of good initial conditions via the GMM procedure, final convergence of the algorithm ensues on average already after the second iteration.

We conducted a thorough Monte Carlo analysis for the evaluation of our estimation methodology. We started with an analysis of the loss in accuracy incurred when employing the misspecified discrete-time MSM model to the incomplete record generated by the discretely sampled ‘continuous’ time counterpart. After that, we have evaluated the performance of the SMM methodology under the two alternative sets of moments. Simulations show that our SMM implementation is able to deliver reliable results both for data based on aggregated returns and data based on the ‘observable’ RV_t . It also turned out that using high-frequency

information via computation of RV provides a computationally convenient way to increase efficiency upon the performance of estimates based on data aggregated over larger time steps (e.g. daily).

We have also explored the applicability of the continuous-time MSM model empirically. For this, we have tested the goodness of fit of the SMM estimator on foreign-exchange and stock-index time series. In case of the set of moments based on the absolute value of the continuously compounded daily-return process $R(t)$, the continuous-time MSM model cannot be rejected as the DGP of the foreign-exchange time series and the DAX, CAC40 and FT100 time series. Having said that, we obtain strong rejections for all stock indexes when extracting information from the daily RV series. This indicates that the SMM estimator based on the second set of moments appears more sensitive to the differences between the return variation of the observed samples series and the modeling assumptions of the MSM model. Note that the latter result might also be the consequence, at least partially, of inaccuracies arising from unmodeled sources of fluctuations present in the DGP of the empirical data.

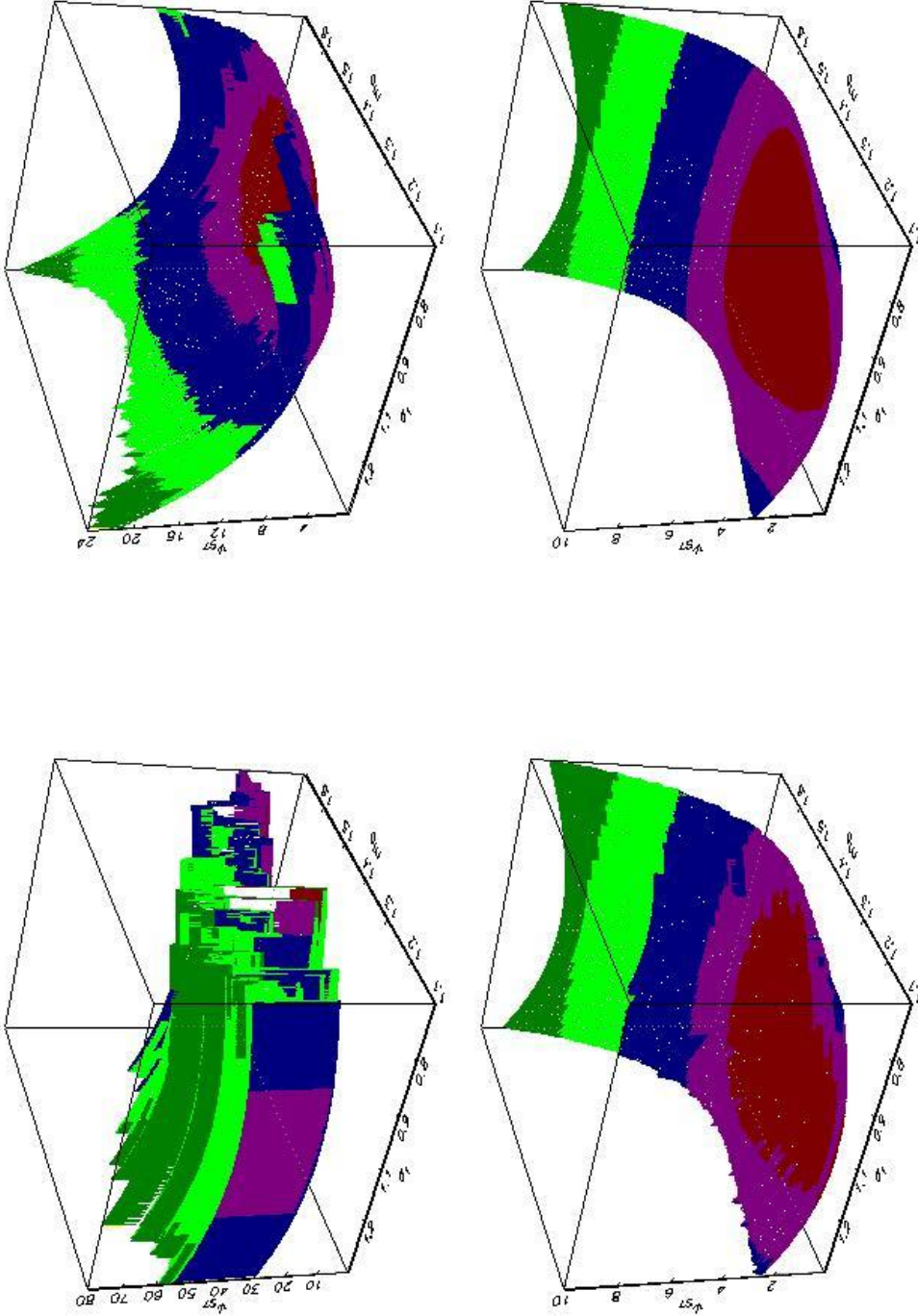


Figure 3.1: Objective function $\Psi_{s\tau}(\varphi)$ for $\bar{k} = 3$ and $S = 10$ -top left-, $\bar{k} = 8$ and $S = 10$ -bottom left-, and $\bar{k} = 8$ and $S = 150$ -bottom right-. The artificial data were generated based on $\Delta t = 0.1$ and a MSM process with Bernoulli distributed factors $M_k(t)$, where $(m_0, \sigma_1, b, \gamma_k) = (1.4, 1, 2, 3)$. $\Psi_{s\tau}(\varphi)$ includes the moment conditions $\mathfrak{M}_{\text{om}}(\ell, q)$ of eq. (3.8) for $q = 1, 2$, together with $\mathbb{E}[R(t)^2]$, where warmer colors indicate lower values. In the top-left panel, the surface $\Psi_{10\tau}$ shows frequent jumps of large size with respect to m_0 . In the top-right panel, jumps are of a milder size and a certain convexity is obtained with $S = 150$. In the case of $\Psi_{10\tau}$ and $\bar{k} = 8$, we depart from jumps of milder size in comparison to the extreme case of $\bar{k} = 3$ though uneven areas can still be found. Finally, the surface $\Psi_{150\tau}$ in the bottom-right panel shows very mild jumps and a much smoother shape from every starting point towards the minimum.

Table 3.1: Monte Carlo results for GMM estimation based on discrete-time MSM, for different discretization steps Δt of the continuous-time MSM with Bernoulli factors ($\gamma_{\bar{k}} = 3$).

Δt		$m_0 = 1.3$			$m_0 = 1.4$			$m_0 = 1.5$		
		T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
1	\bar{m}_0	1.275	1.295	1.306	1.386	1.395	1.400	1.488	1.498	1.499
	FSSE	.085	.048	.030	.057	.032	.023	.044	.026	.017
	RMSE	.089	.049	.031	.058	.032	.023	.045	.026	.017
	$\bar{\sigma}_1$	0.998	1.000	1.001	0.999	1.001	1.000	1.001	1.001	1.000
	FSSE	.047	.034	.022	.065	.047	.031	.087	.058	.041
	RMSE	.047	.034	.022	.065	.047	.031	.087	.058	.041
0.1	\bar{m}_0	1.239	1.257	1.265	1.336	1.351	1.356	1.433	1.442	1.449
	FSSE	.091	.057	.041	.067	.039	.025	.050	.031	.021
	RMSE	.110	.071	.054	.093	.062	.051	.083	.066	.055
	$\bar{\sigma}_1$	0.998	0.998	1.001	1.001	1.002	0.999	1.005	0.998	0.998
	FSSE	.048	.033	.022	.061	.043	.031	.081	.054	.040
	RMSE	.048	.033	.022	.061	.044	.031	.081	.054	.040
0.01	\bar{m}_0	1.241	1.260	1.265	1.336	1.350	1.356	1.430	1.440	1.446
	FSSE	.092	.064	.036	.067	.042	.027	.047	.030	.020
	RMSE	.110	.075	.050	.093	.065	.052	.084	.067	.058
	$\bar{\sigma}_1$	1.000	1.001	1.001	0.998	1.001	1.002	1.002	0.998	0.999
	FSSE	.043	.032	.022	.064	.039	.031	.084	.057	.039
	RMSE	.043	.032	.022	.064	.039	.031	.084	.057	.039

NOTE: All simulations are based on a process with $k = 8$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. \bar{m}_0 and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters based on the moments from Lux^[8] for a discrete-time MSM. Data has been generated according to Δt , the assumed discretization step-size in the Euler approximation of the underlying continuous-time fluctuation. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. GMM was executed using lags $\ell = 1, 5, 10, 50$. For each case, 400 Monte Carlo runs have been carried out.

Table 3.2: Monte Carlo results for GMM estimation based on discrete-time MSM, for different discretization steps Δt of the continuous-time MSM with Bernoulli factors ($\gamma_{\bar{k}} = 0.69315$).

Δt		$m_0 = 1.3$			$m_0 = 1.4$			$m_0 = 1.5$		
		T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
1	\bar{m}_0	1.245	1.276	1.303	1.371	1.391	1.400	1.476	1.495	1.499
	FSSE	.133	.096	.051	.090	.047	.032	.070	.037	.025
	RMSE	.143	.099	.051	.094	.048	.032	.074	.037	.025
	$\bar{\sigma}_1$	1.001	0.998	1.002	0.997	1.004	0.999	0.994	0.992	0.998
	FSSE	.090	.061	.043	.121	.087	.061	.151	.109	.078
	RMSE	.090	.061	.043	.121	.087	.061	.151	.109	.078
0.1	\bar{m}_0	1.229	1.262	1.280	1.344	1.374	1.383	1.460	1.472	1.474
	FSSE	.137	.100	.055	.111	.054	.035	.064	.040	.027
	RMSE	.154	.107	.059	.124	.060	.039	.075	.049	.037
	$\bar{\sigma}_1$	0.998	1.004	1.002	1.003	0.996	1.000	0.984	0.995	0.995
	FSSE	.090	.064	.046	.122	.085	.061	.154	.109	.083
	RMSE	.091	.064	.046	.122	.085	.061	.155	.109	.083
0.01	\bar{m}_0	1.236	1.261	1.274	1.345	1.369	1.380	1.455	1.473	1.474
	FSSE	.128	.096	.066	.108	.055	.037	.075	.041	.028
	RMSE	.144	.103	.071	.122	.063	.042	.087	.049	.038
	$\bar{\sigma}_1$	0.998	1.000	1.000	0.994	1.005	0.994	0.995	1.001	0.993
	FSSE	.084	.056	.043	0.115	.085	.058	0.150	.097	.076
	RMSE	.084	.056	.043	0.115	.086	.058	0.150	.097	.076

NOTE: All simulations are based on a process with $k = 8$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. \bar{m}_0 and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters based on the moments from Lux^[8] for a discrete-time MSM. Data has been generated according to Δt , the assumed discretization step-size in the Euler approximation of the underlying continuous-time fluctuation. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. GMM was executed using lags $\ell = 1, 5, 10, 50$. For each case, 400 Monte Carlo runs have been carried out.

Table 3.3: Monte Carlo results for SMM estimation under alternative starting weighting matrices.

		$\bar{\varphi}$	FSSE	RMSE	ITER
SMM ₁	m_0	1.366	0.070	0.078	2.248
	σ_1	1.000	0.063	0.063	
SMM ₂	m_0	1.332	0.123	0.141	3.496
	σ_1	1.000	0.062	0.062	

NOTE: All simulations are based on a process with $k = 8$, $(m_0, \sigma_1, b, \gamma_{\bar{k}}) = (1.4, 1, 2, 3)$, and a sample length of $T = 2,500$. $\bar{\varphi}$ is the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. ITER represents the average number of iterations needed for convergence. The same Monte Carlo samples have been used for both estimation strategies. The discretization step-size has been set to $\Delta t = 0.1$. SMM₁ was initialized using the weighting matrix resulting from the prior application of GMM and executed using lags $\ell = 1, 5, 10, 50$ and $S = 150$. SMM₂ was initialized using an identity matrix. The moment conditions $\mathfrak{M}\text{om}(\ell, q)$ of eq. (3.8) for $q = 1, 2$ have been used, together with the raw moment $\mathbb{E}[R(t)^2]$. For each case, a total of 400 Monte Carlo runs have been carried out.

Table 3.4: Monte Carlo results for SMM estimation based on raw discretely sampled data from the continuous-time MSM with Bernoulli factors ($\gamma_{\bar{k}} = 3$).

	$m_0 = 1.3$			$m_0 = 1.4$			$m_0 = 1.5$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
\bar{m}_0	1.289	1.291	1.297	1.366	1.384	1.395	1.474	1.490	1.497
FSSE	.084	.063	.039	.070	.046	.028	.067	.040	.023
RMSE	.085	.064	.039	.078	.048	.029	.072	.041	.023
$\bar{\sigma}_1$	1.001	1.002	0.999	1.000	0.998	1.000	1.003	0.999	0.999
FSSE	.051	.031	.022	.063	.046	.030	.079	.061	.041
RMSE	.051	.031	.022	.063	.046	.030	.079	.061	.041

NOTE: All simulations are based on a process with $k = 8$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. \bar{m}_0 and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. Data has been generated according with discretization step-size $\Delta t = 0.1$ in the Euler approximation of the underlying continuous-time process. The moment conditions $\mathfrak{M}\text{om}(\ell, q)$ of eq. (3.8) for $q = 1, 2$ have been used, together with the raw moment $\mathbb{E}[R(t)^2]$. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. SMM was initialized using results from a GMM approximation and executed using lags $\ell = 1, 5, 10, 50$ and $S = 150$. For each case, 400 Monte Carlo runs have been carried out.

Table 3.5: Monte Carlo results for SMM estimation based on raw discretely sampled data from the continuous-time MSM with Bernoulli factors ($\gamma_{\bar{k}} = 0.69315$).

	$m_0 = 1.3$			$m_0 = 1.4$			$m_0 = 1.5$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
\bar{m}_0	1.254	1.276	1.289	1.359	1.384	1.396	1.479	1.492	1.494
FSSE	.110	.079	.052	.103	.061	.039	.073	.042	.028
RMSE	.119	.083	.053	.111	.063	.040	.076	.043	.028
$\bar{\sigma}_1$	0.996	0.998	0.997	0.999	0.986	0.994	0.978	.979	.987
FSSE	.090	.064	.046	.121	.084	.060	.153	0.107	.082
RMSE	.090	.064	.046	.121	.085	.060	.154	0.109	.084

NOTE: All simulations are based on a process with $k = 8$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. \bar{m}_0 and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. Data has been generated according with discretization step-size $\Delta t = 0.1$ in the Euler approximation of the underlying continuous-time process. The moment conditions $\mathfrak{Mom}(\ell, q)$ of eq. (3.8) for $q = 1, 2$ have been used, together with the raw moment $\mathbb{E}[R(t)^2]$. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. SMM was initialized using results from a GMM approximation and executed using lags $\ell = 1, 5, 10, 50$ and $S = 150$. For each case, 400 Monte Carlo runs have been carried out.

Table 3.6: Monte Carlo results for SMM estimation based on raw discretely sampled data from the continuous-time MSM with Bernoulli factors ($k = 10, 15, 20$).

\bar{k}		m_0				m_0		
		1.3	1.4	1.5		1.3	1.4	1.5
10	\bar{m}_0	1.274	1.380	1.486	$\bar{\sigma}_1$	0.995	0.972	0.979
	FSSE	.073	.060	.046	FSSE	.121	.155	.214
	RMSE	.079	.064	.048	RMSE	.122	.158	.215
15	\bar{m}_0	1.290	1.374	1.488	$\bar{\sigma}_1$	0.961	0.898	0.846
	FSSE	.071	.064	.047	FSSE	.303	.437	.461
	RMSE	.072	.069	.049	RMSE	.305	.449	.486
20	\bar{m}_0	1.270	1.378	1.479	$\bar{\sigma}_1$	0.872	0.799	0.752
	FSSE	.073	.061	.047	FSSE	.404	.561	.661
	RMSE	.079	.065	.052	RMSE	.424	.596	.706

NOTE: Simulations are based on a process with $\sigma_1 = 1$, a sample length of $T_2 = 5,000$, and alternative $k = 10, 15, 20$. \bar{m}_0 and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. Data has been generated according with discretization step-size $\Delta t = 0.1$ in the Euler approximation of the underlying continuous-time process. The moment conditions $\mathfrak{Mom}(\ell, q)$ of eq. (3.8) for $q = 1, 2$ have been used, together with the raw moment $\mathbb{E}[R(t)^2]$. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. SMM was initialized using results from a GMM approximation and executed using lags $\ell = 1, 5, 10, 50$ and $S = 150$. For each case, 400 Monte Carlo runs have been carried out.

Table 3.7: Monte Carlo results for SMM estimation based on raw discretely sampled data from the continuous-time MSM with Lognormal factors ($\gamma_{\bar{k}} = 0.69315$).

	$\lambda_0 = 0.05$			$\lambda_0 = 0.10$			$\lambda_0 = 0.15$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$\bar{\lambda}_0$	0.048	0.050	0.051	0.079	0.091	0.096	0.132	0.145	0.146
FSSE	.029	.021	.016	.046	.033	.020	.046	.029	.019
RMSE	.029	.021	.016	.051	.034	.020	.049	.029	.020
$\bar{\sigma}_1$	1.006	1.007	1.000	1.008	1.000	1.000	1.012	0.997	1.001
FSSE	.102	.062	.047	.141	.108	.076	.204	.134	.099
RMSE	.102	.062	.047	.141	.108	.076	.204	.134	.099

NOTE: All simulations are based on a process with $k = 8$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0$ and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. Data has been generated according with discretization step-size $\Delta t = 0.1$ in the Euler approximation of the underlying continuous-time process. The moment conditions $\mathfrak{M}om(\ell, q)$ of eq. (3.8) for $q = 1, 2$ have been used, together with the raw moment $\mathbb{E}[R(t)^2]$. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. SMM was initialized using results from a GMM approximation and executed using lags $\ell = 1, 5, 10, 50$ and $\mathcal{S} = 150$. For each case, 400 Monte Carlo runs have been carried out.

Table 3.8: Monte Carlo results for SMM estimation based on the QV of the continuous-time MSM ($\gamma_{\bar{k}} = 0.69315$).

	$\lambda_0 = 0.05$			$\lambda_0 = 0.10$			$\lambda_0 = 0.15$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$\bar{\lambda}_0$	0.048	0.048	0.049	0.096	0.097	0.098	0.143	0.145	0.147
FSSE	.004	.003	.002	.008	.005	.004	.010	.007	.005
RMSE	.005	.004	.002	.009	.006	.004	.013	.009	.006
$\bar{\sigma}_1$	1.003	1.006	0.999	1.003	0.995	0.999	1.004	0.992	0.995
FSSE	.100	.061	.046	.134	.104	.073	.184	.126	.092
RMSE	.100	.062	.046	.134	.104	.073	.184	.127	.092

NOTE: All simulations are based on a process with $k = 8$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0$ and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. Data for the synthetic sample of RV_i as well as for the simulated moments has been generated according with discretization step-size $\Delta t = 0.1$ in the Euler approximation of the underlying continuous-time process. The moment conditions $\mathfrak{M}om(\ell, q)$ of eq. (3.9) for $q = 1, 2$ have been used, together with the raw moment $\mathbb{E}[[R](t)]$. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. SMM was initialized using results from a GMM approximation and executed using lags $\ell = 1, 5, 10, 50$ and $\mathcal{S} = 150$. For each case, 400 Monte Carlo runs have been carried out.

Table 3.9: Empirical parameter estimates. Foreign exchange asset returns, discrete daily observations.

	Bernoulli						Lognormal										
	SMM5		SMM10		SMM15		SMM20		SMM5		SMM10		SMM15		SMM20		
	\hat{m}_0 (SE) J_T																
CND	1.275 (.051) <u>15.99</u>	1.228 (.064) 16.79	1.271 (.057) <u>15.07</u>	1.224 (.065) 18.68	0.045 (.004) <u>14.29</u>	0.046 (.006) <u>15.14</u>	0.051 (.008) 16.14	0.034 (.004) 16.05	$\hat{\lambda}_0$ (SE) J_T	0.045 (.004) <u>14.29</u>	0.046 (.006) <u>15.14</u>	0.051 (.008) 16.14	0.034 (.004) 16.05	$\hat{\lambda}_0$ (SE) J_T	0.045 (.004) <u>14.29</u>	0.046 (.006) <u>15.14</u>	0.051 (.008) 16.14
YEN	1.454 (.033) <u>5.49</u>	1.427 (.031) <u>5.25</u>	1.438 (.031) <u>5.51</u>	1.424 (.033) <u>5.15</u>	0.112 (.010) <u>5.52</u>	0.109 (.014) <u>5.56</u>	0.107 (.010) <u>5.14</u>	0.109 (.016) <u>5.26</u>	$\hat{\lambda}_0$ (SE) J_T	0.112 (.010) <u>5.52</u>	0.109 (.014) <u>5.56</u>	0.107 (.010) <u>5.14</u>	0.109 (.016) <u>5.26</u>	$\hat{\lambda}_0$ (SE) J_T	0.112 (.010) <u>5.52</u>	0.109 (.014) <u>5.56</u>	0.107 (.010) <u>5.14</u>
SEK	1.302 (.049) <u>1.83</u>	1.279 (.051) <u>2.34</u>	1.277 (.053) <u>2.31</u>	1.269 (.052) <u>1.99</u>	0.045 (.014) <u>1.95</u>	0.038 (.011) <u>2.34</u>	0.041 (.008) <u>2.02</u>	0.042 (.015) <u>2.13</u>	$\hat{\lambda}_0$ (SE) J_T	0.045 (.014) <u>1.95</u>	0.038 (.011) <u>2.34</u>	0.041 (.008) <u>2.02</u>	0.042 (.015) <u>2.13</u>	$\hat{\lambda}_0$ (SE) J_T	0.045 (.014) <u>1.95</u>	0.038 (.011) <u>2.34</u>	0.041 (.008) <u>2.02</u>
CHF	1.262 (.046) <u>9.17</u>	1.248 (.060) <u>10.64</u>	1.257 (.043) <u>10.30</u>	1.201 (.075) <u>12.42</u>	0.031 (.003) <u>9.26</u>	0.012 (.002) <u>10.92</u>	0.019 (.003) <u>11.51</u>	0.021 (.002) <u>10.27</u>	$\hat{\lambda}_0$ (SE) J_T	0.031 (.003) <u>9.26</u>	0.012 (.002) <u>10.92</u>	0.019 (.003) <u>11.51</u>	0.021 (.002) <u>10.27</u>	$\hat{\lambda}_0$ (SE) J_T	0.031 (.003) <u>9.26</u>	0.012 (.002) <u>10.92</u>	0.019 (.003) <u>11.51</u>
AUD	1.396 (.036) <u>12.06</u>	1.410 (.033) <u>14.31</u>	1.394 (.035) <u>13.00</u>	1.377 (.034) <u>13.94</u>	0.099 (.009) <u>11.39</u>	0.074 (.006) <u>13.06</u>	0.082 (.008) <u>13.13</u>	0.091 (.014) <u>13.71</u>	$\hat{\lambda}_0$ (SE) J_T	0.099 (.009) <u>11.39</u>	0.074 (.006) <u>13.06</u>	0.082 (.008) <u>13.13</u>	0.091 (.014) <u>13.71</u>	$\hat{\lambda}_0$ (SE) J_T	0.099 (.009) <u>11.39</u>	0.074 (.006) <u>13.06</u>	0.082 (.008) <u>13.13</u>
DM	1.410 (.036) <u>5.02</u>	1.382 (.038) <u>4.22</u>	1.400 (.036) <u>5.16</u>	1.388 (.036) <u>4.14</u>	0.093 (.015) <u>4.65</u>	0.085 (.016) <u>4.86</u>	0.083 (.015) <u>4.79</u>	0.081 (.015) <u>4.62</u>	$\hat{\lambda}_0$ (SE) J_T	0.093 (.015) <u>4.65</u>	0.085 (.016) <u>4.86</u>	0.083 (.015) <u>4.79</u>	0.081 (.015) <u>4.62</u>	$\hat{\lambda}_0$ (SE) J_T	0.093 (.015) <u>4.65</u>	0.085 (.016) <u>4.86</u>	0.083 (.015) <u>4.79</u>
UKP	1.356 (.039) <u>12.05</u>	1.325 (.037) <u>12.35</u>	1.340 (.038) <u>11.82</u>	1.328 (.038) <u>11.48</u>	0.071 (.011) <u>11.77</u>	0.062 (.003) <u>10.32</u>	0.062 (.005) <u>11.96</u>	0.061 (.013) <u>11.29</u>	$\hat{\lambda}_0$ (SE) J_T	0.071 (.011) <u>11.77</u>	0.062 (.003) <u>10.32</u>	0.062 (.005) <u>11.96</u>	0.061 (.013) <u>11.29</u>	$\hat{\lambda}_0$ (SE) J_T	0.071 (.011) <u>11.77</u>	0.062 (.003) <u>10.32</u>	0.062 (.005) <u>11.96</u>
Gold	1.476 (.028) <u>11.67</u>	1.472 (.032) <u>12.14</u>	1.464 (.032) <u>12.89</u>	1.446 (.033) <u>12.93</u>	0.145 (.018) <u>11.22</u>	0.137 (.015) <u>12.04</u>	0.117 (.012) <u>13.04</u>	0.127 (.009) <u>12.69</u>	$\hat{\lambda}_0$ (SE) J_T	0.145 (.018) <u>11.22</u>	0.137 (.015) <u>12.04</u>	0.117 (.012) <u>13.04</u>	0.127 (.009) <u>12.69</u>	$\hat{\lambda}_0$ (SE) J_T	0.145 (.018) <u>11.22</u>	0.137 (.015) <u>12.04</u>	0.117 (.012) <u>13.04</u>

NOTE: Empirical estimates of Binomial and Lognormal models are obtained from a SMM with $k = 5, 10, 15, 20$. SMM was initialized using results from a GMM approximation and executed using lags $\ell = 1, 5, 10, 50$ and $S = 150$. In each case, the moment conditions $\mathfrak{M}(\ell, q)$ of eq. (3.8) for $q = 1, 2$, together with the raw moment $\mathbb{E}[R(t)^2]$, were used on a sample length of $T = 5, 500$. Simulated data has been generated according to $\Delta t = 0.05$. SE are the standard errors of the pertinent estimates and the entry J_T gives the value of the corresponding J -test statistics, where the underlined entries indicate when the null-hypothesis cannot be rejected.

Table 3.10: Empirical parameter estimates. Stock index returns.

	Results based on $[R](t)$				Results based on $ R(t) $			
	SMM5	SMM10	SMM15	SMM20	SMM5	SMM10	SMM15	SMM20
DAX	$\hat{\lambda}_0$	0.039	0.040	0.039	0.040	0.028	0.031	0.026
	(SE)	(.003)	(.003)	(.003)	(.003)	(.003)	(.006)	(.010)
	J_T	59.26	57.72	57.71	57.64	<u>13.55</u>	<u>14.14</u>	<u>13.90</u>
CAC40	$\hat{\lambda}_0$	0.036	0.036	0.036	0.036	0.021	0.012	0.013
	(SE)	(.003)	(.003)	(.003)	(.003)	(.003)	(.002)	(.012)
	J_T	53.04	52.42	52.29	52.30	<u>10.00</u>	<u>10.30</u>	<u>10.01</u>
FTSE	$\hat{\lambda}_0$	0.038	0.040	0.040	0.040	0.042	0.019	0.033
	(SE)	(.003)	(.004)	(.004)	(.004)	(.008)	(.008)	(.005)
	J_T	34.14	33.27	33.12	33.30	<u>14.91</u>	<u>14.27</u>	<u>13.32</u>
NYSE	$\hat{\lambda}_0$	0.056	0.055	0.054	0.055	0.041	0.051	0.026
	(SE)	(.002)	(.002)	(.002)	(.002)	(.003)	(.009)	(.007)
	J_T	117.82	116.68	116.83	116.68	39.07	39.55	37.71
S&P500	$\hat{\lambda}_0$	0.052	0.050	0.050	0.050	0.045	0.051	0.041
	(SE)	(.002)	(.002)	(.002)	(.002)	(.008)	(.009)	(.004)
	J_T	130.72	132.10	132.10	132.04	23.51	23.80	24.03

NOTE: Empirical estimates of RV_t and Lognormal MSM models are obtained from a SMM with $k = 5, 10, 15, 20$. SMM was initialized using results from a GMM approximation and executed using lags $\ell = 1, 5, 10, 20$ and $S = 150$. On the right-hand side, the moment conditions $\mathfrak{M}(\ell, q)$ of eq. (3.8) for $q = 1, 2$ have been used, together with the raw moment $\mathbb{E}[R(t)^2]$. On the left-hand side, the moment conditions $\mathfrak{M}(\ell, q)$ of eq. (3.9) together with $\mathbb{E}[|R(t)|]$ have been employed. The sample sizes are $T = 2614$ for the DAX (for the period 04/01/1999-30/04/2009), $T = 2445$ for the CAC40 (for the period 13/06/2000-09/02/2010), $T = 1637$ for the FTSE100 (for the period 01/07/2003-04/01/2010), $T = 5682$ for the NYSE Composite (for the period 02/01/1987-09/02/2010), and $T = 6692$ for the S&P500 (for the period 01/02/1983-04/01/2010). Simulated data has been generated according to $\Delta t = 0.05$. SE are the standard errors of the pertinent estimates and the entry J_T gives the value of the corresponding J -test statistics, where the underlined entries indicate when the null-hypothesis cannot be rejected.

3.A Appendix: Formulation of the Moments

In this section, we provide details on the computation of the moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.8) and $\mathfrak{Mom}(\ell, q)$ of eq. (3.9) with $q = 1, 2$ and ℓ an integer instant among $1, \dots, T - 1$. To this end, we consider two alternative sets of functions based on $\zeta(t, \ell)$ and $\xi(t, \ell)$, respectively:

1. The function $\zeta(t, \ell)$ yields

$$\begin{aligned} \zeta(t, \ell) &= \ln |R(t)| - \ln |R(t - \ell)| \\ &= \frac{1}{2} \ln R(t)^2 - \frac{1}{2} \ln R(t - \ell)^2 \\ &= \frac{1}{2} \ln \left(\int_{t-1}^t \sigma(M(s)) dB(s) \right)^2 - \frac{1}{2} \ln \left(\int_{t-\ell-1}^{t-\ell} \sigma(M(s)) dB(s) \right)^2, \end{aligned} \tag{3.A.1}$$

so that the moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.8) become

$$\begin{aligned} \mathbb{E} [\zeta^q(t + \ell, \ell) \cdot \zeta^q(t, \ell)] &= \\ \left(\frac{1}{4}\right)^q \mathbb{E} &\left[\left[\ln \left(\int_{t+\ell-1}^{t+\ell} \sigma(M(s)) dB(s) \right)^2 - \ln \left(\int_{t-1}^t \sigma(M(s)) dB(s) \right)^2 \right]^q \right. \\ &\times \left. \left[\ln \left(\int_{t-1}^t \sigma(M(s)) dB(s) \right)^2 - \ln \left(\int_{t-\ell-1}^{t-\ell} \sigma(M(s)) dB(s) \right)^2 \right]^q \right]. \end{aligned}$$

Note the impossibility to separate the effect of the Gaussian shocks from the ‘latent’ volatility process in these moments. This feature distinguishes our current setting from that of Lux^[8] for the discrete-time MSM. When high-frequency observations are available, the use of the second set of moments may prove more adequate.

3.A Appendix: Formulation of the Moments

2. Application of $\xi(t, \ell)$ leads to

$$\begin{aligned}\xi(t, \ell) &= \ln [[R](t)]^{1/2} - \ln [[R](t - \ell)]^{1/2} \\ &= \frac{1}{2} \ln [R](t) - \frac{1}{2} \ln [R](t - \ell),\end{aligned}\tag{3.A.2}$$

so that the moments $\mathfrak{Mom}(\ell, q)$ of eq. (3.9) become

$$\begin{aligned}\mathbb{E} [\xi^q(t + \ell, \ell) \cdot \xi^q(t, \ell)] &= \left(\frac{1}{4}\right)^q \mathbb{E} \left[\left[\ln [R](t + \ell) - \ln [R](t) \right]^q \right. \\ &\quad \left. \times \left[\ln [R](t) - \ln [R](t - \ell) \right]^q \right].\end{aligned}$$

Both choices of moments seek to capture the switches of the vector $M(t)$ between the times $t + \ell$ and t , and t and $t - \ell$. In contrast to Lux^[8] though, the independence of the factors $M_k(t)$ for different $k' \neq k$ and different $t' \neq t$ cannot longer be exploited analytically. Despite of this, the choice of different lag sizes will expose the switching nature of the vector $M(t)$ at different frequencies. In the constrained case with b and γ_k fixed, these moments allow us to capture the variability of the factors $M_k(t)$.

Note that the logarithmic differences in both eq. (3.A.1) and eq. (3.A.2) remain unaffected by σ_1 , the scale factor in eq. (3.5) that is typically needed to match the order of magnitude of the data under scrutiny. As a consequence, we include an additional raw moment like $\mathbb{E}[R(t)^2]$ to the moments of eq. (3.A.1), and $\mathbb{E}[[R](t)]$ to the moments of eq. (3.A.2), which will allow for the identification of σ_1 .

Putting all together, the dimension of the vector of moments

3.A Appendix: Formulation of the Moments

$$k(\varphi) = \begin{pmatrix} \mathfrak{M}\text{om}(\ell, q) \\ \mathbb{E}[\cdot] \end{pmatrix}$$

is q times ℓ plus one, where we have used $\mathbb{E}[\cdot]$ to refer to either $\mathbb{E}[R(t)^2]$ or $\mathbb{E}[[R](t)]$.

The next step consists in approximating $k(\varphi)$ above by replacing on the interval $[0, T]$ the random variables $R(t)$ of eq. (3.3) by

$$R_t \equiv \sum_{j=1}^{1/\Delta t} \delta_{\Delta t} X(t-1 + (j-1)\Delta t), \quad (3.A.3)$$

and $[R](t)$ of eq. (3.4) by

$$[R]_t \equiv \sum_{j=1}^{1/\Delta t} [\delta_{\Delta t} X(t-1 + (j-1)\Delta t)]^2, \quad (3.A.4)$$

where the increments $\delta_{\Delta t} X(t) = X(t + \Delta t) - X(t)$, for $\Delta t = \frac{T}{n}$ and $n > 0$, approximate the infinitesimal increment of $X(t)$ between the times t and $t + \Delta t$. Note that the bias of this approximation is relatively small and vanishes for $n \rightarrow \infty$, as it will be argued below.

The lack of closed-form solutions leads us to the use of $\frac{1}{sT} \sum_{s=1}^S \sum_{t=1}^T \tilde{k}(u_{t,\Delta t}^s; \varphi)$ for the approximation of the moments. $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ denotes here the simulator based on the discrete process of eq. (3.A.5) below, with transition law $P_n^{\Delta t}$ that weakly converges to the stationary distribution of the continuous-time MSM as $n \rightarrow \infty$ (cf. Mao *et al.*^[145]). For each s , $u_{t,\Delta t}^s$ is a $1/\Delta t$ -sequence of draws from

3.A Appendix: Formulation of the Moments

the vector distribution of orthogonal elements collecting all random sources used in the simulation of the ‘high-frequency’ log-returns between $t - 1$ and t : the Gaussian increments $\delta_{\Delta t} B(\cdot)$, the \bar{k} switching frequencies $\gamma_k \Delta t$, and the \bar{k} factors M_k (all of them to be specified below).

In the practical implementation of eq. (3.11), we first select for $s = 1, \dots, S$ the number of grid points n and generate subject to eq. (3.A.5) below (and the Markov property of the components $M_k(t)$) a ‘high-frequency’ log-return for each of the resulting segments Δt . According to the definitions of R_t and $[R]_t$ above, we then obtain $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ for $t = 1, \dots, T$ and $s = 1, \dots, S$. Following, we average $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ across T , smoothing out the switches from one state of the embedded Markov to the next, and then across S , in order to obtain the mean from all possible states the vector M_t visits.¹⁷

We turn at last to the description of the discretization scheme for the simulation of the continuous-time MSM. With a number of \bar{k} components, we select a constant η such that on the bounded interval $[0, T]$ we can define a sequence of partitions $\Delta t = \frac{T}{n}$ with $n = \eta^{\bar{k}}$ (cf. Calvet and Fisher^[1]).¹⁸ For fixed η and \bar{k} , the Euler approximation of the price change follows

$$\delta_{\Delta t} X(t) \equiv \sigma(M_{t+\Delta t}) \cdot \varepsilon_{t+\Delta t}, \quad (3.A.5)$$

¹⁷ A numerical issue in the calculation of $\tilde{k}(u_{t,\Delta t}^s; \varphi)$ refers to the fact that on any given t , the simulated \tilde{R}_t may actually turn out to be zero, conflicting with the evaluation of the log moments. This may result from the drawing of an unlucky combination of random numbers, falling within the machine precision around zero. It has though little to do with the estimation procedure per se. We will therefore amend in our implementation the definition of R_t to $R'_t = R_t$, for $R_t \neq 0$, and to $R'_t = \epsilon^2$, for $R_t = 0$, being ϵ a very small positive number.

¹⁸ Note that $\eta \rightarrow \infty$ immediately implies that $n \rightarrow \infty$.

3.A Appendix: Formulation of the Moments

with $\sigma(\cdot)$ depicted as in eq. (3.5), and $\varepsilon_{t+\Delta t} \equiv \delta_{\Delta t} B(t) \stackrel{d}{=} N(0, \sqrt{\Delta t})$. In this setting, the geometric probability of a switch for each of the $M_{k,t+\Delta t}$ factors is $\gamma_{k,\Delta t} = 1 - \exp[-\gamma_k \cdot \Delta t] \approx \gamma_k \cdot \Delta t$, the latter term becoming increasingly more precise as $\eta \rightarrow \infty$. The dynamics of the $M_{k,t+\Delta t}$ are defined by:

$$\begin{aligned}
 M_{k,t+\Delta t} \text{ drawn from distribution } M & \quad \text{with probability } \gamma_{k,\Delta t} \\
 M_{k,t+\Delta t} = M_{k,t} & \quad \text{with probability } 1 - \gamma_{k,\Delta t}.
 \end{aligned}$$

For the issue of convergence, let first M be a discrete distribution with finite first four moments. Mao *et al.*^[145] show that the $P_n^{\Delta t}$ induced by eq. (3.A.5) weakly converges to the stationary distribution of the continuous-time diffusion of eq. (3.1). To see this, note that the continuous-time MSM model with a finite number of volatility states can be classified amidst the family of regime-switching diffusions (cf. Yin and Zhu^[146]). In particular, the MSM model is a diffusion with no feedback influence of $X(t)$ on $\sigma(\cdot)$ nor on the intensities γ_k .

In the case of a Lognormally distributed M , we must recur to more pragmatic arguments. Note that the approximating step Δt does not alter the distribution of the $M_{k,t+\Delta t}$ components but only their switching probabilities. Our prior arguments must therefore remain valid as long as M has finite first four moments and a discrete number of states. If so, consider the discretization of the Lognormally cumulative distribution function (*CDF*) in m bins of equal probability mass in a way that the first four moments of both distributions coincide. Without much loss of precision, one could apply the latter distribution in our investigation and derive the $P_n^{\Delta t}$ as in the Bernoulli case.

Chapter 4

Implied MSM Volatility Forecasting

4.1 Introduction

Since Clark^[24] first introduced a compound process for the modeling of information arrivals, a large set of literature has focused on the mixing-distribution hypothesis of asset returns (cf. Chapter 1). The main argument of this line of research is that the well-established rejection of the Lognormal paradigm is the result of unobserved non-linearities that could be captured by introducing latent state-variables (cf. Breeden and Litzenberger^[147], Hull and White^[48], Hamilton^[94], Amin and Ng^[148], Garcia and Renault^[149], Garcia *et al.*^[150], and Chabi-Yo *et al.*^[151]). Among the most popular examples of models applying this construction are the stochastic volatility (SV) models, which offer a convenient approach for the formalization of time-varying, and possibly random, volatility. The success of these models stems from the possibility to bring together analyti-

cal results from stochastic calculus and inference and forecasting tools from their discrete-time counterparts (cf. Andersen *et al.*^[32], Shephard^[33], and Shephard and Andersen^[23] for a review of the history and applications of SV models).

During the last 20 years, the constantly increasing availability of exchange data regarding derivative contracts has prompted authors towards the inclusion of this information in the study of return fluctuations. Despite a large and increasing number of publications focusing on SV models, a number of unresolved issues still exist for this framework (cf. Ghysels *et al.*^[152], Benzoni^[37], and Garcia *et al.*^[51]). In a nutshell, traditional SV models provide a much better in-sample fit for stock price fluctuations than constant-variance Lognormal processes. They are, however, unable to fully accommodate the perceived jumps and hyperbolic decline in volatility, nor the excess kurtosis caused by large unexpected price changes. Because of these deficiencies, several enhancements have been suggested. Noteworthy are the fractionally integrated processes, jump processes in returns and/or in volatility, and the multiple volatility factor processes (cf. Garcia *et al.*^[51]). The problem with the increasing complexity of these models though, is the arising difficulty in identifying and efficiently estimating the resulting constructions. In their analysis of several of these more complex models, Chernov *et al.*^[38] have concluded, for instance, that even when using a 100-year long series of daily prices some parametric specifications remained unidentifiable.

In another strand of research, authors have started to explore how relevant the information from option prices is for the forecasting of future volatility. Different strategies to this purpose can be found in the early studies of Jorion^[101], Bakshi *et al.*^[102], Christensen and Prabhala^[103], Christensen and Hansen^[104], Fleming^[105], and Blair *et al.*^[106], among others.

When working with parametric models, the advantage of complementing option-price with historical-return time series resides in the possibility to combine forecasts generated from each source of information. This strategy has been employed, for instance, by De Jong and Lehnert^[153] in the context of ARCH-type models. One may argue, on the other hand, that a more efficient way to process information would be to generate only one set of estimates from both sources of information. The general strategy in this kind of study is to update during inference the market premia of price and (jump) volatility risk from one trading day to the next within the combined sample (cf. Calvet *et al.*^[100], Chernov and Ghysels^[36], and Pan^[39]). Although a more coherent approach (and indeed more effective in-sample), one still faces here the risk of disregarding most of the forward-looking information contained in the option prices, provided that in practice only a few near-the-money observations per day can be included in the sample. The use of option prices thence, does not necessarily alleviate the trade-off between obtaining a good in-sample fit and a satisfactory out-of-sample performance.

A new type of SV model that accounts for long-term dependence and jumps in volatility is the recently introduced Markov Switching Multifractal (MSM) model of Calvet and Fisher^[1]. The MSM model embodies volatility as a multiplicative stochastic process of \bar{k} positive random variables with heterogeneous arrival times. As its predecessor the Poisson Multifractal Model (PMM) (cf. Chapter 1 for more details), the model depicts a complex nature of volatility, nonetheless outstanding by its parsimony. Calvet and Fisher^[7] and Lux^[8] have shown that this model often outperforms other well-known candidates like GARCH, FIGARCH and MS-GARCH for various historical return data. In this chapter, we investigate

in turn how informative option prices are in the forecasting of future volatility. We will additionally evaluate how information from the different data sources progressively unfolds in a sample period that includes intervals of normal market activity as well as of financial crisis. We contribute finally to the few available studies focusing on the predictability of long-term volatility based on information retrieved from derivative contracts.

Before turning to the forecasting analysis that is the core of this chapter, we will discuss how the parameters are estimated. Specifically, we are required to implement an inference method based solely upon contingent claims data. We will focus here on a version of the well-known mixing-solution approach (cf. Hull and White^[48], and Romano and Touzi^[154]). In line with the literature, the property required for the fair pricing of contingent claims, absence of arbitrage, will be derived directly from the distributional assumptions of the process, leading to a pricing scheme that is coherent with the Black-Scholes formula (cf. Garcia *et al.*^[51]). Due to the Markov nature of the volatility process, moreover, our pricing strategy can be related to other approaches employing integrals of the Black-Scholes prices over the latent state-variables as a valuation mechanism. Experiments with this approach in the context of traditional SV models show, for instance, that the resulting option pricing formulas are well suited for the application of Monte Carlo tools, where only the alternative paths of the latent state variables need to be simulated (cf. Willard^[155]). The pricing scheme acts therefore as a conditional Monte Carlo variance reduction technique, leading to stable results at a relatively low computational cost.

The rest of this chapter is organized as follows. In section 4.2 we introduce the MSM model for the description of the evolution of asset returns in the usual ‘sta-

tistical' world. Next is derived the risk-neutral version of the model that allows to dispense with the risk-preference parameter under an equivalent martingale measure. The corresponding inference methodologies for both types of processes are presented in section 4.3, together with the forecasting scheme. Thereafter, Monte Carlo results of the calibration approach are presented. Based on the selection of the DAX as the underlying, we discuss in section 4.4 some issues with respect to data preparation for the empirical analysis. In section 4.5 we turn to the empirical evaluation in-sample. We will see that despite the MSM model being a symmetric process with U-shaped smiles, we are able to capture the main differences between the 'statistical' and the 'preference-free' volatilities of the selected underlying.¹ For the out-of-sample analysis, we evaluate up to 100-day-ahead forecasts of quadratic variation (QV), taken from both estimated and calibrated parameter sets. We will also show under what circumstances the inclusion of the risk-neutral volatility into the forecasting scheme will lead to an enhancement in the forecasting performance. In section 4.6 we conclude with some final remarks.

4.2 Model Description

The main goal in this chapter is to compare volatility forecasts based on *ex-ante* and *ex-post* sources of information. To this end, daily as well as intra-daily time series are required. Following Heston and Nandi^[158], we fulfill all these requirements by selecting a stock index for which there is daily information

¹ Note that empirical volatility surfaces may exhibit not only skewed smiles, but also smirks and frowns. There is though some controversy in the literature about the importance of capturing these effects of daily volatility smiles (cf. Renault^[156], and Hilliard and Hilliard^[157]).

on option prices, expected dividends as well as high-frequency data of futures contracts available. This will allow us to recover a dividend-adjusted spot price Y_t as the underlying. The derivation of Y_t can be performed recalling the well-known future-spot relationship formula:

$$F_{t,T} = e^{r_f \cdot (T-t)} \cdot [S_t - PV(D_T)] = e^{r_f \cdot (T-t)} \cdot Y_t, \quad (4.1)$$

where S_t denotes the spot value, $F_{t,T}$ the price at t of the future with maturity T , $PV(D_T)$ the present value of the dividends to be obtained at time T , and r_f the deterministic risk-free interest rate, which may be obtained from a time-dependent rate by $r_f = \frac{1}{T} \sum_{i=t+1}^T r_{f,i}$.

The futures market is one of the most liquid markets available nowadays, and due to its close integration with the derivative markets, it works *de facto* as a channel for the settlement and risk management of derivative contracts. Having said that, the statistical properties of futures time series may diverge from those of the original underlying depending on the official rules for the calculation of the stock index as well as the degree to which dividends can be predicted.

To simplify our analysis, we will focus here on performance indexes like the DAX only. This will allow us to evaluate Y_t and any contingent claim on its future value with just one stochastic process after a deterministic proxy for the expected dividend payments $PV(D_T)$ is calculated.²

² An alternative derivation of Y_t in the literature of financial economics is $Y_t \equiv S_t \cdot e^{-q \cdot (T-t)}$, with q being a known instantaneous dividend yield.

The ‘Statistical’ Process

We proceed with a description of the dynamics of the dividend-adjusted spot price Y_t of eq. (4.1). On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the returns of $X_t \equiv \ln(Y_t)$, for $1 \leq t \leq T$, are assumed to obey the following specification

$$x_t = X_t - X_{t-1} \equiv \mu(M_t) - \frac{1}{2} \sigma^2(M_t) + \sigma(M_t) \cdot \varepsilon_t, \quad (4.2)$$

where $\mu(M_t)$ and $\sigma(M_t)$ are the drift and volatility process, respectively. The sequence $\{\varepsilon_t\}_{t=1}^T$ is i.i.d. standard Gaussian $N(0, 1)$, while $\{M_t\}_{t=1}^T$ is generated by a first-order Markov state-vector with \bar{k} components

$$M_t \equiv (M_{1,t}, M_{2,t}, \dots, M_{\bar{k},t}) \in \mathbb{R}_+^{\bar{k}}.$$

The dynamics of this Markov process M_t follows the setting from Calvet and Fisher^[1] where each component of M_t is drawn from a common marginal distribution M but is updated at different frequencies γ_k . Given the value of the state-vector at $t - 1$, the dynamics of the factors $M_{k,t}$ are

$$\begin{aligned} M_{k,t} \text{ drawn from distribution } M & \quad \text{with probability } \gamma_k \\ M_{k,t} = M_{k,t-1} & \quad \text{with probability } 1 - \gamma_k, \end{aligned}$$

where the switching events and new draws from M are assumed to be mutually independent across k and t . A well-behaved volatility process also requires $M \geq 0$ and $\mathbb{E}(M) = 1$. For convenience, we chose here M to be a Bernoulli distribution, taking two values m_0 or $2 - m_0$ ($1 \leq m_0 < 2$) with equal probability. The state-

4.2 Model Description

vector M_t has consequently finitely many states $m^1, \dots, m^d \in \mathbb{R}_+^{\bar{k}}$, with $d = 2^{\bar{k}}$.

In the MSM framework, the transition probabilities $\gamma \equiv (\gamma_1, \gamma_2, \dots, \gamma_{\bar{k}})$ are hierarchically connected by

$$\gamma_k = 1 - (1 - \gamma_1)^{(b^{k-1})}$$

where $\gamma_1 \in (0, 1)$ and $b \in (1, \infty)$. As $\gamma_1 < \dots < \gamma_{\bar{k}} < 1 < b$ holds, the pair $(b, \gamma_{\bar{k}})$ alone specifies the set of transition probabilities $a_{i,j} = \mathbb{P}(M_{t+1} = m^j | M_t = m^i)$.

The latter are collected in the transition matrix $\mathcal{A} = (a_{i,j})_{1 \leq i,j \leq 2^{\bar{k}}}$, which will govern the dynamics of the Markov vector process.

For the specification of $\mu(M_t)$, we follow Heston and Nandi^[158] using a linear coefficient of risk aversion α and model the drift term as³

$$\mu(M_t) \equiv r_f + \alpha \cdot \sigma^2(M_t). \tag{4.3}$$

To require a high variability of $\mu(M_t)$ may look at first unnecessary. However, it allows to potentially offset the high fluctuation arising from the Itô correction term $-\frac{1}{2}\sigma^2(M_t)$, resulting from the standard Lognormal formulae.⁴ Furthermore, the specification of eq. (4.3) follows the spirit of the well-known GARCH-M and SV-M models (cf. Jondeau *et al.*^[19]).

³ The use of a dividend-adjusted underlying Y_t also makes more complicated schemes like the one in Calvet and Fisher^[112] unnecessary.

⁴ Duan^[159,160] opts for a drift like $\mu(M_t) = r_f + \alpha \cdot \sigma(M_t)$ in turn, which is commonly associated to the CAPM. Our preliminary exploration has shown that little is gained in our case from this alternative. Noteworthy is though that Duan's $\mu(M_t)$ allows for the direct estimation of the market price of risk, albeit a very simple version of it.

The depiction of the stochastic volatility follows Calvet and Fisher^[1], namely

$$\sigma(M_t) \equiv \sigma_1 \left(\prod_{k=1}^{\bar{k}} M_{k,t} \right)^{1/2},$$

where σ_1 is a positive constant. The volatility components of the Markov vector M_t interact multiplicatively, giving rise to a rich spectrum of dynamic evolutions that can to some extent account for long-term persistence of autocorrelation in measures of volatility (squared, absolute returns, etc.) and stochastic jumps in regime switching.

The ‘Risk-Neutral’ Process

With the specifications of $\mu(M_t)$ and $\sigma(M_t)$ in mind, the description of the return process in eq. (4.2) is sufficient for its application to historical time series. The contingent nature of derivative contracts, on the other hand, requires the absence of arbitrage possibilities for their evaluation. It is well known that this assumption is tantamount in technical terms to the existence of a probability measure $\tilde{\mathbb{P}}$ on (Ω, \mathcal{F}) , equivalent to \mathbb{P} , under which discounted price processes are martingales (cf. Harrison and Kreps^[161]). Such a probability $\tilde{\mathbb{P}}$ is unique if and only if all risk sources are tradeable in the market. In the presence of latent state variables, however, additional assumptions regarding risk premia required by the *representative* agent are necessary for the determination of the market price of a derivative contract under $\tilde{\mathbb{P}}$.

In case of the MSM model, the mutual independence of the random variables ε_t and M_t implies that $\mathbb{P} = \mathbb{P}^\varepsilon \cdot \mathbb{P}^M$, that is, the probability measure \mathbb{P} factors

between the measures of ε and of M . This property is connected to the ‘no leverage-effect’ assumption, usually well-suited for foreign-exchange assets and for large performance indexes, to a lesser degree. The set of assumptions described in the following lines will allow us to retain this independence under $\tilde{\mathbb{P}}$, resulting in $\tilde{\mathbb{P}} = \tilde{\mathbb{P}}^\varepsilon \cdot \tilde{\mathbb{P}}^M$.

We consider first that for the events $\mathcal{F}_t \in \mathcal{F}$ the risk-neutral measure $\tilde{\mathbb{P}}$ has the Radon-Nikodym variable $\xi_t \equiv \Lambda_t^\varepsilon \cdot \Lambda_t^M$. The relation between $\tilde{\mathbb{P}}$ and \mathbb{P} is determined by

$$\tilde{\mathbb{P}} = \int \xi_t d\mathbb{P} = \int \int \Lambda_t^\varepsilon \cdot \Lambda_t^M d\mathbb{P}^\varepsilon d\mathbb{P}^M, \quad (4.4)$$

where Λ_t^ε adjusts the drift of x_t , and Λ_t^M influences the size and the arrival times of M_t .⁵ The independence argument allows us also to treat each probability measure separately.

The change from $\tilde{\mathbb{P}}^\varepsilon$ to \mathbb{P}^ε can be performed by characterizing the (positive) process

$$\Lambda_t^\varepsilon \equiv \exp \left[- \sum_{i=1}^t \theta_i^\varepsilon \varepsilon_i - \frac{1}{2} \sum_{i=1}^t \theta_i^{\varepsilon 2} \right], \quad (4.5)$$

where the process θ_t^ε is adapted to the filtration \mathcal{F}_t and satisfies the ‘integrability’ condition $\sum_{i=1}^t \theta_i^{\varepsilon 2} < \infty$. It follows by Girsanov’s theorem, that the process $\tilde{\varepsilon}_t \equiv \varepsilon_t + \theta_t^\varepsilon$ is $N(0, 1)$ distributed under $\tilde{\mathbb{P}}^\varepsilon$, with θ_t^ε being specified after eq. (4.7) below.

⁵ Note that $\int \xi d\mathbb{P}$ is a Lebesgue integral. Over the sample space of ε , the latter calculation can be performed by the usual Riemann integral based on the density function f_ε . Over the finite sample space of M though, the Lebesgue integral must be interpreted as a sum over each element in the space, whereas Λ_t^M is a ratio between the probabilities $\tilde{\mathbb{P}}^M$ and \mathbb{P}^M for each state in the sample space (cf. Shreve^[67,68] and Cont and Tankov^[46] for more details, and Chapter 1 of this thesis for an example of the probability change in a simple finite state binomial model).

The existence of the latent state vector M_t , on the other hand, requires the determination of an additional (positive) process Λ_t^M for the full specification of $\tilde{\mathbb{P}}$. In fact, for the underlying Y any price

$$C_t(T, K) = e^{-r_f \cdot (T-t)} \cdot \tilde{\mathbb{E}}_t^{(\Lambda)} [f(Y_T, K, T-t)] \quad (4.6)$$

is an admissible price process of a plain-vanilla option on Y_T given a suitable choice of Λ_t^M , where we have used $\tilde{\mathbb{E}}_t^{(\Lambda)}[\cdot]$ to denote that the expectation is taken under the unspecified $\tilde{\mathbb{P}}$. The impossibility to directly observe the state variables M_t , however, makes the use of additional requirements necessary for its detailed definition. We will take into account that volatility risks remain uncompensated; the risks in M_t are exogenous, and so, not directly tradable. It follows that a zero market premium for volatility risk, or $\Lambda_t^M = 1$, is embodied.⁶

The above assumption leads us to a well-defined density function ξ_t . This can be verified by computing

$$\mathbb{E} [\Lambda_t^M \cdot \Lambda_t^\varepsilon] = \mathbb{E}^M \left[\mathbb{E}^\varepsilon [\Lambda_t^\varepsilon | \mathcal{F}_t^M] \right] = \mathbb{E}^M \left[\tilde{\mathbb{E}}^\varepsilon [1 | \mathcal{F}_t^M] \right] = \tilde{\mathbb{E}} [1] = 1,$$

where the second term arises by Fubini's theorem, the third by Girsanov's theorem, and $\tilde{\mathbb{E}}$ denotes the expectation with respect to $\tilde{\mathbb{P}}$, with $\tilde{\mathbb{P}}^M = \mathbb{P}^M$.

⁶ Several arguments have been formulated for the support of this assumption in the context of traditional SV models. Pham and Touzi^[162] have shown, for instance, that the choice of a zero market premium for volatility risk is consistent with an intertemporal additive equilibrium with logarithmic preferences. Föllmer and Schweizer^[163], on the other hand, have interpreted this assumption as the one that would provide the shortest 'distance' between the equivalent measures $\tilde{\mathbb{P}}$ and \mathbb{P} (cf. Renault and Touzi^[164]). For comparison, note that under the traditional SV model and provided M is Normally distributed, Λ_t^M would have a similar structure to that in eq. (4.5), with θ_t^M instead of θ_t^ε . Our proposition would consequently lead to $\theta_t^M \equiv 0$, leaving the long-term mean volatility under the risk-neutral measure likewise unaffected.

Once proven how both measures $\tilde{\mathbb{P}}$ and \mathbb{P} are connected, we can focus on the evaluation of $C_t(T, K)$ under $\tilde{\mathbb{P}}$. For this, it remains to show the dynamic of the underlying asset $Y_t \equiv \exp(X_t)$ under $\tilde{\mathbb{P}}$. The returns x_t follow

$$x_t = r_f - \frac{1}{2}\sigma^2(M_t) + \sigma(M_t) \cdot \tilde{\varepsilon}_t, \quad (4.7)$$

where the discounted asset price $Y_T \cdot e^{-r_f \cdot (T-t)}$ is a $\tilde{\mathbb{P}}$ -martingale if and only if $\theta_t^\varepsilon = \theta^\varepsilon(M_t) \equiv \frac{\mu(M_t) - r_f}{\sigma(M_t)}$, for $1 \leq t \leq T$.⁷ The process θ_t^ε is commonly known as the market premium of price risk.

The evaluation of eq. (4.6) can be performed by

$$\begin{aligned} C_t(T, K) &= e^{-r_f \cdot (T-t)} \cdot \tilde{\mathbb{E}}_t [f(Y_T, K, T-t)] \\ &= e^{-r_f \cdot (T-t)} \cdot \mathbb{E}_t^M [\tilde{\mathbb{E}}_t^\varepsilon [f(Y_T, K, T-t) | \mathcal{F}_T^M]] \\ &= \mathbb{E}_t^M [BS(Y_t, K, T-t, \bar{\sigma}^2(M_t), r_f)], \end{aligned} \quad (4.8)$$

where $BS(\cdot)$ stands for the Black-Scholes formula with $\bar{\sigma}^2(M_t) = \frac{1}{T-t} \sum_{i=t+1}^T \sigma^2(M_i)$ for a given outcome of the path $\{M_i\}_{i=t+1}^T$.⁸

In summary, the existence of $\tilde{\mathbb{P}}$ (under which $\tilde{\varepsilon}_t$ is independent of M_t and $\tilde{\varepsilon}_t \stackrel{d}{=} N(0, 1)$) enables us the use of the Black-Scholes formula when conditioning by \mathcal{F}_T^M (cf. Hull and White^[48], Romano and Touzi^[154], and Ghysels *et al.*^[152], among others). This procedure is part of the classical literature on option pricing, that

⁷The returns x_t in eq. (4.7) are derived by adding $r_f - r_f$ in eq. (4.2), and setting $\theta_t^\varepsilon \equiv \frac{\mu(M_t) - r_f}{\sigma(M_t)}$ and $\tilde{\varepsilon}_t \equiv \varepsilon_t + \theta_t^\varepsilon$. It follows by the definition of the log-increments x_t and the Lognormal property that $\mathbb{E}_t^\varepsilon [Y_T e^{-r_f \cdot (T-t)} | \mathcal{F}_T^M] = Y_t \cdot \exp \left[\sum_{i=t+1}^T \mu(M_i) - r_f \cdot (T-t) \right]$. The martingale property is fulfilled under $\tilde{\mathbb{P}}$ provided $\tilde{\mathbb{E}}_t^\varepsilon [Y_T e^{-r_f \cdot (T-t)} | \mathcal{F}_T^M] = Y_t$ (cf. Chapter 1 for more details).

⁸ Recall that for a given path $\{M_i\}_{i=1}^T$, application of the Black-Scholes formula only requires a diffusion parameter $\bar{\sigma}^2$ representing the average volatility of the period.

we have only adapted to the case of MSM models. Alternative procedures like the stochastic discount factor (SDF) approach exist though, that allow one to obtain an equivalent pricing rule under different arguments. In case of the SDF approach, one would directly evaluate $C_t(T, K)$ under \mathbb{P} by means of a SDF defined as function of ξ_t and $e^{-r_f \cdot (T-t)}$.⁹ It can further be shown that this approach leads to Garcia *et al.*'s^[51] version of the Black-Scholes formula after a few adjustments.

4.3 Model Implementation

The main goal of this chapter is to evaluate how informative a set of option prices generating a volatility smile is for forecasting future volatility in the context of MSM models. To this end, we will employ the following rolling-window scheme: on a given day $\ell \in \mathcal{T}$, (i) we estimate the parameters of eq. (4.2) based on the historical sample series of size T , (ii) we calibrate the process of eq. (4.7) to the sample \mathcal{N}_ℓ of options available on that day, and (iii) we forecast QV based upon both sets of parameters as a proxy for future volatility. Once the results have been obtained, we roll over the sample by one trading day and repeat the previous steps. In the estimation procedure, the rolling window leads to a new observation being included at the front of the sample while the observation furthest away in the history is dropped. In the calibration scheme, the rolling window replaces all contracts in the sample of the day by those of the next day instead. The procedure is continued until no trading days for calibration remain available. According to the data filtering rules described below in section 4.4, the sample will consist of

⁹ Note that in this setting one needs to keep track not only on the process x_t but also on the SDF as the evaluation procedure updates from one observation to the next in the time series.

a total of $\mathcal{T} = 596$ observations, a sufficient number for a statistical evaluation of our forecasts.

For the rest of this chapter we will fix $(b, \gamma_{\bar{k}}) = (2, 0.5)$ as in Lux^[8] and Lux *et al.*^[9] and consider three alternative choices for the numbers of multipliers $\bar{k} = 3, 6, 9$. Our parameter space will consist of $\varphi = (m_0, \sigma_1, \alpha) \in \mathbb{R}^3$ for the estimation procedure and of $\tilde{\varphi} = (m_0, \sigma_1) \in \mathbb{R}_+^2$ for the calibration.

Estimation Methodology

The time series estimation is performed via the ML estimation procedure developed by Calvet and Fisher^[7] on the historical return series.¹⁰ For this, let us define $\Pi_t^i \equiv \mathbb{P}(M_t = m^i | x_t)$ over the unobserved states m^1, \dots, m^d , $d = 2^{\bar{k}}$. The return x_t , conditional on M_t , is distributed with density $f_x(x_t | M_t = m^i) = [\sigma(m^i)]^{-1} f_\varepsilon((x_t - \mu(m^i) + \frac{1}{2}\sigma^2(m^i))/\sigma(m^i))$, where f_ε is the standard Gaussian distribution. The vector Π_t stacks the conditional probabilities Π_t^i for $i = 1, \dots, d$ and is calculated recursively by Bayes' rule

$$\Pi_t = \frac{\omega(x_t) * (\Pi_{t-1} \mathcal{A})}{\omega(x_t)' (\Pi_{t-1} \mathcal{A})},$$

where $\omega(x_t)$ is the vector stacking $f_x(x_t | M_t = m^i)$ for $i = 1, \dots, d$, \mathcal{A} is the transition matrix for regime changes between the d volatility states, and $*$ is the element by element multiplication operator. For the choice of the initial Π_0 we select the ergodic distribution of the Markov chain (cf. Calvet and Fisher^[7]). Given a sample of size T , the estimates $\hat{\varphi}_T$ are obtained from the maximization

¹⁰ As it will be explained in the next section, we will actually employ the historical return series of the DAX futures rather than of the index for estimation.

of the log-likelihood

$$\widehat{\varphi}_T = \arg \max_{\varphi \in \mathbb{R}^3} \ln L(x_1, \dots, x_T; \varphi) = \sum_{t=1}^T \ln [\omega(x_t)' (\Pi_{t-1} \mathcal{A})]. \quad (4.9)$$

This optimization is performed using traditional gradient-based methods. Once executed, the procedure comes along with the identification of conditional probabilities $\widehat{\Pi}_t$ of the current volatility states. Together with the transition matrix \mathcal{A} , these conditional probabilities can be used to perform forecasts according to Bayes's rule.

Calibration Methodology

The calibration is performed via a Non-Linear Least Squares (NLS) optimization procedure for the following objective function

$$\widehat{\varphi} = \arg \min_{\widetilde{\varphi} \in \mathbb{R}_+^2} \Psi(\widetilde{\varphi}) = \sum_{i=1}^{\mathcal{N}_\ell} w_i |C_{i,\ell}^{\widetilde{\varphi}}(T_i, K_i) - C_{i,\ell}|^2, \quad (4.10)$$

where $\widehat{\varphi}$ is the resulting calibrated set of parameters for the day $\ell \in \mathcal{T}$, \mathcal{N}_ℓ denotes the total number of derivative contracts $C_{i,\ell}$ available on this day, w_i the weighting function, and $C_{i,\ell}^{\widetilde{\varphi}}(T_i, K_i)$ the value obtained in $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$ from eq. (4.8). In the absence of closed-form solutions for option prices in a market with MSM assets, the evaluation of $C_{i,\ell}^{\widetilde{\varphi}}(T, K)$ in eq. (4.8) uses a Monte Carlo integration of MSM sample paths for the evaluation of the $BS(\cdot)$ formula. The procedure consists of (i) the simulation of \mathcal{S} parallel paths of $\{M_i\}_{i=t+1}^T$, (ii) the calculation of $\bar{\sigma}^2(M_t) = \frac{1}{n} \sum_{i=t+1}^T \sigma^2(M_i)$ for each path $s \in \mathcal{S}$, (iii) the evaluation of the BS formula employing the previous result for each path $s \in \mathcal{S}$, and finally, of

(*iv*) the averaging across the \mathcal{S} simulated paths (cf. Javaheri^[165]). Given the first-order Markov property of the vector M_t , the simulation of the process can be initiated given the conditional probabilities $\hat{\Pi}_t$ obtained as a by-product from the estimation procedure.

Figure 4.1 exhibits an example of the objective function $\Psi(\tilde{\varphi})$ of eq.(4.10) using the empirical data of the day 12/06/2007. As one can see, the function appears smooth, which allows the application of gradient-based optimization methods.

For the numerical optimization we have used the Newton-Raphson and BFGS algorithms included in the software package GAUSS. A new set of common pseudo-random numbers has been drawn for each calibration day. The convergence of the procedure has never been a problem despite the fact that strict convexity of the objective function in an empirical study like ours cannot be guaranteed. The choice of the number \mathcal{S} of simulated paths has been based on the tradeoff between the gain in accuracy, measured as the reduction of the variance of calibrated parameters, and the cost of increased calibration times plus the cost of the physical hard-drive space used.

Table 4.1 shows, as an illustration, the reduction in variability obtained by increasing the number of simulated paths \mathcal{S} from 10.000 to 50.000, and from 50.000 to 100.000 when calibrating to one day of artificial data. To get some idea on the performance of the calibration algorithm, we have generated an artificial sample of \mathcal{N}_t option values for subsequent calibration, based on the strikes K , the risk-free interest rate r_f , the time to maturity $T - t$, and the current price of the DAX index Y_t available in the market for the date 12/06/2007, as well as the latent conditional probabilities $\hat{\Pi}_t$ obtained from the historical data for that day. The generation of the artificial data follows the same procedure used for

the evaluation of $C_{i,\ell}^{\tilde{\varphi}}(T, K)$ above. The underlying model has parameter values $m_0 = 1.4$, $\sigma_1 = 1.5$, and $\bar{k} = 9$. We have used weights $w_i = 1/C_i^2$ to capture the accuracy of reported option prices for this analysis (cf. below for more details about alternative choices of w_i). The resulting data was calibrated 200 times using for each calibration a different set of random numbers for the simulated paths \mathcal{S} .

We notice in Table 4.1 that the average of the calibrated parameters coincides almost perfectly with the true parameter values for all chosen values of \mathcal{S} . In addition, a reduction in the mean across the 200 calibrations of the mean squared error (MMSE) (measured as the mean of $\frac{100^2}{\mathcal{N}} * \Psi(\hat{\varphi})$) is obtained when increasing \mathcal{S} . This reduction is accompanied by a reduction in the standard deviation (SD) of all calibrated parameters as well as of the MMSE. Provided that rather accurate and stable results have been obtained with $\mathcal{S} = 50.000$, we will employ this number of scenarios in all our further experiments.

We now turn to the weighting function w_i . This function embodies the importance of the individual contracts according to their moneyness. As Cont and Tankov^[46] argue, the choice of the relative weights should also reflect our confidence in the ‘accuracy’ of the individual option prices, which is usually determined by the liquidity, that is, the size of the bid-ask spread for each given option. The lack of high-quality information in our database regarding spreads precludes us though from employing this first strategy. Note then again that by minimizing the difference of implied volatilities one may get errors proportional to bid-ask spreads when the sample of options is not too far from the money. Since such a procedure may also be accompanied by computational difficulties, one may recur instead to a first order approximation of the minimization of the difference of implied volatil-

ities. This can be conducted by minimizing eq.(4.10) with $w_i = 1/\text{Vega}(I_i)^2$, that is, by choosing the reciprocal of the squared Black-Scholes Vegas evaluated at the implied volatility I_i of the market option prices as weights (cf. Cont and Tankov^[46]). Despite this choice of w_i producing seemingly good calibration results in our preliminary analysis, the procedure appeared to disregard valuable information when forecasting. As such, these results are not going to be discussed any further here. We opted instead to proceed with a simple rule like $w_i = 1/C_i^2$. As it will be shown in section 4.5, this weighting scheme contributes to a favorable performance out-of-sample.

Table 4.2 exhibits Monte Carlo results from 200 calibrations using weights $w_i = 1/C_i^2$, where each calibration has been performed for independent samples ℓ , $\ell = 1, 2, \dots, 200$, of artificial prices $C_{i,\ell}$. As in the previous exercise, we make use of the market information previously collected for our study to design the Monte Carlo generation of the artificial data. We select the first 200 trading days of our dataset and identify each of these days with an instance of the index ℓ . The resulting procedure can be summarized as follows: for each sample day ℓ , (i) we select all the empirical information available for that day,¹¹ (ii) we replace the empirical prices $C_{i,\ell}$ with artificial ones generated according to the procedure used for the evaluation of $C_{i,\ell}^{\tilde{\omega}}(T, K)$, the T_i and K_i associated to the empirical $C_{i,\ell}$, a set of pseudo-random numbers, and the latent conditional probabilities $\hat{\Pi}_t$ for the underlying model with $\bar{k} = 9$, together with parameter values $m_0 = 1.4$ and $\sigma_1 = 1.5$; and (iii) we execute the calibration of eq. (4.10) based on $w_i = 1/C_i^2$, the conditional probabilities $\hat{\Pi}_t$, and an independent set of common

¹¹ This data consists of the number \mathcal{N}_ℓ of derivative contracts $C_{i,\ell}$ traded, their corresponding strikes K_i , time to maturity $T - t$ and price $C_{i,\ell}$, as well as the term structure of interest rates r_f and the current price of the DAX index Y_t .

pseudo-random numbers.

As shown in the top panel, the procedure achieves to replicate on average the parameters from the data generating process (DGP), with a very low dispersion of the calibrates. The bottom panel exhibits an analysis of the calibration errors (cf. Bakshi *et al.*^[102], Madan *et al.*^[166] and Javaheri^[165]). We have collected from each day ℓ of the 200 available samples of artificial option prices, the number \mathcal{N}_ℓ of calibration errors $\epsilon_{i,\ell}$ and run the following regression

$$\epsilon_j = \beta_0 + \beta_1 \mathfrak{M}_j + \beta_2 (T - t)_j + \beta_3 r_{f_j} + u_j, \quad (4.11)$$

where the index j runs over the pooled data for i and ℓ , i.e. $j \in \{1, \dots, \mathcal{N}_1, \mathcal{N}_1 + 1, \dots, \mathcal{N}_1 + \mathcal{N}_2, \mathcal{N}_1 + \mathcal{N}_2 + 1, \dots, \sum_{\ell=1}^{200} \mathcal{N}_\ell\}$, \mathfrak{M}_j denotes moneyness, β_s , for $s = 0, 1, 2, 3$, are the least squares coefficients, and u_j the residuals of the OLS. The t-test values show that only β_2 is statistically significantly different from zero, with a p-value of 0.04. On the other hand, the low R^2 together with a high p-value of the F -test of all $\beta_s = 0$ simultaneously leads us to conclude that our calibration leaves no significant sources of linear predictability in ϵ_j .¹² In short, the good performance of our calibration method supports its further application with empirical data.

Note as a last remark that we lack information about the asymptotic properties of $\widehat{\varphi}$ when calibrating. In fact, the distance or bias of the calibrated parameters with respect to the population values remains uncertain when the sampling period is constrained to one day's data at a time. A verification via Monte Carlo analysis of

¹² Based on the results from White's test for heteroscedasticity, and Ljung-Box and Godfrey-Breusch tests for autocorrelation of the residuals u_j , the mentioned t- and F -tests have been calculated using the Newey-West^[137,167] heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator.

these asymptotic properties may require not only an increasingly larger number of trading days to be included in the sampling period but also a much more sophisticated inference procedure. By using just one day's data in each window instead, we focus on the most recent information only, which may lead to a high dispersion of the calibrated parameters across \mathcal{T} .¹³ Provided prior information like the time series estimate $\hat{\varphi}_{\mathcal{T}}$ of eq. (4.9) is available, it would be possible to apply a regularization technique as a partial solution. The latter may penalize, for instance, large deviations from these prior estimates, leading to a lower dispersion of the calibrated parameters along \mathcal{T} (cf. Cont and Tankov^[46,168] and Engl *et al.*^[169]). This ad-hoc refinement may clearly lead to a smoother development of the calibrated parameters, but it would also alter the nature of our study. In the absence of conspicuous numerical problems in our calibration exercise, we opted to avoid this kind of *ex post* refinement and to focus instead on the evaluation of results from unrestricted calibrations only.

Forecasting Methodology

The MSM model produces point-wise forecasts $\hat{\mathbb{E}}_t[\sigma_{t+N}^2]$ based upon the conditional probabilities of future states $\hat{\Pi}_{t+N} \equiv \mathbb{P}(M_{t+N}|x_t)$. The latter are obtained by Bayesian updating N times the conditional probabilities Π_t with the transition matrix \mathcal{A} :

$$\hat{\Pi}_{t+N} = \Pi_t \mathcal{A}^N. \tag{4.12}$$

¹³ Note that when working with empirical data, the total number of contracts \mathcal{N}_ℓ available for calibration may fluctuate strongly for each $\ell \in \mathcal{T}$. This may depend on the activity level of the market as well as on the number of maturities, strikes, and type of contracts (puts and calls) effectively traded.

4.3 Model Implementation

Note that $\widehat{\mathbb{E}}_t[\sigma_{t+N}^2]$ is a proxy for the expected QV metric of x_{t+N} in eq. (4.2), that is,

$$\widehat{\mathbb{E}}_t[\sigma_{t+N}^2] \equiv \widehat{\mathbb{E}}_t[\sigma^2(M_{t+N}) \cdot \varepsilon_{t+N}^2].$$

The fact that the drift $\mu(M_t)$ in eq. (4.2) depends on M_t may alert the careful reader about the potential approximation error obtained when employing expected QV as a proxy for the expected return variation of x_{t+N} . Our preliminary data analysis of the historical time series shows that based on the estimate of α obtained from the ML estimation procedure, the return variation from $\mu(M_t) - \frac{1}{2}\sigma(M_t)^2$ in eq. (4.2) is two orders of magnitude smaller than that arising from QV so that the omission from the former would not alter the conclusions of this work.

Noteworthy is also that the *direct* comparison of the volatility forecasts QV generated from each set of parameter estimates resides in the assumption that volatility risks remain uncompensated. Technically, this led to $\Lambda_T^M = 1$ in eq. (4.4), so that no adjustments are required for the application of the volatility process under \mathbb{P}^M that has been calibrated under $\widetilde{\mathbb{P}}^M$.

The identification of the conditional probabilities Π_t in eq. (4.12) arises as a by-product of the ML estimation procedure. Once we have obtained the parameter estimates from the times series and the calibrated parameters from the option prices, point-wise multiple-day-ahead forecasts $\widehat{\mathbb{E}}_t[\sigma_{t+N}^2]$ are straightforward to perform. Accumulated forecasts $\widehat{\mathbb{E}}_t[\sigma_{t+1:t+N}^2]$ may be derived in the same way by

$$\widehat{\mathbb{E}}_t[\sigma_{t+1:t+N}^2] = \frac{1}{N} \sum_{i=1}^N \widehat{\mathbb{E}}_t[\sigma_{t+i}^2], \quad (4.13)$$

where we have chosen to scale the sum of point-wise forecasts by N for better comparison. Our focus will be restricted in this case to horizons of up to 20 days, as the accumulated return variation may start diverging from our proxy for longer ranges (cf. Andersen *et al.*^[32,170]).

Finally, we employ the realized volatility (RV) metric calculated from high-frequency data of the DAX future prices as a benchmark for the evaluation of our results.

4.4 The Data

The DAX index measures the development of the 30 largest German companies and represents around 80% of the market capitalization of the German stock market. It is a performance index in the sense that cash dividend payments are assumed to be reinvested twice a year. This leads to a straightforward evaluation of eq. (4.1), provided the adjustment for dividends can be dispensed with.

Our empirical study requires two complementary data sources: a full-sample data set of intra-day 1-minute-interval DAX futures transaction prices, and a data set with the closing prices from DAX index options, together with the corresponding closing price of the DAX index and the EURIBOR rates.

The DAX futures series

The DAX futures series (henceforth simply the FDAX series) will be employed for the estimation exercise as well as for the calculation of the QV proxy. The futures sample period comprises data for all maturities from January 1997 until Octo-

ber 2010. The expiration months are the three nearest calendar months within the cycle March-June-September-December. The contracts are available in three different formats in our data set, namely with original starting and expiration dates, and as adjusted and unadjusted continuous series. In order to avoid the effect from spurious returns produced by the jump from one future at expiration day to the next one, we will employ in our study the continuous adjusted series only.

For the estimation exercise, we focus directly on the daily return of the futures $\ln(\frac{F_{t,T}}{F_{t-1,T}})$. Notice with help of eq. (4.1) and the insertion of eq. (4.3) in eq. (4.2) that the resulting returns allow us to proxy the dynamics of the underlying without an input for r_f . The in-sample estimation period consists of $T = 2636$ observations, starting in January 1997 and ending in June 2007 for the first window in \mathcal{T} . Futures have been traded until the first of June of the year 2000 up to 17:30 hours. After that, trading hours have been increased until 22:00 hours in 2010. Since the trading period in the derivative markets in EUREX finishes shortly before 17:45, there is a disparity in the closing hours of the options and future markets. The need to have a coherent point of comparison between these two contracts leads us to the following selection: up to June 2000, our sample consist of futures closing-day prices; afterwards, we select the closing price of the 1-minute interval belonging to 17:45 hours, which also coincides with the timestamp of the DAX index closing time.

As a benchmark for the evaluation of our forecasting results, we employ the RV metric calculated from the adjusted continuous FDAX price series containing 840 Δt_1 -prices on each trading day t , where we have denoted by Δt_l the l -minute-interval. To fix notation, let us also define by $r_{t-1,i\Delta t}$ the Δt_5 -log-return

at intervals $i = 1, \dots, \frac{1}{\Delta t}$ in $[t - 1, t]$. Disregarding overnight returns, a daily realized volatility metric $RV_t = \sum_{i=1}^{1/\Delta t} r_{t-1, i\Delta t}^2$ computed from a simple Δt_5 time-grid would start with the first observation corresponding to the time grid for that day and employ 20 percent of the available data only. Instead, we apply the refined estimator from Zhang *et al.*^[132]. This estimator has proven to be smoother and makes a more efficient use of the data. For this, we specify five different regularly-spaced grids starting at minutes 1, 2, 3, 4, and 5 of each trading day t . We calculate a RV_t^j estimator for each subgrid j and average them over this index to obtain a more robust estimator in the presence of noise and non-synchronicity.¹⁴

Note, however, that the necessity to match the time-stamp between the option- and futures-closing periods would impose in our setup a restriction on the amount of observations available in each subgrid. To avoid losing too much data, we consider instead the returns $r_{t-2, i'\Delta}$ as belonging to the next trading day, with i' indicating the intervals after 18:00 and reaching just before 22:00 for each regularly-spaced subgrid j . RV_t^j consequently follows as

$$RV_t^j \equiv RV(late)_{t-1}^j + RV(daily)_t^j,$$

where $RV(late)_{t-1}^j$ is the RV for the period between the option- and futures-closing times on the day before, and $RV(daily)_t^j$ is the RV calculated for the current day up to the closing period of the option market. As a result, this procedure leads to the following proxy of daily QV: $\tilde{\sigma}_t^2 \equiv \frac{1}{5} \sum_{j=1}^5 RV_t^j$.

¹⁴ Lux *et al.*^[9] employ instead the 30-minute-interval returns for the calculation of RV_t . Like in that study, our estimator adjusts very fast within at most five minutes after an overnight, weekend or holiday pause.

In coherence with eq. (4.13), the metric for integrated variance exceeding the variation from one trading day is obtained by

$$\tilde{\sigma}_{t+1:t+N}^2 = \frac{1}{N} \sum_{i=1}^N \tilde{\sigma}_{t+i}^2. \quad (4.14)$$

Figure 4.2 exhibits the time-varying evolution of $\tilde{\sigma}_{t+1:t+N}^2$ at four different horizons N . Clearly, all samples are characterized by volatility clustering as it should be expected for a measure of price variation. As we would also expect, the larger N the smoother the evolution of $\tilde{\sigma}_{t+1:t+N}^2$ is.

The DAX-index option series

The second data set has been provided by IVolatility.com and comprises raw daily data for a period from June 2007 until January 2010. Together with the option closing prices, the data includes the transaction volumes, bid-ask quotes as well as Black-Scholes Greeks. The latter were calculated by the provider employing the value of the DAX index at closing time (17:45), the EURIBOR as the proxy for the risk-free rate r_f , and a zero instantaneous dividend yield q . The options on this index have the following maturities: the three nearest successive calendar months, then the three following quarterly months of the March, June, September and December cycle, after that the four following semi-annual months of the June and December cycle and, finally, the two following annual months of the December cycle.

The calibration procedure is performed on the closing prices of both puts and calls for each calibration day in the sample. The latter is defined as a trading

day for which an adequate number of contracts remains available for use after filtering. The long memory features of our model suggest the inclusion of as many option maturities per day as possible. A thorough study of our sample has brought up, however, that maturities after 175 calendar days are scarcely traded, with just a handful number of strikes per expiration.¹⁵ Since their inclusion in the calibration could distort our results, we exclude options with less than 7 and more than 175 calendar days until maturity. This leaves us on every trading day with at least 3, and a maximum of 4 maturities.

Further filter rules to the raw data include the turnover and minimum-price rules as suggested by de Jong and Lehnert^[153], a maximum-width of bid-ask spread rule in which the mean price has to be higher or equal than the ask-bid spread, a Vega rule in which options with a Vega amount less than 100 Euro are discarded and, finally, a minimum-volume rule in which options with less than 10 traded contracts are excluded.

The maximum-width of bid-ask spread rule disregards those options from the sample for which the ask price is high while the bid price is close to zero. The turnover from the traded contracts may still be higher than the minimum required, but its mean price -the value to which we would try to calibrate our parameters- would represent a too noisy measure of the market's fair value of the contract.

The Vega rule is a compound filter in the sense that it disregards those options from the sample that are too deep in- or out-of-the-money for a given maturity, i.e. it eliminates those options that have a very low probability of reversal of their value. As such, very deep in- or out-of-the-money options that still convey

¹⁵ This number may, in addition, strongly fluctuate from one trading day to the next.

valuable information regarding the variability of the underlying remain present in our sample.

The minimum-volume rule filters out those options that for being too deep in-the-money manage to pass the turnover rule despite very few contracts being traded. By applying this rule, we increase the probability that more than one trader is at each side of the order book and avoid therewith an overproportional influence of single traders in our data.

The application of these rules leads to a more reliable sample of derivative contracts. After preliminary work with this sample though, we noticed that some additional adjustments to the daily index level might be needed, before initializing the calibration on each trading day. In particular, we have found evidence that despite the EURIBOR being the most commonly used proxy for the risk-free rate r_f , non-negligible differences can be obtained from time to time between the market prices of the FDAX and the values predicted by eq. (4.1) when the market price of the DAX index has been used as a proxy for Y_t . These differences are particularly perceivable for the two longest maturities.¹⁶ Due to the close integration between the futures and the option markets, it is reasonable to expect that the same differences arise between the market price of the options and the price predicted by eq. (4.8). To deal with this situation, we supply eq. (4.1) with the closing price of the DAX index as well as with the corresponding market values of the EURIBOR and the FDAX for each maturity and perform an adjustment for each future contract. Namely, we compute a second-order polynomial interpola-

¹⁶ Clearly, the fact that the FDAX trades at different maturities on any given day and, on the other hand, that the adjustments for cash dividends of the DAX index is performed only twice a year may lead the market expectations about the price evolution of its constituents, and specifically about the size of their potential dividends, to influence the forward value of the FDAX contracts between maturities (cf. Lehnert^[171]).

tion curve based on the previously obtained values and the available maturities of the DAX options. The resulting rate may be understood as an *artificial* dividend yield that can be applied directly to the DAX index price before the latter is inserted in the *BS* formula of eq. (4.8). For convenience, we denote this rate by q but notice that this rate is not restricted to take positive values only, as it is in its traditional definition as a instantaneous dividend yield.¹⁷

After experimenting with this methodology, we have noticed that on certain trading days an abnormally large forward ‘cost of carry’ $r_f - q$ may be detected between the longer maturities of the FDAX, being here q a nonzero but possibly negative instantaneous ‘yield’. We have also noticed that these days usually coincide in our data set with a period where a large negative market shock has occurred. On these days, this ‘cost of carry’ represents a large deterministic drift in the ‘risk-neutral’ evolution of the index price that may affect the calibration of those options with longest expirations. In general though, the integrity of the historical time series would remain intact, provided that the continuous adjusted series is built on the future contracts with the highest trading volume, which in case of the DAX has always been the future with the closest expiration. In short, the size of the forward ‘cost of carry’ $r_f - q$ has been employed as a final filter rule, with those trading days containing an annualized absolute value of $r_f - q$ higher than 10 percent being discarded from the sample. About 10 percent of the total number of trading days originally available in our data set has consequently

¹⁷ Some authors have avoided these inaccuracies altogether by computing implied interest rates directly from option prices (cf. Duan^[172] and Lehnert^[171]). The method consists in applying a modified version of the put-call parity regression of Shimko^[173]. Preliminary results with this method though have proven that this approach is rather unreliable in our case. The reason for this resides in the large amount of puts *and* calls with identical strikes that is necessary for a reliable inference. This requirement has hardly been met in our data set, which would have led us to discard nearly one third of all available trading days.

been disregarded due to this rule.

In summary, a fairly small amount of trading days has been excluded through filtering, leading to a total sample size of $\mathcal{T} = 596$ days. On each resulting calibration day $\ell \in \mathcal{T}$, a sensible minimum of 10 contracts per maturity T_h , $h \in \mathcal{H}$, has been maintained. The inclusion of $\mathcal{H} = 3$ to 4 maturities per day has ensured a minimum of 53 and a maximum of 205 contracts \mathcal{N}_ℓ per trading day. The resulting sample presents a rich variety of related events: a period of low uncertainty of approximately 300 days before a major market crash, two posterior periods with high level of uncertainty and a final period of recovery. As for the interest rate, an initial stable period of high interest rates is followed by a stable period of low rates in the time following the market crash.

4.5 Empirical Evidence

In this section we analyze the performance of our methodology in terms of in-sample fit as well as of out-of-sample forecast accuracy.

In-sample analysis

The first results from our in-sample analysis are presented in Table 4.3. The analysis of the historical time series shows an improvement of the fit as \bar{k} increases, together with a decrease in the average value of m_0 . These results are pretty much in harmony with previous applications of the MSM (cf. Calvet and Fisher^[7], Lux^[8], Lux *et al.*^[97,98]). We note also that the average estimate of α not only is the same for all \bar{k} but also offsets the variation produced by the Itô

term. We should note, however, that the standard deviation of α is not negligible. Visual inspection of the \mathcal{T} estimates of α indicate that the later tend to slowly switch from a higher value than 0.5 to a lower one and back, as the update of information in the rolling window of size $T = 2636$ incorporates the peaks of the fluctuations associated with the aforementioned market events.

Similarly, we observe an improvement of the in-sample fit in the calibration results as \bar{k} increases, where not only the average of m_0 but also of σ_1 decreases in value. The figures for the latter differ also noticeably from their estimated counterparts and point towards a high level of volatility (cf. Javaheri^[165]). The reason for this could be that the higher dispersion in the calibrating process reflects the uncertainty of market expectations about the future value of the asset (cf. Aït-Sahalia *et al.*^[174]). Note also that calibrated parameters show a much higher dispersion than the time series estimates, which certainly is due to the complete replacement of daily samples within the rolling window scheme.

The mean-MSE (MMSE) figures across all calibrations also appear quite high. Several factors may contribute to this. The first and more obvious reason is the fact that the MMSE averages square figures, so that a few poor calibration results may dominate the final outcome. A high standard deviation of the MMSE and a visual inspection of the calibrated parameters across \mathcal{T} confirm this assertion.¹⁸ Second, a comparison to the previous simulation analysis brings about two potential sources of calibration errors not contemplated before. On the one hand, the synchronization between the spot close price and the option close price was taken care of automatically when generating artificial data. Renault^[156] shows

¹⁸ Though relatively few, poor calibration results tend to concentrate at those periods immediately before and after a large market movement.

via simulations of a Hull and White^[48] model though, that even minor differences between the price of the underlying used for closing the option contract and the one used for calibration may produce asymmetries corresponding to the smirks and frowns widely documented in the literature (cf. Garcia *et al.*^[51]). Confirming the initial argument from Bates^[175], Hilliard and Hilliard^[157] have recently shown that the existence of asynchronous observations among derivative contracts and their corresponding underlying asset does matter when calibrating a model. Thus, provided that our procedure is performed on a daily basis only, a portion of the calibration errors may very well be due to the embedded non-synchronicity and may not be completely eliminated.

The other source of calibration error is model misspecification, and more specifically, the potential mismatch between the level of skewness generated by the MSM model in the ‘risk-neutral’ world and the one captured by the sample of option prices. We note first of all, that the MSM model recovers fairly well the statistical properties of the DAX index in the ‘statistical’ world (cf. Lux *et al.*^[9]). For the analysis of the modeling mismatch in the ‘risk-neutral’ world, we proceed with the evaluation of the in-sample fit in two ways. The first method consists of a comparison of the average MMSE measures between the calibrated MSM and a benchmark model. For the latter, we consider the ad-hoc Black-Scholes model of Dumas *et al.*^[176] with 2 variables:

$$\sigma_{i,\ell} = \beta_{0,\ell} + \beta_{1,\ell} \mathfrak{M}_{i,\ell} + \beta_{2,\ell} T_{i,\ell} + u_{i,\ell}, \quad (4.15)$$

where $\sigma_{i,\ell}$ is the implied volatility for strike $K_{i,\ell}$ and maturity $T_{i,\ell}$, $\mathfrak{M}_{i,\ell}$ is the moneyness, $\beta_{s,\ell}$ for $s = 0, 1, 2$ are the least squares coefficients estimated on each

$\ell \in \mathcal{T}$, and $u_{i,\ell}$ the residuals of the OLS for $i \in \mathcal{N}_\ell$. Before turning to the comparison of results, note that the calibration errors of the MSM model are expressed in monetary terms, or more precisely, in percentages of monetary units when $w_i = 1/C_i^2$ are used as weights. We have also recalculated *ex-post* the results obtained in eq. (4.10) using $w_i = 1/\text{Vega}(I_i)^2$ (labeled MMSE'). By choosing these weights, we obtain a first order approximation to the problem of minimizing the square difference between the market and the model implied volatilities (cf. Cont and Tankov^[46]), leading to a better comparison to the residuals $u_{i,\ell}$ of eq. (4.15).¹⁹ The results in Table 4.3 show that the ad-hoc Black-Scholes model widely outperforms the MSM calibration. This relies on the fact that the ad-hoc Black-Scholes model specifically considers the shape of the moneyness function of the options, while the MSM assumes a symmetric smile. It is interesting to see though that in comparison to the MMSE figures obtained in the simulation exercise in the previous section, the ad-hoc Black-Scholes exhibits large MMSE figures as well, with an even larger standard deviation. As a result, the calibration errors due to the non-synchronicity of spot prices are likely to constitute a large portion of the total mismatch. We should mention, on the other hand, that de Jong and Lehnert^[153] also report high levels of parameter dispersion and an underperformance of their implied volatilities based on a ARCH-type time series model with respect to those of the ad-hoc Black-Scholes model, though to a lower degree. This occurs despite the fact that the EGARCH in their study does take into account the 'leverage' effect. Further results from Heston and Nandi^[158] give reason to believe that the ad-hoc Black-Scholes model may achieve better

¹⁹ Note also that because $\widehat{\varphi}$ has remained fixed in this recalculation, the resulting $\Psi(\widehat{\varphi})'$, say, is suboptimal under the weighting scheme $w_i = 1/\text{Vega}(I_i)^2$.

in-sample fit in the first place by over-fitting the data.

To deepen the analysis of the in-sample under-performance of the MSM, we continue with an evaluation of the calibration errors (cf. Table 4.4). As in the simulation analysis carried out in section 4.3, we pool the calibration errors from each day $\ell \in \mathcal{T}$ and each option i . We consequently run the regression

$$\epsilon_j = \beta_0 + \beta_1 \delta_j + \beta_2 \mathfrak{M}_j + \beta_3 (\delta_j \cdot \mathfrak{M}_j) + \beta_4 (T - t)_i + \beta_5 (r_f - q)_j + u_j, \quad (4.16)$$

where j runs over the pooled data for i and ℓ , i.e. $j \in \{1, \dots, \mathcal{N}_1, \mathcal{N}_1+1, \dots, \mathcal{N}_1+\mathcal{N}_2, \mathcal{N}_1+\mathcal{N}_2+1, \dots, \sum_{\ell=1}^{\mathcal{T}} \mathcal{N}_\ell\}$, \mathfrak{M}_j denotes moneyness, $(r_f - q)_j$ the cost of carry, β_s the least squares coefficients for $s = 0, 1, \dots, 5$, and u_j the OLS residuals. We have added in this case a dummy variable δ to contemplate the possibility that put options are traded in the market in a different way than calls are. Thus, we use $\delta_j = 1$ for puts, and $\delta_j = 0$ otherwise. The reasoning here is that the R^2 from eq. (4.16) would capture the linear predictability of ϵ_j resulting from all possible sources of calibration errors. Given that the MSM generates a symmetric volatility smile though, statistically significant coefficients β_1 and β_3 would indicate that the market treats put contracts unlike call contracts, a well-known sign for the presence of a pronounced negative skewness in the data.

As shown in Table 4.4, the R^2 for all \bar{k} is low though statistically significant, with the t-statistic for all coefficients β being highly significant.²⁰ Simply put, our calibration procedure leaves on average exploitable information behind. We notice moreover that we have obtained a statistically significant indication that puts are traded differently from calls in the market, a feature to which our model

²⁰ Note that in comparison to our calibrated data used in Table 4.2, the empirical sample is almost three times as large, which favors higher significance levels.

does not present a remedy in its current format.²¹

Out-of-sample comparisons

We proceed in the next step with the comparison of the out-of-sample forecasts. Following the literature on economic forecasting, we present also an analysis of forecast combination and encompassing (cf. Newbold and Harvey^[177]). For the evaluation of our results, we will employ the RV metric as a benchmark, calculated from the adjusted continuous FDAX price series.

Table 4.5 provides an assessment of point-wise forecasts $\widehat{\mathbb{E}}_t[\sigma_{t+N}^2]$ for horizons $N = 1, 10, 20, 50,$ and 100 . The first evaluation regarding the predictive abilities of the models uses two traditional summary statistics: the mean squared (MSE) and the mean absolute errors (MAE). These are presented in the top panel of Table 4.5 relative to the corresponding statistics obtained from a ‘naive’ forecast using the historical variance from the detrended daily-return FDAX series (labeled as forecasts from a random walk model (RW)). Table 4.5 also presents combined forecasts (labeled $(E + C)_1$) taken as the simple average between those from time series (labeled E) and those from calibrations (labeled C).²² To quantify the robustness of our findings, a series of tests based on the dispersion of our forecasts has also been conducted. For each level \bar{k} , the following comparisons have been made: all forecasting methods against RW forecasts, E -forecasts and C -forecasts against each other and against combined $(E + C)_1$ -forecasts, and

²¹ We note that despite all coefficients β being statistically significant, inaccuracies related with the cost of carry $(r_f - q)_s$ and time to maturity $(T - t)_s$ may have more to do with imperfections of the employed proxies than with modeling.

²² Cf. Makridakis *et al.*^[178] and Makridakis and Winkler^[179] for early references on the remarkable effectiveness of this strategy in economic scenarios.

finally, forecasting encompassing. The comparison against the RW forecasts is made under MSE and is based on the test statistic for nested models of Clark and West^[124,125]. The comparison of E -forecasts and C -forecasts between each other, and against combined $(E + C)_1$ -forecasts, are based on the modified Diebold and Mariano^[126] test statistic by Harvey *et al.*^[127]. The test for forecast encompassing is based upon Harvey *et al.*^[180]. All tests have been calculated according to the Newey-West^[137] heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator.

We can assert, to start with, that for horizons up to 20 days, the forecasting accuracy of the MSM model under MSE increases monotonically with \bar{k} and is statistically significantly better than the performance of the RW model. C -forecasts for $\bar{k} = 9$ provide the best results, and E -forecasts for $\bar{k} = 3$ the worst.²³ At the shortest horizons $N \leq 10$, C -forecasts outperform all others - even with statistical significance for $N = 1$. Combined $(E + C)_1$ -forecasts, on the other hand, improve only upon E -forecasts at these horizons. At longer horizons, that is, for $N = 50, 100$, $(E + C)_1$ -forecasts seem to provide for all \bar{k} a better performance than C -forecasts. At these horizons, however, none of the multifractal forecasts

²³ We are aware that the construction of our forecasts may not obey the assumptions underlying some of these tests to the letter, and, as such, remarks about the statistical significance of our results should be taken with a grain of salt. Consider, for instance, that while no assumption about the origin of the forecasts is required for the Diebold and Mariano^[126] test (cf. Diebold^[181]), the test of nested models requires the use of asymptotic properties that are unknown to us when forecasts based on calibrations are used. In addition, one may wonder whether the RW forecasts are a good candidate for the test of nested models against C -forecasts or one should use forecasts based on a Black-Scholes model as a more correct proxy. Note, on the other hand, that the asymptotic irrelevance of parameter uncertainty may apply for the E -forecasts provided $\mathcal{T} = 596$ and a sample size of $T = 2636$ observations for the historical ML estimation are used. Hence, one may consider the use of the standard critical values of the Diebold and Mariano^[126] test as reasonable in this case (cf. West^[182,183] for a reference to this ‘rule of thumb’). Alternatively, one could refer to the asymptotic properties of Giacomini and White^[184] for the tests of nested models and for encompassing.

is statistically significantly better than RW forecasts anymore. The results of the encompassing tests indicate that at the shortest horizons C -forecasts are significantly different from E -forecasts. As the forecasting horizon increases, however, forecasts smooth out and become increasingly similar. This can be seen by the fact that $\widehat{\Pi}_t \mathcal{A}^N$ in eq. (4.12) converges to the ergodic distribution as $N \rightarrow \infty$, letting $\widehat{\mathbb{E}}_t[\sigma_{t+N}^2] \rightarrow \sigma_1^2$. Results under MAE appear to sustain in general our previous remarks. In this case, however, we have only compared E -forecasts and C -forecasts against each other and against $(E + C)_1$ -forecasts, given the difficulty in obtaining reliable critical values for comparison of nested models under this evaluation criterion.

For a further assessment of our results at all horizons, we provide in the bottom panel of Table 4.5 the heteroscedasticity-adjusted error statistics (cf. Andersen *et al.*^[170]):

$$\text{HMSE} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \left[1 - \frac{\widehat{\mathbb{E}}_t[\sigma_{t+N}^2]}{\tilde{\sigma}_{t+N}^2} \right]^2 \quad \text{HMAE} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \left| 1 - \frac{\widehat{\mathbb{E}}_t[\sigma_{t+N}^2]}{\tilde{\sigma}_{t+N}^2} \right|,$$

where $\tilde{\sigma}_{t+N}^2$ is the RV metric as defined in the paragraph immediately before eq. (4.14). Once again, the statistics are presented in relation to the corresponding statistics obtained from RW forecasts.²⁴

As a first remark, we can assert that C -forecasts for $\bar{k} = 9$ also outperform all others under HMSE and HMAE, but only at horizons up to 10 days. For $\bar{k} = 3$ and 6 though, $(E + C)_1$ -forecasts do seem to improve upon C -forecasts,

²⁴ The HMSE and HMAE statistics (also known as the mean square percent error (MSPE) and mean absolute percent error (MAPE), respectively) penalize the direction of the forecast errors asymmetrically (cf. Diebold and Lopez^[185]). To our knowledge, there aren't any available tests designed specifically for these evaluation functions.

uniformly under HMSE and with only one exception under HMAE. At longer horizons, E -forecasts for $\bar{k} = 6$ perform best, with E -forecasts for $\bar{k} = 3$ being only slightly worse. The reason for this change in results is that under HMSE and HMAE, large deviations from $\tilde{\sigma}_{t+N}^2$ are less heavily penalized than under MSE and MAE, so that an improvement upon RW forecasts can be obtained more easily with a comparably higher number of accurate forecasts. This is, for instance, the case at the shortest horizon $N = 1$, where the close influence of $\hat{\Pi}_t$ in $\hat{\Pi}_t \mathcal{A}$ leads to a higher number of accurate forecasts for all methods. As the forecast horizon N increases, however, $\hat{\Pi}_t \mathcal{A}^N$ approaches the ergodic distribution making the forecasts more homogeneous. In doing so, the rate at which all forecasts $\hat{\mathbb{E}}_t[\sigma_{t+N}^2]$ will approach the unconditional moment σ_1^2 will depend on the size of \bar{k} . Hence, for $N \geq 20$ the size of the parameter σ_1 tends to become decisive in our setting. Note, for instance, that the average calibrate $\bar{\sigma}_1$ in Table 4.3 is much higher in value than its historical time series counterpart. Also in-sample, the lower \bar{k} is, the closer the average time series estimate $\bar{\sigma}_1$ will be to the average historical volatility, while the average calibrate $\bar{\sigma}_1$ diverges from it. Consequently, E -forecasts at longer horizons N will produce smoother forecasts the smaller \bar{k} , with potentially the same large errors at infrequent times as the RW forecasts, and more time-dependent forecasts for higher \bar{k} , which by their comparably higher dependence on $\hat{\Pi}_t$, and on m_0 , may lead to errors more frequently penalized under HMSE and HMAE. On the other hand, the convergence of C -forecasts to a larger unconditional moment σ_1^2 makes this method less successful at longer horizons. It remains to analyze the source of the sudden increase in forecasting errors for low \bar{k} of the C -forecasts for $N = 10$. Noteworthy in Table 4.5 is, for instance, that for $\bar{k} = 3$ the HMSE jumps from 0.692 at $N = 1$ to 2.09 at $N = 10$, while

decreasing only very slowly back at longer horizons. In-sample, not only the average calibrated parameters \bar{m}_0 and $\bar{\sigma}_1$ of Table 4.3 increase the lower \bar{k} is, but also their dispersion does. The reason for this resides in the higher variability of the calibrated parameters required to fit the daily implied volatility surfaces across \mathcal{T} , which may occasionally contain very large parameter values m_0 and σ_1 for low \bar{k} . Then again, when forecasting, the shifting of probability mass from $\widehat{\Pi}_t \mathcal{A}$, for $N = 1$, to a more ‘uniformly’ distributed vector $\widehat{\Pi}_t \mathcal{A}^N$, for $N = 10$, allocates more weight to those states with very high values. For low \bar{k} , this leads more frequently to erroneous projections of high variability. The combination of forecasts leads, in this case, to a decrease in the parameter values, which helps generally improve upon C -forecasts alone. As \bar{k} increases, however, not only the average calibrated parameters \bar{m}_0 and $\bar{\sigma}_1$ decrease, but also the amount of probability mass that is shifted as we move from $N = 1$ to $N = 10$. As a result, C -forecasts become more stable and the effectiveness of combined forecasts disappears.

In Table 4.6, we turn to the study of accumulated fluctuations for daily, weekly, biweekly, and monthly periods. We also extend our previous analysis of combined forecasts by allowing a more complex weighting scheme of the E - and the C -forecasts. Under MSE, the optimal weighting vector may arise from the regression

$$\tilde{\sigma}_{t+1:t+N}^2 = \beta_0 + \beta_1 \widehat{\mathbb{E}}_{1,t}[\sigma_{t+1:t+N}^2] + \beta_2 \widehat{\mathbb{E}}_{2,t}[\sigma_{t+1:t+N}^2] + u_{t+1:t+N},$$

where $u_{t+1:t+N}$ are the usual OLS errors, β_s for $s = 0, 1, 2$ are the optimal weights to be estimated, $\tilde{\sigma}_{t+1:t+N}^2$ is specified in eq. (4.14), $\widehat{\mathbb{E}}_{1,t}[\sigma_{t+1:t+N}^2]$ and $\widehat{\mathbb{E}}_{2,t}[\sigma_{t+1:t+N}^2]$ represent the E - and the C -forecasts according to eq. (4.13), respectively, and $N = 1, 5, 10, 20$ (cf. Granger and Ramanathan^[186]). Clearly, the coefficients β are

unknown at the time of performing the forecasts, but by applying on them certain restrictions, this framework allows us to assess the performance of the alternative forecasting strategies. Note, for instance, that when we set $\beta_2 = 0$, we can test whether $(\beta_0, \beta_1) = (0, 1)$ for the E -forecasts, which is equivalent to test for *efficiency* according to the definition of Mincer and Zarnowitz^[187]. Alternatively, we restrict $\beta_1 = 0$ and test whether $(\beta_0, \beta_2) = (0, 1)$ for the C -forecasts. For completeness, we perform the same analysis for the equally weighted $(E + C)_1$ -forecasts by treating them as single forecasts. We fix consequently $\beta_2 = 0$, say, and test $(\beta_0, \beta_1) = (0, 1)$. Finally, the assessment of the optimal contributions of the E - and C -forecasts is applied as an alternative test for encompassing (labeled $(E + C)_2$). If the hypothesis $(\beta_0, \beta_1, \beta_2) = (0, 1, 0)$ cannot be rejected, we say that E -forecasts encompasses C -forecasts, and if the hypothesis $(\beta_0, \beta_1, \beta_2) = (0, 0, 1)$ cannot be rejected, the opposite is stated. For any other $(\beta_0, \beta_1, \beta_2)$ values, both models contain useful information about $\tilde{\sigma}_{t+1:t+N}^2$ (cf. Diebold and Lopez^[185]). For the evaluation of these hypotheses, we use a Wald form of the OLS F -test for m hypotheses. For the assessment of efficiency in the Mincer-Zarnowitz sense, the statistic is asymptotically χ^2 distributed with $m = 2$ degrees of freedom, and alternatively, for encompassing, it is asymptotically χ^2 distributed with $m = 3$ degrees of freedom.²⁵ In evaluating the statistics, we apply the Newey-West^[137,167] heteroscedasticity- and autocorrelation-consistent (HAC) estimator for the calculation of the covariance matrix of the OLS coefficients, where we have used

²⁵ Note that for the test of efficiency in the Mincer-Zarnowitz sense, the adjustments for parameter uncertainty simplify to a constant to be multiplied to the covariance matrix of the OLS coefficients. For the E -forecasts, we have argued that this constant is not significantly different from one (cf. West and McCracken^[188]). The use of this procedure for the tests based on C -forecasts, on the other hand, may be arguably unnecessary. We have opted, therefore, to omit the adjustments for any of the tests involving combined forecasts.

prewhitening with a VAR(1) process prior to the calculation.²⁶ Once the inference is performed, we have recolored the HAC estimator to obtain a VAR-prewhitened kernel estimator of the covariance matrix (cf. Andrews and Monahan^[189], Newey and West^[167], and Sul *et al.*^[190]).

Based on the R^2 prediction coefficient, we can see that C -forecasts always outperform E -forecasts, confirming that the calibrated MSM does incorporate forward-looking information effectively, at least up to $N = 20$. C -forecasts at $\bar{k} = 9$ are also generally efficient in the Mincer-Zarnowitz sense. At lower \bar{k} though, they may be *inefficient* for $N \leq 5$. E -forecasts, on the other hand, are always *efficient* despite their OLS coefficients being larger in absolute value.

Regarding combined forecasts, we notice first of all that the $(E + C)_1$ -forecasts are the least *efficient* sets of predictions across \bar{k} . The combination of forecasts by simple averaging leads here to larger OLS coefficients than those associated with the C -forecasts, but less dispersed residuals. As a result, we obtain larger values of the test statistics, and the efficiency hypothesis of the $(E + C)_1$ -forecasts gets often rejected. In terms of predictive ability, $(E + C)_1$ -forecasts perform uniformly worse for $\bar{k} = 9$ than C -forecasts. Moreover, the R^2 levels for $(E + C)_2$ -forecasts, together with the size of the coefficients β_2 along N , show signs of C -forecasts being the only necessary source of information. This seems to be confirmed by the large p-values found for the encompassing test $(\beta_0, \beta_1, \beta_2) = (0, 0, 1)$, indicating that the hypothesis of C -forecasts encompassing E -forecasts cannot be rejected.

The assertion that E -forecasts encompass C -forecasts, on the other hand, is not

²⁶ First suggested by Andrews and Monahan^[189], prewhitening has become a standard tool in estimations based on integrated (overlapping) observations. In our setting, the partial autocorrelation functions (PACF) of the cross-product between the ‘instruments’ $\widehat{E}_{1,t}[\sigma_{t+1:t+N}^2]$ and $\widehat{E}_{2,t}[\sigma_{t+1:t+N}^2]$ and the residuals $u_{t+1:t+N}$ confirm the existence of a significant temporal dependence at the first lag that increases in importance with N .

supported by the data. For $\bar{k} = 3$, and 6, combined forecasts increase R^2 slightly upon C -forecasts, with $(E + C)_1$ and $(E + C)_2$ yielding practically the same results. Note that the fact that both $(E + C)_1$ and $(E + C)_2$ possess the same R^2 figures, while their abscissas β_0 differ only slightly in relative terms, implies that the weighting scheme employed in $(E + C)_1$ is close to being ‘optimal’. The strong rejections found for the encompassing test $(\beta_0, \beta_1, \beta_2) = (0, 1, 0)$ indicate however that these increases in R^2 are not significant. The exception is found for $N = 1$, where neither forecasts encompasses the other, and both sources of information become useful for generating predictions.

Results from Tables 4.5 and 4.6 lead us to conclude that at shorter horizons the MSM model calibrated to DAX option prices conveys more information about the future QV of the DAX index than the traditionally estimated MSM model does. As the forecasting horizon increases, however, the information included in option prices becomes less precise and historical time series provide for a better coverage to the long-memory feature of the model. Our results suggest also that the effectiveness of forecast combination depends greatly on the evaluation scheme applied. The evidence of a general enhancement in forecasting accuracy by combination of historical and forward-looking information is therefore inconclusive.

4.6 Conclusion

In this chapter, we have investigated a method for extracting information from option prices about future volatility. We have also analyzed the potential increase

in forecasting accuracy of the MSM model when exploiting this source of information, as compared to forecasts based on historical time series only.

As a point of departure, we introduced a MSM model for the evolution of asset returns in the usual ‘statistical’ world, adding a risk premium to the traditional MSM model from Calvet and Fisher^[7]. Following, we have derived an alternative version of the model factoring out the risk-preference parameter under the assumed equivalent martingale measure. A small Monte Carlo simulation analysis of the calibration scheme has demonstrated its ability to recover the parameters of the process.

The MSM model is consistent with a symmetric process producing U-shaped smiles. The OLS analysis of the calibration errors though, provides evidence that DAX put contracts are traded differently than call contracts in the market. Clearly, the baseline MSM model is unable to accommodate such a *stylized fact*, and the calibration procedure leaves information behind in this case. The application of the calibration methodology is nonetheless successful in that we are able to capture the main differences between the ‘statistical’ and the ‘preference-free’ volatilities of our underlying. This is confirmed out-of-sample by the differences in sign and size of the projection errors obtained when forecasting with both sets of parameters.

Forecast evaluations have been performed for up to 100-day-ahead forecasts of QV under different loss functions. For shorter horizons, the MSM model calibrated to DAX option prices conveys more information about the future QV of the DAX index than the traditionally estimated MSM model does. This is confirmed by all error statistics as well as by the regression analysis. As the forecasting horizon increases though, the information extracted from option prices becomes less

precise and historical time series provide a better coverage to the long-memory feature of the model. In any case, the MSM model does not seem to outperform the RW model at longer horizons. Our results suggest also, that the usefulness of forecast combination depends greatly on the evaluation function and time horizon applied. The evidence of a general enhancement in forecasting accuracy by combination of historical and forward-looking information is therefore weak, at best. An hypothetical investor would consequently be better-off using predictions based on one information source only, depending on her time horizon.

Several extensions to this line of work are planned for the future. First of all, we would like to consider the inclusion of an additional dividend-adjusted index in order to draw comparisons as in de Jong and Lehnert^[153]. This extension would allow us to obtain more general results as well as to test the causal directions between the forward-looking and the historical sources of data, to see whether the forecasting accuracy can be predetermined, for instance, by Granger-causality of one series upon the other. Further, a thorough comparison against long-memory stochastic volatility (SV) and/or GARCH, FIGARCH and MS-GARCH processes would allow to assess the relative effectiveness of forward-looking information when incorporated into the MSM and the other models. The fact that the MSM model was unable to retrieve some of the features observed in the volatility smiles stimulates us, on the other hand, to search for a more complex model. The study of a MSM model with asymmetric volatility responses may, therefore, represent a promising research avenue in the future (cf. Appendix B for more details).

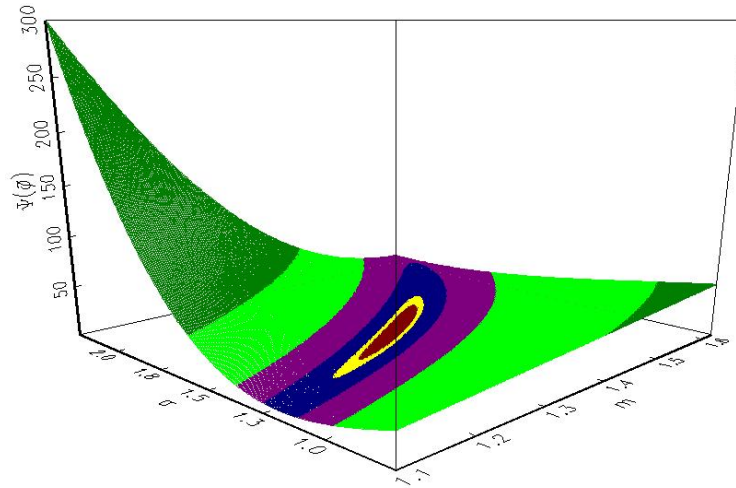


Figure 4.1: Objective function $\Psi(\tilde{\varphi})$ based on empirical data of 12/06/2007, weights $w_i = 1/C_i^2$, and a transition matrix \mathcal{A} based on $(b, \gamma_k) = (2, 0.5)$.

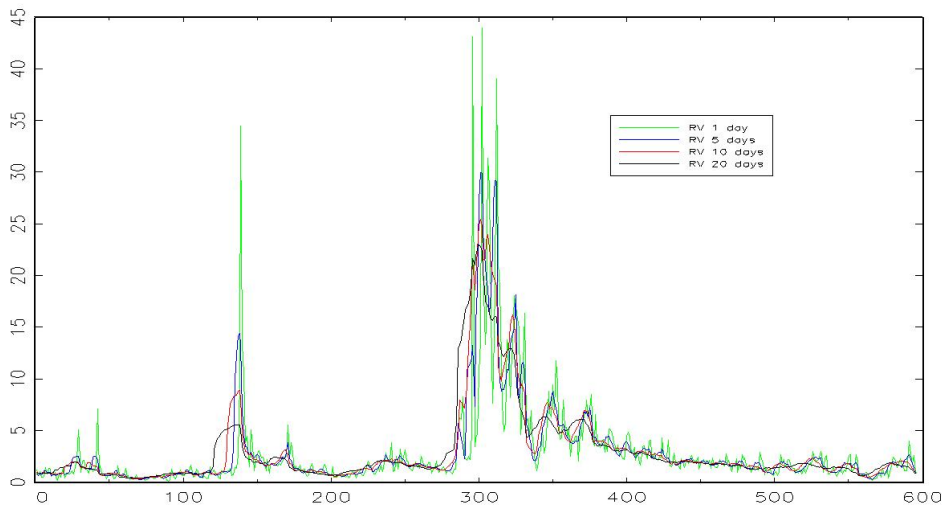


Figure 4.2: Integrated realized Volatility $\tilde{\sigma}_{t:t+N-1}^2$ for four different horizons N as defined in eq. (4.14). The sample of intraday FDAX returns starts in July 2007 and runs until February 2010.

4.6 Conclusion

Table 4.1: Simulated calibration exercise.

S=10.000			S=50.000			S=100.000		
MMSE	\bar{m}_0	$\bar{\sigma}_1$	MMSE	\bar{m}_0	$\bar{\sigma}_1$	MMSE	\bar{m}_0	$\bar{\sigma}_1$
0.080	1.399	1.488	0.051	1.399	1.498	0.020	1.399	1.497
(.085)	(.003)	(.011)	(.043)	(.002)	(.006)	(.019)	(.001)	(.004)

NOTE: Mean and standard deviation of 200 sets of calibrated parameters, obtained under alternative simulation sample sizes \mathcal{S} . One sample of artificial prices C_i has been generated for each size \mathcal{S} and calibrated 200 times using weights $w_i = 1/C_i^2$. For each of these calibrations, a different set of pseudo-random numbers was employed. The artificial prices were generated according to the number of contracts \mathcal{N}_ℓ , the strikes K , the risk-free interest rate r_f , the time to maturity $T - t$, and the current price of the DAX index available in the market on the 12/06/2007, as well as $m_0 = 1.4$, $\sigma_1 = 1.5$, and the latent conditional probabilities $\hat{\Pi}_t$ from the historical data for an underlying model with $\bar{k} = 9$. \bar{m}_0 and $\bar{\sigma}_1$ denote the average among the calibrated parameters, while MMSE denotes the average mean squared calibration error $\frac{100^2}{\mathcal{N}_\ell} * \Psi(\hat{\varphi})$ among the 200 samples.

Table 4.2: Monte Carlo results for calibration of parameters.

			MMSE	\bar{m}_0	$\bar{\sigma}_1$			
			0.069	1.400	1.500			
			(.112)	(.002)	(.006)			
β_0	t-stat	β_1	t-stat	β_2	t-stat	β_3	t-stat	R^2
-6.440	-1.398	3.361	0.737	23.075	2.050	1.512	0.127	0.010
(.162)		(.461)		(.040)		(.898)		(.296)

NOTE: Monte Carlo results from 200 calibrations. Each calibration has been performed for an independent sample ℓ , $\ell = 1, 2, \dots, 200$, of artificial prices $C_{i,\ell}$. These samples were generated based on the available market information, by linking together a trading day in our data set with an instance of the index ℓ . Starting on 12/06/2007, the procedure consisted of (i) selecting all the empirical information available for the day, (ii) replacing the empirical prices $C_{i,\ell}$ with artificial ones generated according to the procedure used for the evaluation of $C_{i,\ell}^{\hat{\varphi}}(T, K)$ in eq. (4.8), a set of pseudo-random numbers, and the latent conditional probabilities $\hat{\Pi}_t$ for an underlying model with $\bar{k} = 9$, together with the information collected in the previous step and parameter values $m_0 = 1.4$ and $\sigma_1 = 1.5$; and (iii) executing the calibration with weights $w_i = 1/C_i^2$.

(top panel) Mean and standard deviation among the 200 sets of calibrated parameters. MMSE denotes the average mean squared error $100^2/\mathcal{N}_\ell \cdot \Psi(\hat{\varphi})$ across the 200 samples, where \mathcal{N}_ℓ and $\Psi(\hat{\varphi})$ indicate the number of contracts and the sum of squared errors resulting from each sample, respectively.

(bottom panel) Parameters, t-statistics, p-values (in parentheses), and R^2 figures from the OLS of eq. (4.11) performed on the total amount of calibration errors ϵ_j across the 200 samples. R^2 has been adjusted for the number of variables included. The t-statistic has $\sum_{\ell=1}^{200} \mathcal{N}_\ell - 4$ degrees of freedom. The F -test on $R^2 = 0$ is asymptotically χ^2 distributed with 4 degrees of freedom according to its Wald form. The t- and F -tests have been calculated according to the Newey-West^[137,167] heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator.

Table 4.3: In-sample Analysis: Empirical estimates and Goodness-of-fit.

		Estimated MSM			Calibrated MSM			Ad-hoc B-S					
		$\bar{k} = 6$		$\bar{k} = 9$		$\bar{k} = 3$		$\bar{k} = 6$		$\bar{k} = 9$			
Param	Value	Param	Value	Param	Value	Param	Value	Param	Value	Param	Value		
\bar{m}_0	1.484 (.034)	\bar{m}_0	1.320 (.021)	\bar{m}_0	1.245 (.015)	\bar{m}_0	1.248 (.236)	\bar{m}_0	1.210 (.182)	\bar{m}_0	1.198 (.169)	β_0	31.901 (22.857)
$\bar{\sigma}_1$	1.292 (.090)	$\bar{\sigma}_1$	1.275 (.083)	$\bar{\sigma}_1$	1.277 (.094)	$\bar{\sigma}_1$	1.450 (.367)	$\bar{\sigma}_1$	1.431 (.350)	$\bar{\sigma}_1$	1.415 (.328)	β_1	-4.170 (22.508)
α	.500 (.007)	α	.500 (.007)	α	.500 (.007)							β_2	-2.476 (13.796)
L	-1.559 (.036)	L	-1.527 (.033)	L	-1.515 (.031)	MMSE	1185.049 (372.391)	MMSE	1164.072 (367.208)	MMSE	1156.226 (360.796)	MMSE	29.808 (37.524)

NOTE: Mean and standard deviations for the estimates from ML estimation, the calibrated parameters from the Simulated (S)-NLS, and the ad-hoc estimates from the Black Scholes model. MMSE denotes the mean of $\frac{100^2}{N_\ell} * \Psi(\hat{\varphi})$, for $\ell \in \mathcal{T}$. Average values were obtained over a rolling-window period of $\mathcal{T} = 596$ days. L denotes the mean maximized likelihood for a sample of 2636 days. MMSE' represents MMSE recalculated in terms of implied volatilities for better comparison against the MMSE of the ad-hoc Black Scholes model.

Table 4.4: In-sample Error Analysis for Calibrated Option Prices.

\bar{k}	β_0	t-stat	β_1	t-stat	β_2	t-stat	β_3	t-stat	β_4	t-stat	β_5	t-stat	R^2
3	161.754	10.264 (.000)	-290.218	-15.640 (.000)	-164.059	-10.153 (.000)	282.841	14.590 (.000)	-69.353	-22.115 (.000)	-175.551	-8.422 (.000)	0.203 (.000)
6	161.056	10.251 (.000)	-287.350	-15.614 (.000)	-162.860	-10.107 (.000)	279.939	14.564 (.000)	-73.511	-27.663 (.000)	-169.719	-8.320 (.000)	0.207 (.000)
9	160.441	10.226 (.000)	-286.909	-15.597 (.000)	-161.534	-10.040 (.000)	279.452	14.544 (.000)	-76.610	-28.143 (.000)	-169.592	-8.322 (.000)	0.210 (.000)

NOTE: Parameters, t-statistics, p-values (in parentheses), and R^2 figures from the OLS of eq. (4.16) performed on the sum of calibration errors ϵ_i across $\mathcal{T} = 596$ days. R^2 has been adjusted for the number of variables included. The t-statistic has $\sum_{\ell=1}^{\mathcal{T}} \mathcal{N}_{\ell}-4$ degrees of freedom. The F -test on $R^2 = 0$ is asymptotically χ^2 distributed with 6 degrees of freedom according to its Wald form. The t- and F -tests have been calculated according to the Newey-West [137,167] heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator.

Table 4.5: Out-of-sample Forecasting Results. Error Statistics.

\bar{k}	Method	MSE					MAE				
		N = 1	N = 10	N = 20	N = 50	N = 100	N = 1	N = 10	N = 20	N = 50	N = 100
3	E	.839*	.986*	.997*	1.000	1.001 [†] C	.879	.984	.996	1.000	1.002
	C	.661* [†] E	.903* [†]	.923*	.998*	1.035	.745 [†] E	.931	.973	1.062	1.129
	$(E + C)_1$.731* [†] E	.930*	.943*	.986*	1.005	.775 [†] E	.928	.950	.989 [†] C	1.028
6	E	.717*	.913*	.952*	.995	1.003 [†] C	.797	.924	.959	.992	1.001
	C	.564* [†] E	.832* [†]	.888*	1.001	1.031	.669 [†] E	.868	.932	1.048	1.109
	$(E + C)_1$.623* [†] E	.863*	.908*	.987 [†] C	1.005 [†] C	.702 [†] E	.875	.923	.992 [†] C	1.023 [†] C
9	E	.636*	.840*	.895*	.988	1.008	.757	.894	.952	1.026	1.049
	C	.481* [†] E	.736*	.828*	.994	1.031	.607 [†] E	.797	.887	1.040	1.116
	$(E + C)_1$.524* [†] E	.776*	.850*	.983	1.012	.651 [†] E	.830 [†] E	.902	1.016	1.066

\bar{k}	Method	HMSE					HMAE				
		N = 1	N = 10	N = 20	N = 50	N = 100	N = 1	N = 10	N = 20	N = 50	N = 100
3	E	.633	.965	1.004	1.010	1.021	.849	.988	1.001	1.005	1.009
	C	.692	2.090	1.900	1.523	1.513	.857	1.189	1.225	1.239	1.294
	$(E + C)_1$.623	1.378	1.327	1.190	1.188	.827	1.061	1.084	1.088	1.110
6	E	.488	.822	1.002	.988	.975	.776	.940	.994	.997	.993
	C	.462	1.313	1.477	1.424	1.442	.726	1.047	1.147	1.214	1.252
	$(E + C)_1$.454	.993	1.164	1.136	1.132	.731	.977	1.052	1.082	1.087
9	E	.475	.821	1.117	1.134	1.424	.781	.962	1.066	1.116	1.186
	C	.341	.726	1.089	1.232	1.438	.632	.875	1.023	1.163	1.282
	$(E + C)_1$.385	.749	1.071	1.146	1.375	.687	.908	1.032	1.127	1.216

NOTE: (top panel) Mean squared errors (MSE), mean absolute errors (MAE), (bottom panel) heteroscedasticity-adjusted counterparts (HMSE and HMAE) from time series estimated MSM (E), from implied calibrated MSM (C), and from a combination of both ($(E + C)_1$) with equal weights as a % of the pertinent error statistics from a 'naive' forecast using the historical variance from the detrended daily-return series -referred to as random walk (RW). The statistics for each the forecasting horizons $N = 1, 10, 20, 50, 100$ were calculated over a sample period of $T = 596$ days.

(top panel) All comparisons against RW under MSE are based on the test statistic for nested models of Clark and West^[124,125]. Comparisons of C against E , and of C and E against $(E + C)_1$ are based on the modified Diebold and Mariano^[126] test statistic by Harvey *et al.*^[127]. The test for forecast encompassing under MSE is based on Harvey *et al.*^[180].

* denotes an improvement against the RW model which is significant at the 95% level (under MSE only). ** denotes an improvement against the RW model which is significant at the 99% level (under MSE only). [†] i denotes an improvement significant at the 95% level (only among E, C , and $(E + C)_1$, with subscripts $E, C, E + C$ respectively). ^{††} i denotes an improvement significant at the 99% level (only among E, C , and $(E + C)_1$, with subscripts $E, C, E + C$ respectively). [‡] denotes a significance at the 95% level that E does not encompass C (under MSE only). # denotes a significance at the 99% level that E does not encompass C (under MSE only).

Table 4.6: Out-of-sample Forecasting Results. Regression Analysis.

\bar{k}	Method	N = 1				N = 5				N = 10				N = 20							
		β_0	β_1	β_2	p-val	R^2	β_0	β_1	β_2	p-val	R^2	β_0	β_1	β_2	p-val	R^2	β_0	β_1	β_2	p-val	R^2
3	E	-2.747	2.776		(.15)	.242	-3.572	3.439		(.16)	.189	-3.854	3.770		(.26)	.134	-2.459	3.106		(.37)	.060
	C	-1.483		1.792	(.02)	.382	-1.055		1.696	(.14)	.308	-0.580		1.548	(.32)	.237	0.227		1.215	(.60)	.151
	$(E + C)_1$	-2.980	2.631		(.01)	.395	-2.953	2.778		(.01)	.329	-2.535	2.682		(.03)	.253	-1.349	2.155		(.07)	.155
	$(E + C)_2$	-2.619	.930	1.487	(.00)	.397	-2.947	1.384	1.390	(.00)	.328	-2.824	1.564	1.291	(.00)	.252	-1.150	.933	1.101	(.03)	.154
6	E	-2.357	2.312		(.07)	.378	-2.484	2.509		(.18)	.340	-2.706	2.739		(.28)	.300	-2.438	2.733		(.30)	.215
	C	-1.008		1.518	(.00)	.459	-0.928		1.565	(.04)	.423	-0.941		1.635	(.12)	.387	-0.436		1.462	(.32)	.274
	$(E + C)_1$	-1.926	1.994		(.00)	.465	-1.958	2.115		(.01)	.430	-2.100	2.271		(.04)	.393	-1.659	2.148		(.08)	.286
	$(E + C)_2$	-1.628	.627	1.207	(.00)	.467	-1.643	.691	1.251	(.00)	.431	-1.748	.747	1.325	(.00)	.395	-1.560	.973	1.117	(.01)	.285
9	E	-1.499	1.737		(.07)	.403	-1.365	1.752		(.22)	.389	-1.289	1.784		(.39)	.351	-0.756	1.608		(.53)	.254
	C	-0.136		1.079	(.40)	.493	-0.091		1.123	(.62)	.481	-0.229		1.229	(.41)	.488	-0.025		1.186	(.54)	.395
	$(E + C)_1$	-0.883	1.417		(.00)	.488	-0.821	1.460		(.04)	.457	-0.925	1.561		(.11)	.463	-0.608	1.477		(.22)	.363
	$(E + C)_2$	-0.425	.253	.955	(.00)	.494	-0.340	.227	1.010	(.00)	.482	-0.061	-0.157	1.309	(.00)	.488	.438	-0.446	1.413	(.00)	.399

NOTE: Parameter estimates of the OLS regression of RV on integrated volatility forecasts from time series estimated MSM (E), from implied calibrated MSM (C), and from a combination of both ($(E + C)_2$) for the forecasting horizons $N = 1, 5, 10, 20$ and a sample period of $T = 596$ days. $(E + C)_1$ refers to forecasts obtained as simple averages of E and C , while $(E + C)_2$ refers to optimal weights arising from the OLS regression. R^2 has been adjusted for the number of variables included. The values in parenthesis state the p-value of the F -statistic that is asymptotically χ^2 distributed with 2 degrees of freedom for E , C , and $(E + C)_1$, and it is asymptotically χ^2 distributed with 3 degrees of freedom for the encompassing tests under $(E + C)_2$ (the first row is used for $(\beta_0, \beta_1, \beta_2) = (0, 1, 0)$ and the second one for $(\beta_0, \beta_1, \beta_2) = (0, 0, 1)$). The F -tests have been calculated according to the Newey-West^[137,167] heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator after prewhitening of the residuals by a VAR(1) process.

Conclusions and Future Research Avenues

Conclusion

Multifractal models arise as an important alternative to the mainstream literature of financial time series analysis. These models stand out by their parsimony and their focus on temporal multiscaling. In their causal form, not only fit these models into the literature of regime-switching and Itô diffusions, but also they may achieve to replicate almost all the *stylized facts* of the return time series.

A number of original contributions to the existing literature of multifractal models has been made in this thesis. In Chapter 1, we have focused on the theoretical construction of multifractal processes, setting the foundations for the rest of this work. The following chapters are dedicated to independent, self-contained topics regarding the estimation and forecasting of multifractal processes.

Chapter 2 has proposed a GMM approach for estimation of Lognormal cascade processes, which are traditionally important in the literature on turbulent flows. The procedure compares favorably with previously proposed estimators of that body of research. The numerical analysis performed in this chapter suggests that

the GMM estimator is indeed consistent and asymptotically Normally distributed. The employed methodology allows further to retrieve the cascade parameter value with high accuracy even when the number of levels of the cascade is unknown. To apply the estimates obtained to forecast the future evolution of a cascade, a methodology based on the Levinson-Durbin algorithm for best linear forecasts has been developed. The methodology circumvents the statistical problems related to the definition of a cascade process on a bounded interval by allowing for a new initialization of the process each time the endpoint of the cascade is reached. It is also shown that the size of the interval, i.e. the number of cascade steps, has virtually no influence on the estimated intermittency parameter. Similarly, the predictive power of forecasts based on past realizations is relatively insensitive to the number of steps beyond some threshold. The applicability of the procedure is confirmed by an extensive simulation analysis. The empirical application consists in the estimation of the intermittency parameter and the forecasting of volatility for various foreign exchange markets and the gold market. Results finally suggest that cascade models, even with their grid-bound nature of volatility components, capture a non-trivial part of the variability of price fluctuations.

Chapter 3 has proposed a two-step mixed SMM estimation approach for the continuous-time MSM model with finitely many levels. The first step consists in the application of Lux's^[8] GMM procedure. Results from this step are incorporated as starting values in an iterative SMM algorithm. Due to the prior execution of the GMM procedure to the same sample, final convergence of the algorithm ensues on average already after the second iteration. To confirm this, a thorough Monte Carlo analysis for the evaluation of the methodology is pursued. As a first step, an analysis of the loss in accuracy incurred when employing the

misspecified GMM estimator of Lux^[8] to the incomplete record generated by the discretely sampled ‘continuous’ time process is carried out. Following, the introduction of the SMM methodology is evaluated. Monte Carlo experiments show that the SMM implementation is able to deliver reliable results for both sets of moments, those applied to the daily returns and those applied to the RV_t series. The applicability of the continuous-time MSM model is also explored empirically, on foreign-exchange and stock-index time series. As it turns out, when the set of moments based on daily returns are used, the MSM model works well in the foreign exchange and the DAX, CAC40 and FT100 time series. Results obtained by the SMM estimator based on QV moments, on the other hand, show that this version of the MSM model does not accommodate the stock index series tightly. A reason for this may be the misspecification arising from unmodeled sources of return fluctuation found in the empirical data. Note, then again, that the more efficient use of intradaily information when using the RV series may be better suited to detect differences between the data and the hypothesized data generating process. In this respect, the rejections observed with the moments applied to the RV series would merely be an indication that, like any model, the one based on the RV series would be rejected when a sufficient amount of data is available. Further analysis of the high-frequency observations in this data set should shed light on this issue in the future.

Chapter 4 has investigated first a method for extracting information contained in option prices. After this, the potential increase in forecasting accuracy of the MSM model is analyzed, when this source of data is complemented with the available return series. The empirical implementation of the MSM model shows that the model is unable to accommodate the asymmetry of odds between pos-

itive and negative returns present in the ‘risk-neutral’ world of the DAX index. That is, the calibration procedure cannot retrieve all the information available in the option prices. The application is nonetheless successful in capturing the main differences between the ‘statistical’ and the ‘preference-free’ volatilities of the underlying. This is confirmed out-of-sample by the differences in sign and size of the projection errors obtained when forecasting with both sets of retrieved parameters. Evaluations have been performed for up to 100-day-ahead forecasts of QV under different loss functions. For shorter horizons, the calibrated MSM model conveys better information about the future QV of the DAX index than the traditionally estimated MSM model does. This is confirmed by all error statistics as well as the Mincer-Zarnowitz regressions. As the forecasting horizon increases, however, the information contained in option prices becomes less precise and historical time series provides a better input to the long-memory feature of the model. These results suggest, finally, that the effectiveness of forecast combination depends greatly on the evaluation function and time horizon applied. Moreover, the evidence of a general enhancement in forecasting accuracy by combination of historical and forward-looking information appears negligible. This would lead one to conclude that an hypothetical investor would be better-off using predictions based on one information source only, depending on her time horizon.

Future research avenues

A number of issues left aside in this thesis constitute avenues for possible future research. Among them, the asymmetry of volatility feedbacks on past return

innovations deserves special attention from my point of view. I will elaborate merely on this issue here and dedicate Appendix B at the end of this thesis for more details.

The importance of capturing the asymmetry of volatility responses to past return innovations resides not only in the obvious fact that as a well documented statistical regularity it should be incorporated in the analysis of asset returns but also in that the methodology employed to depict it may facilitate the study of other related, though more difficult to replicate, attributes of financial time series. This is particularly true when adopting a time-series modeling approach. In a return series exhibiting negative skewness and an increase of volatility after a large negative shock, for instance, one may be able to distinguish parametrically the asymmetric volatility response in the martingale component from a term possibly causing the one-sided jumps of the return distribution. The approach may also lead to distinguish further more subtle regularities of volatility. Among the latter, one should highlight that asymmetric volatility responses may also vary according to the magnitude of the return innovations, not only its sign, and that the effect of certain negative innovations may be more persistent than other (cf. Wu and Xiao^[191] and Yu^[192]).

Note from the literature on multifractal models and the contributions presented in this thesis, that the symmetry of the volatility responses belongs to one of the pillars of MSM models. This feature arises as a consequence of **Condition 3** introduced in the last proposition of Chapter 1, and it seems also to be shared by most types of multifractal models, signaling how deep-rooted it is in the construction (cf. Zhong and Zhao^[93]). A modification of this property for the inclusion of ‘asymmetric’ effects represents hence a great challenge that may not

be easily undertaken without altering other highly appreciated properties of the model. Because of this, Calvet and Fisher^[112] suggest to directly avoid the issue to turn, alternatively, to a link between volatility and prices by using equilibrium valuation methods. Their approach, however, is focused on dividend-paying stocks and requires the acceptance of equilibrium-based valuation relationships that are otherwise troublesome to extrapolate to other type of assets without further adjustments. Future work from my part will consequently focus on alternative approaches within the context of pure time-series modeling of the MSM process.

One approach to this problem that has already been effectively tested, but not fully completed by the time this thesis had been submitted, is presented in Appendix B. It consists in relaxing the assumption about the marginal distribution of the volatility shocks when new arrivals occur, providing a wide spectrum of potential responses from the volatility vector. Further research in this area involves the Monte Carlo evaluation in the immediate future of the estimation procedure presented in Appendix B.

Following work in the asymmetric version of the MSM model contemplates a series of empirical evaluations of the model and a comparison with the up-to-date most prominent contestants. In this venue, the adoption of a drift component dependent on volatility, as it has previously been used in Chapter 4, may demonstrate worthwhile. This would allow us to compare this construction with the well-known GARCH-M and SV-M models, adapted for asymmetric volatility responses (cf. Jondeau *et al.*^[19]).

Note also that a large amount of literature dedicated to the analysis of option pricing has led to suggest that the theoretical design of a parsimonious model

for the pricing of contingent claims, that contemplates asymmetric and jumpy volatility responses, volatility clustering, and hyperbolic decline in volatility autocorrelations, is a conundrum. The inclusion of these features in the design of an option pricing model based on the asymmetric version of the MSM process, together with an efficient valuation methodology, may thus prove highly relevant (cf. Pan^[39], Eraker^[41], Garcia *et al.*^[51], and Chernov *et al.*^[38], among others). Extensions to a multivariate field may also prove fruitful. Note, for instance, that a potential bivariate asymmetric MSM process may be applicable not only to the study of return time series of small portfolios but also to model the dependence between volatility and trading volume of single assets (cf. Calvet *et al.*^[107] and Liu and Lux^[108] for an approach towards bivariate symmetric MSM models, and Jondeau *et al.*^[19] and Liesenfeld^[193], among others, for an introduction to the study of volatility and trading volume comovements). The latter case may even allow for heterogenous levels of adjustment in volatility and trading activity, a feature hardly contemplated in current studies.

Appendix A

Additional Estimation Results for Chapter 2

In this appendix, complementary results for the GMM estimation procedure presented in Chapter 2 are exhibited, together with figures for the $\widehat{\lambda}_q^2$ estimator of Kiyono *et al.* plus the simple sample estimator $\widehat{\sigma}_1$ for better comparability. Additionally results are included for the $\widehat{\lambda}_q^2$ and $\widehat{\chi}^2$ estimators under different choices of q , and bin type and number of bins, respectively. For a more comprehensive analysis of the results, all tables present the mean of the estimated parameters, the finite sample standard errors (FSSE) and the root mean squared errors (RMSE).

Table A.1 displays GMM results for $\lambda_0^2 = 0.01, 0.025, \dots, 0.15$ and $n = 8, \dots, 16$ with intermediate parameter values among those selected in the main text. Table A.2 displays the results for the $\widehat{\lambda}_q^2$ and $\widehat{\sigma}_1$ estimators under similar parameter choices.

Table A.3 introduces the results for the $\widehat{\lambda}_q^2$ estimator under different q values. The data has been generated considering a $\lambda_0^2 = 0.05$ and $\sigma_1 = 1$. Results have

been normalized by n to ease comparison to the values from Table A.2.

Table A.4 displays estimation outcomes for the $\widehat{\chi}^2$ estimator under different number of bins $w = 10, \dots, 100$ of equal probability mass each. As before, the data has been generated by choosing $\lambda_0^2 = 0.05$, and $\sigma_1 = 1$. In all cases, results have been normalized by n to better comparison.

At last, Table A.5 displays estimation outcomes for the $\widehat{\chi}^2$ estimator under different number of bins $w = 10, \dots, 100$. In this case, bins are chosen by fixing the support-size $[x_{j-1}, x_j]$ and equally distributing the support space by the number of bins w .

Table A.1: Monte Carlo results for the GMM estimator.

n	$\lambda_0^2 = 0.01$			$\lambda_0^2 = 0.025$			$\lambda_0^2 = 0.05$			$\lambda_0^2 = 0.1$			$\lambda_0^2 = 0.15$			
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	
8	$\bar{\lambda}_0^2$	0.010	0.011	0.011	0.024	0.025	0.026	0.047	0.049	0.050	0.094	0.098	0.099	0.142	0.146	0.148
	FSSE	.009	.008	.005	.012	.009	.005	.017	.010	.007	.019	.013	.009	.023	.016	.011
	RMSE	.009	.008	.005	.012	.009	.005	.017	.011	.007	.019	.013	.009	.024	.016	.011
	$\bar{\sigma}_1$	1.000	1.002	1.000	0.999	1.003	1.000	1.000	0.995	0.998	1.002	1.007	0.999	0.969	0.967	0.993
	FSSE	.039	.027	.019	.061	.044	.028	.094	.066	.048	.161	.133	.073	.214	.158	.139
	RMSE	.039	.027	.019	.061	.044	.028	.094	.066	.048	.161	.133	.073	.216	.161	.139
10	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.023	0.026	0.026	0.047	0.048	0.049	0.094	0.098	0.098	0.142	0.146	0.148
	FSSE	.009	.008	.005	.013	.009	.006	.016	.011	.007	.021	.014	.009	.024	.015	.011
	RMSE	.009	.008	.005	.013	.009	.006	.017	.011	.007	.021	.014	.009	.025	.016	.011
	$\bar{\sigma}_1$	0.993	1.001	1.001	0.999	1.000	0.997	0.995	0.987	0.987	0.986	0.981	0.991	0.950	0.934	0.976
	FSSE	.069	.047	.036	.105	.077	.056	.162	.122	.084	.268	.199	.150	.357	.292	.214
	RMSE	.069	.047	.036	.105	.077	.056	.162	.122	.085	.268	.200	.151	.360	.299	.215
13	$\bar{\lambda}_0^2$	0.011	0.011	0.011	0.024	0.025	0.025	0.044	0.048	0.049	0.094	0.097	0.099	0.142	0.146	0.148
	FSSE	.010	.007	.005	.013	.009	.006	.016	.010	.007	.020	.013	.010	.022	.016	.012
	RMSE	.010	.007	.005	.013	.009	.006	.017	.010	.007	.021	.014	.010	.024	.016	.012
	$\bar{\sigma}_1$	0.997	0.998	1.000	0.976	0.986	0.992	0.942	0.954	0.981	0.871	0.917	0.974	0.797	0.821	0.875
	FSSE	.164	.128	.090	.236	.193	.148	.386	.301	.204	.515	.423	.359	.599	.475	.468
	RMSE	.164	.128	.090	.237	.194	.148	.390	.304	.205	.531	.431	.360	.632	.508	.485
16	$\bar{\lambda}_0^2$	0.010	0.011	0.011	0.023	0.025	0.026	0.047	0.049	0.050	0.093	0.098	0.099	0.142	0.146	0.148
	FSSE	.010	.007	.005	.013	.009	.006	.016	.011	.007	.020	.014	.009	.023	.017	.012
	RMSE	.010	.007	.005	.013	.009	.006	.016	.011	.007	.021	.015	.009	.024	.017	.012
	$\bar{\sigma}_1$	0.976	0.979	0.993	0.960	0.944	0.937	0.916	0.900	0.917	0.814	0.808	0.862	0.611	0.661	0.740
	FSSE	.223	.206	.182	.351	.301	.279	.514	.422	.362	.803	.638	.570	.552	.681	.798
	RMSE	.224	.207	.182	.354	.306	.287	.521	.433	.372	.824	.666	.586	.676	.761	.840

NOTE: All simulations are based on a process with $\varepsilon \sim N(0, \sigma_1^2)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2, 500$, $T_2 = 5, 000$, and $T_3 = 10, 000$. λ_0^2 and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. GMM was executed using lags $\ell = 1, 14, 64$. For each case, 400 Monte Carlo runs have been carried out.

Table A.2: Monte Carlo results for $\widehat{\lambda}_q^2$ estimator.

n	$\lambda_0^2 = 0.01$			$\lambda_0^2 = 0.025$			$\lambda_0^2 = 0.05$			$\lambda_0^2 = 0.1$			$\lambda_0^2 = 0.15$			
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	
8	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.025	0.025	0.025	0.049	0.049	0.050	0.097	0.100	0.099	0.142	0.144	0.147
	FSSE	.002	.002	.001	.004	.003	.002	.007	.005	.004	.015	.014	.009	.024	.018	.017
	RMSE	.002	.002	.001	.004	.003	.002	.007	.005	.004	.015	.014	.009	.025	.019	.017
	$\bar{\sigma}_1$	1.000	1.001	1.000	0.999	1.002	1.000	0.997	0.992	0.996	1.002	1.007	0.997	0.973	0.970	0.993
	FSSE	.037	.026	.019	.058	.042	.027	.084	.063	.046	.154	.127	.071	.222	.153	.141
	RMSE	.037	.026	.019	.058	.043	.027	.084	.063	.046	.154	.127	.071	.223	.156	.141
10	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.024	0.025	0.025	0.048	0.048	0.049	0.093	0.095	0.097	0.136	0.138	0.144
	FSSE	.002	.002	.001	.005	.003	.002	.009	.007	.005	.017	.014	.011	.030	.023	.019
	RMSE	.002	.002	.001	.005	.003	.002	.009	.007	.005	.019	.015	.012	.033	.026	.019
	$\bar{\sigma}_1$	0.993	1.000	1.001	0.998	0.999	0.996	0.991	0.985	0.986	0.982	0.981	0.990	0.963	0.934	0.979
	FSSE	.067	.046	.035	.103	.076	.055	.155	.119	.083	.262	.194	.150	.389	.287	.210
	RMSE	.067	.046	.035	.103	.076	.055	.155	.119	.084	.263	.195	.151	.391	.294	.211
13	$\bar{\lambda}_0^2$	0.008	0.009	0.009	0.021	0.023	0.023	0.040	0.043	0.046	0.077	0.084	0.089	0.112	0.120	0.129
	FSSE	.002	.002	.002	.005	.005	.004	.010	.008	.008	.017	.017	.015	.024	.022	.024
	RMSE	.003	.002	.002	.007	.005	.004	.014	.011	.009	.029	.023	.019	.045	.037	.032
	$\bar{\sigma}_1$	0.997	0.997	1.000	0.975	0.985	0.991	0.942	0.953	0.980	0.869	0.914	0.974	0.801	0.823	0.875
	FSSE	.161	.126	.089	.229	.191	.146	.379	.297	.205	.496	.412	.359	.578	.468	.467
	FSSE	.161	.126	.089	.231	.191	.147	.383	.300	.206	.513	.421	.360	.611	.500	.484
16	$\bar{\lambda}_0^2$	0.007	0.007	0.008	0.017	0.019	0.020	0.033	0.036	0.040	0.064	0.071	0.075	0.090	0.100	0.111
	FSSE	.002	.002	.002	.005	.004	.004	.008	.007	.008	.016	.015	.016	.022	.022	.021
	RMSE	.004	.003	.003	.009	.008	.007	.019	.016	.013	.039	.033	.030	.064	.055	.044
	$\bar{\sigma}_1$	0.973	0.979	0.992	0.943	0.942	0.966	0.879	0.899	0.932	0.754	0.807	0.882	0.637	0.680	0.758
	FSSE	.226	.205	.183	.328	.298	.261	.447	.416	.391	.633	.633	.743	.834	.843	.658
	RMSE	.228	.206	.184	.333	.303	.263	.463	.429	.397	.679	.661	.752	.910	.901	.701

NOTE: All simulations are based on a process with $\varepsilon \sim N(0, \sigma_1^2)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. Kiyono *et al.*'s estimator $\widehat{\lambda}_q^2$ was calculated using $q = 0.5$ after the series was filtered by the sample estimate of σ_1 . The results for $\widehat{\lambda}_q^2$ have been normalized by n . For each case, 400 Monte Carlo runs have been carried out.

Table A.3: Monte Carlo results for $\hat{\lambda}_q^2$ estimator with different q values and $\lambda_0^2 = 0.05$.

k	$q = -0.5$			$q = 0.01$			$q = 0.5$			$q = 1$			$q = 1.5$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
8	$\bar{\lambda}_0^2$	0.049	0.050	0.050	0.049	0.050	0.049	0.049	0.050	0.049	0.049	0.050	0.049	0.049	0.050
	FSSE	.010	.008	.005	.007	.005	.004	.007	.005	.004	.008	.005	.008	.006	.005
	RMSE	.010	.008	.005	.007	.005	.004	.007	.005	.004	.008	.005	.008	.006	.005
10	$\bar{\lambda}_0^2$	0.048	0.049	0.049	0.048	0.049	0.048	0.048	0.049	0.047	0.048	0.049	0.047	0.048	0.048
	FSSE	.009	.008	.005	.008	.007	.004	.009	.007	.005	.009	.008	.010	.008	.006
	RMSE	.009	.008	.005	.008	.007	.005	.009	.007	.005	.010	.008	.011	.009	.006
13	$\bar{\lambda}_0^2$	0.040	0.044	0.047	0.041	0.043	0.046	0.040	0.043	0.046	0.040	0.043	0.046	0.042	0.045
	FSSE	.010	.009	.008	.010	.008	.007	.010	.008	.008	.010	.009	.008	.011	.008
	RMSE	.014	.011	.008	.013	.010	.008	.014	.011	.009	.014	.011	.009	.015	.010
16	$\bar{\lambda}_0^2$	0.034	0.037	0.040	0.034	0.037	0.040	0.033	0.036	0.040	0.033	0.036	0.039	0.036	0.039
	FSSE	.008	.008	.008	.008	.008	.008	.008	.007	.008	.008	.008	.008	.008	.008
	RMSE	.018	.015	.013	.018	.015	.013	.019	.016	.013	.019	.016	.014	.019	.014

NOTE: All simulations are based on a process with $\varepsilon \sim N(0, \sigma_1^2)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, $\lambda_0^2 = 0.05$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0^2$ is the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. Kiyono *et al.*'s estimator $\hat{\lambda}_q^2$ was calculated using different values of q after the series was filtered by the estimate of σ_1 (not shown here). The results for $\hat{\lambda}_q^2$ have been normalized by n . For each case, 400 Monte Carlo runs have been carried out.

Table A.4: Monte Carlo results for $\hat{\chi}^2$ estimator under different number w of bins of equal probability mass and $\lambda_0^2 = 0.05$.

n	$w = 10$			$w = 20$			$w = 50$			$w = 100$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
8	$\bar{\lambda}_0^2$	0.050	0.049	0.050	0.050	0.050	0.051	0.051	0.051	0.052	0.051	0.051
	FSSE	.009	.006	.004	.008	.006	.004	.008	.005	.008	.005	.004
	RMSE	.009	.006	.004	.008	.006	.004	.008	.005	.008	.005	.004
	$\bar{\sigma}_1$	1.020	1.010	1.019	1.024	1.019	1.024	1.032	1.025	1.035	1.028	1.034
	FSSE	.113	.077	.053	.110	.074	.051	.108	.074	.108	.074	.051
	RMSE	.115	.077	.057	.112	.076	.057	.112	.078	.113	.079	.062
16	$\bar{\lambda}_0^2$	0.032	0.034	0.039	0.034	0.036	0.040	0.035	0.037	0.035	0.038	0.041
	FSSE	.012	.013	.011	.012	.012	.011	.012	.012	.012	.011	.010
	RMSE	.021	.021	.016	.020	.018	.015	.019	.018	.019	.016	.013
	$\bar{\sigma}_1$	0.861	0.916	0.939	0.857	0.919	0.910	0.924	1.004	0.935	1.029	1.012
	FSSE	.359	.385	.349	.391	.403	.340	.437	.507	.461	.540	.448
	RMSE	.385	.394	.354	.417	.411	.351	.443	.507	.465	.540	.449

NOTE: All simulations are based on a process with $\varepsilon \sim N(0, \sigma_1^2)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, $\lambda_0^2 = 0.05$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2, 500$, $T_2 = 5, 000$, and $T_3 = 10, 000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. $\hat{\chi}^2$ estimator was calculated using different total number of bins w of equal probability mass. For a given number w , a varying support-size $[x_{j-1}, x_j]$, with $j = 1, \dots, w$, is selected so that in each bin a probability mass $F_{\lambda, \sigma_0}(x_j) - F_{\lambda, \sigma_0}(x_{j-1}) \approx 1/w$ is obtained. All entries referring parameter λ_0^2 were obtained by normalizing by n the resulting estimates. For each case, 400 Monte Carlo runs have been carried out.

Table A.5: Monte Carlo results for $\hat{\chi}^2$ estimator under different number of bins w and $\lambda_0^2 = 0.05$.

n	$w = 10$			$w = 20$			$w = 50$			$w = 100$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
8	$\bar{\lambda}_0^2$	0.071	0.069	0.068	0.066	0.065	0.068	0.067	0.066	0.070	0.068	0.068
	FSSE	.023	.015	.010	.017	.011	.008	.011	.008	.018	.012	.008
	RMSE	.031	.025	.020	.023	.019	.017	.020	.018	.026	.022	.020
$\bar{\sigma}_1$		1.157	1.136	1.130	1.144	1.129	1.128	1.144	1.144	1.180	1.159	1.159
	FSSE	0.199	.137	.087	.177	.127	.081	.133	.087	.201	.141	.092
	RMSE	0.254	.193	.156	.228	.181	.151	.196	.168	.270	.213	.184
16	$\bar{\lambda}_0^2$	0.064	0.065	0.068	0.050	0.056	0.060	0.055	0.058	0.050	0.055	0.059
	FSSE	0.048	.038	.032	.022	.024	.021	.021	.019	.020	.021	.019
	RMSE	0.050	.041	.037	.022	.024	.023	.021	.021	.020	.022	.021
$\bar{\sigma}_1$		1.298	1.499	1.464	1.312	1.525	1.384	1.434	1.409	1.332	1.458	1.419
	FSSE	0.900	1.091	1.032	1.095	1.294	0.990	1.170	1.053	1.201	1.206	1.053
	RMSE	0.948	1.200	1.132	1.139	1.396	1.062	1.248	1.130	1.247	1.290	1.134

NOTE: All simulations are based on a process with $\varepsilon \sim N(0, \sigma_1^2)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, $\lambda_0^2 = 0.05$ and $\sigma_1 = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}_1$ are the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. $\hat{\chi}^2$ estimator was calculated using different total number of bins w of equal support-size $[x_{j-1}, x_j]$. All entries referring parameter λ_0^2 were obtained by normalizing by n the resulting estimates. For each case, 400 Monte Carlo runs have been carried out.

Appendix B

The Asymmetric MSM model

This appendix summarizes some of the building blocks in the design of an asymmetric MSM model. The focus is mainly on the theoretical construction of the model and on the analysis of its properties. An estimation methodology to be employed in future work is proposed after that. To help understand the subtle but complex changes made to the traditional model, we depart from the symmetric construction of Calvet and Fisher^[1].

The traditional model

On the $(\Omega, \mathcal{F}, \mathbb{P})$ probability space, the discrete-time increments of the log-price X_t , for $1 \leq t \leq T$, have the standard formulation

$$x_t = X_t - X_{t-1} \equiv \sigma(M_t) \cdot \varepsilon_t \equiv \sigma_1 \left(\prod_{k=1}^{\bar{k}} M_{k,t} \right)^{1/2} \cdot \varepsilon_t. \quad (\text{B.1})$$

The sequence $\{\varepsilon_t\}_{t=1}^T$ is i.i.d. standard Gaussian $N(0, 1)$, while $\{M_t\}_{t=1}^T$ is gener-

ated by a first-order Markov state-vector with \bar{k} components

$$M_t \equiv (M_{1,t}, M_{2,t}, \dots, M_{\bar{k},t}) \in \mathbb{R}_+^{\bar{k}}.$$

Recall that the components $M_{k,t}$ are drawn from the same marginal distribution M but change with different frequencies γ_k . Given the value of the state-vector at $t - 1$, the dynamics of the factors $M_{k,t}$ are

$$\begin{aligned} M_{k,t} \text{ drawn from distribution } M & \quad \text{with probability } \gamma_k \\ M_{k,t} = M_{k,t-1} & \quad \text{with probability } 1 - \gamma_k, \end{aligned}$$

where the switching events and new draws from M are assumed to be mutually independent across k and t . A well-behaved volatility process also requires $M \geq 0$. Volatility clustering and mean reversion of volatility are obtained when $\mathbb{E}(M) = 1$. The transition probabilities $\gamma \equiv (\gamma_1, \gamma_2, \dots, \gamma_{\bar{k}})$ are hierarchically connected by

$$\gamma_k = 1 - (1 - \gamma_1)^{(b^{k-1})}$$

where $\gamma_1 \in (0, 1)$ and $b \in (1, \infty)$.

For convenience, the first choice for M is to be a Bernoulli distribution as in Calvet and Fisher^[7], taking two values m_0 or $2 - m_0$ ($1 \leq m_0 < 2$) with probabilities p_0 and $1 - p_0$. The state-vector M_t has consequently finitely many states $m^1, \dots, m^d \in \mathbb{R}_+^{\bar{k}}$. As $\gamma_1 < \dots < \gamma_{\bar{k}} < 1 < b$ holds, the pair $(b, \gamma_{\bar{k}})$ alone specifies the set of transition probabilities $\mathbb{P}(M_{t+1} = m^j | M_t = m^i)$. The latter are collected in the transition matrix \mathcal{A} , which will govern the dynamics of the Markov vector process.

The other well-known choice for M is the Lognormal distribution $\text{LN}(\tilde{\mu}, \tilde{\sigma}^2)$, leading to an infinite number of states for each $M_{k,t}$. Lux^[8] determines that in this case mass conservation requires

$$\exp\left(\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2\right) = 1, \quad (\text{B.2})$$

so that the specification of the Lognormal distribution can also be reduced to one parameter, say λ_0 , with $\tilde{\mu} = -\lambda_0$ and $\tilde{\sigma} = \sqrt{2\lambda_0}$. The resulting models are fully described by the parameter spaces $\varphi \equiv (m_0, \sigma_1, b, \gamma_{\bar{k}})$ and $\varphi' \equiv (\lambda_0, \sigma_1, b, \gamma_{\bar{k}}) \in \mathbb{R}_+^4$, respectively.

Introducing asymmetry

Note that the multidimensional nature of the vector M_t and the need to specify the evolution of its components via the Markov property contrast with the approach followed by the class of SV models. In the latter, the marginal distribution of the volatility process may be directly obtained from the joint distribution between the return and the volatility innovations. Because of this, I suggest to introduce asymmetries to the MSM model by altering the conditional distribution of M to make it dependent on the outcome of ε_{t-1} .

The dynamic of the factors $M_{k,t}$ becomes

$$\begin{array}{ll} M_{k,t} \text{ drawn from } \mathbb{P}[M|\varepsilon_{t-1}] & \text{with probability } \gamma_k \\ M_{k,t} = M_{k,t-1} & \text{with probability } 1 - \gamma_k. \end{array}$$

where $\mathbb{P}[M|\varepsilon_{t-1}]$ represents the conditional distribution of M on ε_{t-1} . In terms of Bayes' rule, this leads to $\mathbb{E}(M) \neq \mathbb{E}[M|\varepsilon_{t-1}]$, and $\mathbb{E}(M)$ can now be calculated by integrating $\mathbb{E}[M|\varepsilon_{t-1}]$ with respect to the distribution of ε_{t-1} . By introducing such a dependency, we keep the strictly stationarity property of the innovations and avoid the need to analytically specify the joint distribution between $M_{k,t}$ and ε_{t-1} , which may be rather difficult to find for other distributions of $M_{k,t}$ than the Lognormal. Note, finally, that $\mathbb{P}[M|\varepsilon_{t-1}]$ exhibits temporary shifts in probability mass. Clearly, the outcome of $\mathbb{P}[M|\varepsilon_{t-1}]$ depends on the outcome of ε_{t-1} so that past negative shocks may increase the contemporaneous probability mass of drawing high volatility shocks while positive ones may reduce it, i.e. $\mathbb{E}[M|\varepsilon_{t-1}] \leq 1$.

The simplest way to implement the above premise for the two distributions of M employed so far is to substitute the parameters p_0 and $\tilde{\mu}$ by some functions $p_t \equiv p_0 + h(\varepsilon_{t-1})$ and $\tilde{\mu}_t \equiv \tilde{\mu} + h(\varepsilon_{t-1})$, respectively. For simplicity, we will consider here only the case

$$h(\varepsilon_{t-1}) \equiv \frac{1}{2} - CDF(\rho \cdot \varepsilon_{t-1}),$$

where CDF refers to the cumulative distribution function of ε , and $\rho \geq 0$ is a parameter assigning the level of asymmetry. In the Bernoulli case, this leads to the draw of the high-state and low-state values, m_0 and $2 - m_0$, with probabilities $p_t = 1 - CDF(\rho \cdot \varepsilon_{t-1})$ and $1 - p_t = CDF(\rho \cdot \varepsilon_{t-1})$, respectively. Hence, for a $\rho > 0$, a negative shock ε_{t-1} increases the probability of drawing a high-state m_0 , while a positive ε_{t-1} reduces it. In the limits $\pm\infty$, the chances of drawing m_0 are asymptotically 1 and 0, respectively. Note also that for $\rho = 0$, the asymmetry

disappears and the traditional MSM construction emerges. Finally, a draw from $\mathbb{P}[M|\varepsilon_{t-1}]$ will only occur for those variables for which a new arrival time has occurred. All $m^1, \dots, m^d \in \mathbb{R}_+^{\bar{k}}$ states are as a result achievable, where we have only modified the set of transition probabilities to $\mathbb{P}(M_t = m^j | M_{t-1} = m^i, \varepsilon_{t-1})$. In case of the Lognormal distribution, $\tilde{\mu}_t \equiv \tilde{\mu} + h(\varepsilon_{t-1})$ leads to $\tilde{\mu}_t = -\lambda_0 + \frac{1}{2} - CDF(\rho \cdot \varepsilon_{t-1})$. Note that when $\tilde{\sigma} = \sqrt{2\lambda_0}$ remains unaltered, $-CDF(\rho \cdot \varepsilon_{t-1})$ moves to the right the center axis of the Gaussian distribution when a negative shock ε_{t-1} is drawn, and to the left otherwise. As such, the shift in probability mass due to the outcome of ε_{t-1} is qualitatively similar to that in the Bernoulli case.¹ Alternatively, one could employ the same construction as in traditional SV models. In this case, however, the correlation parameter ρ would appear not only in eq. (B.2) via $\tilde{\mu}$, but also through $\tilde{\sigma}$, which could complicate its identification in practical implementations.

Properties

We turn now to a review of the properties arising from this new construction. We consider first the property of mass conservation on average, that leads to volatility clustering. As mentioned before, $\mathbb{E}(M)$ can be derived from $\mathbb{E}[M|\varepsilon_{t-1}]$ and the distribution of ε_{t-1} . It follows that when the presented choices of $f(\varepsilon_{t-1})$ are adopted, $\mathbb{E}(M) = 1$ still holds provided $\mathbb{E}(\varepsilon) = 0$. In turn, note that $M_{k,t}$ remains a random variable independent of ε_t and $M_{k',t}$, for $k \neq k'$. This implies that the martingale property of the returns would follow trivially. On the other hand, the Markov property of the state vector M_t has not been lost completely, but

¹ The differential nature of the distributions leads, on the other hand, to subtle differences in the maximum and minimum values that $\mathbb{E}[M|\varepsilon_{t-1}]$ can take under each scheme.

rather, been shifted from an order one to a second order. From the key features depicting the traditional MSM model, it remains to consider stationarity and the hyperbolic decline of volatility across time. For the former, we notice that the introduction of a new conditional distribution of $M_{k,t}$ with respect to ε_{t-1} does not alter the causal nature of the model based on information arrivals. Provided also that none of these probability laws depend on time, the process retains its stationary properties. For an analysis of the long memory features of the model, the proof follows immediately from Proposition 1 and Appendix A from Calvet and Fisher^[7] by substituting x_t of their notation to $\sigma(M_t)$. For brevity we refer the reader to the mentioned literature.

We proceed now with an illustration of the process. Figure B.1 presents three samples of 10.000 observations of the MSM model with $\rho = 0$ (top panel), $\rho = 1$ (middle panel), and with $\rho = 3$ (bottom panel). All series were created based on the same random seed so that the differences between them are solely product of the introduced dependency. The asymmetries are observed by the increase in outliers both positive and negative on the middle and bottom panels. All series start, for instance, with a large negative shock, but while on the top panel the next 1,000 points mildly fluctuate between positive and negative values, a much intensified fluctuation is observed on the other two. According to eq. (B.1), outliers are generated by a simultaneous presence of a large return innovation ε_t and a large number of factors in M_t switching to m_0 . This implies that in these first 1,000 observations, a multitude of large ε_t were met on the top panel by a majority of $M_{k,t}$ switching to the low state $2 - m_0$ at each t , while the outcome of ε_{t-1} modified the probabilities of M_t on the other two panels allowing to switches to the high state and leading consequently to mentioned outliers. On the other

hand, between observations 1,500 and 2,000 the top panel presents a large negative shock absent on the other two panels. This is the flip-side of introducing asymmetries in that an eventual $\varepsilon_{t-1} > 0$ reduces the switching probabilities to states m_0 , which would else carry out their influence in the symmetric MSM and yield high volatility states. Note also that whilst the initial outlier on the lower two panels is followed by an intensified fluctuation for the next 1,000 points, other outliers are followed by a milder fluctuation resembling closely the return variation on the top panel. The asymmetric MSM may seem to capture the more subtle features of asymmetric responses that are otherwise difficult to replicate by other models (cf. Wu and Xiao^[191] and Yu^[192]). Note, finally, that the process on the bottom panel provides the largest negative outliers, despite the fact that some of the outliers have disappeared when switching from $\rho = 1$ to $\rho = 3$.

The issue of quantifying the level of connection between return and volatility innovations is, on the other hand, arduous. The fact that the new parameter describing the volatility dependency modifies only the probability laws of the \bar{k} elements $M_{k,t}$ presents some challenges. In turn, we will consider the following proposition.

Proposition (Asymmetric responses of volatility) The *asymmetric MSM model* with a $\rho > 0$ provides of a *negative covariance* between x_t^2 and x_{t-1} .

The proof follows by inspection of the second moments and the evaluation of signs of the resulting terms. For $\rho > 0$ and a process x_t departing from the ergodic

distribution, i.e. $\mathbb{E}[M_0] \equiv 1$ and $\varepsilon_0 \equiv 0$, we obtain

$$\begin{aligned}
\text{COV}[x_t^2, x_{t-1}] &= \mathbb{E}[(x_t^2 - \sigma_1^2) \cdot (x_{t-1} - 0)] \\
&= \mathbb{E}[\varepsilon_t^2] \cdot \mathbb{E}[\sigma^2(M_t) \cdot \sigma(M_{t-1}) \cdot \varepsilon_{t-1}] \\
&= \mathbb{E}[\mathbb{E}_{t-1}[\sigma^2(M_t) \cdot \sigma(M_{t-1})] \cdot \varepsilon_{t-1}] \\
&= \sigma_1^3 \cdot \mathbb{E}\left[\prod_{k=1}^{\bar{k}} \mathbb{E}_{t-1}[M_{k,t} \cdot M_{k,t-1}^{1/2}] \cdot \varepsilon_{t-1}\right] \\
&= \sigma_1^3 \cdot \mathbb{E}\left[\prod_{k=1}^{\bar{k}} \left[(1 - \gamma_k) M_{k,t-1}^{3/2} + \gamma_k M_{k,t-1}^{1/2} \mathbb{E}[M_{k,t}|\varepsilon_{t-1}]\right] \cdot \varepsilon_{t-1}\right] \\
&\quad + \quad \quad \quad + \quad \quad \quad + \quad \quad \quad + \quad \quad \quad > 1 \quad \quad - \\
&\quad + \quad \quad \quad + \quad \quad \quad + \quad \quad \quad + \quad \quad \quad < 1 \quad \quad + \\
&< 0,
\end{aligned}$$

where first line follows from the definition of COV , $\mathbb{E}(x_t) = 0$, and $\mathbb{E}(x_t^2) = \sigma_1^2$; the second line arises from the independence of ε_t and M_t , and the definition of x_t in eq. (B.1); the third one is the result of the law of total expectation, where $\mathbb{E}_{t-1}[\cdot]$ is taken with respect to all information available at $t - 1$ and $\sigma(M_{t-1})$ is left inside of the expectation for convenience; the fourth line follows from the definitions of $\sigma(M_t)$ in eq. (B.1) and the mutually independence of factors M across k and t ; and the fifth one follows from the switching rule of M_t . The symbols on the last two lines refer to the sign evaluation of the equation. For this, notice that all signs but those for σ_1^3 and γ_k are associated to a probability of occurrence. By our choice of the distribution of M and the independence between ε_{t-1} and each $M_{k,t-1}$, however, the only term that matters in that expectation is the joint presence of $\mathbb{E}[M_{k,t}|\varepsilon_{t-1}]$ and ε_{t-1} .² Provided that all the variables

² Clearly, terms including ε_{t-1} but not $\mathbb{E}[M_{k,t}|\varepsilon_{t-1}]$ have zero contributions.

except ε_{t-1} must be positive, $\text{COV} < 0$ results from the fact that the distribution of ε_{t-1} assigns the same probability mass to its positive and negative values, and since the latter are magnified by $\mathbb{E}[M_{k,t}|\varepsilon_{t-1}] > 1$ while the former are shrunk, the overall result is negative.

Note that for a given \bar{k} the introduced dependency is controlled by all parameters. Clearly, the larger ρ , the larger the asymmetry will be. The rest of the parameters, on the other hand, allows to scale the level of dependence. The inclusion of an additional volatility factor, that is, the change from \bar{k} to $\bar{k} + 1$, increases likewise the level of potential dependence on ε_{t-1} . This can be seen in all the terms that appear multiplying ε_{t-1} . These terms can be classified in those exerting the direct effect $\prod_{k=1}^{\bar{k}} \gamma_k M_{k,t-1}^{1/2} \mathbb{E}[M_{k,t}|\varepsilon_{t-1}]$ and in those including the cross products $\prod_{k'=1}^j (1 - \gamma_{k'}) M_{k',t-1}^{3/2} \cdot \prod_{k=1}^{\bar{k}-j} \gamma_k M_{k,t-1}^{1/2} \mathbb{E}[M_{k,t}|\varepsilon_{t-1}]$, for $1 < j < \bar{k}$ and $k \neq k'$. In the end, the incremental level of dependency due to $M_{\bar{k}+1,t}$ is weighted by $\gamma_{\bar{k}+1}$.

An Estimation Methodology

The procedure suggested here for application in future research includes some changes to the approach introduced by Calvet and Fisher^[7] to contemplate the asymmetric responses of volatility. When the distribution of M is Bernoulli, the state-vector M_t has finitely many states $m^1, \dots, m^d \in \mathbb{R}_+^{\bar{k}}$. As $\gamma_1 < \dots < \gamma_{\bar{k}} < 1 < b$ holds, the pair $(b, \gamma_{\bar{k}})$ and $\{\varepsilon_t\}_{t=1}^T$ specify the set of transition probabilities $a_{i,j} \equiv \mathbb{P}(M_t = m^j | M_{t-1} = m^i, \varepsilon_{t-1})$. The latter are collected in the transition matrix $\mathcal{A}_t = (a_{i,j})_{1 \leq i,j \leq d}$, which will govern the dynamics of the vector volatility process.³ Specification of the elements $a_{i,j}$ in a single equation as in Calvet and

³ Note that we have used the subscript t only to signal that \mathcal{A} now also depends on ε_{t-1} . The probability law, however, does not change in time.

Fisher^[7] is, on the other hand, no longer possible. The reason for this is that the transition probabilities depending on both the departure and arrival states must be now specified individually (cf. the impact of ε_{t-1} on the resulting transition probabilities in Table B.1).

In practice, all states are accessible with positive probability, which will allow us to compute $\Pi_t^i \equiv \mathbb{P}(M_t = m^i | x_t)$ over the unobserved states m^1, \dots, m^d . Note that the $CDF(\rho \cdot \varepsilon_{t-1})$ in \mathcal{A}_t must be computed by conditioning on M_{t-1} provided that ε_{t-1} is unobservable. It follows that $CDF(\rho \cdot \varepsilon_{t-1}) = CDF(\rho \cdot (x_{t-1}/\sigma(m^i)))$, for $i = 1, \dots, d$. The rest follows as usual. The return x_t , conditional on M_t , is distributed with density $f_x(x_t | M_t = m^i) = [\sigma(m^i)]^{-1} f_\varepsilon(x_t/\sigma(m^i))$, where f_ε is the standard Gaussian distribution. The vector Π_t stacks the conditional probabilities Π_t^i for $i = 1, \dots, d$ and is calculated recursively by Bayes' rule

$$\Pi_t = \frac{\omega(x_t) * (\Pi_{t-1} \mathcal{A}_t)}{\omega(x_t)' (\Pi_{t-1} \mathcal{A}_t)},$$

where $\omega(x_t)$ is the vector stacking $f_x(x_t | M_t = m^i)$ for $i = 1, \dots, d$, $*$ is the element by element multiplication operator, and $-$ represents the element by element division.

The conditioning in $\omega(x_t)$ and \mathcal{A}_t makes the choice of the initial Π_0 and ε_0 important. We opt for $\varepsilon_0 = 0$, and Π_0 to be selected as in the symmetric case (cf. Calvet and Fisher^[7]). This leads to $\mathcal{A}_1 = \mathcal{A}$ as in the symmetric case, and consequently, to Π_2 being the first vector presenting asymmetries.

For a sample of size T , the log-likelihood follows as

$$\ln L(x_1, \dots, x_T; \varphi) = \sum_{t=1}^T \ln [\omega(x_t)' (\Pi_{t-1} \mathcal{A}_t)].$$

We obtain as in the symmetric case, a closed-form likelihood that can be optimized using traditional gradient-based direction methods. For $T \rightarrow \infty$, consistency and asymptotic normality of the estimators follow as usual.

Evaluation of this methodology, together with the design of an estimation methodology for the Lognormally distributed vector M_t , is part of future research. Note also that we have restricted ourselves to the driftless case. As such, following work in asymmetric MSM models may contemplate a series of empirical evaluations of the model with and without a drift component dependent on volatility. Figure B.2 shows, for instance, a sample series of the asymmetric MSM (bottom panel) enhanced by the mentioned construction, with drift $\mu(M_t) = \alpha \cdot \sigma^2(M_t)$ and parameter values as specified in the figure. It is easy to appreciate the remarkable resemblance to the S&P500 series (top panel), which has been one of the most widely used assets for evaluation of the always increasingly more complex extensions of the traditional SV models (cf. Bates^[35], Chernov and Ghysels^[36], Benzoni^[37], Chernov *et al.*^[38], Pan^[39], Eraker *et al.*^[40], Eraker^[41]).

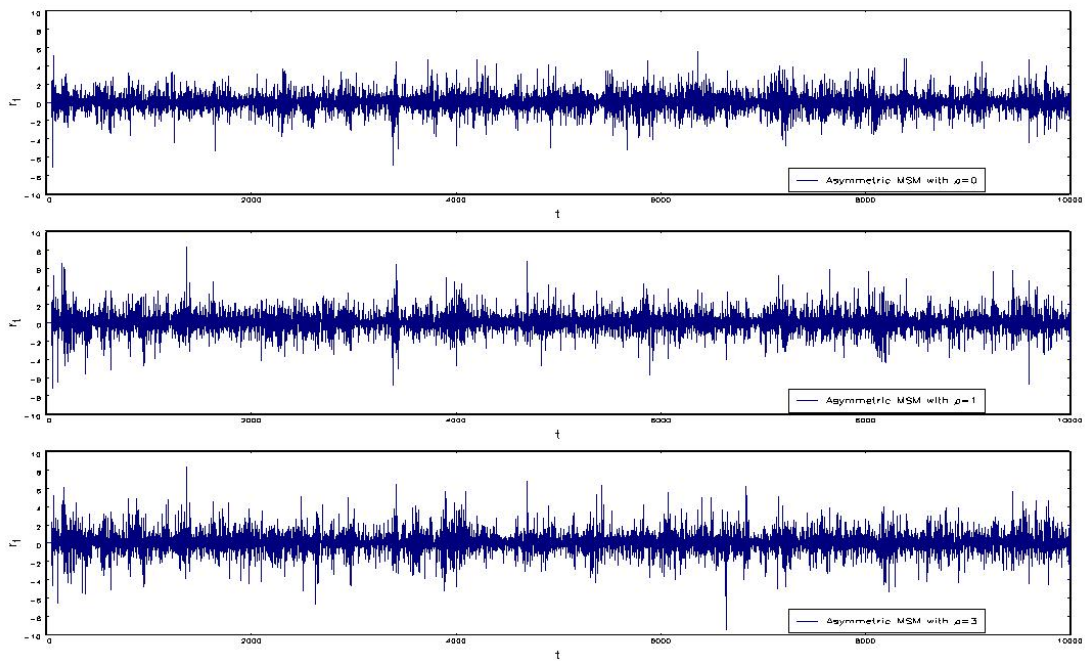


Figure B.1: One sample of $T = 10,000$ observations of the asymmetric MSM model with $\rho = 0$ (top panel), $\rho = 1$ (middle panel), and with $\rho = 3$ (bottom panel). All series were created based on the same random seed, number of cascade levels $k = 6$, and parameter space $(m_0, \sigma_1, b, \gamma_{\bar{k}}) = (1.4, 1, 3, 0.95)$.

Table B.1: Transition probabilities for $M_{k,t}$.

	m_0	$2 - m_0$
m_0	$1 - \gamma_k CDF(\rho \cdot \varepsilon_{t-1})$	$\gamma_k CDF(\rho \cdot \varepsilon_{t-1})$
$2 - m_0$	$\gamma_k [1 - CDF(\rho \cdot \varepsilon_{t-1})]$	$1 - \gamma_k [1 - CDF(\rho \cdot \varepsilon_{t-1})]$

NOTE: Transition probabilities for $M_{k,t}$ according to ε_{t-1} . Rows denote initial states at $t - 1$, whereas columns indicate destination states at t . Note that while summations across columns add up to one, they do not across rows. The latter quantify the total probability of arriving at a certain state and are the result of the asymmetries introduced.

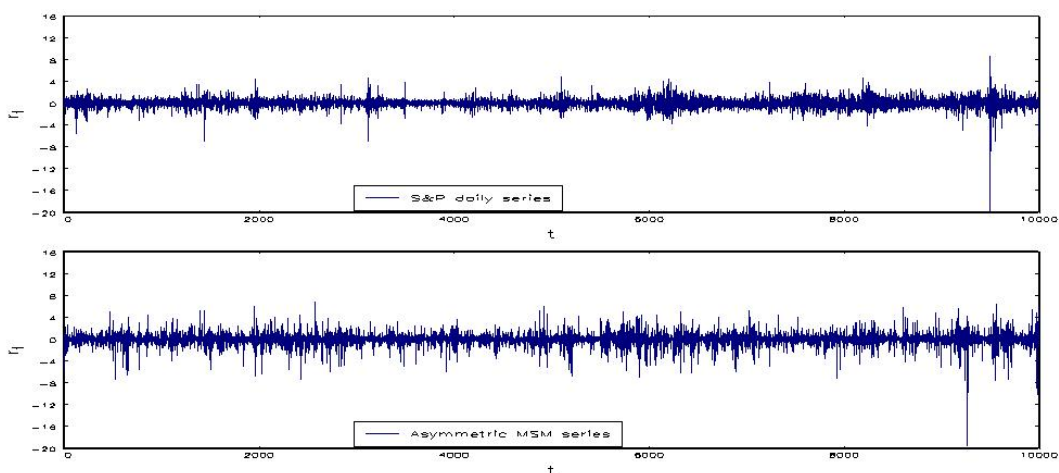


Figure B.2: A sample of 10,000 observations of the S&P500 index (top panel) and the asymmetric MSM process (bottom panel). The asymmetric MSM series was created using $k = 10$ cascade components $M_{k,t}$, which are Bernoulli distributed, and adding to the traditional formulation of the process a drift like $\mu(M_t) = \alpha \cdot \sigma^2(M_t)$. The resulting parameter space is $(\alpha, m_0, \sigma_1, b, \gamma_{\bar{k}}, \rho) = (-0.1, 1.4, 1, 2, 0.95, 1)$.

Bibliography

- [1] L. E. Calvet and A. Fisher. Forecasting Multifractal Volatility. *Journal of Econometrics*, 105:27–58, 2001.
- [2] A. E. Leövey and T. Lux. Parameter Estimation and Forecasting for Multiplicative Lognormal Cascades. Kiel Working Papers 1746, Kiel Institute for the World Economy, 2011.
- [3] A. E. Leövey and T. Lux. Parameter Estimation and Forecasting for Multiplicative Lognormal Cascades. *Physical Review E*, 85:046114, 2012.
- [4] B. Castaing, Y. Gagne, and E. J. Hopfinger. Velocity Probability Density Functions of High Reynolds Number Turbulence. *Physica D*, 46(177), 1990.
- [5] C. Beck. Superstatistics in Hydrodynamic Turbulence. *Physica D*, 193:195–207, 2004.
- [6] K. Kiyono, Z. R. Struzik, and Y. Yamamoto. Estimator of a Non-Gaussian Parameter in Multiplicative Lognormal Models. *Physical Review E*, 76:41113, 2007.
- [7] L. E. Calvet and A. Fisher. How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes. *Journal of Financial Econometrics*, 2:49–83, 2004.
- [8] T. Lux. The Markov-Switching Multifractal Model of Asset Returns. *Journal of Business & Economic Statistics*, 26(2):194–210, 2008.

- [9] T. Lux, L. Morales-Arias, and C. Sattarhoff. A Markov-switching Approach to Forecasting Realized Volatility. Kiel Working Papers 1737, Kiel Institute for the World Economy, 2011.
- [10] R. Cont. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, 1(2):223–236, 2001.
- [11] B. B. Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4):394–419, 1963.
- [12] E. Fama. Mandelbrot and the Stable Paretian Hypothesis. *Journal of Business*, 36:420–429, 1963.
- [13] E. Fama. The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105, 1965.
- [14] A. R. Pagan. The Econometrics of Financial Markets. *Journal of Empirical Finance*, 3(1):15–102, 1996.
- [15] J. Campbell, A. Lo, and C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, 1997.
- [16] T. Lux. Stochastic Behavioral Asset Pricing Models and the Stylized Facts. *Handbook of Financial Markets: Dynamics and Evolution*, 1:161, 2009.
- [17] R. F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007, 1982.
- [18] T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.
- [19] É. Jondeau, S. H. Poon, and M. Rockinger. *Financial Modeling under Non-Gaussian Distributions*. Springer finance. Springer, 2007.
- [20] F. C. Drost and T. E. Nijman. Temporal Aggregation of GARCH Processes. *Econometrica*, 61:909–927, 1993.

- [21] S. J. Taylor. Financial Returns Modelled by the Product of Two Stochastic Processes, a Study of Daily Sugar Prices, 1961-1979. In O. E. Anderson, editor, *Time Series Analysis: Theory and Practice*, pages 203–226. North-Holland, Amsterdam, 1982.
- [22] S. J. Taylor. *Modelling Financial Time Series*. Wiley, New York, 1986.
- [23] N. Shephard and T. G. Andersen. Stochastic Volatility: Origins and Overview. In T G Andersen, R.A. Davis, J.-P. Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*. Springer, 2009. Forthcoming.
- [24] P. K. Clark. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*, 41:135–156, 1973.
- [25] C. S. Jones. Bayesian Estimation of Continuous-Time Finance Models. Working paper, Simon School of Business. University of Rochester, 1998.
- [26] B. Eraker. MCMC Analysis of Diffusion Models With Application to Finance. *Journal of Business & Economic Statistics*, 19(2):177–191, 2001.
- [27] G. O. Roberts and O. Stramer. On Inference for Partially Observed Nonlinear Diffusion Models using the Metropolis-Hastings algorithm. *Biometrika*, 88(3):603–621, 2001.
- [28] O. Elerian, S. Chib, and N. Shephard. Likelihood Inference for Discretely Observed Non-Linear Diffusions. *Econometrica*, 19(2):959–993, 2001.
- [29] M. W. Brandt and P. Santa-Clara. Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets. *Journal of Financial Economics*, 63:161–210, 2002.
- [30] G. B. Durham and A. R. Gallant. Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. *Journal of Business & Economic Statistics*, 20(3):297–316, 2002.
- [31] G. B. Durham. Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate. *Journal of Financial Economics*, 70:463–487, 2003.

- [32] T. G. Andersen, T. Bollerslev, P. F. Christoffersen, and F. X. Diebold. Volatility and Correlation Forecasting. *Handbook of Economic Forecasting*, I:778–878, 2006.
- [33] N. Shephard. Stochastic Volatility. In Steven Durlauf and Lawrence Blume, editors, *New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2 edition, 2008.
- [34] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and Forecasting Realized Volatility. *Econometrica*, 71(2):579–625, March 2003.
- [35] D. S. Bates. Post-'87 Crash Fears in the S&P 500 Futures Option Market. *Journal of Econometrics*, 94:181–238, 2000.
- [36] M. Chernov and E. Ghysels. A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Option Evaluation. *Journal of Financial Economics*, 56:407–458, 2000.
- [37] L. Benzoni. Pricing Options under Stochastic Volatility: An Empirical Investigation. Technical report, Carlson School of Management, 2002.
- [38] M. Chernov, A. R. Gallant, E. Ghysels, and G. Tauchen. A new Class of Stochastic Volatility Models with Jumps: Theory and Estimation. Discussion Paper, 2000.
- [39] J. Pan. The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics*, 63:3–50, 2002.
- [40] B. Eraker, M. Johannes, and N. Polson. The Impact of Jumps in Volatility and Returns. *The Journal of Finance*, 58(3):1269–1300, 2003.
- [41] B. Eraker. Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *The Journal of Finance*, 59(3):1367–1403, 2004.
- [42] Y. Aït-Sahalia. Disentangling Diffusion from Jumps. *Journal of Financial Economics*, 74(3):487–528, 2004.

- [43] T. G. Andersen and T. Bollerslev. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, 39(4):885–905, November 1998.
- [44] O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based Models and some of their uses in Financial Economics. *Journal Of The Royal Statistical Society Series B*, 63(2):167–241, 2001.
- [45] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 96:42–55, 2001. Correction published in 2003, volume 98, page 501.
- [46] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC Fin. Math. Series, London, 2004.
- [47] F. Black and M. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.
- [48] J. C. Hull and A. White. The Pricing of Options on Assets with Stochastic Volatility. *Journal of Finance*, 42:281–300, 1987.
- [49] E. M. Stein and J. C. Stein. Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies*, 4:727–752, 1991.
- [50] S. L. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 1993.
- [51] R. Garcia, E. Ghysels, and E. Renault. The Econometrics of Option Pricing. In Y. Aït-Sahalia and L. P. Hansen, editors, *Handbook of Financial Econometrics*, North Holland, Amsterdam, 2007.
- [52] M. Chernov, A. R. Gallant, E. Ghysels, and G. Tauchen. Alternative Models for Stock Price Dynamics. *Journal of Econometrics*, 116(1-2):225–257, 2003.

- [53] P. Billingsley. *Probability and Measure*. John Willey & Sons, New York, second edition, 1986.
- [54] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, second edition, 2000.
- [55] B. B. Mandelbrot. Possible Refinements of the Lognormal Hypothesis Concerning the Distribution of Energy Dissipation in Intermittent Turbulence. In M. Rosenblatt and C. Van Atta, editors, *Statistical Models and Turbulence*. Springer Verlag, 1972.
- [56] B. B. Mandelbrot. Intermittent Turbulence in Self-Similar Cascades; Divergence of High Moments and Dimension of the Carrier. *Journal of Fluid Mechanics*, 62:331–358, 1974.
- [57] B. B. Mandelbrot. *The fractal geometry of nature*. W.H. Freeman, 1982.
- [58] B. B. Mandelbrot. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer, New York, 1997.
- [59] L. E. Calvet, A. Fisher, and B. Mandelbrot. A Multifractal Model of Asset Returns. Discussion Papers 1164-1166, Cowles Foundation, 1997.
- [60] L. E. Calvet and A. Fisher. *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press Advanced Finance Series. Academic Press, 2008.
- [61] E. Bacry, J. Delour, and J. F. Muzy. Multifractal Random Walk. *Physical Review E*, 64:026103, 2001.
- [62] B. Pochart and J. P. Bouchaud. The Skewed Multifractal Random Walk with Applications to Option Smiles. *Quantitative Finance*, 2(4):303–314, 2002.
- [63] A. Saichev and D. Sornette. Generic Multifractality in Exponentials of Long Memory Processes. *Physical Review E*, 74:011111, 2006.

- [64] A. Saichev and V. Filimonov. On the Spectrum of Multifractal Diffusion Process. *Journal of Experimental and Theoretical Physics*, 105:1085–1093, 2007.
- [65] E. Bacry, L. Duvernet, and J.F. Muzy. Continuous-time Skewed Multifractal Processes as a Model for Financial Returns. *Journal of Theoretical and Applied Finance*, 2010. in press.
- [66] V. Filimonov and D. Sornette. Self-excited Multifractal Dynamics. *Europhysics Letters*, 94(4):46003, 2011.
- [67] S. E. Shreve. *Stochastic Calculus for Finance I - The Binomial Asset Pricing Model*. Springer, New York, 2004.
- [68] S. E. Shreve. *Stochastic Calculus for Finance II - Continuous-Time Models*. Springer, New York, 2004.
- [69] L. Sorriso-Valvo, V. Carbone, P. Veltri, G. Consolini, and R. Bruno. Intermittency in the Solar Wind Turbulence through Probability Distribution Functions of Fluctuations. *Geophys. Res. Letters*, 26:1801, 1999.
- [70] K. Kiyono, Z. R. Struzik, N. Aoyagi, S. Sakata, J. Hayano, and Y. Yamamoto. Critical Scale Invariance in a Healthy Human Heart Rate. *Physical Review Letters*, 93:178103, 2004.
- [71] K. Kiyono, Z. R. Struzik, N. Aoyagi, F. Togo, and Y. Yamamoto. Phase Transition in a Healthy Human Heart Rate. *Physical Review Letters*, 95:58101, 2005.
- [72] S. Carius and G. Ingelman. The Lognormal Distribution for Cascade Multiplicities in Hadron Collisions. *Physical Letters B*, 252(4), 1990.
- [73] E. Bullmore, A. Barnes, D. S. Bassett, A. Fornito, M. Kitzbichler, D. Menier, and J. Suckling. Generic Aspects of Complexity in Brain Imaging Data and other Biological Systems. *NeuroImage*, 47(3):1125 – 1134, 2009.

- [74] P. Sinha-Ray, L. Borda de Agua, and H. J. Jensen. Threshold Dynamics, Multifractality and Universal Fluctuations in the SOC Forest-Fire: Facets of an Auto-Ignition Model. *Physica D: Nonlinear Phenomena*, 157(3):186 – 196, 2001.
- [75] M. S Baptista, L. O. B. de Almeida, J. F. W. Slaets, R. Kberle, and C. Grebogi. A Complex Biological System: the Flys Visual Module. *Phil. Trans. R. Soc. A*, 366(1864):345–357, 2008.
- [76] B. B. Mandelbrot, L. E. Calvet, and A. Fisher. A Multifractal Model of Asset Returns. Discussion Paper 1164, Cowles Foundation, 1997.
- [77] L. E. Calvet and A. Fisher. Multifractality in Asset Returns: Theory and Evidence. *Review of Economics and Statistics*, 84:381–406, 2002.
- [78] A. A. Demos, C. Vassilicos, and F. Tata. No Evidence of Chaos but Some Evidence of Multifractals in the Foreign Exchange and Stock Markets. In A.J. Crilly, R.A. Earnshaw, and H. Jones, editors, *Application of Fractals and Chaos. The Shape of Things*. Springer-Verlag, 1993.
- [79] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge. Turbulent Cascades in Foreign Exchange Markets. *Nature*, 381:767–770, 1996.
- [80] A. Arneodo, J. F. Muzy, and D. Sornette. ‘Direct’ Causal Cascade in the Stock Market. *The European Physical Journal B - Condensed Matter and Complex Systems*, 2:277, 1998.
- [81] J. F. Muzy, J. Delour, and E. Bacry. Modelling Fluctuations of Financial Time Series: from Cascade Process to Stochastic Volatility Model. *The European Physical Journal B - Condensed Matter and Complex Systems*, 17:537, 2000.
- [82] N. Kozuki and N. Fuchikami. Dynamical Model of Financial Markets: Fluctuating ‘Temperature’ Causes Intermittent Behavior of Price Changes. *Physica A*, 329(222), 2003.

- [83] M. Ausloos and K. Ivanova. Dynamical Model and Nonextensive Statistical Mechanics of a Market Index on Large Time Windows. *Physical Review E*, 68:046122, 2003.
- [84] K. Kiyono, Z. R. Struzik, and Y. Yamamoto. Criticality and Phase Transition in Stock-price Fluctuations. *Physical Review Letters*, 96:68701, 2006.
- [85] A. P. Nawroth, R. Friedrich, and J. Peinke. Multi-Scale Description and Prediction of Financial Time Series. *New Journal of Physics*, 12(8):083021, 2010.
- [86] A. Fisher, L. E. Calvet, and B. Mandelbrot. Multifractality of Deutschemark / US Dollar Exchange Rates. Discussion Paper 1166, Cowles Foundation, 1997.
- [87] L. E. Calvet, A. Fisher, and B. Mandelbrot. Large Deviations and the Distribution of Price Changes. Discussion Paper 1165, Cowles Foundation, 1997.
- [88] J. Fillol. Multifractality: Theory and Evidence an Application to the French Stock Market. *Economics Bulletin*, 3(31):1–12, 2003.
- [89] R. Yalamova. Wavelet Test of Multifractality of Asia-Pacific Index Price Series. *Asian Academy of Management Journal of Accounting and Finance*, 2(1):64–83, 2006.
- [90] Z. Eisler and J. Kertész. Multifractal Model of Asset Returns with Leverage Effect. *Physica A: Statistical Mechanics and its Applications*, 343(0):603–622, 2004.
- [91] A. Pantanella and A. Pianese. Minimum risk portfolios using MMAR. In *Proceedings of the 10th WSEAS Int. Conference on Mathematics and Computers in Business and Economics*, 2009.
- [92] H-W Chen, M-H Huang, and S-H Huang. Multifractality Analysis for Stock Market Characteristics. *Middle Eastern Finance and Economics*, pages 64–74, 2011.

- [93] Jian Zhong and Xin Zhao. A Comparison of Multifractal Models for Assets Returns in Economic Perspective. In *Business Intelligence and Financial Engineering (BIFE)*, pages 320–324, 2011.
- [94] J. D. Hamilton. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:357–384, 1989.
- [95] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton New Jersey, 1994.
- [96] T. Lux and T. Kaizoji. Forecasting Volatility and Volume in the Tokyo Stock Market: Long Memory, Fractality and Regime Switching. *Journal of Economic Dynamic and Control*, 31(6):1808–1843, 2007.
- [97] T. Lux and L. Morales-Arias. Forecasting Volatility under Fractality, Regime-Switching, Long Memory and Student-t Innovations. *Computational Statistics and Data Analysis*, 54(11):2676–2692, 2010.
- [98] T. Lux and L. Morales-Arias. Relative Forecasting Performance of Volatility Models: Monte Carlo Evidence. Kiel Working Papers 1582, Kiel Institute for the World Economy, 2010. Forthcoming in: *Quantitative Finance*.
- [99] T. Lux, T. Di-Matteo, and R. Liu. Multifractality and Long-Range Dependence of Asset Returns: The Scaling Behavior of the Markov-Switching Multifractal Model with Lognormal Volatility Components. *Advances in Complex Systems*, 11(5):669–684, 2008.
- [100] L. E. Calvet, M. Fearnley, A. Fisher, and M. Leippold. What’s Beneath the Surface: Option Pricing with Multifrequency Latent States. HEC Paris Research Papers 969, HEC Paris, 2013.
- [101] P. Jorion. Predicting Volatility in the Foreign Exchange Market. *Journal of Finance*, 50(2):507–528, 1995.
- [102] G. Bakshi, C. Cao, and Z. Chen. Empirical Performance of Alternative Option Pricing Models. *The Journal of Finance*, 52(5):2003–2049, 1997.

- [103] B. J. Christensen and N. R. Prabhala. The relation between implied and realized volatility. *Journal of Financial Economics*, 50:125–150, 1998.
- [104] B. J. Christensen and C. S. Hansen. New Evidence on the Implied-Realized Volatility Relation. *The European Journal of Finance*, 8(2):187–205, 2002.
- [105] J. Fleming. The Quality of Market Volatility Forecasts Implied by S&P 100 Index Option Prices. *Journal of Empirical Finance*, 5(4):317–345, 1998.
- [106] B. J. Blair, S.-H Poon, and S. J. Taylor. Forecasting S&P 100 Volatility: the Incremental Information Content of Implied Volatilities and High Frequency Index Returns. *Journal of Econometrics*, 105:5–26, 2001.
- [107] L. E. Calvet, A. Fisher, and S. B. Thompson. Volatility Comovement: a Multifrequency Approach. *Journal of Econometrics*, 131(1-2):179–215, 2006.
- [108] R. Liu and T. Lux. Flexible and Robust Modelling of Volatility Comovements: A Comparison of Two Multifractal Model. Kiel Working Papers 1594, Kiel Institute for the World Economy, 2010.
- [109] J. Idier. Correlation: A Markov-Switching Multifractal Model with Time-Varying Correlations. Technical report, Bank of France, 2010.
- [110] J. Baruník, N. Shenaiz, and F. Žikeš. Modeling and Forecasting Persistent Financial Durations. Technical report, Academy of Sciences of the Czech Republic and Imperial College London Business School, .
- [111] F. Chen, F. X. Diebold, and F. Schorfheide. A Markov-Switching Multifractal Inter-Trade Duration Model, with Application to U.S. Equities. Technical report, Huazhong University and University of Pennsylvania, 2012.
- [112] L. E. Calvet and A. Fisher. Multifrequency News and Stock Returns. *Journal of Financial Economics*, 86:178–212, 2007.
- [113] L. E. Calvet and A. Fisher. Multifrequency Jump-Diffusions:an Equilibrium Approach. *Journal of Mathematical Economics*, 44:207–226, 2008.

- [114] K. Kiyono. Log-amplitude Statistics of Intermittent and Non-Gaussian Time Series. *Physical Review E*, 79():031129, 2009.
- [115] C. Beck and E.G.D. Cohen. Superstatistics. *Physica A*, 322:267–275, 2003.
- [116] A. Arneodo, N. Decoster, and S. G. Roux. Intermittency, Lognormal Statistics, and Multifractal Cascade Process in High-Resolution Satellite Images of Cloud Structure. *Physical Review Letters*, 83(6):1255, 1999.
- [117] B. B. Mandelbrot. *Multifractals and 1/f Noise: Wild Self-Affinity in Physics*. Springer, New York, 1999.
- [118] W. A. Gardner, A. Napolitano, and L. Paura. Cyclostationarity: Half a Century of Research. *Signal Processing*, 86(4):639–697, 2006.
- [119] D. Harris and L. Mátyás. *Generalized Method of Moment Estimation*. University Press, Cambridge, 1999.
- [120] L. P. Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, 1982.
- [121] L. P. Hansen, J. Heaton, and A. Yaron. Finite-sample Properties of some Alternative GMM Estimators. *Journal of Business & Economic Statistics*, 14:262–280, 1996.
- [122] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, 1991.
- [123] P. J. Brockwell and R. Dahlhaus. Generalized LevinsonDurbin and Burg algorithms. *Journal of Econometrics*, 118:129–149, 2004.
- [124] T. E. Clark and K. D. West. Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis. *Journal of Econometrics*, 135:155–186, 2006.
- [125] T. E. Clark and K. D. West. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics*, 138:291–311, 2007.

- [126] F. X. Diebold and R. S. Mariano. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263, 1995.
- [127] D. Harvey, S. Leybourne, and P. Newbold. Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, 13(2):281–291, 1997.
- [128] C. Gourieroux, A. Monfort, and E. Renault. Indirect Inference. *Journal of Applied Econometrics*, 8:85–118, Dec 1993.
- [129] A. R. Gallant and G. Tauchen. Which Moments to Match? *Econometric Theory*, 12(04):657–681, 1996.
- [130] D. Duffie and K. Singleton. Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, 61:929–952, 1993.
- [131] M. Carrasco and J.-P. Florens. Simulation-Based Method of Moments and Efficiency. *Journal of Business & Economic Statistics*, 20(4):482–492, 2002.
- [132] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.
- [133] L. E. Calvet and A. Fisher. Regime Switching and the Estimation of Multifractal Processes. Technical report, Harvard University, 2002.
- [134] H. Zhou and T. Bollerslev. Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility. *Journal of Econometrics*, 109:33–65, 2002.
- [135] V. Corradi and W. Distaso. Estimating and Testing Stochastic Volatility Models using Realized Measures. Working paper, Warwick Business School, Financial Econometrics Research Centre, Coventry, 2004.
- [136] C. Gourieroux and A. Monfort. *Simulation-Based Econometric Methods*. Oxford University Press, NY, 2002.

- [137] W. Newey and K. West. A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, (55): 703–708, 1987.
- [138] B. W. Brown and W. K. Newey. Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference. *Journal of Business & Economic Statistics*, 20(4):507–517, 2002.
- [139] A. Inoue and M. Shintani. Bootstrapping GMM Estimators for Time Series. *Journal of Econometrics*, (133):531–555, 2006.
- [140] P. Winker, M. Gilli, and V. Jeleskovic. An Objective Function for Simulation Based Inference on Exchange Rate Data. *Journal of Economic Interaction and Coordination*, 2:125–145, 2007.
- [141] R. Franke and F. Westerhoff. Estimation of a Structural Stochastic Volatility Model of Asset Pricing. *Computational Economics*, 38(1):53–83, 2011.
- [142] W. K. Newey and D. McFadden. Large Sample Estimation and Hypothesis Testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier, 1994.
- [143] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes: The Art of Scientific Computing*. Cambridge University Press, 2007.
- [144] N.C. Mark. *International Macroeconomics and Finance: Theory and Econometric Methods*. Blackwell Publishers, 2001.
- [145] X. Mao, C. Yuan, and G. Yin. Numerical Method for Stationary Distribution of Stochastic Differential Equations with Markovian Switching. *Journal of Computational and Applied Mathematics*, 174(1):1–27, 2005.
- [146] G. Yin and C. Zhu. *Hybrid Switching Diffusions. Properties and Applications*, volume 63 of *Stochastic Modelling and Applied Probability*. Springer New York, 2010.

BIBLIOGRAPHY

- [147] D. T. Breeden and R. H. Litzenberger. Prices of State-Contingent Claims Implicit in Option Prices. *Journal of Business*, 51(4):621–651, 1978.
- [148] K. I. Amin and V. K. Ng. Option Valuation with Systematic Stochastic Volatility. *Journal of Finance*, 48(3):881–910, 1993.
- [149] R. Garcia and E. Renault. Latent Variable Models for Stochastic Discount Factors. In E. Jouini, J. Cvitanic, and M. Musiela, editors, *Option Pricing, Interest Rates and Risk Management*, Handbooks in Mathematical Finance. Cambridge University Press, 2001.
- [150] R. Garcia, E. Renault, and G. Tsafack. Proper Conditioning for Coherent VaR in Portfolio Management. *Management Science*, 53(3):483–494, 2007.
- [151] F. Chabi-Yo, R. Garcia, and E. Renault. State Dependence Can Explain the Risk Aversion Puzzle. *The Review of Financial Studies*, 21(2):973–1011, 2008.
- [152] E. Ghysels, A. Harvey, and E. Renault. Stochastic Volatility. In G. S. Madala, editor, *Statistical Methods in Finance*, volume 14 of *Handbook of Statistics*, North Holland, Amsterdam, 1996.
- [153] C. de Jong and T. Lehnert. Implied GARCH Volatility Forecasting. Working paper 004, LIFE, 2003.
- [154] M. Romano and N. Touzi. Contingent Claims and Market Completeness in a Stochastic Volatility Model. *Mathematical Finance*, 7(4):399–410, 1997.
- [155] G. A. Willard. Calculating Prices and Sensitivities for Path-Independent Derivatives Securities in Multifactor Models. *The Journal of Derivatives*, 5(1):45–61, 1997.
- [156] E. Renault. Econometric Models of Option Pricing Errors. In D. Kreps and K. Wallis, editors, *Advances in Economics and Econometrics: Theory and Applications*, volume 3. Cambridge University Press, 1997.

- [157] J. Hilliard and J. Hilliard. Matching Non-Synchronous Observations in Derivative Markets: Choosing Windows and Efficient Estimators. *Quantitative Finance*, 12(1):49–60, 2012.
- [158] S. L. Heston and S. Nandi. A Closed-Form GARCH Option Pricing Model. *Review of Financial Studies*, 13:585–625, 2000.
- [159] J.-C. Duan. The GARCH Option Pricing Model. *Mathematical Finance*, 5:13–32, 1995.
- [160] J.-C. Duan. Conditionally Fat-Tailed Distributions and the Volatility Smile in Options. Working paper, 1999.
- [161] J. M. Harrison and D. M. Kreps. Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory*, 20:381–408, 1979.
- [162] H. Pham and N. Touzi. Intertemporal Equilibrium Risk Premia in a Stochastic Volatility Model. *Mathematical Finance*, 6:215–236, 1996.
- [163] H. Föllmer and M. Schweizer. Hedging of Contingent Claims under Incomplete Information. In M.H.A. Davis and R.J. Elliot, editors, *Applied Stochastic Analysis*, volume 5 of *Stochastic Monographs*, pages 389–414. Gordon and Breach, New York, 1991.
- [164] E. Renault and N. Touzi. Option Hedging and Implied Volatilities in a Stochastic Volatility Model. *Mathematical Finance*, 6(3):279–302, 1996.
- [165] A. Javaheri. *Inside Volatility Arbitrage: The Secrets of Skewness*. Wiley Finance, 2005.
- [166] D. B. Madan, P. P. Carr, and E. C. Chang. The Variance Gamma Process and Option Pricing. *European Finance Review*, 2:79–105, 1998.
- [167] W. Newey and K. West. Automatic Lag Selection in Covariance Matrix Estimation, The Review of Economic Studies. *Review of Economic Studies*, (61):631–653, 1994.

- [168] R. Cont and P. Tankov. Calibration of Jump-Diffusion Option Pricing Models: A Robust Non-Parametric Approach. Working Paper 490, CMAP, Ecole Polytechnique, 2002.
- [169] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publishers, 1996.
- [170] T. G. Andersen, T. Bollerslev, and S. Lange. Forecasting Financial Market Volatility: Sampling Frequency vis-a-vis Forecast Horizon. *Journal of Empirical Finance*, 6(5):457–477, 1999.
- [171] T. Lehnert. Mandelbrot and the Smile. Working paper, LIFE, 2005.
- [172] J.-C. Duan. GARCH and Stochastic Volatility Option Pricing. Manuscript, 2000.
- [173] D. Shimko. Bounds of Probability. *RISK*, 6:33–37, 1993.
- [174] Y. Aït-Sahalia, Y. Wang, and F. Yared. Do Option Markets correctly Price the Probabilities of Movement of the Underlying Asset? *Journal of Econometrics*, 102:67–110, 2001.
- [175] D. S. Bates. Empirical Option Pricing: a Retrospection. *Journal of Econometrics*, 116:387–404, 2003.
- [176] B. Dumas, J. Fleming, and R. E. Whaley. Implied Volatility Functions: Empirical Tests. *The Journal of Finance*, 53:2059–2106, 1998.
- [177] P. Newbold and D. I. Harvey. *Forecast Combination and Encompassing*, chapter 12, pages 268–283. Blackwell Publishing Ltd., 2007.
- [178] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. L. Winkler. The Accuracy of Extrapolation (time series) Methods: Results of a Forecasting Competition. *Journal of Forecasting*, 1(2):111–153, 1982.

BIBLIOGRAPHY

- [179] S. Makridakis and R. L. Winkler. Averages of Forecasts: Some Empirical Results. *Management Science*, 29(9):987–996, 1983.
- [180] D. Harvey, S. Leybourne, and P. Newbold. Test for Forecast Encompassing. *Journal of Business and Economic Statistics*, 16(2):254–259, 1998.
- [181] F. X. Diebold. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. Working Paper 12–035, Penn Institute for Economic Research, 2012.
- [182] K.D. West. Asymptotic Inference about Predictive Ability. *Econometrica*, 64:1067–1084, 1996.
- [183] K. D. West. Chapter 3 Forecast Evaluation. In G. Elliott, C.W.J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 1, pages 99–134. Elsevier, 2006.
- [184] R. Giacomini and H. White. Tests of Conditional Predictive Ability. *Econometrica*, 74(6):1545–1578, 2006.
- [185] F. X. Diebold and J. A. Lopez. Forecast Evaluation and Combination. In G.S. Maddala and C.R. Rao, editors, *Handbook of Statistics*, volume 14, pages 241–268. Amsterdam: North-Holland., 1996.
- [186] C. W. J. Granger and R. Ramanathan. Improved Methods of Combining Forecasts. *Journal of Forecasting*, 3(2):197–204, 1984.
- [187] J. A. Mincer and V. Zarnowitz. The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER Chapters, pages 1–46. National Bureau of Economic Research, Inc, 1969.
- [188] K. D. West and M. W. McCracken. Regression-Based Tests of Predictive Ability. Working Paper 226, National Bureau of Economic Research, 1998.
- [189] D. W. K. Andrews and J. C. Monahan. An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica*, 60(4):953–966, 1992.

BIBLIOGRAPHY

- [190] D. Sul, P. C. B. Phillips, and C-Y Choi. Prewhitening Bias in HAC Estimation. *Oxford Bulletin of Economics and Statistics*, 67(4):517–546, 2005.
- [191] G. Wu and Z. Xiao. A Generalized Partially Linear Model of Asymmetric Volatility. *Journal of Empirical Finance*, 9(3):287 – 319, 2002.
- [192] J. Yu. A Semiparametric Stochastic Volatility Model. *Journal of Econometrics*, 167(2):473–482, 2012.
- [193] R. Liesenfeld. A Generalized Bivariate Mixture Model for Stock Price Volatility and Trading Volume. *Journal of Econometrics*, 104(1):141–178, 2001.

Affirmation

I hereby affirm that I have completed my doctoral thesis entitled ‘Multifractal Models: Estimation, Forecasting and Option Pricing’ entirely on my own and unassisted, and that I have specially marked all of the quotes I have used from other authors as well as those passages in my own work that are extremely close to the thoughts presented by other authors, and listed the sources in accordance with the regulations I have been given.

Hagen, 31. July 2014

(Andrés Esteban Leövey)

Curriculum Vitae

Birth date & place: 22.02.1979 in Buenos Aires, Argentina.
Email: leoevey.andres@yahoo.de.

Work experience

Dec 2012 - Present **ENERVIE A.G.** (Hagen, Germany). Quantitative Analyst. Energy Trading.
Jul 2011- Jun 2012 **CAU zu Kiel.** (Kiel, Germany). Research Associate. DFG project on Multifractal Models.
May 2007- Nov 2008 **HSH NordBank A.G.** (Kiel-Hamburg Germany, London UK). Trainee Global Markets - Quantitative Analyst. Capital Markets.
Jul 2006- Oct 2006 **HSH NordBank A.G.** (Kiel-Hamburg, Germany). Intern. Market Risk Management.

Studies

Nov 2008-Present **CAU zu Kiel.** (Kiel, Germany). Ph.D. Candidate. Quantitative Economics Program.
Apr 2007 **HU zu Berlin.** (Berlin, Germany). M.Sc. in Economics & Management.
Dec 2002 **University of Buenos Aires.** (Buenos Aires, Argentina). Licentiate Degree in Business Administration.
Dec 1996 **Instituto Santa María de los Ángeles.** (Buenos Aires, Argentina). Baccalaureate degree.

Awards & Grants

2009-2011 Schleswig-Holstein Scholarship for Junior Researchers. Germany.
2005, 2007 Top 5 student and Dean's List. Humboldt Universität zu Berlin. Germany.
2002 Cum Laude Merit. Universidad de Buenos Aires. Argentina.
1996 U-18 National Volleyball Team of Argentina.

Languages

Spanish: Mother tongue.
English: Excellent. TOEFL (280 from 300). GRE (verbal skills 590).
German: Excellent.