ESSAYS IN VOLATILITY RESEARCH

by

TRISTAN LINKE

MSc in Quantitative Finance, Lancaster University (2011)

THESIS SUBMITTED TO THE DEPARTMENT OF ACCOUNTING AND FINANCE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF

DOCTOR OF PHILOSOPHY IN FINANCE

at

THE MANAGEMENT SCHOOL,
LANCASTER UNIVERSITY

2017

Supervisor:

Stephen J. Taylor, Professor of Finance

Thesis Advisor:

Prof. Ser–huang Poon, Professor of Finance

External Examiner:

Prof. Nick Taylor, Professor of Finance

Internal Examiner:

Prof. Ingmar Nolte, Professor of Finance and Econometrics

Preface

This thesis has been written during my time as a Ph.D. candidate in Finance at Lancaster University and as a visiting Ph.D. scholar at the Econometrics Department of the University of Amsterdam. This Ph.D. thesis starts with an introduction to finance for a general audience. Followed by an extensive literature overview, which I have continued to enrich over the years, of interesting topics in the econometrics–intense field of finance so I felt the need to incorporate it into my Ph.D.—which is a documentation of my work over the last years. I then start with a very brief teaser regarding finite sample properties of the classical skewness estimator and a robust alternative. This paves the way to the first paper in this thesis on Realized Skewness, Asymmetric Volatility and Risk Management which is co–authored with my supervisor and was influenced by my time spent in Amsterdam. The next paper, asymmetric dynamics in index volatility and constituent correlation was my first work after completing my M.Sc. in Quantitative Finance at Lancaster University on jump detection methodologies. It largely developed during the first period of my Ph.D. with the great support of my supervisor and my thesis advisor. These two papers are linked by an overarching theme: the asymmetric effects of returns on future volatility and vice versa. The last chapter, co–authored with Cisil Sarisoy, developed more recently and concentrates on transformations of the covariance matrix, also known as the beta, in a noisy setting.

Acknowledgements

Of course this thesis would not have existed if it were not for the support from so many different people over the years. I would here like to take the opportunity to thank in particular my supervisor Stephen J. Taylor for his invaluable guidance, his advice with applications, and his insightful contributions to our joint research projects.

Without knowing of my existence, Stephen already guided me through my undergraduate thesis on volatility spreads at the ESB Business School through his book Asset Price Dynamics, Volatility and Prediction, which ultimately draw me into the newly opened master degree programme in quantitative finance at Lancaster and I felt like Robert Cole visiting Ibn Sina in the book the Physician by Noah Gordon when he offered to supervise me for my M.Sc. and later on for my Ph.D. dissertation. Very special thanks also goes to my thesis advisor Ser–huang Poon at Manchester Business School who was guiding me very well and shared lots of interesting research ideas over the course of my studies and providing me the opportunity to present my work at Manchester, and to my sponsor at the University of Amsterdam Peter Boswijk, who enabled me to stay for nine months at the University at Amsterdam and always seemed curious in my research and gave many insightful comments and suggestions for improvements. I would also like to thank Mark Shackleton and Ingmar Nolte for guiding me through the literature and commenting on many research ideas and for their open door policy.

The Department of Accounting and Finance at Lancaster University, the Department of Econometrics at the University of Amsterdam, and the Department of Finance at Manchester Business School have all been very stimulating and interesting environments to work in. The great seminar series and brown bag seminars have been very inspiring for my work. I am also grateful for the financial support throughout my studies provided by the Department of Accounting and Finance at Lancaster, the scholarship by Lancaster University Management School, scholarshop by the Northwest Doctoral Training Center of the ESRC and in particular I would like to thank Steve Young or enabling me to teach on the departmental programmes, and the trust in me in supervising several master dissertations.

No matter where I have been, I have had the privilege to be surrounded by great Ph.D. colleagues, and I have very much enjoyed their company during this endeavor. Thank you to my officemates Rui Fan, Xi Fu, Yang Liu, Andrei Lalu and in particular Tobias Langenberg for academic and non–academic discussions.

Most importantly, I would like to thank Cisil Sarisoy for all her support and all our inspiring research talks over all these years.

Lastly, I wish to thank my friends, my parents and my sister for your patience and support over the years, and for bearing over with me being slightly absentminded at times.

Declaration

This thesis is submitted to Lancaster University in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented including data generated and data analysis was carried out by the author.

Contents

Introduction

Introduction to the Financial Economic Setting

IN REAL–LIFE we are exposed to a variety of possibly interrelated risks as is the case in the domain of finance with the difference being that in the latter case risks may be directly tradable. If so, one should reasonably expect those risks to be priced, albeit in sometimes fragmented markets through in–transparent market mechanisms with divers market participants trading at different time horizons both in terms of execution and holding period.

In the case of exchange–listed as opposed to over–the–counter traded financial instruments researchers can more easily obtain a set of information from which a priced compound risk premium can be inferred provided the underlying (pricing) model is correct. Model misspecification plays in fact a very important role but it also raises notoriously difficult questions to answer, which is why it often remains unaddressed.

Risk premia are then further related to economically interpretable risk factors whose individual, relative contribution to total risk and, thus the absolute exposure to risk may well change over time.

The variation in returns calculated from a financial asset price process is non–stationary. It is stochastic. Although theoreticians may argue that when conditioning on the realized asset price path the variation might in fact be deterministic and, going further, piecewise constant. The investor faces exposure to the instantaneous states such as variance, derived from a relatively short realized asset price path, and to the stochasticity thereof. However, we can still gain insights when simplifying assumptions are made. Particularly over long horizons time–varying volatility becomes less important.

Recent financial econometric work has also brought forward evidence that theoretical, often affine financial asset price models may need to feature price discontinuities, i.e.,

jumps, which are superimposed upon an otherwise continuous realized asset price path. Whenever these are non–diversifiable, analogously there should exist a premium based on the prevailing instantaneous arrival rate, i.e., the probability of occurrence, its size distribution, and for the stochasticity thereof—in terms of intensity and parameters related to its size distribution. Currently, there is an ongoing debate regarding the contribution of discontinuities to the total variation of the price price process, ranging from 20% to only 1% percent—depending on how inference is conducted, and where these jumps occur in parameterized models. The equity (price) risk premium can then be understood as the compendium of exposures to risks.

From a macrofinance perspective, risk premia can be linked to macroeconomic variables and preference parameters such as risk aversion. Hence establishing a link between the macroeconomy and asset prices in a joint way. Moreover, there is a strand of literature trying to link the aversion of the economic agent with the uncertainty attached to asset prices and hence establishing an intuitive link between the dynamics of asset prices and uncertainty aversion. Based on this, models of macrofinance should provide insights and/or generate the dynamics of asset prices.

It is important to remind ourselves that all price models are approximations to the unknown stochastic process which generates observed prices. Financial econometric empiricism and financial economic theory are rarely advanced at the same time, both of which can progress substantially without relying too much on the other while eventually each disciplines the other. Measurement is important as it builds the foundation and key inputs for theoretical models which are built towards isolating signal in noiseless, controlled settings.

This thesis focusses in great detail on the measurement part, in particular on the dynamics and co–dependencies between asset price paths and volatility in univariate, bivariate and in simplified multivariate settings.

Financial Econometric Rationale for Chapters 1, 2, 3

Chapter 1: Realized Skewness, Asymmetric Volatility, and Volatility Feedback

Higher moments, identification, and misspecification: Overall, the interest in this thesis lies on obtaining a better understanding of the higher moments of the density gen-

erated by the underlying asset price dynamics and their relation to one another through non–linear transformations. The second moment has received a great deal of attention over the last decades from the recognition of its time–varying nature to linking innovations in measures of variation back to realizations in the return residual process with dependent innovations either occurring concurrently or lagged. It then depends on the sampling frequency and noise inherent in the data whether the econometrician is able to correctly identify and moreover correctly specify the underlying economic interdependencies of functions of the return generating process.

Lead lag co–variation terms of functions of returns and spot variation naturally extend to the definition of skewness: It is key to recognize that the asymmetric nature of the link between realizations in the return residual process with innovations in the variance process and vice versa explain the stylized facts regarding the third moment and its non–linear transformation such as skewness which we document and explain here from both a measuring point of view in noisy settings and from a theoretical model based (parametric) point of view.

Parsimonious univariate measures: Aggregate multivariate constructs, in which stocks are weighted constituents of large index tracking portfolios, are generally more parsimoniously modeled and measured in their univariate form due to errors–in–variables issues when aggregating estimates of large multivariate systems under the alternative. The univariate way in which we propose to measure non–parametrically certain functional forms of the third moment under very mild assumptions in Chapter 1 is in congruence with that line of thought, both at the index level and at constituent stock level.

Harvesting the Realized Methodology: We extend the class of estimators belonging to the realized methodology framework from a previous focus on measures of variation to the aforementioned measures of co–variation between functions of returns and spot variation which extend to the definition of skewness, here and henceforth denoted realized skewness estimators. Existing proposed increasingly noise–robust and more efficient estimators, under certain assumptions, are applied, examined and compared in terms of their results for realized skewness estimates. Note that the conventional measure of skewness is a subset of the realized skewness estimator suggested here under simple assumptions and at a certain sampling frequency. The realized skewness estimator builds on the aggrega-

tional properties of the data generating process under certain assumptions. As such the information set is richer which is again reflected in a smooth evolution of realized skewness estimates as a function of extending return horizons compared side–by–side to the conventional measure of skewness, which evolves erratic and noisy. The conventional measure of skewness suffers from the small sample size, in which the sampling frequency dictates the measuring frequency as we explain later on, which would result in large standard errors when the return horizon is long.

Empirical Findings: From an empirical perspective, we examine the contribution of the three components, i.e., the instantaneous skewness, asymmetric volatility, and volatility feedback, to total realized skewness for various proposed model–free realized skewness estimators across different sampling frequencies ranging from five to 55 minutes for index returns and constituent stock returns. We show how at the stock market index level realized skewness builds up over time, becoming more negative, and then slowly decreases in magnitude towards zero. While for stock market index returns asymmetric volatility appears to be the main driver of realized skewness across sampling frequencies and realized skewness estimators, which makes it an interesting candidate for the reduced–form semi–parametric realized skewness estimators, the attribution for constituent stocks looks different: We classify stocks into distinct realized skewness categories based on attribution by the three components (stocks with asymmetric volatility, stocks with asymmetric volatility and volatility feedback effects, stocks with large instantaneous skewness and stocks whose return multi–horizon return distribution appears to be largely symmetric).

Remarks: While we show in a theoretical way by relying on an extension of a standard commonly applied parametric model how skewness builds up over time in line with standard assumptions of the data generating process, we do not explicitly derive the standard errors of this class of parametric reduced–form estimators at this stage. However, these can be derived from stated moment conditions. The application of the proposed reduced–form parametric estimators becomes interesting when the researcher is willing to superimpose structure on the data generating process without necessarily being willing to truly belief in or estimate the most restricted version, i.e., the full parametric model.

Recall that the realized methodology framework relies on within trading day open market data. If the researcher is willing to distinctly incorporate overnight returns, which

we do not pursue at this stage, the proposed reduced–form estimators can be extended to include these terms and resulting cross–terms under the assumption that the correlation structure estimated from intraday data remains intact and applies. Then, overnight terms and cross–terms are simply scaled by the variances of non–open market period returns.

Chapter 2: Reconciling Asymmetric Dynamics in Index Volatility and Constituent Correlations

Upon the relevance of ARCH–type models: Autoregressive conditional heteroskedasticity–type models are still very much the workhorse in the domain of finance for practitioners and academics beyond the narrow focus group of financial econometricians alike—due to the models tractability, (quasi) maximum likelihood estimates are readily available, and modeling dynamics in multivariate settings (portfolios) can easily be broken down into univariate estimation of the volatility series with variance targeting and leaving much flexibility to impose almost any correlation structure on standardized (or simulated) residuals.

Realized measures are interesting in the sense of being able to observe an estimate of variance conditional on the realized asset price path, particularly as realized variance has been shown to be a credible proxy for the variance of open–to–close returns, i.e., under empirically realistic conditions, the conditional expectation of quadratic variation is equal to the conditional variance of returns (Andersen, Bollerslev, Diebold, and Labys, 2003, Corollary 1), which could be further extended to allow for overnight returns (Hansen and Lunde, 2006). However, realized variance does not yield an asset price dynamic generator and it is instead used as a plug in estimate in ARCH–type models or HAR–type reduced form volatility models, where in the latter case further assumptions need to be made to produce a financial economic (asset price) scenario generator as in Fan, Taylor, and Sandri (2017).

Frequency Domain Considerations and Temporal Aggregation: Empirical financial econometric research and findings evolve along with the availability and depth of the transaction history records. High frequency data for constituent stocks are only available since approx. two decades from the mid 1990s, with credible high frequency data for only the most liquid stocks for a much shorter time period. Once the research interest extends beyond that time period, it is no longer feasible to employ realized measures that rely on

intraday data.

The benefit of using daily and less coarse data frequencies are that single–period returns sum up to multi–period returns as there are no gaps in contrast to using intermittent high frequency data. Note, however, that both applications exercise interesting use cases: While the former addresses the exposure that an investor faces over the course of consecutive trading days when closing out her position over night, the latter case is applicable for a long term investor who stays invested over longer periods of time. ARCH–type models do not temporally aggregate, meaning that when estimated at one frequency and then simulated, the properties of multi–period returns do generally not equate to those that follow from the same model which is estimated at a lower frequency theoretically; refer to Drost and Werker (1996) for continuous time GARCH diffusions instead.

However, this property serves as a useful control mechanism when inference is conducted at different sampling frequencies; bearing in mind that point estimates of parameters come with a standard error which is larger when less data points are used and that the information set overall is smaller when sparse sampling is applied. Still the properties of the data generating process which we are interested in here hold for those sampling frequencies at which the models are estimated respectively such as daily and weekly returns. For properties of multi–horizon returns the analysis in the previous chapter shall lead to more easily interpretable results as, e.g., the skewness of the density of monthly returns based on the simulation of estimated ARCH–type models at more granular frequency may be misspecified.

Asymmetric Volatility & Asymmetric Constituent Correlations Methodology: We consequently set out in Chapter 2 to explore the inherent dependence structure among constituent stocks in order to reconcile the large magnitude of asymmetric volatility response at the aggregate portfolio (index) level with the less pronounced univariate asymmetric volatility effects at the individual stock level.

We revert to univariate estimation of model based conditional asymmetric volatility based on its own univariate residual process. We also tested to include lagged market shock residuals as explanatory variable(s) in each constituent's heteroskedasticity process. The rationale for doing so rests on the observation that (asymmetric) ARCH–type processes were originally designed to be used at the index level and consecutively applied one to one to common stocks. Particularly, we were hoping to find that the asymmetry governing

parameters of common stock variance processes is related to negative market residuals which turns out not to be strongly supported by the data; this part is thus excluded from the analysis for brevity but available upon request.[1]

The most common choice for asymmetric ARCH–type models are the EGARCH model (Nelson, 1991) and the GJR–GARCH model (Glosten, Jagannathan, and Runkle, 1993) with the main difference being in the news impact curve and the benefit of ensuring positivity without imposing additional constrains for the exponential garch formulation. While we test other specifications, we focus on these models which have proven to withstand the test of time and to save ourselves from (joint) data and model dredging. In order to ensure ourselves against potentially flat likelihood objective functions and thus numerical convergence issues, we estimate parameters based off a vast dimensional grid of starting value and also test parameters estimates from previous rolling windows, which results in several hundred thousands of estimation sets.

For index and constituent data covering almost one century, we estimate univariate conditional variance models based on forward rolling ten year windows at daily and weekly sampling frequency. The window size is chosen in order to balance having a large enough sample size for these observation driven models while keeping the horizon short enough to reasonably ensure stationarity

In order to facilitate the analysis of showing how asymmetric constituent correlations lead to asymmetric volatility effects at the aggregate portfolio level, inference results need to rely on exactly the same information set of index returns and constituent returns. We address this issue in several ways by a novel reconstitution methodology: Firstly, as the Dow Jones Industrial Average is a price weighted index, we create two additional indices. An equally weighted index rebalanced daily and a market capitalization index rebalanced daily—both keeping track of stock splits, mergers, name changes, new issuances etc. Sampling takes place at the daily level which can be aggregated to coarser frequencies such as weekly. Note that at each rolled forward estimation date we get the Dow Jones constituents as of that date and create both equally weighted and market capitalization weighted portfolios of these constituents over the backward looking ten year period.

We then propose one simple estimator of the average threshold exceedance correlation across all portfolio constituents, where the threshold can be set to zero or specified as a

[1] In continuing research, we have started to extend the framework to allow for asymmetries using the realized beta garch model of Hansen, Lunde, and Voev (2014), which was presented at the 2nd Konstanz Lancaster Financial Econometrics Workshop in 2015 in Konstanz, which is, however, not part of this thesis.

function of standardized constituent residuals. This natural estimator conditions on the estimated constituent return series volatilities but is otherwise intentionally left unconditional. The estimator can thus be used as a constraint in constructing portfolios via constituent selection and weight optimization which minimizes the downside exceedance correlations in an unconditional way for long term investors that does not active market timing.

Empirical Findings: Since the asymmetric components in conditional volatility models were first discovered in the 90s, we witnessed a gradual and strong increase in the asymmetric effect at the index level, which holds for all major US stock market indices, to the extend that almost all future updating of the one step ahead conditional variance comes from negative lagged return residuals.

To the best of our knowledge, for the first time we go beyond that time period and document that there have been equally high asymmetric periods throughout the 1900s and the magnitude of the effect is time–varying—this is robust against a range of alternative asymmetric specifications which are available from the author upon request, but not reported here for brevity except for the GJR and EGARCH specification. We also note that the effect is more pronounced at the weekly sampling frequency opposed to the daily sampling frequency which is in line with the results in the previous chapter. The effects are weaker for individual firms than for aggregate portfolios: Correlations between constituent returns increase when the market falls, which explains the higher index asymmetry. Thus the index has more asymmetry than the firms. Asymmetry has increased over the last three decades both for the index and the mean firm.

Chapter 3: The Role of Noise in the Estimation of Betas

The variance and bias trade-off: In the last Chapter, we consider the relation between classic OLS betas and realized betas: It has been an interest to use highly frequent data in estimating assets' sensitivities to systemic risks, β, in order to benefit from statistical precision gains.

Sheppard (2006) notes that estimates of realized covariance sampled at five minute frequency produce a strong bias towards zero even for the most liquid stocks, with sam-

pling frequencies higher than 20 minutes still leading to a significant bias towards zero. Sheppard (2006) further revisits two models for their ability to explain common empirical regularities finding that for standard models where prices are contaminated with stochastically independent noise are unable to explain the behavior of realized covariance as the sampling frequency increases.

However, we show that when the prices are contaminated by i.i.d. microstructure noise, estimates of the betas are biased and derive a closed–form expression for it. The estimates are biased towards zero, the magnitude of the bias rises with sampling more frequently. We conduct a simulation analysis to examine the properties of the estimates of the betas. We demonstrate that there is a trade-off caused by bias and the variance and lead to an optimal sampling frequency (in the RMSE sense) among a wide range of frequencies examined. We conduct a simulation analysis as we were not yet able to derive the optimal sampling frequency analytically.

Financial Econometric Literature Review

Before we proceed to the three core chapters, we will review relevant financial econometric building blocks that should help the reader to put the thesis and its contributions into the relevant context. As the first and second chapter both focus on asymmetric volatility and skewness we view it as more parsimonious to conduct one large literature review in the introductory chapter rather than having two largely repetitive literature reviews in this thesis. We start by reviewing stochastic volatility models which leads us to the model considered in Chapter 1: While we do not feature discontinuities in an otherwise smooth model, we favour a discrete exponential multiplicative component volatility model with can likewise produce very rough paths. As many stochastic volatility models are driven by the intend to more adequately option price dynamics we also review this strand of literature incl. non–parametric risk–neutral skewness research from implied volatility curves. Eventually real–world and risk–neutral dynamics should be theoretically connected but it is hard to quantify the exact linkages. [2] We then review the measurement theory and applications falling under the realized methodology as it is relevant to understand the realized skewness estimator we propose in Chapter 1 under the various asset price dynamic assumptions (jump diffusion setting etc.). We then turn to ARCH–type models

[2] Moreover, estimates of risk premia derived from risk–neutral/real–world dynamics are not tested for significance.

in univariate and review multivariate settings which sets the analysis in Chapter 2 into the appropriate context. We conclude the literature review with economic explanations regarding the asymmetric volatility and skewness empirical regularities.

In volatility research we can broadly distinguish three mutually influencing and now partly overlapping concepts: One, stochastic volatility; two, autoregressive conditional hetereoskedasticity–type models; three, forward–looking option–implied volatility. More an underlying theme than a modelling technique influencing all three areas is the sampling frequency which has evolved from daily to intraday data (high frequency up to one minute), tick data (ultra high frequency below one minute). There is always a trade–off between non–trivial residuals which may arise due to (complex) microstructure noise at high frequencies vs. loss of information at low frequencies and inefficiencies due to transaction costs.

A Brief Note about Time Deformation

Inspired by Houthakker, Mandelbrot (1963) first established the notion of volatility clustering, whose associated non–stationarity property of the return generating process has also been perceived by Black and Scholes (1972) in empirical tests of their option pricing formula. Probably also due to the elegance of their derivation and a not well–defined economic understanding of the variance risk dynamics, the detected apparent *anomaly* has not been considered as significant enough to impact pricing and outweigh transaction costs at that time and was consequently neglected. Almost twenty years later, Mandelbrot (1982) briefly comments on his seminal 1963 work re–emphasizing how his approach based on invariance principles decouples prices formation from investor inertia and fits constant jumps and swings without necessarily having to introduce (time–varying) co–dependence among the model building blocks, which might have in turn motivated Merton's (1976) idiosyncratic jump-diffusion model with constant instantaneous (an average implied) variance.

The financial economist Clark (1973) introduced the theoretical idea of Bochner's (1949) time–deformed Wiener process W_τ, as both a latent and a deterministic function of variables such as trade volume, into finance. Assuming independence in W and τ, log returns p are a mixture of normal distributions when observed in *calendar time*, $p_t = W_{\tau_t} \quad \rightarrow \quad p_t|\tau_t \sim N(0, \tau_t)$. Follow up articles include Tauchen and Pitts (1983), who distinguish between information arrival, divergent market opinions of trading agents and

the number of active traders, Andersen (1996), who allows for informational asymmetries and liquidity effects to impact the trade process in response to news, and Ane and Geman (2000), who advocate the number of trades as a better proxy for market activity in their jump–diffusion framework. Recently, Andersen, Bondarenko, Kyle, and Obizhaeva (2015) find that return variation per transaction is log–linearly related to trade size in contrast to previous studies which are trying to relate volatility to transaction counts or trading volume.

The trading frequency explained through a time deformation from calendar to *asset intrinsic* time by the opportunity to execute a trade either independent of volume or as a fixed amount of e.g. shares outstanding/ percentage of free-float should in the long run not influence the intrinsic value of that share as all business activities take place in calendar time.

More importantly, the time deformation via τ is simply a non–negative, non–decreasing, non–autocorrelated process which introduces variation in volatility but cannot distinctly address empirically observed volatility clustering.

Foundations of Stochastic Volatility

Origins of Stochastic Volatility: The first published research on stochastic volatility with clustering effects induced by autocorrelation can be found in Taylor (1980, 1982) in which returns are modelled as a non–linear product process of the form

$$R_t - \mu_X = V_t \cdot U_t,$$

where volatility shocks V_t are always positive via setting $V_t = \exp(h_t/2)$ with h_{t+1} being an AR(1) Gaussian white noise process, i.e.

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t.$$

U_t, a zero mean unit variance process, determines the news impact effect in magnitude and sign. The random variables η_t and U_t are here assumed to be independent. Relaxing this assumption yields a model which can deal with asymmetric effects and produce different levels of skewness as a function of the horizon. The discussion takes place in discrete time and motivated by the fact that news are produced at a finite rate. Extensions of the model are widely used today such as in Huang and Tauchen (2005), Barndorff–Nielsen, Hansen, Lunde, and Shephard (2009) and related papers.

Stochastic Volatility in Continuous Time: The continuous time equivalent to the model above, which also addresses asymmetries and thus skew effects, is the Hull and White (1987) model, building on Vasicek's (1977) application of a mean–reverting Ornstein–Uehlenbeck process in modelling interesting rates. Hull and White write the variance process in the two forms shown in below, where ω is a scaling factor,

$$d\sigma^2 = \alpha(\sigma^2)dt + \omega(\sigma^2)dW,$$

$$d\log(\sigma^2) = \alpha(\mu - \log(\sigma^2))dt + \omega dW,\ \alpha > 0.$$

The Heston (1993) model theoretically overcomes the drawback of obtaining negative values (as previously encountered in the standard Hull–White variance process above) by modeling a square–root process. For S_t the diffusion process and v_t the variance process we write the model in real-world dynamics as

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dz_{1,t},$$

$$d\sqrt{v_t} = -\beta\sqrt{v_t}dt + \delta dz_{2,t},$$

where μ is the drift parameter, $dz_{1,t}$ is a standard Wiener process correlated with $dz_{2,t}$ via ρ and scaled by $\sqrt{v_t}$. Itô's lemma applied to second equation above yields the CIR variance process. The risk–neutral stochastic dynamics for the log asset price $\ln[S_t] = x_t$ can thus be summarized as

$$dx_t = \left[r - \frac{1}{2}v_t\right]dt + \sqrt{v_t}dz_{1,t},$$

$$dv_t = \kappa\left[\theta - v_t\right]dt + \sigma_V\sqrt{v_t}dz_{2,t},$$

where κ is the mean–reversion parameter, θ the long–run mean of the variance, and σ the volatility of volatility parameter. When Euler discretization is applied to the variance process negative values can be obtained. In such cases the variance can either be set to zero or inverted. Other methods include sampling from the exact transition law of the process. However, when the discretization step is very small, negative values are hardly obtained and there are virtually no biases arising.

Asymmetric Volatility and Skewness in Stochastic Volatility Models: The negative ρ, governing the skewness of the conditional return distribution, produces downward sloping Black–Scholes implied volatility (IV) curves across a range of strikes which suits

markets dominated by long hedgers. Furthermore, the specification equips the conditional return distribution with excess kurtosis and BS IV curves with concavity increasing w.r.t. $|\rho|$ and σ_V, and IV curves can both be upward or downward sloping depending on the initial value at V_0; see Das and Sundaram (1999) for respective volatility smile and smirk properties; refer to the following subsection: Risk–Neutral Skewness from Implied Volatility Curves.

Heston–type models still serve today as the workhorse for volatility modelling, see Gatheral (2005). The application might not yield an adequate description of the asset price dynamics but obtained estimates are often accurate proxy enough to determine reasonable option prices due to high bid–ask–spreads and investor's willingness to pay for protection. Via Fourier inversion the solution is available in closed form making it particularly attractive. In practice, the model is often re–calibrated daily to only match option prices, i.e. minimize option pricing errors, at a target maturity.

The Heston (1993) Fourier inversion of the conditional characteristic function to calculate risk–neutral expectations in closed form encouraged the theoretical development of new models with rare jumps in prices due to the following shortcoming of the model.

Deutsche Bank (2002) states that in unconstrained parameter estimation Heston–type models barely match the earlier addressed Feller (1951) constraint and even if matched in appropriately calibrated market models the distribution of realised variance is highly peaked near zero with a heavy tail. Market calibration is consequently only achieved by outweighing the probability of long low volatility periods with very high volatility probabilities Jaeckel (2002). In the various available estimation methods using price and options data, the correlation coefficent $|\rho|$ is assigned high values (≥ 0.7) to produce the distinct skewness. In terms of variance risk requirements such a high variance–spot correlation has not been verified by econometric analysis. Among others, Benzoni (1998) rejects the specification as skewness and excess kurtosis cannot be simultaneously matched. Another related issue occurs with the earlier addressed peaks in the distribution of realised variance at low levels which can be circumvented by increasing the mean–reversion parameter to increase convergence speed towards an assumed "core stationary distribution" (in a mixture of distribution setting), see Fouque (2000) for an empirical work. While this approach has the advantage to allow for temporary variance upsurges it will be even harder to replicate the distinct skewness.

Multi–factor Stochastic Volatility Models and Jump Processes The smile of Black–Scholes implied volatilities is consequently indicating the heavy–tailed risk–neutral distribution. Bates (1996) is probably the first influential study which showed the need to incorporate a jump component in the price process with regard to FX data; followed by Scott (1996), Bates (1997), Scott (1997), Chernov and Ghysels (1998), Pan (1998), Bakshi et al. (1997, 2000), Bakshi and Madan (2000) and many more.

Duffie, Pan, and Singleton (2000) specify a general class of affine jump-diffusion models which is adaptable to the $\mathbb{Q}$ measure using their proposed transform theorem. Their generalized affine stochastic volatility model with state-dependent, correlated jumps (SVSCJ) for the logarithmic asset price process, $P_t = log(S_t)$, solves:

$$dP_t = \mu_{t,p}\, dt + \sqrt{V_{t-}} dW_{1,t} + J_t^P dN_t^P,$$

$$dV_t = \kappa\left(\theta - V_{t-}\right) dt + \sigma_V \sqrt{V_{t-}} \Big(\rho dW_{1,t} + \sqrt{(1-\rho^2)}\Big) dW_{2,t} + J_t^V dN_t^V,$$

where for the càdlàg (continué à droite, limité à gauche) specification $V_{t-} = \lim_{s \uparrow t} V_s$, the drift parameter $\mu_{t,p}$ has locally bounded variation and W_t is standard Brownian Motion in $\mathbb{R}^2$ for which the scaled diffusion terms in dP and dV are correlated via a constant instantaneous correlation coefficient ρ. Consequently, the instantaneous log price variance process is modelled as a one–factor square–root process plus a (counting) Poisson process N_t^V modelling the occurence of jumps with magnitudes coming from the distribution J_t^V. N_t^P is a Poisson process generating jumps in the asset return process, whose size follows the distribution J_t^P. For a complete definition and characterization of the affine framework we refer to the very technical document, Duffie (2003).

As special case of the affine class proposed by DPS, Barndorff-Nielsen and Shephard (2001), henceforth BNS, suggest pure non–Gaussian (positive) jump processes of Ornstein-Uehlenbeck type as the building blocks for volatility models. In the simplest form they write the following SDE for the variance process

$$d\sigma_t^2 = -\lambda \sigma_t^2 dt + dz(\lambda t),\ \lambda > 0,$$

where z is a non–negative process with independent and stationary increments. Volatility can then be modelled as a weighted sum of independent Ornstein–Uehlenbeck processes with different persistencies. From their equation we can see that the autocorrelation function is a weighted sum of exponentials which can also mimic quasi–long–range dependence.

This can be regarded as the Lévy response to Comte and Renault (1998)'s long–range dependent stochastic volatility model where the log volatility is constructed via fractional Brownian Motion. Interesting and stimulating for further research is also Section 6.3 in BNS, which outlines the idea to directly link tick–by–tick data to their proposed model dynamics; resorting to the framework by Rydberg and Shephard (2000). For more extensive research into option pricing using the BNS model I refer to Nicolato and Venardos (2003).

In a series of papers Carr, Geman, Madan, Wu and Yor generalise the time–changed stochastic process idea proposed by Clark (1973) to a general Lévy process. Starting at zero, any increasing pure jump process is suitable to model the time–deformed random clock as time cannot turn negative; I suppose unless we model the instantaneous (log) time change instead of determining the passage of time directly. Carr, Geman, Madan, and Yor (2003) state that the clock would be locally deterministic in the absence of jumps. Ruling out this undesired feature, it can be concluded that jumps in time then trigger jumps in the asset price process. An equivalent argument requiring a diffusion component in the price process is not obvious (ibid.).

Early on in their study, Geman, Madan, Yor (2001) remind us that diffusions should only be applicable in the case where prices adjust continuously to the information flow resulting in instantaneous equilibria. Discontinuous information flow should thus be modelled by discontinuous stochastic processes exhibiting finite variation unlike Brownian Motion. Under this new setting jump–diffusion can be interpreted as infinite–activity[3] pure jump processes of finite variation where large innovations exhibit a lower (finite) intensity than their smaller complements. From my point of view the question remains why there should be infinitely many jumps instead of a finite number per time interval.

The resulting model can be linked to dependently and heterogeneously occurring evaluations of a continuous diffusion process in which the random evaluation times can itself show dependence to the price process. Consequently, when the intensity rate is high enough continuity can emerge in an *activity–based* measure of time which does not need to hold for the price process in calendar time. Interesting is their modelling of the price process as the difference between two increasing, here gamma, processes, one for price increases with mean and variance $\mu_1,\ \nu_1$ and one for price decreases with $\mu_2,\ \nu_2$ such that

[3]For a high–activity level the price is non–constant for all time intervals, however the arrival rate for innovations strictly bounded away from zero is finite: "All but finitely many of the infinitely many price moves are arbitrarily small in size" Geman, Madan, Yor (2001, p81).

the log price process p_t is described as

$$p_t = \frac{\nu_1}{\mu_1}\gamma\Big(\frac{\mu_1^2}{\nu_1}t\Big) - \frac{\nu_2}{\mu_2}\gamma\Big(\frac{\mu_2^2}{\nu_2}t\Big),$$

where γ is a unit mean and variance gamma process.

Carr, Geman, Madan, Yor (2002), henceforth CGMY, empirically investigate into index and equity return dynamics and suggest that indices tend to be well described by an infinite–activity pure jump process with finite variation under both the real–world and risk–neutral measure as potentially existing idiosyncratic diffusion risk contributions from individual stocks seem to be diversified away. However, CGMY impose caps at a lower boundary for very small changes and respectively an upper boundary for large changes in approximating the finite activity process to minimize the degree of equivalence between the measure. The cost and the potential impact of doing so onto the measure change from real–world to risk–neutral remains unclear to me at this stage. Further CGMY put forward an interesting and intuitive hypothesis that open interest in OTM puts might be a plausible cause of the negative skewness observed in option–implied distributions (ibid p.307), which moves the discussion towards equilibrium rather than arbitrage pricing.

In contrast to BNS (2001) who focused on variance modelling exclusively, i.e. the price process process is still driven by Brownian Motion, the previously addressed studies are obviously concerned with a joint modelling of the price and variance process as a pure jump process in terms of a simple homogenous Lévy framework. Consequently, Carr, Geman, Madan, and Yor (2003) introduce the BNS modelled features such as asymmetries and mean–reverting volatilities into their time–changing process which governs the discontinuous Lévy process—primarily to match option prices across maturities.

Due to the distinctly separate modelling of price and variance processes in stochastic volatility models and subsequently encountered difficulties in theoretical and econometrical work even for univariate models, multivariate models are not widely spread in the SV domain—the exception probably being Fiorentini et al. (2004) which is based on the discrete–time model by Diebold and Nerlove (1989). It has influenced asset pricing factor models when volatility clusters.

Risk–Neutral Skewness from Implied Volatility Curves

Implied Volatility Curves: An implied volatility curve, i.e. the cross–section of implied volatilities calculated numerically by inverting the Black–Scholes–Merton formula to

match mid quotes of exchange traded options at one maturity on individual stock and plotted against a simple, non–standardised moneyness measure,[4] exhibits a 'smile', i.e. is almost symmetric, at short horizons and a 'smirk', i.e. is asymmetric, at longer horizons.[5]

To overcome truncation, discretisation, extrapolation biases, which arise when approximating the continuum of option prices from sparse discrete data, Jian and Tiang (2005, 2007) recommend to fit cubic splines to mid prices of index option trades and quotes over a simple but wide moneyness range.

Gatheral and Jacquier (2004, 2005, 2012, 2013) come up with a flexible parametric specification of the volatility surface which aims at eliminating static and recently also dynamic arbitrage jointly at multiple maturities in the presence of stochastic volatility and jumps. Such a formulation enables us to further analyse the dynamics of the implied volatility surface in terms of their model parameters with regard to dynamics of the underlying security. Bedendo and Hodges (2009) employ discrete and linear Kalman filter updating of the volatility skew and assess the models capability of producing good density forecasts. They claim an outperformance of their model compared to the sticky–delta and vega–gamma alternatives. Zhang and Xiang (2008) employ a second–order polynomial function to quantify the implied volatility curve and then relate level, slope, and curvature coefficients to the properties (cumulants) of the implied risk-neutral distribution of asset returns and show how such links can be used to calibrate and compare different option pricing models.

The slope of an implied volatility curve or differential slopes of left (out–of–the–money put) vs. right (out–of–the–money call) 'wing' of the implied volatility curve set in relation to at–the–money implied volatility ultimately determine but do not directly equate to the standardised third moment, i.e. skewness, of the risk–neutral return distribution. By building on the Bakshi and Madan (2000), Carr and Madan (2001) result on how to replicate any claim from a continuum of option prices in the cross–section of moneyness of one asset at one target expiry $(T-t)$, BKM show how to replicate implied moments, moreover, implied skewness and implied kurtosis of the risk–neutral return distribution, henceforth RNS and RNK, over that forward–looking horizon $(T-t)$.

[4] such as K/X, $log(K/X)$ where K denotes strike price and X denotes spot or future price

[5] We refer to short, medium and long horizons for $8 \leq T-t \leq 90$, $91 \leq T-t \leq 365$ and $T-t > 365$ calendar days.

Approximating RNS at Index and Stock Level: Slopes of individual equity options are much less negative than the index (Bakshi, Kapadia, Madan (2003) as equivalent analysis of options on stock indices such as the S&P500 reveals a strongly pronounced asymmetry at short horizons which is in turn almost flat at very long horizons under the same (simple) moneyness measure. It seems to be yet unclear whether these results hold in the case of standardised moneyness measures such as Black–Scholes–Merton (BSM) delta.[6] E.g. Wu (2012) provides some, although weak empirical evidence for index implied volatility curves as a function of BMS delta resembling the shape of a smirk also at long horizons. Robust inference at long horizons is especially hampered when the covered standardised moneyness range measured by e.g. BMS delta is shrinking over time even when the range of covered strike prices remains unchanged, which is the case under usual market conditions for index and individual stock options. Then, tails of the risk–neutral density can no longer be spanned by option prices and the density is consequently truncated, not necessarily symmetrically.

Neumann and Skiadopolous (2012) make repeated use of smooth cubic splines over a truncated delta grid to obtain constant implied maturity moments for S&P500 index options. The integrals are evaluated using trapezoidal approximation. The authors further note that extrapolating in the time dimension creates artificial spikes in implied moments from which they refrain themselves. However, in absence of a formal sensitivity test thereof, consequently created large gaps in the time series of constant maturity moments[7] can themselves lead to spurious spikes. The CBOE implied volatility and skewness reference indices VIX and Skew Index are both calculated based on a discrete strike grid with two consecutive zero bid option prices as cut–off rule for the strike price range and linear inter–/extrapolation in the time dimension to yield constant 30 calendar days moments. The CBOE Skew Index seems to be more subject to extrapolation errors as regularly recurring spikes are visible in its time series plot unlike for the VIX Index.

In order to calculate implied skewness of individual stocks, Conrad, Dittmar, Ghysels (2013) rely on procedures similar to those used by the CBOE but require an equal number of put and call options entering their calculation and impose stricter exclusion criteria. The number of qualifying out–of–the–money puts and calls on single name stocks can be sparse which emphasises the need for good approximation methods of the inte-

[6] Alternative standardised moneyness measures are the BMS $d1$, BMS $d2$ or $\frac{log(K/X)}{\sigma\sqrt{(T-t)}}$ where $T-t$ denotes the time to maturity and σ can be either at–the–money implied volatility, implied volatility or 1 (Wu 2012).

[7] when nearby contracts fall below the minimum threshold maturity and would require extrapolation

gral over a continuum of strike prices. In their cross–sectional analysis of RNS among U.S. firms, Taylor et al. (2009) require a minimum of five option prices and then employ quadratic approximation to the implied volatility curve in order to avoid less meaningful, non–arbitrage–free shapes sometimes imposed by spline methods. While the former mentioned shapes, which are a result of splining exactly through each option mid price, can distort the efficient estimation of implied moments, they might still be arbitrage–free under incorporation of large bid–ask spreads.

Skewness Trading Strategies Bali and Murray (2012) do for skewness what Goyal and Saretto (2009) did for the second moment.[8] They isolate the exposure to time variation of risk-neutral skewness by creating delta– and vega–neutral trading strategy based on 'skewness assets', i.e. portfolios of equity options and stock combinations which mimic skewness. They find a strong negative relation between risk-neutral skewness and 'skewness assets' consistent with a positive skewness preference. Schneider (2012) develops a simple skew swap trading strategy which exploits the skew in S&P 500 futures options. The strategy is executed in the futures market after initial option transactions, avoiding transaction costs and lending itself to high-frequency data to assess realized skew. The fixed leg of the simple skew swap is highly correlated with the fear index from Bollerslev and Todorov (2011b).

From Implied Volatility Surface back to Data Generating Process

Implied volatility 'smiles' ('smirks') generate non–normality in terms of altering the tail probability (asymmetry), i.e. fat–tailedness (skewness), of the risk–neutral return distribution. As mentioned earlier implied skewness essentially depends on the differential pricing of left (OTM puts) versus right wing (OTM calls) of the implied volatility curve and on standardisation by the overall level of implied volatility.[9] However, it is a common fact, see e.g. Neuberger (2012) with regard to skewness, that the ratio of two efficient estimators is not in itself an efficient estimator of the ratio. This results complicates the analysis but moreover certainly calls for a disentangling and test for significance of the effects jointly affecting implied skewness.

The different shapes of implied volatility curves seem to be mainly caused via the

[8] Goyal and Saretto (2009) find that the difference between realised and implied volatility predicts future straddle returns.

[9] $Skewness = m_3/m_2^{1.5}$ where m_p denotes the p'th moment of the distribution.

following two mechanisms: Price jumps generate non–normality for the innovation distribution. Any form of stochastic volatility generates non–normality through mixing over multiple periods.[10] Over very short time intervals only significant jumps cause departures from normality.

In absence of intraday volatility pattern but in the presence of concurrent price–volatility cojumps, Ait–Sahalia et al. (2013) have shown that strong negative correlation between the continuous or high–activity components of an asset's price and its variance process exists also at high–frequency for S&P500 futures at the one–minute frequency and for one large cap US stock, Microsoft. Consequently, the aforementioned short intervals over which only jumps contribute to non–normality might have to be limited to intraday horizon and we can complement the second mechanism, stochastic volatility, by negatively correlated return–volatility innovations which contribute, albeit less pronounced, to skewness of the risk–neutral density. Christoffersen, Heston, Jacobs (2009) note that the slope and the level of the smirk fluctuate largely independently such that a more credible model should be able to generate stochastic correlation between volatility and stock returns. They propose a two–factor stochastic volatility model but do not provide a benchmark test against other models from e.g. the Duffie, Pan, Singleton (2000) framework or other pure jump or mixture Levy models.

Convergence: In Asymptotics and Measure–wise. Under the functional central limit theorem on stochastic processes, henceforth FCLT, a return distribution converges to normality with aggregation under weak conditions such as finite return variance; originally due to Donsker (1951) under rather simple assumptions. As stock option maturity increases, the 'smirk' should flatten whereas evidence suggests it steepens. Nevertheless, FCLT seems to hold statistically fine under the real–world measure for index returns and individual stocks.

Measurement of real–world skewness is heavily plagued by outliers with only few observations available (Kim and Kong, 2004). As real–world skewness can be hardly assessed, there is consequently less research about how risk–neutral information can be used in forecasting real–world skewness. Moreover, asymptotic convergence results are well–established for power functions of σ^{2p} for even moments (realised variance and quarticity, p = 1, 2) in the absence of leverage effect in a series of papers by Barndorff–Nielsen and

[10] When volatility is assumed to be piecewise constant under consideration of the U–shaped intraday volatility pattern.

Shephard and many others[11] but not well–established for uneven moments, in particular p = 1.5. Further, it is not immediately obvious how to aggregate the obtained skewness of HF returns at the daily level, as skewness unlike variance does not scale nicely with time. Nevertheless, Buckle, Chen and Williams (2012) test the stability of realised higher moments of Dow Jones 30 stocks and analyse its impact on portfolio optimisation.

Outside the field of financial econometrics, Brys, Hubert, and Struyf (2003, 2004, 2004) suggest robust alternative measures of skewness based on the median over certain kernels which are defined on couples. However, the convergence towards the traditional measure of skewness is yet unclear to me but could easily be evaluated via Monte Carlo methods.

Neuberger (2012) suggests the concept of 'realised skewness'. The term 'realised' does not refer to high–frequency (HF) returns but should be interpreted relative to the horizon, thus daily returns qualify as HF returns. The term 'skewness' refers to a hybrid of real–world returns and option prices such that it is contaminated by risk premia—for which arising biases are analysed. Further a re–definition of realised variance based on differences in log and arithmetic returns is required. Then, 'Neuberger skewness' of equity index returns increases with horizons up to one year with economically important magnitude.

Measurement Theory & Applications: Realized Methodology

The research into risk–neutral (option–implied) dynamics as introduced in the previous section is only one strand of literature. Due to encountered frictions in derivatives markets, the informative value of these studies to make inference about dynamics under the $\mathbb{P}$ measure seems to be somewhat limited. Available high–frequency data enables us to explore the asset price dynamics under the statistical measure more accurately.

Quadratic variation: In view of incorporating the information content inherent in intraday data Andersen and Bollerslev (1998), Comte and Renault (1998), Meddahi (2002) and Barndorff-Nielsen and Shephard (2002) mainly contributed to the theory of the so–called realised variance measure. While daily squared returns are an unbiased estimator as a proxy for daily volatility, they contain a high degree of noise due to the underlying asymmetric χ^2 distribution Lopez (2001). Under the assumption that a GARCH(1,1) process is underlying the risky part of returns in $V_t U_t$, or in adapted notation $\sigma_t \epsilon_t$ equivalently, Andersen and Bollerslev (1998) show that the R^2 of the ex–post squared daily

[11] Renault and Comte (1998), Andersen and Bollerslev (1998), Mykland and Zhang (2012), Podolski, Ait-Sahalia, Jacod and their co–authors.

return–volatility regression,

$$r^2_{t+\Delta} = a + b\sigma^2_{t+\Delta} + u_{t+\Delta},$$

"simply measures the extent of idiosyncratic noise in squared returns relative to the mean which is given by the (true) conditional return variance" Andersen and Bollerslev (1998, p7). In particular, they show that the R^2 coefficient is bounded by the reciprocal of the finite unconditional kurtosis, i.e. for conditional Gaussian errors by $\frac{1}{3}$ and for more fat–tailed error distribution by even less. In their Eq. (12) and (13) they show when correcting the forecast errors for sampling frequency bias and the induced covariance thereof, their more appropriate statistics attributes a much higher explanatory power (theoretical up to $\frac{1}{2}$ for their time series) in assessing the daily variance in one–step–ahead forecasts. Blair, Poon, Taylor (2001) report a more than tripling in the coefficient of determination when 5–minute squared returns instead of daily returns are used; using absolute instead of squared variables in the above equation they report an R^2 of 0.285.

Under the strong assumption that discretely sampled returns are serially uncorrelated and the variance sample path is continuous, it follows from Karatzas and Shreve (1988) for the probability limit with regard to Geometric Brownian Motion that

$$p \lim_{\Delta \to 0} \Big(\int_t^{t+1} \sigma^2_{t+\tau} d\tau - \sum_{j=1}^{M=1/\Delta} r^2_{t+j\cdot\Delta} \Big) = 0.$$

In a discrete setting, the sum of M squared intraday returns over a fractional period of the day of length delta (Δ), denoted $r_{t,\Delta} \equiv p_t - p_{t-\Delta}$, corresponds to the daily realised variance measure

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{M=1/\Delta} r^2_{t+j\cdot\Delta,\Delta}.$$

Consequently, variance assessed by the cumulative sum of squared intraday returns refines the actual measurement and thus the forecasting power. Andersen, Bollerslev, Diebold, and Labys (2000) and Andersen, Bollerslev, Diebold, and Ebens (2001) report three major properties of the newly estimated realised volatility and so did Areal and Taylor (2002) w.r.t. FTSE100 HF returns. One, for many markets daily returns standardised by realised volatility ($r_t/\hat{\sigma}_t$) are approximately normally distributed; two, the distribution of $\hat{\sigma}_t$ over time is approximately lognormally distributed; three the autocorrelations of volatility seem to decay slowly probably at a hyperbolic rate.

Now assume the logarithmic asset price p_t follows generic jump diffusion of the form

$$dp_t = (\mu_t dt +)\ \sigma_t dW_t + J_t dN_t,$$

where the drift term is usually negligible for sufficiently small increments as in the high-frequency domain. Moreover, we are here only interested in the risky part of returns and not in the instantaneous reward component. In the absence of microstructure noise the total ex–post return variability (quadratic variation) is then defined to be

$$QV_{t+1} = \int_t^{t+1} \sigma_s^2 ds + \sum_{t<s\leqslant t+1} J_s^2. \tag{0.0.1}$$

In the presence of discontinuities, Barndorff-Nielsen and Shephard (2002) show that (under weak regularity condition) RV asymptotically converges to QV as $\Delta \to 0$. When jumps are not present the RV measure is consistent for the total return variation coming from the diffusion component alone, denoted integrated variance IV.

Christensen and Podolski (2007) suggest high–frequency sampled squared realised price ranges which by definition inspect all available data points regardless of the interval length. Their non–parametric measure of quadratic variation is defined as

$$RRV_{t+1}(\Delta) \equiv \frac{1}{\lambda_2} \sum_{j=1}^{M=1/\Delta} s^2_{p_{t+j\cdot\Delta,\Delta}},$$

where $\{s^2_{p_{t+j\cdot\Delta,\Delta}}\}_{j=1,\ldots,M}$ are the observed intraday ranges of the log price p and λ_2 is a normalising constant which also removes downward bias induced by infrequent trading.

Nevertheless, the variance process itself remains latent as: One, discontinuities (jumps) in the price process emit additional variation such that RV estimates total variation (QV) while it is important to distinguish between diffusive and jump–induced variation being two fundamentally different sources of risk. Two, microstructure effects render the limiting case to be infeasible and thus question the reliability of the asymptotically valid result. In response, two relatively separate strands of literature have evolved dealing with jumps and microstructure noise, which are just being combined with surprising outcomes. I will start with the latter one as it affects the jump detection procedure inherently.

Microstructure noise–reduction techniques With the empirical inter–trade interval being at just a few fractions of seconds length on average for highly liquid securities, the minimum tick size movement superimposed upon the asset is regularly larger than feasible

diffusion price variation over those short horizons. Consequently, most price changes are either zero (on a discrete time scale) or too large in comparison with the expected price variation and trading under incorporation of bid–ask spreads only intensifies this effect. Moreover, in the presence of market microstructure noise (u_t) the fundamental price process (dp_t^*) remains latent,

$$p_t \equiv p_t^* + u_t,$$

$$r_{t,\Delta} \equiv p_t^* + u_t - p_{t-\Delta}^* + u_{t-\Delta} \equiv r_{t,\Delta}^* + \epsilon_{t,\Delta},$$

$$\lim_{\Delta \to 0} E[(r_{t,\Delta}^*)^2] = 0, \text{ but } \lim_{\Delta \to 0} E[(\epsilon_{t,\Delta}^2)] > 0, \text{i.e.}$$

the noise term dominates for $\Delta \to 0$ such that $RV_t(\Delta)$ is inconsistent in exploiting the limiting behaviour and it is thus counterproductive to use the entire set of price information of $\Delta \to 0$. Taylor (2005) notes that the noise introduced into prices inflates the measure of volatility.

For index related assets Andersen, Bollerslev, and Diebold (2007) advocate two-minute returns while for liquid equities the frequency is usually decreased to fifteen- to twenty-minute returns following the concept of "volatility signature", originated in Fang (1996), where for fairly constant values of the average daily realised variance plotted against the return interval length the RV measure is assumed to be free of microstructure noise bias ($E[(r_{t,\Delta})^2] \approx E[(r_{t,\Delta}^*)^2]$). *Nota bene*, the word "bias" only implies that the level of microstructure noise has been reduced to an approximately constant proportion. It does not mean that RV is entirely free of microstructure noise. From the shape of the signature plot we can learn about the properties of the noise process itself.

Another idea is to aggregate RV measures across sampling frequencies and starting points (sub–sampling) in order to infer, reduce or, asymptotically, eliminate the degree of microstructure distortion. Bandi and Russel (2008) work with mean squared errors in deriving the optimal frequency, which indicates that the advocated five–minute returns understates the optimal frequency for liquid stocks; also see Zhang (2006) for an extension from two– to a multi–scale realised variance estimator. Christensen, Podolski, Vetter (2008) also propose a bias–correction making the realised range–based estimator more robust to simple microstructure noise and they further show how to optimally split up the high–frequency data.

Inspired by Zhou (1996), Hansen and Lunde (2006) apply realised kernel-based estimators to US trade and quote equity data and point out that instead of the usual i.i.d. noise assumption, noise correlates with the "efficient" price and is time–dependent. Other methods have been proposed such as the realised kernel approach for equidistant time conversion by Barndorff–Nielsen et al. (2008), an early–stage extension thereof Li et al. (2009) and the pre-averaging approach of Jacod et al. (2009) which has received attention for ultra high-frequencies. Parallel to the noise–reduction techniques a number of jump robust volatility estimators have been proposed.

Jump robust variation measures In the presence of jumps Barndorff-Nielsen and Shephard (2004) define the concept of standardised realised bipower variation, abbreviated RBV or simply BV, which asymptotically only measures the variation IV which is attributable to the diffusion component. By showing that the expectation of the product of consecutive absolute return increments is proportional to their variance, Barndorff-Nielsen and Shephard (2004) establish the bipower variation measure as

$$BV_{t+1}(\Delta) \equiv \frac{\pi M}{2(M-1)} \sum_{j=2}^{M} |r_{t+(j-1)\cdot\Delta,\Delta}||r_{t+j\cdot\Delta,\Delta}|. \tag{0.0.2}$$

For the limiting behaviour we can summarize

$$BV_{t+1} \to IV_{t+1} = \int_t^{t+1} \sigma_s^2 ds \qquad \text{and} \qquad RV_{t+1} \to QV_{t+1} = IV_{t+1} + \sum_{t<s\leqslant t+1} J_s^2.$$

Using the definition of quadratic variation under jump–diffusion in the equation above leads to the conjecture that subtracting the integrated variance (bipower variation) from the quadratic variation (realised variance) yields an estimate of the jump component. However, doing so comes with a series of caveats; refer to the later sections.

Inspired by the theory of bipower variation a number of related robust–to–jump measures which essentially build upon this framework or on transformed power functions thereof have evolved. Mancini (2004a, 2006) and Jacod (2008) propose threshold realised variance (TRV), which is defined as follows

$$TRV_{t+1}(\Delta) = \sum_{j=1}^{M} |r^2_{t+j\cdot\Delta,\Delta}| 1_{\{|r_{t+j\cdot\Delta,\Delta}|<cM^{-\varpi}\}}, \text{ for } \varpi \in (0, 1/2),$$

where $\varpi = 0.47$ and $c = 6\sqrt{IV}$ Ait–Sahalia and Jacod (2009) and IV is estimated by bipower variation. Note that lowering c introduces a downward bias while gaining further

robustness towards jumps on the other hand. The measure is unbiased in the absence of jumps (Christensen, Oomen and Podolski, 2009). Andersen, Dobrev, and Schaumburg (2012) propose MedRV where c is a normalising constant

$$MedRV_{t+1}(\Delta) = c \sum_{j=2}^{M-1} \text{median}\big(|r_{t+(j-1)\cdot\Delta,\Delta}||r_{t+(j)\cdot\Delta,\Delta}||r_{t+(j+1)\cdot\Delta,\Delta}|\big)^2,$$

$$\text{where } c = \frac{\pi}{6 - 4\sqrt{3} + \pi} \frac{M}{M-2}.$$

Newer Concepts in the Realized Methodology: In light of these advances, researchers are currently setting out to integrate both issues under a unifying framework as jump robust variation measures are not noise robust and vice versa. Christensen, Oomen and Podolski (2011) introduce the concept of quantile–based realised variance, abbreviated to QRV, as an efficient and robust measure to noise and jumps simultaneously. Another interesting property is that it also filters out outliers. Based on tick data preferably, the idea builds upon inverting the relationship between quantiles and the spread of the normal distribution in the limit where return data is calculated at ever finer frequencies over given small time intervals. By analysing the behaviour of the resulting sample return distribution quantiles over respective finer time intervals volatility can be backed out and thus jumps or outliers can be filtered out when at most one jump occurs per (small) time interval and volatility is locally constant, see (ibid., p2f). The underlying idea is consequently similar to TRV (Mancini, 2004, Mancini, 2006, Jacod, 2008) where the global threshold for a given day is replaced by a local threshold which is fitted to the magnitude of intra–interval returns. As such it can also be regarded as a generalisation of the MedRV and MinRV (not elaborated upon here) suggested by Andersen, Dobrev, and Schaumburg (2010) but in the presence of microstructure noise. Note that sub–sampling can be applied to increases efficiency even further. For the test statistic and how to choose the time interval length, quantiles and quantile weights I refer to the article directly.

Due to space constraints we would like to allude to Hansen and Horel (2009) for Markov Chain–based estimation, Diebold and Strasser (2009) with regard to market microstruce theory related to the variation measure and Large (2011) and Oomen (2005) for business time sampling. Also note the recent advances into multivariate covariation measures going by e.g. Barndorff–Nielsen, Hansen, Lunde, and Shephard (2011), Christensen, Kinnebrock,

Podolskij (2011), and Zhang (2011) with regard to non–synchronous trading or quotes.

Test Statistics for Jumps, Identification, Implications

Note that with regard to all current jump detection procedures that we are aware of, the formulation of the null and alternative hypothesis differ from the conventional set–up as it can only be detected whether a realised discretely observed sample path contains discontinuities (jumps) and not whether the underlying (system of multivariate) processes, possibly continuous, contains a discontinuity shock term generator, i.e., it is not a formal test of the property of the generating model.

Thus, tests can determine whether a jump component is either present or absent and in the former case determine whether it is of finite or infinite activity along with an estimate of its degree of activity. Consequently, it is also possible to invert the test and estimate whether the continuous component is necessary in the presence of infinite activity jumps. Ait–Sahalia (2004) shows in theory that the continuous component can be disentangled from noise caused by infinite Cauchy–distributed jumps under a maximum likelihood framework and vice-versa, i.e. detecting infinite activity jumps imposed on the continuous model Ait–Sahalia and Jacod (2010b).

Variation–based jump test statistics Under the arguable assumption that jumps occur "rarely", i.e. restricting the generating model by excluding infinitely–active jump terms over finite time horizons, and thus with occurring jumps presumably dominating the return process at the daily level, a number of jump detection tests have been established based on variations measures.

Barndorff-Nielsen and Shephard (2004) derive the joint asymptotic behaviour of RV and BV under the null hypothesis of a diffusion process which initiated the development of new jump detection tests at the daily level based on statistically scaled differences in the variation measures. Huang and Tauchen (2005) recommend comparing the relative jump measure,

$$RJ_t = \frac{RV_t - BV_t}{RV_t}$$

with a consistent estimate of its standard errror. Building upon the jump robust bipower variation, Andersen et al. (2007) show that a randomly selected intraday return $r_{t+\xi\cdot\Delta,\Delta}$ maintains these properties over small increments which allows for the construction of a jump detection test at the increment level.

Without explicitly deriving the actual distribution, the knowledge of conditional mean and variance of a randomized intraday return, which are of an order of their daily equivalents, are assumed sufficient to let each standardized intraday return asymptotically follow a normal distribution under the assumption of a diffusion process. Being aware that $E[r_{t+1}]$ is non-zero under leverage ABD (2007b, p133-134) formally state for any standardized intraday return $\frac{r_{t+\xi\cdot\Delta,\Delta}}{\sqrt{\Delta}}$:

$$\lim_{\Delta\to 0} E\Big[\frac{r_{t+\xi\cdot\Delta,\Delta}}{\sqrt{\Delta}}\Big] = \lim_{\Delta\to 0}\frac{1}{\sqrt{\Delta}}\Big(\Delta\cdot E[r_{t+1}]\Big) = 0$$

$$\lim_{\Delta\to 0} V\Big[\frac{r_{t+\xi\cdot\Delta,\Delta}}{\sqrt{\Delta}}\Big] = \lim_{\Delta\to 0}\frac{1}{\Delta}\Big(\Delta\cdot IV_{t+1} - \big(\Delta\cdot E[r_{t+1}]\big)^2\Big) = IV_{t+1}$$

$$\frac{r_{t+\xi\cdot\Delta,\Delta}}{\sqrt{\Delta}} \sim N(0, IV_{t+1})\Big(\sim N\big(0,\Delta\cdot BV_{t+1}(\Delta)\big)\Big)$$

Under the assumption of constant daily variance the ABD test identifies jumps during day $t+1$ according to the decision rule:

$$J_s(\Delta) = r_{t+s\cdot\Delta,\Delta}\cdot I\Big[|r_{t+s\cdot\Delta,\Delta}| > \Phi_{1-\beta/2}\cdot\sqrt{\Delta\cdot BV_{t+1}(\Delta)}\Big], s \;=\; 1,2,\ldots,1/\Delta$$

$$\beta = 1-(1-\alpha)^{\Delta}$$

The test depends heavily on the choice of the "stochastic threshold level" determined by the daily significance level α and thus the intraday level β. ABD choose a very small significance level with $\alpha = 10^{-5}$ to insure themselves against type one errors under the null of a diffusion process, as empirically observed considerable intraday variation in the level of variance gives rise to heteroskedasticity effects and the conditional distribution of scaled increments will thus come from a mixture of (potentially normal) distributions.

The plain ABD test allocates an equal $(1/\Delta)$ proportion of the daily BV to any intraday return increment $r_{t+\xi\cdot\Delta,\Delta}$ which neglects the U–shaped intraday volatility pattern (high volatility during market opening and closing and lower than average volatility during the remaining trading hours), as documented for equity-related assets by Wood et al. (1985) and Harris (1986). With regard to FX data, Taylor and Xu (1997) suggest to estimate the mean proportion of variance attributable to each j'th return increment by averaging the sum of all j'th squared intraday return increments across days to gauge and thus correct for the arising intraday volatility pattern. Note that this is already pointing towards the use of the sums of squared intraday returns as consistent and unbiased daily variance estimator, i.e. later termed realised variance. For m denoting the total number of days

$(t = 1, 2, 3, \ldots, m)$ and leaving all other variable definitions unchanged, we define the j'th variance proportion as:

$$s^2_{j,M} = \sum_{t=1}^{m} r^2_{t+j\cdot\Delta,\Delta} \Big/ \sum_{j=1}^{M=1/\Delta} \sum_{t=1}^{m} r^2_{t+j\cdot\Delta,\Delta} \ .$$

Based on the asymptotic theory of bipower variation, Bollerslev, Law and Tauchen (2008) advocate the use of products of adjacent absolute return increments instead of consecutive squared return increments to exclude the contribution of jumps to the intraday volatility pattern, which is assumed to be caused by the continuous component. Further, we have to be aware that the incorporation of the distinct, empirically observed, average variance proportion per return increment in an extended ABD test might put the assumption of conditional normality for scaled return increments into question.

Ait–Sahalia and Jacod (2009) compare variability at different time scales by considering the sum of absolute return increments to the p'th power. Adapting slightly to their notation

$$B_{t+1}(p, \Delta) = \sum_{j=1}^{M} |r_{t+j\cdot\Delta,\Delta}|^p \text{ for } p > 2.$$

The test nests on the limit of the sum in Eq. () when p is larger than two and Δ goes towards zero. In the presence of jumps, Ait–Sahalia and Jacod (2009) show that B converges to a finite, non–zero limit

$$B_{t+1}(p, \Delta) \xrightarrow{\text{P}} \sum_{i=1}^{M} |c_i|^p \text{ for } p > 2.$$

When the underlying process is continuous the convergence property towards zero depends on the sequence of Δs going to zero. While the limit is clearly zero, appropriate scaling yields a positive finite limit under discrete sampling,

$$\frac{\Delta^{1-p/2}}{\mu_p} B_{t+1}(p, \Delta) \xrightarrow{\text{P}} \int_t^{t+1} |\sigma_s|^p ds = A(p)_{t+1},$$

where μ_p is the p'th absolute moment of the standard normal distribution. Ait–Sahalia and Jacod (2009) then precede to compare the measure over two time–scales, Δ and $m\Delta$, for some integer value of $m > 1$, such that the ratio of the two yields the jump test statistic $S(p, m, \Delta)$

$$S(p, m, \Delta) = \frac{B_{t+1}(p, m\Delta)}{B_{t+1}(p, \Delta)} \xrightarrow{\text{P}} \begin{cases} \frac{\Delta^{1-p/2}}{(m\Delta)^{1-p/2}} = m^{p/2-1} & \text{in absence of a jump,} \\ 1 & \text{in presence of a jump.} \end{cases}$$

If the sums $B_{t+1}(p, m\Delta)$ and $B_{t+1}(p, \Delta)$ are of the same(!) size ($S \to 1$), the test statistic suggests the presence of a discontinuity, whereas for continuous innovations only, the statistic converges towards 2 for their recommended values of $p = 4$ and $m = 2$. Appropriately scaled S converges to a standard normal distributed variable and the test can be applied for an appropriate significance level.

Motivated by Bollerslev, Law and Tauchen (2008), who study co–jumps between individual stocks and portfolios of these stocks and Andersen et al. (2007), Huang and Tauchen (2005) who also also suggest a common systematic jump component, Jacod and Todorov (2008) derive a test based on power variation measures for common jump arrivals in multidimensional processes, which we will not elaborate upon for brevity. As the time interval Δ is shortened, the possible variation of the continuous component decreases proportional to the square–root of time such that jumps can be detected for sufficiently short intervals as they are too large to be in line with feasible continuous component innovations. Tests based on this idea determine how much the process can vary over such intervals by describing the distribution of the largest continuous innovations while at the same time taking account of stochastic volatility and the properties that come along with it. The key insight is, somewhat similar to variation–based tests, that the convergence order of diffusion terms, $O_p\big(\sqrt{2\Delta \log(1/\Delta)}$, versus the order of the jump term, $O_p(1)$, which does not depend on Δ, can be used to construct the jump robust measure even when allowing for (bounded) volatility Mancini (2004). Inference results to detect jumps based on distributional extreme value assumptions have consequently been established by, mostly notably, Lee and Mykland (2008) based on local variance estimates over short observation windows and the "big" Lévy jump detection paper by Lee and Hanig (2010), both of which provide the usual extreme value theory drawbacks as the timeframe and the threshold level must be chosen.

Further interesting papers on different methodologies are Jiang and Oomen (2008), who propose a jump detection test based on the (daily) profit and loss (P&L) profile of variance swap contracts (the hedging error of its replication strategy) and Fan and Wang (2006), who introduced a wavelet–based approach.

Empirical Evidence from Jump Tests & Variation Measures with Implications: This section sheds light on the decomposition of model–free variation in high–frequency S&P500 returns for spot and tradable futures data using the above described methodol-

ogy or variations thereof. In particular, Table 1 reports the ratio $\overline{BV(\Delta)}$ / $\overline{RV(\Delta)}$ (or an adequate, equivalent ratio) when the sampling frequency is between two and five minutes inclusive and hopes this estimates the ratio $\overline{IV}$ / $\overline{QV}$, i.e. the percentage variation attributable to the continuous component (abbreviated to CC) where the remainder is assumed to be attributable to the jump component (JC). The two- to five-minute return sampling frequency builds upon a compromise between the assumed higher information content in a larger series of prices at ever finer intervals and microstructure noise arising at higher frequencies which invalidates the asymptotic theory of the BV and RV measure; as obviously $\Delta = 2$ minutes behaves rather differently than the asymptotic limit $\Delta \to 0$.

Table 1: Decomposition of model–free variation in high-frequency ($\leqslant$ 5mins) S&P500 returns

Type	Date	JC	CC	Article
Spot	2005 - 1986	5.40%	94.60%	Tauchen and Zhou (2011)
Spot	2002 - 1997	7.30%	92.70%	Huang and Tauchen (2005)
Future	2004 - 1990	4.40%	96.60%	Andersen et al. (2007) (ABD)
Future	2005 - 1990	4.90%	95.10%	Andersen, Bollerslev, and Huang (2007)
Future	2007 - 1990	6.00%	94.00%	Corsi and Reno (2012)
Future	2002 - 1982	4.40%	95.60%	Huang and Tauchen (2005)
Future	2004 - 1985	6.80%	93.20%	Bollerslev et al. (2009a)

In contrast to using daily primary and/or derivatives data as in previous parametric studies such as Bates (1996), Eraker (2004) and for weekly data Pan (2002), the model–free variation attributable to jumps decreases to approx. 4.4–7.3%. As results obtained hold only in the infeasible limiting case, obtained estimates have to be treated with special care. Note that is feasible to calculate $BV_t > RV_t$, I assume which occurs primarily when jumps are not rare and large. Furthermore, as I have shown previously in my MSc dissertation that the average ratio $\overline{BV(\Delta)}$ / $\overline{RV(\Delta)}$ on non–jump days according to the ABD test is nevertheless in the region of 96% using 1–minute SPY returns over more than ten years while the ratio is almost exactly 100% for a simple DPS Monte Carlo setting without explicitly modelling noise.

Progress and Contradicting Evidence from the Ultra High–frequency Domain Christensen, Oomen, and Podolskij (2014), henceforth COP, apply their pre-averaging methodology to tick-by-tick data and attribute only 1% of total return, i.e. quadratic

variation, to the jump component thereby giving rise to their hypothesis that jumps in volatility at high frequencies might be falsely identified as jumps in prices for lower frequencies (ibid. p1, p25). The authors refer to what Kirilenko et al. (2011) describe as a "flash crash", i.e. a "brief period of extreme market volatility".

By using tick data, COP actually look at the number of trades occurring during flash crash periods as a proxy for market activity to argue for a continuous sample path, whereas for hedging purposes executable, tradable volume at the respective price is the important figure to focus on. Following a news announcement or a flash crash, the number of trades is likely to shoot up. It is however questionable whether the intertick price distance will increase in length (a possible indication of a discontinuity) given the high proportion of automated algorithm trading especially in the high-frequency domain (informed investors) and less agile/responsive (uninformed) investors who are still trading on valuations formed prior to the news announcement, i.e. the price path might appear to remain continuous but it would have to be examined whether these prices are executable for high volumes under the assumption that uninformed investors are low volume traders by the majority or that high volume orders are executed during low volatile periods or in the close.

Moreover, such a burst in volatility has to be investigated further as to whether the actual dispersion of traded prices (up and down movements) along the pre-averaged sample path has increased. If not, it is likely that a jump is observed gradually at ultra-high frequency and thus not detected. Here, the price distance between consecutive trades may add explanatory power: If this bandwidth for a series of consecutive prices is large with positive and negative innovations it is likely to be continuous. If this bandwidth is small and changes are primarily of one sign, the process can be indicative of an occurring but gradually observed jump over multiple increments.

All these reservations aside, the distortion caused by bursts in volatility and its implications on assessing the variation attributable to a jump component in the return generating process are intriguing. Moreover, it is of general interest in which way the occurrence of jumps in volatility hinders the exact identification of jumps: Here, it is possible to argue both ways: In line with COP who suggest that volatility jumps alone can falsely be detected as jumps by prevailing test statistics and thus simultaneously occurring jumps in the variance and price process should theoretically increase the number of detected jumps.

On the other hand, when the test statistic builds upon neighbouring increments many jumps might not be detected as the bi-power variation, which asymptotically only measures

the variation attributable to the continuous component, is subsequently increased following a volatility jump. This increases the probability for small jumps of not being detected.

Introduction to ARCH–type Models and Long Memory Properties

Parallel to the origins of stochastic volatility the discrete–time autoregressive conditional heteroskedasticity (ARCH) model was proposed by Engle (1982), in which the one–step ahead volatility is conditionally deterministic and thus partly forecastable. In words, the model can be described as such that the time–varying variance of observed returns (heteroskedasticity) depends (conditional) on the variability of the process in form of a linear regression on past squared residuals (an autoregression)

$$X_t \sim \mu + \epsilon_t,\ \epsilon_t \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2. \tag{0.0.3}$$

The success of ARCH–type models can probably be traced back to the relatively easy to implement estimation procedure through the likelihood function L determined by the product of conditional densities. This is also the main feature in differentiating ARCH from SV models in which the predictable component of the density of future returns is specified implicitly through the dynamics of the model, i.e. the variance can be modelled separately leaving much flexibility to superimpose almost any (often affine) dependence structure. However, the likelihood function for SV models is not readily available whereas the log likelihood (l), exemplarily given for the ARCH(1) model with a parameter vector $\theta = (\mu, \omega, \alpha_1)$, is straightforward to obtain

$$L(\theta) = f(x_2, \ldots, x_n | x_1, \theta) = \prod_{t=2}^{n} f(x_t | x_{t-1}, \theta). \tag{0.0.4}$$

To procede numerical methods are required for maximising the log likelihood $l(\theta)$,

$$\max\ l(\theta) = \sum_{t=2}^{n} -\frac{1}{2} log(\omega + \alpha_1 (x_{t-1} - \mu)^2) - \sum_{t=1}^{n} \frac{(x_t - \mu)^2}{2[\omega + \alpha_1 (x_{t-1} - \mu)^2]}. \tag{0.0.5}$$

The extension to GARCH (1,1) which lets the conditional variance depend on its own lag followed respectively in Bollerslev (1986) and Taylor (1986) who independently defined and derived properties for the model; here stated in the (p, q) representation

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{0.0.6}$$

with boundaries $\omega,\ \alpha,\ \beta > 0$ and $\alpha + \beta < 1$ to be a sensible GARCH(1,1) model; for GARCH(p,q) boundary conditions refer to Nelson and Cao (1992). Perhaps inspired by the work of Hull and White (1987) and Stein and Stein (1991) in the continuous time literature, Nelson (1991) introduced the idea of asymmetric effects into ARCH–type models and provided empirical evidence by estimating an exponential, generalised ARCH (EGARCH) model, which was subsequently followed by e.g. Glosten et al. (1993), whose GJR model likewise aims at exploiting the sign information carried through via the residual ϵ_t in a simpler specification.

Taylor (1986) notes that the absolute return series exhibits a longer memory effect than the corresponding squared returns, which has subsequently and commonly been referred to as the *Taylor effect*. The long memory property in volatility is associated with volatility shocks assessed by autocorrelations decaying at a hyperbolic rate. **??** compare risk–neutral option price properties from various ARCH–type formulations and find that their Fractionally Integrated EGARCH–model (FIEGARCH) outperforms other ARCH–specifications with regard to the mean–reverting behaviour of volatility displayed by S&P500 long-term equity anticipation securities (leaps). Taylor (2000) extends their option pricing results. In general, we can say that long–memory effects do not always positively impact the option premium but that moreover the premium is highly influenced by the spot variance level: the speed of mean–reversion and the unconditional variance level then determines the pricing impact over the time to expiry. Some heat was taken out of the discussion, when Alizadeh et al. (2002) noticed that the sum of two AR(1) processes, which is a short memory process, can mimic the behaviour of a long memory process. It is this observation that motivates the extension to the proposed two component stochastic volatility model in Chapter 1.

Multivariate ARCH–type models and Correlation Modelling

In order to describe multi–asset dynamic dependence structures ARCH–type models have been extended to the multivariate case which enables us to capture time–varying conditional pair–wise correlations. Engle and Kroner (1995) reformulate their BEKK model for the conditional variance and correlation dynamics for asset i is as follows

$$\sigma^2_{i,t} = \omega_i + \alpha\epsilon^2_{i,t-i} + \beta\sigma^2_{i,t-1},$$

$$\sigma^2_{ij,t} = \omega_{ij} + \alpha\epsilon_{i,t-1}\epsilon_{j,t-1} + \beta\sigma^2_{ij,t-1}.$$

In words, the conditional BEKK covariance is constructed as a weighted average of the long–run covariance, the product of previous cross return shocks and previous conditional covariance. Note that two restrictions need to be imposed for feasible convergence: One, there is no asymmetric effect, as above Equations must be of the same functional form, and two, the persistency determining parameters α and β are constant for the conditional variance–covariance matrix to be positive definite.

The key insight in the Engle (2002) and Tse and Tsui (2002) DCC model is that the estimation procedure can be separated into fitting a series of univariate ARCH–type volatilities from which the dynamic conditional correlation (DCC) series is estimated allowing each asset i to maintain its own conditional variance and pair–wise correlation persistency with any other asset.

Multivariate GARCH In the multivariate case we now have an $N \times 1$ (column) vector of residuals, denoted E_t, with an $N \times N$ dimensional (assumed to be positive definite) conditional covariance matrix $\mathbf{\Sigma_t}$, i.e. $\mathbb{E}_{t-1}(E_t E_t')$. For expositional purposes denote the n'th asset with $a, b, c, \ldots, N$. The convolution of the heteroskedasticity processes is represented by the elements in $\mathbf{\Sigma_t}$ as follows,

$$\mathbf{\Sigma_t} = \mathbb{E}_{t-1}(E_t E_t') = \begin{array}{c} \\ a \\ b \\ c \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & \cdots & N \end{array} \\ \begin{pmatrix} h_t^{aa} & h_t^{ab} & h_t^{ac} & \cdots & h_t^{aN} \\ h_t^{ba} & h_t^{bb} & h_t^{bc} & \cdots & h_t^{bN} \\ h_t^{ca} & h_t^{cb} & h_t^{cc} & \cdots & h_t^{cN} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_t^{Na} & h_t^{Nb} & h_t^{Nc} & \cdots & h_t^{NN} \end{pmatrix} \end{array},$$

with Cholesky factor $\mathbf{\Sigma_t^{1/2}}$ and $N(N+1)/2$ distinct elements in the (upper) triangle of the matrix. A general multivariate GARCH(1,1) model can consequently be written as

$$vech(\mathbf{\Sigma_t}) = vech(\mathbf{\Omega}) + \mathbf{A} vech(E_{t-1} E_{t-1}') + \mathbf{B} vech(\mathbf{\Sigma_{t-1}}),$$

where the *vech* operator converts the upper triangle of a symmetric matrix into an $N(N+1)/2 \times 1$ column vector such that

$$vech(\mathbf{\Sigma_t}) = \begin{pmatrix} h_t^{aa} \\ h_t^{ab} \\ h_t^{bb} \\ h_t^{ac} \\ h_t^{bc} \\ h_t^{cc} \\ \vdots \\ h_t^{NN} \end{pmatrix} = \begin{pmatrix} \omega^{aa} \\ \omega^{ab} \\ \omega^{bb} \\ \omega^{ac} \\ \omega^{bc} \\ \omega^{cc} \\ \vdots \\ \omega^{NN} \end{pmatrix} + \mathbf{A} \begin{pmatrix} \varepsilon_{t-1}^{a}\varepsilon_{t-1}^{a} \\ \varepsilon_{t-1}^{a}\varepsilon_{t-1}^{b} \\ \varepsilon_{t-1}^{b}\varepsilon_{t-1}^{b} \\ \varepsilon_{t-1}^{a}\varepsilon_{t-1}^{c} \\ \varepsilon_{t-1}^{b}\varepsilon_{t-1}^{c} \\ \varepsilon_{t-1}^{c}\varepsilon_{t-1}^{c} \\ \vdots \\ \varepsilon_{t-1}^{N}\varepsilon_{t-1}^{N} \end{pmatrix} + \mathbf{B} \begin{pmatrix} h_{t-1}^{aa} \\ h_{t-1}^{ab} \\ h_{t-1}^{bb} \\ h_{t-1}^{ac} \\ h_{t-1}^{bc} \\ h_{t-1}^{cc} \\ \vdots \\ h_{t-1}^{NN} \end{pmatrix},$$

where $\mathbf{A}$ and $\mathbf{B}$ are matrices both of dimension $N(N+1)/2 \times N(N+1)/2$, here exemplarily for $\mathbf{A}$,

$$\mathbf{A} = AA' = \begin{array}{c} \\ aa \\ ab \\ bb \\ \vdots \\ NN \end{array} \begin{array}{c} \begin{array}{ccccc} aa & ab & bb & \cdots & NN \end{array} \\ \begin{pmatrix} \alpha^{aa,aa} & \alpha^{aa,ab} & \alpha^{aa,ac} & \cdots & \alpha^{aa,NN} \\ \alpha^{ab,aa} & \alpha^{ab,ab} & \alpha^{ab,bb} & \cdots & \alpha^{ab,NN} \\ \alpha^{bb,aa} & \alpha^{bb,ab} & \alpha_t^{bb,bb} & \cdots & \alpha^{bb,NN} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha^{NN,aa} & \alpha^{NN,aa} & \alpha^{NN,bb} & \cdots & \alpha^{NN,NN} \end{pmatrix} \end{array},$$

and respectively for $\mathbf{B}$. In the most general case each heteroskedastic covariance process depends on a constant, all other pairwise cross–products of residual innovations and all lagged covariance processes.

Parsimony can be introduced by restricting the coefficient matrix, previously denoted $\mathbf{A}$ ($\mathbf{B}$), to be diagonal matrices D^A and D^B,

$$\mathbf{D^A} = \mathbf{I} \odot A \text{ and } \mathbf{D^B} = \mathbf{I} \odot B,$$

where $\mathbf{I}$ is the identity matrix, A (B) is the $N(N+1)/2 \times 1$ vector of ARCH (GARCH) coefficents and $\odot$ the Hadamard (element by element) product operator such that diagonal GARCH can be written as

$$vech(\mathbf{\Sigma_t}) = vech(\mathbf{\Omega}) + \mathbf{D^A} vech(E_{t-1}E_{t-1}') + \mathbf{D^B} vech(\mathbf{\Sigma_{t-1}}),$$

which imposes a simple GARCH(1,1) structure for the elements, here discrete processes

$h_t^{nm} = h_t^{aa}, h_t^{ab}, h_t^{bb}, \ldots, h_t^{NN}$, in the covariance matrix,

$$h_t^{nm} = \omega^{nm} + \delta^{\alpha^{nm}} \varepsilon_{t-1}^n \varepsilon_{t-1}^m + \delta^{\beta^{nm}} h_{t-1}^{nm},$$

with the only difference to univariate estimation lying in the multivariate joint likelihood optimization.

The multivariate approach comes at the cost of having to impose restrictions on parameters to ensure positive definiteness of the covariance matrix. BEKK–type GARCH models as postulated by Engle and Kroner (1995) build on the insight that the sum of a positive semi–definite matrix and a positive definite matrix is positive definite. In matrix notation,

$$\mathbf{\Sigma_t} = \mathbf{L}(\mathbf{\Omega})\mathbf{L}(\mathbf{\Omega}') + \mathbf{A}E_{t-1}E_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{\Sigma_{t-1}}\mathbf{B}',$$

where the L operator denotes the lower triangular matrix with $\mathbf{\Omega}$, $\mathbf{A}$ and $\mathbf{B}$ being $N \times N$ parameter matrices. The dynamics of the model quickly become intractable for the purpose it has originally been designed for, i.e. in vast multivariate dimensions. Consider, for example, the conditional variance of asset a in a bivariate BEKK which evolves as

$$\begin{aligned} h_t^{aa} \;=\; & (\omega^{aa})^2 \\ & + (\alpha^{aa})^2 \quad (\varepsilon_{t-1}^a)^2 \\ & + 2\alpha^{aa}\alpha^{ab} \quad (\varepsilon_{t-1}^a \varepsilon_{t-1}^b) \\ & + (\alpha^{ab})^2 \quad (\varepsilon_{t-1}^b)^2 \\ & + (\beta^{aa})^2 \quad h_{t-1}^{aa} \\ & + 2\beta^{aa}\beta^{ab} \quad h_{t-1}^{ab} \\ & + (\beta^{ab})^2 \quad h_{t-1}^{bb} \end{aligned} .$$

Consequently, BEKK–type GARCH models can similarly be constrained to be diagonal,

$$\mathbf{\Sigma_t} = \mathbf{L}(\mathbf{\Omega})\mathbf{L}(\mathbf{\Omega}') + \mathbf{D^A}E_{t-1}E_{t-1}'\mathbf{D^{A'}} + \mathbf{D^B}\mathbf{\Sigma_{t-1}}\mathbf{D^{B'}},$$

then

$$h_t^{nm} = \omega^{nm} + \alpha^n \alpha^m \varepsilon_{t-1}^n \varepsilon_{t-1}^m + \beta^n \beta^m h_{t-1}^{nm},$$

which differs in the construction to diagonal GARCH such that conditional covariances do not maintain their own parameters but are determined by those of their respective components directly—which is a testable assumption. But the main drawback lies in the fact that often convergence is not feasible once an asymmetric effect is introduced.

Diagonal GARCH and BEKK GARCH can both be further restricted to be scalar in their parameter matrices A and B, i.e. all processes share one common α and β, i.e. it is same across the diagonal, such that vast dimensional cases can be addressed; especially under application of covariance targeting.

Nevertheless, multivariate ARCH–type models remain notoriously complicated and difficult to estimate for the covariance matrix to remain positive semi–definite when more than a dozen assets are consider.[12]

Covariance Models: The simplest suitable benchmark model for the covariance matrix is the k–period moving average (unconditional) covariance estimator,

$$\mathbf{\Sigma_t} = k^{-1} \sum_{i=1}^{k} E_{t-i} E_{t-i}^{'},$$

for arbitrary k.

The equally–weighted moving average (EWMA) covariance estimator puts more weight on recent observations,

$$\mathbf{\Sigma_t} = (1-\lambda)(E_{t-1}E_{t-1}^{'}) + \lambda \mathbf{\Sigma_{t-1}} = (1-\lambda) \sum_{i=1}^{\infty} \lambda^{i-1} (E_{t-i}E_{t-i}^{'}),$$

with $0 \leqslant \lambda \leqslant 1$ and where $\Sigma_{t=1}$ can be set equal to the unconditional covariance estimate. Riskmetrics (1994) recommends $\lambda = 0.94$ (0.97) for daily (monthly) data.

To model hyperbolic rather than exponential decay in the autocorrelation structure of the covariances, RiskMetrics (2006) advocates to compute the covariance estimator as the average of k_{max} recent EWMA covariances over increasing time horizons where the weights w_k decay logarithmically such that the lagged autocorrelation of absolute residuals also decays logarithmically rather than exponentially, i.e. with more weight on recent and far distant data and consequently less weight therein between,

[12] Multivariate stochastic volatility models are even more complicated with Fiorentini et al. (2004), which is based on the discrete–time model by Diebold and Nerlove (1989), probably being the only common model.

$$
\begin{aligned}
\mathbf{\Sigma_t^{eff.}} &= \sum_{k=1}^{k_{max}} w_k \mathbf{\Sigma_{k,t}}, \\
\mathbf{\Sigma_{k,t}} &= (1-\lambda_k) E_{t-1} E_{t-1}^{'} + \lambda_k \mathbf{\Sigma_{k,t-1}}, \\
w_k &= \frac{1}{C}\Big(1 - \frac{log(\tau_k)}{log(\tau_0)}\Big), \\
\lambda_k &= exp\Big(-\frac{1}{\tau_k}\Big), \\
\tau_k &= \tau_1 \rho^{k-1}, \qquad k = 1, 2, \ldots, k_{max},
\end{aligned}
$$

where Zumbach (2007) recommends $\tau_0 = 1560 = 6$ years for the logarhithmic decay factor, $\tau_1 = 4$ days, $\rho = \sqrt{2}$, and C a normalisation constant such that the weights sum to unity.

Basic Setup for Modelling Conditional Correlation: Residuals standardised only in terms of their (own) variance, i.e. 'auto–standardised' or 'devolatised' residuals, are defined as

$$
z_t^n \equiv \frac{\varepsilon_t^n}{\sqrt{\mathrm{E}_{t-1}[\varepsilon_t^n \varepsilon_t^n]}} \equiv \frac{\varepsilon_t^n}{\sqrt{h_t^{nn}}} \qquad n = a, b, \ldots, N \qquad t = 1, 2, \ldots,
$$

or equivalently,

$$
Z_t = E_t \oslash H_t \qquad t = 1, 2, \ldots,
$$

where $H_t = diag(\mathbf{\Sigma_t})^{1/2}$ is an $N \times 1$ vector of conditional volatilities and $\oslash$ is the (element by element) Hadamard division operator.[13] Now defining ζ_t as an $N \times 1$ vector of multivariate standaridised residuals, in standard matrix notation,

$$
\zeta_t = (\mathbf{\Sigma_t})^{-\mathbf{1/2}} E_t \qquad t = 1, 2, \ldots,
$$

implies that ζ_t is both 'devolatised' and 'decorrelated' such that the conditional expectation of its vector cross product equals the identity matrix with size of the number of assets N,

$$
\mathrm{E}_{t-1}(\zeta_t \zeta_t^{'}) = \mathbf{I_N},
$$

due to unit expectations on the diagonal and the all cross–correlations having been accounted for such that the off–diagonals must all be zero. Put differently, the conditional correlation matrix, denoted by capital ρ at time t, i.e. $\mathbf{P_t}$, can be explicitly parametrised as the (conditional) covariance matrix which is standardised by the matrix of its respec-

[13] e.g. the dot division operator in Matlab

tive (conditional) cross–volatility products such that only an estimator of each pairwise correlation remains on the off–diagonals with ones on the diagonal,

$$\mathrm{E}_{t-1}(Z_t Z_t') \equiv \mathbf{P_t} = \mathbf{\Sigma_t} \oslash (H_t H_t'),$$

or equivalently in matrix notation,

$$\mathbf{P_t} = \mathbf{D_t^{-1}} \mathbf{\Sigma_t} \mathbf{D_t^{-1}}$$

and

$$\mathbf{\Sigma_t} = \mathbf{D_t} \mathbf{P_t} \mathbf{D_t}$$

where $\mathbf{D_t}$ is the corresponding matrix to H_t with conditional volatilities on the diagonal and zeros elsewhere, here exemplarily,

$$\mathbf{D_t} = \begin{pmatrix} \sqrt{h_t^{aa}} & 0 & \cdots & 0 \\ 0 & \sqrt{h_t^{bb}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{h_t^{NN}} \end{pmatrix}.$$

Correlation Models: Equicorrelation models are restricted covariance estimators which superimpose one common correlation, denoted ρ_t, onto all pairwise asset combinations. Then, the covariance matrix is defined as

$$\mathbf{\Sigma_t} = \begin{pmatrix} h_t^{aa} & \rho_t \sqrt{h_t^{aa}} \sqrt{h_t^{bb}} & \cdots & \rho_t \sqrt{h_t^{aa}} \sqrt{h_t^{NN}} \\ \rho_t \sqrt{h_t^{bb}} \sqrt{h_t^{aa}} & h_t^{bb} & \cdots & \rho_t \sqrt{h_t^{bb}} \sqrt{h_t^{NN}} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_t \sqrt{h_t^{NN}} \sqrt{h_t^{aa}} & \rho_t \sqrt{h_t^{NN}} \sqrt{h_t^{bb}} & \cdots & h_t^{NN} \end{pmatrix}$$

In the simplest case, the heteroskedastic variance process of the n'th asset can be replaced by a k–period moving average variance estimator, i.e. the diagonal of Eq. (),

$$Var_t^n = k^{-1} \sum_{i=1}^{k} (\varepsilon_{t-i}^n)^2 \qquad n = a, b, \ldots, N.$$

Let $\{e_{\Pi,t}\}$ denote the return process of an equally weighted portfolio, then the variance thereof is given by

$$\mathrm{E}_t[(e_t^{\Pi})^2] = N^{-2}\sum_{n=1}^{N} Var_t^n + 2N^{-2}\rho_t \sum_{n=1}^{N}\sum_{m=n+1}^{N} \sqrt{Var_t^n}\sqrt{Var_t^m}$$

where the equicorrelation ρ_t was taken in front of the double sum as it is by definition common to all covariances. In a moment–based estimation the equicorrelation sample estimate can be computed as

$$\hat{\rho}_t(k) = \frac{\widehat{Var}_t^{\Pi} - N^{-2}\sum_{n=1}^{N}\widehat{Var}_t^n}{2N^{-2}\rho_t \sum_{n=1}^{N}\sum_{m=n+1}^{N}\sqrt{\widehat{Var}_t^n}\sqrt{\widehat{Var}_t^m}}.$$

Alternatively, the conditional equicorrelation can be obtained by maximising the multivariate, here Gaussian, likelihood of the 'devolatised' or 'autostandardised' or but not 'decorrelated' residuals Z, then

$$L(\rho_t; Z) \quad = \quad -\frac{1}{2}\sum_{i=1}^{T} Nlog(2\pi) + log(|\mathbf{P_t}|) + Z'_{t-i}\mathbf{P_t^{-1}}Z_{t-i}$$

where the second term is the log of the determinant of the correlation matrix and the last term corresponds to the sum of the multivariate standardised and thus decorrelated residuals, previously denoted with ζ, squared. The use of 'autostandardised' residuals leaves much flexibility to model the variance of the each process. Note that P_t and not P_{t-i} is used in the maximum likelihood estimator, the properties of $\{\rho_t\}$ over time can be exploited by choosing a rolling time window of observations with e.g. 252 days. There is then a trade–off between significance and adjustment speed of the estimator.

Constant and Dynamic Conditional Correlation Models: Building upon the insight in the previous section, the Bollerslev (1990) CCC GARCH decomposes the conditional covariances into conditional variances and constant conditional correlations. Then we have

$$\mathbf{\Sigma_t} = \mathbf{D_t}\ \mathbf{P}\ \mathbf{D_t},$$

where P is now time invariant. Based on a two step estimation procedure, in step one almost any specification can be used to model the conditional variances without being restricted to using the same functional form and then in a second step obtain pairwise correlation estimates based on standardised residuals. Opposed to the covariance matrix in equicorrelation model, correlations are now allowed to differ across assets and similarly

a rolling window methodology can be applied.

DCC GARCH revokes the restriction of a constant correlation matrix P and allows for earlier introduced scalar dynamics to be imposed onto the conditional correlation matrix. Recall,

$$\mathbf{\Sigma_t} = \mathbf{D_t}\ \mathbf{P_t}\ \mathbf{D_t},$$

and elaborate

$$\mathbf{P_t} = \mathbf{Q_t}\ \oslash\ \mathbf{Q_t^*},$$

where $\mathbf{Q_t}$ follows a scalar diagonal (or equivalently scalar BEKK) GARCH structure,

$$\mathbf{Q_t} = (1 - \alpha - \beta)\overline{P} + \alpha(Z_{t-1}Z_{t-1}^{'}) + b\mathbf{Q_{t-1}},$$

where correlation targeting can be applied by using a consistent estimate of the long–run average correlation, and where

$$\mathbf{Q_t^*} = \begin{pmatrix} q_t^{aa} & \sqrt{q_t^{aa}q_t^{bb}} & \cdots & \sqrt{q_t^{aa}q_t^{NN}} \\ \sqrt{q_t^{aa}q_t^{bb}} & q_t^{bb} & \cdots & \sqrt{q_t^{bb}q_t^{NN}} \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{q_t^{NN}q_t^{aa}} & \sqrt{q_t^{NN}q_t^{aa}} & \cdots & q_t^{NN} \end{pmatrix},$$

which is a trick that forces the diagonals in the correlation matrix to unity and rescales all other entries accordingly. The key insight in the Engle (2002) and Tse and Tsui (2002) DCC model is that the estimation procedure can be separated into fitting a series of univariate ARCH–type volatilities from which the dynamic conditional correlation (DCC) series is estimated allowing each asset n to maintain its own conditional variance persistency with any other asset.

Economic Explanations of Asymmetric Volatility and Skewness

Asymmetric volatility effects have economically been motivated by primarily two theories. One, the leverage effect (Black, 1976; Christie, 1982) is related to the gearing of a company. When the market value of equity decreases, its volatility must increase under the assumption that total firm volatility stays *constant* which is the underlying assumption in Christie (1982). Two, not news related to or resulting in instantaneous return shocks but news affecting the perceived level of future volatility will trigger a sell–off until the expected return–risk ratio has been re–established. However, this demand by risk–averse

investors induces itself a rise in volatility, termed volatility feedback. Thus, markets can start declining due to an anticipated expected rise in volatility, negative innovations (either self–excited or news–induced) then increase this effect (Campbell and Hentschel, 1992). Duffee (1995) used a wide sample of US firms to show that the degree of volatility asymmetry is not mirrored by firm size, capital structure or leverage (Bekaert and Wu, 2000). Wu (2001) extends the setting and allows for dividend innovations regarding news and growth which are important factors in tranquil periods while the volatility feedback dominates in high volatility periods.

Endogenous Skew in Asset Pricing Other empirical studies testing the ability of skewness (or related measures) to predict cross–sectional variation in stock returns have produced mixed results. Xing, Zhang, and Zhao (2010), Cremers and Weinbaum (2010), and Rehman and Vilkov (2010) find a theoretically contradictory positive relation between skewness and future returns, while Bali, Cakici, and Whitelaw (2011) and Conrad, Dittmar, and Ghysels (2009) find a theoretically consistent negative relation. Boyer, Mitton, and Vorkink (2010) demonstrate that historical based estimates of skewness provide poor forecasts of future skewness. Albuquerque (2012) builds a stationary asset pricing model of firm announcement events where firm returns display positive skewness and then shows that cross–sectional heterogeneity in firm announcement events can lead to conditional asymmetric stock return correlations and negative skewness in aggregate returns.

Amaya, Christoffersen, Jacobs and Vasquez (2013) show how realized skewness predict the cross–section of equity returns. Jha and Kalimipalli (2010) and Ait-Sahalia, Wang, and Yared (2001) develop higher order risk–neutral moment strategies based on deviations between implied moments and forecasts for the realised moments in the former case and those extracted from the time series of the underlying asset in a risk-neutral world in the letter case. A number of recent papers furthermore shows that option-implied moments are very useful and successful in the prediction of future returns as well as factor exposures, see, e.g., Christoffersen, Jacobs, and Vainberg (2009), Duan and Wei (2009), Han (2008), and Rehman and Vilkov (2009). These authors show that on the basis of quantities like option-implied skewness and, ultimately, option–implied betas, investment strategies with very favorable risk-return trade–offs can be constructed. However, their construction of option–implied beta estimates requires very strict assumptions which have been heavily

criticised. The handbook chapter by Christoffersen, Jacobs, Young (2012) goes beyond the univariate option implied density and also consider results on option–implied covariance, correlation and beta forecasting, as well as the use of option-implied information in cross-sectional forecasting of equity returns. It also involves a discussion of how option–implied information can be adjusted for risk premia to remove biases in forecasting regressions.

Demand–based Option Pricing: At index level and under the plausible assumption that aggregate net exposure to the market and to changes therein is positive, investors are willing to pay for protection against large downward movements in the market up to medium–term horizons. A hedge against long–term downside exposure using options is expensive and would eliminate much of the equity risk premium. Accordingly, there should be low demand for deep–out–of–the–money put options with long time to expiry. As addressed earlier, the risk–neutral density is observed only truncated as there might be more attractive investing or hedging instruments at long horizons. At stock level, rationalising aggregate net exposure to changes in the stock price of up to medium–term horizons is less straightforward taking primary and derivative security markets into account.

Garleanu, Pedersen, and Poteshman (2009) show theoretically how demand pressure increases an option's price by an amount proportional to the variance (covariance) of the unhedgeable part of the option (two options). Empirically, they show how demand helps explain the overall expensiveness and skew patterns of index options and single-stock options in the cross–section. Also Carr, Geman, Madan, and Yor (2003) hypothesize that large open interest in out-of-the-money puts are a possible source of the negative skewness observed in option-implied distributions. For their CGMY model they conjecture that the structure of open interest in the market is an important determinant of the shape of market risk premia as reected in the constructing of their measure change function.

Aggregation Properties: Individual and Market Skew At long horizons default risk generates a steep skew in the cross–section of options, evidenced by high correlation with CDS spreads based on the unit recovery claim methodology introduced by Carr and Wu (2010). New information is first reflective in the CDS market due to higher liquidity compared to long–term options with large spreads (Giamouridis, Angelopoulos and Nikolakakis 2013). This potential mechanism of index implied volatility shapes might not be so interesting when default risk is assumed to be diversifiable, deteriorating stocks

leave the index prior to bankruptcy or a too big to fail argument holds for e.g. the Dow Jones 30 constituents.

BKM have shown that the risk–neutral index distribution is always more negatively skewed than those for constituent stocks which they attribute to risk–averse investor and physical fat–tailed not necessarily skewed distributions (BKM 2003, BM 2006) over horizons of up to 120 calendar days.

Determinants of RNS While BKM conclude that variations in the risk–neutral skew are instrumental in the differential pricing of individual equity option, Taylor et al. (2009) pin down that cross–sectionally RNS varies negatively with firm characteristics, market volatility, positively with market skewness with most of the variation in individual RNS explained by firm–specific rather than market–wide factors.[14] Surprisingly, in Taylor et al. volatility asymmetry is not relevant for RNS. In a formal setting, BKM demonstrate that leverage effect imparts a negative skew to individual stocks and aggregate index but predictions for the skew magnitudes are at odds with their theorem (3), i.e. index skew will be less negative compared to individual stocks. Probably the leverage effect for individual stocks is not as pronounced in the real–world as assumed in their model and the higher index skewness is stemming from other factors: demand–hedging–related option pricing which might well be related to P–measure fat–tails and risk–aversion. It would thus be in line with Taylor 's finding that greater option market liquidity (not underlying security liquidity is associated with more negative RNS.

In a P– to Q–measure comparison, physically detected jumps seem to be mean zero and even jump tails, which avoid the peso–type problems of detecting jumps, appear to be symmetric for index and stocks (Bollerslev Todorov 2011), whereas only large negative jumps with a high arrival rate could generate large skewness at short horizons.

Nevertheless, in a Q–measure only setting, judging from only slightly negative sloping individual stock implied volatility curves and thus almost symmetric, though fat–tailed risk–neutral distribution such jump features of the index dynamics are questionable.

Revisiting RNS and Asymmetric Correlations Instead, the following point, also related to demand and hedging arguments which are neglected in a univariate framework, focusses on asymmetric correlations. In light of higher individual implied volatilities than

[14] Other findings are that higher overall volatility comes along with more negative skewness for stocks and index.

at index level any sort of downside risk is less manageable to be diversiable (hedgable). At short to medium term horizon we might equally well disentangle the proportion of skewness of the risk–neutral distribution stemming from jumps and the unexplained proportion stemming from asymmetric correlations among constituents return series which might mimic the features of downward jumps at index level when option prices are observed at daily frequency only.[15] Consequently, at index level we might want to add a third generating mechanism determining the shape of index implied volatility curves and thus the moments of its risk–neutral density: Implied (heterogeneous or asymmetric) correlation between constituent stock return series.

Fengler, Herwartz, and Werner (2012) suggest to go the other direction and try to recover index implied volatility from option prices computed from discrete time risk–neutralised copula asymmetric DCC multivariate process which are estimated under the P measure. They corroborate the common jump hypothesis of Branger and Schlag (2004) as their model contributes to but explains only partly implied index skew. Suggestions are to go beyond the exponentially affine pricing kernel and employ the more sophisticated specifications as suggested in Christoffersen et al. (2011) and Monfort and Pegoraro (2011). Second, given the importance of dependence beyond correlation, one could employ alternative copula models which strictly separate the marginal volatility processes and the copula, as suggested by Granger et al. (2006), Patton (2006), Kim et al. (2008), Fantazzini (2009), Jin (2009), Christoffersen, Errunza, Jacobs and Langlois (2011). Unfortunately, the risk–neutral asset price process is unobservable, only option prices (implied volatilities) are observable. Heterogeneous or asymmetric implied correlations might not only be relevant among asset price return series but also among their volatility processes which determines option prices. Nevertheless this would not explain why constituent stocks implied volatility curves are relatively flat.

[15] One might want to argue that instead of asymmetric correlation among all stock return series, just higher correlation between certain stocks (heterogeneous correlations) might be sufficient or indistinguishable from asymmetric correlation.

A review of the classical skewness estimator, estimates, and a robust alternvative(?)

Daily Skewness and its Quantile–based Variations

Point estimates of skewness are very sensitive to large returns in magnitude of either sign, which are amplified due to the third power function in the numerator of the standard measure of skewness, while, on the other side, the impact of small returns seems to be almost diminishing. The former apparent stylised fact can be inferred from Fig. 1 when large returns leave the one year rolling observation window. There also appears to be a common factor driving the skewness of all three representative market proxies.[16] In the following, we should be able to concentrate on inference results based on the S&P 500 only without much loss of generality. All returns are excess returns over the one–month T-bill rate.

In order to analyse daily measures of skewness more precisely while at the same time knowingly sacrificing some time–variation, we increase the length of the observation window from one to ten years of daily returns, rolled forward on a monthly basis, and bootstrap each time series to create confidence intervals.

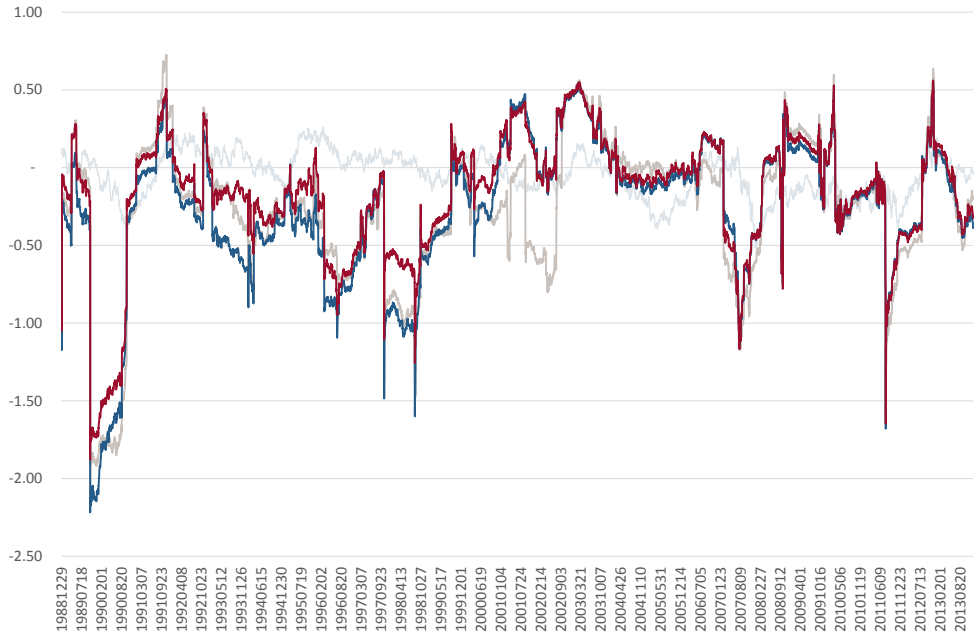
Then the average estimate of skewness is -0.076 over 193 overlapping samples, up from previously -0.176 (-0.154 for non–overlapping samples) for daily skewness based on yearly observation windows in Fig. 1. Skewness is a measure around the first moment: Here, it is negatively correlated with the sample mean (-0.262), non–centralised skewness estimates are slightly positively correlated (0.126).

In their Monte Carlo study Kim and White (2004) criticise the robustness and thus usefulness of conventional measures of skewness and suggest looking beyond the standard textbook formula of skewness to gain further insight into market dynamics. Robust alternative measures of skewness based on deviations of all return pair combinations from the median have been suggested by Brys, Hubert, and Struyf (2004), which we will review and put to the test in the next section. Another way is to truncate the distribution and model their conditional multi–quantiles jointly as in Kim, Manganelli, and White (2008). Ghysels, Plazzi, and Valkanov (2012) find that a quantile–based measure of conditional

[16] The figure also reveals that estimates of the standard textbook measure of skewness based on daily S&P 500 close–close returns seems to be a good proxy for the skewness of the broader Fama–French definition of the market return (correlation of 0.93), and it is in line with the Dow Jones Industrial Average (correlation of 0.97).

Figure 1: Skewness estimates of daily US indices and aggregate market returns

X–AXIS: Denotes the end of the one year backward–looking return periods, rolled forward daily over data spanning Jan. 1988 to Dec. 2013. RED: S&P 500. BLUE: DJIA. GREY: Fama–French definition of the market return. LIGHT BLUE: Interquartile skewness estimates of the S&P 500 index..



skewness (or asymmetry) of returns also varies significantly over time and demonstrate its economic significance in an international portfolio setting.

So–called 'robust' interquartile skewness seems to be significantly negative from the ten year period ending in Feb. 2008 onwards, see Fig. 3. When return innovations are small but the unconditional mean return is large, skewness can become more negative versus the non–centralised skewness measure. Nevertheless, correlation between centralised skewness and the first moment of the interquartile range is 0.556 (0.427 in first differences) computed over all overlapping sample estimates. Over shorter rolling observations windows, also interquartile–based skewness is very erratic as can be inferred from the lightblue line in Fig. 1.

The tails of the return distributions in Fig. 3 are mean negative but positively skewed in recent periods and vice versa in early periods, although the bootstrapped standard error bands under the null of no skewness suggest no significance. The return period ending Oct. 2002 throughout to Oct. 2007 are virtually flat with the Oct. 2008 Lehman bankruptcy pushing up the confidence intervals whose shapes resemble those seen in the previous figure. Centralised skewness to first moment correlations are -0.723 (0.196 in first

Figure 2: Skewness estimates of daily S&P 500 close–to–close returns

X–AXIS: Denotes the end of the ten year backward–looking return periods, rolled forward monthly over data spanning Jan. 1988 to Dec. 2013. RED: Skewness estimates. RED DOTTED: Non–centralised skewness estimates. GREY: Two standard deviation confidence intervals obtained by bootstrapping each original time series 10,000 times letting the sign of returns follow a Bernoulli random variable with probability half, i.e. we randomly flip the sign via the two point distribution, which preserves other features such as volatility clustering etc. BLUE: Mean per annum.

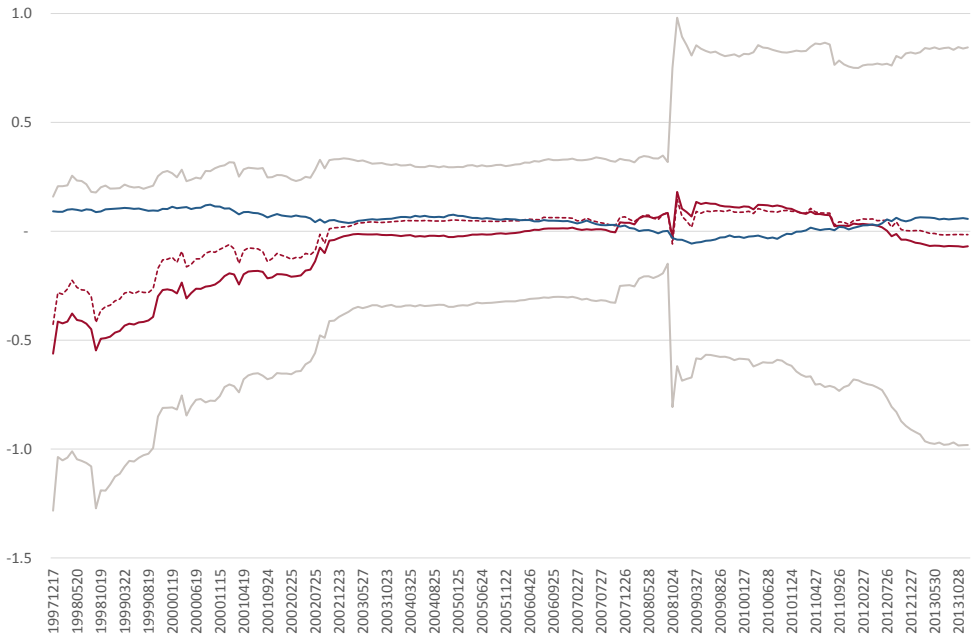
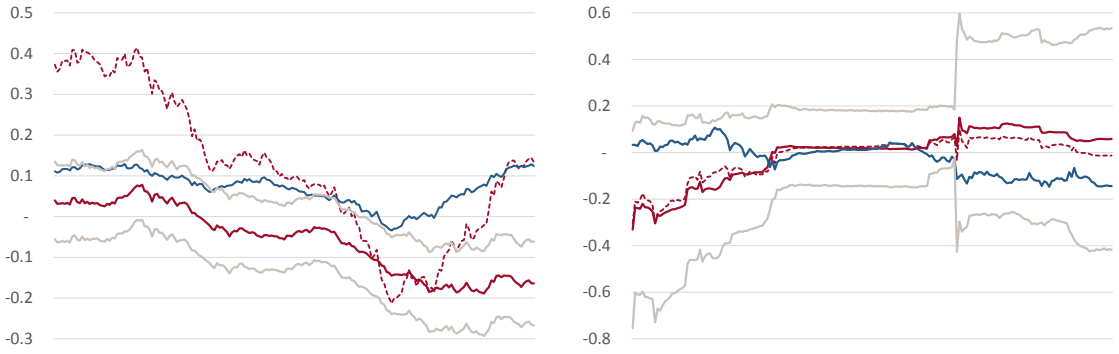


Figure 3: Interquartile and corresponding tail skewness estimates

Interquartile (LHS) and corresponding tail (RHS), joint 1st and 4th quartile, skewness estimates based on monthly rolling forward ten year empirical S&P 500 daily return distribution corresponding to Fig. 2. The mean (BLUE) refers to the mean of the respective empirical quantiles only. Legend as in Fig. 2.



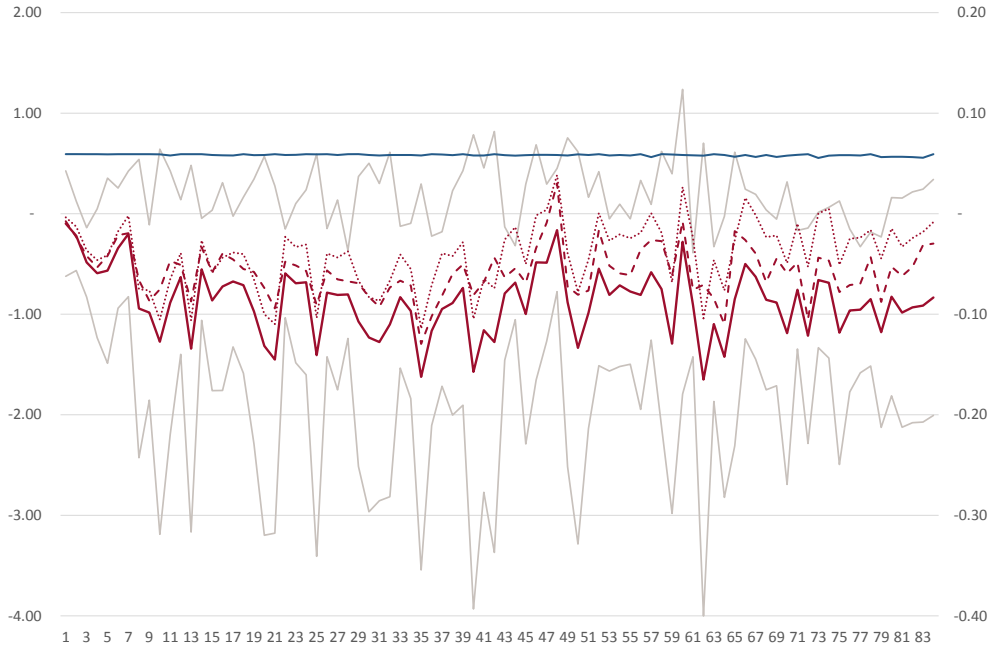
differences).

Skewness as a Function of the Return Horizon

The previous figures, particular Fig. 2, put the unconditional negative skewness of daily index returns into question. Fig. 4 shows skewness as a function of the return horizon ranging from one day to four months (84 trading days) with bootstrapped two standard deviations confidence intervals based on previously used daily S&P 500 returns from 1988–2013. The null hypothesis of zero skewness can be rejected more often than the expected type II error rate.

Figure 4: Skewness estimates over extending return horizons up to four months

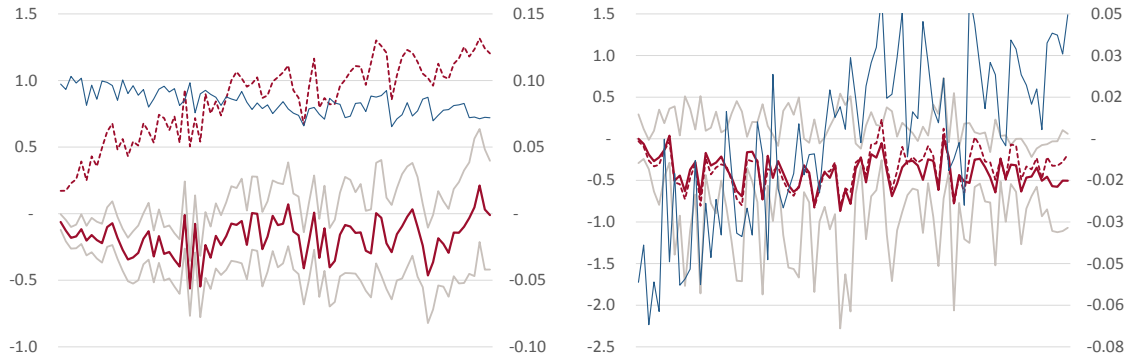
X–AXIS: Denotes the length of the return horizon with one denoting one trading day up to 84 trading days (four months). RED: Skewness estimates of entire S&P 500 time series from January 1988 to December 2013 corresponding to the respective return horizon. RED DOTTED: Non–centralised skewness estimate. RED DASHED: Average skewness estimate for each return horizon (one day to four months) based on one time unit rolling return periods of the respective frequency. GREY: ± 2 std. confidence intervals calculated by bootstrapping each aggregated return series eight times its length using the same methodology as before. When the bootstrapping is performed on the highest (here daily) frequency before the return series are aggregated, the standard deviations are on average 2.1 times higher. BLUE: Mean p.a. of 5.8% on secondary secondary axis.



Also, there seems to be significant interquartile skewness of up to at least one month return horizons, Fig. 4. Overall, skewness is more pronounced when all observations are considered versus its tail skewness measure in Fig. 5. It is interesting to note that the tails are mean negative at short horizon but positive with although insignificant negative skewness at long horizons. When bootstrapping, the obtained mean tail skewness estimates

Figure 5: Interquartile & tail skewness estimates corresponding to Fig. 4

Legend see Fig. 1. Standard deviations are on average 1.1 (2.4) times higher for interquartile (tail) estimates when bootstrapping is performed before return series are aggregated.



as a function of the horizon decreases linearly from almost zero at one day to -0.15 at four month horizon which might be due to the fact that the ratio of negative to positive returns of the demeaned tail series decreases to approx. 45% at four month horizon.

Skewness of Large Cap Stocks at Monthly Return Horizon

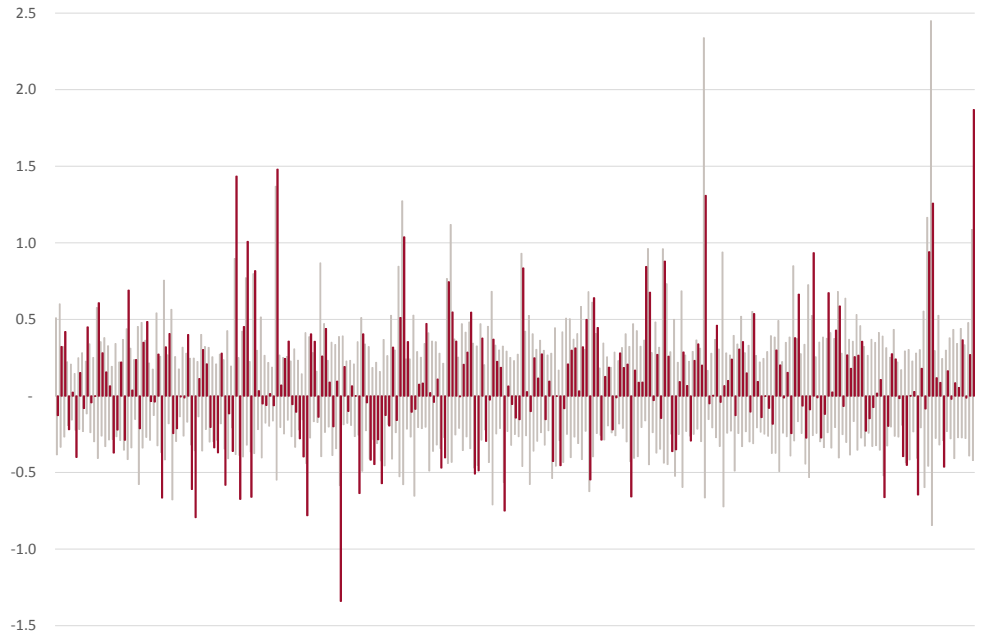
In a theoretical context of a stationary asset pricing model, Albuquerque (2012) reconciles negatively skewed index returns with positively skewed stock returns via heterogeneity in firm announcement events which can lead to conditional asymmetric stock return correlations and further to negative skewness in aggregate returns.

In Fig. 6, we examine the average skewness of monthly returns over rolling ten year periods of all 247 stocks which were constituents of the S&P 500 throughout the 1988 to 2013 period. we examine only surviving firms, which is indeed the case in the theoretical analysis of Albuquerque (2012) as there appears to be no default probability.

Then, as before skewness is driven by the largest returns in magnitude. The point estimate of skewness across all series is 0.080 with s.e. 0.026 (0.184 with s.e. 0.052 when skewness is calculated once per series and not over rolling ten year periods of monthly returns). Repeating the analysis over rolling three year instead of ten year periods yields a point estimate of skewness across all series of only 0.039 with s.e. 0.014, not shown here for brevity. Bootstrapped confidence intervals are not reported as the returns are so large in magnitude that a null of zero skewness can clearly not be rejected.

Figure 6: Point estimates of monthly return horizon skewness of S&P 500 constituents

RED: Each bar represents the average skewness for one S&P 500 constituent stock, computed by averaging the skewness estimates obtained from monthly rolling forward over–lapping ten year periods. GREY: Each bar shows the maximum and minimum monthly return for each constituent throughout the 1988 to 2013 period.



Finite Sample Properties and Alternative (Robust) Skewness Measure

In the following, we are investigating the finite sample properties of the moment–based skewness estimator versus the medcouple, an alternative, more robust, measure of asymmetry.

The medcouple, denoted MC_n (Brys et al., 2004), is not based on the third and second moment as classical skewness but defined as a scaled median difference of the right and left half of a continuous univariate distribution. For n independently sampled and sorted observations $X_n = \{x_1, \ldots, x_n\}$ such that $x_1 \leq \ldots \leq x_n$ with median m_n of X_n, the medcouple is defined as

$$\mathrm{MC}_n = \underset{x_i \leq m_n \leq x_j}{\mathrm{med}} h(x_i, x_j), \tag{0.0.7}$$

where $\forall\ x_i \neq x_j$ the kernel function h applies,

$$h(x_i, x_j) = \frac{(x_j - m_n) - (m_n - x_i)}{x_j - x_i}.$$

For the important situation when $x_i = x_j = m_n$, as might easily be the case for discretely sampled financial returns when several observations are zero, denote the indices of those

observations by $m_1 < \dots m_k$; i.e. $x_{m_l} = m_n \; \forall \; l = 1, \dots k$. The kernel function is then

$$h(x_{m_i}, x_{m_j}) = \begin{cases} -1 & \text{if} \quad i+j-1 < k \\ 0 & \text{if} \quad i+j-1 = k \\ 1 & \text{if} \quad i+j-1 > k \end{cases} .$$

Thus, the kernel measures the (standardized) difference between the distances of x_j and x_i to the median, and due to the denominator MC_n lies between -1 and 1. It is location and scale invariant and, unlike classical moment–based skewness, can be computed for distribution without finite moments as it is based on ranks only.

However, the investor's interest might rather lie in the classical skewness measure, when say a risk premium earned is assumed to be linear in skewness or a distribution is to be estimated in terms of its moments rather than its medcouple, which belongs to the class of incomplete L–statistics.[17] Nevertheless, it might be possible to establish a finite sample link between the medcouple and the standard skewness estimator through Monte Carlo simulation. Therefore, we consider realization from the skewed generalized t distribution (Theodossiou, 1998) whose pdf is given by

$$f_{SGT}(x; \mu, \sigma, \lambda, p, q) = \frac{p}{2v\sigma q^{1/p} B\big(\frac{1}{p}, q\big)\big(\frac{|x-\mu+m|^p}{q(v\sigma)^p(\lambda sign(x-\mu+m)+1)^p} + 1\big)^{\frac{1}{p}+q}} \tag{0.0.8}$$

with restrictions $\sigma, p, q > 0, -1 < \lambda < 1$. Setting

$$\begin{aligned} m &= 2v\sigma\lambda q^{\frac{1}{p}} B(\tfrac{2}{p}, q - \tfrac{1}{p}) B(\tfrac{1}{p}, q)^{-1}, \\ v &= q^{-\frac{1}{p}} \Big((3\lambda^2 + 1) \tfrac{B(\frac{3}{p}, q - \frac{2}{p})}{B(\frac{1}{p}, q)} - 4\lambda^2 \tfrac{B(\frac{2}{p}, q - \frac{1}{p})^2}{B(\frac{1}{p}, q)^2} \Big)^{-1/2} \end{aligned}$$

makes μ and σ^2 the mean and variance of the SGT distribution with the h'th moment only being defined if $pq > h$, in which case

$$\begin{aligned} Skewness_{f_{GST}} \quad = \quad & \frac{2q^{3/p}\lambda(v\sigma)^3}{B(\frac{1}{p}, q)^3} \times \\ & \Big(8\lambda^2 B(\tfrac{2}{p}, q - \tfrac{1}{p})^3 - 3(1 + 3\lambda^2) B(\tfrac{1}{p}, q) B(\tfrac{2}{p}, q - \tfrac{1}{p}) B(\tfrac{3}{p}, q - \tfrac{2}{p}) \\ & + 2(1 + \lambda^2) B(\tfrac{1}{p}, q)^2 B(\tfrac{4}{p}, q - \tfrac{3}{p}) \Big). \end{aligned} \tag{0.0.9}$$

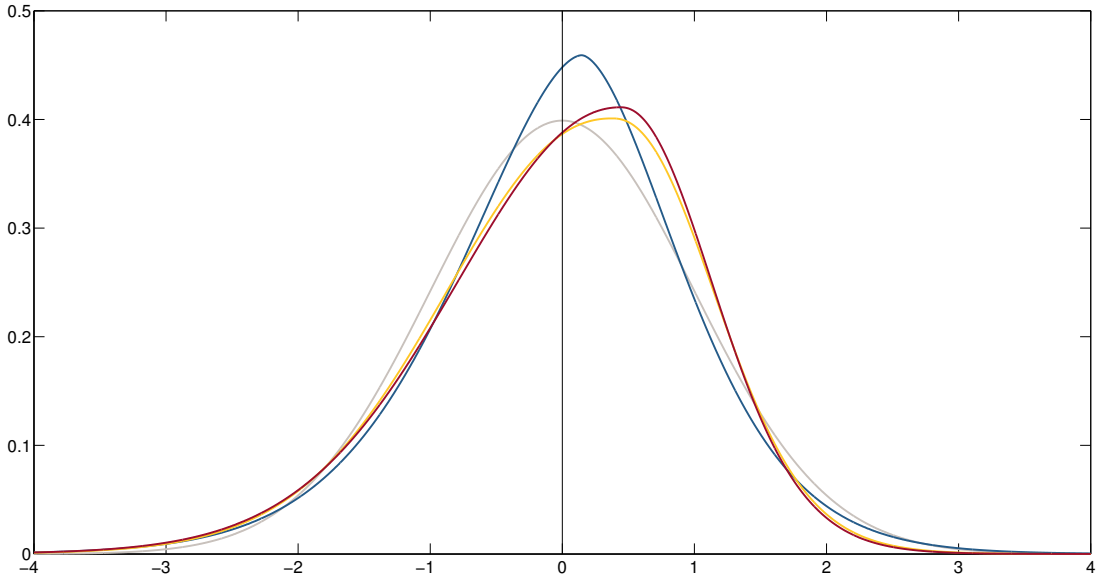
The SGT distribution has recently been applied in a parametric setting to model the residuals in various GARCH in mean specifications uncovering a more stable risk return link in intertemporal asset pricing models, see Theodossiou and Savva (2015). In Fig.

[17] Note MC_n is not applied to all possible pairwise combinations (x_i, x_j).

7, we plot their respective EGARCH–M residual densities for daily, weekly, and monthly market excess returns versus the standard normal distribution, and compute analytically their respective coefficients of skewness based on Eq. (0.0.9)—which are reported in the first row in Table 2.

Figure 7: Daily, weekly, monthly SGT EGARCH–M residual distributions

BLUE (daily): SGT($\mu = 0, \sigma = 1, \lambda = -0.095$ [s.e. 0.008], $p = 1.706, q = 11.285$). ORANGE (weekly): SGT($\mu = 0, \sigma = 1, \lambda = -0.242$ [s.e. 0.022], $p = 2.225, q = 8.038$). RED (monthly): SGT($\mu = 0, \sigma = 1, \lambda = -0.285$ [s.e. 0.048], $p = 2.111, q = 9.788$). Parameter estimates from Theodossiou and Savva (2015) when CRSP market excess returns are winsorized at plus/minus four standard deviations. GREY: Benchmark standard normal distribution, SGT($\mu = 0, \sigma = 1, \lambda = 0, p = 2, q \to \infty$).



In order to grasp the finite sample properties, we then simulate 100,000 times ten[18], five, two years, and three months of daily, weekly, and monthly return residuals to obtain the standard deviations of the empirical distributions of skewness estimates around their means based on their respective sample sizes of, e.g. in case of the ten year horizon 2,500 daily, 520 weekly, and 120 monthly draws. We look at GARCH residuals rather than raw return data directly as these are free of volatility clustering effects.

As noted in Hsieh (1989), the conditional distribution of the data is expected to be less skewed (and leptokurtic) than the unconditional distribution as can be inferred from Jensen's inequality. As the sample size decreases the downward bias becomes more severe and finite sample standard deviations are decreasing as skewness cannot build up from

[18]Thus, we refrain ourselves from using horizons longer than ten years after which the time series might not be stationary any more.

relatively few observations whenever it is not caused by outliers. While the simple $6/n$ approximation formulae[19] provides the finite sample variance of skewness under the hypothesis of zero skewness, the simulations provide the standard deviation over different time horizons and sampling frequencies under presence of the respective skewness.

When the data is weakly skewed to the left (negative) as in the case of daily data, the finite sample standard deviations are much larger compared to those when the underlying distribution is standard normal, by factors of 1.71, 1.69, 1.66, 1.41 over the four horizons. It is also clear that moment–based skewness is not a very efficient estimator but the medcouple, despites being more robust (not tested here), appears to be even less efficient.

In order to illustrate the large confidence intervals of the medcouple estimator in terms of the moment–based skewness measure, we consider Hansen's (1994) skewed t distribution, varying its moment–based skewness between minus one and zero in steps of 0.05.[20] Simulating from these 21 distributions provides consistent estimates of medcouples along with estimates of their standard deviations. Somewhat surprisingly, the latter seem to be invariant across the different degrees of asymmetries considered in finite sample (ten year of daily realizations, N=2,500), and there appears to be a linear relationship between the medcouple and moment–based skewness. Based on this linear relation, we can then infer the plus/minus two standard deviations confidence intervals of moment-based skewness from those of the medcouple; shown in the right/left graph in Fig. 8.

Preliminary Take–away

Based on the standard textbook definition of skewness, it is hard to obtain precise estimates with low standard errors both at short and long horizons. Large returns in magnitude drive up the standard deviations of the bootstrapped skewness distribution and even advocated tail–truncated interquartile daily skewness measures are not significantly negative throughout two thirds of the sample. There is some evidence although weak in significance that skewness becomes more negative as the return horizon increases.

[19] The more precise formulae for the variance of the skewness estimator used here is $\frac{6n(n-1)}{(n-2)(n+1)(n+3)}$.

[20] The skewed t distribution is a special limiting case of the SGT distribution with zero mean, unit variance, and fixed $p = 2$. The asymmetry of the distribution is defined over the plus/minus one range, see SGT distribution, so we first need to solved Eq. (0.0.9) for λ for given values of skewness and degrees of freedom q.

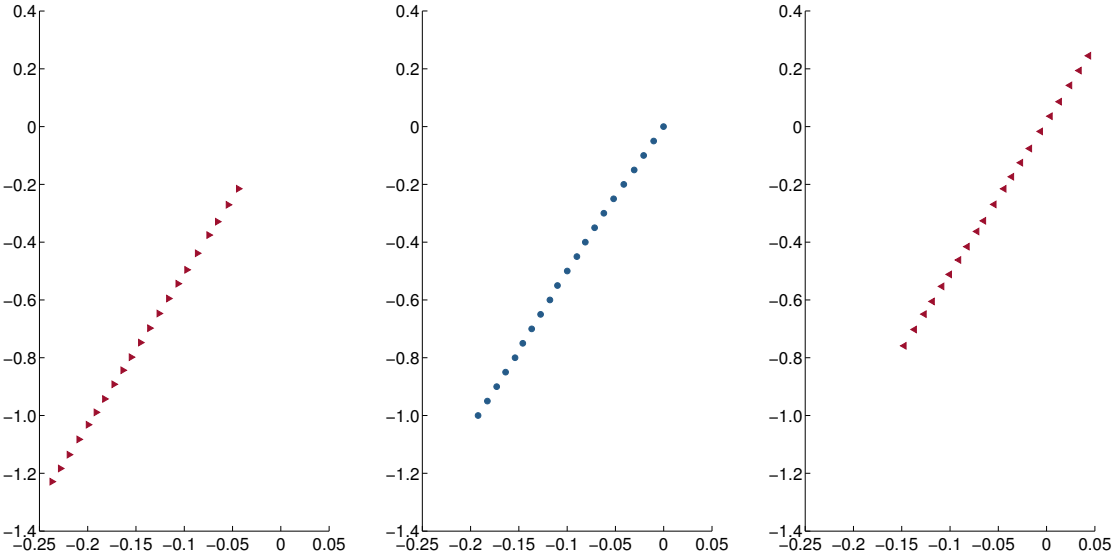
Table 2: Finite sample properties of moment–based skewness and medcouple

Finite sample properties of classical moment–based skewness (Sk) measure and medcouple (MC) under the three parameterizations of the SGT distribution as in Fig. 7. Reported are theoretical skewness, Sk SGT (theo.), of the distribution based on Eq. (0.0.9). The moment–based skewness means, Sk m., and standard deviations, Sk std., refer to the empirical distribution of skewness estimates from simulating 100,000 times over each of the four time horizons at the three sampling frequencies (daily, weekly, and monthly); e.g. for the ten year horizon under the daily sampling scheme we simulate 250 million realizations. Equivalently, we conduct the same experiment for the medcouple based on the same seed for which we obtain MC m. and MC std.. As benchmark we also report the finite sample standard deviation of skewness for the standard normal distribution, denoted Sk std. ND (theo.).

Skew stats.	**d'ly**	**wk'ly**	**m'ly**	**d'ly**	**wk'ly**	**m'ly**
Sk SGT (theo.)	-0.230	-0.403	-0.487	-0.230	-0.403	-0.487
10 yrs / 5 yrs	n=2,500	n=520	n=120	n=1,250	n=260	n=60
Sk m. SGT (sim.)	-0.230	-0.397	-0.456	-0.228	-0.390	-0.423
Sk std. SGT (sim.)	0.084	0.129	0.257	0.117	0.179	0.339
Sk std. ND (theo.)	0.049	0.107	0.221	0.069	0.151	0.309
MC m. SGT (sim.)	-0.045	-0.073	-0.089	-0.044	-0.073	-0.087
MC std. SGT (sim.)	0.022	0.050	0.101	0.032	0.070	0.140
2 yrs / 3 m	n=500	n=104	n=24	n=63	n=12	n=3
Sk m. SGT (sim.)	-0.225	-0.372	-0.349	-0.197	-0.210	-0.027
Sk std. SGT (sim.)	0.181	0.267	0.447	0.426	0.496	0.272
Sk std. ND (theo.)	0.109	0.237	0.472	0.302	0.637	1.225
MC std. SGT (sim.)	-0.044	-0.072	-0.083	-0.042	-0.060	-0.029
MC std. SGT (sim.)	0.049	0.108	0.207	0.140	0.259	0.277

Figure 8: Medcouple confidence intervals in terms of moment–based skewness

X–AXIS: Medcouple. Y–AXIS: Moment–based skewness. MIDDLE (BLUE): Empirical means of medcouple estimates for realizations of 21 SGT distributions with zero mean, unit variance, fixed $p = 2$, $q = 30$ degrees of freedom, and λs to yield moment–based skewness measures ranging from minus one to zero in steps of 0.05 based on re–arranging Eq. (0.0.9). 10,000 simulation loops with $n = 2,500$ for each of the 21 parameterizations. RIGHT/LEFT (RED): Upper/lower 95% confidence interval bounds based on the estimated standard deviation of medcouple estimates (via simulation), moment–based coefficients of skewness are obtained by plugging in the computed upper and lower medcouple confidence interval bounds into the estimated linear regression equation of theoretical skewness on the empirical means of the medcouples, i.e. through the blue data points, $\text{Skewness}_{f_{SST(\lambda)}} = 0.01697 + 5.261 \times \text{MC}_n(\lambda)$; R-square = 0.9993.



Chapter 1

Realized Skewness, Asymmetric Volatility, and Volatility Feedback

Co–Author: Stephen J. Taylor[1]

Abstract

Over multi–day and longer horizons aggregate stock market returns appear to display negative skewness, the propensity to generate large negative returns with greater probability than suggested by a symmetric distribution. So far the literature on realized skewness is sparse and has either focussed on the skewness of intraday returns or lent itself to option prices. We provide a measure for long horizon skewness by making use of the dependence structure inherent in intraday asset returns only. The measure shows the accumulation of skewness over time as it builds up from the asymmetry between return and variance at the intraday level. As estimates of spot variance are generally noisy, we show how more robust volatility proxies can better be used instead. In a theoretical model analysis we propose how one could impose stepwise more structure on the covariance terms between return and volatility proxy in order determine skewness at long horizons more precisely in the context of a very general two factor non–affine log linear stochastic volatility model and its reduced form specifications.

Keywords: model–free/semi–parametric realized skewness, realized power variations, asymmetric volatility/statistical leverage effect, discrete time multifactor stochastic volatility models

[1]Stephen Taylor is Professor of Finance at the Department of Accounting and Finance at The Mangement School, Lancaster University, United Kingdom.

1.1 Introduction: A New Estimator of Realized Skewness

A large and still growing body of literature focuses on estimating the second moment more precisely using high–frequency data. This has brought forward the 'realized' methodology with different estimators horse–racing between efficiency gains and microstructure noise robustness in computing the quadratic variation or integrated variance of a realized asset price path over any horizon, see e.g. Aït–Sahalia and Jacod (2012) for a review of different estimators.

The conventional realized variance set–up relies on aggregating selected positive quantities.[2] An analogy for the third moment does not follow intuitively, as aggregating third powers of returns does not provide useful limits as the (intraday) sampling frequency increases.

Consider instead the p'th N–period return moment when the N–period return is the sum of N high–frequency returns over observation window t,

$$\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^p\Big] = \mathrm{E}\Big[\Big(r_{t,1} + r_{t,2} + \ldots + r_{t,N}\Big)^p\Big],$$

Then, N–period skewness over the specified time frame can be defined as the ratio,

$$Skewness_t^{(N)} = \frac{\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^3\Big]}{\Big(\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^2\Big]\Big)^{3/2}},$$

where the denominator requires a suitable estimator of quadratic variation to the power three over two and the realized third moment can be estimated from high–frequency returns in the form

$$\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^3\Big] = \sum_i \mathrm{E}[r_{t,i}^3] + 3\sum_{i<j} \mathrm{E}[r_{t,i} r_{t,j}^2] \tag{1.1.1}$$

under the parsimonious assumptions

$$\underset{i<j<k}{\mathrm{E}}[r_{t,i} r_{t,j} r_{t,k}] = \underset{i<j}{\mathrm{E}}[r_{t,i}^2 r_{t,j}] = 0. \tag{1.1.2}$$

In particular, the second term on the right–hand side of Eq. (1.1.1) applies for many well–known asymmetric volatility models in both discrete time, such as the class of general

[2] i.e. integrated power functions of increasingly finer sampled innovations of the underlying asset price process of the form $\int \sigma_s^{2p} ds$, where σ_s can be thought of as spot volatility, $p = 1$ defines quadratic variation, and $p = 2$ defines quarticity used in hypothesis testing.

stochastic volatility models in Harvey and Shephard (1996) and its multi–factor extensions, and in continuous time such as the Heston (1993) model or, more generally the broad class of affine models falling under the Duffie et al. (2000) framework.

Under asymmetric volatility, negative skewness will build up over time from symmetric high–frequency returns. Then, skewness is a function of the horizon, first decreasing, then increasing towards zero due to some functional central limit theorem when the time horizon goes to infinity; see Das and Sundaram (1999), Taylor (2005, Chapter 11).

The cross–correlation structure between returns and volatility proxies at different lead orders satisfying the assumptions in Eq. (1.1.2) has previously been documented by Bollerslev, Litvinova, and Tauchen (2006) for 5–minute S&P 500 futures. Semi–parametric return volatility asymmetries (new impact curves) have been further revisited at high–frequency index level data by Chen and Ghysels (2011). Patton and Sheppard (2015) provide further insight on the different persistency of negative versus positive shocks onto future volatility for the S&P 500 index and constituent stocks. Nevertheless, work on asymmetric volatility at high frequencies (relative to the target frequency of interest) has not yet been extended to realized skewness in a model–free or semi–parametric way.

The research objective is to provide robust insight into the apparent prevailing stylized fact of negatively skewed equity index returns and positively skewed constituent stock returns, see Albuquerque (2012) and the many references therein. Similar to what realized variance does for the second moment, we expect to gain from efficiency, i.e. in terms of obtaining a more precise estimator, by using high–frequency intraday data for inference about the third moment, and thus in estimating realized skewness over longer horizons. While previously one would have to rely on long time series of data at one frequency to estimate a coefficient of skewness for that frequency, it is possible to document skewness more precisely by building on the aggregational properties of functions of the return series only. When the underlying asset price path is filtered at one frequency, such as the standard choice of 5–minutes for index data, and e.g. 11–minutes for liquid constituent stocks to avoid arising microstructure noise, it is possible to construct a completely model–free estimator. Going further, we suggest a framework under which it is possible to construct a more precise semi–parametric skewness estimator taking into account the information captured in the asset price path sampled at potentially multiple (different) intraday frequencies.

In the multi–frequency approach, we first look at high–frequency M_1, i.e. when the

intraday sampling frequency M is large, and then stepwise reduce M for which we hope the underlying model dynamics hold approximately. If the frequency is too high the results may be mainly driven by MMN not present at a lower frequency.[3] On the other hand, when the sampling frequency is too low we might average out some of those potentially very short lived effects occurring at high–frequency. Consequently, there is a small subset of M.

As per definition of the discrete time nature of our semi–parametric model, skewness for very short horizons (at one or two lags) will be different and depend on the sampling or, moreover, the generating frequency at which realizations of the process are being produced.[4] However, we are interested in skewness over *longer* horizons relative to the sampling/generating frequency such as the skewness of monthly returns relying on finely sampled 5–minute returns, and we show in our analysis that it does not make a difference between the asymmetric effect being generated at any of the most frequently encountered (different) values of M.

1.2 Literature Review: Skewness, Realized Skewness, and the Statistical Leverage Effect

The standard textbook formula of skewness measures asymmetry of the generating process at one sampling frequency. It does not relate the dynamics of functions of high–frequency returns to the properties of low–frequency returns over various time horizons by addressing asymmetric return–volatility effects, i.e. the so–called 'leverage effect'. Consequently, the literature review is structured as follows: We briefly summarise the properties, empirical and theoretical findings related to conventional skewness measures from coarse to fine frequencies. We will then point out advances in the asymmetric volatility literature at high frequency and show the gap in its application to skewness.

In a theoretical context of a stationary asset pricing model, Albuquerque (2012) reconciles negatively skewed index returns with positively skewed stock returns via heterogeneity in firm announcement events which can lead to conditional asymmetric stock return correlations and further to negative skewness in aggregate returns.

[3] Note that correlation between two random variables will be downward biased whenever the series are observed with white noise.

[4] Note the sampling frequency can be different (usually coarser) compared to the generating frequency.

The empirical measurement of a standard coefficient of skewness is heavily plagued by outliers, in particular when there are only few few observations available as is the case for monthly returns, which puts the negative skewness of index index returns into question (Kim and White, 2004). In their extensive Monte Carlo study they criticise the robustness and thus usefulness of conventional measures of skewness and suggest looking beyond the standard textbook formula of skewness to gain further insight into market dynamics.

Robust alternative measures of skewness based on deviations of all return pair combinations from the median have been suggested by Brys et al. (2004). However, a comparable interpretation of their measure is vague; refer to the last section in the introduction for a detailed investigation. Another way is to truncate the distribution and model their conditional multi–quantiles jointly as in Kim et al. (2008). Ghysels et al. (2012) find that a quantile–based measure of conditional skewness (or asymmetry) of returns varies significantly over time and demonstrate its economic significance in an international portfolio setting. But in both studies the relation between their (reduced–form) processes and the unconditional or conditional third moment and thus skewness of returns is not straightforward.

In the context of high–frequency data, Amaya, Christoffersen, Jacobs, and Vasquez (2015) find their estimator of 'realized skewness' in Eq. (1.2.3) below to be priced in the cross–section of the most liquid US stocks when computed at the five–minute level per trading day and then averaged over the course of a week,

$$Avg.\ Skew_{t-0:-4,\Delta} = \frac{1}{5}\sum_{i=0}^{4} Skew_{t-i,\Delta} \quad \text{where} \quad Skew_{t,\Delta} = \frac{\sqrt{N}\sum_{j=1}^{N} r_{t,j}^3}{\left(\sum_{j=1}^{N} r_{t,j}^2\right)^{3/2}}. \tag{1.2.3}$$

The measure circumvents the fact that skewnesss in high–frequency returns decays very fast. If returns $\{r_{t,j}\}_{j=1,2,\cdots,N}$ were i.i.d., skewness at high–frequency Δ, $Skew_{t,\Delta}$, i.e. the third moment over the second moment to the power 3/2, would be related to skewness at low frequency t by a factor of $1/\sqrt{N}$,

$$Skew_t = \frac{N\mathrm{E}(r_{t,j} - \mathrm{E}r_{t,j})^3}{(N\mathrm{E}(r_{t,j} - \mathrm{E}r_{t,j})^2)^{\frac{3}{2}}} = \frac{1}{\sqrt{N}} Skew_{t,\Delta}, \tag{1.2.4}$$

with moments of N–period returns being proportional to N.

Consequently, Eq. (1.2.3) and related measures tells us something about the joint pricing of sign and activity level of large returns over relatively 'short–lived' (weekly) horizons from an asset pricing perspective. Bollerslev and Todorov (2011) point out that

positive and negative jump tails at S&P 500 index level appear to be largely symmetric, which, in combination with the fast decay rate in Eq. (1.2.4) put the hypothesis of the source of negatively skewed equity index returns calculated from very long time series into question. The standard textbook formula of skewness applied to high–frequency data cannot generate the pronounced skew patterns we would expect to encounter. Skewness attributable to high–frequency realized third moment only indirectly contributes to but does not directly measure realized skewness of the return distribution over the horizon of interest.

Realized skewness over longer horizons stems from two sources, skewness at high frequency and asymmetric volatility, i.e. negative correlation between returns and consecutive volatility innovations. Bollerslev et al. (2006), henceforth BLT, find these cross–correlations to be significantly negative for several days, consistent with a prolonged 'leverage effect', whereas the reverse cross–correlations, which would refer to some sort of 'volatility feedback effect', are generally negligible for index level data. It is this finding which paves the path towards aggregating skewness according to Eq. (1.1.1) under the assumptions in Eq. (1.1.2).

A non–parametric measure for the 'leverage effect' has recently been studied by Mykland and Wang (2014) under the stochastic volatility model, based on the ratio of changes in covariations of the form

$$\frac{\mathrm{d}\langle X, \sigma_t^2\rangle_t}{\mathrm{d}\langle X, X\rangle_t} = \frac{\mathrm{d}\langle X, \log\sigma\rangle}{dt} = 2f_t \tag{1.2.5}$$

where X refers to the log asset price process, σ its spot volatility, and $2f_t$ refers to the time–varying 'leverage' parameter of interest. But, as noted in their Section 7, their corresponding estimator of 'leverage' does not yield a method of estimating skewness as it is defined over the correlation between returns and changes in spot volatility at high frequency,[5] while we define asymmetric volatility as in BLT via the correlation between returns and the level of volatility, which naturally extends to the calculation of skewness. Nevertheless, their analysis reveals a theoretical connection between skewness and their 'leverage effect', previously pointed out by Meddahi and Renault (2004).

Also Aït–Sahalia, Fan, and Li (2013) focus on the correlation between returns and changes in volatility, however, at the daily level and coarser frequencies. They show

[5]Changes in spot volatility are inherently difficult to estimate empirically, especially under the diurnal intraday volatility pattern.

different asymptotic biases which arise in estimating the 'leverage effect' once allowing for 'more realistic' estimators of integrated volatility[6] and propose a simple correction based on the assumption that biases are linear in the horizon. Our conjecture from these studies is that using more sophisticated integrated volatility estimators is counterproductive in estimating asymmetric volatility effects, particularly at the intraday level when arising biases might not be that simple (linear). Further light is shed on the non–parametric efficiency of the standard 'natural' realized volatility estimator of Mykland and Zhang (2009) in Renault, Sarisoy, and Werker (2016). We expect our measure of realized skewness to be time–varying as Bandi and Renò (2012) find economically important time–variation in 'leverage' with more negative values associated with higher variance levels.

On the risk–neutral side, Bakshi, Kapadia, and Madan (2003), henceforth BKM, show how to replicate implied skewness of the risk–neutral return distribution from a continuum of option prices in the cross–section of moneyness at a target expiry. The implied skewness measure put forward by BKM naturally extends to the fixed leg of a skew swap contract, $E^Q[Skew]$. But no consensus has yet been reached on the definition of the floating leg of such a swap contract. Neuberger (2012) suggests a concept of skewness, denoted realized skewness, which is a hybrid real/risk–neutral measure as it lends itself to option prices at daily frequency, and is thus biased in the presence of risk premia. Recall that our definition of realized skewness will instead be based on functions of the high–frequency return series only and should thus be an unbiased estimator. Then, 'Neuberger skewness' of equity index returns increases with horizons up to one year with economically important magnitude. The dependence of negative skewness on the time horizon proportinal to the (negative) correlation between return and volatility innovations, first increasing and then decreasing from a central limit theorem, has previously been pointed out in the context of a general stochastic volatility model by Taylor (2005, Chapter 11).

The floating leg of Schneider (2012)'s skew swap is based on high–frequency returns instead of mixed inference of daily returns and option prices as previously employed. It builds on functional differences of log and arithmetic returns such that local deviations from one another are captured through higher order terms. It does not relate the dynamics of functions of high frequency returns to the properties of low frequency returns over the horizon of intert for the construction of the floating leg of such a swap. His approach

[6]Such as the pre–averaging volatility (PAV) estimator of Jacod et al. (2009) and the two scales realized volatility (TSRV) estimator of Aït-Sahalia, Mykland, and Zhang (2005).

resembles the idea used by Jiang and Oomen (2008) who apply a similar methodology, however, to detect jumps in high–frequency data.

The literature on forecasting realized skewness is sparse in the absence of a suitable estimator and has previously focussed on predictable patterns in risk–neutral skewness, henceforth RNS. In their multi–horizon study, Neumann and Skiadopolous (2013) find one–day ahead RNS to be predictable but not exploitable under transaction costs. Consequently, in line with volatility forecasting studies, we might expect RNS to contain incremental information about future realised skewness conditionally on the forecast horizon.

1.3 Methodology: Designing Realized Skewness Estimators

1.3.1 Model–free Realized Skewness

We here provide a non–parametric results for long–horizon realized skewness and its non–centralized counterpart which nests the parametric example in Taylor (2005, Chap. XI, p. 296, Eq. (11.58)) as a special case. Note that we refer to *long–horizon*, e.g., a week or a month, as being long relative to the sampling times, or if we impose a model in the next section the generating frequency at which realizations are being drawn from a discrete time stochastic process, such as five minutes.

In order to illustrate the idea, ignore possibly non–zero expected returns for now and consider the p'th N-period return moment when the N-period return is the sum of N high–frequency returns over time t,

$$\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^p\Big] = \mathrm{E}\Big[\Big(r_{t,1} + r_{t,2} + \cdots + r_{t,N}\Big)^p\Big], \tag{1.3.6}$$

where t could exemplarily refer to the one month horizon from January 1$^{\text{st}}$ to January 31$^{\text{st}}$ 2016, which can be disentangled into $N = D \times M$ high–frequency returns when there are D trading days with M equally spaced high–frequency returns per day.

Then, N-period skewness over time t can be defined as the ratio,

$$\text{Skewness}_t^{(N)} = \frac{\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^3\Big]}{\Big(\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^2\Big]\Big)^{3/2}}. \tag{1.3.7}$$

Note that Eq. (1.3.7) may be biased because the ratio of two unbiased estimators is not in general an unbiased estimator of the ratio, which can be shown via Jensen's inequality; $E(x\frac{1}{y}) = \mathrm{E}x\mathrm{E}\frac{1}{y} \geq \mathrm{E}x\frac{1}{\mathrm{E}y} = \frac{\mathrm{E}x}{\mathrm{E}y}$. Nevertheless, the coefficient can be interpreted as a lower (upper) bound conditional upon prior knowledge of its sign, i.e. the numerator being positive (negative). Also note that in principle for known joint characteristic function of numerator and denominator, with the latter one being positive, the distribution function of the ratio could be reconstructed via Fourier inversion techniques. We here focus throughout on the simpler case above in Eq. (1.3.7).

The denominator is well researched in the literature, with Andersen and Bollerslev (1998), Comte and Renault (1998), Andersen et al. (2001), Barndorff-Nielsen and Shep-

hard (2002) and Meddahi (2002), among others, having established the theoretical basis and empirical properties of estimators of quadratic variation for a broad class of stochastic processes in finance. In particular, Andersen et al. (2003, their Corollary 1) shows that, under empirically realistic conditions, the conditional expectation of quadratic variation is equal to the conditional variance of returns. The numerator of Eq. (1.3.7) and related objects instead is less well researched and we present analysis for both, plain and centralized moments.

Multi–period building blocks of long–horizon skewness. With the literature offering a variety of estimators for the denominator, each having its own merits and caveats in terms of microstructure noise robustness and efficiency gains, essentially, we are interested in the third moment of long–horizon returns.

Denote the expected N–period return by

$$\mathrm{E}\Big[r_t^{(N)}\Big] = \mu_t^{(N)}$$

such that the centralized 3^{rd} N–period return moment can be disentangled into the following long–horizon building blocks

$$\mathrm{E}\Big[\big(r_t^{(N)} - \mu_t^{(N)}\big)^3\Big] \;=\; \mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2, r_t^{(N)}\Big) - 2\mu_t^{(N)}\mathrm{Var}\Big(r_t^{(N)}\Big), \tag{1.3.8}$$

see Eq. (1.A.1) in Appendix 1.A. The last term on the right–hand side of the Eq. (1.3.8) can be thought of as centralization correction when we are interested in the centralized third moment, i.e. the skewness around the first moment. On the other hand, the non–centralized counterpart equals

$$\mathrm{E}\Big[\big(r_t^{(N)}\big)^3\Big] \;=\; \mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2, r_t^{(N)}\Big) + \mu_t^{(N)}\mathrm{Var}(r_t^{(N)}) + \big(\mu_t^{(N)}\big)^3,$$

see Eq. (1.A.2) in Appendix 1.A, with the resulting difference consequently been given by

$$\mathrm{E}\Big[\big(r_t^{(N)}\big)^3\Big] - \mathrm{E}\Big[\big(r_t^{(N)} - \mu_t^{(N)}\big)^3\Big] = 3\mu_t^{(N)}\mathrm{Var}(r_t^{(N)}) + \big(\mu_t^{(N)}\big)^3. \tag{1.3.9}$$

Naturally, if the expected return $\mu_t^{(N)} > 0$, then the non–demeaned third moment (NDTM) is larger than the centralized third moment (CTM) and vice versa. Basically, we have to add back on all those terms which were previously subtracted in the covariance terms, therefore it makes more sense to define the non–centralized third moment in terms of simple expectations of products of returns and functions thereof rather than covariances,

see Eq. (1.3.14ff.) further below in this section.

The resulting differences in the two skewness measures as a function of the expected return and total variation over horizons from one day to three months are shown in Table 1.B.1 in Appendix 3.A. The difference (NDTM - CTM) in terms of its contribution to units of skewness accumulates to 0.28 on a quarterly (0.04 on a daily) horizon when the expected return is 3% with volatility being at the typically encountered long–run average value of 16%. Consider the situation when the expected return is 8% p.a. with volatility at 40% p.a. the resulting difference accumulates to only 0.04 units at the daily level, whereas it is, *ceteris paribus*, around 0.3 units at the quarterly horizon. Situations in which low levels of volatility meet high expected returns in Table 1.B.1 are empirically less relevant as are long–run volatility levels in the lower range reported.

Realized third moment of long–horizon returns relative to the sampling frequency

A high–frequency decomposition of covariances. Regarding the first term on the right–hand side, which is identical in both Eq. (1.3.8) and Eq. (1.3.9), we have the following covariance decomposition, with intermediate steps provided in Appendix 1.A.1, Eq. (1.A.3ff.),

$$\begin{aligned}&\mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2, r_t^{(N)}\Big)\\&= \mathrm{Cov}\Big(\sum_{i=1}^{N} r_{t,i}^2, \sum_{j=1}^{N} r_{t,j}\Big) + 2\ \mathrm{Cov}\Big(\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} r_{t,i}r_{t,k}, \sum_{j=1}^{N} r_{t,j}\Big).\end{aligned} \tag{1.3.10}$$

Now under the empirically realistic assumption that covariance terms spanning three random, non–overlapping ($i \neq j \neq k$) high–frequency returns are zero when the efficient return process is assumed to be uncorrelated,

$$\underset{i\neq j}{\mathrm{corr}}(r_{t,i}, r_{t,j}) = 0,$$

$$\underset{i\neq j\neq k}{\mathrm{Cov}}(r_{t,i}r_{t,k}, r_{t,j}) = \mathrm{E}[r_{t,i}r_{t,k}r_{t,j}] - \mathrm{E}[r_{t,i}, r_{t,k}]\mathrm{E}[r_{t,j}] = \mu_t^3 - \mu_t^2\mu_t = 0,$$

where $\mu_{t,i} = \mu_{t,k} = \mu_{t,j} = \mu_t$. Then, the second part on the right–hand side of the covariance decomposition in Eq. (1.3.10) reduces to those cases when either $j = i$ or

$j = k$, i.e.

$$\begin{aligned}&\mathrm{Cov}\Big(\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} r_{t,i}r_{t,k}, \sum_{j=1}^{N} r_{t,j}\Big)\\ &= \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Big(\mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,i}\big) + \mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,j}\big)\Big),\end{aligned}$$

where we have further grouped covariance terms according to their return lead–lag orders, i.e. those involving return combinations of the form $r_{t,i}^2 r_{t,j}\ \forall\ j < i$, which refers to the volatility feedback effect (VFB), and those involving $r_{t,i}r_{t,j}^2\ \forall\ j > i$, which refers to the asymmetric volatility effect (AVL). Further note that

$$\begin{aligned}\mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,i}\big) &= \mathrm{E}\big[r_{t,i}^2 r_{t,j}\big] - \mathrm{E}[r_{t,i}r_{t,j}]\mathrm{E}[r_{t,j}] = \mathrm{E}\big[r_{t,i}^2 r_{t,j}\big] - \mu_t^3,\\ \mathrm{Cov}\big(r_{t,i}^2, r_{t,j}\big) &= \mathrm{E}\big[r_{t,i}^2 r_{t,j}\big] - \mathrm{E}[r_{t,i}^2]\mathrm{E}[r_{t,j}] = \mathrm{E}\big[r_{t,i}^2 r_{t,j}\big] - [\mu_t^3 + \mu_t\mathrm{Var}(r_{t,i})],\\ \mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,i}\big) &= \mathrm{Cov}\big(r_{t,i}^2, r_{t,j}\big) + \mu_t\mathrm{Var}(r_{t,i}),\end{aligned}$$

and as

$$\sum_{i=1}^{N}\mathrm{Cov}\Big(r_{t,i}^2, r_{t,i}\Big) = \sum_{i=1}^{N}\Big(\mathrm{E}[(r_{t,i}-\mu_t)^3] + 2\mu_t\mathrm{Var}(r_{t,i})\Big),$$

combining all terms leads to the results

$$\begin{aligned}&\mathrm{E}\Big[\big(r_t^{(N)} - \mu_t^{(N)}\big)^3\Big]\\ &= \sum_{i=1}^{N}\mathrm{E}[(r_{t,i}-\mu_t)^3] + 3\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{Cov}\Big(r_{t,i}^2, r_{t,j}\Big) + 3\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{Cov}\Big(r_{t,j}^2, r_{t,i}\Big), \quad (1.3.11)\\ &= -2N\mu_t^3 + 3\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{Cov}\Big(r_{t,i}^2, r_{t,j}\Big) + 3\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{Cov}\Big(r_{t,j}^2, r_{t,i}\Big), \quad (1.3.12)\\ &= -2N\mu_t^3 + 3\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{Cov}\Big(r_{t,j}^2, r_{t,i}\Big), \quad (1.3.13)\end{aligned}$$

where

- the second equality follows, without loss of generality for long horizons, from the assumption that demeaned high–frequency returns are symmetric, which is the standard theoretical assumption in continuous time and seems to hold well empirically near the in–fill limit even in the presence of discontinuities (jumps) superimposed on an otherwise continuous price path, see Bollerslev and Todorov (2011, their Table III). To further illustrate the negligible impact of this assumption hypothetically

consider the situation in which we impose high–frequency return residuals were genuinely asymmetric at, e.g., one minute frequency, coming from a Skewed T Distribution with skewness equal to -0.20,[7] then the contribution of the instantaneous skewness to long–horizon skewness (relative to the observation times) of aggregated five minute, hourly and daily returns equals -0.089, -0.026, -0.010, and is virtually zero for longer horzions of interest.

- the last equality stems from the assumption that the current level of volatility is uncorrelated with future returns, which is the standard assumption in discrete and continuous time models (Harvey and Shephard, 1996; Duffie et al., 2000), and has been shown to hold well empirically (Bollerslev et al., 2006). Also note that the first term on the right–hand side directly related to the expected return very small such that skewness at long horizons builds up from the asymmetry between returns and the future volatility at the intraday level almost exclusively.

A high–frequency decomposition of expectations. As previously mentioned, with regard to Eq. (1.3.9), it is perhaps more natural to define the non–centralized third moment in terms of its expectation decomposition only rather than in terms of covariances. Then, in analogy to the derivation above, and with intermediate steps provided in Appendix 1.A.1, Eq. (1.A.4ff.),

$$\mathrm{E}\Big[\big(r_t^{(N)}\big)^3\Big] \;=\; \sum_{j=1}^{N}\sum_{i=1}^{N} E\Big[r_{t,i}^2 r_{t,j}\Big] + 2\sum_{j=1}^{N}\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} \mathrm{E}\Big[r_{t,i}r_{t,k}r_{t,j}\Big], \qquad (1.3.14)$$

where combining terms for $j = i$ and $j = k$ leads to the latter term on the right–hand side above being equal to

$$2\left(\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} \mathrm{E}\Big[r_{t,i}r_{t,k}r_{t,i}\Big] + \sum_{i=1}^{N-1}\sum_{k=i+1}^{N} \mathrm{E}\Big[r_{t,i}r_{t,k}r_{t,k}\Big] + \sum_{j=1}^{N}\sum_{\substack{i=1\\ j\neq i,k\neq j}}^{N-1}\sum_{k=i+1}^{N} \mathrm{E}\Big[r_{t,i}r_{t,k}r_{t,k}\Big]\right).$$

Combining these terms and stepwise imposing the same assumptions as before, starting with uncorrelatedness of randomly sampled returns such that $\underset{j\neq i,k\neq j}{\mathrm{E}}\big[r_{t,i}r_{t,k}r_{t,k}\big] = \mu_t^3$,

[7] For example, $f_{SGT}(r_{t,i}; \mu = 0, \sigma, \lambda - 0.124113711767968, p = 2, q = 200)$, where the level of σ is irrelevant.

yields the following results for the non–centralized third moment,

$$\begin{aligned}
&\mathrm{E}\Big[\big(r_t^{(N)}\big)^3\Big] \\
&= (N-2)(N-1)N\mu_t^3 + \sum_{i=1}^{N}\mathrm{E}[r_{t,i}^3] + 3\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{E}\Big[r_{t,i}^2 r_{t,j}\Big] + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{E}\Big[r_{t,i} r_{t,j}^2\Big]\right), \qquad (1.3.15)\\
&= \big(1+(N-1)^3\big)\mu_t^3 + 3\sum_{i=1}^{N}\mu_t\mathrm{Var}(r_{t,i}) + 3\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{E}\Big[r_{t,i}^2 r_{t,j}\Big] + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{E}\Big[r_{t,i} r_{t,j}^2\Big]\right) \qquad (1.3.16)\\
&= \big(1+(N-1)^3\big)\mu_t^3 + 3\sum_{i=1}^{N}\mu_t\mathrm{Var}(r_{t,i}) + 3\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{E}\Big[r_{t,i} r_{t,j}^2\Big], \qquad (1.3.17)
\end{aligned}$$

where as before the second equality stems from the assumption of symmetrically distributed return residuals at the intraday level, and the last equality follows from the non–significant volatility feedback effect. The small contribution from the primarily the first of the two terms of the righ–side hand above can be inferred from Table 1.B.2, where even at the quarterly horizon the contribution to non–centralized skewness stemming from these two terms combined accumulates to only 0.0158 when the long–run volatility is at the typically encountered value of 16% p.a. for the S&P 500 and expected returns are at relatively high 8% p.a.

Candidate functional forms of covariation Albeit very appealing for the analysis beforehand, the most straightforward estimator of covariation between returns and future variation, i.e. when $\theta = 1$ and $\lambda = 2$ below, is likely to be extremely noisy. Recall, we are interested in covariation measures of the form

$$\mathrm{Cov}\Big(\mathrm{sign}(r_{t,i})|r_{t,i}|^{\theta}, |r_{t,j}|^{\lambda}\Big) \quad \forall \quad i<j, \qquad (1.3.18)$$

and we will focus on the interesting, and potentially less noisy case, when $\theta = \lambda = 1$ in the semi–parametric part of this paper which requires additional assumptions (structure) to relate these functional forms of covariation back to the original measure of interest.

There are also a variety of non–parametric candidate functions for the second argument above, which essentially requires an estimator of spot variation, under very mild assumptions. In particular, there is the class of multipower variation (MPV) measures,

derived in a series of papers[8] by Barndorff–Nielsen and Shephard and their co–authors, which are defined via m products of adjacent absolute return increments raised to the r/m'th power. The following product term delivers an unbiased estimate of the respective power of spot variation,

$$\text{MPV}(m;r) = \mu_{r/m}^{-m} \prod_{j=1}^{j+m-1} |r_{t,j}|^{\frac{r}{m}}, \tag{1.3.19}$$

with $\mu_{r/m} = \text{E}[|u_j|^p] = 2^{p/2}\pi^{-1/2}\Gamma(p+1)/2)$, $U \sim N(0,1)$, if adjacent returns are i.i.d. Gaussian. Eq. (1.3.19) also nests the building block of the natural realized variance estimator, $\text{MPV}(2;1)$, and the first estimator of integrated variance, bipower variation $\text{MPV}(2;1)$, which is "somewhat robust" to rare jumps. Potentially more interesting are the tripower variations $\text{MPV}(3;2)$ and $\text{MPV}(3;1)$, as first explored in Huang and Tauchen (2005), and the MedRV estimator suggested in Andersen, Dobrev, and Schaumburg (2012). from a theoretical perspective, while these are asymptotically less efficient, they appear to better dampen the impact of large returns, i.e. in a continuous time setting discontinuities superimposed on an otherwise continuous price path, the latter measure has an asymptotic distributional result under the alternative unlike bipower variation, and works in particular better in finite samples when some of the discretely sampled returns are in fact zero[9] due to stale prices and the discrete price grid. The MedRV estimator has building blocks of the form,

$$\text{MedRV} = \frac{\pi}{6-4\sqrt{3}+\pi}\text{med}\big(|r_{t,j-1}|,|r_{t,j}|,|r_{t,j+1}|\big)^2, \tag{1.3.20}$$

and delivers an unbiased estimate of the underlying spot variance under the standard assumption that the high–frequency returns are locally a sequence of i.i.d. Gaussian variables.

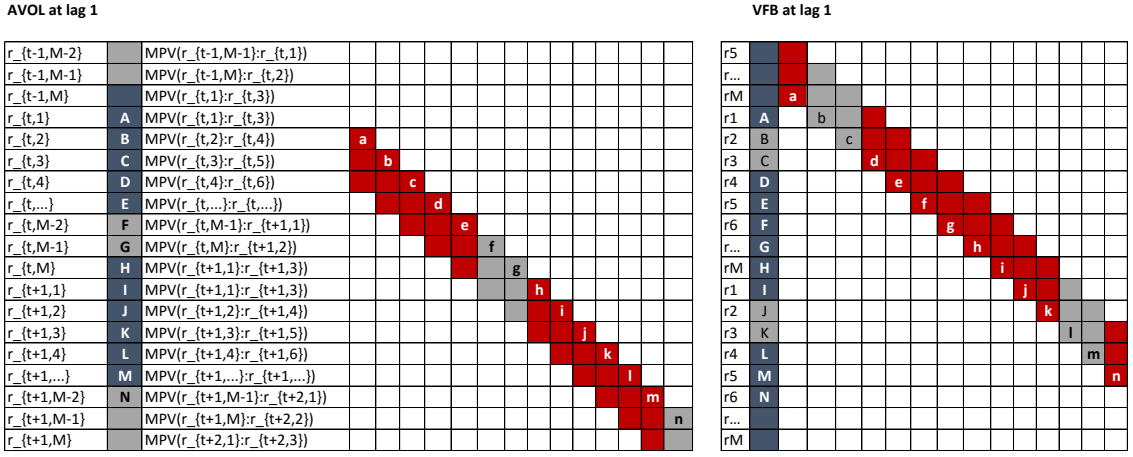
In terms of implementing multi–power variation measures, it seems advisable to not construct estimates of the level of spot volatility when the local increments are spanning different days. Figure 1.1 visualizes the chosen computational algorithm, i.e. the matching pairs of return and volatility proxy at lead lag order one for computing realizations of the asymmetric volatility and the volatility feedback effect, denoted AVL and VFB, from return increments spanning three different days.

[8]The reader is referred to Barndorff-Nielsen, Shephard, and Winkel (2006) and the many references therein.

[9]In practice, this also restricts the parameter m to be of low order despite its theoretical appeal to be large.

Figure 1.1: Computing multi–power variation–based realizations of AVL and VFB

Consider three consecutive days $t-1, t, t+1$ each with M equally spaced observations. The left figures refers to the computation of realizations (via simple multiplication of return and volatility proxy) of the asymmetric volatility effect (AVL) at lag one only and the right figure refers to the computation of realizations of the volatility feedback effect. The return increments are marked on the left side of each subfigure and highlighted in blue from A to N, the corresponding chosen volatility proxies span the multi–power variation function over the corresponding red increments from a to n. For the grey return increments F, G and volatility proxies f, g realizations are chosen to not be computed at lead lag order one for the aforementioned reason. However, the neglected returns will be useful again for the computation of realizations of the respective measure at lead lag order two, three etc.



Estimators of Quadratic Variation

The denominator should be a suitable estimator of the quadratic variation of the asset price sample path taking the previously addressed autocovariance structure into account. Realized kernels introduced by Barndorff–Nielsen, Hansen, Lunde, and Shephard (2008); Barndorff–Nielsen et al. (2009), henceforth BNHLS, are particularly well suited to estimate the quadratic variation of the efficient price process from noisy high frequency data. Their estimator takes the form

$$K(X) = \sum_{h=-H}^{H} k\Big(\frac{h}{H+1}\Big)\gamma_h, \quad \gamma_h = \sum_{j=|h|+1}^{m} r_j r_{j-h}, \tag{1.3.21}$$

with their preferred choice of $k()$ being the Parzen kernel following Gallant (1987),

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & \text{for} \quad 0 \leq x \leq 1/2 \\ 2(1-x)^3 & \text{for} \quad 1/2 < x \leq 1 \\ 0 & \text{for} \quad x > 1 \end{cases}$$

which satisfies the smoothness conditions, $k'(0) = k'(1) = 0$, guarantees a non–negative estimate, and has a faster convergence rate than e.g. the Bartlett kernel. The optimal bandwidth H in terms of achieving the best asymptotic bias variance trade-off, see BNHLS, is when $H \propto m^{3/5}$, in particular

$$H^* = c^* \xi^{4/5} m^{3/5}, \text{ with } c^* = \left(\frac{k''(0)^2}{k_\bullet^{0.0}} \right)^{1/5}, \text{ and } \xi^2 = \frac{\omega^2}{\sqrt{T \int_0^T \sigma_u^4 \mathrm{d}u}}, \tag{1.3.22}$$

where $c^* = (12^2/0.269)^{1/5} = 3.5134$ for the Parzen kernel, and ξ^2 is a data driven ratio of noise variance to volatility of volatility over the course of one day. We estimate ξ^2 following BNHLS as

$$\hat{\xi}^2 = \hat{\omega}^2 / \hat{\mathrm{IV}}, \tag{1.3.23}$$

where we hope that $IV^2 = \left(\int_0^T \sigma_u^2 \mathrm{d}u \right)^2 \simeq T \int_0^T \sigma_u^4 \mathrm{d}u$ is more precise than trying to estimate $\sqrt{T \int_0^T \sigma_u^4 \mathrm{d}u}$ directly.[10] In fact, we can regard $\hat{\mathrm{IV}}$ as an estimator for the lower bound of the integrated quarticity. In specific, we use

$$\hat{\mathrm{IV}} = \mathrm{RV}_{\mathrm{SS,sparse}} \tag{1.3.24}$$

where $\mathrm{RV}_{\mathrm{SS,sparse}}$ is the subsampled, sparse RV estimator introduced by Zhang, Mykland, and Aït-Sahalia (2005) as a simple way to improve efficiency of the simple RV estimator,

$$\mathrm{RV}_{\mathrm{sparse}} = \sum_{i=1}^{M_{\mathrm{sparse}}} r^2_{t,i/M_{\mathrm{sparse}}}. \tag{1.3.25}$$

The subsampling technique involves averaging over the $\mathrm{RV}_{\mathrm{sparse}}$ estimates obtained by rolling the time grid of price observations in small steps forward while keeping the length of the high frequency return period fixed.[11] Following BHNLS, the noise variance ω^2 is computed by averaging over the dense realized variance estimators using trades every q'th second under which the noise increments are approximately free of autocorrelation,

$$\hat{\omega}^2_{(i)} = \frac{\mathrm{RV}^{(i)}_{\mathrm{dense}}}{2n_{(i)}}, i = 1, \ldots, q, \quad \hat{\omega}^2 = \frac{1}{q} \sum_{i=1}^{q} \hat{\omega}^2_{(i)} \tag{1.3.26}$$

[10] Alternatively, one could explore the use of nearest neighbourhood trunction in estimating integrated quarticity as in Andersen, Dobrev, and Schaumburg (2014) in order to determine the bandwidth of the realised kernel estimator but this is beyond the scope of the work presented here.

[11] Exemplarily, $\mathrm{RV}_{\mathrm{sparse}}$ can be computed using 5–minute returns ending 9:35, 9:40, ... 16:00 EST. But we can also compute it using 5–minute returns starting at 9:31 and ending 9:36, 9:41, ... etc. which eliminates the 'arbitrary' choice of a particular set of grid points.

Note that upward biased estimates of ω^2 increase the bandwidth, and thus robustness to the autocovariance structure.

1.3.2 Addressing the periodicity component in intraday volatility

Genuine high–frequency volatility time series Genuine high–frequency volatility time series, as opposed to daily volatility estimators based on aggregating high–frequency returns such as realized variance, are subject to a pronounced periodicity component which has a strong impact on the dynamics of the high–frequency returns and functions thereof. Standard volatility models which do not address these systematic patterns in volatility across, e.g., a trading day, often fail to converge in terms of their volatility factor persistences and volatility–of–volatility estimates. Thus, when interested in imposing structure on the return–volatility dynamics, the arising systematic pattern needs to be precisely accounted for to uncover and preserve the fine complex structure of asset returns at very short lead and lag orders.

High–frequency return decomposition In a model–free setting, i.e., without imposing any structure on the high–frequency return dynamics, we do not expect the intraday periodicity component to fundamentally change a non–parametric estimate of the asymmetric volatility effect, i.e. between returns and the level of volatility.[12] In that case, the estimate will just be driven primarily by large returns occurring during the deterministically high volatility periods, and we expected to achieve a more precise estimate by accounting for the periodicity factor in a preliminary step.

For clarity of exposition, let t now denote one day with subscripts i running from $i = 1, 2, \ldots$ to M only.[13] Then, the data generating process for the high–frequency return subject to intraday periodicity effects can be written as

$$R_{t,i,M} = S_{d,i,M}\sigma_{t,i,M}u_{t,i,M}, \tag{1.3.27}$$

[12] Intuitively and loosely speaking, this follows from the fact that at ultra high–frequency, near the in–fill limit, adjacent returns are symmetrically distributed. Then, the product of symmetric ultra high–frequency returns with subsequent spot levels of volatility involving a deterministic periodicity component will only be scaled by the periodicity factor. As the periodicity factor repeats and we multiply returns from deterministically low (high) volatility periods with volatility from high (low) volatility periods, the periodicity should be averaged out approximately.

[13] Recall previously t denotes the observation window and subscripts run from $i = 1, 2, \ldots, N$, where $N = T \times M$ with T counting days.

where $S_{d,i,M}$ is a deterministic function of the periodicity component based on the time of the day indicated by subscript i of M, and with d distinguishing weekdays and recurring macroeconomic announcement days, i.e., the periodicity component does *only* vary across weekdays and announcement days with scheduled news releases.

Non–stochastic periodicity factors There is inconclusive evidence that level and shape of the periodicity component depend directly on the characteristics of each trading day such as a measure of its volatility, especially once we allow for an intra*week* periodicity component as in Boudt, Croux, and Laurent (2011) and Lahaye, Laurent, and Neely (2011), plus distinctly filtered announcement cyclicalities. However, high–frequency periodicity components estimated from raw high–frequency returns might be distorted through high and low volatility regimes (volatility clustering). As we group observations across days, it seems reasonable to control for changing volatility in general in order to isolate high–frequency periodicity effects. Following Andersen and Bollerslev (1998) and subsequent studies, we estimate the deterministic component in intraday volatility from high–frequency returns standardized by an estimate of the average volatility in the local neighbourhood, i.e. here the same day in line with Boudt et al. (2011),

$$\bar{R}_{t,i,M} = R_{t,i,M}/\bar{\sigma}_t, \tag{1.3.28}$$

where $\bar{\sigma}_t$ is estimated using an estimator of quadratic variation such as a standard realized variance or the more efficient realized kernel. The standardized high–frequency return $\bar{R}_{t,i/M}$ has variance approximately equal to the squared periodicity factor.[14]

The Taylor and Xu (1997)–type, henceforth TX, periodicity estimator, here for standardized high–frequency returns, for any element i in cycle d is given by

$$\hat{S}^2_{d,i,M} = \frac{\frac{1}{N_d}\sum_{t\in D_d}\bar{R}^2_{t,i,M}}{\frac{1}{M}\sum_{i=1}^{M}\frac{1}{N_d}\sum_{t\in D_d}\bar{R}^2_{t,i,M}}, \tag{1.3.29}$$

where D_d denotes the collection of indices t which are subject to the same cyclicality component $d = 1, 2, 3, \ldots$ and with N_d denoting the number of elements in set D_d. The standardization condition in the denominator above ensures that the mean squared peri-

[14]Note its distribution is actually truncated over the $[-\sqrt{M}, \sqrt{M}]$ interval and cannot be truly Gaussian even under simplest assumptions as pointed out by Andersen et al. (2007).

odicity factor equals one over the respective cycle d,

$$\frac{1}{M}\sum_{i=1}^{M} S^2_{d,i,M} = 1. \tag{1.3.30}$$

This setting might be considered less restrictive than imposing the absolute periodicity factor (not its square) to be one over a local window such as one day (not over the entire cycle) which resembles the notion of constant volatility through the day. Using squares, i.e. variances, allows us to use simple aggregation results over different sampling frequencies.

The Nadaraya–Watson (NW) kernel regression is particularly well suited to obtain a smooth non–parametric intraday periodicity pattern when very finely sampled data is available, and it is thus particularly popular in the financial duration literature, see e.g., Nolte, Taylor, Zhao (2016) for applications. The Nadaraya–Watson estimate of the unknown regression function can be written as

$$\hat{s}^2_{h,K}(d,i,M) = \frac{\frac{1}{N_d}\sum_{t\in D_d}\frac{1}{h}\frac{1}{M}\sum_{j=1}^{M} K\left(\frac{i-j}{h}\right)(\bar{R}^2_{t,j,M})}{\frac{1}{N_d}\sum_{t\in D_d}\frac{1}{h}\frac{1}{M}\sum_{j=1}^{M} K\left(\frac{i-j}{h}\right)},$$

which can be re–written as a weighted average,

$$\hat{s}^2_{h,K}(d,i,M) = \sum_{t\in D_d}\sum_{j=1}^{M} w_j(\bar{R}^2_{t,j,M}), \quad \text{where} \quad w_j = \frac{K\left(\frac{i-j}{h}\right)}{\sum_{t\in D_d}\sum_{j=1}^{M} K\left(\frac{i-j}{h}\right)},$$

with $\sum_{t\in D_d}\sum_{j=1}^{M} w_j = 1$. As before, the factors average one when we set

$$\hat{S}^2_{h,K}(d,i,M) = M\frac{\hat{s}^2_{h,K}(d,i,M)}{\sum_{i=1}^{M}\hat{s}^2_{h,K}(d,i,M)}, \tag{1.3.31}$$

and Eq. (1.3.30) applies. Once the NW kernel regression is estimated at very high frequency such as, e.g., three or 15 seconds to avoid so–called bounce back effects and prevent domination of the periodicity components by microstructure noise effects, it can be further aggregated at a lower target frequency of choice. This is particularly useful when considering the class of average realized (co)–variation estimators computed by alternating the starting grid points. The piecewise periodicity components can be computed to match the respective grid points exactly; refer to Section 1.3.1, Footnote 11 regarding the sub–sampling methodology.

High–frequency returns adjusted for deterministic intraday volatility factors
For the main analysis, we are interested in sparsely sampled high–frequency returns such

as, e.g., five–, eleven–minute and even coarser frequencies. For any random t in a set D_d the respective sparsely sampled high–frequency returns of interest, adjusted for intraday periodicity effects, are given by

$$\hat{r}_{t\in D_d,i,M_{\text{sparse}}} = \frac{R_{t\in D_d,i,M_{\text{sparse}}}}{\hat{\bar{S}}_{h,K}(d,i,M_{\text{sparse, all}})}, \tag{1.3.32}$$

where the denominator, $\hat{\bar{S}}_{h,K}(d,i,M_{\text{sparse, all}})$, is the resulting 'realized' M_{sparse} periodicity component, computed over the M_{dense}, i.e., "all", periodicity component estimates in the corresponding M_{sparse} interval.

When overnight and opening period returns are disregarded say over the first $m \geq 1$ equally spaced observation times, the adjustment factor,

$$\hat{a}^2_{h,K}(d,m,M) = \frac{M-m}{M-\hat{Y}_{h,K}(d,m,M)}, \quad \text{with} \quad \hat{Y}_{h,K}(d,m,M) = \sum_{j=1}^{m} \hat{S}^2_{h,K}(d,j,M),$$

applies to the periodicity component for the $R_{t\in D,i,M_{\text{exo}}}$, where $M_{\text{exo.}} = (M-m)$ denotes the total number of equally spaced return increments over primary market trading hours.[15] For clarity of exposition, we set the first i'th return free from overnight and opening effects to be the $(m+1)$'th original observation time. Then, we have under the sparse sampling scheme,

$$\hat{r}_{t\in D_d,i,M_{\text{sparse,exo}}} = \frac{R_{t\in D_d,i,M_{\text{sparse,exo}}}}{\hat{a}_{h,K}(d,m,M)\hat{S}_{h,K}(d,i,M_{\text{sparse,exo}})}, \tag{1.3.33}$$

where as before we use all available data (dense sampling) for the construction of the periodicity component.

1.3.3 Semi–parametric and model–based approaches

Continuous vis–à–vis discrete time general stochastic volatility return processes

The pragmatic choice of a lognormal distribution for stochastic volatility has gained credibility several years after its adoption for daily data based on the evidence reported in several high–frequency studies, e.g. Andersen et al. (2000), Andersen et al. (2001), and Areal and Taylor (2002). The general lognormal stochastic volatility model below is often preferred over the square–root process formulation as it also produces quite rough volatility paths through the exponential function. It is the benchmark model when the

[15] As it is straightforward to cover the after market closure period as well.

focus is on real–world asset price dynamics rather than fast option pricing formula under risk–neutrality and has been used in simulation of high frequency returns in Huang and Tauchen (2005), henceforth HT, and BNHLS, among many others. Its continuous time formulation in standard notation reads

$$\begin{aligned} \mathrm{d}p_t &= \mu \mathrm{d}t + \sigma_t \mathrm{d}W_t \\ \sigma_t &= \exp(\beta_0 + \beta_1 \varrho_t) \\ \mathrm{d}\varrho_t &= \kappa \varrho_t \mathrm{d}t + \mathrm{d}B_t \\ \mathrm{corr}(\mathrm{d}W_t, \mathrm{d}B_t) &= \rho \end{aligned} \tag{1.3.34}$$

Assume an integer value of M intervals over the $[0,1]$ time span such that $\Delta = 1/M$. Then, the set of Equations (1.3.35) defines a discrete time process for times $[0, 1/M]$, here denoted **model M***,

$$\begin{aligned} p_{(j+1),M} &= p_{j,M} + \mu\Delta + \sigma_{j,M}\sqrt{\Delta}\varepsilon_j^W \\ \sigma_{j,M} &= \exp(\beta_0 + \beta_1 \varrho_{j,M}) \\ \varrho_{(j+1),M} &= \exp(\kappa\Delta)\varrho_{j,M} + \frac{1}{\sqrt{-2\kappa}}\sqrt{1-\exp(2\kappa\Delta)}\varepsilon_j^B, \end{aligned} \tag{1.3.35}$$

which follows from the General SDE solution in Appendix 1.A.2 with

$$\begin{pmatrix} \varepsilon_j^W \\ \varepsilon_j^B \end{pmatrix} \sim i.i.d.\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \bullet & 1 \end{pmatrix} \right).$$

Rearranging terms for the volatility process as shown in App. 1.A.2 yields

$$\ln(\sigma_{j,M}) = \beta_0 + \exp(\kappa\Delta)\Big(\log(\sigma_{(j-1),M}) - \beta_0 \Big) + \frac{\beta_1}{\sqrt{-2\kappa}}\sqrt{1-\exp(2\kappa\Delta)}\varepsilon^B_{(j-1),M}. \tag{1.3.36}$$

While the Ornstein–Uhlenbeck process, $\{\mathrm{d}\varrho\}$, has an exact iterative discretization scheme[16] based on the solution in 1.A.2, the set of Eqs. (1.3.34) cannot be discretized in the sense that the distribution of the $p_{1,M}, p_{2,M}, \ldots$ is precisely that the the continuous time process at times $j = 1, 2, \ldots$ for the same value of $p_{0,M}$ and $\sigma_{0,M}$. However, when the discretization step is small, the discretization bias will be small. The impact for the analysis presented here will be quantified exemparily in the following Section 1.3.3.

The discrete time lognormal SV model originated in Taylor (1982, 1986), and subse-

[16] A slightly simpler Euler scheme entails some discretization error, which can be inferred from the results in Glasserman (2004).

quently popularized in Harvey and Shephard (1996), can be written in the following way for times [0, 1/M], here denoted **model M**,

$$\begin{aligned} r_{j,M} &= \mu\Delta + \sigma_{j,M}\sqrt{\Delta}u_{j,M} \\ \log(\sigma_{j,M}) - \alpha &= \phi_M\Big(\ln(\sigma_{(j-1),M}) - \alpha\Big) + \eta_{j,M}, \end{aligned} \tag{1.3.37}$$

where the variables $(u_{j,M}, \eta_{(j+1),M})'$ are bivariate normally distributed with

$$\begin{pmatrix} u_{j,M} \\ \eta_{(j+1),M}) \end{pmatrix} \sim i.i.d.\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta\sigma_{\eta,M} \\ \bullet & \sigma^2_{\eta,M} \end{pmatrix}\right).$$

Model M and **model M*** are the same if

$$\begin{aligned} \alpha &= \beta_0, \\ \phi_M &= \exp(\kappa\Delta), \\ \sigma_{\eta,M} &= \frac{\beta_1}{\sqrt{-2\kappa}}\sqrt{1-\exp(2\kappa\Delta)}. \end{aligned}$$

The first line above equalizes the scaling factors, the second equality refers to the sensible choice when we set $\ln(\phi_M) = \kappa\Delta = \kappa/M$,[17] and the third line stems from the fact that the log volatility in Eq. (1.3.37) follows an AR(1) process with mean α and variance β^2,

$$\mathrm{Var}\Big[\ln(\sigma_{j,M})\Big] = \beta^2 = \frac{\sigma^2_{\eta,M}}{1-\phi^2_M},$$

with β^2 set equal to $\frac{\beta_1^2}{-2\kappa}$ such that $\sigma_{\eta,M} = \beta\sqrt{1-\phi^2_M} = \frac{\beta_1}{\sqrt{-2\kappa}}\sqrt{1-\exp(2\kappa\Delta)}$.

As shown in Yu (2005) the contemporaneous formulation is inferior to the general specification, as the return process then ceases to be a martingale difference sequence and a strict interpretation of the statistical leverage effect is not obvious. Henceforth, we will focus on the discrete time version, **model M**, with parameters $\alpha, \phi_M, \beta, \delta$ with discretization step M, as it is more natural to the way we observe discrete returns and proxy the latent volatility under sparse sampling due to arising microstructure noise.

[17] More clearly, if e.g. $\phi_{M=1} = 0.98$, then $\phi_{M=2} = 0.98^{\Delta=\frac{1}{2}}$.

Skewness in the one factor general SV model: Analytical results

Assume $\mathrm{E}[r_{t,i,M}] = 0$ for clarity of exposition for now, then as in the setup for model–free inference in the Section 1.3.1, Eq. (1.3.6) and Eq. (1.3.7), with N indexing return increments and M denoting the assumed integer number of equally spaced intraday returns within one day, realized skewness for $N \geq 2$ equals

$$\begin{aligned} \mathrm{RSkew}_t^{(N)} &= \frac{\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^3\Big]}{\Big(\mathrm{E}\Big[\Big(r_t^{(N)}\Big)^2\Big]\Big)^{3/2}} = \frac{3\sum_{i<j}\mathrm{E}[r_{t,i,M}r_{t,j,M}^2]}{\Big(\sum_i \mathrm{E}[r_{t,i,M}^2]\Big)^{3/2}} \\ &= \frac{3\sum_{j=2}^{N}(N-j+1)\mathrm{Cov}[r_{t,1,M}, r_{t,j,M}^2]}{\Big(N\mathrm{Var}[r_{t,i,M}]\Big)^{3/2}}, \end{aligned} \tag{1.3.38}$$

where the second equality follows due to variance stationarity for any random i

$$\mathrm{Var}\Big[r_{t,i,M}\Big] = M^{-1}\exp(2\alpha + 2\beta^2), \tag{1.3.39}$$

and covariance stationarity in the third order sense for any random i and $j \geq 2$, with

$$\mathrm{Cov}\Big[r_{t,1,M}, r_{t,j,M}^2\Big] = M^{-\frac{3}{2}}2\delta\sigma_{\eta,M}\phi_M^{j-1}\ \exp(3\alpha + \frac{1}{2}(5+4\phi_M^j)\beta^2), \tag{1.3.40}$$

which follows from the respective proofs in App. 1.A.2. Also note that the dependence on M of the two equations above cancels out in Eq. (1.3.38).

Still as preluded in Section 1.3.3, for fixed $\phi_{@M=1}$, invariant δ, and invariant standard deviation of the log volatility β, skewness depends on the generating frequency M as it accumulates over time. Interestingly, Figure 1.2 shows that even for the $N = M$ case, that is skewness of the one day return horizon, there is still a substantial skewness being produced for typically encountered values of M in the empirical integrated variance literature due to arising microstructure noise at very fine frequencies.

Table 1.1 shows that the (intraday) generating frequency M is of negligible importance for the skewness of longer return horizons, i.e. N being a multiple of M, which is the focus of interest here. Then, the magnitude of skewness of the three month return horizon is almost regardless of asymmetry being generated at either high frequency or at as coarse as daily frequency. Even for small integer values of M, the magnitude of skewness of the two day return distribution are virtually the same.

When the asymmetric effect is being generated at a frequency higher than can empiri-

Figure 1.2: Analytical skewness for a one day return horizon, as a function of the intraday *generating* frequency M

X–AXIS: Denotes the generating frequency $M = 390, 389, \ldots, 1$ corresponding to one minute returns down to one daily open–close return at which there is zero skewness, note here $N = M$. DASHED BLUE LINE: Analytical one day return horizon skewness, Eq. (1.3.38) as a function of the generating frequency M. SOLID GREY LINE: Benchmark analytical one day return skewness when the generating frequency is one second ($N = M = 23,400$) with a value of -0.15196. The underlying model is the one factor general SV model, Eq. (1.3.37), with parameters $\phi_{@M=1} = 0.98, \beta = 0.4, \delta = -0.5$.

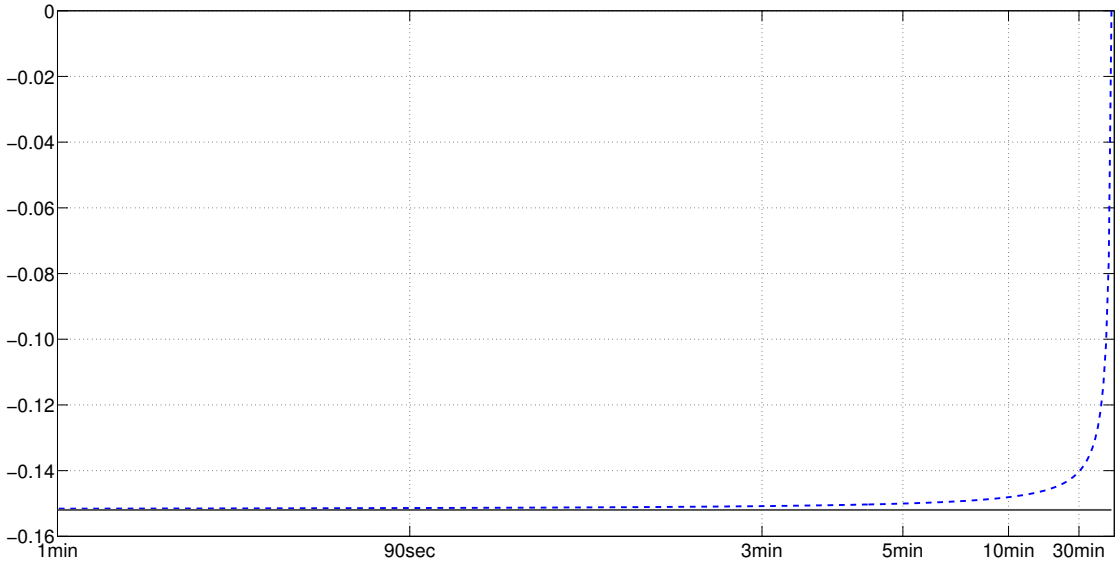


Table 1.1: Comparison of analytical skewness as a function of the *generating* frequency M across extending return horizons

Gen. Freq. denotes the generating frequency with $M = 23,400, 390, 78, 39, 26, 13, 1$ for 1sec, 1min, 5min etc. and multiplied by 1, 2, 5, 10, 21, 42, 63 to yield 1 day, 2 days, ..., 3 months return horizons with parameter values as in Figure 1.2.

Gen. Freq.	1 day	2 days	1 week	2 weeks	1 month	2 months	3 months
1sec	-0.1520	-0.2130	-0.3282	-0.4449	-0.5902	-0.7158	-0.7641
1min	-0.1516	-0.2128	-0.3280	-0.4448	-0.5901	-0.7157	-0.7641
5min	-0.1500	-0.2117	-0.3273	-0.4443	-0.5898	-0.7156	-0.7640
10min	-0.1481	-0.2103	-0.3265	-0.4438	-0.5895	-0.7153	-0.7638
15min	-0.1461	-0.2089	-0.3257	-0.4432	-0.5891	-0.7151	-0.7637
30min	-0.1403	-0.2049	-0.3232	-0.4415	-0.5881	-0.7145	-0.7632
1x Daily	0.0000	-0.1066	-0.2628	-0.4009	-0.5627	-0.6993	-0.7525

cally be anticipated due to arising microstructure noise, the high–frequency return at the respective (lower) sampling frequency will appear to be skewed—which may also be due to genuine skewness. On the other side, Meddahi and Renault (2004) note the presence of the leverage effect at a low frequency implies the same feature at higher frequencies (their Proposition 3.4).

Multifactor GSV models: More analytical results

Although appealing for coarser frequencies, one volatility factor seems to oversimplify the volatility dynamics at high frequency implying almost a unit root for volatility shocks in contrast to the relatively more short–lived shocks reported in the high–frequency literature.[18] Therefore, consider model M with modified volatility dynamics

$$\ln(\sigma_{j,M}) = \ln(\sigma_{1,j,M}) + \ln(\sigma_{2,j,M}), \tag{1.3.41}$$

where each of the, here $k = 2$, volatility components is independent from one another and follows, as before in the set of equations (1.3.37),

$$\ln(\sigma_{k,j,M}) - \alpha_k = \phi_{k,M}\Big(\ln(\sigma_{k,(j-1),M}) - \alpha_k\Big) + \eta_{k,j,M},$$

with mean α_k and variance β_k^2. The random variables $(u_{j,M},\ \eta_{1,(j+1),M},\ \eta_{2,(j+1),M})'$ have a multivariate normal distribution with

$$\begin{pmatrix} u_{j,M} \\ \eta_{1,(j+1),M} \\ \eta_{2,(j+1),M} \end{pmatrix} \sim i.i.d.\ N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta_1\sigma_{\eta_1,M} & \delta_2\sigma_{\eta_2,M} \\ \bullet & \sigma^2_{\eta_1,M} & 0 \\ \bullet & \bullet & \sigma^2_{\eta_2,M} \end{pmatrix}\right). \tag{1.3.42}$$

The definition of the unconditional realized skewness holds and it is given as before by Eq. (1.3.38) but now covariance terms are driven by the dynamics of two volatility components with the unscaled return innovations $u_{j,M}$,

$$\begin{aligned} \mathrm{Cov}(r_{i,M}, r^2_{(i+j),M}) \;=\; & M^{-\frac{3}{2}} \sum_{k=1}^{2} 2\delta_k\sigma_{\eta_k,M}\phi^{(j-1)}_{k,M} \\ & \times \exp\left(\tfrac{1}{2}\sum_{k=1}^{2}\Big[6\alpha_k + (5+4\phi^j_{k,M})\beta^2_k\Big] + 4\prod_{k=1}^{2}\delta_k\sigma_{\eta_k,M}\phi^{(j-1)}_{k,M}\right), \end{aligned} \tag{1.3.43}$$

[18] Such as macroeconomic news or, more generally, market participants with different decision horizons, see Taylor (2005, Chapter 12) for a review.

from Eq. (1.A.10), and the unconditional variance for any i equals

$$\text{Var}\Big[r_{t,i/M}\Big] = M^{-1}\exp\Big(2\sum_{k=1}^{2}\big[\alpha_k + \beta_k^2\big]\Big), \tag{1.3.44}$$

following the same logic as before in 1.A.2, and with the direct dependence on M cancelling in the definition of skewness. Note that also the mean level of volatility, α_k, is just a scaling factor leaving the skewness estimator unchanged.

Realizing precision gains through imposing stepwise more structure

The above covariance terms between plain and consecutive squared returns are appealing in the sense that they map one–to–one into our definition of skewness, $\mathbf{URS}_t^{(N)}$. However, the correlation structure between a white noise process and a persistent autoregressive process might be better inferred for a certain transformation of the series, i.e. when the power function of the white noise series is near or here simply equal to one, and a positive real–valued power function of a consecutive random absolute return at lead order j which is smaller than one.[19] Denote these generalized statistical leverage terms as

$$\text{Cov}(r_{i/M}, |r_{(i+j)/M}|^{\lambda}) = \zeta_{j/M}(\lambda, \Theta),$$

where Θ denotes a parameter vector. The exact analytical expression for a generalized stochastic volatility model with (potentially) two factors, denote it by GSV2F, is stated in Eq. (1.A.10) in Appendix 1.A.2.

Consequently, it seems reasonable to make inference about the correlation structure when it is easiest from a statistical point of view, and the challenge is then to relate it back to the object of interest, i.e. when λ equals two, in order to reconstruct an estimator of realized skewness. There are different approaches available subject to the dependencies which show up in the data eventually.

A weakly specified semi–parametric estimator:

It turns out that there is a convenient scaling result when going from say $\zeta_{j/M}(\xi, \theta)$ to $\zeta_{j/M}(\lambda, \theta)$ when one of the volatility processes is orthogonal to innovations in the return process, i.e. say $\eta_{k=2,i+1} \perp u_i$, write $\zeta_{O,j}(\xi, \theta)$. Then, as outlined in Eq. (1.A.11), the

[19] Some evidence, although applying for functions of volatility in the standard SV model can be found in Taylor (2005, Figure 11.2) whose autocorrelation attain maxima for power functions below one for different values of the invariant standard deviation of the log volatility distribution. The same intuition should hold for the generalized SV model.

ratio

$$\frac{\zeta_{O,j/M}(\lambda,\Theta)}{\zeta_{O,j,M}(\xi,\Theta)}$$

only depends on the invariant standard deviation β_k of each of the here two volatility processes, their persistencies $\phi_{k,M}$, and the mean level of volatility $\alpha = \sum_{k=1}^{2}\alpha_k$—but it does not depend on the correlation between the (white noise) shock term generator in the return process and innovations in the volatility process.

Consequently, the $\zeta_{O,j}(\lambda,\Theta)$ can be written as a function of the lag j grouping constant terms together,

$$\begin{aligned}&\mathrm{Cov}(r_{i/M},|r_{(i+j)/M}|^{\lambda}) \qquad (1.3.45)\\ &= M^{-\frac{1}{2}(\lambda+1)}\underbrace{\varsigma(\lambda,\alpha_{k=1,2})}_{\text{const.}}\underbrace{C(\lambda,\beta_{k=1,2})}_{\text{const.}}\Big(\sum_{k=1}^{2}\delta_k\sigma_{\eta_k,M}\phi_{k,M}^{j-1}\Big)\exp\Big(\lambda\sum_{k=1}^{2}\beta_k^2\phi_{k,M}^{j}\Big),\end{aligned}$$

which follows from the results in Appendix 1.A.2, and furthermore also grouping multiplicative components together such that we relax the tight structure of the model, we can fit for say $\xi = \frac{1}{4}, \frac{1}{2}, \frac{1}{2}$, or 1

$$\begin{aligned}&\mathrm{Cov}(r_{i/M},|r_{(i+j)/M}|^{\xi})\\ &= M^{-\frac{1}{2}(\xi+1)}S(\xi,A)(C_{\xi,1,M}v_{1,M}^{j-1}+C_{\xi,2,M}v_{2,M}^{j-1})\exp\Big(\xi(B_1v_{1,M}^{j}+B_2v_{2,M}^{j})\Big)\end{aligned}$$

with $S()$ the estimator of $\varsigma()$, which includes A the estimator of the sum of the mean volatility levels, $\alpha_1+\alpha_2$, B_k denotes the estimator for β_k^2, and $C_{\xi,k,M}$ denotes the estimator for the product terms $\delta_k\sigma_{\eta_k,M}\times C(\xi,\beta_{k=1,2})$, $v_{k,M}$ the estimator for $\phi_{k,M}$, and use these coefficients for the case $\lambda = 2$ in the following way,

$$\begin{aligned}&\mathrm{Cov}(r_{i/M},|r_{(i+j)/M}|^{\lambda})\\ &= M^{-\frac{1}{2}(\lambda+1)}S(\lambda,A)C_{\lambda,1,M}(v_{1,M}^{j-1}+\frac{C_{\xi,2,M}}{C_{\xi,1,M}}v_{2,M}^{j-1})\exp\Big(\lambda(B_1v_{1,M}^{j}+B_2v_{2,M}^{j})\Big),\end{aligned}$$

for which there is only one unknown, $C_{\lambda,1,M}$, the estimator for the product term $\delta_1\sigma_{\eta_1,M}\times C(\lambda,\beta_{k=1})$; as $C_{\lambda,2,M} = C_{\lambda,1,M}\frac{C_{\xi,2,M}}{C_{\xi,2,M}}$. Then, $\Theta = (A, B_{k=1,2}, C_{\xi,k=1,2,M}, C_{\lambda,k=1,M}, v_{k=1,2,M})$ has eight parameters, with restrictions $0 \le v_{k=1,2,M} \le 1$, $B_{k=1,2} \ge 0$ for which we use information in the series of $\xi = \frac{1}{4}, \frac{1}{2}, \frac{1}{2}$, or 1 and $\lambda = 2$.

One could also argue that the case $\lambda = 2$ is uninformative and base inference solely

on absolute power functions ξ smaller than two. We can plug in B_k into the $C_{\xi,k,M}$, and evaluate say $C_{\xi,k,M}^{B_{k=1,2}}$ at $\delta_k \sigma_{\eta_k,M} \times C(\xi, B_{k=1,2})$, i.e. we use knowledge of the term $C(\xi, \beta_{k=1,2})$ and in effect only estimate the two products of correlation and volatility–of–volatility. Then, $\Theta = (A, B_{k=1,2}, C_{\xi,k,M}^{B_{k=1,2}}, v_{k=1,2,M})$ has seven parameters for which we use information in the series of $\xi = \frac{1}{4}, \frac{1}{2}, \frac{1}{2}$, or 1 only, imposing the same restrictions as before.

Note that while $C(\lambda, \beta_{k=1,2})$ does not depend on M, $C_{\xi,k,M}$ and $C_{\xi,k,M}^{B_{k=1,2}}$ depend on M such that we cannot address multiple frequencies, e.g. $M = 78, 39, 26, \ldots$, at the same time in the spirit of MIDAS regressions (Ghysels, Santa–Clara, and Valkanov, 2005). However, the dependency only occurs through the terms $\sigma_{\eta_k,M}$.

A more tightly specified prodecure based only on covariance terms:

Imposing even more structure, we could impose the model equality $\sigma_{\eta_k,M} = \beta_k(1 - \phi_{k,M}^2)^{\frac{1}{2}}$. Then, $\Theta = (A, d_{k=1,2}, B_{k=1,2}, v_{k=1,2,M})$ has seven parameters, where d_k estimates δ_k for which we use information in the series of only $\xi = \frac{1}{4}, \frac{1}{2}, \frac{1}{2}$, or 1, potentially also including $\lambda = 2$, in this more tightly parameterized version, as higher power function should tell us more about $\beta_{k=1,2}$, with additional restriction $-1 \leq \delta_{k=1,2} \leq 1$. Note that this procedure now allows for multiple frequencies as we can relates difference frequencies to one another—$\phi_{k,M}$ is scaleable in the frequency domain M.

Imposing moments of the unconditional distribution of volatility

Another way to address the expectation of the log volatility directly, while at the same time taking a step further towards parametrization lies in specifying more moments, in particular moments of the unconditional distribution of volatility. We can extend the result from 1.A.2, recall $\ln(\sigma_k^\theta) = \theta \ln(\sigma_k)$ with distribution $N(\alpha_k \theta, \theta^2 \beta_k^2)$, to k independent volatility factors, whose distribution is of course also independent of M. More information about $\alpha = \sum_{k=1}^2 \alpha_k$ and $\beta^2 = \sum_{k=1}^2 \beta_k^2$ can be obtained via $\mathrm{E}[|r_{j/M}|^\theta]$, in specific

$$\mathrm{E}[|r_{j/M}|^\theta] = M^{-\frac{1}{2}\theta} \mathrm{E}[|u_{j/M}|^\theta] \exp(\alpha\theta + \frac{1}{2}\theta^2\beta^2) \tag{1.3.46}$$

where

$$\mathrm{E}[|u_{j/M}|^\theta] = 2^{\theta/2} \pi^{-1/2} \Gamma((\theta + 1)/2),$$

for $\theta = 1, 2$, and possibly also for $\theta = \frac{1}{2}, \frac{3}{2}$, where $\Gamma()$ denotes the gamma function. These can serve as additional moment conditions and help in determining A and $B = \sum_{k=1}^2 B_k$.

A fully specified model with variance and covariance dynamics

The fully specified model should yield the highest precision, although being potentially more biased. In particular the interest lies in incorporating the persistence in volatility and variance in the simple form when $\theta = \lambda$, i.e.

$$\begin{aligned}\mathrm{E}\big[|r_{i/M}|^{\theta=1}|r_{(i+j)/M}|^{\lambda=1}\big] &= \mathrm{E}\big[M^{-\frac{1}{2}}|u_{j/M}|\exp(\lambda\phi_M^p\eta_{(j+1)/M})\big] \\ &= M^{-\frac{1}{2}} \times \mathrm{E}[A] \times \Big(\sqrt{2/\pi}\exp\big(-\tfrac{1}{2}\lambda^2\phi_M^{2p}\delta^2\sigma_{\eta,M}^2\big) \\ &\quad - \lambda\phi_M^p\delta\sigma_{\eta,M}\big(1-2\Phi(-\lambda\phi_M^p\delta\sigma_{\eta,M})\big)\Big),\end{aligned}$$

$$\begin{aligned}&\mathrm{E}\big[|r_{i/M}|^{\theta=2}|r_{(i+j)/M}|^{\lambda=2}\big] \\ &= M^{-2}\sum_{k=1}^{2}\big(1+\lambda^2\phi_k^{2(j-1)}\delta_k^2\sigma_{\eta_k}^2\big)\varsigma(\theta,\lambda,\alpha_{k=1,2})\exp\Big(\sum_{k=1}^{2}\frac{1}{2}(1+2\lambda\phi_k^j+\lambda^2)\beta_k^2\Big),\end{aligned}$$

as shown in Appendix 1.A.2.

1.4 Empirical Results

1.4.1 Data

Available financial instruments The empirical analysis is based on US equity–related market data. The primary interest lies in determining and differentiating between instantaneous, lead and lag effects at high frequency as well as, in case of the latter two, the joint decay pattern from which more important aggregational results for lower frequencies follow. When different tradable financial instruments are available for the same underlying such as the S&P 500 index, which is a mathematical construct, the most liquid one should be chosen in order to minimize the effect of market microstructure noise onto the inference results.

However, it is less obvious *a priori* which financial instrument to choose over which time period and to what extend the choice of instrument changes inference results and makes established results invalid. The large SP futures contract tends to lead the underlying cash market in terms of price discovery according to Hasbrouk (2003). Since then, newly available CME E–mini futures on the S&P 500 index, denoted ES, have quickly gained in popularity due to their smaller notional value compared to standard SP futures contracts. However, recently the cash market has also brought forward highly liquid exchange traded funds (ETFs) such as the State Street Global Advisors SPYDR S&P 500 ETF, which also tracks the level of the S&P 500 index.[20] A standard no–arbitrage condition, which depends on the respective interest rate, the index dividend yield, the bid–ask spreads, and potentially the depth of the order book (liquidity in the respective market) ties the prices of these instruments together, see Taylor (2015) for some new results in this direction.[21]

The grey area in Fig. 1.3, commencing with the 2003 H futures contract, marks the chosen cut–off after which we deem ES futures trading more informative and liquid than to SP futures trading following a large drop in the latter's intraday liquidity, and ES trading picking up pace (not shown here for brevity). The SP futures contract has become a (large) position holding instrument with the S&P 500 e–mini futures being the active (intra–)day trading vehicle.

[20]The State Street Global Advisors SPYDR S&P 500 ETF is the largest ETF on the S&P 500 index with 872.3mn shares outstanding vs. the iShares Core S&P 500 ETF with 333.7mn shares outstanding as of November 2015.

[21]Discrete rebalancing with a remaining cash position and accruing of management fees may result in a tracking error of the ETF versus the index construct but this is not expected to be important for intraday no–arbitrage arguments.

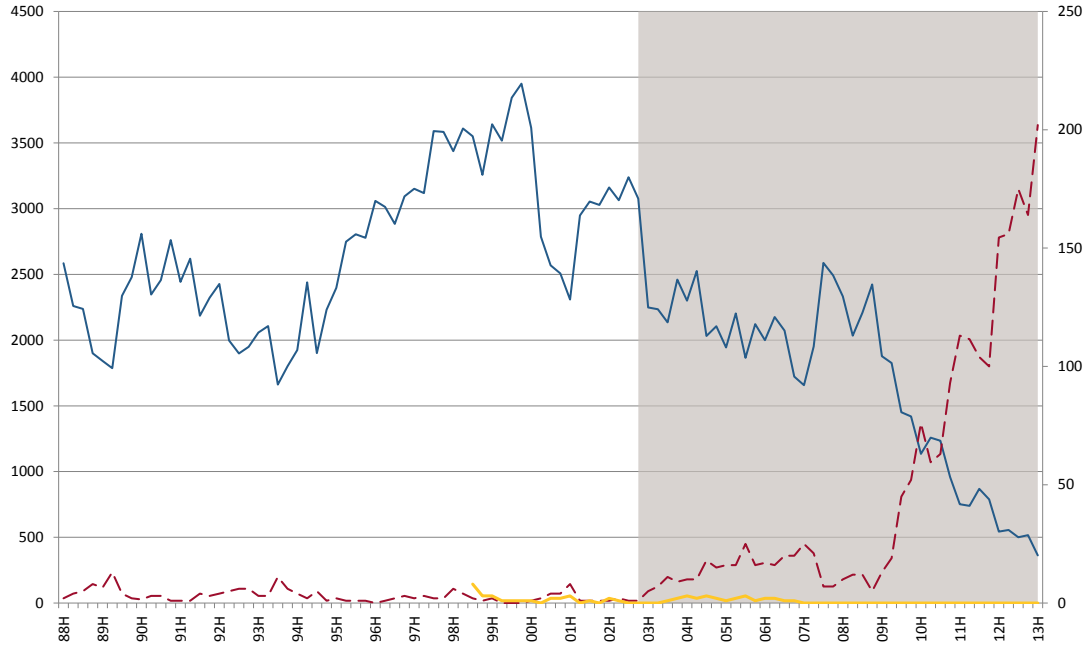
Table 1.2: Liquid S&P 500 financial instruments over time

SP and ES futures contracts data bought by Lancaster University from Price Data Inc. SPY data until end of 2013 downloaded from the TAQ database, which Lancaster University subscribes to. SPY data before 2003 is usually disregarded when inference is to be conducted at high frequency such as at 5–minute intervals and we here start using it along with the ES e–mini futures.

Financial Instrument	Start Date	End Date	Frequency
SPY ETF	Dec. 13th, 2002	Dec. 31th, 2013	Tick, second time stamp
ES futures	Dec. 13th, 2002	March 6th, 2013	Tick, minute time stamp
SP futures	Jan. 4th, 1988	Dec. 12th, 2002	Tick, minute time stamp

Table 1.3: CME S&P 500 futures contracts (SP and ES) liquidity measures, 1988–2013H

SOLID BLUE (LHS scale): Median number of transactions per ordinary trading day between 9.30 and 16.00 EST for the most liquid traded CME S&P 500 futures contract from 1988H to 2013H, where H, M, U and Z denote contracts with expiry in March, June, September and December respectively. DASHED RED (RHS scale): CME S&P 500 futures contract median number of one–minute intervals without any transaction taking place out of 390 minutes. SOLID ORANGE (RHS scale): ES futures contract median number of one–minute intervals without any price change taking place out of 390 minutes. Rollover dates can be found in Appendix 3.A, Table 1.B.3.



In order to validate and extend results previously brought forward in the literature, we concentrate on instruments used in those studies over the same time period but extend the analysis to cover newer instrument which show higher liquidity today.

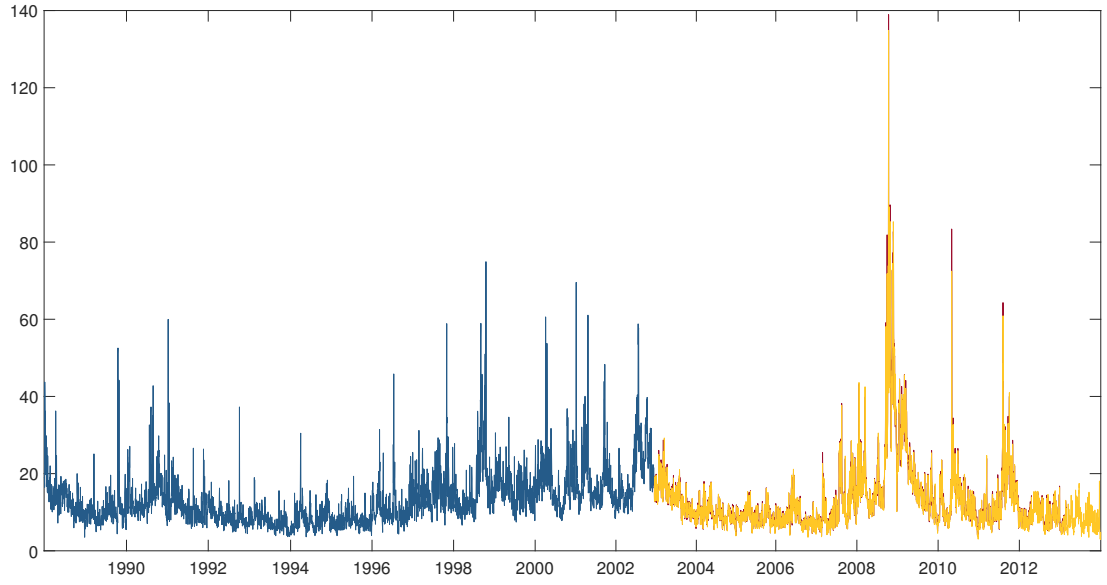
Trading Hours The primary NYSE cash market trading hours extend from 9.30 until 16.00 EST during which the underlying stocks are traded and recorded in the Trade and Quote (TAQ) database, which is why we limit ourselves to this daily observation window only. Moreover, opening returns are usually disregarded following the lengthy discussion in Andersen and Bollerslev (1997, their Appendix A.2), which is also in line with the related study on high–frequency return–volatility asymmetries by Bollerslev et al. (2006). From 2003 onwards we can further include trading days with early closure at 13.00 EST based on the NYSE special closing list (NYSE, 2013), see our Table 1.B.8. Nevertheless, it is possible to trade the SPY contract outside this time frame when we consider trades across all exchanges.

Sampling scheme The sampling scheme is throughout in "calendar" time rather than in "tick" time. These two schemes represent different views on the dependence structure between price innovations and observation times. In tick time, there is a one–to–one mapping between the observation times and the ex–post variation of the realized asset price path. In calendar time, sampling times are imposed exogenously. The latter assumes independence between the trade arrival process and changes in spot volatility which can largely be alleviated by sampling at a frequency lower than data is available. The resulting efficiency loss can be compensated by sub–sampling the estimator, as originally proposed in Zhang et al. (2005). Moreover, while tick time sampling could potentially diminish the systematic component in intraday volatility, it turns out to persist when sampling takes place at either all ticks or at a one second tick grid (Andersen et al., 2012).

Realized variation over time For illustrative purposes, we plot estimates of quadratic variation, here five–minute realized variance, sub–sampled every minute, over all 6,464 days excluding holidays in our sample from 1988 until 2013. Despite the obvious spikes during the financial crises, a visual inspection seems to suggest that volatility behaves more smoothly throughout the second half of the period while there were more sudden short–lived volatility busts in earlier years in tranquil market environments. Also the e–mini futures appear to show a higher level of variation on particular volatile days compared

Table 1.4: Realized variances (sub–sampled) of S&P 500 instruments, 1988–2013

For consistency purposes, we here show five–minute realized variance, sub–sampled every minute, in volatility p.a. $\times$ 100, instead of using a more efficient kernel–based estimator due to the availability of only minute time stamps for the SP futures. BLUE: SP futures. RED: ES futures. GOLD: SPY ETF.



to the SPY ETF, albeit the standard error of any estimator of quadratic variation or integrated variance might be very large on those days as well.

1.4.2 Intra–weekday and macroeconomic announcement day volatility patterns

We consider ordinary weeksdays (Mondays to Fridays) and the following scheduled U.S. macroeconomic announcements in Table 1.5, which occur during primary cash market trading hours from 9.30–16.00 EST, based on a comprehensive list of major scheduled U.S. macroeconomic announcements (Andersen, Bollerslev, Diebold, and Vega, 2007, their Table 4).

Real activity indicators such as non–farm payroll information and GDP reports fall outside these boundaries and will thus be reflected in the opening return. Once accounted for weekday and announcement day effects we do not expect the level and shape to vary significantly across days. However, high volatility regimes might bias the calculation such that it is advisable to account for the overall level of volatility as in AB and later studies such as Boudt et al. (2011).

Table 1.5: Selected U.S. macroeconomic announcements during primary cash market trading hours, 2003–2013

FOMC, MTS, and CC announcement observations (Obs.) over the sample period from January, 1st 2003 to December 31st 2013. When announcements occur on the same day, FOMC $\cap$ MTS = 7, FOMC $\cap$ CC = 2, MTS $\cap$ CC = 0, we always allocate it to the FOMC set as it constitutes the dominating effect. We further separate plain FOMC funding level announcement days from those on which the FOMC releases their projection material at 12.30 EST. In more recent years, FOMC announcements are released at 14.00 EST. Sources: Bureau of the Fiscal Service (BFS) available at www.fiscal.treasury.gov, and Federal Reserve Board (FRB) available at www.federalreserve.gov. Dates are listed in Appendix 3.A, Tables 1.B.5, 1.B.6, and 1.B.7. For the earlier period w.r.t. FOMC announcements for SP futures, 1988–2002, refer to Appendix 3.A, Table 1.B.4.

Announcement	Obs.	Source	Time	Frequency
Federal Funds Target Rate (FOMC)	97	FRB	∼14.15 EST	six–weekly
Consumer Credit (CC)	132	FRB	@15.00 EST	monthly
Treasury Statement (MTS)	132	BFS	@14.00 EST	monthly

To ratify the procedures outlined in Section 1.3.2, we exemplarily present detailed step–by–step results for, here, FOMC announcement days only.[22] The significant role of FOMC announcements for the equity premium and its negative price of risk have recently been highlighted in Lucca and Moench (2015) and Sarisoy (2015). Overall, we find that

- working with devolatized returns, Eq. (1.3.28), for the construction of the periocidity component eliminates the effects of occasional (randomly–timed) large returns in magnitude distorting the periodicity estimates at high, see Fig. 1.3 Panel A, and also at l frequencies,

- working with finely sampled increments such as 15–second returns, and then aggregating the squared quantities at a lower target frequency such as five minutes, increases the smoothness of the obtained intraday volatility pattern versus working with coarser frequencies directly, without having to rely on additional filtering rules as in, e.g. Boudt et al. (2011), see Fig. 1.3 Panel B. Furthermore,

    - applying a kernel to squared return increments in the local neighbourhood helps in determining the periodicity component more precisely when working at very high frequencies, which is in turn also important when a sub–sampling technique is to be applied looping over coarser frequencies at consecutive starting times,

    - a large kernel bandwidth choice spanning, e.g., five minutes, might lead to over

[22] Detailed inference results for other days are available upon request from the author.

smoothing abrupt changes in the diurnal volatility pattern; then, computed periodicity estimates overstate (understate) the true cyclicality component before (after) such scheduled news releases, or smooth out short–lived volatility busts such as the 10.00 am news release completely, see the iceblue and orange graphs in Fig. 1.3 Panel B,

- a kernel regression should be refrained from being conducted at highest available frequency (one second) as it is prone to bounce effects showing up in the periodicity components,
- choosing a sampling frequency is a trade–off between obtaining a more precise estimate (sampling more frequently within the coarse frequency bin) versus capturing more (high–frequency) noise variance not being present at lower frequency—inflating coarse low periodicity and deflating coarse high periodicity factors as can be inferred from Fig. 1.3 Panel B.

Thus, the choice for the NW kernel regression made here covers throughout squared 15–second returns, devolatized by the realized kernel estimator, with a 60–second bandwidth, which can be aggregated at any coarser target frequency of choice. Somewhat surprisingly, monthly treasury statements (MTS) and consumer credit (CC) release days do not exhibit very large announcement effects in terms of their intraday volatility patterns.[23] For the earlier time period from 1988–2002, with regard to SP futures contract trading, we thus consider only FOMC announcement days and ordinary weekdays for distinct intraday volatility cyclicalities with dates given in in Appendix 3.A, Table 1.B.4.

The intra–weekday periodicity factors in Fig. 1.4 obtained via the advocated choice of kernel regression estimation exhibit a sharp decline just before news releases such as at 10.00am which immediately shoot up on the time stamp of periodic news releases. At coarser frequencies the effect is captured with some delay even when compared to the five–minute aggregate kernel regression estimates based on finely sampled data; not shown here for brevity. It is further interesting to note certain spikes not captured using coarser frequencies such as on Wednesdays, which are likely to be related to roll–over and contract expiries in the derivative markets and which persist throughout sub–samples of the data. Vice versa, there appear to be several outliers in the periodicity factors when inference is conducted at coarse frequency not showing up in the 15–second squared return NW

[23]For brevity, results are not reported for these two cases. Available from the author upon request.

Figure 1.3: Smoothing the intraday volatility periodicity factor by example of FOMC announcement days, SPY ETF 2003–2013

PANEL A Factors based on Eq. (1.3.29) for plain (GREEN) and devolatized (BLUE) 15–second SPY returns, Eq. (1.3.28) using realized kernels. **PANEL B** NW Epanechnikov kernel regression based on Eq. (1.3.31) using previously filterd devolatized returns. BLUE (preferred choice): 15–second increments, bandwidth $h = 60$ seconds. RED: 3–second increments, $h = 60$ seconds. ICEBLUE: 15–second increments, $h = 300$ seconds. ORANGE: 3–second increments, $h = 300$ seconds. GREY: 300–second devolatized increments, Eq. (1.3.29) estimate. The procedure based on the automatic Silverman bandwidth selection oversmooths. X–AXIS: Denotes the cash market trading hours from 9:35 EST to 16:00 EST in 15–second increments. BLACK DOTTED: Benchmark constant periodicity factor of one across the trading day. Realized kernels are constructed using 20–minute RV sub–sampled very minute as integrated quarticity estimates and one–minute RV sub–sampled every 15 seconds as noise variance estimate.

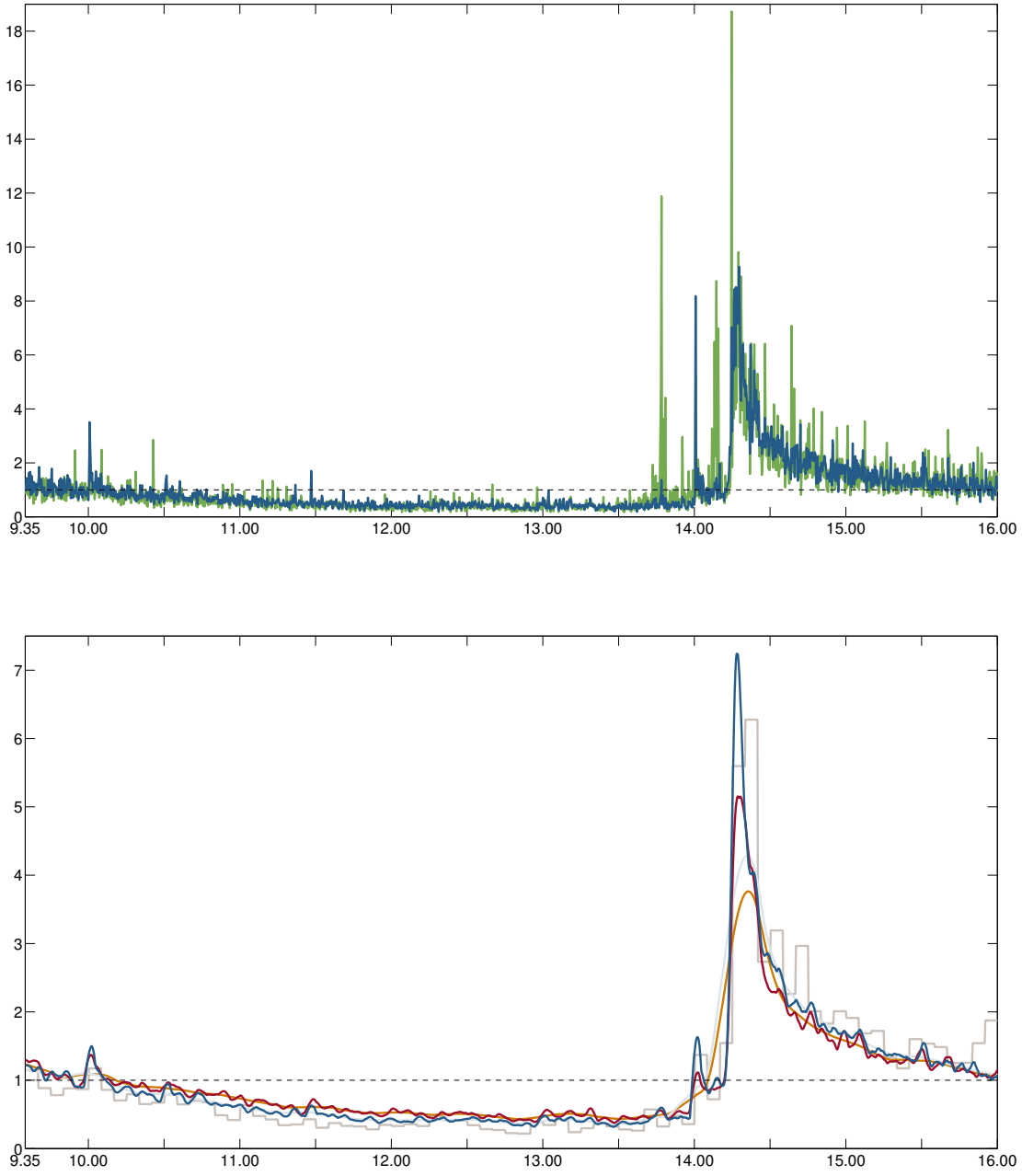
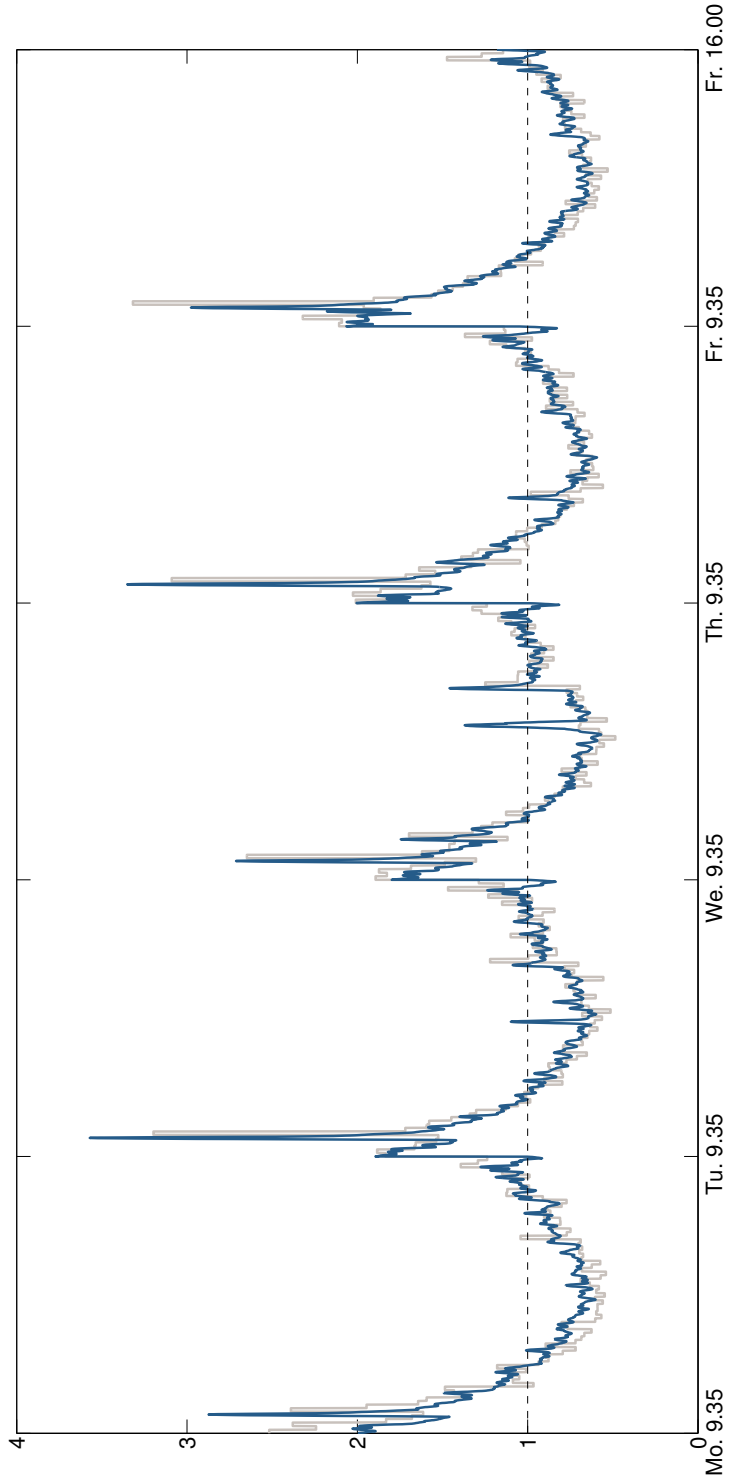


Figure 1.4: Intra–weekday volatility periodicity factors, SPY ETF 2003–2013

NW Epanechnikov kernel regression based on Eq. (1.3.31) using devolatized returns on the SPY contract from the TAQ database. BLUE (preferred choice): 15–second increments, bandwidth $h = 60$ seconds. GREY: 300–second devolatized increments, TX periodicity estimates, Eq. (1.3.29). X–AXIS: Denotes the cash market trading hours Mondays to Fridays from 9.35 EST to 16.00 EST in 15–second increments. BLACK DOTTED: Benchmark constant periodicity factor of one across the trading day. Realized kernels are constructed using 20–minute RV sub–sampled very minute as integrated quarticity estimates and half–a–minute RV sub–sampled every 15 seconds as noise variance estimate. No. of observations excl. announcement days and early closures: Mo. 471, Tu. 488, We. 467, Th. 496, Fr. 475. Estimates are virtually identical when the so–called Flash Crash days 2010–05–06, the following day 2010–05–07, and 2013–04–23 are removed.



regression. As the latter methods yields 20 times more data points (it is less affects by single outliers) plus there is the kernel smoothing effect, it is the preferred choice.

The intraday volatility factors when the NYSE closes regularly at 1.00pm, days provided in Appendix 3.A Table 1.B.8, exhibit two hikes just after the offical NYSE close at 1.00.45pm (at a level of 2.09×) and at 1.06.45pm (at a level of 2.98×) with a low of 0.88 in between when multipliers are defined to average one between 9.35am and 1.15pm. We also cover these half trading days yielding 44, 20, 6, and 4 return increments per day at 5–, 11–, 35–, and 55–minute sampling frequency.[24]

The same analysis in terms of intraday volatility cyclicalities has been conducted for the SP futures and the ES e–mini futures which we choose not to present here for clarity of exposition but are available from the author upon request. As is the case for the SPY ETF weekday cyclicalities, it turns out that also for the SP futures the difference from one day to another is mainly driven by the magnitude of the volatility surge around 10.00am, with different news being released on different days of the week.

1.4.3 The Class of Non–parametric Realized Skewness Estimators

Remark: Numerical truncation error. As all estimators of realized skewness employ sums of power functions of high frequency returns, which are very small in magnitude, the computational accuracy of an estimator, i.e the accuracy of the sum of floating point numbers, is limited by the numerical truncation error.[25] The error grows with the number of input elements N such that the error of a sum of N standard normal variates, `sum(randn(N,1))`, is about `eps*(N/10)`, where `eps` denotes the floating–point relative accuracy, i.e. the distance from 1.0 to the next larger double–precision number, which is, `eps` $= 2^{-52} \approx$ 2.2204e-16. Sums, which enter the covariance terms are numerically more precise when calculating sums with error compensation such as when using the Knuth or Knuth2 algorithm, see Knuth (1998), which yields approximately 15 and 30 more valid digits equivalent to a 128 and 190-bit float. However, even in the latter case covariance estimates between very finely sampled high frequency returns and high frequency volatility proxies at different lags differ only by 1 to 100 × e-25 when covariance estimates are

[24] In case of 35–minute sampling, we let the day end at 1.05pm rather than at 1.40pm and hope that most of the variation has already been captured. Periodicity factors then average one between 9.35am and 1.05pm.

[25] As an extreme example, `sum([-1e16,1,1e16])` in Matlab returns zero instead of one, while `sum([-1e16,1e16,1])` returns one.

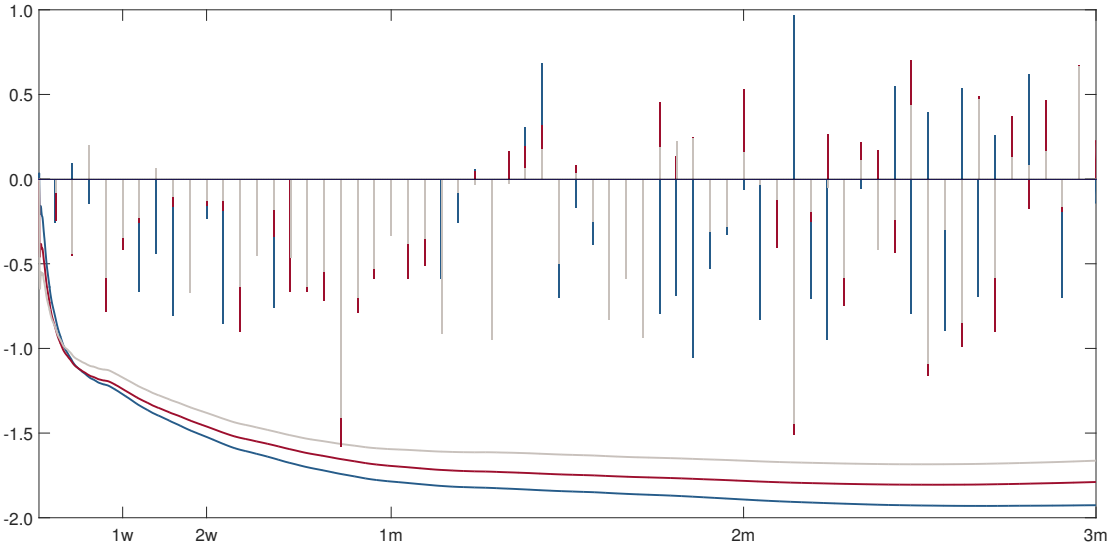
between 1 and 100 × e-12, and can thus be safely ignored.

Unconditional Realized Skewness

We have (jointly) reliable data available for the E–mini futures contract and the SPY and DJIA ETF over the period Dec., 12th 2002 – Mar., 6th 2013. Figure 1.5 shows the most straight forward standard realized skewness estimator for the E–mini futures, the SPY contract and the Dow Jones industrial average ETF vis–à–vis their classical skewness estimators. All index instruments share the same pronounced negative realized skewness pattern with the DJIA being slightly less pronounced in magnitude.

Figure 1.5: Realized Skewness vs. classical skewness over extending horizons

SOLID LINES: Realized skewness as a function of the horizon based on the aggregation formula, Eq. (1.3.13) for ES (BLUE), SPY (RED), and DJIA (GREY) with distinct intraday volatility pattern adjustment for weekdays and announcements days. The denominator of the ratio is based on the realized kernel methodology. BARS: Classic skewness estimator for respective series and horizon (for integer number of days only). One week (1w) contains 5 × 77 increments up to 4,651 increments for the three month (3m) period. Data spans from Dec., 12th 2002 – Mar., 6th 2013.



It is also apparent that the classical skewness estimator behaves very erratic, in particular when less and less data points become available as the return horizon increases, even changing sign when the return horizon is altered by as little as one day.

The efficiency gain of the realized skewness estimator rests on the use of intraday data, moreover relying on its inherent dependence structure, it is a richer information set, and excluding noise terms which are zero on average. This explains why we have a nice and

clean accumulation over time.

Remark: Non–centralized realized skewness. We carry out the same experiment for the non–centralized realized skewness estimator, Eq. (1.3.16), yielding the same albeit slightly less pronounced pattern with differences being at -0.0118 (1d), -0.0175 (2d), -0.0280 (1w), -0.0388 (2w), -0.0566 (1m), -0.0789 (2m), -0.0918 (3m) for the horizons stated in parenthesis.[26]

Remark: Cubic HF returns. While we assume the expectation of the third power of high–frequency returns to be zero, i.e. ultimately the instantaneous skewness at sampling frequency such as five minutes, sample estimates thereof are very noisy for the same reasons as outlined before and removing one large observation can change the sign of the sample estimate. The instantaneous skewness at 5–minute frequency for the SPY ETF equals -0.2448 and its contribution to long return horizon skewness relative to the sampling frequency vanishes at a fast rate yielding a contribution to realized skewness of hourly returns of only -0.0704 and -0.0277 at the daily horizon, and is of negligible empirical relevance thereafter. Equivalent estimates for the ES e–mini futures are -0.3644, -0.1047, -0.0413, and -0.0509, -0.0143, -0.0056 for the DJIA ETF.

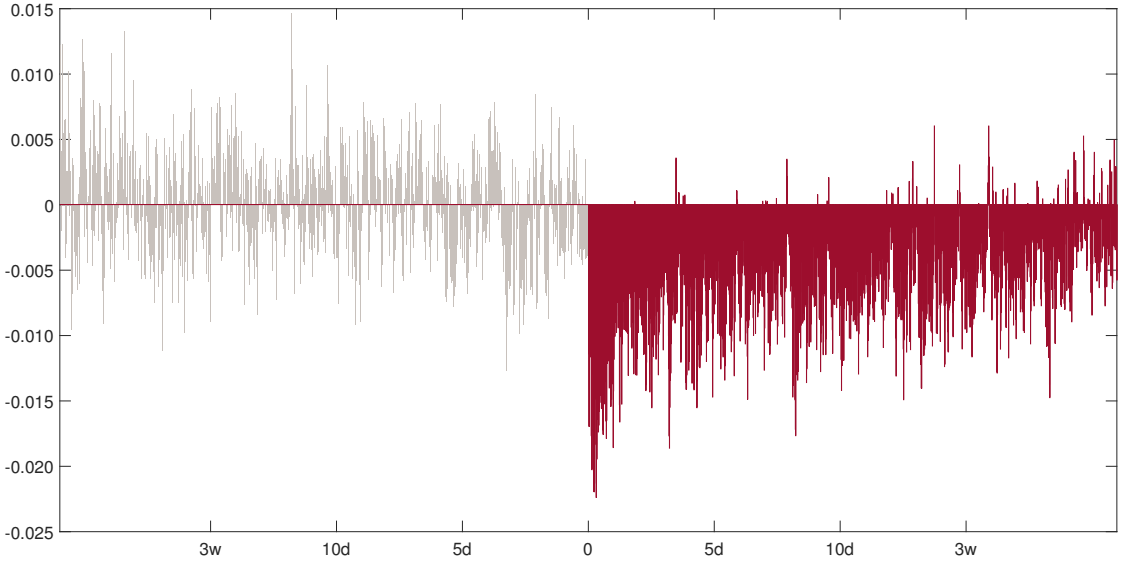
Remark: Volatility feedback effect. Incorporation of the volatility feedback *noises up* the realized skewness estimator when we are in fact trying to isolate the signal. Note that including all cross–product terms of the cubic expansion of the sum of high–frequency returns, as elaborated upon in Section 1.3.1, Eq. (1.3.6), should be equivalent to and thus be as noisy as the classical skewness estimator. Figure 1.6 shows the asymmetric volatility effect and the volatility feedback effect in terms of correlations when the absolute return is used as a simple volatility proxy. In the subsequent analysis, we will thus follow the literature and focus on the asymmetric volatility effect as the building block of long–horizon skewness relative to the sampling frequency.

Remark: The use of multi–power variations. Using squared returns as a proxy for the local level of volatility directly extends to the definition of realized skewness but it is also very noisy. In this case, the equivalent asymmetric volatility feedback effect with

[26]Here, results are for the SPY ETF with a sample mean of only 0.83% p.a. over the period considered, otherwise the impact would be more severe.

Figure 1.6: Asymmetric volatility and volatility feedback correlation at +/- 1 month lead–lag order

RED: Asymmetric volatility effect in terms of correlation between return and subsequent volatility innovation at up to one month lead order, i.e. $\underset{j>i}{\text{corr}}[r_{t,i}, |r_{t,j}|]$. GREY: Volatility feedback effect in terms of correlation between return and preceding volatility innovation at up to one month lag order, $\underset{j<i}{\text{corr}}[r_{t,i}, |r_{t,j}|]$. One week (1w) contains 5×77 increments up to 1,617 increments for the three month (1m) period. Based on 5–minute SPY ETF data from Dec., 12th 2002 – Mar., 6th 2013. The equivalent figure for the ES e–mini futures and the DJIA ETF look virtually identical and are available from the author upon request.



regard to the right–hand side of Figure 1.6 is also much less persistent, i.e. it reverts faster towards zero, as shown theoretically based on the parametric SV model specification. Instead of using squared or absolute returns as simple volatility proxies in Figure 1.6, we revert to using multi–power variation measures as elaborated upon in Section 1.3.1 , Eq. (1.3.19f.), in specific the tripower–variation, bi–power variation and the MedRV estimators, which can be used more directly in place of squared returns entering the asymmetric volatility realizations in a non–parametric way. Note over long time spans using these MPV–type estimator produces virtually the same, i.e. insignificantly different results as can be inferred from Figure 1.C.1 in Appendix 1.C, which compares in the upper panel the standard realized skewness estimator against those using more sophisticated MPV measures for the same asset for the SPY ETF over a return horizon of three months. In the lower panel, we compare the RSkew measure based on the RV–type against the RSkew measure based on tripower variation (TPV). While RV–type based estimators

produce slightly different estimator for realized skewness at horizons of up to one day only, the TPV–type RSkew measure produces remarkably similar estimates for the two different assets considered (SPY ETF vs. ES E–mini futures in the lower panel only). These become more important when considering shorter time spans of data, see the next subsection further down below.

Remark: Optimal sampling frequency. Within the framework of the realized methodology, the optimal sampling frequency has been well researched and established for the *natural* realized variance estimator, which ranges from around one to two minutes for highly liquid index tracking investment vehicles to around five to 17.5 minutes for liquid and less liquid stocks.

In terms of advisable lower sampling interval bounds for the univariate realized skewness case there are no non–synchronicity issues. If co–skewness is considered between the market return and its impact on future stock return volatility, we can revert to popular lower bounds in the co–jump detection and realized beta literature to mitigate the non–synchronicity effects. These are ranging from eleven minutes (Gilder, Shackleton, and Taylor, 2014) to 22.5 (Bollerslev and Todorov , 2010) and 25 minutes (Patton and Verardo, 2012) up to 75 minutes (Bollerslev, Li, and Todorov, forthcoming). The optimal sampling frequency for realized variance (RV) has historically been based on volatility signature plots, originated in Fang (1996), for which as the length of the sampling interval increases, the microstructure noise bias of RV remains constant or is absent in the best case, and more recently based on the model confidence set methodology as employed in (Liu, Patton, and Sheppard, 2015). We will leave the latter case for future research and focus on the most straight forward sampling frequencies when the trading days is considered over the 385 minutes from 9.35–16.00 EST. Further, increasing the length of the return interval from five minutes upwards is expected to help eliminate microstructure noise issues related to zero return increments caused by the discrete price grid and infrequent trading, bounce back effects.

While the absence of autocovariance is clearly important, it is secondly also important to, loosely speaking, allow some time for the market to respond to negative and positive news as can also be inferred from the time lag in Andersen, Bondarenko, and González–Pérez (2015), although their asymmetric volatility effect is defined based on the correlation

between E–mini future contract returns and high–frequency option data.[27] Thirdly and most importantly, realizations of the asymmetric volatility effect depend on the sign of the return innovation whose magnitude must be relevant and large enough to move the level of volatility over subsequent periods. While near the in–fill limit the effect can hardly be determined or is not present at all as assumed in most theoretical non–parametric frameworks (Mykland and Zhang, 2009), sampling not frequently enough might render the estimation of these potentially very short–lived effects at low lead–lag order infeasible. The last two points above are linked but different: The former describes the presence of the time lag and the latter allows for an accumulation of the lagged effect over very short periods of time.

In Table 1.6 we present standard and MPV–type RSkew estimates over horizons ranging from one day up to three months based on the long time span of SPY ETF data for different sampling frequencies. Overall, all estimators perform remarkably well in terms of estimates produced lining up nicely across frequencies and across estimators.

Taking 5–minute standard RSkew estimates as the benchmark, the magnitude of skewness is decreasing with sparser sampling when the return horzion is as short as one day only—which is anticipated based on the theoretical results presented earlier, and it is generally advisable to sample frequently relative to the return horizon of interest for which an RSkew estimate shall be obtained.

In terms of MPV–typed RSkew measures, the MedRV–based RSkew estimator produces slightly more negative estimates for dense sampling frequencies but slightly less pronounced estimates when the return horizon to be considered is long under the sparse 55–minute sampling scheme. Leaving the question of significance in differences aside, at dense sampling frequencies this estimator might be less prone to outliers while at sparse sampling the estimator misses the relative immediate impact related to the asymmetric volatility effect by construction which is why we do not pursue investigating further into this direction.

Note that when we include all terms, the volatility feedback effect and the instantaneous skewness of high–frequency returns, the standard RSkew estimates at longest horizon considered (here three months) are down by 0.1%, 3.8%, 6.3%, and 9.4% as the sampling

[27] In a parametric setting (Yu, 2005) points out that the asymmetric volatility effects appears well to be lagged rather than instantaneous when estimating SV–type models.

Table 1.6: Standard and MPV–type RSkew when altering the sampling frequency

Standard and MPV–type RSkew estimates over horizons ranging from one day (1d) up to three months (3m) based on the SPY ETF data from Dec., 12th 2002 – Mar., 6th 2013 with distinct intraday volatility pattern adjustment for weekdays and announcements days. The denominator of the ratio is based on the realized kernel methodology at a 15–second frequency. Classic skewness estimates obtained via sampling at the respective return horizon of interest directly for integer number of intervals (observations) only, thus N_{floor}. The four panels provide estimates when sampling takes place at 5–minute, 11–minute, 35–minute, and 55–minute intervals. For 5–minute and 11–minute frequency the the MPV–type RSkew estimates are calculated based on the algorithm illustrated in Figure 1.1. As for the 35– and 55–minute frequency this would translate in a serious loss of data, we construct MPV–based spot measures from data spanning consecutive days as well.

Horizon	**1d**	**2d**	**3d**	**1w**	**2w**	**3w**	**1m**	**6w**	**2m**	**3m**
					5–min					
RSkew	-0.88	-1.08	-1.16	-1.24	-1.46	-1.60	-1.69	-1.74	-1.78	-1.79
$RSkew_{TPV}$	-0.74	-0.96	-1.07	-1.19	-1.42	-1.58	-1.68	-1.74	-1.79	-1.81
$RSkew_{BV}$	-0.80	-1.01	-1.10	-1.20	-1.42	-1.56	-1.66	-1.71	-1.75	-1.76
$RSkew_{MedRV}$	-0.90	-1.09	-1.16	-1.25	-1.49	-1.64	-1.74	-1.79	-1.82	-1.82
					11–min					
RSkew	-0.65	-0.94	-1.05	-1.17	-1.39	-1.54	-1.63	-1.67	-1.70	-1.73
$RSkew_{TPV}$	-0.70	-0.95	-1.10	-1.24	-1.45	-1.61	-1.70	-1.72	-1.76	-1.78
$RSkew_{BV}$	-0.65	-0.94	-1.07	-1.20	-1.40	-1.56	-1.64	-1.66	-1.69	-1.71
$RSkew_{MedRV}$	-0.71	-0.98	-1.11	-1.23	-1.43	-1.59	-1.67	-1.69	-1.73	-1.77
					35–min					
RSkew	-0.63	-0.89	-1.04	-1.18	-1.42	-1.54	-1.61	-1.63	-1.66	-1.69
$RSkew_{TPV}$	-0.68	-0.83	-1.00	-1.14	-1.38	-1.51	-1.57	-1.58	-1.63	-1.66
$RSkew_{BV}$	-0.78	-0.96	-1.12	-1.25	-1.44	-1.56	-1.62	-1.62	-1.66	-1.69
$RSkew_{MedRV}$	-0.87	-1.04	-1.17	-1.26	-1.45	-1.55	-1.59	-1.60	-1.66	-1.70
					55–min					
RSkew	-0.46	-0.83	-1.01	-1.18	-1.48	-1.65	-1.76	-1.79	-1.85	-1.88
$RSkew_{TPV}$	-0.55	-0.80	-0.91	-1.09	-1.33	-1.47	-1.58	-1.66	-1.74	-1.78
$RSkew_{BV}$	-0.52	-0.79	-0.92	-1.09	-1.35	-1.50	-1.61	-1.68	-1.76	-1.79
$RSkew_{MedRV}$	-0.50	-0.71	-0.79	-0.96	-1.19	-1.31	-1.43	-1.52	-1.59	-1.63
$Skew_{classic}$	-0.24	-0.45	0.12	-0.41	-0.16	-0.66	-0.23	-0.43	0.53	0.23
N_{floor}	2,523	1,261	841	504	252	168	120	81	60	40

frequency decreases from 5– to 55–minutes, i.e. while there is almost no effect present from introducing the noise terms at high frequency, realizations appear to be non–zero when sampling very sparsely, as it is then less feasible to disentangle the effects happening at higher frequencies which are only seen in aggregation.

Remark: The large SP futures contract For the large SP futures contract we consider the same time span in terms of number of days/number of increments as we have data available for the more recent period. This results in the time span from Sept., 18th 1992 – Dec., 11th 2002 (SP_{II}). And we also consider the data from the earliest period onwards for which we have data available, from Jan., 5th 1988 – Mar., 9th 1999 (SP_I). The two time spans are overlapping by approximately 2/3, nevertheless from Figure 1.C.2 we can see that the asymmetric volatility effect appears to be stronger but less persistent resulting in more pronounced skewness at shorter horizon and less in magnitude at longer horizons. Recall that the Black Monday on Oct., 19th 1987 just precedes the earlier period. Noteworthy are the different magnitudes at -0.65 and -0.89 for the earlier and later period respectively for monthly and longer return horizons. This compares to a staggering -1.7 (approximately) for monthly returns based on the latest period for the SPY ETF. These differences can also to some extend be visually inferred from Figure 1.4 in which sudden spikes in volatility before 1999 appear to occur more frequently and are less persistent than in later periods. Note there are almost no volatility busts in the period 2003 – 2007 and after 2012 unlike in earlier periods. The average realized volatility p.a. over the three periods are 12.96% (SP_I), 15.32% (SP_{II}), 16.01% (ES), 16.16% (SPY). But the net impact on RSkew estimates is not at all straightforward as the (quadratic) variation of return affects both numerator and denominator. Further, trading at the end of the 1980s was less automated and took place primarily through the open outcry market so results might not be straightforward to compare.[28]

Realized Skewness of Dow Jones stocks

In this section, we set out to shed light on the realized skewness of individual stocks. We analyze all stocks which were at some point constituent of the Dow Jones index over the same time period as in the previous section. Focusing on these stocks only largely mitigates the impact of microstructure noise onto our inferences results as these are among the most

[28] I'm thankful to Ingmar Nolte for pointing this out.

liquid US stocks.

In specific, for each of the 36 stocks in our market representative sample, we analyse the contribution of the three components asymmetric volatility (AVL), volatility feedback (VFB) and the instantaneous skewness (IA) to the realized skewness of enlargening return horizons, which are long relative to the sampling frequency. Sampling at a high intraday frequency leads to more precise estimates of the different effects at work which might otherwise be only observed jointly at a coarser frequency and are then potentially offsetting each other. This can be identified when at a relatively high sampling frequency such as 5–minutes, returns are symmetric but appear to be either positively skewed when VFB>AVL, or negatively skewed when AVL>VFB at a coarser intraday sampling frequency, keeping in mind that the expected instantaneous return at high frequency is usually assumed symmetric which has been empirically documented for price discontiunities as well, see the relevant section in the literature review. Sampling too densely, below five minutes, often renders the identification of the effects infeasible.[29]

We consider the impact of the sampling frequency, 5–, 11–, 35–, and 55–minutes, onto the realized skewness estimates for nine different return horizons, 1–, 2–, 3–day(s), 1–, 2–3–week(s), 1–, 2–, 3–month(s) when using different proxies for the spot variation measure, where the latter ones are, as before, of type realized variance (RV), tripower variation (TPV), bipower variation (BV), and median realized variance (MedRV).[30]

While for aggregate stock market returns there appears to be only the asymmetric volatility effect at work in terms of creating (negative) skewness in the unconditional return distribution over extending return horizons, the notion for single name stocks is rather different and we establish the following six categories into which we group the selection of stocks. In order to be classified into a certain category the respective effects have to be consistent throughout all of the 16 cases considered, i.e. 4 frequencies $\times$ 4 different RSkew (multi–)power variation measures, for each of the 36 individual stocks.

Category A: Negative AVL, no or almost no VFB.

A-1: Negative AVL, no VFB. The following five stocks CAT, GE, MO, IBM, and DD (see Fig. 1.C.4 – 1.C.9 in Appendix 1.C) exhibit strong asymmetric volatility effects resulting in large negative realized skewness at all return horizons ranging from one

[29] Not reported for brevity.

[30] We do not report results for the bipower variation–based RSkew estimator as it resembles very close similarity with the TPV–based RSkew estimator.

day to one quarter without consistent patterns in terms of volatility feedback effects, which are changing sign around zero across measures and frequencies.

A-2: Negative AVL, Z–shaped VFB. The stocks categorized in this subsection, i.e. MSFT, KO, KRFT, T, MRK, BAC (Fig.1.C.10 – 1.C.15), are labeled as featuring a Z–shaped volatility feedback effect. In contrast to A-1, these stocks appear to show a positive contribution of VFB to the realized skewness of short horizon returns spanning less than two weeks time which turns into a negative contribution at monthly and quarterly return horizons throughout measures and frequencies examined. This translates into shocks to the level of current volatility being positively correlated with future returns at the short end and negatively correlated at the long end of the horizons considered. We could interpret this as investors rewarding risk (or risk taking) of companies such as Microsoft, Coca Cola, and Kraft Foods in the short run, while requiring a premium for this in the long run.

Category B: Negative AVL dominating positive VFB at long return horizons

B-1: Negative AVL larger in magnitude than positive VFB. Stocks MMM, CSCO, HON, PFE, UNH, and INTC (see Fig. 1.C.15 – 1.C.22) have consistently negative AVL and consistently positive VFB effects with AVL being the dominating factor. In category B-1, the sum of the three individual components results in consistently negative realized skewness for long return horizons and produces either negative estimates at the short end or these are scattered around zero such that we summarize RSkew ≤ 0 for the stocks in this subcategory at all horizons considered.

B-2: Negative AVL, positive VFB, Z–shaped RSkew. Here, the difference to the previous subcategory is given by the larger magnitude of VFB versus AVL effects at very short horizons only resulting in combined positive RSkew estimates for horizons shorter than one week with, however, AVL being the dominating factor at longer horizons resulting in a Z–shaped RSkew pattern (see Fig. 1.C.23 – 1.C.28 which presents estimates for MCD, UTX, VZ, BA, and JNJ).

Category C: RSkew not consistently present. For only one stock, HPQ (Fig. 1.C.29), we attribute RSkew as not being consistently present with AVL, VFB and IA all semingly fluctuating randomly in a narrow bandwidth around zero.

Category D: Positive VFB dominating negative AVL at all horizons. Category D builds the counterpart to category B, with AVL and VFB still being negative and positive respectively but with VFB dominating AVL effects in terms of being more pronounced in magnitude at all horizons. This holds for stocks CVX, HD, WMT, DIS, AXP, and JPM; see Fig. 1.C.30 – 1.C.35 in the appendix.

Category E: AVL and VFB being both negative. This category groups stocks AA, TRV and C all of which feature both negative AVL and negative VFB effects, see Fig. 1.C.36 – 1.C.38. In terms of economic reasoning, for Citigroup (C), a distressed investment banking stock, we might deduce that any increase in risk, i.e. changes in the current level of volatility are negatively correlated with future return innovations, or in other words there is a negative premium attached to increases in the level of volatility of distressed stocks which accumulates in subsequent periods. The negative VFB effect also appears to be present for Alcoa (AA), a commodity–related stock with primary activities as a lightweight metal producer and manufacturer of products thereof. Here, negativity in VFB might be related to volatile commodity prices which induces subsequent lower returns. The company has undergone a corporate turnaround in 2007 shifting its focus on higher–value added products. Note that the effects appear to be very short–lived as these are hardly present at the 55–minute sampling frequency. The Travelers Companies (TRV) is a writer of U.S. commercial property casualty and personal insurance, previously owned by Citigroup, alleged of anticompetitive practices, which might explain negative VFB effects, however, these are not as stable throughout measures and frequencies as for the other two stocks.

Category OTHER:

GM: Potentially multiple correlated volatility processes with different signs. General Motors (GM), Fig. 1.C.39, features negative AVL effects and thus negative RSkew estimates across measures and frequencies for three week and longer horizons. However, for shorter horizons there appears to be large positive AVL effects when sampling densely (5–35 minutes) which are almost not present at the coarsest frequency considered, i.e. finely sampled positive returns appear to increase (i.e. are on average positively correlated with) volatility at very short lead lad orders (below one hour), cumulative returns over 55–minutes then show the typically encountered

negative AVL patterns from the two day return horizon onwards which are increasing in magnitude over longer horizons. Thinking in a parametric way, a credible model, which is capable of explaining the observed RSkew patterns, should consequently feature at least two volatility processes, one highly persistent, low volatility–of–volatility process which is negatively correlated with return innovations, and one which is positively correlated with higher vol–of–vol and low persistency parameter. Note that GM filed for a government–backed bankruptcy under Chapter 11 on June 8th, 2009 from which it emerged through the Troubled Asset Relief Program (TARP) of the US Treasury and further measures. As trading of the stock was consequently suspended, we do not have data after this point and the effect of the rescue program are consequently not identifiable, while investors' speculation of a potential bailout prior to the actual filing might well be present in the data, which might give some economic intuition of why large positive returns could increase future volatility at short lead lag order.

AIG: Negative RSkew at long horizons, short–lived positive VFB effects at short horizons. A similar picture in terms of the combined (AVL+VFB+IA) RSkew estimates in Fig. 1.C.40 emerges for AIG which are positive for horizons up to three weeks for the two finer sampling frequencies considered and increasingly negative thereafter for all sampling frequencies. However, unlike for GM, the positive combined RSkew estimates come from short–lived positive VFB effects which are most pronounced at the 11–minute frequency and not present or identifiable at the coarsest frequency.

XOM: Potentially multiple negatively correlated volatility processes with positive VFB. Exxon Mobile (XOM), see Fig. 1.C.41, has negative AVL and positive VFB effects resulting in combined negative RSkew estimates up to three week horizons and slightly positive RSkew estimates thereafter in a narrow bandwidth around minus/plus a half. At the 55–minute sampling frequency, the contribution of AVL to RSkew is slowly increasing in magnitude from near zero whereas for finer sampling frequencies such as 5–minutes the contribution of AVL to RSkew starts high (near one) and is declining thereafter towards a half which points to the existence of at least two volatility factors with one of them being more negatively correlated with return innovations than the other and exhibiting very low persistence.

Dispersion of RSkew estimates across stocks In order to compare RSkew estimates across stocks we analyze the dispersion of these in terms of boxplots, grouping stocks for each of the nine horizons ranging from one day to three months together, see Fig. 1.7. We then analyze the dispersion for different horizons in terms of their median values and their interquartile ranges. Note that interquartile ranges are constructed based on winsorized data as explained in the figure notes with classified outliers being plotted individually; the standard technique for presenting boxplots.

The first three quartiles of $\text{RSkew}_{\text{AVL}}$ estimates across stocks are all in the negative domain from the one day horizon onwards and increasing in magnitude with the medians of one day, one week and two month horizons being significantly different from one another based on the boxplot notches with standard assumptions.

The upper three quartiles of the $\text{RSkew}_{\text{VFB}}$ estimates across stocks are in the positive domain up to and including the two week return horizon.

The combined (AVL + VFB + IA) RSkew estimates are centered around zero at the one day horizon with the first three quartiles being in the negative domain from the one week horizon onwards. These results for the one day horizon stand in clear contrast to the strongly negative RSkew estimates of index returns at short horizons. The strong index asymmetry is determined by the dependence structure among the constituent stocks and cannot be reconciled through simply averaging RSkew estimates across constituents. In this part of the thesis we will not focus on reconciling index and constituent stock asymmetries.

1.5 Conclusions

Various measures of skewness have been suggested in the literature mostly focusing on achieving robustness but neglecting arising biases.

The literature on realized skewness is sparse and has either focused on the skewness of intraday returns or, contrary to what the name might suggest, lent itself to option prices. Most importantly, suggested estimators usually estimate a coefficient of skewness at a target sampling frequency but then naturally fail to relate estimates to the dynamics governed by the DGP.

Instead we provide a measure for long horizon returns (relative to the sampling frequency) by making use of the dependence structure inherent in intraday asset returns only.

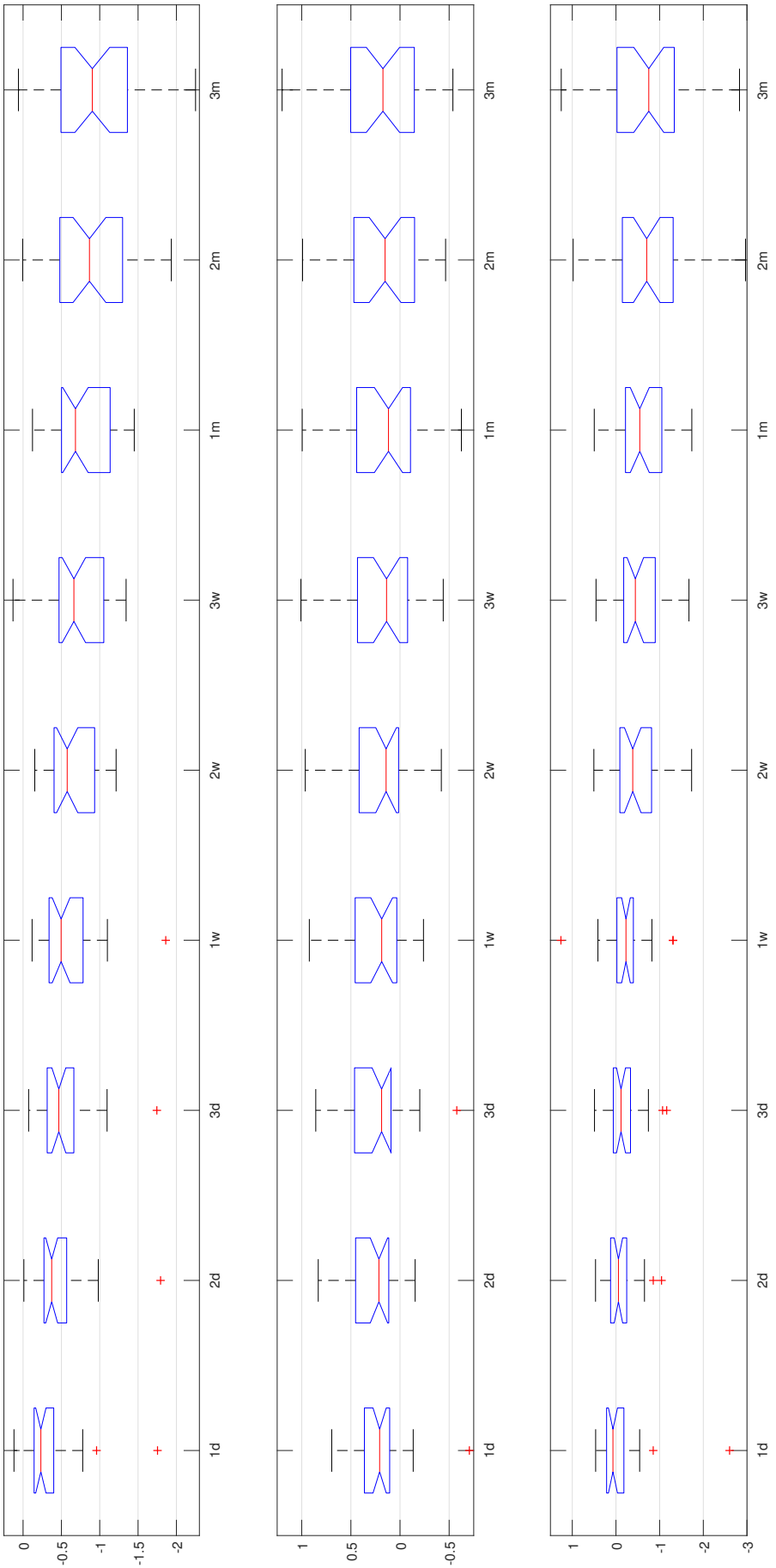


Figure 1.7: Boxplots of RSkew$_{\text{AVL}}$, RSkew$_{\text{VFB}}$ and combined RSkew estimates across stocks from 1d to 3m horizons

Central marks are the medians with the edges of the boxes indicating 25th and 75th percentiles; whiskers extend to the most extreme data points not considered outliers. Outliers are classified as observations larger (smaller) than $q_3 + w(q_3 - q_1)$ ($q_1 - w(q_3 - q_1)$), where w=1.5 and q_x denotes the x'th quartile. Two medians are signif. different at the 5% level if their intervals do not overlap. The interval endpoints are the extremes of the notches which correspond to $q_2 - 1.57(q_3 - q_1)/\sqrt{n}$ UPPER PANEL presents boxplots of RSkew$_{\text{AVL}}$ estimates across stocks for horzions ranging from one day to three months, MIDDLE PANEL presents RSkew$_{\text{VFB}}$ and the LOWER PANEL present the combined (AVL+ VFB+ IA) RSkew estimates. All based on 5–minute data.

The measure shows the accumulation of skewness over time as it builds up from three components at the intraday level: 1.) asymmetric volatility 2.) volatility feedback effects, and 3.) instantaneous skewness. All these terms are functions of the horizon and have to be analysed separately. Statements about positive or negative coefficients of skewness based on one sampling frequency are vague and have perhaps led to contradicting evidence in the literature.

We present an extensive empirical analysis of these, sometimes offsetting, effects at various frequencies and horizons. As estimates of spot variance are generally noisy, we show how more robust volatility proxies can better be used instead. Patterns across individual assets are dispersed and we group stocks into distinct categories showing the same behaviour. Portfolios of these stocks should lead to important implications for asset pricing studies. Aggregate stock market returns appear to be largely subject to asymmetric volatility effects only with negligible volatility feedback effects and are also less subject to instantaneous skewness, which is in line with the current literature. Index return series are then suitable candidates for semi–parametric skewness estimators which we suggest. In a theoretical model analysis, thus suitable for aggregate market returns or stocks not subject to volatility feedback effects such as, e.g., CAT, GE, MO, IBM, and DD, we propose how one could impose stepwise more structure on the covariance terms between return and volatility proxy in order determine skewness at long horizons more precisely in the context of a very general two factor non–affine log linear stochastic volatility model and its reduced form specifications. An estimation of these semi–parametric estimators is beyond the scope of the current paper and left for future research.

Appendix

1.A Proofs for Chapter 1

1.A.1 Decompositional & aggregational results of realized skewness measures

Multi–period building blocks of long–horizon skewness and its non–centralized counterpart. The centralized 3rd N-period return moment can be disentangled into the following long–horizon building blocks

$$\begin{aligned}
\mathrm{E}\Big[\big(r_t^{(N)}-\mu_t^{(N)}\big)^3\Big] &= \mathrm{E}\Big[\big(r_t^{(N)}-\mu_t^{(N)}\big)^2\big(r_t^{(N)}-\mu_t^{(N)}\big)\Big] \\
&= \mathrm{Cov}\Big(\big(r_t^{(N)}-\mu_t^{(N)}\big)^2,\big(r_t^{(N)}-\mu_t^{(N)}\big)\Big) \\
&\quad + \mathrm{E}\Big[\big(r_t^{(N)}-\mu_t^{(N)}\big)^2\Big]\mathrm{E}\Big[\big(r_t^{(N)}-\mu_t^{(N)}\big)\Big] \\
&= \mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2-2r_t^{(N)}\mu_t^{(N)}+\big(\mu_t^{(N)}\big)^2,\ r_t^{(N)}\Big) \\
&= \mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2,r_t^{(N)}\Big)-2\mu_t^{(N)}\mathrm{Var}\Big(r_t^{(N)}\Big). \qquad (1.A.1)
\end{aligned}$$

On the other hand, the non–centralized counterpart equals

$$\begin{aligned}
\mathrm{E}\Big[\big(r_t^{(N)}\big)^3\Big] &= \mathrm{E}\Big[\big(r_t^{(N)}\big)^2\big(r_t^{(N)}\big)\Big] \\
&= \mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2,r_t^{(N)}\Big)+\mathrm{E}\Big[\big(r_t^{(N)}\big)^2\Big]\mathrm{E}\Big[r_t^{(N)}\Big] \\
&= \mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2,r_t^{(N)}\Big)+\mu_t^{(N)}\mathrm{Var}(r_t^{(N)})+\big(\mu_t^{(N)}\big)^3, \qquad (1.A.2)
\end{aligned}$$

■

Realized skewness: High–frequency decomposition of covariances. Regarding the first term on the right–hand side, which is identical in both in Eq. (1.3.8) and Eq. (1.3.9), we have

the covariance decomposition

$$\begin{aligned}
&\mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2, r_t^{(N)}\Big) \\
&= \mathrm{Cov}\Big((r_{t,1}, r_{t,2}, \ldots, r_{t,N})^2, r_{t,1}, r_{t,2}, \ldots, r_{t,N}\Big) \\
&= \mathrm{Cov}\Big(\sum_{i=1}^{N} r_{t,i}^2 + 2\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} r_{t,i}r_{t,k}, \sum_{j=1}^{N} r_{t,j}\Big) \\
&= \mathrm{Cov}\Big(\sum_{i=1}^{N} r_{t,i}^2, \sum_{j=1}^{N} r_{t,j}\Big) + 2\ \mathrm{Cov}\Big(\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} r_{t,i}r_{t,k}, \sum_{j=1}^{N} r_{t,j}\Big). \qquad (1.A.3)
\end{aligned}$$

Now under the empirically realistic assumption

$$\underset{i\neq j}{\mathrm{corr}}(r_{t,i}, r_{t,j}) = 0,$$

$$\underset{i\neq j\neq k}{\mathrm{Cov}}\,(r_{t,i}r_{t,k}, r_{t,j}) = \mathrm{E}[r_{t,i}r_{t,k}r_{t,j}] - \mathrm{E}[r_{t,i}, r_{t,k}]\mathrm{E}[r_{t,j}] = \mu_t^3 - \mu_t^2\mu_t = 0,$$

where $\mu_{t,i} = \mu_{t,k} = \mu_{t,j} = \mu_t$. Then, the second part on the right hand side of the covariance decomposition in Eq. (1.A.3) reduces to those cases when either $j = i$ or $j = k$, i.e.

$$\begin{aligned}
&2\ \mathrm{Cov}\Big(\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} r_{t,i}r_{t,k}, \sum_{j=1}^{N} r_{t,j}\Big) \\
&= 2\sum_{j=1}^{N}\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} \mathrm{Cov}\big(r_{t,i}r_{t,k}, r_{t,j}\big) \\
&= 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \Big(\mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,i}\big) + \mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,j}\big)\Big).
\end{aligned}$$

Further note that

$$\begin{aligned}
\mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,i}\big) &= \mathrm{E}\big[r_{t,i}^2 r_{t,j}\big] - \mathrm{E}[r_{t,i}r_{t,j}]\mathrm{E}[r_{t,j}] = \mathrm{E}\big[r_{t,i}^2 r_{t,j}\big] - \mu_t^3, \\
\mathrm{Cov}\big(r_{t,i}^2, r_{t,j}\big) &= \mathrm{E}\big[r_{t,i}^2 r_{t,j}\big] - \mathrm{E}[r_{t,i}^2]\mathrm{E}[r_{t,j}] = \mathrm{E}\big[r_{t,i}^2 r_{t,j}\big] - [\mu_t^3 + \mu_t\mathrm{Var}(r_{t,i})], \\
\mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,i}\big) &= \mathrm{Cov}\big(r_{t,i}^2, r_{t,j}\big) + \mu_t\mathrm{Var}(r_{t,i}).
\end{aligned}$$

Then, Eq. (1.A.3) reads

$$
\begin{aligned}
&\mathrm{Cov}\Big(\big(r_t^{(N)}\big)^2, r_t^{(N)}\Big) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}\Big(r_{t,i}^2, r_{t,j}\Big) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Big(\mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,i}\big) + \mathrm{Cov}\big(r_{t,i}r_{t,j}, r_{t,j}\big)\Big) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}\Big(r_{t,i}^2, r_{t,j}\Big) \\
&\quad + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Big(\mathrm{Cov}\big(r_{t,i}^2, r_{t,j}\big) + \mu_t\mathrm{Var}(r_{t,i}) + \mathrm{Cov}\big(r_{t,j}^2, r_{t,i}\big) + \mu_t\mathrm{Var}(r_{t,j})\Big) \\
&= \sum_{i=1}^{N}\mathrm{Cov}\Big(r_{t,i}^2, r_{t,i}\Big) + 3\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{Cov}\Big(r_{t,i}^2, r_{t,j}\Big) + 3\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathrm{Cov}\Big(r_{t,j}^2, r_{t,i}\Big) \\
&\quad + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Big(\mu_t\mathrm{Var}(r_{t,i}) + \mu_t\mathrm{Var}(r_{t,j})\Big).
\end{aligned}
$$

The last single line above can be re–written in the following way,

$$
2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Big(\mu_t\mathrm{Var}(r_{t,i}) + \mu_t\mathrm{Var}(r_{t,j})\Big) = 2\mu_t^{(N)}\mathrm{Var}(r_t^{(N)}) - 2\sum_{i=1}^{N}\mu_t\mathrm{Var}(r_{t,i}),
$$

and as

$$
\begin{aligned}
&\sum_{i=1}^{N}\mathrm{Cov}\Big(r_{t,i}^2, r_{t,i}\Big) \\
&= \sum_{i=1}^{N}\Big(\mathrm{E}[r_{t,i}^3] - \mathrm{E}[r_{t,i}^2]\mathrm{E}[r_{t,i}]\Big) = \sum_{i=1}^{N}\Big(\mathrm{E}[r_{t,i}^3] - (\mathrm{Var}(r_{t,i}) + \mu_t^2)\mu_t\Big) \\
&= \sum_{i=1}^{N}\Big(\mathrm{E}[(r_{t,i}-\mu_t)^3] + 2\mu_t\mathrm{Var}(r_{t,i})\Big),
\end{aligned}
$$

combining all terms leads to the result in Eq. (1.3.11). ■

Non–centralized realized skewness: High–frequency decomposition of expectations. In analogy to the derivation above,

$$
\begin{aligned}
\mathrm{E}\Big[\big(r_t^{(N)}\big)^3\Big] &= \mathrm{E}\Big[\big(\sum_{i=1}^{N}r_{t,i}\big)^2\big(\sum_{j=1}^{N}r_{t,j}\big)\Big] \\
&= \mathrm{E}\Big[\Big(\sum_{i=1}^{N}r_{t,i}^2 + 2\sum_{i=1}^{N-1}\sum_{k=i+1}^{N}r_{t,i}r_{t,k}\Big)\times\sum_{j=1}^{N}r_{t,j}\Big] \\
&= \mathrm{E}\Big[\sum_{i=1}^{N}r_{t,i}^2\sum_{j=1}^{N}r_{t,j}\Big] + 2\,\mathrm{E}\Big[\sum_{i=1}^{N-1}\sum_{k=i+1}^{N}r_{t,i}r_{t,k}\sum_{j=1}^{N}r_{t,j}\Big]. \qquad (1.A.4)
\end{aligned}
$$

The first term of the last line above equals

$$E\Big[\sum_{i=1}^{N} r_{t,i}^2 \sum_{j=1}^{N} r_{t,j}\Big] = E\Big[\sum_{j=1}^{N}\sum_{i=1}^{N} r_{t,i}^2 r_{t,j}\Big] = \sum_{j=1}^{N} E\Big[\sum_{i=1}^{N} r_{t,i}^2 r_{t,j}\Big] = \sum_{j=1}^{N}\sum_{i=1}^{N} E\Big[r_{t,i}^2 r_{t,j}\Big].$$

The second term can be re–written as

$$\begin{aligned}
&2\ \mathrm{E}\Big[\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} r_{t,i}r_{t,k}\sum_{j=1}^{N} r_{t,j}\Big] \\
&= 2\ \sum_{j=1}^{N}\mathrm{E}\Big[\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} r_{t,i}r_{t,k}r_{t,j}\Big] = 2\ \sum_{j=1}^{N}\sum_{i=1}^{N-1}\sum_{k=i+1}^{N}\mathrm{E}\Big[r_{t,i}r_{t,k}r_{t,j}\Big],
\end{aligned}$$

combining terms for $j = i$ and $j = k$,

$$= 2\left(\sum_{i=1}^{N-1}\sum_{k=i+1}^{N}\mathrm{E}\Big[r_{t,i}r_{t,k}r_{t,i}\Big] + \sum_{i=1}^{N-1}\sum_{k=i+1}^{N}\mathrm{E}\Big[r_{t,i}r_{t,k}r_{t,k}\Big] + \sum_{j=1}^{N}\sum_{\substack{i=1\\ j\neq i,k\neq j}}^{N-1}\sum_{k=i+1}^{N}\mathrm{E}\Big[r_{t,i}r_{t,k}r_{t,k}\Big]\right).$$

Under the same assumption as before, $\underset{j\neq i,k\neq j}{\mathrm{E}}\big[r_{t,i}r_{t,k}r_{t,k}\big] = \mu_t^3$, thus

$$\sum_{j=1}^{N}\sum_{\substack{i=1\\ j\neq i,k\neq j}}^{N-1}\sum_{k=i+1}^{N}\mathrm{E}\big[r_{t,i}r_{t,k}r_{t,k}\big] = (N-2)\frac{(N-1)N}{2}\mu_t^3,$$

as there were previously $\sum_{j=1}^{N}\sum_{i=1}^{N-1}\sum_{k=i+1}^{N} 1 = N\frac{(N-1)N}{2}$ summands[31] from which we separated those having either $j = i$ or $j = k$. Combining these terms, yields the result in Eq. (1.3.15f). ■

1.A.2 Properties and analytical moments of SV–type models

Proof of Eq. (1.3.35). The set of Eq. (1.3.34) is a special case of the general Gaussian Markov process specified by

$$\mathrm{d}\varrho(t) = [g(t) + h(t)\varrho(t)]\mathrm{d}t + \sigma(t)\mathrm{d}W(t)$$

with g, h, and σ deterministic functions of time with solution, see e.g. (Glasserman, 2004),

$$\varrho(t) = \exp(H(t))\varrho(0) + \int_0^t \exp(H(t) - H(s))g(s)\mathrm{d}s + \int_0^t \exp(H(t) - H(s))\sigma(s)\mathrm{d}W(s)$$

[31] This simply follows from the fact that $1, 2, \cdots + N = N(N+1)/2$.

with

$$H(t) = \int_0^t h(s)\mathrm{d}s.$$

For the Ornstein–Uhlenbeck process in Eq. (1.3.34) the general equation above specializes to

$$\varrho(t) = \exp(\kappa t)\varrho(0) + \int_0^t \exp(\kappa(t-s))\mathrm{d}B(s),$$

i.e. with $\kappa < 0$ the innovations are conditionally normal distributed

$$\varrho_t|\varrho_u \sim N\Big(\exp(\kappa(t-u))\varrho_u, \frac{1}{-2\kappa}(1-\exp(2\kappa(t-u))\Big),$$

for $\varrho_t|\varrho_{u=0}$ with limiting mean zero and limiting variance

$$\lim_{t\to\infty}\mathrm{Var}[\varrho(t)] = \lim_{t\to\infty}\frac{1}{-2\kappa}\Big(1-\exp(2\kappa t)\Big) = \frac{1}{-2\kappa}.$$

■

Proof of Eq. (1.3.36). From the second line of the set of Eq. (1.3.35), note that

$$\varrho_{j-1} = \frac{\ln(\sigma_{j-1}) - \beta_0}{\beta_1},$$

such that ϱ_j can be written in terms of $\ln(\sigma_{j-1})$

$$\varrho_j = \exp(\kappa\Delta)\frac{\ln(\sigma_{j-1}) - \beta_0}{\beta_1} + \frac{1}{\sqrt{-2\kappa}}\sqrt{1-\exp(2\kappa\Delta)}\varepsilon^B_{j-1},$$

and thus as shown in Eq. (1.3.36)

$$\ln(\sigma_j) = \beta_0 + \exp(\kappa\Delta)\Big(\ln(\sigma_{j-1}) - \beta_0\Big) + \frac{\beta_1}{\sqrt{-2\kappa}}\sqrt{1-\exp(2\kappa\Delta)}\varepsilon^B_{j-1}.$$

■

Proof of Eq. (1.3.39). The random variable $\ln(\sigma_t)$ is normally distributed independent of the generating frequency with mean β_0 and variance $\frac{\beta_1^2}{-2\kappa}$ such that we can suppress the subscript M or Δ respectively.[32] More generally, as $\ln(\sigma_t^\theta) = \theta\ln(\sigma_t)$, it has distribution

[32] It can be interpreted in a similar manner to the way interest rates are modeled and quoted in annualized terms.

$\mathrm{N}(\theta\beta_0, \theta^2 \frac{\beta_1^2}{-2\kappa})$. Consequently, by the property of the lognormal distribution, σ_t^θ has mean

$$\begin{aligned} \mathrm{E}[\sigma_t^\theta] &= \mathrm{E}\int_0^1 [\exp(\beta_0 + \beta_1 \mathrm{d}\varrho_t)]^\theta \mathrm{d}t \\ &= \exp\Big(\theta\beta_0 + \frac{\theta^2}{2}\big(\frac{\beta_1}{\sqrt{-2\kappa}}\big)^2\Big), \end{aligned}$$

i.e. $\mathrm{E}[\sigma_t^2] = \exp\Big(2\beta_0 + 2\big(\frac{\beta_1}{\sqrt{-2\kappa}}\big)^2\Big)$, and variance

$$\begin{aligned} \mathrm{Var}[\sigma_t^\theta] &= \mathrm{E}\int_0^1 [\exp(\beta_0 + \beta_1 \mathrm{d}\varrho_t)]^{2\theta} \mathrm{d}t - \Big[\mathrm{E}\int_0^1 \exp(\beta_0 + \beta_1 \mathrm{d}\varrho_t)^\theta \mathrm{d}t\Big]^2 \\ &= \exp\Big(2\theta\beta_0 + \theta^2\big(\frac{\beta_1}{\sqrt{-2\kappa}}\big)^2\Big)\Big(\exp\Big(\theta^2\big(\frac{\beta_1}{\sqrt{-2\kappa}}\big)^2\Big) - 1\Big), \end{aligned}$$

i.e. $\mathrm{Var}[\sigma_t^2] = \exp\Big(4\beta_0 + 4\big(\frac{\beta_1}{\sqrt{-2\kappa}}\big)^2\Big)\Big(\exp\Big(4\big(\frac{\beta_1}{\sqrt{-2\kappa}}\big)^2\Big) - 1\Big)$.

Note for $\beta_0 = \frac{\beta_1^2}{2\kappa}$, $\mathrm{E}[\sigma_t^2]$ integrates to unity with variance $\mathrm{Var}[\sigma_t^2] = \exp\Big(4\big(\frac{\beta_1}{\sqrt{-2\kappa}}\big)^2\Big) - 1$, which is a useful result to compare it to constant volatility cases. ■

Analytical moments of the General SV model with one volatility factor

Proof of Eq. (1.3.40). When interested in the third N–period return moment (Eq. (1.3.6)) under **model M** (Eqs. (1.3.37)), we only need to evaluate

$$\mathrm{E}[r_{t,i,M} r_{t,(i+j),M}^2] = M^{-\frac{3}{2}} \mathrm{E}[u_{t,i,M} \sigma_{t,i,M} \sigma_{t,(i+j),M}] \text{ for } j > 0 \qquad (1.A.5)$$

as $u_{t,(i+j),M}$ is independent of $\{u_{t,i,M}, \sigma_{t,i,M}, \sigma_{t,(i+j),M}\}$, and as the following terms are all zero in expectation;

$$\begin{aligned} \mathrm{E}[r_{t,i,M}^3] &= 0, \text{ because } u_{t,i,M} \text{ is indep. of } \sigma_{t,i,M}, \\ \mathrm{E}[r_{t,i,M} r_{t,(i+j),M} r_{t,(i+k),M}] &= 0, \text{ because } u_{t,(i+k),M} \text{ is indep. of} \\ &\qquad \underbrace{\{u_{t,i,M}, u_{t,(i+j),M}, \sigma_{t,i,M}, \sigma_{t,(i+j),M}, \sigma_{t,(i+k),M}\}}_{0<j<k}, \\ \mathrm{E}[r_{t,i,M}^2 r_{t,(i+j),M}] &= 0, \text{ because } u_{t,(i+j),M} \text{ is indep. of } \{\sigma_{t,i,M}, u_{t,i,M}, \underbrace{\sigma_{t,(i+j),M}}_{0<j}\}. \end{aligned}$$

With $\alpha = 0$ for now, note that the log volatility process at lead order j can be rewritten as

$$\begin{aligned}
\log(\sigma_{t,(i+1),M}) &= \phi_M \log(\sigma_{t,i,M}) + \eta_{t,(i+1),M}, \\
\log(\sigma_{t,(i+2),M}) &= \phi_M \log(\sigma_{t,(i+1),M}) + \eta_{t,(i+2),M} \\
&= \phi_M^2 \log(\sigma_{t,i,M}) + \eta_{t,(i+2),M} + \phi_M \eta_{t,(i+1),M}, \\
&\vdots \\
\log(\sigma_{t,(i+j),M}) &= \phi_M^j \log(\sigma_{t,i,M}) + \underbrace{\eta_{t,(i+j),M} + \phi_M \eta_{t,(i+j-1),M} + \ldots + \phi_M^{(j-1)} \eta_{t,(i+1),M}}_{j \text{ terms}},
\end{aligned}$$

for $j \geq 1$. Consequently, the volatility process reads

$$\sigma_{t,(i+j),M} = \sigma_t^{\phi_M^j} \exp\Big(\eta_{t,(i+j),M} + \ldots + \phi_M^{(j-1)} \eta_{t,(i+1),M}\Big),$$

and the product term on the RHS of Eq. (1.A.5) above for the calculation of the third moment under **model M** has the explicit form

$$u_{t,i,M} \sigma_{t,i,M} \sigma_{t,(i+j),M}^2 = A \times B \times C,$$

where

$$\begin{aligned}
A &= \sigma_{t,i,M}^{1+2\phi_M^j}, \\
B &= \Big[u_{t,i,M} \exp(2\phi_M^{(j-1)} \eta_{t,(i+1),M})\Big], \\
C &= \exp\Big(\underbrace{2(\eta_{t,(i+j),M} + \ldots + \phi_M^{(j-2)} \eta_{t,(i+2),M})}_{j-1 \text{ terms}}\Big).
\end{aligned}$$

The expectation of this term can be evaluated as follows: Using the result of the proof of Eq. (1.3.39) above, note that

$$\mathrm{E}[A] = \mathrm{E}[\sigma_{t,i,M}^{1+2\phi_M^j}] = \exp\Big(\tfrac{1}{2}(1 + 2\phi_M^j)^2 \beta^2\Big). \tag{1.A.6}$$

For $\mathrm{E}[B]$, we can use the fact that the term inside the exponential operator follows a normal distribution,

$$2\phi_M^{(j-1)} \eta_{t,(i+1),M} \sim N(0, 4\phi_M^{2(j-1)} \sigma_{\eta,M}^2).$$

Then, the expectation of the product of the terms of interests is the covariance of the bivariate normal–lognormally distributed variables. From the **Auxiliary results** in the last subsection of this Appendix 1.A.2 further down below, for the case $\mathbf{g}(\mathbf{u_{i,M}}) = \mathbf{u_{i,M}}$,

it is given by

$$\begin{aligned} \mathrm{E}[B] &= \mathrm{E}\Big[u_{t,i,M}\exp\Big(2\phi_M^{(j-1)}\eta_{t,(i+1),M}\Big)\Big] \\ &= 2\delta\phi_M^{(j-1)}\sigma_{\eta,M}\exp\Big(2\phi_M^{2(j-1)}\sigma_{\eta,M}^2\Big). \end{aligned} \tag{1.A.7}$$

The last term equals

$$\begin{aligned} \mathrm{E}[C] &= \mathrm{E}\Big[\exp\big(2(\eta_{t,(i+j),M}+\ldots+\phi_M^{(j-2)}\eta_{t,(i+2),M})\big)\Big] \\ &= \mathrm{E}\Big[\exp\big(2\sigma_{\eta,M}^2[\underbrace{1+\phi_M^2+\ldots+\phi_M^{2(j-2)}}_{D}]\big)\Big], \end{aligned} \tag{1.A.8}$$

where based on partial sum formula $D=\Big[\frac{1-\phi_M^{2(j-1)}}{1-\phi_M^2}\Big]$. Thus,

$$\mathrm{E}[ABC] = 2\delta\sigma_{\eta,M}\phi_M^{(j-1)}\exp[f_j(\phi_M)\beta^2]$$

with

$$\begin{aligned} f_j(\phi_M) &= 2\phi_M^{2(j-1)}(1-\phi_M^2)+\tfrac{1}{2}(1+2\phi_M^j)^2+2(1-\phi_M^2)\left[\frac{1-\phi_M^{2(j-1)}}{1-\phi_M^2}\right] \\ &= 2\phi_M^{2(j-1)}-2\phi_M^{2j}+\tfrac{1}{2}\Big(1+4\phi_M^j+4\phi_M^{2j}\Big)+2+2\phi_M^{2(j-1)} \\ &= \tfrac{1}{2}\Big(5+4\phi_M^j\Big). \end{aligned}$$

With non–zero α, the covariance terms will be scaled by a factor, $\exp(3\alpha)$, which leads to the result stated in Eq. (1.3.40). ■

Analytical moments of the General SV model with two volatility factors

Proof of Eq. (1.3.43) Part I: Re–visiting and extending the framework.
For clarity of exposition, we now drop the subscript t marking the time window under consideration. Let k_x denotes the x'th volatility process such that we can easily differentiate, e.g., $\sigma_{k_1,i,M}$ from $\sigma_{k_2,i,M}$. Initially assume $\alpha_{k_x}=0$ for $x=1,2$ and consider only $j\geq 1$. For now, we are primarily interested in the case when $\theta=1$ with any positive real λ. Covariance terms are

$$\begin{aligned} \mathrm{Cov}\Big(r_{i,M}^{\theta},|r_{i+j,M}|^{\lambda}\Big) &= M^{-\frac{1}{2}(\theta+\lambda)}\mathrm{E}\Big[u_i^{\theta}\sigma_{i,M}^{\theta}|u_{i+j}|^{\lambda}\sigma_{i+j,M}^{\lambda}\Big] \\ &= M^{-\frac{1}{2}(\theta+\lambda)}\mathrm{E}\Big[u_i^{\theta}\sigma_{i,M}^{\theta}\sigma_{i+j,M}^{\lambda}\Big]\mathrm{E}\Big[|u_{i+j}|^{\lambda}\Big]. \end{aligned}$$

As before the k_x'th logarithmic volatility process can be recursively written as

$$\ln(\sigma_{k_x,i+j,M}) = \phi^j_{k_x,M}\ln(\sigma_{k_x,i,M}) + \underbrace{\eta_{k_x,i+j} + \phi_{k_x,M}\eta_{k_x,i+j-1} + \ldots + \phi^{j-1}_{k_x,M}\eta_{k_x,i+1}}_{j \text{ terms}}.$$

The interest lies in product terms involving power functions of volatility, i.e.

$$\sigma^\theta_{k_x,i,M}\sigma^\lambda_{k_x,i+j,M} = \sigma^{\theta+\lambda\phi^j_{k_x,M}}_{k_x,i,M}\exp\Big(\lambda\phi^{j-1}_{k_x,M}\eta_{k_x,i+1}\Big)\exp\Big(\lambda(\underbrace{\eta_{k_x,i+j} + \ldots + \phi^{j-2}_{k_x,M}\eta_{k_x,i+2}}_{j-1 \text{ terms}})\Big),$$

more specific

$$u^\theta_i\sigma^\theta_{i,M}\sigma^\lambda_{i+j,M} = A \times B \times C,$$

where

$$\begin{aligned}
A &= \sigma^{\theta+\lambda\phi^j_{k_1,M}}_{k_1,i,M}\sigma^{\theta+\lambda\phi^j_{k_2,M}}_{k_2,i,M},\\
B &= u^\theta_i\exp\Big(\lambda(\phi^{j-1}_{k_1,M}\eta_{k_1,i+1} + \phi^{j-1}_{k_2,M}\eta_{k_2,i+1})\Big),\\
C &= \exp\Big(\lambda(\eta_{k_1,i+j} + \ldots + \phi^{j-2}_{k_1,M}\eta_{k_1,i+2} + \eta_{k_2,i+j} + \ldots + \phi^{j-2}_{k_2,M}\eta_{k_2,i+2})\Big).
\end{aligned}$$

Then, the expectation of the first and third term are straightforward, with $\beta^2_{k_x} = \frac{\sigma^2_{\eta_{k_x},M}}{1-\phi^2_{k_x,M}}$,

$$\mathrm{E}[A] = \exp\Big(\tfrac{1}{2}(\theta + \lambda\phi^j_{k_1,M})^2\beta^2_{k_1} + \tfrac{1}{2}(\theta + \lambda\phi^j_{k_2,M})^2\beta^2_{k_2}\Big),$$

$$\mathrm{E}[C] = \exp\Big(\tfrac{1}{2}\lambda^2\sum_{x=1}^{2}\sigma^2_{\eta_{k_x},M}\Big[1 + \phi^2_{k_x,M} + \ldots + \phi^{2(j-2)}_{k_x,M}\Big]\Big).$$

For the second term we need to evaluate

$$\mathrm{E}[B] = \mathrm{E}\Big[u^\theta_i\exp\Big(\lambda(\phi^{j-1}_{k_1,M}\eta_{k_1,i+1} + \phi^{j-1}_{k_2,M}\eta_{k_2,i+1})\Big)\Big].$$

It is convenient to define

$$\eta^*_{i+1,j} = \lambda(\phi^{j-1}_{k_1,M}\eta_{k_1,i+1} + \phi^{j-1}_{k_2,M}\eta_{k_2,i+1}),$$

i.e. we group the shocks to volatility together, then

$$\begin{pmatrix} u_i \\ \eta^*_{i+1,j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \mathrm{E}[u_i\eta^*_{i+1,j}] \\ \mathrm{E}[u_i\eta^*_{i+1,j}] & \mathrm{Var}[\eta^*_{i+1,j}] \end{pmatrix}\right)$$

with $\mathrm{E}[u_i\eta_{k_x,i+1}] = \delta_{k_x}\sigma_{\eta_{k_x},M}$, from Eq. (1.3.42), and

$$\mathrm{E}[u_i\eta^*_{i+1,j}] = \sum_{x=1}^{2}\lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M},$$

$$\mathrm{Var}(\eta^*_{i+1,j}) = \sum_{x=1}^{2} \lambda^2 \sigma^2_{\eta_{k_x},M} \phi^{2(j-1)}_{k_x,M}.$$

which follows from the decomposition of correlated random variables in 1.A.2. Generally, we seek

$$\mathrm{E}[g(u_i)\exp(\eta^*_{i+1,j})], \tag{1.A.9}$$

which enters the expectation of the product of $A \times B \times C$, and thus extends to the definition of the third moment, and eventually skewness. ■

Proof of Eq. (1.3.43) Part II: The explicit case of Eq. (1.A.9) leading to Eq. (1.3.43). Continuing the analysis above, we want to evaluate $\mathrm{E}_g[ABC]$ for the specification of Eq. (1.A.9) in which $\mathbf{g(u_i)} = \mathbf{u_i}$. Using the result for Eq. (1.A.9) when $g(u_i) = u_i$, see Eq. (1.A.12) stated in the **Auxiliary results** section of this Appendix, yields

$$\begin{aligned}
&\mathrm{E}_{g(u_i)=u_i}[ABC] \\
&= \underbrace{\Big(\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M}\Big)\exp\Big(\lambda^2\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M}\Big)}_{\text{from } B} \\
&\quad \times \exp\Big(\sum_{x=1}^{2}\Big[\underbrace{\tfrac{1}{2}(1+\lambda\phi^{j}_{k_x,M})^2\tfrac{\sigma^2_{\eta_{k_x},M}}{1-\phi^2_{k_x,M}}}_{\text{from } A} + \underbrace{\tfrac{1}{2}\lambda^2\sigma^2_{\eta_{k_x},M}\phi^{2(j-1)}_{k_x,M}}_{\text{from } B} \\
&\qquad + \underbrace{\tfrac{1}{2}\lambda^2\sigma^2_{\eta_{k_x},M}[1+\phi^2_{k_x,M}+\ldots+\phi^{2(j-2)}_{k_x,M}]}_{\text{from } C}\Big]\Big) \\
&= \Big(\sum_{x=1}^{2}\lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M}\Big)\exp\Big(\lambda^2\prod_{x=1}^{2}\delta_{k_x}\phi^{j-1}_{k_x,M}\sigma_{\eta_{k_x},M}\Big)\exp\Big(\sum_{x=1}^{2}\tfrac{1}{2}\sigma^2_{\eta_{k_x},M}\times G_j(\lambda,\phi_{k_x,M})\Big),
\end{aligned}$$

with

$$\begin{aligned}
G_j(\lambda,\phi_{k_x,M}) &= \frac{(1+\lambda\phi^{j}_{k_x,M})^2}{1-\phi^2_{k_x,M}} + \frac{\lambda^2(1-\phi^{2j}_{k_x,M})}{1-\phi^2_{k_x,M}} \\
&= \frac{1+2\lambda\phi^{j}_{k_x,M}+\lambda^2}{1-\phi^2_{k_x,M}},
\end{aligned}$$

where we use partial sum formula to combine contributions from B and C. Note that with $\alpha_{k_x} \neq 0$ for $x = 1, 2$ the following scaling factor must be included,

$$\varsigma(\lambda,\alpha_{k_1},\alpha_{k_2}) = \exp\Big(\Big[\sum_{x=1}^{2}\alpha_{k_x}\Big]\Big[1+\lambda\Big]\Big),$$

and

$$\begin{aligned}\mathrm{Cov}(r_{i,M},|r_{i+j,M}|^{\lambda}) &= M^{-\frac{1}{2}(1+\lambda)}\mathrm{E}\Big[|u_{i+j}|^{\lambda}\Big]\varsigma(\lambda,\alpha_{k_1},\alpha_{k_2})\lambda\sum_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\\ &\quad\times\exp\Big(\sum_{x=1}^{2}\tfrac{1}{2}(1+2\lambda\phi_{k_x,M}^{j}+\lambda^2)\beta_{k_x}^2+\lambda^2\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big)\\ &= \zeta_j(\lambda,\Theta), \qquad (1.A.10)\end{aligned}$$

where Θ denotes the parameter vector, and recall that the persistencies and the (instantaneous) volatilities of the the two volatility factors are all functions of M. ■

Remark: The one volatility factor case. For only one volatility component and $\lambda = 1, 2$ Eq. (1.A.10) becomes

$$\mathrm{Cov}(r_{i,M},|r_{i+j,M}|) = M^{-1}\sqrt{\tfrac{2}{\pi}}\delta\sigma_{\eta,M}\phi_M^{j-1}\exp\Big(2\alpha+\tfrac{1}{2}(2+2\phi_M^j)\beta^2\Big),$$

$$\mathrm{Cov}(r_{i,M},r_{i+j,M}^2) = M^{-\frac{3}{2}}2\delta\sigma_{\eta,M}\phi_M^{j-1}\exp\Big(3\alpha+\tfrac{1}{2}(5+4\phi_M^j)\beta^2\Big),$$

as Eq. (11.54) for $\lambda = 1$, for the case $j = 1$ only, and Eq. (11.58) for $\lambda = 2$ in Taylor (2005, Chapter 11).

Remark: Scaling results and ratios when covariance terms are linear in delta. Note that

$$\begin{aligned}\frac{\zeta_j(\lambda,\Theta)}{\zeta_j(\xi,\Theta)} &= M^{\frac{\xi-\lambda}{2}}\frac{\mathrm{E}[|u_{i+j}|^{\lambda}]}{\mathrm{E}[|u_{i+j}|^{\xi}]}\frac{\lambda}{\xi}\exp\big(\sum_{x=1}^{2}\alpha_{k_x}(\lambda-\xi)\big) \qquad (1.A.11)\\ &\quad\times\exp\Big(\frac{1}{2}\sum_{x=1}^{2}\Big[\big(\lambda^2-\xi^2+2(\lambda-\xi)\phi_{k_x,M}^{j}\big)\beta_{k_x}^2\Big]+(\lambda^2-\xi^2)\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big)\end{aligned}$$

only depends on $\beta_{k_x}, \phi_{k_x,M}$, and j for one volatility factor. It does depend on δ_{k_x} only when both volatility factors have a non–zero correlation, e.g.

$$\frac{\zeta_j(2,\Theta)}{\zeta_j(1,\Theta)} = M^{-\frac{1}{2}}2\sqrt{2\pi}\exp\Big(\sum_{x=1}^{2}\Big[\alpha_{k_x}+(\tfrac{3}{2}+\phi_{k_x,M}^{j})\beta_{k_x}^2\Big]+3\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x,M}}\phi_{k_x,M}^{j-1}\Big).$$

As a function of j

$$\begin{aligned}\mathrm{Cov}(r_{i,M},|r_{i+j,M}|^{\lambda}) &= M^{-\frac{1}{2}(1+\lambda)}\underbrace{C(\lambda,\beta_{k_1},\beta_{k_2})}_{\text{const.}}\varsigma(\lambda,\alpha_{k_1},\alpha_{k_2})\\ &\quad\times\Big(\sum_{x=1}^{2}\delta_k\sigma_{\eta_k,M}\phi_{k_x,M}^{j-1}\Big)\exp\Big(\lambda\sum_{x=1}^{2}\beta_{k_x}^2\phi_{k_x,M}^{j}+\lambda^2\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_k,M}\phi_{k,M}^{j-1}\Big),\end{aligned}$$

with

$$C(\lambda,\beta_{k_1},\beta_{k_2})=\mathrm{E}\Big[|u_{i+j}|^{\lambda}\Big]\lambda\exp\Big(\tfrac{1}{2}(1+\lambda^2)\sum_{x=1}^{2}\beta_{k_x}^2\Big),$$

for example, for $\lambda=1,2$

$$C(1,\beta_{k_1},\beta_{k_2})=\sqrt{\frac{2}{\pi}}\exp\big(\beta_{k_1}^2+\beta_{k_2}^2\big),\quad\text{and}\quad C(2,\beta_{k_1},\beta_{k_1})=2\exp\big(\tfrac{5}{2}(\beta_{k_1}^2+\beta_{k_1}^2)\big),$$

and for the constant contribution ratio we obtain

$$\frac{C(\lambda)}{C(\xi)}=\frac{\lambda}{\xi}\exp\Big(\frac{1}{2}\sum_{x=1}^{2}\big[\lambda^2-\xi^2\big]\beta_{k_x}^2\Big)\frac{\mathrm{E}[|u_{i+j}|^{\lambda}]}{\mathrm{E}[|u_{i+j}|^{\xi}]}.$$

Further results related to the persistence and autocorrelation in functions of volatility

Using the auxiliary result for the two volatility factor case from the subsection below, i.e. Eq. (1.A.13), it follows that

$$\begin{aligned}\mathrm{E}_{g(u_i)=u_i^2}[ABC] &= \Big(\sum_{x=1}^{2}\big[1+\lambda^2\delta_{k_x}^2\sigma_{\eta_{k_x,M}}^2\phi_{k_x,M}^{2(j-1)}\big]\Big)\\ &\quad\exp\Big(\sum_{x=1}^{2}\big[\tfrac{1}{2}\sigma_{\eta_{k_x},M}^2\times G_j(\theta,\lambda,\phi_{k_x,M})\big]+\lambda^2\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big),\end{aligned}$$

with

$$G_j(\theta,\lambda,\phi_{k_x,M}) = \frac{\theta^2+2\theta\lambda\phi_{k_x,M}^{j}+\lambda^2}{1-\phi_{k_x,M}^{2}},$$

and scaling factor

$$\varsigma(\theta,\lambda,\alpha_{k_1},\alpha_{k_2})=\exp\Big(\Big[\sum_{x=1}^{2}\alpha_{k_x}\Big]\Big[\theta+\lambda\Big]\Big),$$

which produces the result,

$$\begin{aligned}
\mathrm{E}\big[|r_{i,M}|^2|r_{i+j,M}|^\lambda\big] &= M^{-\frac{1}{2}(2+\lambda)}\sum_{x=1}^{2}\big(1+\lambda^2\delta_{k_x}^2\sigma_{\eta_{k_x},M}^2\phi_{k_x,M}^{2(j-1)}\big)\varsigma(2,\lambda,\alpha_{k_1},\alpha_{k_2})\mathrm{E}[|u_{i+j}|^\lambda] \\
&\quad\times\exp\Big(\sum_{x=1}^{2}\Big[\tfrac{1}{2}(2^2+2\theta\lambda\phi_{k_x,M}^{j}+\lambda^2)\beta_{k_x}^2\Big]+\lambda^2\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big) \\
\text{for } \lambda=2 \\
&= M^{-2}\sum_{x=1}^{2}\big(1+4\delta_{k_x}^2\sigma_{\eta_{k_x},M}^2\phi_{k_x,M}^{2(j-1)}\big) \\
&\quad\times\exp\Big(4\Big[\sum_{x=1}^{2}\big[\alpha_{k_x}+(1+\phi_{k_x,M}^{j})\beta_{k_x}^2\big]+\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big]\Big).
\end{aligned}$$

Also note that covariances, $\theta=\lambda=2$, with $\mathrm{E}[|u_x|^2]=1$,

$$\begin{aligned}
\mathrm{Cov}(r_{i,M}^2,r_{i+j,M}^2) &= \mathrm{E}\big[r_{i,M}^2r_{i+j,M}^2\big]-\mathrm{E}\big[r_{i,M}^2\big]\mathrm{E}\big[r_{i+j,M}^2\big] \\
&= \mathrm{E}\big[r_{i,M}^2r_{i+j,M}^2\big]-M^{-2}\exp\Big(4\sum_{x=1}^{2}\big[\alpha_{k_x}+\beta_{k_x}^2\big]\Big) \\
&= M^{-2}\Bigg[\sum_{x=1}^{2}\big(1+4\delta_{k_x}^2\sigma_{\eta_{k_x},M}^2\phi_{k_x,M}^{2(j-1)}\big)\exp\Big(4\Big[\sum_{x=1}^{2}\big[\phi_{k_x,M}^{j}\beta_{k_x}^2\big] \\
&\quad+\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big]\Big)-1\Bigg]\times\exp\Big(4\sum_{x=1}^{2}\big[\alpha_{k_x}+\beta_{k_x}^2\big]\Big).
\end{aligned}$$

And, again using the auxiliary result for the two volatility factor case from the subsec-

tion below, i.e. Eq.(1.A.14), for $\mathbf{g(u_i)} = \mathbf{abs(u_i)}$ we obtain

$$
\begin{aligned}
&\mathrm{E}_{g(u_i)=|u_i|}[ABC] \\
&= \Bigg(\sqrt{2/\pi} \exp\Big(-\tfrac{1}{2}\big(\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M} \times \phi_{k_x,M}^{j-1}\big)^2\Big) - \Big[\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big] \\
&\quad \Big[1 - 2\Phi(-\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1})\Big]\Bigg) \\
&\quad \times \exp\Big(\sum_{x=1}^{2}\Big[\underbrace{\tfrac{1}{2}(1+\lambda\phi_{k_x,M}^{j})^2 \tfrac{\sigma^2_{\eta_{k_x},M}}{1-\phi^2_{k_x,M}}}_{\text{from } A} + \underbrace{\tfrac{1}{2}\lambda^2\sigma^2_{\eta_{k_x},M}\phi^{2(j-1)}_{k_x,M}(1-\delta^2_{k_x})}_{\text{from } B} \\
&\quad \underbrace{+\tfrac{1}{2}\lambda^2\sigma^2_{\eta_{k_x},M}(1+\phi^2_{k_x,M}+\ldots+\phi^{2(j-2)}_{k_x,M})\Big]}_{\text{from } C}\Big) \\
&= \Bigg(\Big[\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big]\Big[2\Phi(-\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}) - 1\Big] \\
&\quad + \sqrt{2/\pi} \exp\Big(-\tfrac{1}{2}\big(\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k}^{j-1}\big)^2\Big)\Bigg) \\
&\quad \times \exp\Big(\tfrac{1}{2}\sum_{x=1}^{2}\big[\big(1+2\lambda\phi^{j}_{k_x,M}+\lambda^2\big)\beta^2_{k_x} - \lambda^2\delta^2_{k_x}\sigma^2_{\eta_{k_x}}\phi^{2(j-1)}_{k_x}\big]\Big),
\end{aligned}
$$

which leads to

$$
\begin{aligned}
\mathrm{E}\big[|r_{i,M}||r_{i+j,M}|^{\lambda}\big] &= M^{-\frac{1}{2}(1+\lambda)}\Bigg(\Big[\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big]\Big[2\Phi(-\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}) - 1\Big] \\
&\quad + \sqrt{2/\pi} \exp\Big(-\tfrac{1}{2}\big(\sum_{x=1}^{2} \lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\big)^2\Big)\Bigg)\varsigma(\theta,\lambda,\alpha_{k_1},\alpha_{k_1}) \\
&\quad \times \exp\Big(\tfrac{1}{2}\sum_{x=1}^{2}\big[\big(1+2\lambda\phi^{j}_{k_x,M}+\lambda^2\big)\beta^2_{k_x} - \lambda^2\delta^2_{k_x}\sigma^2_{\eta_{k_x},M}\phi^{2(j-1)}_{k_x,M}\big]\Big)\mathrm{E}[|u_{i+j}|^{\lambda}],
\end{aligned}
$$

for $\lambda = 1$

$$
\begin{aligned}
&= M^{-1}\sqrt{2/\pi}\Bigg(\Big[\sum_{x=1}^{2} \delta_{k}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big]\Big[2\Phi(-\sum_{x=1}^{2} \delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}) - 1\Big] \\
&\quad + \sqrt{2/\pi} \exp\Big(-\tfrac{1}{2}\big(\sum_{x=1}^{2} \delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\big)^2\Big)\Bigg) \\
&\quad \times \exp\Big(\sum_{x=1}^{2}\big[2\alpha_{k_x} + \big(1+\phi^{j}_{k_x,M}\big)\beta^2_{k_x} - \tfrac{1}{2}\delta^2_{k_x}\sigma^2_{\eta_{k_x},M}\phi^{2(j-1)}_{k_x,M}\big]\Big).
\end{aligned}
$$

Also note that for covariances we have the result,

$$\begin{aligned}
\mathrm{Cov}(|r_{i,M}|,|r_{i+j,M}|) &= \mathrm{E}\big[|r_{i,M}||r_{i+j,M}|\big] - \mathrm{E}\big[|r_{i,M}|\big]\mathrm{E}\big[|r_{i+j,M}|\big] \\
&= \mathrm{E}\big[|r_{i,M}||r_{i+j,M}|\big] - M^{-1}\frac{2}{\pi}\exp\Big(\sum_{x=1}^{2}\big[2\alpha_{k_x}+\beta_{k_x}^2\big]\Big) \\
&= M^{-1}\sqrt{2/\pi}\exp\Big(\sum_{x=1}^{2}\big[2\alpha_{k_x}+\beta_{k_x}^2\big]\Big) \\
&\quad\times\Bigg[\Bigg(\Big[\sum_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\Big]\Big[2\Phi(-\sum_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\big)-1\Big] \\
&\qquad+\sqrt{2/\pi}\exp\Big(-\tfrac{1}{2}\big(\sum_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi_{k_x,M}^{j-1}\big)^2\Big)\Bigg) \\
&\qquad\times\exp\Big(\sum_{x=1}^{2}\big[2\alpha_{k_x}+\big(1+\phi_{k_x,M}^{j}\big)\beta_{k_x}^2-\frac{1}{2}\delta_{k_x}^2\sigma_{\eta_{k_x},M}^2\phi_{k_x,M}^{2(j-1)}\big]\Big) \\
&\qquad-\sqrt{2/\pi}\Big)\Bigg].
\end{aligned}$$

Auxiliary results for analytical moments of certain power functions of correlated normal/lognormal distributed variables

Proof of Eq. (1.A.7) and related product terms in the presence of one volatility factor. Consider the bivariate normally distributed variables $(u_{i,M},\eta_{(i+1),M})'$ from the original set of Equations (1.3.35), where we can exploit the standard property of correlated normal random variables in which

$$\eta_{(i+1),M} = v_{(i+1),M} + \delta\sigma_{\eta,M}u_{i,M},$$

with $v_{(i+1),M}$ being independent of $u_{i,M}$, and having a normal distribution with zero mean, and variance $\sigma_{\eta,M}^2(1-\delta^2)$. Then, the following expectation of interest can be factorized

$$\mathrm{E}\big[g(u_{i,M})\exp(\lambda\phi_M^p\eta_{(i+1),M})\big] = \mathrm{E}\big[V\big]\times\mathrm{E}\big[U\big],$$

with

$$\begin{aligned}
\mathrm{E}\big[V\big] &= \mathrm{E}\Big[\exp\big(\lambda\phi_M^p v_{(i+1),M}\big)\Big] \\
&= \exp\big(\tfrac{1}{2}\lambda^2\phi_M^{2p}\sigma_{\eta,M}^2(1-\delta^2)\big),
\end{aligned}$$

and for $g(u_{i,M}) = u_{i,M}$

$$\begin{aligned}
\mathrm{E}\big[U\big] &= \mathrm{E}\Big[u_{i,M}\exp\big(\lambda\phi_M^p\delta\sigma_{\eta,M}u_{i,M}\big)\Big] \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} u_{i,M}\exp\big(\lambda\phi_M^p\delta\sigma_{\eta,M}u_{i,M}\big)\exp\big(-\tfrac{1}{2}u_{i,M}^2\big)du_{i,M} \\
&= \exp\big(\tfrac{1}{2}\lambda^2\phi_M^{2p}\delta^2\sigma_{\eta,M}^2\big)\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} u_{i,M}\exp\big(-\tfrac{1}{2}(u_{i,M}-\lambda\phi_M^p\delta\sigma_{\eta,M})^2\big)du_{i,M}}_{N(\lambda\phi_M^p\delta\sigma_{\eta,M},1)} \\
&= \lambda\phi_M^p\delta\sigma_{\eta,M}\exp\big(\tfrac{1}{2}\lambda^2\phi_M^{2p}\delta^2\sigma_{\eta,M}^2\big).
\end{aligned}$$

We obtain, for $\mathbf{g(u_{i,M})} = \mathbf{u_{i,M}}$,

$$\mathrm{E}\big[M^{-\frac{1}{2}}u_{i,M}\exp(\lambda\phi_M^p\eta_{(i+1),M})\big] = M^{-\frac{1}{2}}\lambda\phi_M^p\delta\sigma_{\eta,M}\exp\big(\tfrac{1}{2}\lambda^2\phi_M^{2p}\sigma_{\eta,M}^2\big),$$

and likewise for the case of squares, $\mathbf{g(u_{i,M})} = \mathbf{u_{i,M}^2}$,

$$\mathrm{E}\big[M^{-1}u_{i,M}^2\exp(\lambda\phi_M^p\eta_{(i+1),M})\big] = M^{-1}(1+\lambda^2\phi_M^{2p}\delta^2\sigma_{\eta,M}^2)\exp\big(\tfrac{1}{2}\lambda^2\phi_M^{2p}\sigma_{\eta,M}^2\big),$$

which follows from the first moment[33] of the non–central chi–squared distribution.

For the absolute value, i.e. $\mathbf{g(u_{i,M})} = |\mathbf{u_{i,M}}|$,

$$\begin{aligned}
&\mathrm{E}\big[M^{-\frac{1}{2}}|u_{i,M}|\exp(\lambda\phi_M^p\eta_{(i+1),M})\big] \\
&= M^{-\frac{1}{2}}\times\mathrm{E}[V]\times\Big(\sqrt{2/\pi}\exp\big(-\tfrac{1}{2}\lambda^2\phi_M^{2p}\delta^2\sigma_{\eta,M}^2\big)-\lambda\phi_M^p\delta\sigma_{\eta,M}\big(1-2\Phi(-\lambda\phi_M^p\delta\sigma_{\eta,M})\big)\Big),
\end{aligned}$$

which follows from the first moment of the folded normal distribution, $\mu_1 = \frac{2}{\sqrt{2\pi}}\sigma\exp\big(-\frac{1}{2}\theta^2\big)-\mu\big(1-2\Phi(\theta)\big)$ with $\theta = -\frac{\mu}{\sigma}$, from Elandt (1961), and $\mathrm{E}[V]$ as given before. ■

Proof of Eq. (1.A.9) and related product terms in the presence of two volatility factors. Building upon the previous results above for the one volatility factor specification, consider now the multifactor case with k_x for $x = 1, 2$, where for clarity of exposition we drop subscript M of the random variables η, u, and v only, then we can re–write $\eta_{i+1,j}^*$ from Eq. (1.A.9) as

$$\begin{aligned}
\eta_{i+1,j}^* &= \lambda(\phi_{k_1,M}^{j-1}\eta_{k_1,i+1}+\phi_{k_2,M}^{j-1}\eta_{k_2,i+1}), \\
&= \sum_{x=1}^{2}\lambda\phi_{k_x,M}^{j-1}v_{k_x,i+1}+\sum_{x=1}^{2}\lambda\delta_{k_x}\phi_{k_x,M}^{j-1}\sigma_{\eta_{k_x},M}u_i,
\end{aligned}$$

[33] The degrees of freedom times the centrality parameter.

and, following the exposition for the one volatility factor case, for $g(u_i) = u_i$

$$\mathrm{E}\big[g(u_i)\exp(\eta^*_{i+1,j})\big] = \mathrm{E}\big[V^*\big] \times \mathrm{E}\big[U^*\big],$$

with

$$\begin{aligned}\mathrm{E}\big[V^*\big] &= \mathrm{E}\Big[\exp\Big(\sum_{x=1}^{2}\lambda\phi^{j-1}_{k_x,M}v_{k_x,(i+1)}\Big)\Big] \\ &= \exp\Big(\tfrac{1}{2}\sum_{x=1}^{2}\lambda^2\phi^{2(j-1)}_{k_x,M}\sigma^2_{\eta_{k_x},M}(1-\delta^2_{k_x})\Big),\end{aligned}$$

as v_{k_1} and v_{k_2} are independent, and

$$\begin{aligned}\mathrm{E}\big[U^*\big] &= \mathrm{E}\Big[u_i\exp\Big(\sum_{x=1}^{2}\lambda\delta_{k_x}\phi^{j-1}_{k_x,M}\sigma_{\eta_{k_x},M}u_i\Big)\Big] \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}u_i\exp\Big(\sum_{x=1}^{2}\lambda\delta_{k_x}\phi^{j-1}_{k_x,M}\sigma_{\eta_{k_x},M}u_i\Big)\exp\big(-\tfrac{1}{2}u_i^2\big)du_i \\ &= \exp\Big(\tfrac{1}{2}\Big(\sum_{x=1}^{2}\lambda\delta_{k_x}\phi^{j-1}_{k_x,M}\sigma_{\eta_{k_x},M}\Big)^2\Big) \\ &\quad\times\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}u_i\exp\Big(-\tfrac{1}{2}\big(u_i-\sum_{x=1}^{2}\lambda\delta_{k_x}\phi^{j-1}_{k_x,M}\sigma_{\eta_{k_x},M}\big)^2\Big)du_i}_{N\left(\sum_{x=1}^{2}\lambda\delta_{k_x}\phi^{j-1}_{k_x,M}\sigma_{\eta_{k_x},M},1\right)} \\ &= \sum_{x=1}^{2}\lambda\delta_{k_x}\phi^{j-1}_{k_x,M}\sigma_{\eta_{k_x},M}\exp\Big(\tfrac{1}{2}\Big(\sum_{x=1}^{2}\lambda\delta_{k_x}\phi^{j-1}_{k_x,M}\sigma_{\eta_{k_x},M}\Big)^2\Big),\end{aligned}$$

for two factors there remains a non–cancelling cross–product term involving δ_{k_x} from the quadratic expansion, and we obtain for $\mathbf{g}(\mathbf{u_i}) = \mathbf{u_i}$

$$\begin{aligned}\mathrm{E}\big[M^{-\frac{1}{2}}u_i\exp(\eta^*_{i+1,j})\big] &= M^{-\frac{1}{2}}\sum_{x=1}^{2}\lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M} \\ &\quad\times\exp\Big(\tfrac{1}{2}\big(\sum_{x=1}^{2}\lambda^2\sigma^2_{\eta_{k_x},M}\phi^{2(j-1)}_{k_x,M}\big)+\lambda^2\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M}\Big)\end{aligned} \tag{1.A.12}$$

and likewise for the case of squares, $\mathbf{g}(\mathbf{u_i}) = \mathbf{u_i^2}$,

$$\begin{aligned}\mathrm{E}\big[M^{-1}u_i^2\exp(\eta^*_{i+1,j})\big] &= M^{-1}\Big(1+\big(\sum_{x=1}^{2}\lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M}\big)^2\Big) \\ &\quad\times\exp\Big(\tfrac{1}{2}\big(\sum_{x=1}^{2}\lambda^2\sigma^2_{\eta_{k_x},M}\phi^{2(j-1)}_{k_x,M}\big)+\lambda^2\prod_{x=1}^{2}\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M}\Big).\end{aligned} \tag{1.A.13}$$

For the absolute value, i.e. $\mathbf{g}(\mathbf{u_i}) = |\mathbf{u_i}|$,

$$\begin{aligned}\mathrm{E}\big[M^{-\frac{1}{2}}|u_i|\exp(\eta^*_{i+1,j})\big] &= M^{-\frac{1}{2}} \times \mathrm{E}[V^*] \\ &\times\Bigg(\sqrt{2/\pi}\exp\Big(-\tfrac{1}{2}\big(\sum_{x=1}^{2}\lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M}\big)^2\Big) \\ &-\big(\sum_{k=1}^{2}\lambda\delta_k\sigma_{\eta_k,M}\phi^{j-1}_{k,M}\big)\big(1-2\Phi(-\sum_{x=1}^{2}\lambda\delta_{k_x}\sigma_{\eta_{k_x},M}\phi^{j-1}_{k_x,M})\big)\Bigg).\end{aligned} \tag{1.A.14}$$

■

1.B Tables for Chapter 1

Table 1.B.1: Differences in non– to centralized realized skewness as a function of mean and total variation only over horizons up to three months

Differences are reported for $\mathrm{E}\left[\left(r_t^{(N)}\right)^3\right] - \mathrm{E}\left[\left(r_t^{(N)} - \mu_t^{(N)}\right)^3\right]$ based on Eq. (1.3.9) with the annualized expected return being either 3% (Panel A) or 8% (Panel B) scaled by the total variation in the range of 8%–40% volatility in annualized terms to the power 3/2 over the respective horizon of interest.

Vol. p.a.	1 day	2 days	1 week	10 days	1 month	2 months	3 months
Panel A				3% expected return p.a.			
8%	0.07	0.10	0.16	0.22	0.33	0.46	0.57
12%	0.05	0.07	0.11	0.15	0.22	0.31	0.38
16%	0.04	0.05	0.08	0.11	0.16	0.23	0.28
20%	0.03	0.04	0.06	0.09	0.13	0.18	0.23
40%	0.01	0.02	0.03	0.04	0.06	0.09	0.11
Panel B				8% expected return p.a.			
8%	0.19	0.27	0.43	0.61	0.89	1.29	1.63
12%	0.13	0.18	0.28	0.40	0.58	0.84	1.04
16%	0.09	0.13	0.21	0.30	0.44	0.62	0.77
20%	0.08	0.11	0.17	0.24	0.35	0.49	0.61
40%	0.04	0.05	0.08	0.12	0.17	0.25	0.30

Table 1.B.2: The contribution of third powers of high–freqeuncy returns to non–centralized skewness over horzions up to three months

Reported are the contributions to non–centralized skewness measure, $\mathrm{E}\left[\left(r_t^{(N)}\right)^3\right]$, of third powers of high–frequency returns, in specific $\left(1+(N-1)^3\right)\mu_t^3 + 3\sum_{i=1}^{N}\mu_t \mathrm{Var}(r_{t,i})$ in Eq. (1.3.16), scaled by the total variation to the power 3/2 over the respective horizon of interest for expected returns being either 3% (Panel A) or 8% (Panel B) and volatility in the range of 8%–40% in annualized terms.

Vol. p.a.	1 day	2 days	1 week	10 days	1 month	2 months	3 months
Panel A			3% expected return p.a.				
8%	9.2E-04	6.8E-04	5.5E-04	7.0E-04	1.5E-03	3.7E-03	6.7E-03
12%	6.1E-04	4.4E-04	3.1E-04	3.1E-04	5.1E-04	1.2E-03	2.0E-03
16%	4.6E-04	3.3E-04	2.2E-04	2.0E-04	2.6E-04	5.2E-04	8.8E-04
20%	3.6E-04	2.6E-04	1.7E-04	1.4E-04	1.6E-04	2.9E-04	4.7E-04
40%	1.8E-04	1.3E-04	8.2E-05	6.1E-05	5.0E-05	5.7E-05	7.6E-05
Panel B			8% expected return p.a.				
8%	2.7E-03	2.4E-03	3.9E-03	8.6E-03	2.5E-02	6.8E-02	1.3E-01
12%	1.7E-03	1.3E-03	1.5E-03	2.8E-03	7.5E-03	2.0E-02	3.7E-02
16%	1.2E-03	9.4E-04	8.9E-04	1.4E-03	3.3E-03	8.7E-03	1.6E-02
20%	9.8E-04	7.3E-04	6.1E-04	8.1E-04	1.7E-03	4.5E-03	8.1E-03
40%	4.9E-04	3.5E-04	2.4E-04	2.2E-04	3.0E-04	6.2E-04	1.1E-03

Table 1.B.3: CME S&P 500 futures contract rollover dates, 1988H–2013H

Trading floor convention is to roll the expiring quarterly futures contract eight calendar days before the contract expires. From the rollover date on, it is customary to identify the second nearest expiration month as the 'lead month' for the index futures as the nearest expiring contract will terminate soon and will have a less liquid market than the new 'lead month' contract (CME 2013). This convention is followed here unless the lag (lead) month has still (already) higher liquidity in terms of number of transactions per day. The days below refer to the rollover days, i.e. the first trading day of the lead month future contract.

1988-01-04					
1988-03-10	1989-03-09	1990-03-08	1990-03-08	1991-03-07	1992-03-12
1988-06-09	1989-06-08	1990-06-07	1990-06-07	1991-06-13	1992-06-11
1988-09-08	1989-09-07	1990-09-13	1990-09-13	1991-09-12	1992-09-10
1988-12-08	1989-12-07	1990-12-13	1990-12-13	1991-12-12	1992-12-10
1993-03-11	1994-03-10	1995-03-09	1996-03-07	1997-03-13	1998-03-12
1993-06-10	1994-06-09	1995-06-08	1996-06-13	1997-06-12	1998-06-11
1993-09-09	1994-09-08	1995-09-07	1996-09-12	1997-09-11	1998-09-10
1993-12-09	1994-12-08	1995-12-07	1996-12-12	1997-12-11	1998-12-10
1999-03-11	2000-03-09	2001-03-08	2002-03-07	2003-03-13	2004-03-11
1999-06-10	2000-06-08	2001-06-07	2002-06-13	2003-06-12	2004-06-10
1999-09-09	2000-09-07	2001-09-11	2002-09-12	2003-09-11	2004-09-09
1999-12-09	2000-12-07	2001-12-13	2002-12-12	2003-12-11	2004-12-09
2005-03-10	2006-03-09	2007-03-03	2008-03-13	2009-03-12	2010-03-09
2005-06-09	2006-06-08	2007-06-07	2008-06-12	2009-06-11	2010-06-10
2005-09-08	2006-09-07	2007-09-13	2008-09-04	2009-09-10	2010-09-09
2005-12-08	2006-12-07	2007-12-13	2008-12-11	2009-12-10	2010-12-09
2011-03-10	2012-03-08				
2011-06-09	2012-06-07				
2011-09-08	2012-09-13				
2011-12-16	2012-12-12				

Table 1.B.4: FOMC announcement days for SP futures, 1988–2002

Source: Federal Reserve Board (FRB) available at www.federalreserve.gov.

1988-02-04	1988-08-17	1989-02-02	1989-08-16	1990-02-03	1990-08-15
1988-03-30	1988-10-05	1989-03-22	1989-09-20	1990-03-21	1990-09-26
1988-05-18	1988-11-16	1989-05-17	1989-11-15	1990-05-16	1990-11-14
1988-06-30	1988-12-21	1989-07-06	1989-12-20	1990-07-05	1990-12-19
1991-01-31	1991-08-21	1992-01-30	1992-08-19	1993-02-04	1993-08-18
1991-03-27	1991-09-25	1992-03-25	1992-09-23	1993-03-24	1993-09-29
1991-05-22	1991-11-14	1992-05-20	1992-11-12	1993-05-19	1993-11-11
1991-07-05	1991-12-18	1992-07-02	1992-12-16	1993-07-01	1993-12-15
1994-02-03	1994-08-17	1995-02-02	1995-08-23	1996-02-01	1996-08-21
1994-03-30	1994-09-28	1995-03-29	1995-10-04	1996-03-20	1996-10-02
1994-05-18	1994-11-16	1995-05-17	1995-11-15	1996-05-15	1996-11-14
1994-06-30	1994-12-21	1995-06-29	1995-12-20	1996-06-27	1996-12-18
1997-01-30	1997-08-20	1998-01-29	1998-08-12	1999-01-28	1999-08-11
1997-03-19	1997-10-01	1998-03-18	1998-09-23	1999-03-17	1999-09-15
1997-05-14	1997-11-05	1998-05-06	1998-11-05	1999-05-05	1999-10-27
1997-06-26	1997-12-10	1998-06-25	1998-12-09	1999-06-24	1999-12-08
2000-01-27	2000-08-09	2001-02-01	2001-08-08	2002-01-30	2002-08-07
2000-03-15	2000-09-20	2001-03-21	2001-09-19	2002-03-27	2002-09-19
2000-05-03	2000-11-09	2001-05-02	2001-10-31	2002-05-09	2002-10-24
2000-06-29	2000-12-13	2001-06-29	2001-12-12	2002-06-28	2002-12-11

Table 1.B.5: FOMC announcement days for ES futures and SPY ETF, 2003–2013

Source: Federal Reserve Board (FRB) available at www.federalreserve.gov.

2003-01-29	2003-06-25	2003-12-09	2004-06-30	2004-12-14	2005-06-30
2003-03-18	2003-08-12	2004-01-28	2004-08-10	2005-02-02	2005-08-09
2003-03-25	2003-09-16	2004-03-16	2004-09-21	2005-05-03	2005-09-20
2003-05-06	2003-10-28	2004-05-04	2004-11-10	2005-06-22	2005-11-01
2005-12-13	2006-06-29	2006-12-12	2007-06-28	2007-09-18	2008-01-10
2006-01-31	2006-08-08	2007-01-31	2007-08-07	2007-10-31	2008-01-22
2006-03-28	2006-09-20	2007-03-21	2007-08-10	2007-12-06	2008-01-30
2006-05-10	2006-10-25	2007-05-09	2007-08-16	2007-12-11	2008-03-18
2008-04-30	2008-09-29	2009-03-18	2009-08-12	2010-01-27	2010-06-23
2008-06-25	2008-10-29	2009-04-29	2009-09-23	2010-03-16	2010-08-10
2008-08-05	2008-12-16	2009-06-03	2009-11-04	2010-04-28	2010-09-21
2008-09-16	2009-01-28	2009-06-24	2009-12-16	2010-05-10	2010-11-03
2010-12-14	2011-06-22	2011-12-13	2012-06-20	2012-12-12	2013-06-19
2011-01-26	2011-08-09	2012-01-25	2012-08-01	2013-01-30	2013-07-31
2011-03-15	2011-09-21	2012-03-13	2012-09-13	2013-03-20	2013-09-18
2011-04-27	2011-11-02	2012-04-25	2012-10-24	2013-05-01	2013-10-30
2013-12-18					

Table 1.B.6: Consumer credit (CC) announcement days for ES futures and SPY ETF, 2003–2013

Source: Bureau of the Fiscal Service (BFS) available at www.fiscal.treasury.gov.

2003-01-08	2003-05-07	2003-09-08	2004-01-08	2004-05-07	2004-09-08
2003-02-07	2003-06-06	2003-10-07	2004-02-06	2004-06-07	2004-10-07
2003-03-07	2003-07-08	2003-11-07	2004-03-05	2004-07-08	2004-11-05
2003-04-07	2003-08-07	2003-12-05	2004-04-07	2004-08-06	2004-12-07
2005-01-07	2005-05-06	2005-09-08	2006-01-09	2006-05-05	2006-09-08
2005-02-07	2005-06-07	2005-10-07	2006-02-07	2006-06-07	2006-10-06
2005-03-07	2005-07-08	2005-11-07	2006-03-07	2006-07-10	2006-11-07
2005-04-07	2005-08-05	2005-12-07	2006-04-07	2006-08-07	2006-12-07
2007-01-08	2007-05-07	2007-09-10	2008-01-10	2008-05-07	2008-09-08
2007-02-07	2007-06-07	2007-10-05	2008-02-07	2008-06-06	2008-10-07
2007-03-07	2007-07-09	2007-11-07	2008-03-07	2008-07-08	2008-11-07
2007-04-06	2007-08-07	2007-12-07	2008-04-07	2008-08-07	2008-12-05
2009-01-08	2009-05-07	2009-09-08	2010-01-08	2010-05-07	2010-09-08
2009-02-06	2009-06-05	2009-10-07	2010-02-05	2010-06-07	2010-10-07
2009-03-06	2009-07-08	2009-11-06	2010-03-05	2010-07-08	2010-11-05
2009-04-07	2009-08-07	2009-12-07	2010-04-07	2010-08-06	2010-12-07
2011-01-07	2011-05-06	2011-09-08	2012-01-09	2012-05-07	2012-09-10
2011-02-07	2011-06-07	2011-10-07	2012-02-07	2012-06-07	2012-10-05
2011-03-07	2011-07-08	2011-11-07	2012-03-07	2012-07-09	2012-11-07
2011-04-07	2011-08-05	2011-12-07	2012-04-06	2012-08-07	2012-12-07
2013-01-08	2013-05-07	2013-09-09			
2013-02-07	2013-06-07	2013-10-07			
2013-03-07	2013-07-08	2013-11-07			
2013-04-05	2013-08-07	2013-12-06			

Table 1.B.7: Monthly treasury statement (MTS) announcement days for ES futures and SPY ETF, 2003–2013

Source: Bureau of the Fiscal Service (BFS) available at www.fiscal.treasury.gov.

2003-01-21	2003-05-20	2003-09-17	2004-01-15	2004-05-12	2004-09-13
2003-02-21	2003-06-19	2003-10-20	2004-02-12	2004-06-10	2004-10-14
2003-03-20	2003-07-18	2003-11-17	2004-03-11	2004-07-13	2004-11-10
2003-04-18	2003-08-19	2003-12-12	2004-04-13	2004-08-11	2004-12-10
2005-01-12	2005-05-11	2005-09-13	2006-01-12	2006-05-10	2006-09-13
2005-02-10	2005-06-10	2005-10-14	2006-02-10	2006-06-12	2006-10-12
2005-03-10	2005-07-13	2005-11-10	2006-03-10	2006-07-13	2006-11-13
2005-04-12	2005-08-10	2005-12-12	2006-04-12	2006-08-10	2006-12-12
2007-01-12	2007-05-10	2007-09-13	2008-01-11	2008-05-12	2008-09-11
2007-02-12	2007-06-12	2007-10-11	2008-02-12	2008-06-11	2008-10-14
2007-03-12	2007-07-12	2007-11-13	2008-03-12	2008-07-11	2008-11-13
2007-04-11	2007-08-10	2007-12-12	2008-04-10	2008-08-12	2008-12-10
2009-01-13	2009-05-12	2009-09-11	2010-01-13	2010-05-12	2010-09-13
2009-02-11	2009-06-10	2009-10-16	2010-02-17	2010-06-10	2010-10-15
2009-03-11	2009-07-13	2009-11-12	2010-03-10	2010-07-13	2010-11-10
2009-04-10	2009-08-12	2009-12-10	2010-04-12	2010-08-11	2010-12-10
2011-01-12	2011-05-11	2011-09-13	2012-01-12	2012-05-10	2012-09-13
2011-02-10	2011-06-10	2011-10-14	2012-02-10	2012-06-12	2012-10-15
2011-03-10	2011-07-13	2011-11-10	2012-03-12	2012-07-12	2012-11-13
2011-04-12	2011-08-10	2011-12-12	2012-04-11	2012-08-10	2012-12-12
2013-01-11	2013-05-10	2013-09-12			
2013-02-12	2013-06-12	2013-10-15			
2013-03-12	2013-07-11	2013-11-13			
2013-04-10	2013-08-12	2013-12-11			

Table 1.B.8: NYSE early closure trading days at 13.00 EST, 2003–2013

Source: NYSE special closing list until 2011 (available from the author), thereafter NYSE website. Verified with TAQ database trading activity measures, i.e. number of trades and trading volume after 1.00pm.

2003-07-03	2004-11-26	2007-07-03	2008-11-28	2010-11-26	2012-12-24
2003-11-28	2005-11-25	2007-11-23	2008-12-24	2011-11-25	2013-07-03
2003-12-24	2006-07-03	2007-12-24	2009-11-27	2012-07-03	2013-11-29
2003-12-26	2006-11-24	2008-07-03	2009-12-24	2012-11-23	2013-12-24

1.C Figures for Chapter 1

Figure 1.C.1: MPV–type based RSkew vis–à–vis the standard RSkew estimator (long time span)

UPPER PANEL: Realized skewness as a function of the return horizon based on the aggregation formula, Eq. (1.3.13) for the SPY ETF with distinct intraday volatility pattern adjustment for weekdays and announcements days up to 4,651 5–minute increments yielding a maximum return horizon of three months. Data spans the time period from Dec., 12th 2002 – Mar., 6th 2013. RED: Standard RSkew measure, ORANGE: TPV–based RSkew measure. GOLD: BV–based RSkew measure. GREY: MedRV–based RSkew measure. Note GREY is the most negative of the four estimators considered. Here we want to emphasize that estimators behave very similar. **LOWER PANEL:** Zoom into the finite sample behaviour of the different estimator when the return horizon of interest is very short. SOLID LINES (BLUE and RED): Standard RSkew estimates for SPY ETF and ES E–mini futures. DOTTED LINES (BLUE and RED): TPV–based RSkew estimates which virtually coincidence for the two assets considered.

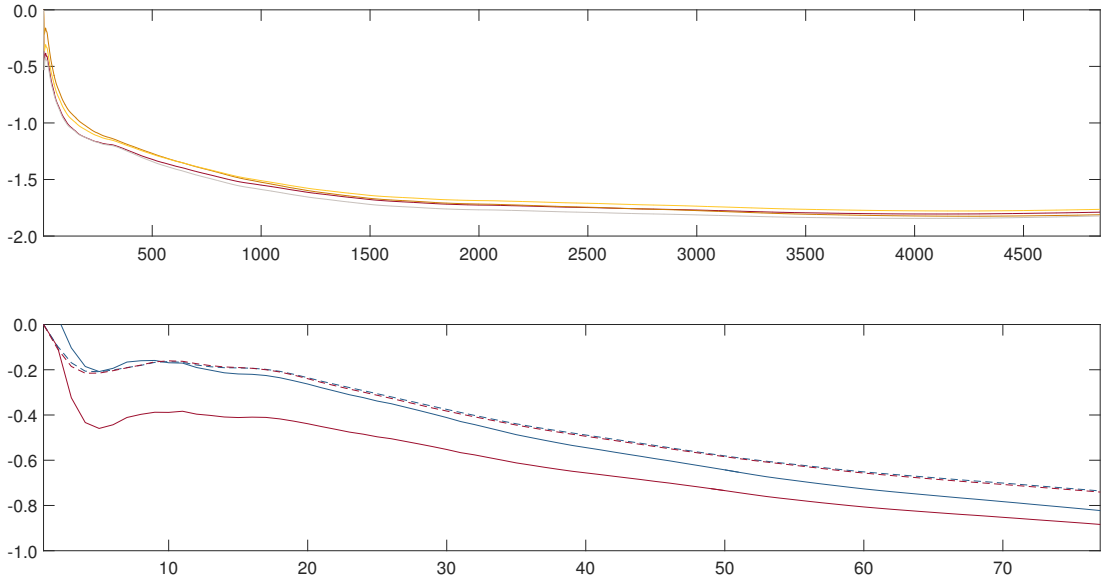
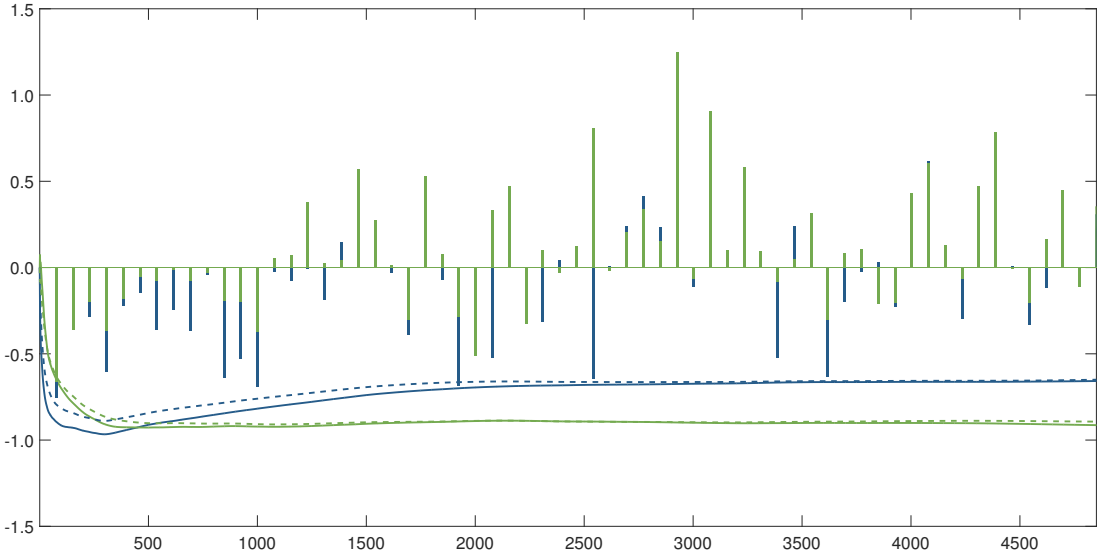


Figure 1.C.2: Standard and MPV–based RSkew vs. classical skewness over extending horizons (SP futures)

This figure is equivalent to Figure 1.5 but for the SP futures data over the immediately preceding ten year period, i.e. from Sept., 18th 1992 – Dec., 11th 2002 shown in BLUE and from Jan., 5th 1988 – Mar., 9th 1999 in GREEN, approximately 2/3 of the data is overlapping. SOLID LINES: Standard RSkew estimates as a function of the horizon based on the aggregation formula, Eq. (1.3.13) with distinct intraday volatility pattern adjustment for weekdays and announcements days. The denominator of the ratio is based on the realized kernel methodology. DOTTED LINES: TPV–based RSkew estimates. BARS: Classic skewness estimator for respective series and horizon (for integer number of days only). One week (1w) contains 5×77 increments up to 4,651 increments for the three month (3m) period. All estimates based on 5–minute sampling frequency.



Notes for Tables 1.C.4 – 1.C.41

For each of the stocks considered here, sorted columnwise unless stated otherwise, we present standard RSkew, i.e. RSkew–RV, RSkew–TPV, and RSkew–MedRV estimates as a function of enlargening return horizons. The four rows from top to bottom refer to the respective estimates when sampling takes place at either 5–, 11–, 35–, or 55–minute frequency. Each of the twelve individual subfigures presented on each page is split into four sets. The first set of nine (DARK BLUE to YELLOW) colored bars in each of the graphs refers to the realized skewness building up from asymmetric volatility (AVL), the second set refers to the realized skewness building up from volatility feedback (VFB), and the third set refers to the realized skewness building up from the instantaneous skewness (IA). The last set represents the sum of these three components, see Eq. 1.3.11 for details. The bars themselves show the estimated magnitudes for enlargening return horizons ranging from 1–, 2–, 3–day(s) in BLUE, 1–, 2–, 3–week(s) in GREEN, and 1–, 2–, 3–month(s) in YELLOW. The data spans the time period from Dec., 12th 2002 – Mar., 6th 2013 with distinct intraday volatility pattern adjustment for weekdays and announcements days for each stock based on the previously employed Nadaraya–Watson kernel regression method (not shown for brevity). The denominator of the ratio is based on the realized kernel methodology at 15–second frequency. For 5–minute and 11–minute frequency the MPV–type RSkew estimates are calculated based on the algorithm illustrated in Figure 1.1. As for the 35– and 55–minute frequency this would translate in a serious loss of data, we construct MPV–based spot measures from data spanning consecutive days as well.

CAT. A STOCKS: A-1: CAT, GE, MO, IBM, DD. **A-2:** MSFT, KO, KRFT, T, MRK, BAC.

Figure 1.C.4: Category A-1 stocks: RSkew–RV measure estimates across frequencies for CAT, GE, and MO

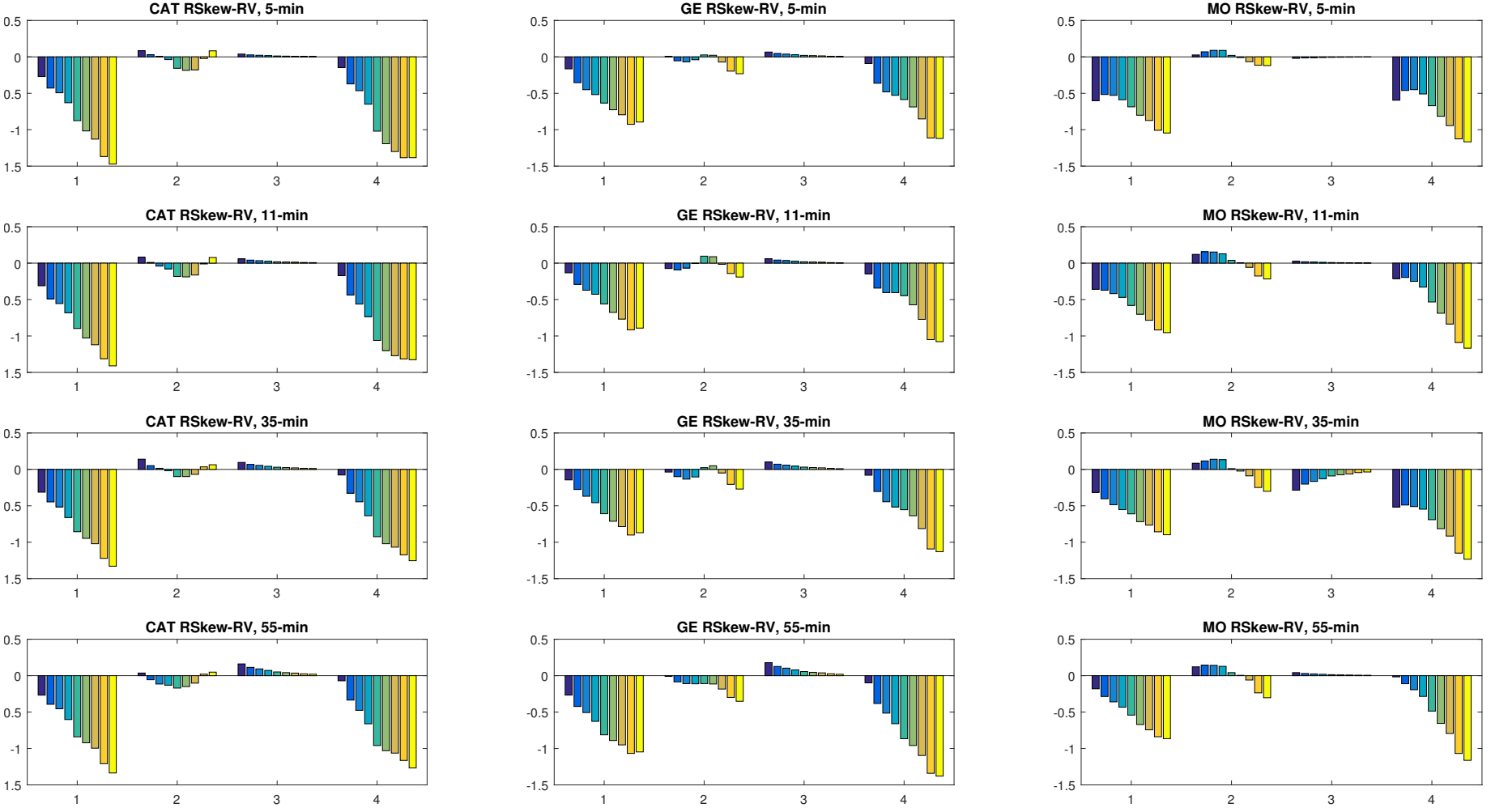


Figure 1.C.5: Category A-1 stocks: RSkew–TPV measure estimates across frequencies for CAT, GE, and MO

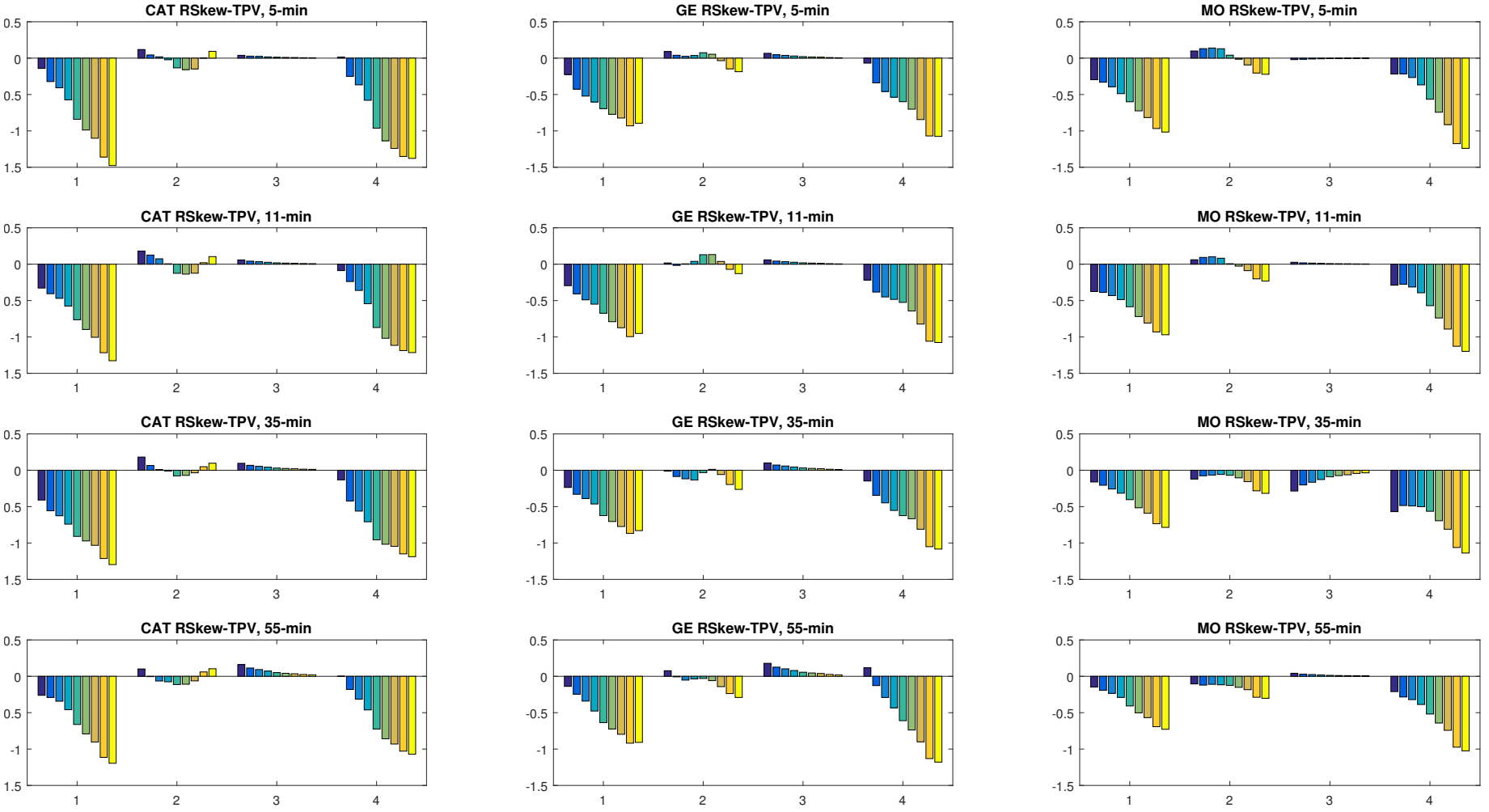


Figure 1.C.6: Category A-1 stocks: RSkew–MedRV measure estimates across frequencies for CAT, GE, and MO

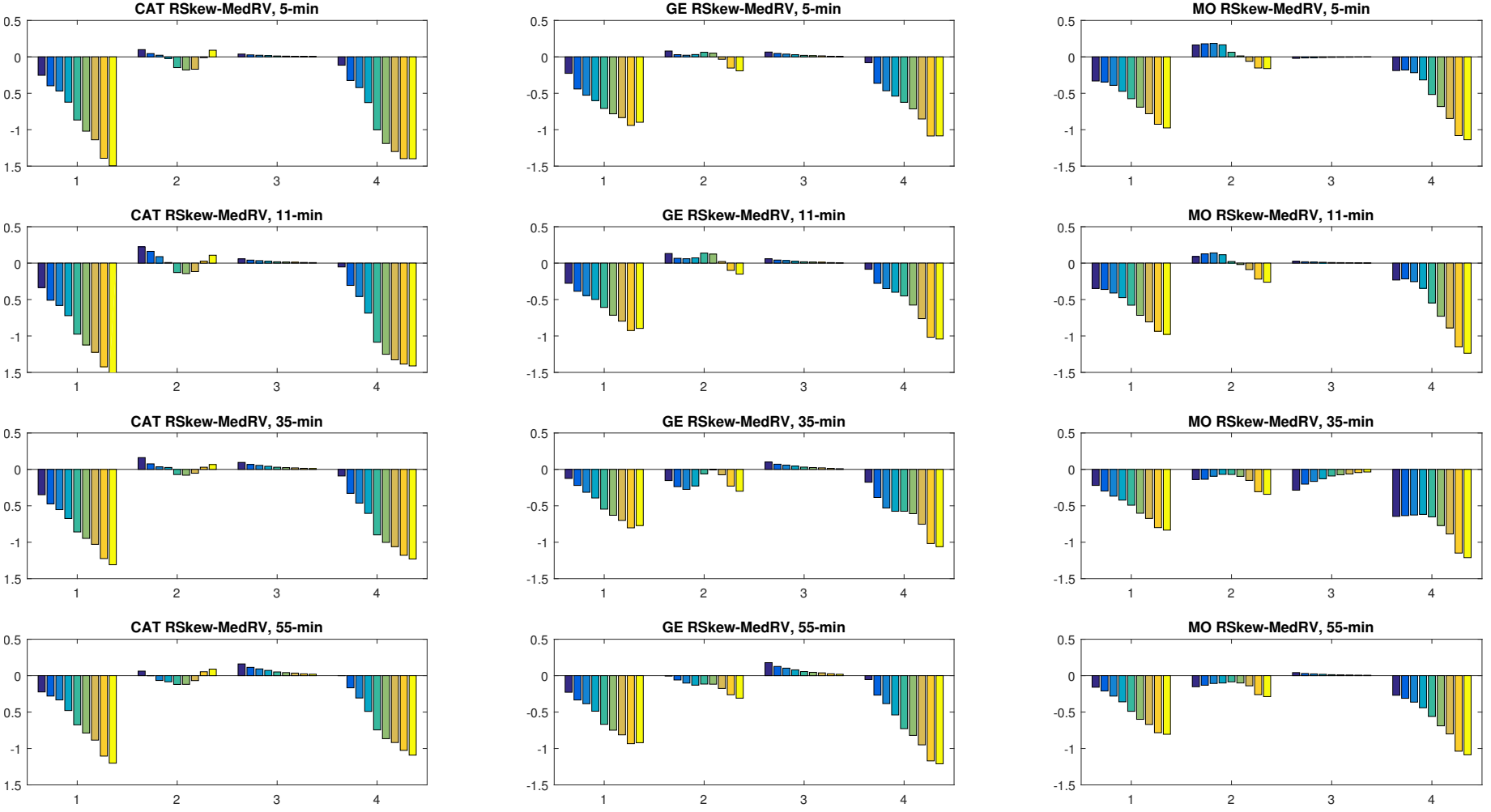


Figure 1.C.7: Category A-1 stocks: RSkew–RV measure estimates across frequencies for IBM and DD

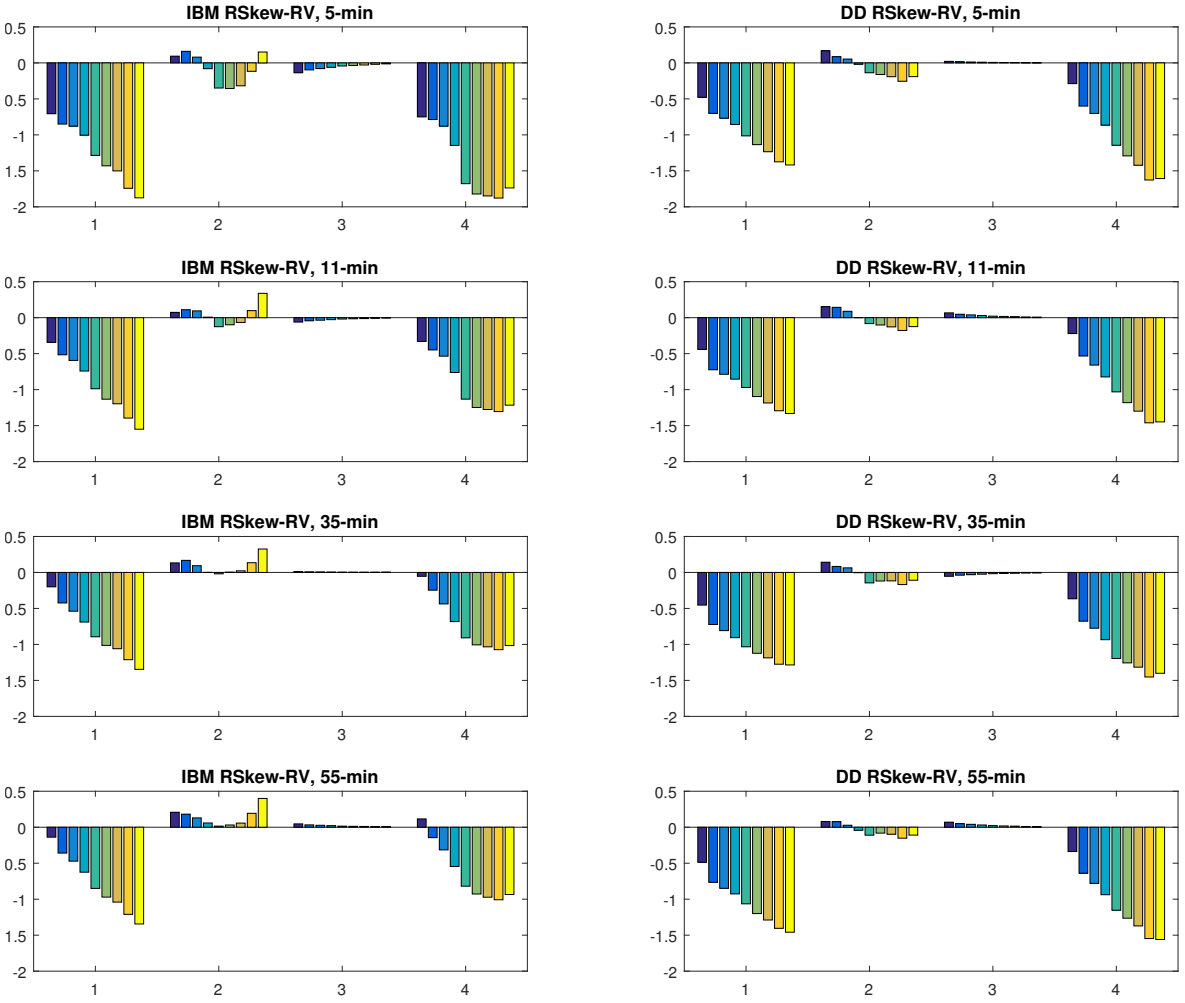


Figure 1.C.8: Category A-1 stocks: RSkew–TPV measure estimates across frequencies for IBM and DD

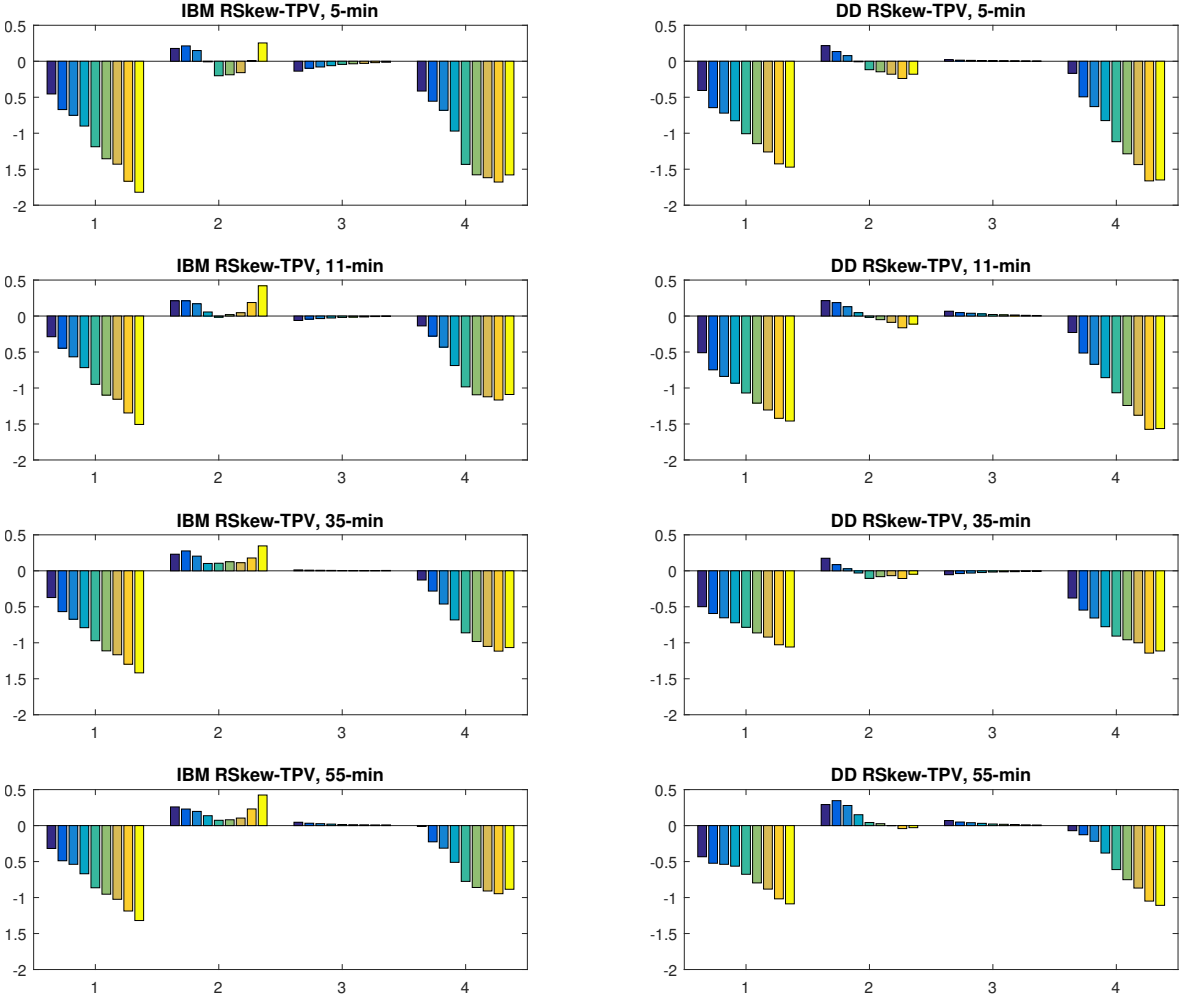


Figure 1.C.9: Category A–1 stocks: RSkew–MedRV measure estimates across frequencies for IBM and DD

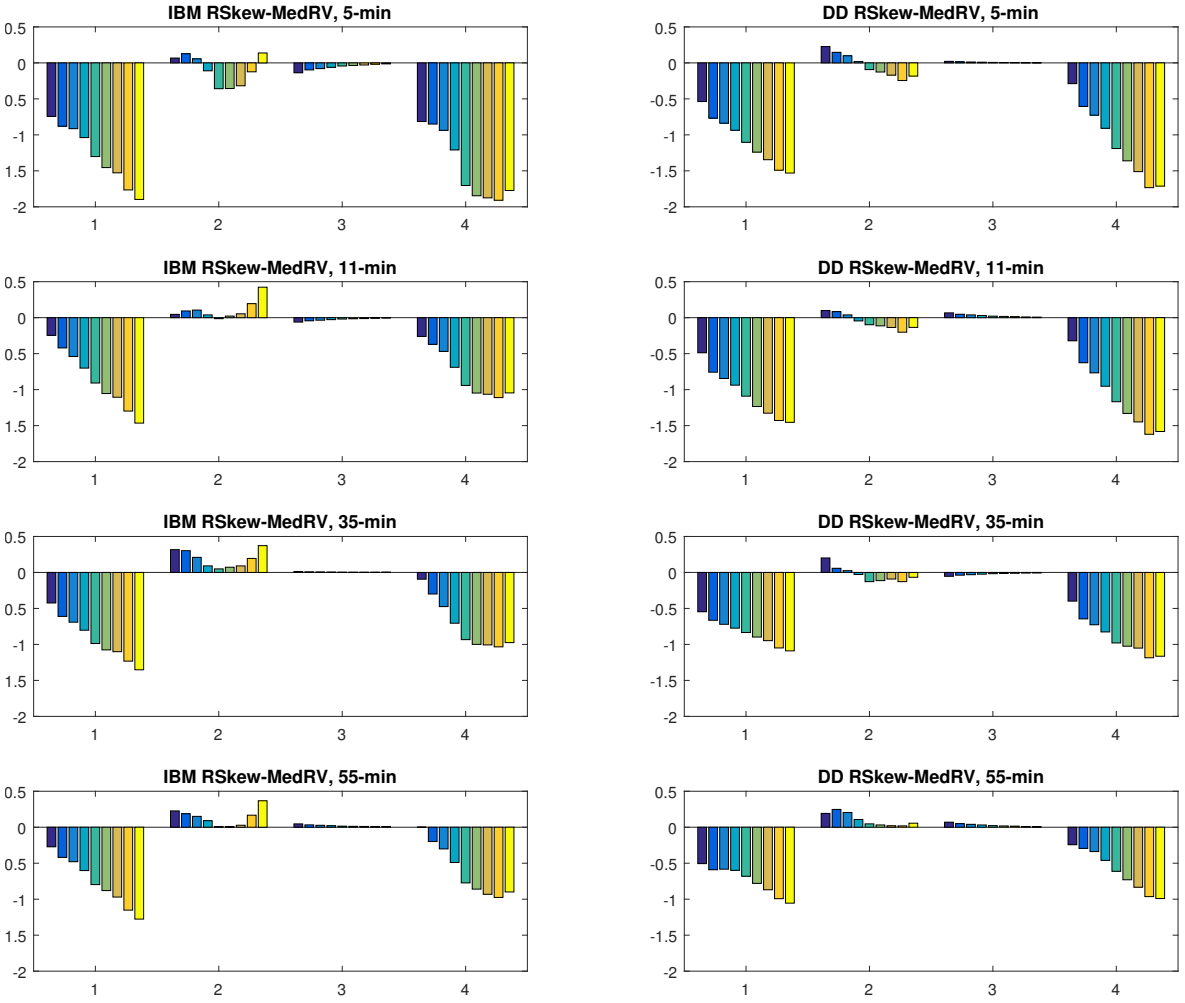


Figure 1.C.10: Category A-2 stocks: RSkew–RV measure estimates across frequencies for MSFT, KO, and KRFT

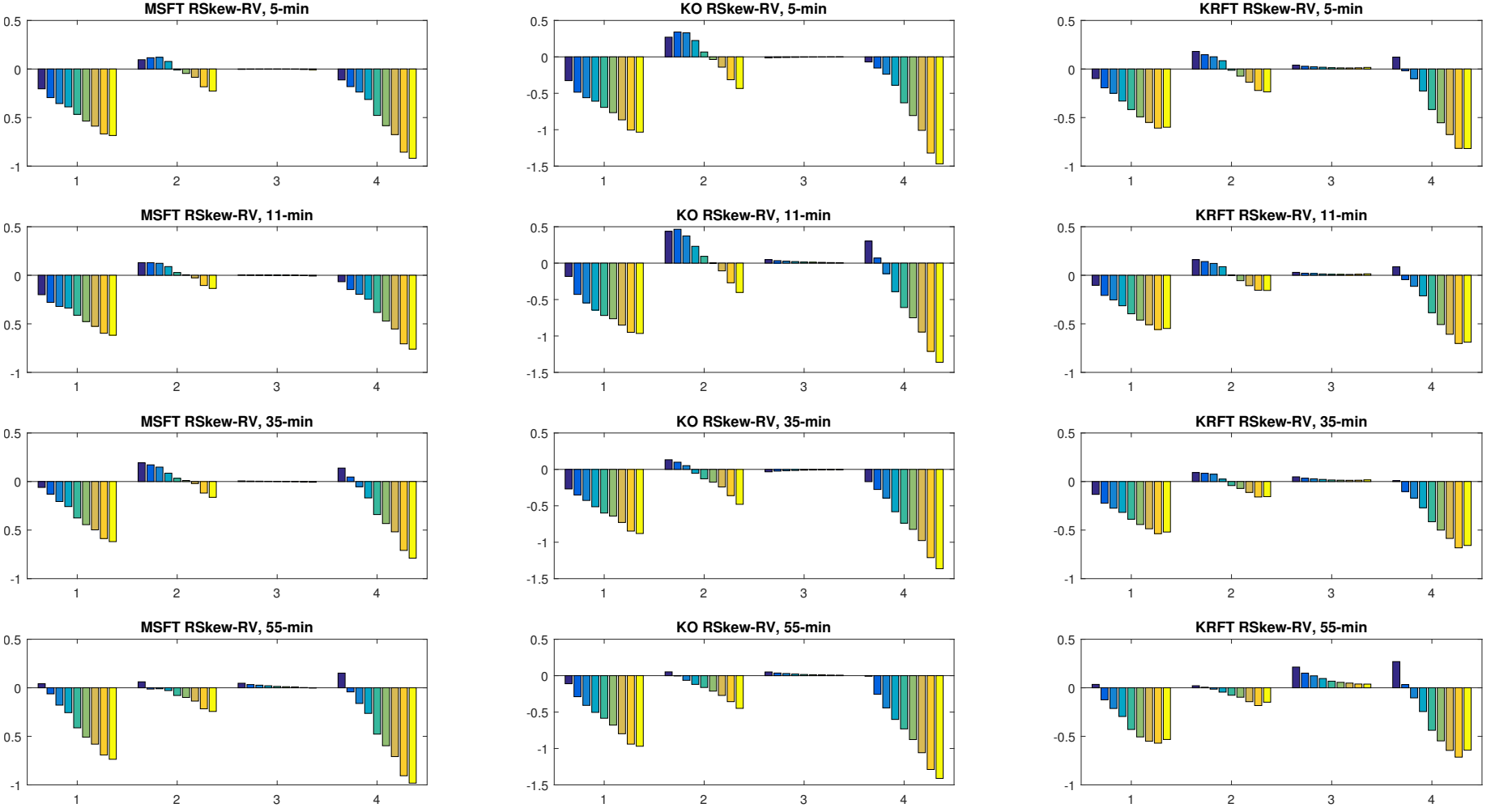


Figure 1.C.11: Category A-2 stocks: RSkew–TPV measure estimates across frequencies for MSFT, KO, and KRFT

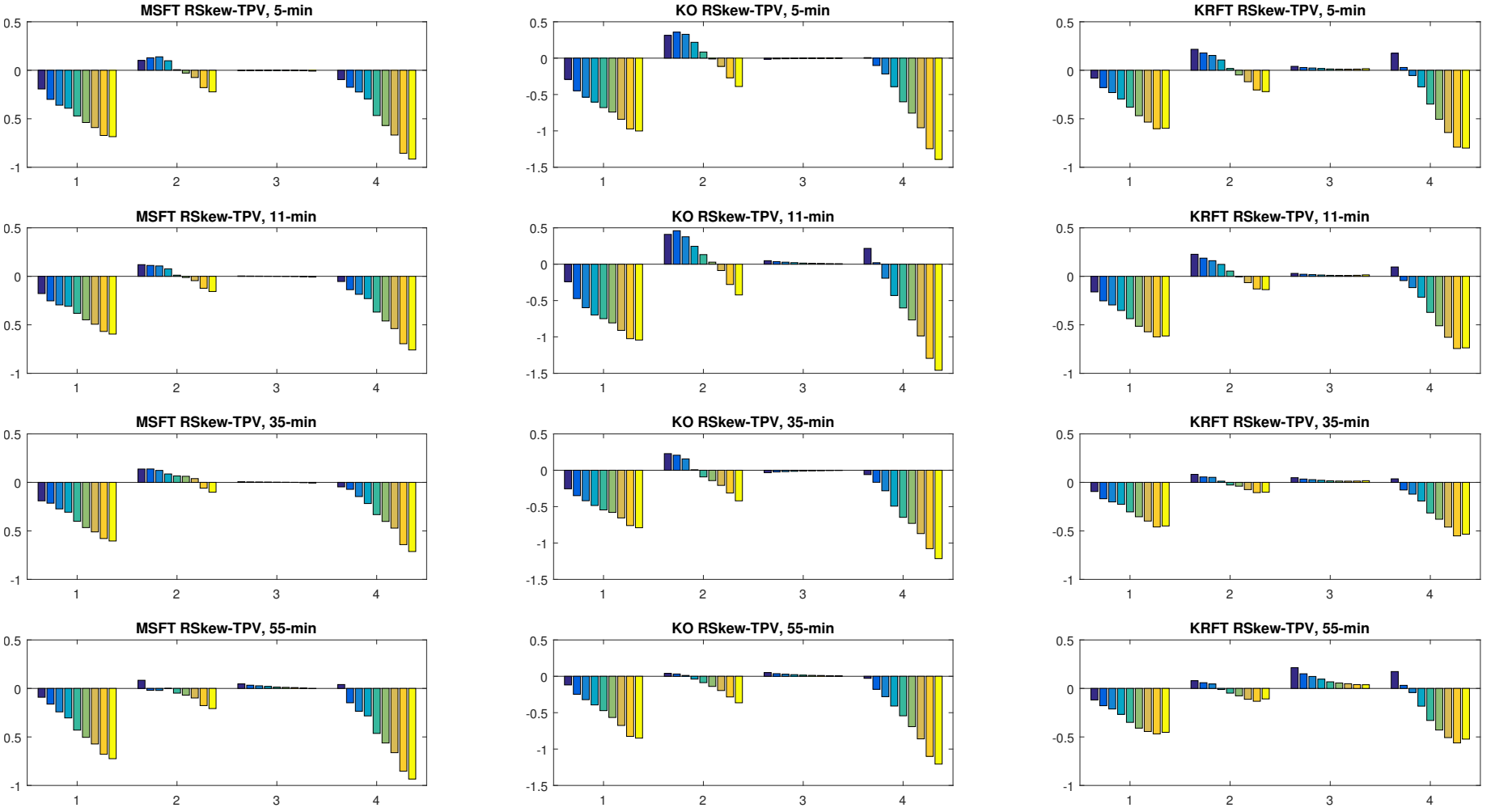


Figure 1.C.12: Category A-2 stocks: RSkew–MedRV measure estimates across frequencies for MSFT, KO, and KRFT

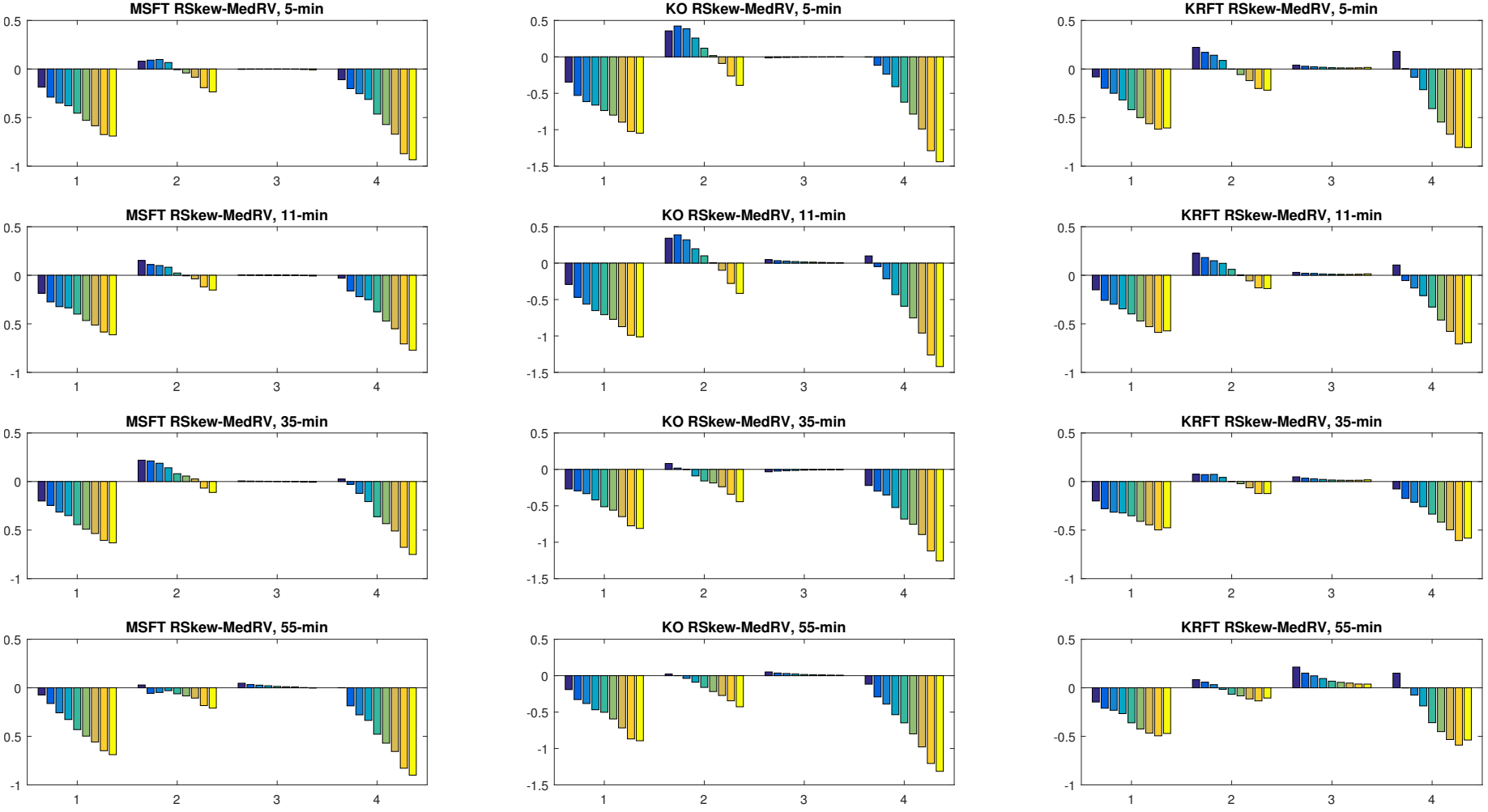


Figure 1.C.13: Category A-2 stocks: RSkew–RV measure estimates across frequencies for T, MRK, and BAC

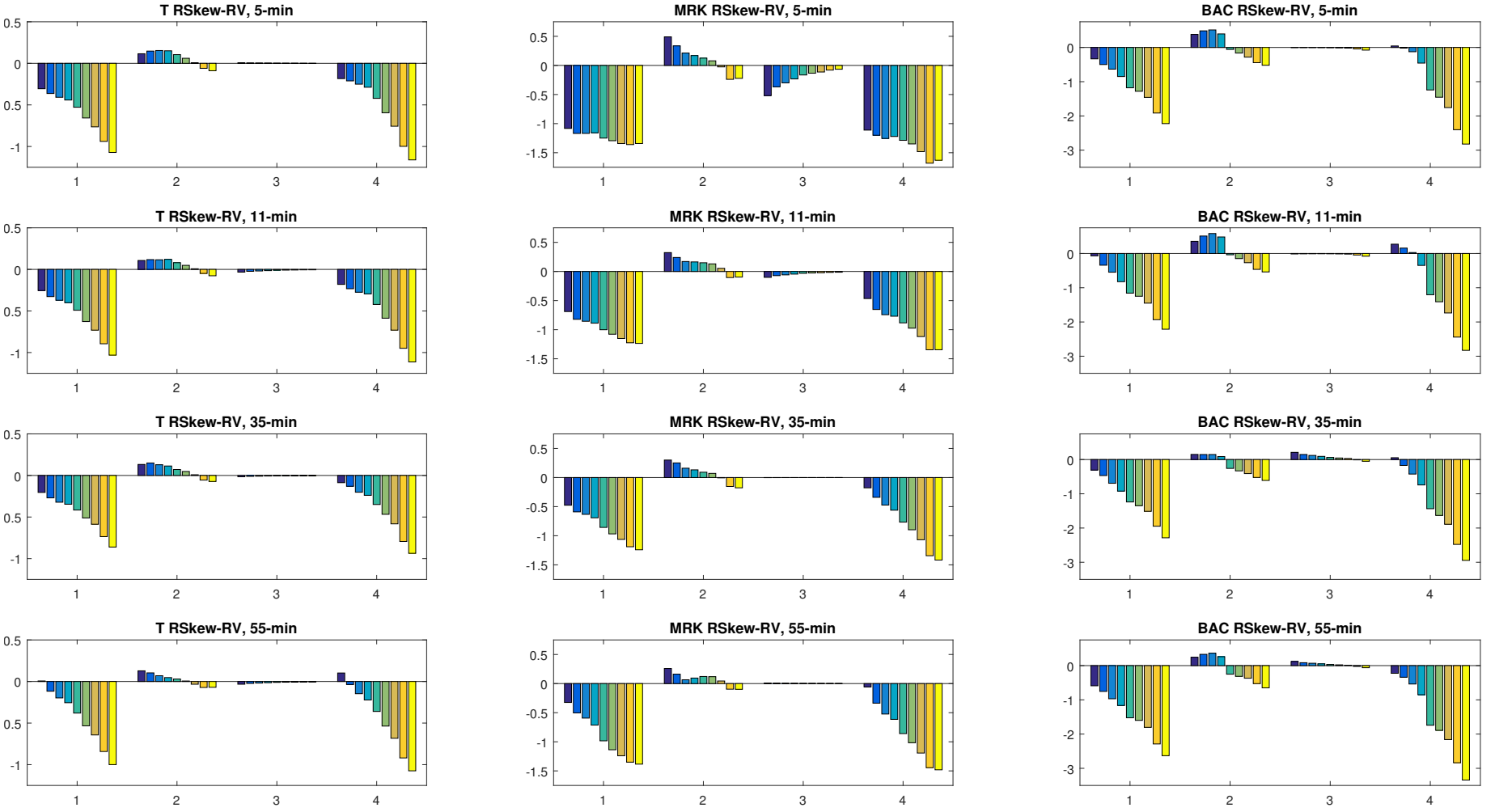


Figure 1.C.14: Category A-2 stocks: RSkew–TPV measure estimates across frequencies for T, MRK, and BAC

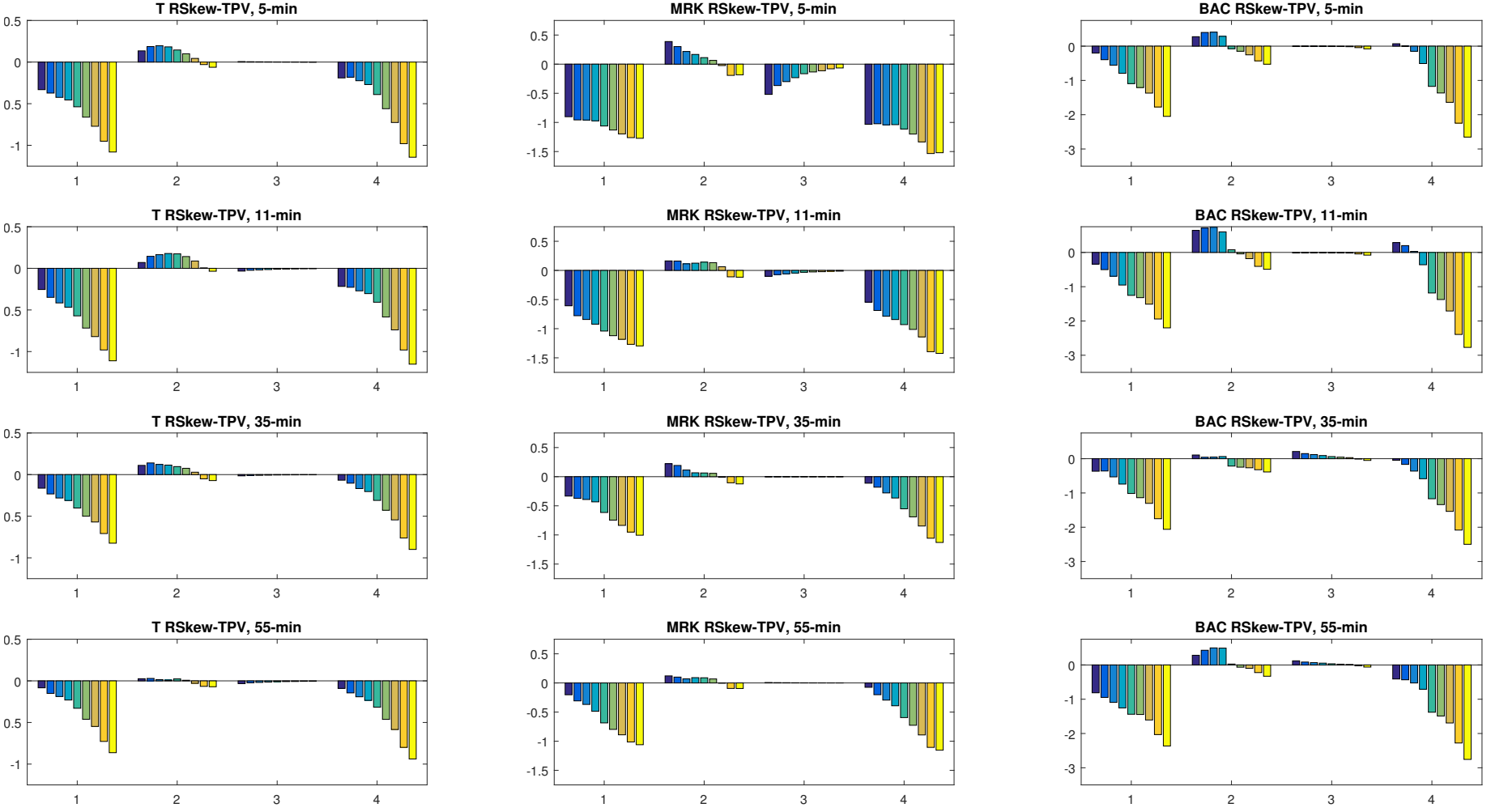
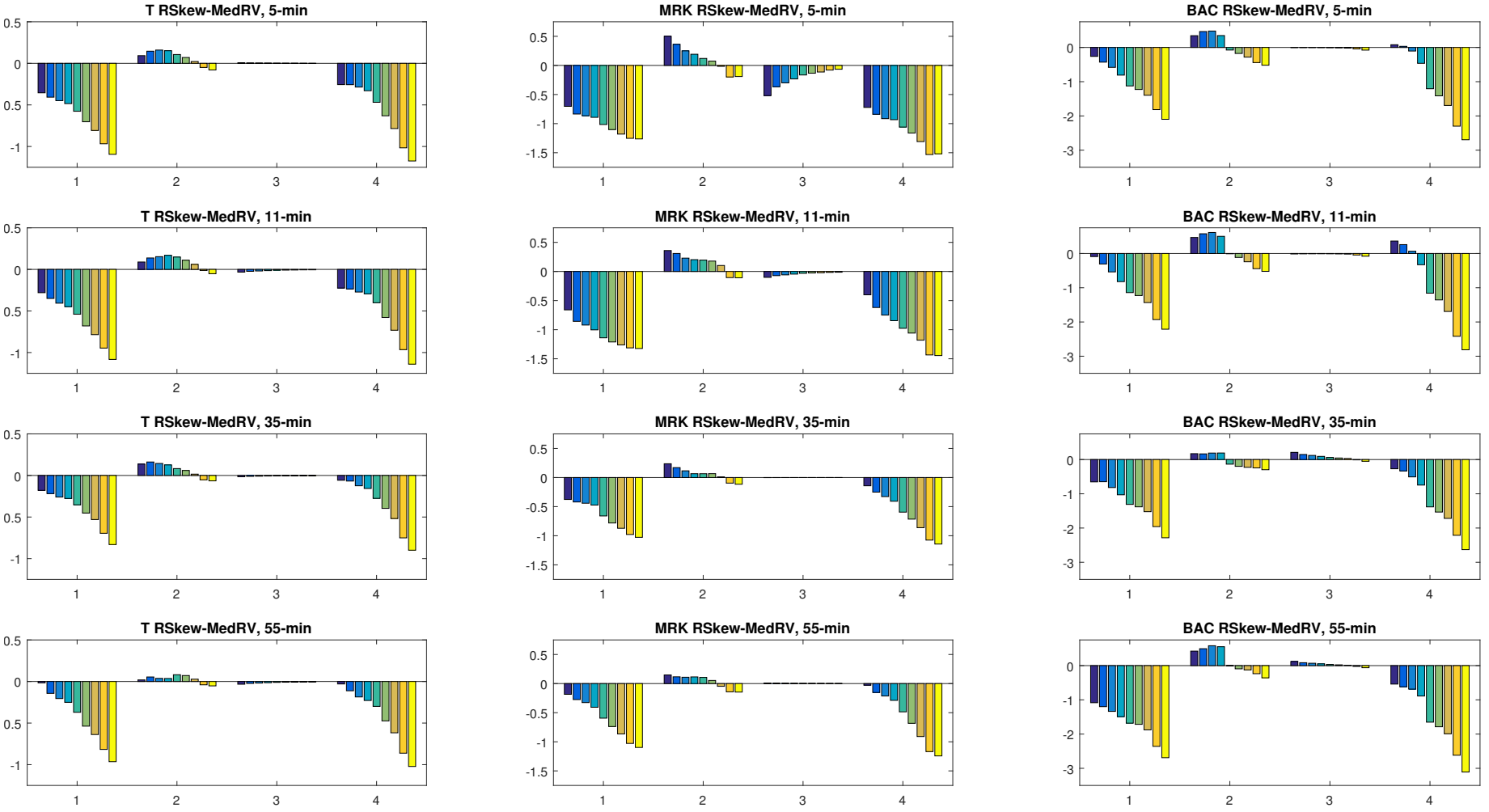


Figure 1.C.15: Category A-2 stocks: RSkew–MedRV measure estimates across frequencies for T, MRK, and BAC



CAT. B STOCKS: B-1: MMM, CSCO, HON, PFE, UNH, INTC. **B-2:** MCD, UTX, VZ, BA, JNJ.

Figure 1.C.16: Category B-1 stocks: RSkew–RV measure estimates across frequencies for MMM, CSCO, and HON

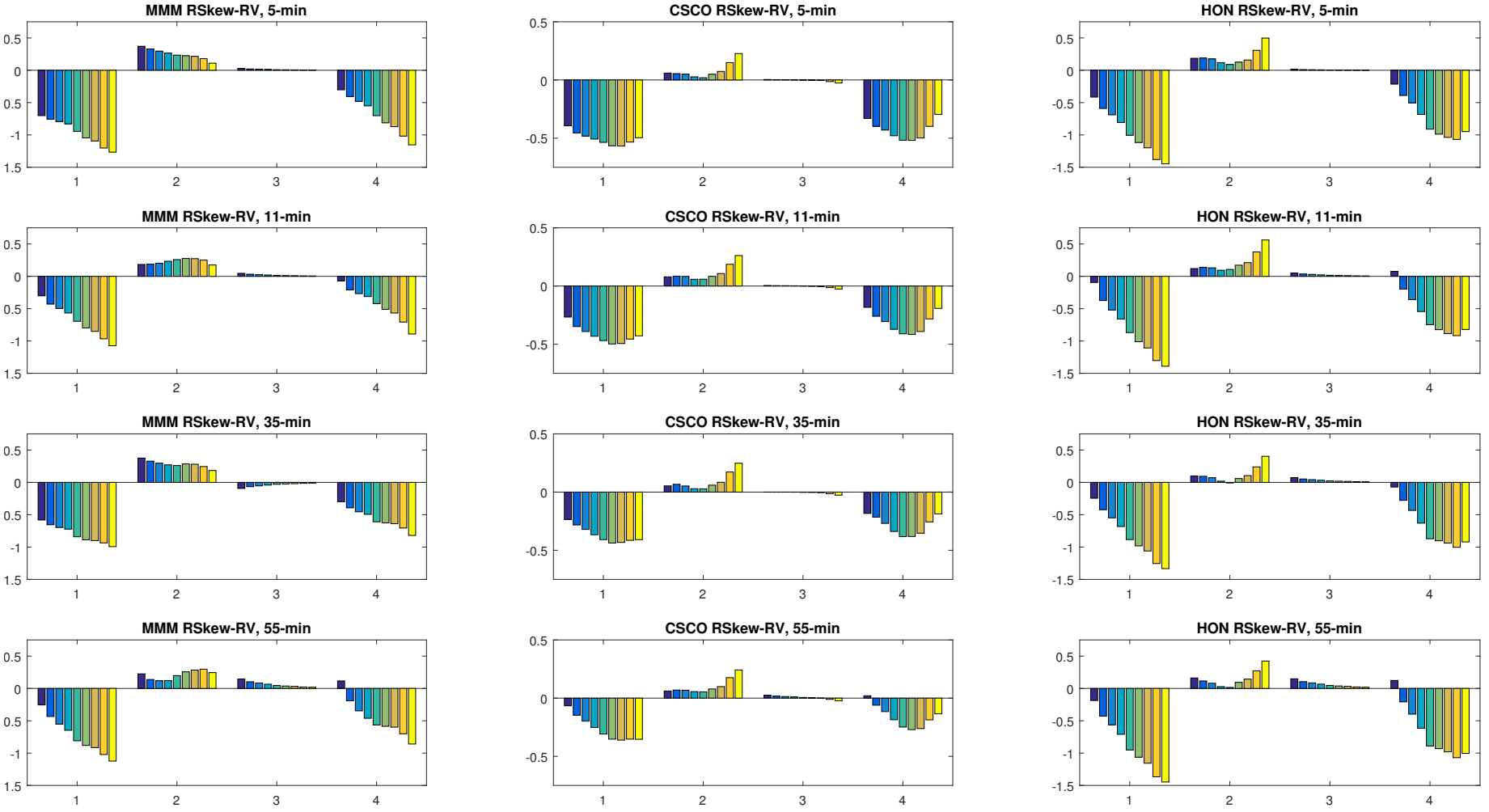


Figure 1.C.17: Category B-1 stocks: RSkew–TPV measure estimates across frequencies for MMM, CSCO, and HON

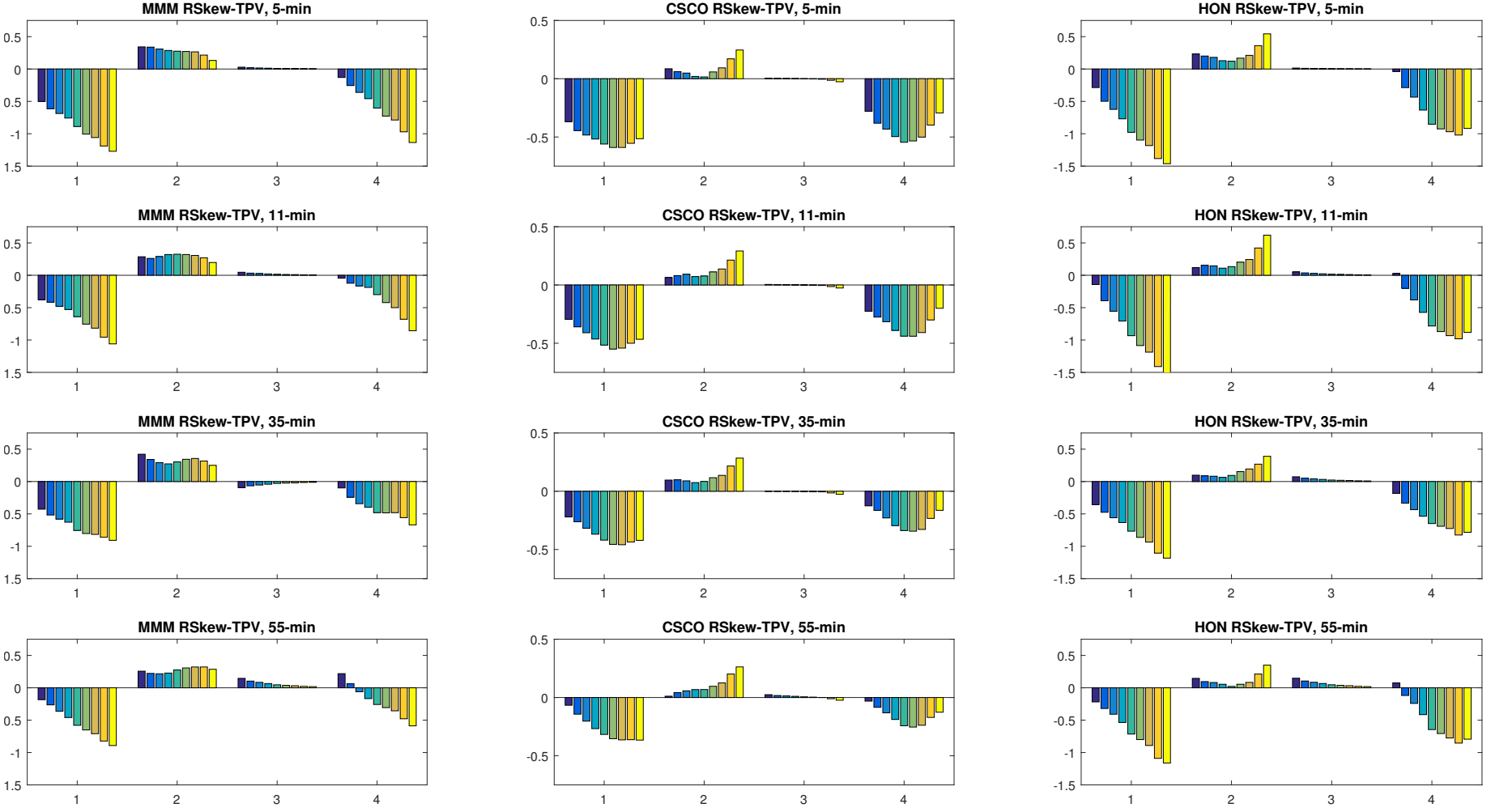


Figure 1.C.18: Category B-1 stocks: RSkew–MedRV measure estimates across frequencies for MMM, CSCO, and HON

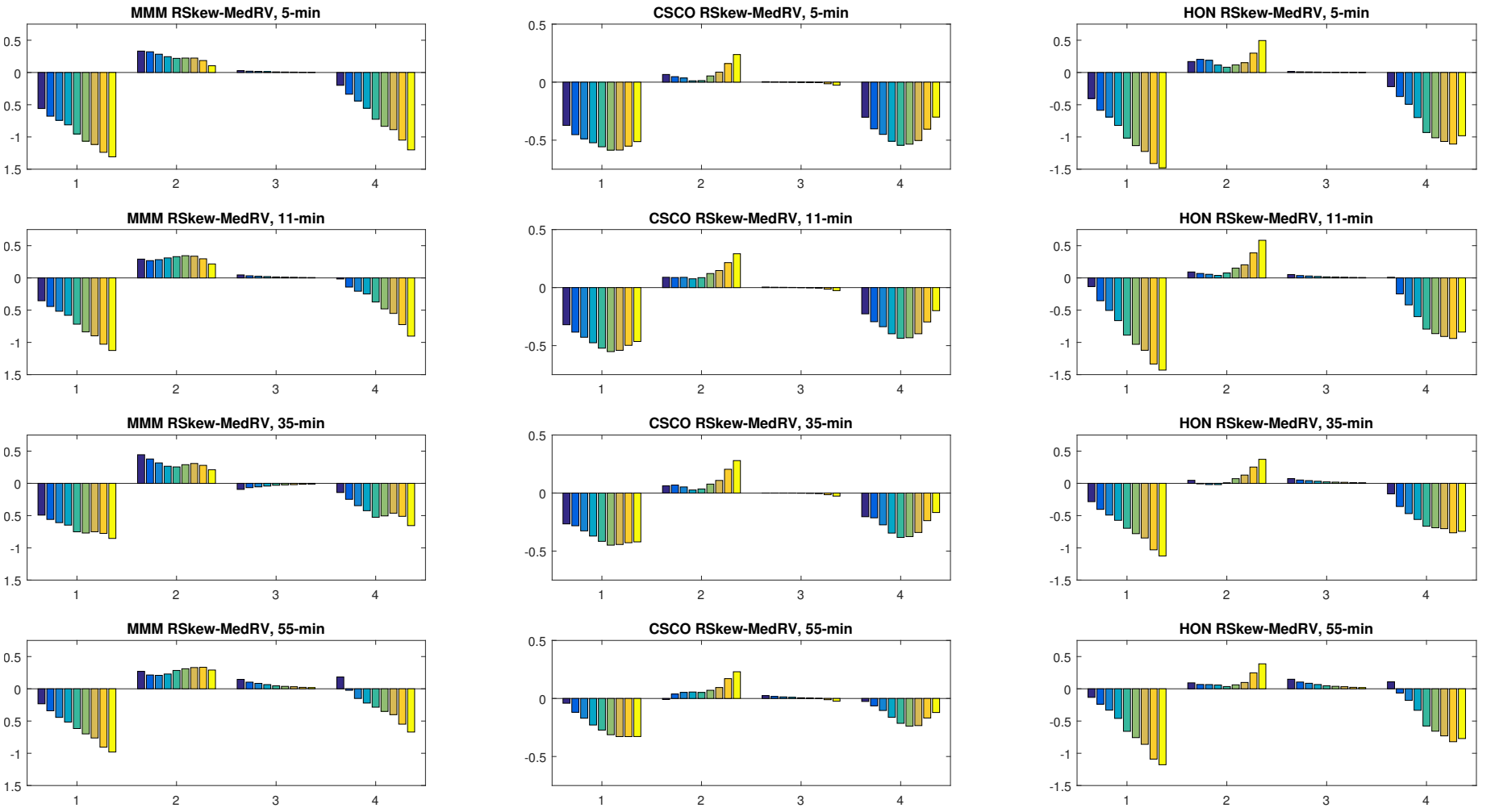


Figure 1.C.19: Category B-1 stocks: RSkew–RV measure estimates across frequencies for PFE, UNH, and INTC

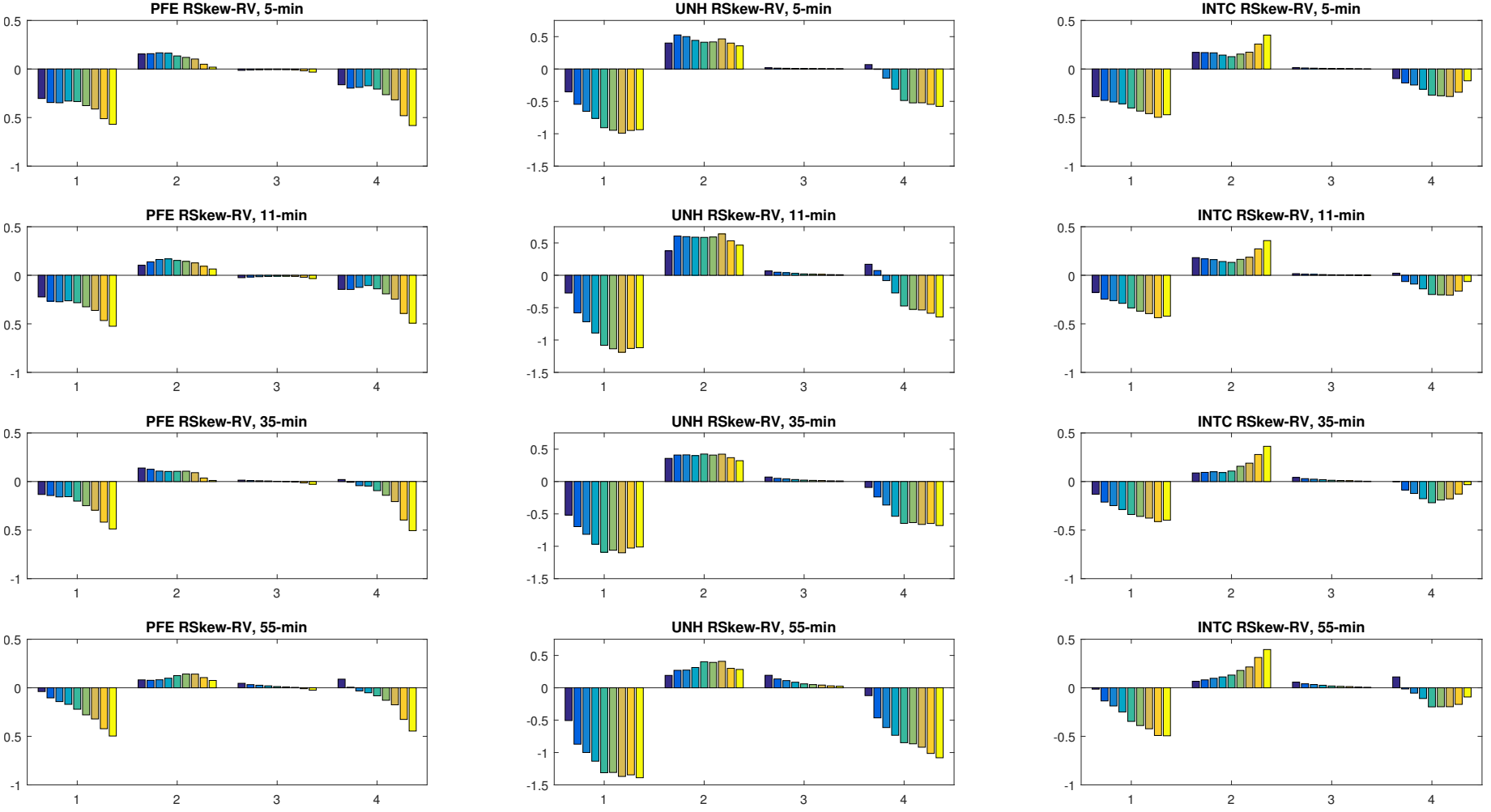


Figure 1.C.20: Category B-1 stocks: RSkew–TPV measure estimates across frequencies for PFE, UNH, and INTC

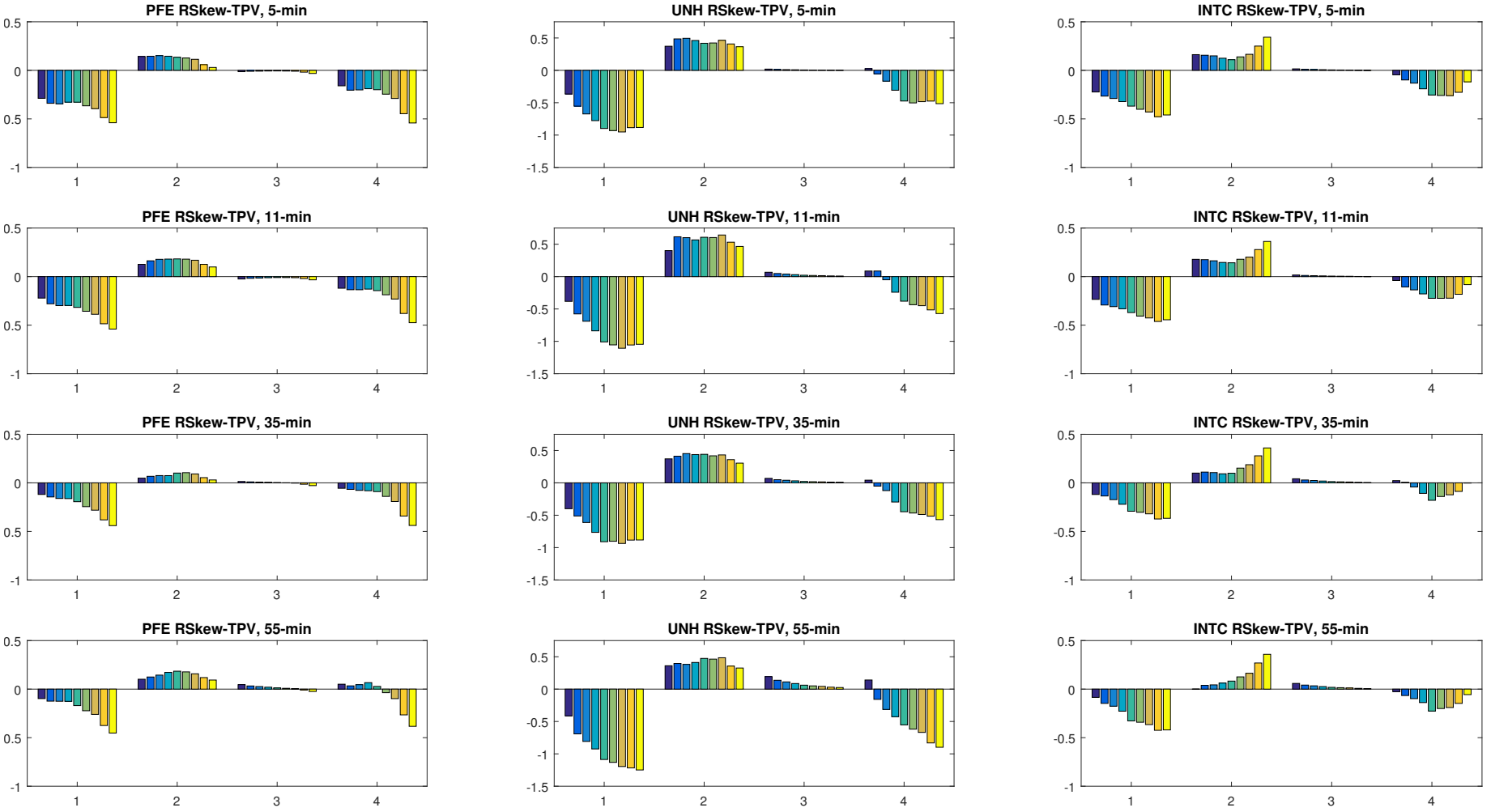


Figure 1.C.21: Category B-1 stocks: RSkew–MedRV measure estimates across frequencies for PFE, UNH, and INTC

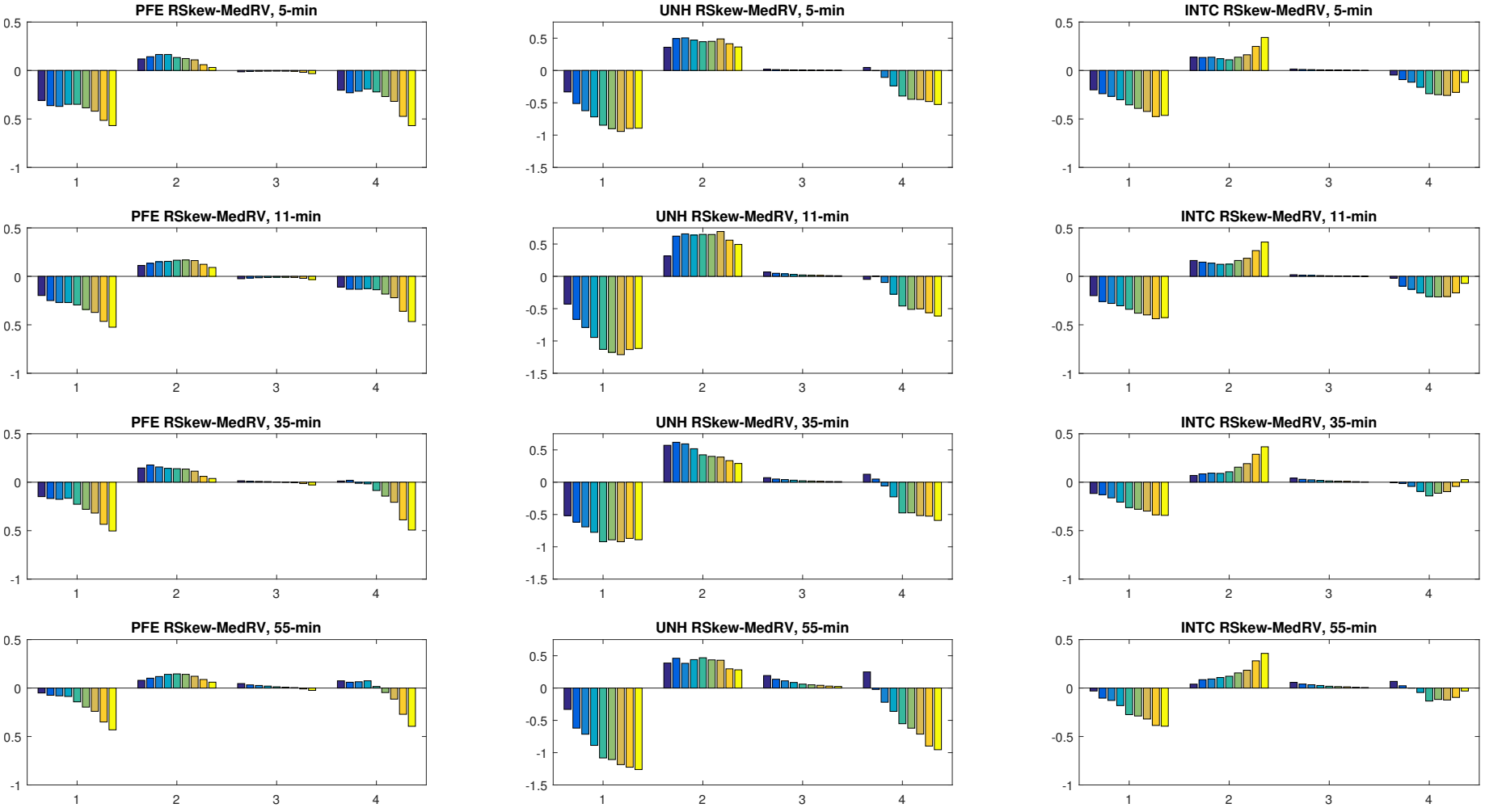


Figure 1.C.22: Category B-1 stocks: RSkew–RV, –MPV measure estimates across frequencies for PG

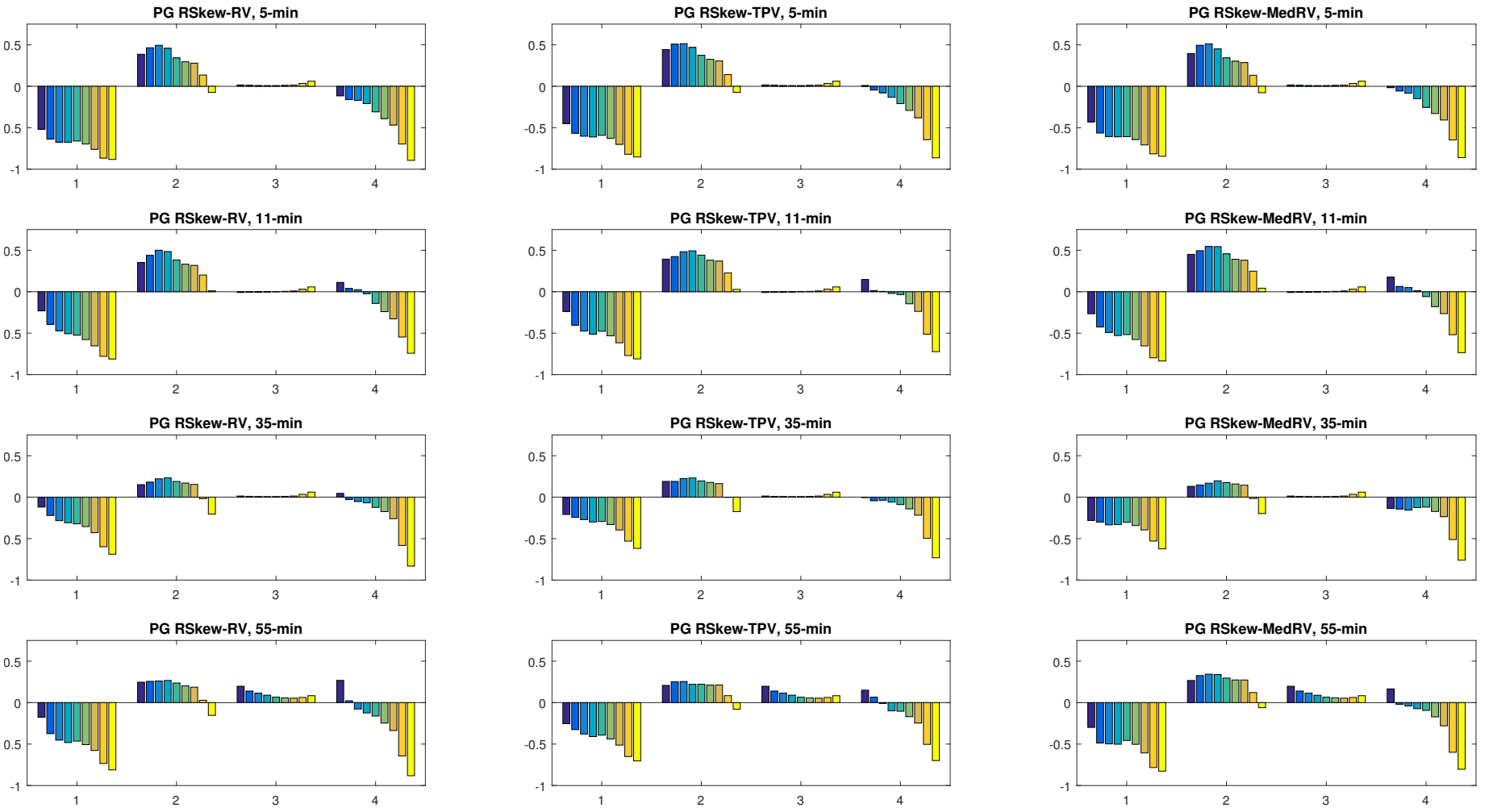


Figure 1.C.23: Category B-2 stocks: RSkew–RV measure estimates across frequencies for MCD, UTX, and VZ

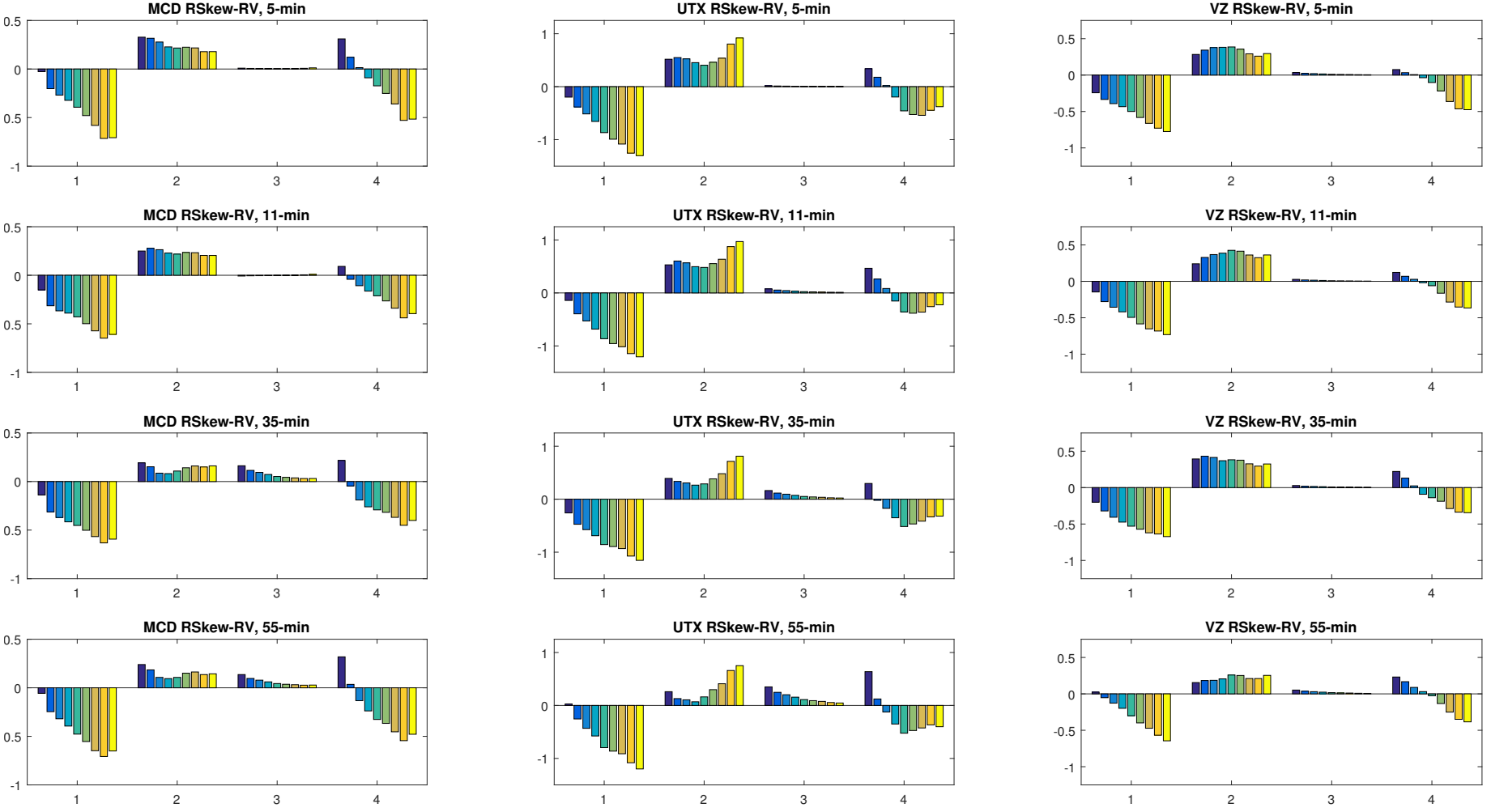


Figure 1.C.24: Category B-2 stocks: RSkew–TPV measure estimates across frequencies for MCD, UTX, and VZ

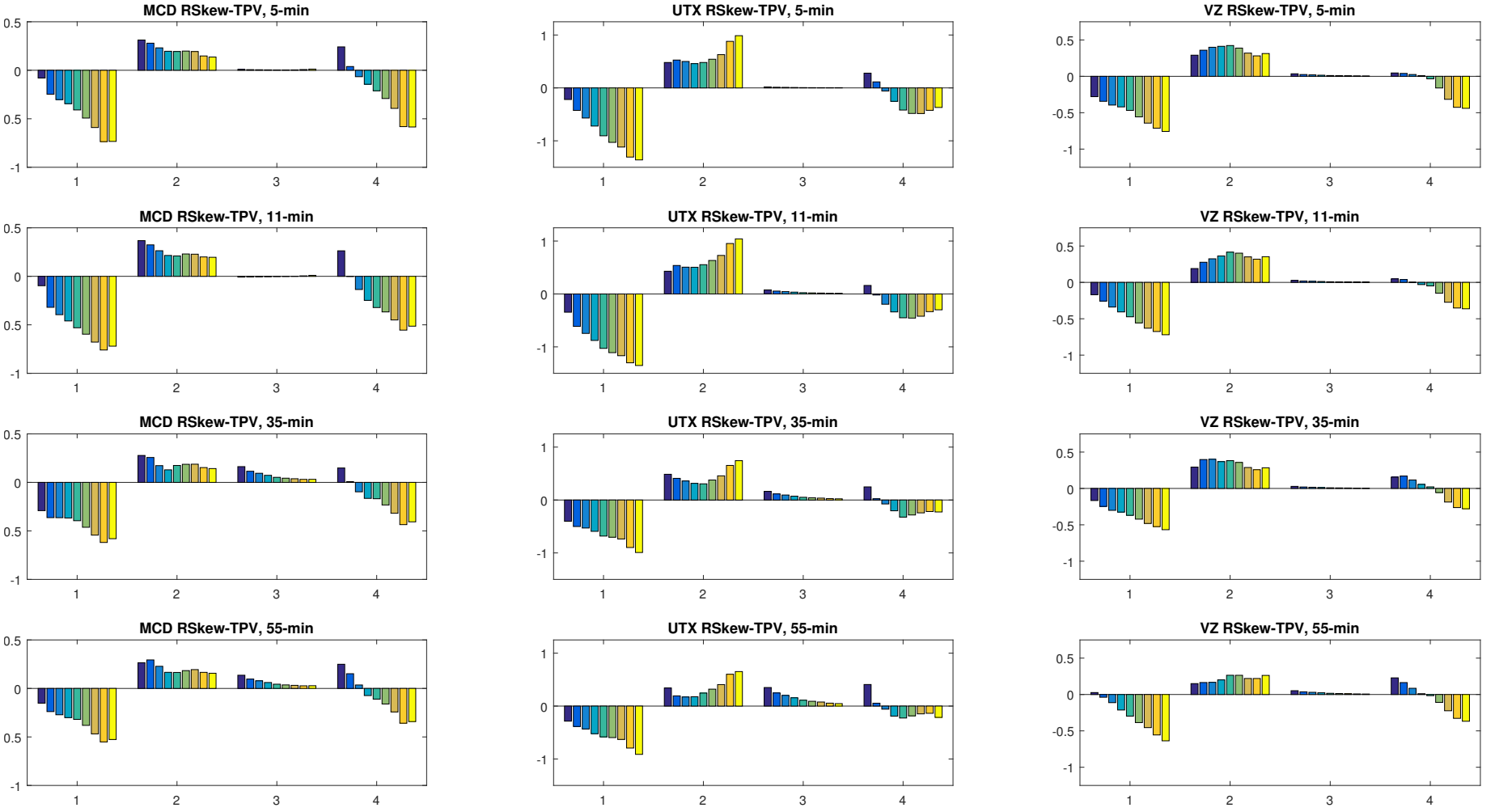


Figure 1.C.25: Category B-2 stocks: RSkew–MedRV measure estimates across frequencies for MCD, UTX, and VZ

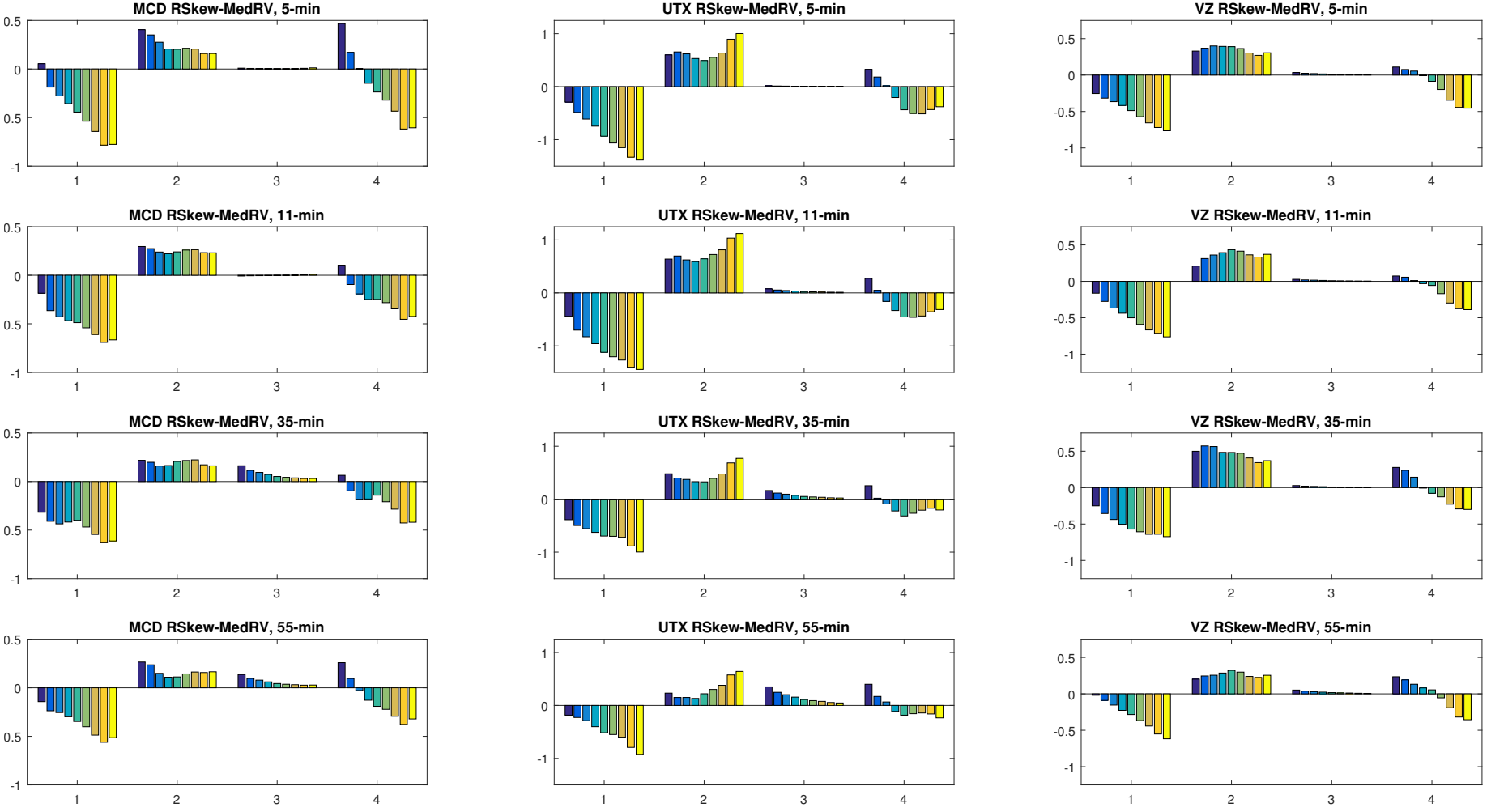


Figure 1.C.26: Category B-2 stocks: RSkew–RV measure estimates across frequencies for BA and JNJ

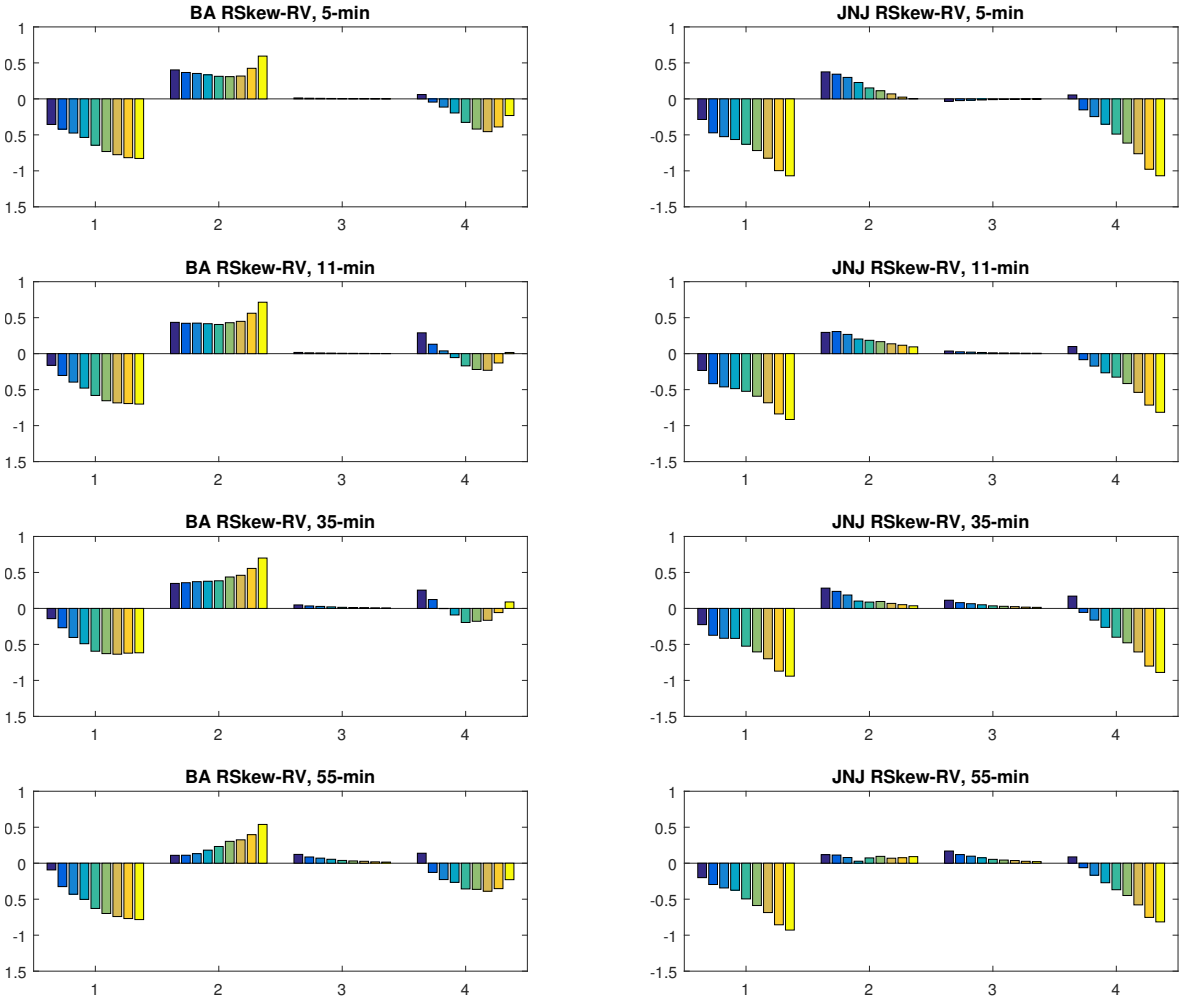


Figure 1.C.27: Category B-2 stocks: RSkew–TPV measure estimates across frequencies for BA and JNJ

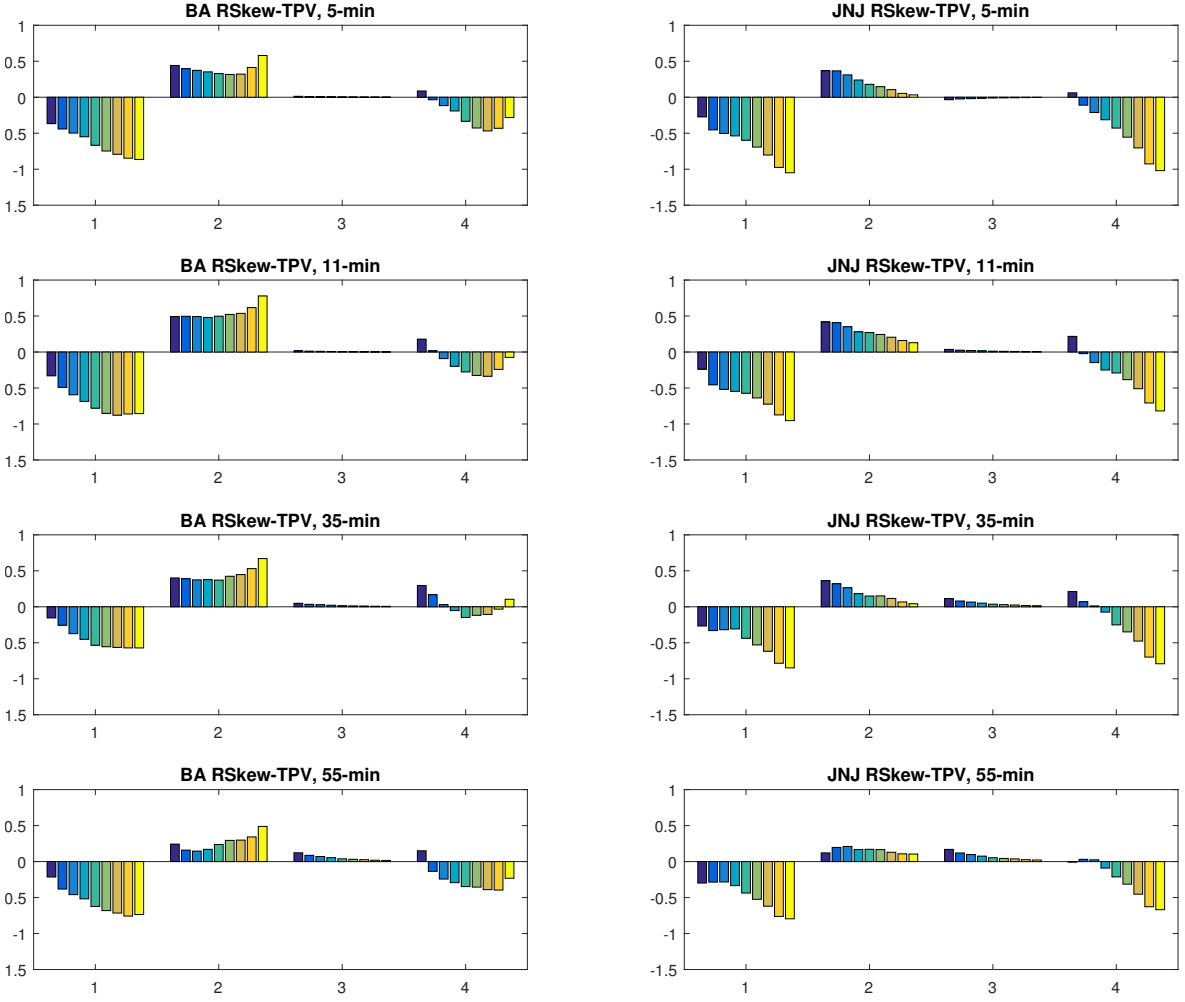
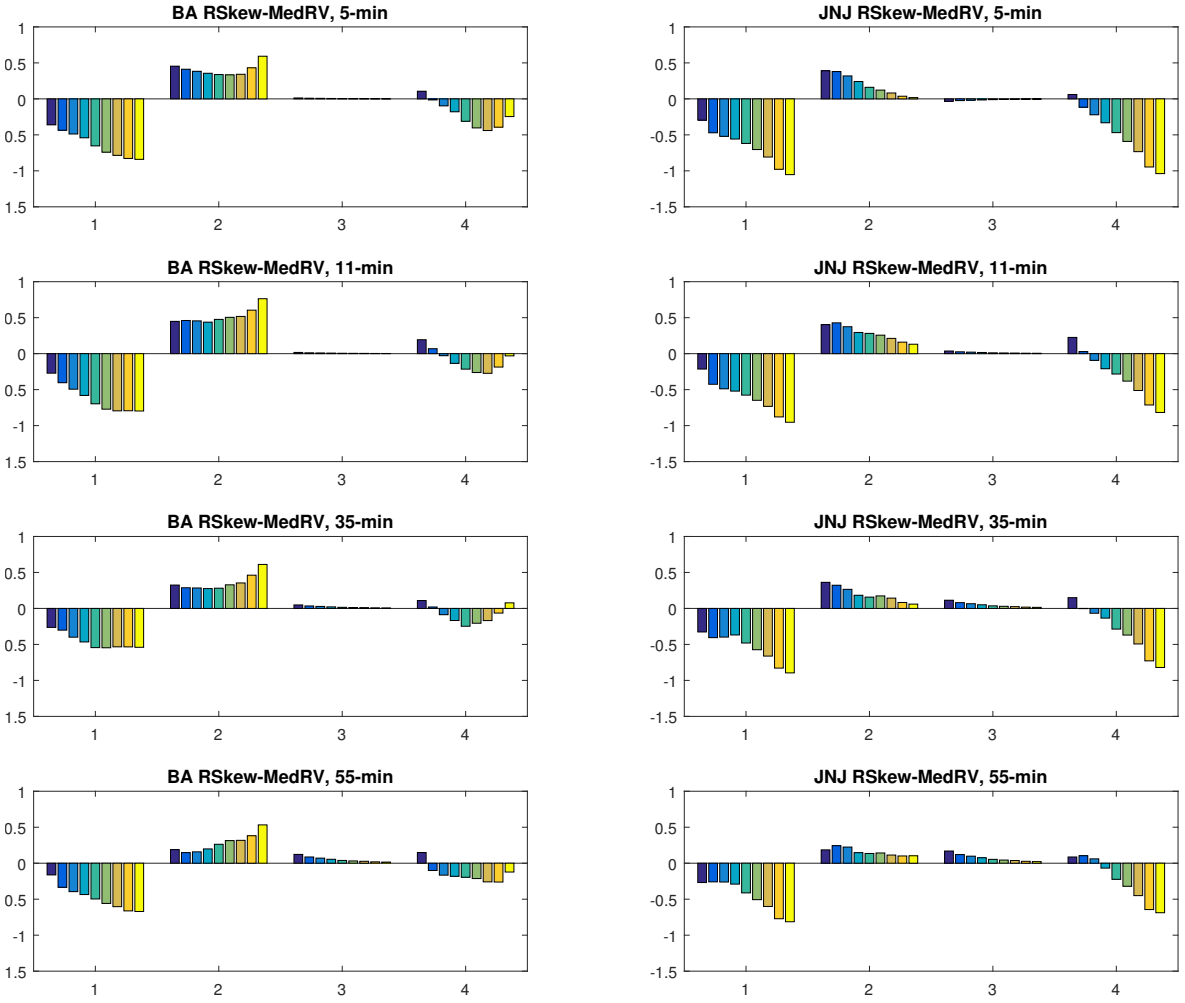
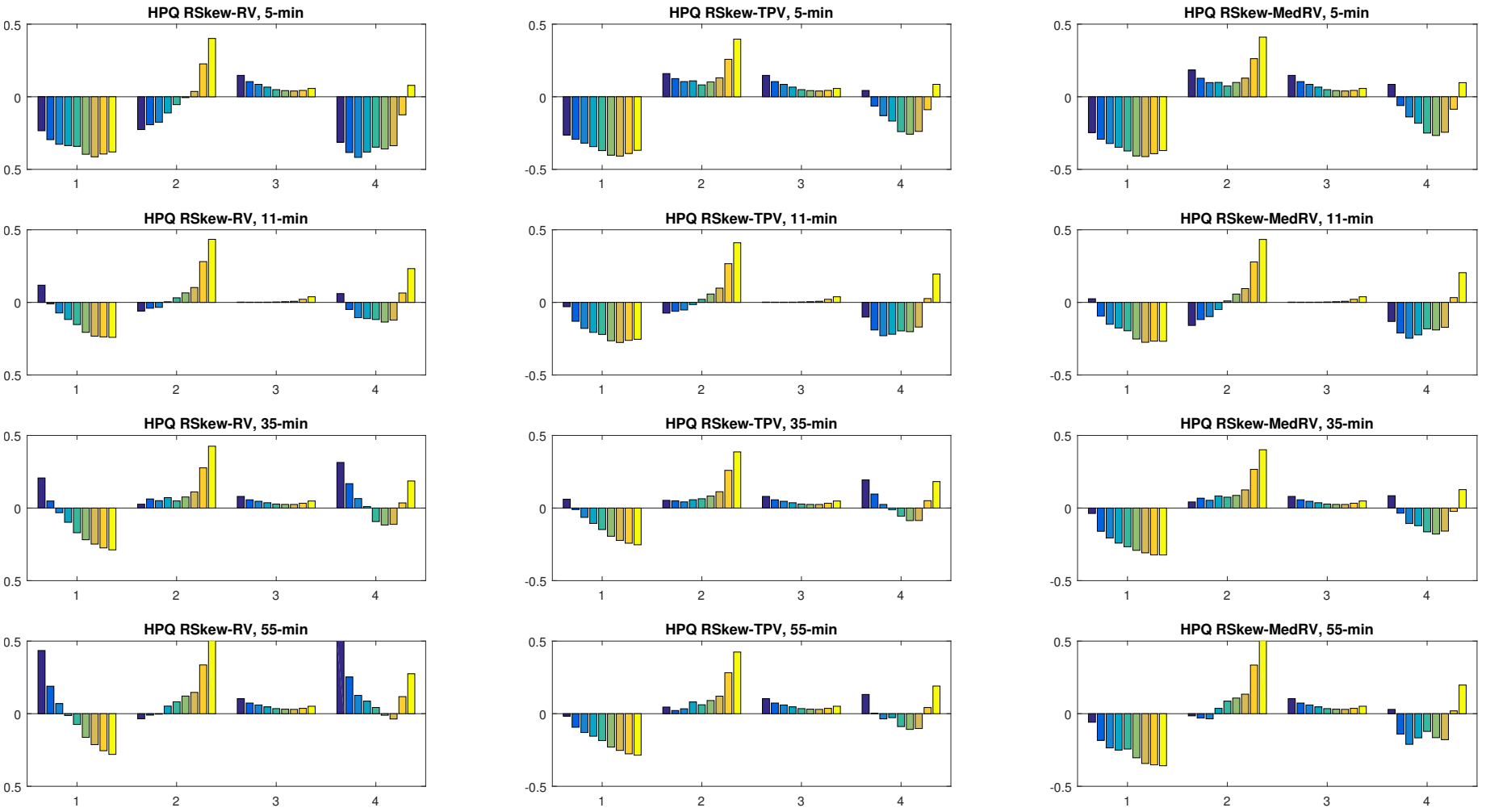


Figure 1.C.28: Category B-2 stocks: RSkew–MedRV measure estimates across frequencies for BA and JNJ



CAT. C STOCKS: HPQ.

Figure 1.C.29: Category C stocks: RSkew–RV, –MPV measure estimates across frequencies for HPQ



CAT. D STOCKS: CVX, HD, WMT, DIS, AXP, JPM.

Figure 1.C.30: Category D stocks: RSkew–RV measure estimates across frequencies for CVX, HD, and WMT

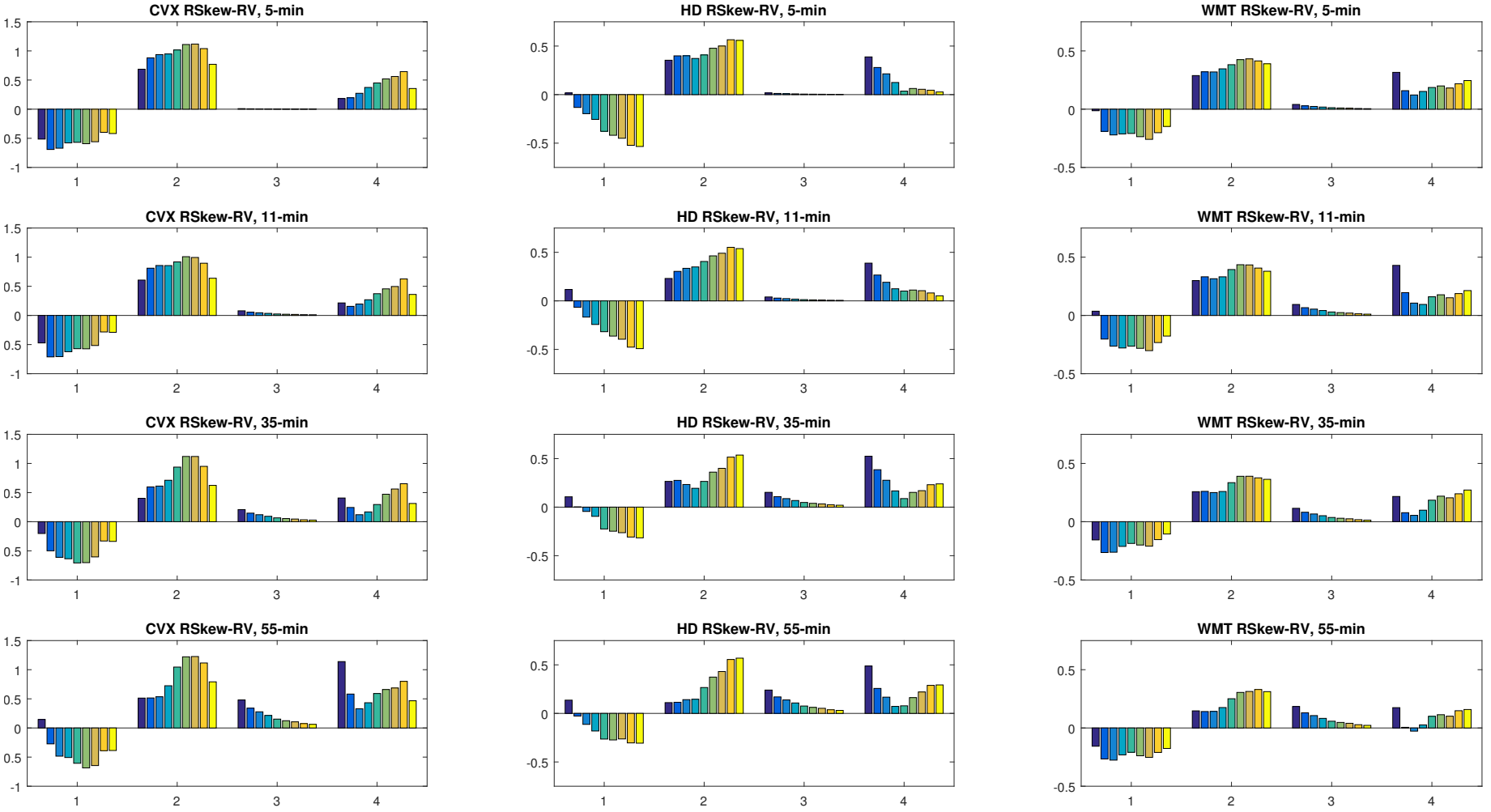


Figure 1.C.31: Category D stocks: RSkew–TPV measure estimates across frequencies for CVX, HD, and WMT

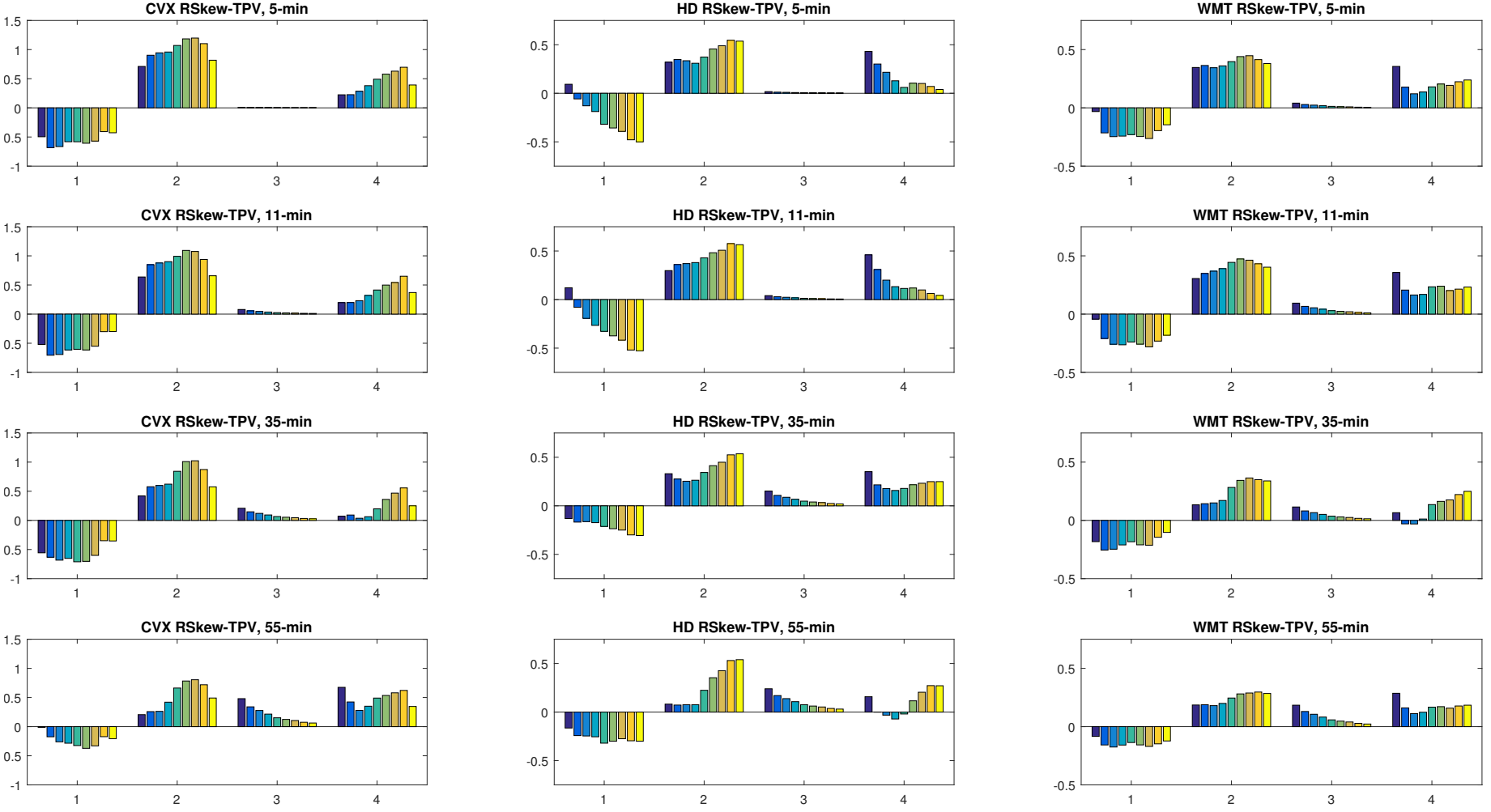


Figure 1.C.32: Category D stocks: RSkew–MedRV measure estimates across frequencies for CVX, HD, and WMT

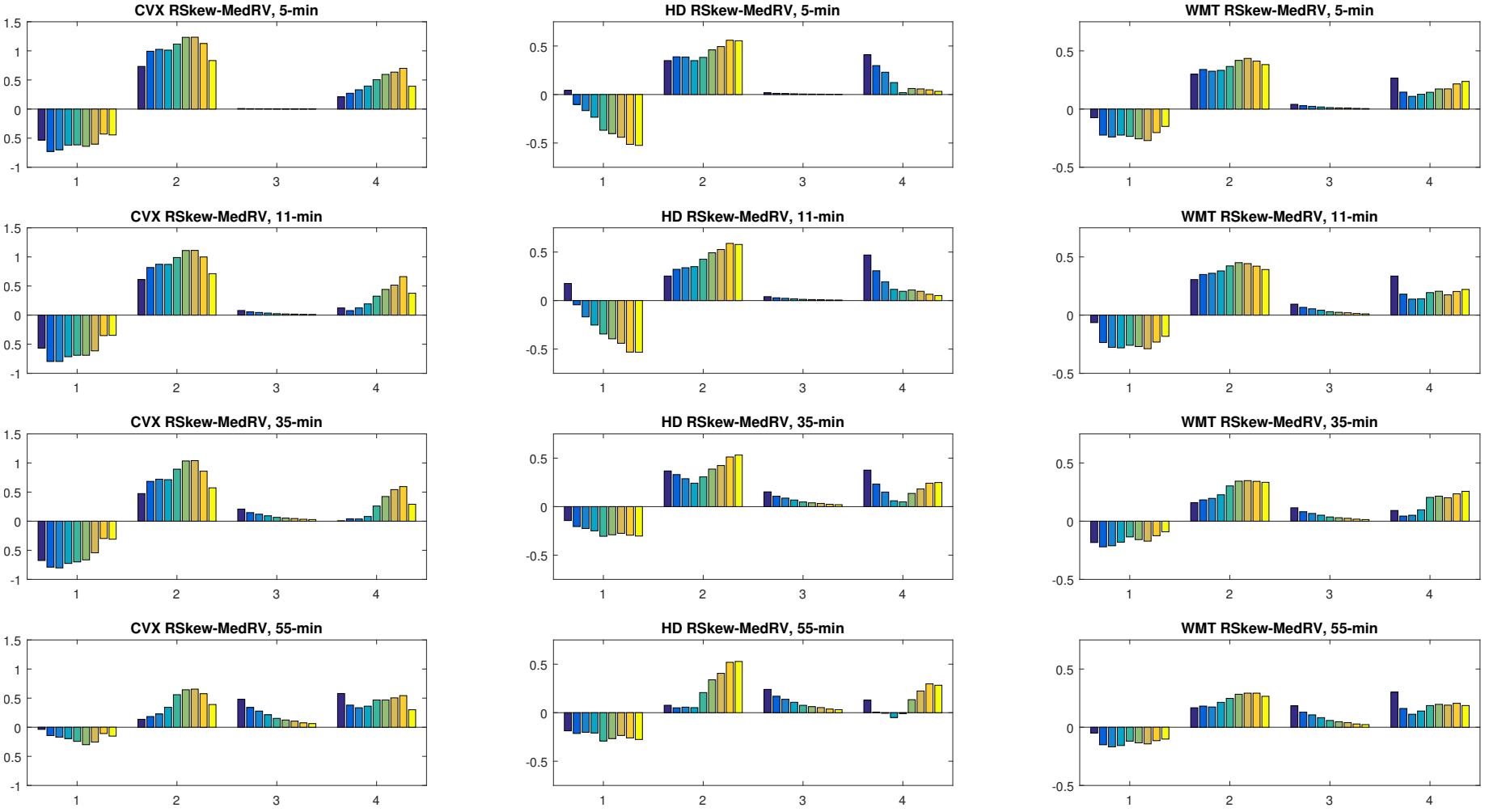


Figure 1.C.33: Category D stocks: RSkew–RV measure estimates across frequencies for DIS, AXP, and JPM

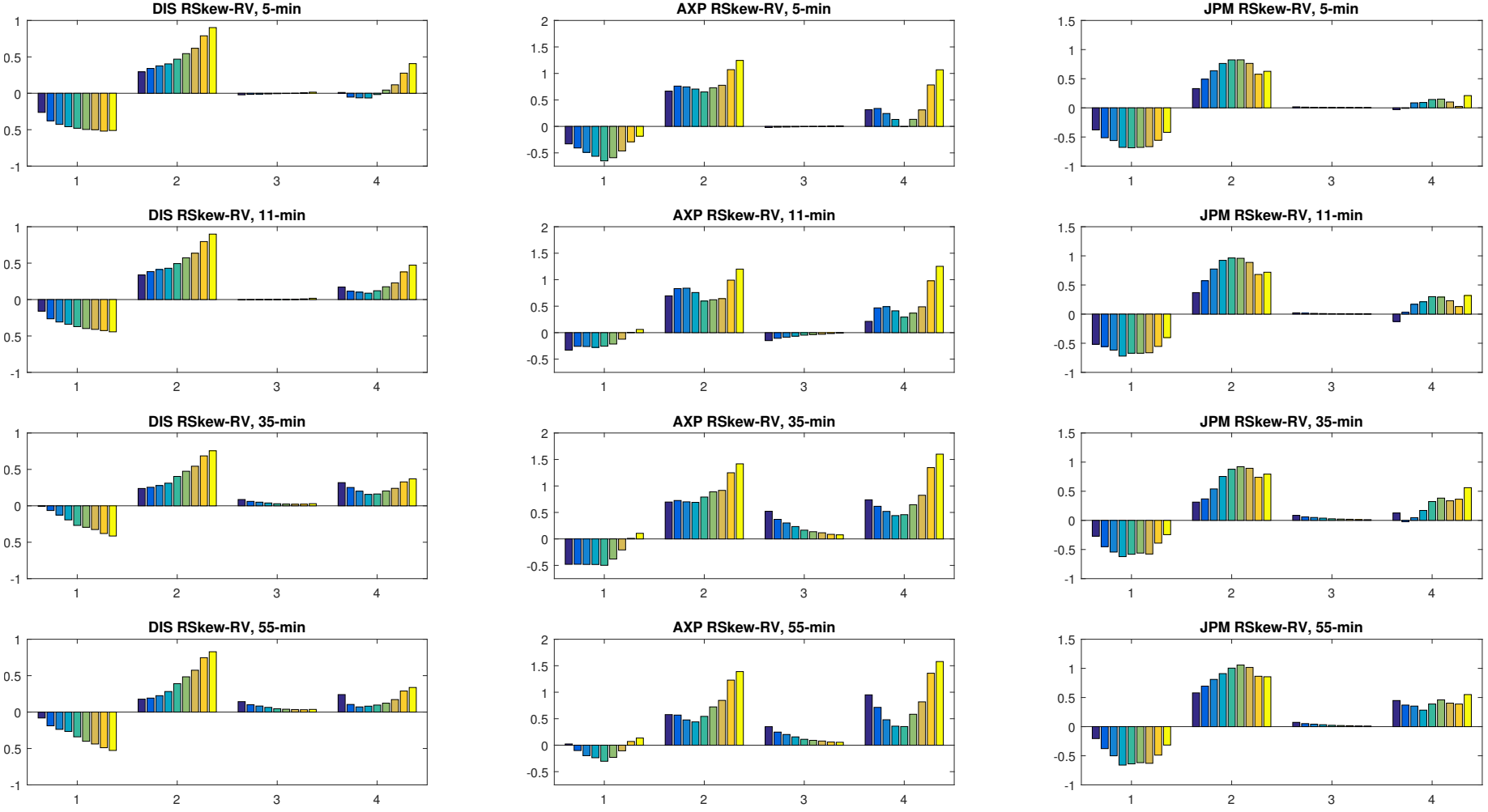


Figure 1.C.34: Category D stocks: RSkew–TPV measure estimates across frequencies for DIS, AXP, and JPM

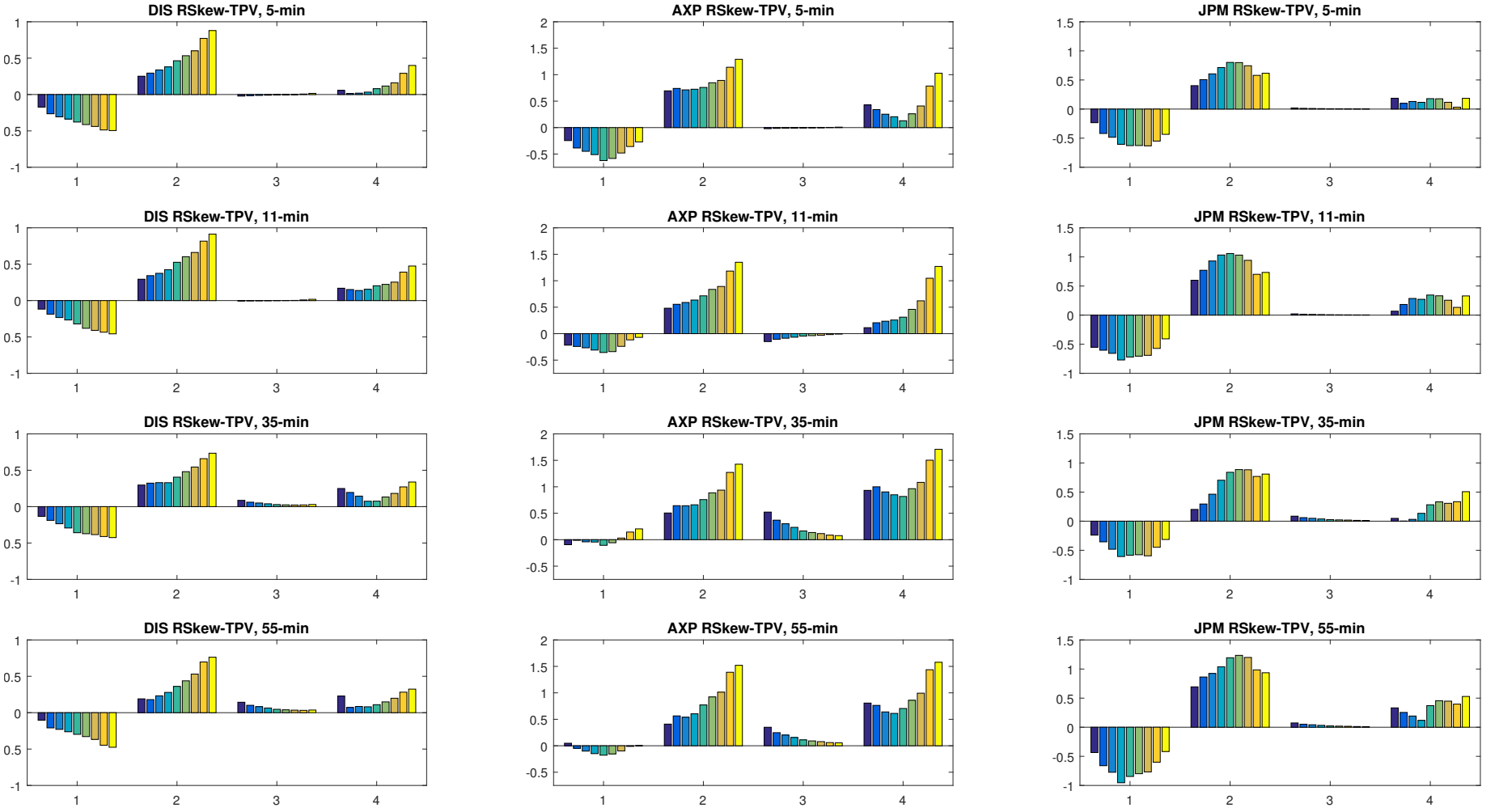
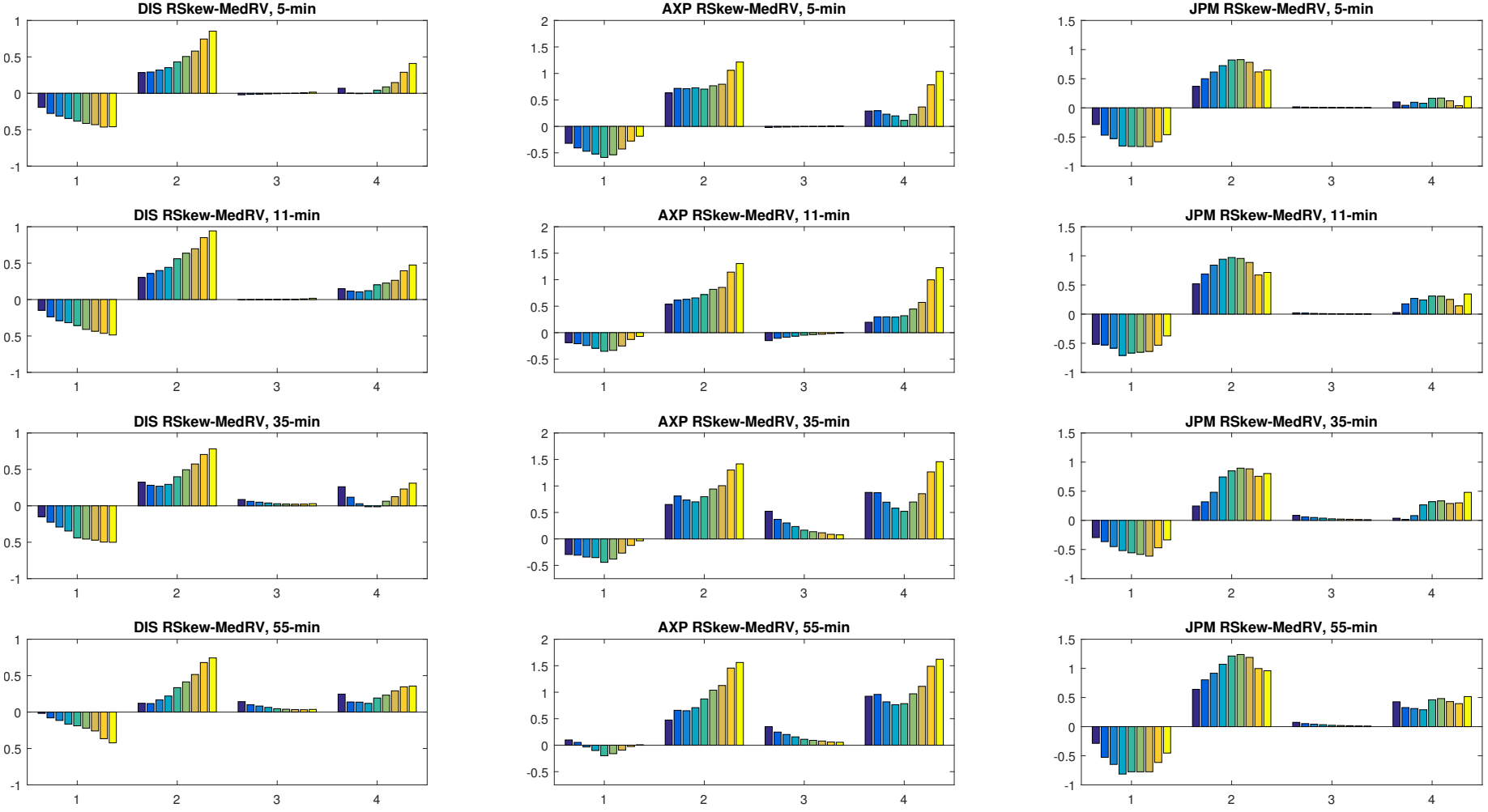


Figure 1.C.35: Category D stocks: RSkew–MedRV measure estimates across frequencies for DIS, AXP, and JPM



CAT. E STOCKS: AA, TRV, C.

Figure 1.C.36: Category E stocks: RSkew–RV measure estimates across frequencies for AA, TRV, and C

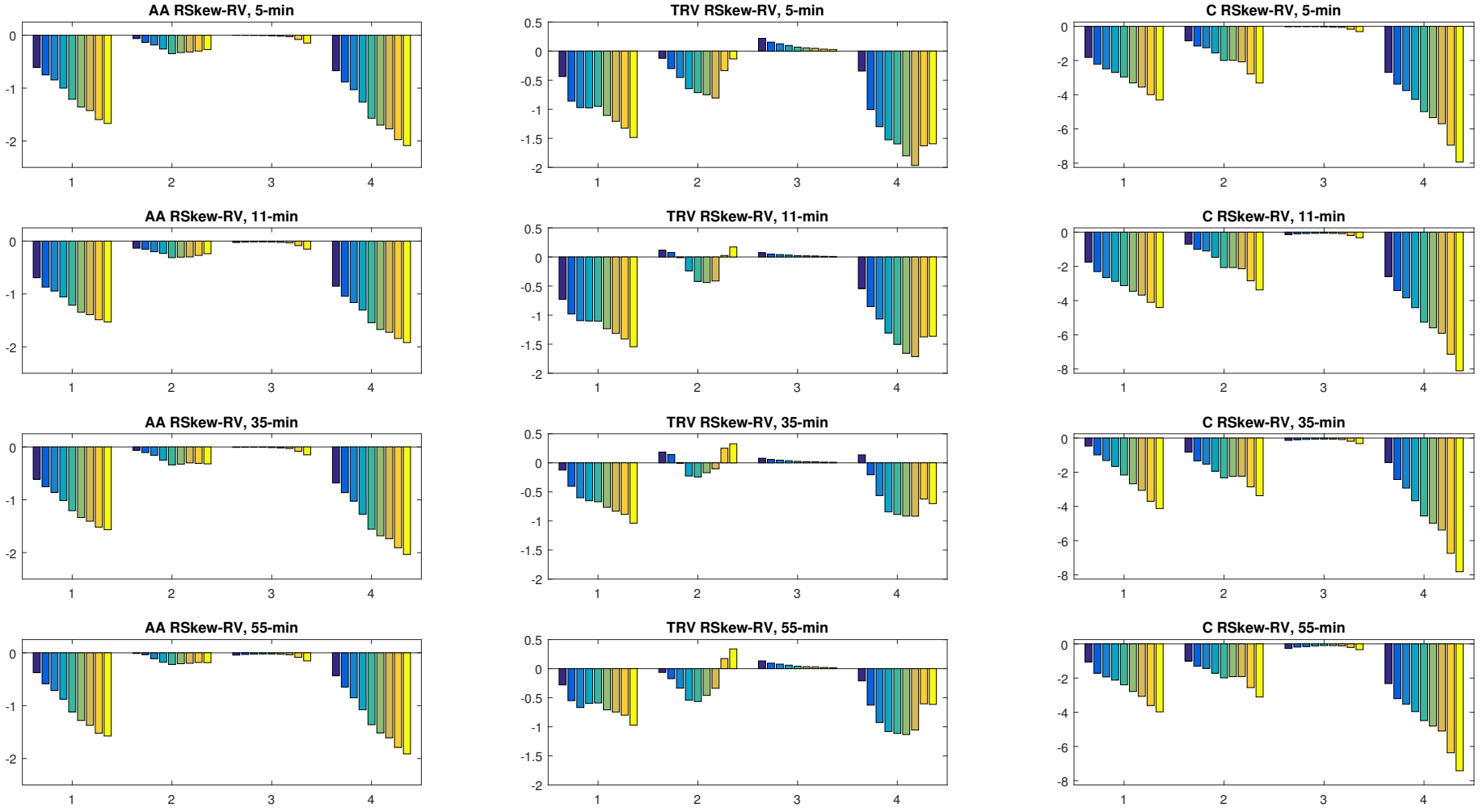


Figure 1.C.37: Category E stocks: RSkew–TPV measure estimates across frequencies for AA, TRV, and C

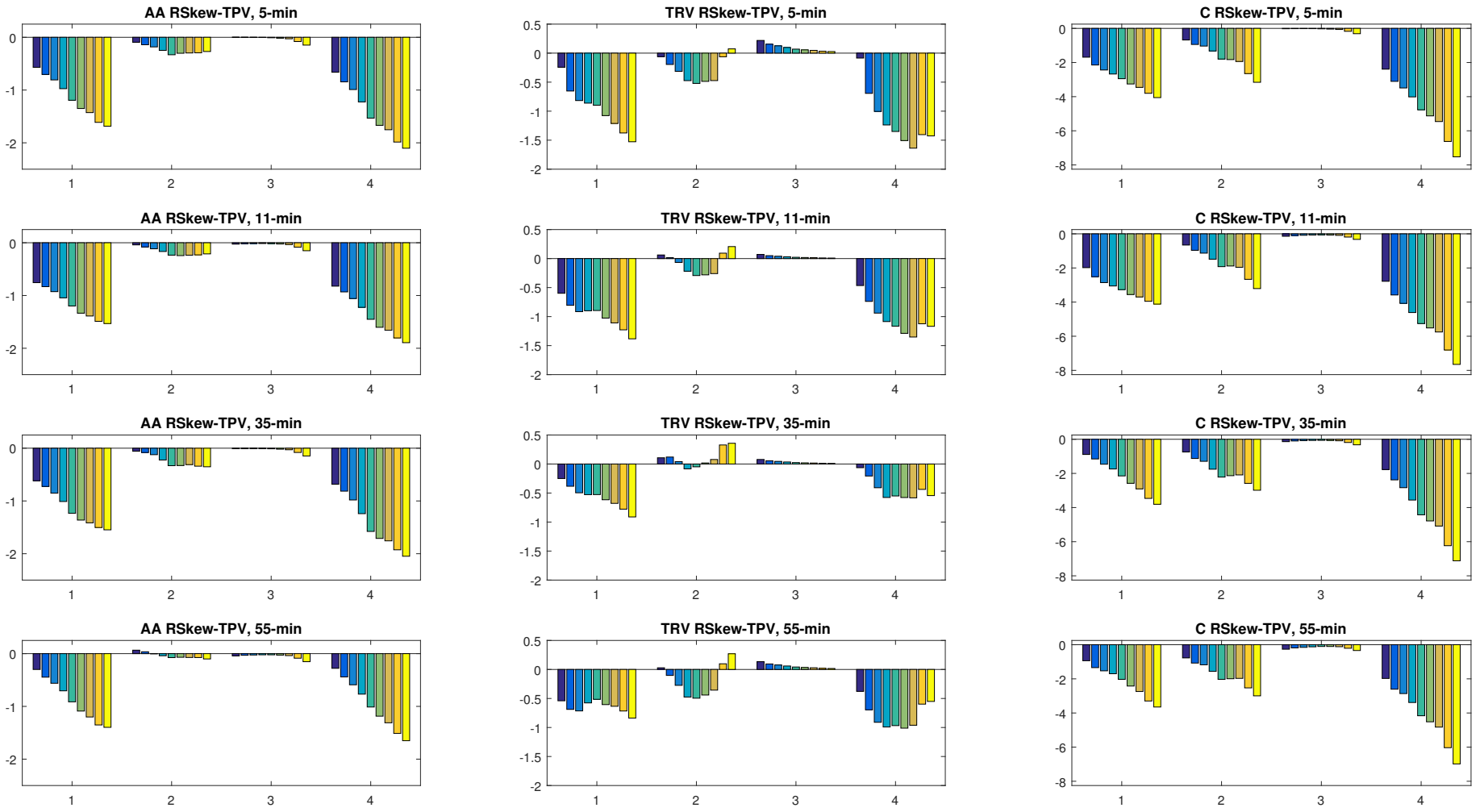
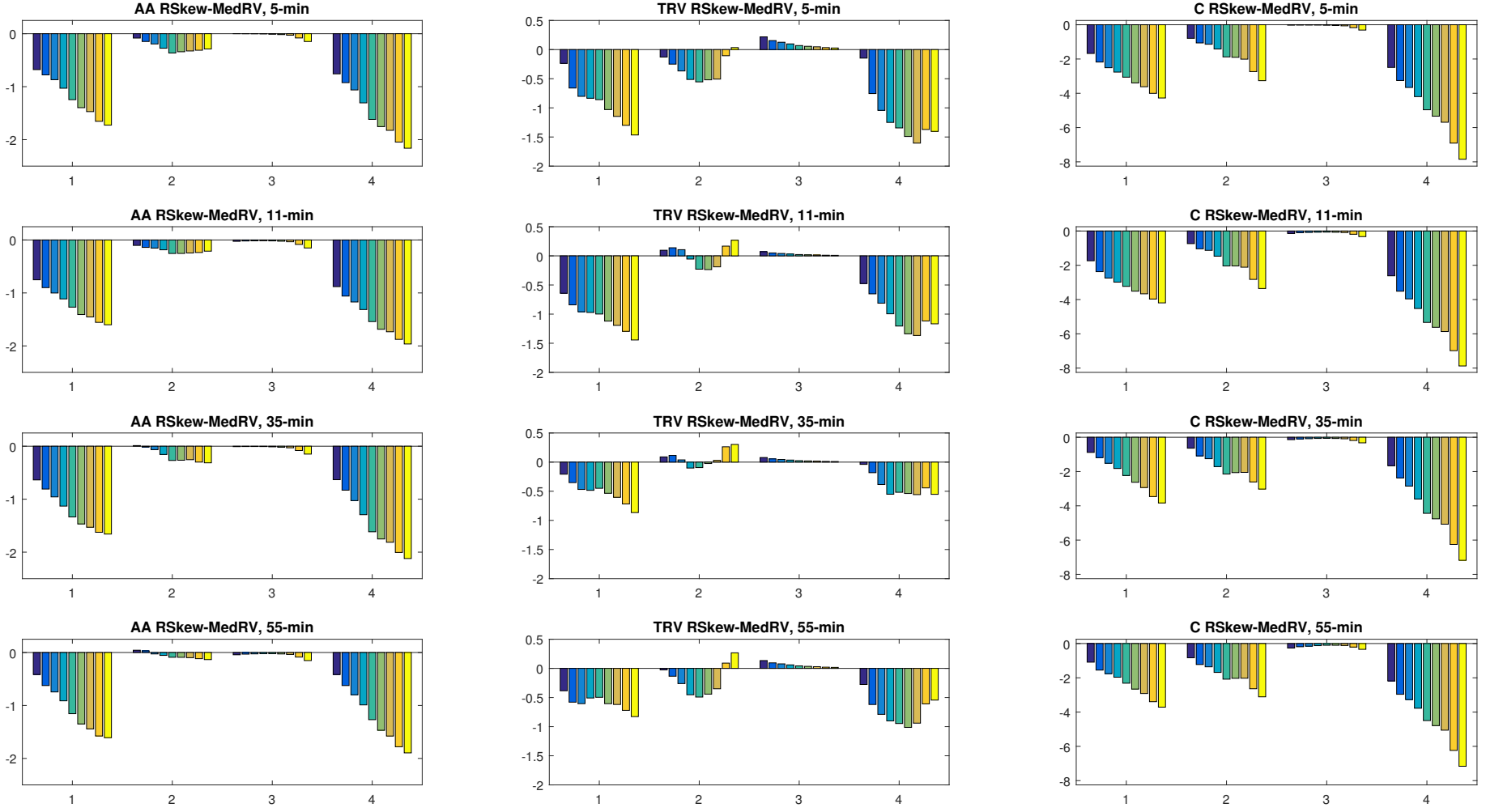


Figure 1.C.38: Category E stocks: RSkew–MedRV measure estimates across frequencies for AA, TRV, and C



CAT. OTHER: GM, AIG, XOM

Figure 1.C.39: Category OTHER: RSkew–RV, –MPV measure estimates across frequencies for GM

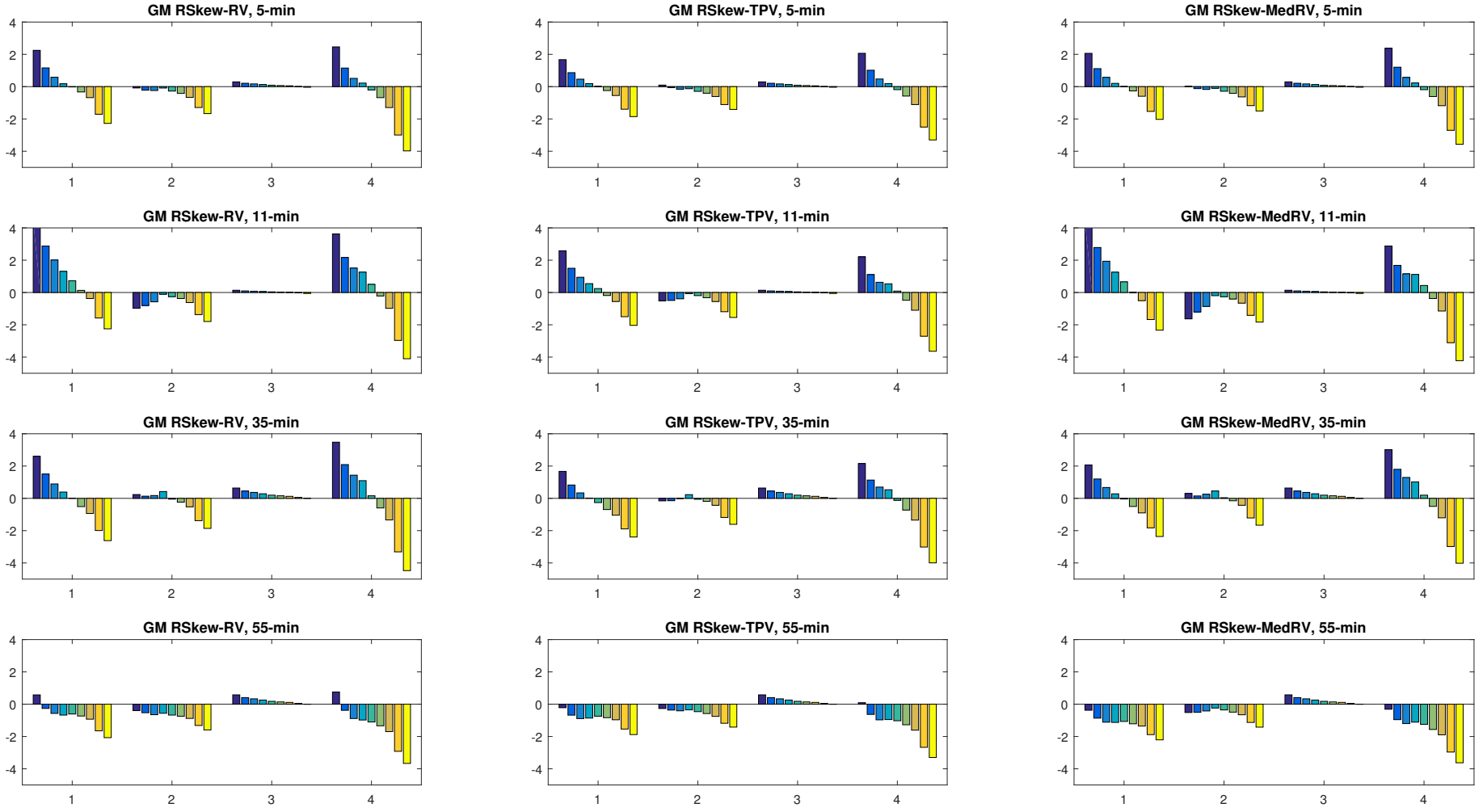


Figure 1.C.40: Category OTHER: RSkew–RV, –MPV measure estimates across frequencies for AIG

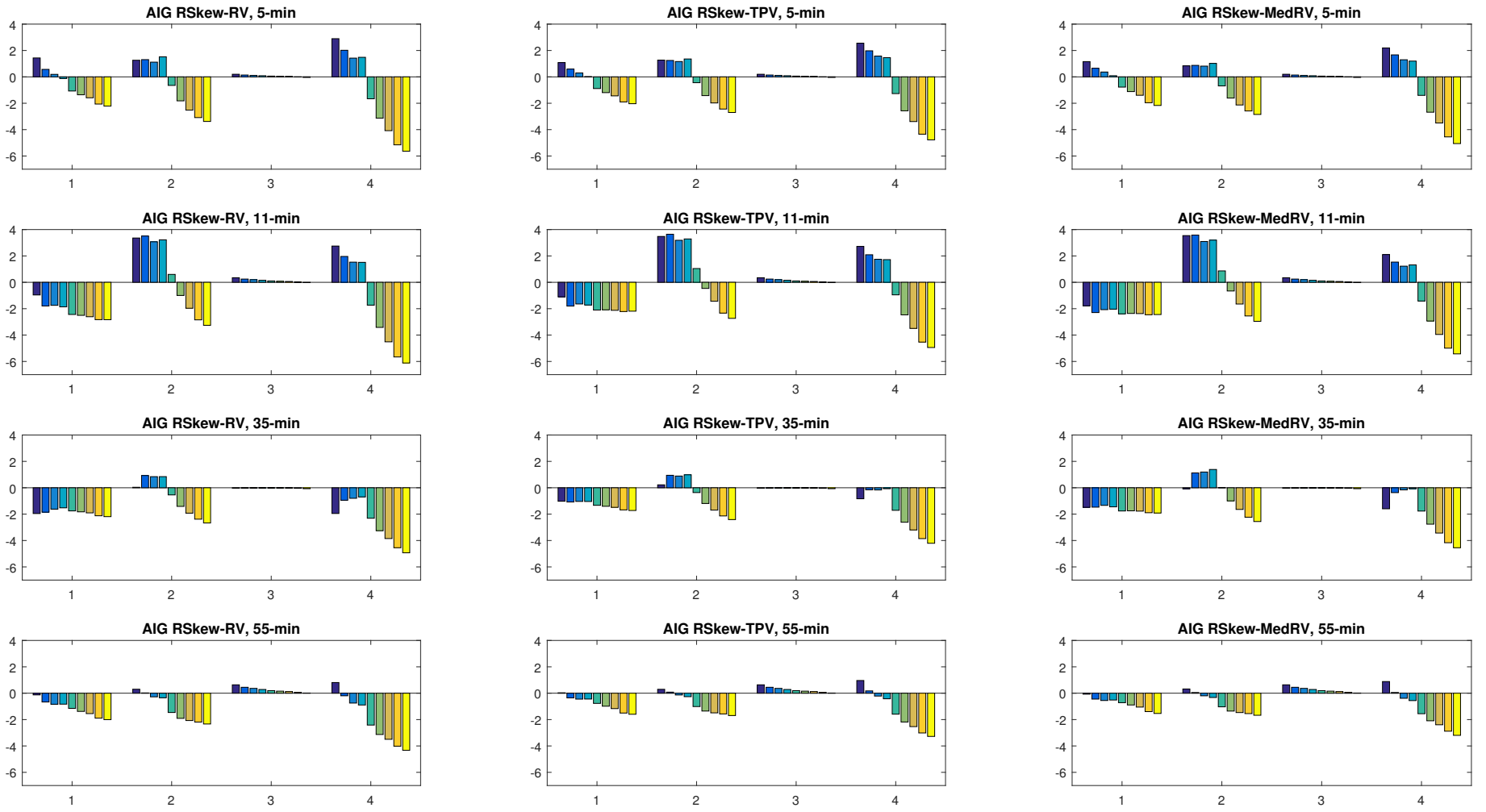
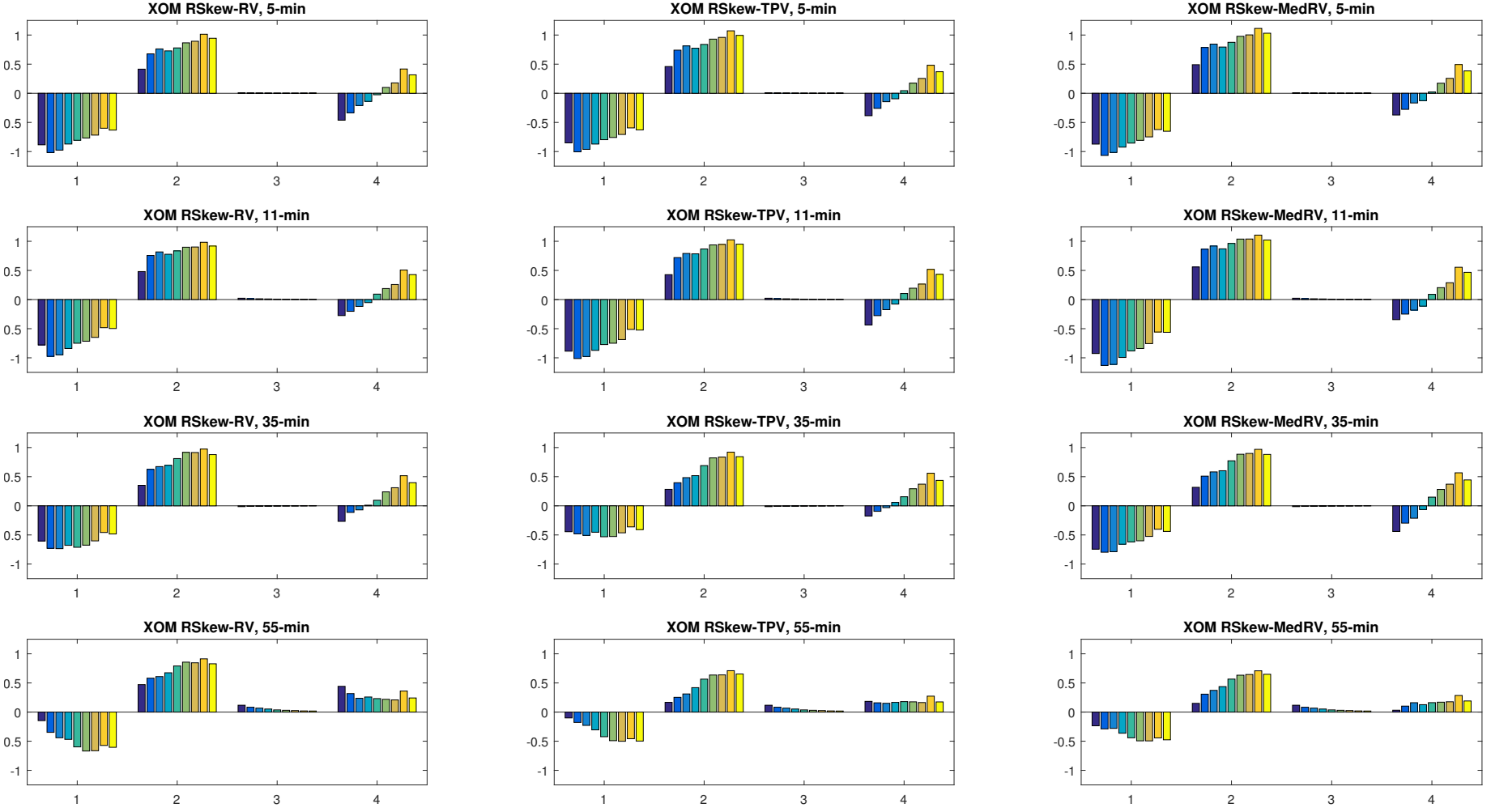


Figure 1.C.41: Category OTHER: RSkew–RV, –MPV measure estimates across frequencies for XOM



1.D Robustness Checks

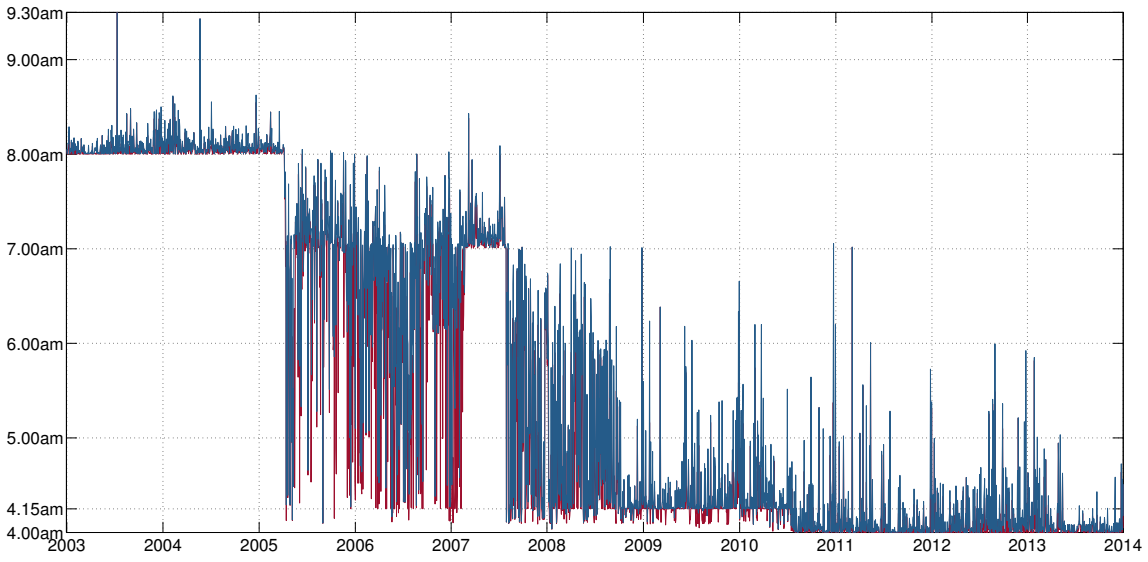
1.D.1 Extended Trading Hours

We also investigate into the return volatility asymmetries and volatility feedback effects related to the following closed market and overnight period. However, as these do not lead to any coherent findings these are excluded from the main anlysis but available from the author upon request.

S&P 500 Depositary Receipt (SPDR) Trust ETF – Trading Hours It is possible to trade the SPY contract outside this time frame. In order to grasp when trading starts effectively the time stamps of the first two trades are exhibited in Figure 1.D.1 for all trading days over the eleven year period from 2003 until end of year 2013 across all exchanges feeding into the TAQ database. Five periods can be distinguished with trading

Figure 1.D.1: Time stamp of 1st and 2nd SPY trade for every day, 2003–2013

RED: Time stamp of the 1st trade during any given day. BLUE: Time stamp of the 2nd trade. Y–AXIS: Time in EST format. X–AXIS: Time line denotes the 2,769 trading days in the TAQ database from 2003 until end of year 2013.

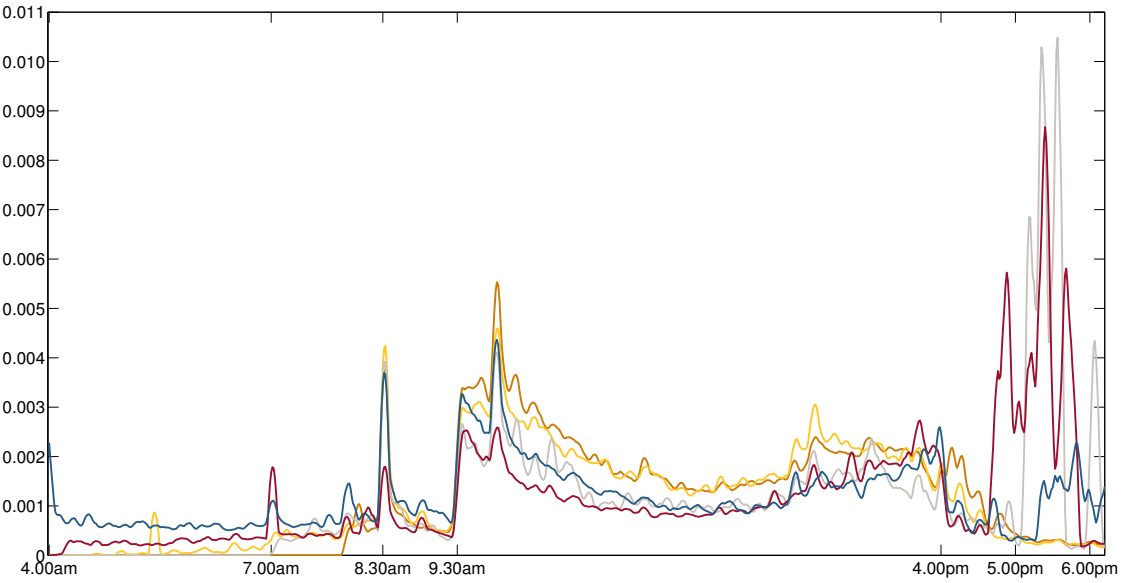


commencing at I.) 8.00am until April 8th, 2005 with few trades taking places before 8.30am, II.) 4.00am until February 23th, 2007 with only few trades actually taking place before 7.00am or even 8.00am, III.) 7.00am until July 31th, 2007 with few trades taking place before 8.00am, IV.) 4.00am until July 30th, 2010 with almost no trades taking place before 4.15am and only few trades before 7.00am, V.) 4.00am from August, 1st 2010

onwards. Based on this information, we do not recommend to make any inference at using trades before at least 8.30am in period I.), 8.00am in period II.) and III.), 7.00am in period IV.), and 6.00am in the remainder. When interested in lead–lag effects at high–frequency we need to be even more conservative in the choice of the intraday observation window.

Figure 1.D.2: Crude dirunal volatility periodicity component extended trading hours, Periods I.-V.

The diurnal volatility periodicity components are calculated using a Naradaya–Watson kernel regression on squared five second return increments with a bandwidth of 90 seconds. The smooth estimated are standardized to yield 1 over each cycle. It is crude in the sense that no preliminary scaling has been performed. BLUE: Period V. RED: Period IV. GREY: Period III. YELLOW: Period II. ORANGE: Period I. Periods and data covered corresponds to the preceding Figure 1.D.1.



The mean intertrade duration during primary cash market trading hours averages approx. two seconds before August 2007 and one and a half seconds thereafter. For the 15–minute period after the NYSE cash market close, the mean duration still averages approx. four and two seconds for the respective time periods.

For a liquidity analysis we divide the trading day into five intervals: Trading before 6.00am is negligible as it represents approx. 0.001% of total volume in the latter period.

Table 1.D.1: SPY liquidity summary statistics 4.00 EST - 18.12 EST

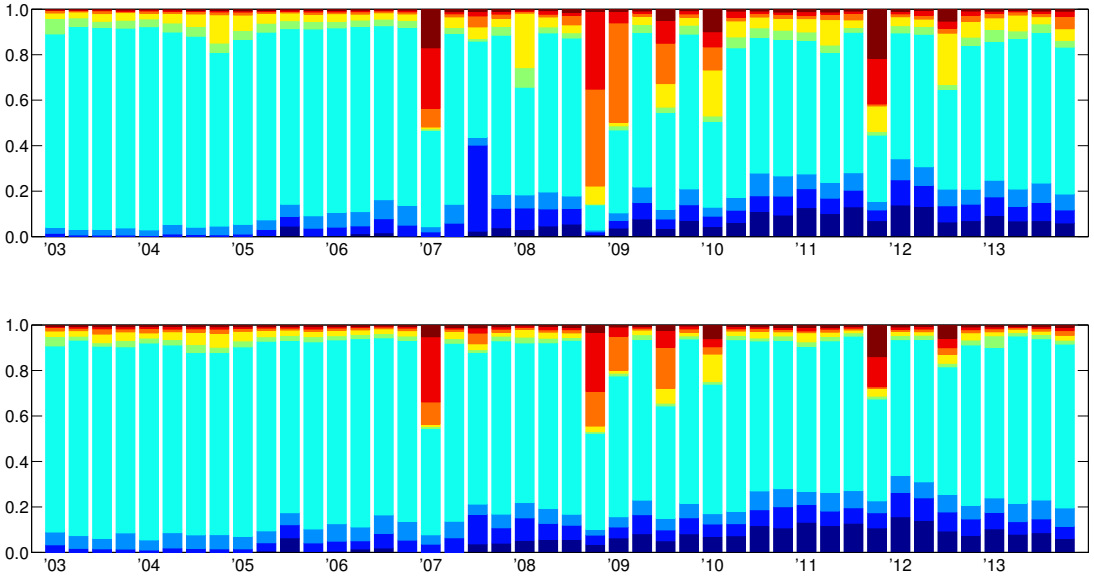
	Period I.)	Period II.)	Period III.)	Period IV.)	Period V.)
Starting	2003-01-02	2005-04-11	2007-02-26	2007-08-01	2010-08-02
Ending	2005-04-08	2007-02-23	2007-07-31	2010-07-30	2013-12-31
A Bin percentage volume as total volume					
4.00-6.00	-	0.00	-	0.01	0.05
6.00-7.00	-	0.02	-	0.04	0.07
7.00-7.30	-	0.06	-	0.18	0.17
7.30-8.00	-	0.12	0.10	0.26	0.23
8.00-8.30	0.11	0.59	0.44	0.83	0.62
8.30-9.00	0.27	1.38	1.13	1.28	1.05
9.00-9.30	0.19	0.89	0.68	0.96	0.99
9.30-16.00	93.29	91.91	93.88	92.45	91.46
16.00-16.15	4.06	3.03	2.54	3.48	5.14
16.15-16.45	1.41	1.39	0.56	0.65	1.12
16.45-17.15	0.38	0.34	0.17	0.28	0.46
17.15-17.45	0.19	0.18	0.20	0.10	0.15
17.45-18.12	0.12	0.10	0.27	0.10	0.19
B Ratio bin vs. average trading hours volume over same interval length					
4.00-6.00	-	0.00	-	0.00	0.00
6.00-7.00	-	0.00	-	0.00	0.00
7.00-7.30	-	0.01	0.01	0.03	0.02
7.30-8.00	-	0.02	0.01	0.04	0.03
8.00-8.30	0.02	0.09	0.06	0.12	0.09
8.30-9.00	0.04	0.20	0.16	0.18	0.15
9.00-9.30	0.03	0.13	0.09	0.14	0.14
9.30-16.00	1.00	1.00	1.00	1.00	1.00
16.00-16.15	1.15	0.87	0.71	0.99	1.47
16.15-16.45	0.20	0.20	0.08	0.09	0.16
16.45-17.15	0.05	0.05	0.02	0.04	0.07
17.15-17.45	0.03	0.03	0.03	0.01	0.02
17.45-18.12	0.02	0.02	0.04	0.02	0.03
C Bin mean duration in seconds					
4.00-6.00	-	503.5	-	393.6	261.3
6.00-7.00	-	300.8	-	156.5	126.6
7.00-7.30	-	136.0	103.8	28.0	30.7
7.30-8.00	-	113.9	94.3	18.3	23.4
8.00-8.30	97.4	54.0	41.3	6.6	9.2
8.30-9.00	69.8	32.4	22.5	5.0	5.7
9.00-9.30	58.1	26.1	21.1	4.7	4.8
9.30-16.00	2.4	2.0	1.6	1.3	1.3
16.00-16.15	3.8	3.8	2.8	2.0	2.4
16.15-16.45	25.5	26.3	41.9	10.6	9.6
16.45-17.15	79.8	93.4	97.7	29.0	24.6
17.15-17.45	116.5	120.9	122.8	46.9	41.5
17.45-18.12	137.9	127.4	138.2	51.4	40.6

Table 1.D.2: SPY liquidity summary statistics 4.00 EST - 18.12 EST (cont'd.)

	Period I.)	Period II.)	Period III.)	Period IV.)	Period V.)
Starting	2003-01-02	2005-04-11	2007-02-26	2007-08-01	2010-08-02
Ending	2005-04-08	2007-02-23	2007-07-31	2010-07-30	2013-12-31
D Bin median duration in seconds					
4.00-6.00	-	117	-	143	89
6.00-7.00	-	98	-	52	45
7.00-7.30	-	55	34	11	14
7.30-8.00	-	48	37	7	11
8.00-8.30	42	20	15	3	4
8.30-9.00	27	10	8	2	3
9.00-9.30	22	10	9	2	2
9.30-16.00	1	1	1	1	1
16.00-16.15	2	2	1	1	1
16.15-16.45	9	8	14	4	4
16.45-17.15	25	35	38	11	11
17.15-17.45	42	41	51	14	15
17.45-18.12	49	36	46	16	14
E Bin median of the longest intertrade duration per day in seconds					
4.00-6.00	-	2252	-	1444	1120
6.00-7.00	-	1000	-	691	549
7.00-7.30	-	453	434	155	162
7.30-8.00	-	401	337	124	137
8.00-8.30	363	259	228	51	64
8.30-9.00	306	182	134	40	43
9.00-9.30	274	162	135	37	41
9.30-16.00	35	32	23	13	14
16.00-16.15	27	31	23	16	16
16.15-16.45	212	225	283	93	77
16.45-17.15	352	381	378	168	144
17.15-17.45	429	447	421	240	222
17.45-18.12	482	503	462	234	215
F Bin median of the second longest intertrade duration per day in seconds					
4.00-6.00	-	802	-	876	765
6.00-7.00	-	518	-	428	380
7.00-7.30	-	266	250	116	122
7.30-8.00	-	255	216	93	106
8.00-8.30	233	183	162	41	52
8.30-9.00	199	135	97	32	35
9.00-9.30	187	124	103	31	33
9.30-16.00	30	28	19	11	12
16.00-16.15	21	25	18	13	13
16.15-16.45	143	151	191	70	60
16.45-17.15	223	239	241	123	111
17.15-17.45	249	253	251	163	158
17.45-18.12	287	275	238	161	155

Figure 1.D.3: Measurable intraday variance proportions over extended trading hours, quarterly 2003–2013

Variance proportions (in percentages) per intraday bin over extended trading hours for non–overlapping quarters 2003–2013 based on NW kernel regression on squared five second returns with a bandwidth choice corresponding to 90 seconds (upper panel) and on squared one minute returns with a bandwidth choise corresponding to two minutes (lower panel). DARK BLUE: 4.00–7.00am. BLUE: 7.00–8.30am. LIGHT BLUE: 8.30–9.30am. TURQUOISE: 9.30am–4.00pm. GREEN: 4.00–4.15pm. YELLOW: 4.15–4.45pm. ORANGE: 4.45—5.15pm. RED: 5.15–5.45pm. DARK RED: 5.45–6.12pm. X–AXIS: Year 'XX denotes the first quarter of the corresponding year with quarters adjusted to match starting and ending periods I.–V. whenever necessary.



Chapter 2

Reconciling Asymmetric Dynamics in Index Volatility and Constituent Correlations

Co–Authors: Stephen J. Taylor[1] Ser–huang Poon[2]

Abstract

Return volatility asymmetric effects are found in several U.S. indices throughout the 1900s. The magnitude of the effect is time-varying and it has become stronger in recent years. The effects are weaker for individual firms than for indices. Correlations between firm returns increase when the market falls, which explains the higher index asymmetry. Thus the index has more asymmetry than the firms. Asymmetry has increased over the last three decades both for the index and the mean firm.

Keywords: asymmetric volatility, asymmetric correlation

[1]Stephen Taylor is Professor of Finance at the Department of Accounting and Finance at The Mangement School, Lancaster University, United Kingdom.

[2]Ser–hunag Poon is Professor of Finance at the Department of Finance at The Manchester Business School, United Kingdom.

2.1 Introduction

Research concerning the asymmetry between, in particular, changes in stock returns and changes in a measure of their volatility has a prolonged history in finance. Asymmetric volatility effects have economically been motivated by primarily two theories, one 'leverage' and two 'volatility feedback'. An older opinion is that debt/equity ratios increase in a falling market and hence equity is more risky and so more volatile. But the asymmetric effect, (1) is very large compared with changes in d/e ratios, (2) occurs for firms which are debt-free.

We investigate into the former issue by estimating asymmetric ARCH–type models over rolling, overlapping windows of data to assess time–variation in the parameter structure for even small observation period changes and through time.

We document large changes in parameter values even when only 21 days are rolled over, i.e. less than 1% of a time window of ten years with 2,500 daily observations. Overall, the return–volatility index and constituent asymmetry seem to be getting stronger which is NOT a new phenomenon. The explanation we suggest links the increase in downside equicorrelation among the devolatized return series of constituent stocks to an increase in index volatility. The analysis is based on a univariate framework in which we examine the aggregate market return in terms of its constituents at each point in time. Doing so puts this study apart from those papers which either model portfolios of stocks in a low dimensional multivariate setting or incorporate the index returns into the multivariate setting, neglecting the fact that the index is a weighted average of constituents with very different persistencies and ARCH–type coefficients, which we point out. Focusing on the Dow Jones index with a small number of constituents we are able to track asymmetric patterns of index versus constituent stocks consistently since the 1920s and show the link between asymmetric index return volatility and asymmetric correlations between the return series of constituent stocks.

Asymmetric ARCH–type models are often assume to have a symmetric residual density at the daily level. There is then not much difference between a unit expectation of an asymmetric (such as a GJR–GARCH) and a symmetric specification (such as a GARCH(1,1)). However, the asymmetric properties in the model specification have important implications for the multi–horizon return distribution which introduces negative skewness over time.

2.2 Literature review: Asymmetric return–volatility innovations

2.2.1 The leverage effect

The term 'leverage effect' was first coined in Black (1976b), which, in its original meaning, refers to the observation that a fall in equity value (a negative return) and thus an increase in the debt–to–equity ratio leads to a rise in firm level risk, i.e. volatility.[3] Christie (1982) provides early evidence for Black's financial leverage story by examining *yearly face value* debt levels (assumed to be riskfree) to *quarterly market values* of equity in a regression/maximum likelihood framework.

If in turn we allowed for multiple tranches of risky debt marked–to–market on a regular basis or by taking credit spreads/credit defaults swaps into account, intuitively we might not expect a large shift and perhaps even a weakening in the associated negative elasticity of equity volatility with respect to its underlying stock price—as changes in denominator of the debt–to–equity ratio will then be anticipated in the same direction by the numerator as well. Loosely speaking, when the market value of equity drops, market price of debt also drops,[4] although in a non–linear way; for a discussion see also Aydemir, Gallmeyer, and Hollifield (2006).

Daouk and Ng (2011) unlever firm–level volatility based on a Merton–KMV and thus market–value based debt approach using quarterly debt data and returns data aggregated at the monthly frequency. Based on cross–sectional regression results this reveals a reduction in the magnitude of the asymmetric effect but overall firm–aggregated unlevered returns remain asymmetric and it does not provide an explanation of firm versus index level asymmetries.

Another consideration to be made are changes in interest rates. In Christie's contingent claim model the sensitivity of volatility to interest rates is clearly negative as the wealth transfer from bond to equity holders lowers the financial leverage, however, only conditional upon the caveat that firm value is independent of the interest rate. Stocks react negative to interest rate changes, i.e. a rise in expected or unexpected inflation (Fama and Schwert,

[3]Note that Black (1976b) uses an early predator to realized variance, i.e. squared daily returns aggregated at the monthly level, which is surprisingly not encountered in follow up studies except French, Schwert, and Stambaugh (1987), who further add twice the sum of products of adjacent daily returns to account for non–synchronous trading.

[4]The elasticity should still be left bounded at minus one while dividends may cause it to be positive when it should be right bounded at zero Black and Cox (1976, pp357-358).

1977), such that the overall interest rate sensitivity of equity volatility depends on the relative magnitude of these two effects and can only be tested empirically—the expected sign is unsure.

Incorporating these ambiguous interest rate effects (Christie's second model) leads to a positive association between interest rates and 'gross volatility' estimated from daily data. Under the premise of no omitted variables, interest rates and financial leverage have a positive impact on volatility at the quarterly horizon but cannot explain the entire magnitude of the asymmetric effect; compare also Schwert (1989). Moreover, changes in leverage are small while the asymmetric effect is comparatively large.

Inspired by Black (1976b) but not receiving much attention, Christie (1982b) also analyses operating leverage (relative to fixed assets) with similar results which he eventually attributes along with swings in dividends and the 'asset mix theory'[5] to firm level volatility being a random variable (Christie, 1982, p428)—putting a feasible disentangling of divergent economic explanatory effects generally into question.

Extending the previous work to finer frequencies and different lead–lag horizons but using even simpler methods, Duffee (1995) finds a *positive contemporaneous* relation between returns and the level of volatility (qualitatively in line with positive skewness) at the firm level while the sign between current returns and *future* one day ahead volatility levels is positive and negative one month ahead. As the contemporaneous effect is larger for both frequencies the effect on the net volatility change over the one month horizon, $log(\frac{\sigma_{t+1}}{\sigma_t})$, turns out to be negative. While here the contemporaneous financial leverage effect disappears when all 2,500 instead of only the largest AMEX and NYSE firms as in Christie (1982) are examined, Duffee (1995) confirms the positive firm size effect of Cheung and Ng (1992), i.e. larger firms exhibit less return–volatility innovation asymmetry, which is driven by the weaker *instantaneous* 'reaction' (correlation) of large firms' volatility level to return innovations.[6] Loosely speaking small firms exhibit more positive skewness than large firms, which Hong, Tu, and Zhou (2007), albeit in a very different methodological setting, view as a compensation due to their significantly larger (asymmetric) downside

[5] A firm holding a mixture of assets with different volatilities exhibits changes in firm level volatility whenever the relative weighting inside the portfolio changes with the exact change determined by the variance–covariance structure of the holding portfolio. Note this is straightforward to view in the light of different volatile projects, clients, target group behaviour etc. The change in total firm volatility can then be well approximated by e.g. Brownian Motion or as news arrive at a finite rate more accurately by a jump process perhaps.

[6] Note that here no separation is made between positive and negative innovations which have previously been shown to have significant impact (Nelson, 1991; Glosten et al., 1993).

correlation[7] at the monthly horizon. Note our analysis is based on the largest US blue–chip firms.

Duffee's findings are interesting in so far as the positive contemporaneous return—volatility innovation asymmetry contrasts the negative correlation asymmetry at the stock market aggregate level of e.g. French et al. (1987), which Duffee tries to explains by showing that in multi–factor models the market is negatively skewed, idiosyncratic firm returns are positively skewed with an additional positive skewness factor attached to small firms.

Nevertheless, all studies seem to have in common that the magnitude of the asymmetric effect seems infeasible large to be explained by changes in financial leverage (Figlewski and Wang , 2001). Moreover, somewhat counter–intuitive the asymmetry is larger for indexes (aggregate market returns) than for individual stocks; see e.g. Kim and Kon (1994).

2.2.2 The volatility feedback effect

Concern regarding the direction of causality from returns to volatility has already been expressed by (Christie, 1982), however, rather by addressing exogenous volatility shocks than a volatility feedback effect. The volatility feedback effect is based on a time–varying risk premia explanation mainly put forward by Pindyck (1984), Poterba and Summers (1986), French et al. (1987) and Campbell and Hentschel (1992):[8] Not news related to instantaneous return shocks but news affecting the perceived level of future volatility will trigger a sell–off in the asset until the expected return–risk ratio has been re–established under the assumption that volatility risk is priced.[9] However, this demand by risk–averse investors induces itself a rise in volatility, termed volatility feedback. Thus, markets can start declining immediately due to an anticipated expected rise in volatility, negative innovations (either self–excited or news–induced) then increase this effect. Using ARCH–M–type models French et al. (1987) document such a large negative contemporaneous return–volatility innovation effect which cannot be explained by relying on the leverage effect alone, however, as their "results are so ambiguous" (ibid., p28) over the 1928–1984 period they refrain from advocating any particular model. Further, one might want to raise stationarity concerns.

[7]Here, correlation between the asset and the market not the return–volatility correlation with regard to the same asset.

[8]Expected (unexpected) returns are positively (negatively) related to the predictable (unpredictable) component of volatility changes.

[9]This is somewhat paradox given that ARCH–type conditional volatility is known in advanced.

While the previously mentioned articles argue in favour of a positive relation between volatility and expected returns by relying on *inter alia* ARCH–M–type models, Nelson (1991) and Glosten et al. (1993) formulate and estimate asymmetric GARCH models and show that volatility increases more conditional on negative than on positive returns and after incorporating this effect the expected returns–volatility relationship at the one day horizon turns insignificant or even negative; although note the difference in sampling frequencies compared to previous studies. Their coefficient, which links volatility to returns, seems to depend on the horizon as it often turns out to be insignificant with low actual impact on volatility (Campbell and Hentschel, 1992). We might want to argue the latter observation might be due to a horizon mismatch, firstly, because the conditional volatility is known in advance, whereas there is some unexpected news on day t (the ARCH–type model problem in general), and secondly because the return might not (only) be related to volatility at the daily lead–lag frequency. The volatility feedback hypothesis involves a complicated economic process that passes through expectations and dividends to validate the negative relation and only (weakly) explains the longer term return–volatility relation.

In addressing this problem, Ghysels et al. (2005) have proposed a mixed frequency data sampling (MIDAS) GARCH approach in which, similar to French et al. (1987), they use daily returns to forecast one–step ahead monthly variance. While their results are in line with French et al. at the one–month horizon, the magnitude of the estimated asymmetric coefficient increases and the coefficient becomes statistically significant when the rolling lagged window is extended to the three– to four–month horizon and it does not change sign and remains significant when an additional asymmetry coefficient is introduced, in contrast to the results of Glosten et al. (1993).

2.2.3 Simultaneous comparison and differentiation of the two effects

On a theoretical basis, Wu (2001) endogenizes the volatility feedback based on the fact that standard equilibirum models with expected utility maximising agents fail to generate time–varying risk premia. In his simulated methods of moments model, uncertainty stems from shocks to dividend growth as in Campbell and Hentschel (1992) but also shocks to dividend volatility which in turn drive both stock returns and stock return volatility. Consequently, when uncertainty rises stock volatility increases and the stock price declines—resulting in asymmetric volatility independent of the leverage effect. The leverage itself is introduced through allowing dividend growth and volatility to be correlated.

Going further, Bekaert and Wu (2000)[10] point out that the volatility feedback effect/ time–varying risk premium for the market portfolio is consistent with the CAPM model but, now at the firm level, asymmetric covariance with the market is the important risk measure to focus on, which responds particularly strong (positively) to negative systematic market volatility changes.[11] This necessary condition, covariance asymmetry to explain the feedback effect at the firm level, can be identified by considering idiosyncratic shocks: Given such a shock the covariance between the market and firm return should not change. Consequently, the risk premium remains unchanged and idiosyncratic shocks generate asymmetry in volatility only through the leverage effect. The strength of the firm–specific feedback effect (market shock, magnitude of increase in covariance) then naturally depends on the firm's exposure to systematic risk. Multivariate results such as those of Bekaert and Wu (2000), unlike this study, are often based on portfolios of stocks due to the well known dimensionality issues.

The volatility and covariance asymmetry is particularly apparent for low leverage portfolios (ibid., p33, Fig. 7), which is counter–intuitive to the leverage effect story. But the absolute responses to changes in volatility/covariance are also considerably smaller for the low leverage portfolio as the response itself is a function of the magnitude of shocks. Now consider that low leverage portfolio shocks are generally much larger in magnitude of either sign compared to the medium and high leverage sorted portfolios.

Moreover, in his theoretical model Wu (2001) has shown that volatility feedback dominates in high volatility periods while dividend innovations are more important in tranquil periods. Bae, Kim, and Nelson (2007) incorporate Markov–switching regimes in the joint ARCH–type modelling of leverage and volatility feedback of Bekaert and Wu (2000) in order to distinctly address the volatility feedback effect of persistent volatility changes. Their results are in line with the theoretical study of Mayfield (2004): agents in low volatility regimes demand an additional up–front risk premium for being exposed to sudden switches.

[10] In contrast to Christie (1982), Bekaert and Wu (2000) use semi–annual (not annual) face value debt levels and daily (not quarterly) returns on portfolios of stocks grouped by leverage (not individual stock returns) with dividends reinvested (instead of ex–dividend returns) which they then aggregate at the weekly frequency over less than ten years of data (1985–1994) in the Japanese (not US) market.

[11] Under the assumption that the volatility feedback effect does not depend on leverage and a constant price of risk in a CAPM world (constant positive betas).

2.2.4 Economic explanations and 'investor sentiment'

The magnitude of the asymmetric volatility effect is too large to be explained solely by relying on volatility feedback and leverage effects in the Epstein–Zin preference model of Li (2003). Schwert (1989) and French and Sichel (1993) argue for asymmetry in the volatility of macroeconomic real variables linked to recessions. Then an anticipated expected downward revision of GDP growth triggers an immediate decline in stock prices followed by higher volatility.

From a very broad perspective, not related to firms but to 49 countries, Talsepp and Rieger (2010) also observe that economic development, GDP and market capitalization are important determinants of the asymmetric effect *per se*. Then, a higher participation of private investors and financial analyst coverage are positively related to asymmetry which in turn suggests 'investor sentiment' as an underlying force of the effect.

A trading–based explanation for daily asymmetric effects at the firm level over the period 1988–1998 is put forward by Avramov, Chordia, and Goyal (2006) who show that daily volatility regressed on lagged returns, market leverage and trading activity varies with selling activity thereby introducing time–variation (as selling activity is time–varying) in the asymmetry effect: in the presence of negative unexpected returns higher selling activity increases one day head volatility and vice versa. Moreover, quite intuitively, sell order initiated trades when returns are positive, denoted 'contrarian, informed sell trades', decrease subsequent volatility while 'uninformed, herding sell trades' increase volatility with price changes tending to be reversed in the latter case[12]—resulting in the small empirically observerd first order autocorrelation. Note that the former, 'contrarian sell trades' (in the presence of positive unexpected returns), have no impact on serial correlation.

In contrast, Ozdenoren and Yuan (2008) point out the incentives of coordinated strategic trading activity of *informed* not uninformed investors. Informed investors' learning then generates self–fulfilling beliefs with resulting multiple equilibria. However, this new source of volatility can be excessive only in illiquid markets when the signalling of informed investors is particularly strong as, e.g., in the housing market, and might thus be not particularly pronounced for very liquid stocks.

[12]For more literature on herding refer to review in Avramov et al. (2006, pp1243–44): Friedman (1953) [irrational destabilisation], Cutler, Poterba, and Summers (1990) and De Long, Shleifer, Summers et al. (1990) [positive feedback investment strategies can result in excess volatility], Froot, Scharfstein, and Stein (1992) [herding on news unrelated to fundamentals]. For uninformed/informed trades refer to the noisy rational expectations models of Hellwig (1980) and Campbell, Grossman, and Wang (1993) in Avramov et al. (2006) [volatility increases with non-informational, liquidity–driven trades]

Back to Avramov et al. (2006), they underpin some general results: 1.) incorporating trading activity measures turns previously significant daily leverage effects insignificant and 2.) high asymmetric effects are found for 117 NYSE stocks with less than 1% leverage. The R^2s for their daily volatility regression lie in the region of 15–22% over the total 2,232 firms considered and increase more than twofold to 38–52% for a subset of the largest 238 firms when realized instead of daily volatility measures are used.[13] However, from their findings we cannot infer the actual degree or magnitude of asymmetry explained. Here, a simple conditioning of the left–hand side variable on the sign of lagged or current returns might lead to further insights, e.g. we might expect a higher conditional R^2 for negative days.

Using a different methodology and without focusing explicitly on trading activity measures to distinguish between arguably herding (here temporary) and contrarian (here permanent) trades, Shively (2007) distinguishes regimes based on one–month lagged positive and negative returns using non–linear bivariate threshold models and then considers impulse response functions to find that more than two third of innovations are of 'temporary' nature during the negative, higher volatility regime while almost all innovations in the positive, lower volatility regime are permanent. Loosely speaking, as temporary innovations can be viewed as pure volatility and these occur extensively in the negative return regime, which is also the high volatility regime, asymmetric and excess volatility are linked together. Adding trading activity measures in the spirit of Avramov et al. (2006) might lead to potential refinements.

[13] Note this is not a forecast exercise as information available on day t is used on both sides of the regression equations.

2.3 Methodology

2.3.1 Generalized ARCH–type model setup

Following the ARCH–type literature, we define ε_t as a function of θ, $\{\varepsilon_t(\theta)\}$, to be a discrete time stochastic process with zero conditional mean,

$$\mathrm{E}_{t-1}[\varepsilon_t(\theta_0)] = 0 \qquad t = 1, 2, \ldots, \tag{2.3.1}$$

where $\mathrm{E}_{t-1}[\cdot]$ denotes the expectation operator conditional on the set of information available at $t-1$ (I_{t-1}) and likewise the conditional variance h_t as a function of θ, $\{h_t(\theta)\}$,

$$h_t(\theta_0) \equiv Var_{t-1}[\varepsilon_t(\theta_0)] = \mathrm{E}_{t-1}[\varepsilon_t^2(\theta_0)] \qquad t = 1, 2, \ldots, \tag{2.3.2}$$

which depends on the information set generated by past observations. Recall that we reserved any mnemonics related to σ_t^2 to refer to the integrated variance in continuous time, i.e. $\int_0^t \sigma_s^2 ds$. Both mean and variance functions are parametrized by the finite dimensional vector $\theta \in \Theta \subseteq R^m$, where θ_0 denotes the true parameters. The stochastic logarithmic return process $\{r_t(\theta)\}$ obtained from the asset price process $\{P_t(\theta)\}$,

$$\begin{aligned} r_t^{log_e}(\theta_0) &\equiv log_e[P_t(\theta_0)/P_{t-1}(\theta_0)] \\ &\equiv p_t(\theta_0) - p_{t-1}(\theta_0) \qquad t = 1, 2, \ldots, \end{aligned} \tag{2.3.3}$$

has conditional mean

$$\mu_t(\theta_0) \equiv \mathrm{E}_{t-1}[r_t(\theta_0)] \qquad t = 1, 2, \ldots. \tag{2.3.4}$$

As both $\mu_t(\theta_0)$ and $h_t(\theta_0)$ are both measurable w.r.t. I_{t-1},

$$\varepsilon_t(\theta_0) \equiv r_t - \mu_t(\theta_0) \qquad t = 1, 2, \ldots \tag{2.3.5}$$

is then defined to follow an ARCH process and the conditional variance for the residual process $\{\varepsilon_t(\theta)\}$ equals the conditional variance for the return process $\{r_t(\theta)\}$. The standardized residual process $\{z_t(\theta)\}$,

$$z_t(\theta_0) \equiv \frac{\varepsilon_t(\theta_0)}{\sqrt{\mathrm{E}_{t-1}[\varepsilon_t^2(\theta_0)]}} \equiv \frac{\varepsilon_t(\theta_0)}{\sqrt{h_t(\theta_0)}} \qquad t = 1, 2, \ldots, \tag{2.3.6}$$

will constitute a conditional mean of zero with time invariant conditional variance of unity.

For a time invariant conditional distribution of z_t with finite fourth moment, it follows

by Jensen's inequality for any real–valued convex function $f(\cdot)$, $\mathrm{E}[f(X)] \geqslant f[\mathrm{E}(X)]$, that

$$\mathrm{E}(\varepsilon_t^4) = \mathrm{E}(z_t^4)\mathrm{E}[(h_t)^2] \geqslant \mathrm{E}(z_t^4)[\mathrm{E}(h_t)]^2 = \mathrm{E}(z_t^4)[\mathrm{E}(\varepsilon_t^2)]^2, \tag{2.3.7}$$

or equivalently

$$\frac{\mathrm{E}(\varepsilon_t^4)}{[\mathrm{E}(\varepsilon_t^2)]^2} \geqslant \mathrm{E}(z_t^4), \tag{2.3.8}$$

where the inequality is strict unless the conditional variance is constant. Imposing a normal distribution on the standardised residuals z_t in Eq. (2.3.6), the unconditional distribution of the residuals ε_t is leptokurtic. More generally, excess kurtosis in the distribution of ε_t can arise from both time variation in h_t and from excess kurtosis in the underlying distribution of ε_t, i.e. z_t.

2.3.2 Empirical Model Specification

Let r_t^i be the return on the asset i and $\mu_t^i = \mathrm{E}[r_t^i|F_t]$ where F_t denotes the information set at time t. Conditional variance of r_t^i is given by $h_t^i = Var[r_t^i|F_t]$. r_t^i displays asymmetric volatility if

$$Var[r_t^i|I_{t-1}, \varepsilon_{t-1}^i < 0] - h_{t-1}^i > Var[r_t^i|I_{t-1}, \varepsilon_{t-1}^i > 0] - h_{t-1}^i, \tag{2.3.9}$$

i.e. unexpected negative innovations can trigger an upward updating of the conditional volatility (depending on the threshold ARCH–type specification), whereas positive innovations lead to less upward or even downward revision of conditional volatility. Consequently, strong asymmetry is present when the left–hand (right–hand) side is smaller (larger) than zero.

In order to examine the return volatilty asymmetries, we use an extension of the ARCH model developed by Glosten et al. (1993) - the GJR–GARCH(1,1,1)-MA(1) model. In particular, return and conditional volatility equations are as follows

$$\begin{aligned} r_t^i &= \mu_t^i + \varepsilon_t^i = \mu_t^i + (h_t^i)^{1/2} z_t^i, \quad z_t^i \sim i.i.d.\ D(0,1) \\ \mu_t^i &= \mu^i + \theta^i \varepsilon_{t-1}^i \\ h_t^i &= \omega^i + (\alpha^i + \alpha^{i,-} S_{t-1}^i)(\varepsilon_{t-1}^i)^2 + \beta h_{t-1}^i, \end{aligned} \tag{2.3.10}$$

$$S_{t-1}^i = \begin{cases} 1 & \text{if} \quad \varepsilon_{t-1}^i < 0 \\ 0 & \text{if} \quad \varepsilon_{t-1}^i \geq 0 \end{cases},$$

where, $i = 1, ..., N, N+1$ with N representing the number of individual assets and $N+1$

is the index and $\omega^i \geq 0$, $\alpha^i >$, $\alpha^i + \alpha^{i,-} > 0$, and $0 < \beta^i < 1$. Note that these parameter restrictions are necessary for the existence and stationarity of the conditional variance process. In specific, the second constraint ensures that $\frac{\partial h_t^i}{\partial \epsilon_{t-1}^i} < 0$ when $\epsilon_{t-1}^i < 0$ and the third constraint allows the parameter $\alpha^{i,-}$ to be either positive or negative as long as $\frac{\partial h_t^i}{\partial \epsilon_{t-1}^i} > 0$ when $\epsilon_{t-1}^i < 0$.

The effect of the squared residual on the conditional variance of the return captures the asymmetries via the parameter, $\alpha^{i,-}$. The squared residual, $(\varepsilon_{t-1}^i)^2$, is multiplied by α^i if the return on asset i is larger than or equal to its conditional expectation, whereas it is multiplied by $\alpha^i + \alpha^{i,-}$ when the return i is lower than its conditional expectation. In other words, if $\alpha^{i,-}$ is positive, then negative residual will increase the conditional volatility more than the positive residual with the exact same size. Moreover, larger value of $\alpha^{i,-}$ indicates the presence of stronger asymmetries.

The conditional distribution of returns are consequently represented by

$$r_t^i | I_{t-1} \sim D(\mu_t^i, h_t^i), \tag{2.3.11}$$

where we consider both the standard normal and the standardized t–distribution for $D(0,1)$, where the latter density function is determined by one parameter, the degrees of freedom ν. It appears that that excess kurtosis is not just a result of stochastic volatility, i.e. is still present after applying a conditional volatility standardization, (Bollerslev, 1987). Then,

$$f(z|\nu) = c(\nu)\Big[1 + \frac{z^2}{\nu - 2}\Big]^{-(\nu+1)/2}$$

with $c(\nu)$ defined in terms of the gamma function, $\Gamma(\cdot)$, by

$$c(v) = \frac{\Gamma(\frac{1}{2}(\nu+1))}{\Gamma(\frac{1}{2}\nu\sqrt{\pi(\nu-2)})},$$

where $\nu > 2$, and the kurtosis being finite when $\nu > 4$, i.e. $3(\nu-2)/(\nu-4)$. As $\nu \to \infty$, the density function converges to that of the standard normal distribution.

Moreover, in order to examine the robustness of our results we specify an exponential GARCH model after Nelson (1991). Our EGARCH(1,1,1)-MA(1) model has log variance dynamics specified by

$$\ln(h_t^i) = \omega^i + \nu^i\Big(\Big|\frac{\varepsilon_{t-1}^i}{\sqrt{h_{t-1}}}\Big| - \mathrm{E}\Big[\Big|\frac{\varepsilon_{t-1}^i}{h_{t-1}^i}\Big|\Big]\Big) + \gamma^i \frac{\varepsilon_{t-1}^i}{\sqrt{h_{t-1}^i}} + \beta_1 \ln(h_{t-1}^i - \omega^i). \tag{2.3.12}$$

with stationarity condition $-1 < \beta^i < 1$ and where for the normal distribution we have

$$\mathrm{E}\Big[\Big|\tfrac{\varepsilon^i_{t-1}}{h^i_{t-1}}\Big|\Big] = \mathrm{E}[z^i_{t-1}] = \sqrt{\frac{2}{\pi}},$$

and analogous for the standardized t–distribution,

$$\mathrm{E}[z^i_{t-1}] = \frac{2\sqrt{(\nu-2)}\Gamma[(\nu+1)/2]}{\sqrt{\pi}(\nu-1)\Gamma[\nu/2]}.$$

2.3.3 Statistical Properties

Measures of Asymmetry We define the measure of asymmetry $A^i \in [-1,1]$ for GJR-GARCH(1,1)-MA(1,1) specification as follows,

$$A^i = 2\Big(\frac{\alpha^i + \alpha^{i,-}}{2\alpha^i + \alpha^{i,-}} - \frac{1}{2}\Big). \tag{2.3.13}$$

Note that

- A^i converges towards one when α^i tends to zero. In particular, $A^i = 1$ if and only if $\alpha^i = 0$,
- A^i is zero in the symmetric case. In particular, $A^i = 0$ if and only if $\alpha^{i,-} = 0$,
- A^i is theoretically bounded at minus one for the reversed asymmetric case, i.e. $\alpha^i > -\alpha^{i,-}$, as $\alpha^{i,-}$ is not restricted to be positive.

We specify a corresponding asymmetry measure for the EGARCH model. It can be defined by first calculating the two slopes of the news impact curve,

$$\alpha^i_{\text{EGARCH}} = \nu^i - \gamma^i \text{ and } \alpha^{i,-}_{\text{EGARCH}} = \nu^i + \gamma^i \tag{2.3.14}$$

an then inserting the slope coefficient in Eq. (2.3.13). Note that, however this ratio is not bounded as the corresponding α^i_{EGARCH} coefficient is not restricted to be positive, which is empirically relevant as we shall later see.

The impact asymmetry measure is then naturally defined as

$$A^i_{Impact} = A^i * \Big(|\alpha^i| + \frac{1}{2}|\alpha^{i,-}|\Big) \tag{2.3.15}$$

The persistency can be calculated as follows

$$\phi^i = \alpha^i + \frac{1}{2}\alpha^{i,-} + \beta^i, \tag{2.3.16}$$

and the unconditional variance, which enables us to reveal non–stationarity, is given by

$$\sigma^{i\,2} = \frac{\omega^i}{1-\phi^i}, \tag{2.3.17}$$

and the half–life of a variance shock is given by

$$H^i = log(0.5)/log(\hat{\phi}^i). \tag{2.3.18}$$

2.3.4 Correlations

Note that the conditional correlation between return on asset i and return on asset j is given by

$$\begin{aligned}
\mathrm{Corr}_t(r^i_{t+1}, r^j_{t+1}) &= \frac{\mathrm{Cov}_t(r^i_{t+1}, r^j_{t+1})}{\sqrt{\mathrm{Var}(r^i_{t+1})}\sqrt{\mathrm{Var}(r^j_{t+1})}} \\
&= \mathrm{Cov}_t\left(\frac{r^i_{t+1}-\mu^i_{t+1}}{\sqrt{h^i_{t+1}}}, \frac{r^j_{t+1}-\mu^j_{t+1}}{\sqrt{h^j_{t+1}}}\right) \\
&= \mathrm{Cov}_t\Big(z^i_{t+1}, z^j_{t+1}\Big) \\
&= \mathrm{Corr}_t\Big(z^i_{t+1}, z^j_{t+1}\Big).
\end{aligned}$$

Hence, conditional correlation between return on asset i and return on asset j is equivalent to the conditional covariance between the standardized residuals, here (z^i_t, z^j_t) and also equivalent to the correlation coefficient between those standardized residuals.

Assume that $\rho^{i,j}_t = \mathrm{corr}_t(r^i_t, r^j_t) = \rho$ is constant over time and the same across all assets mutually i, j. To obtain an estimator for ρ, consider

$$\begin{aligned}
y_t &= \sum_{i=1}^{N} z^i_t \\
y^2_t &= \sum_i z^{i\,2}_t + 2\sum_{i\neq j} z^i_t z^j_t
\end{aligned}$$

$$
\begin{aligned}
\tfrac{1}{2}(y_t^2 - \sum_i {z_t^i}^2) &= \sum_{i \neq j} z_t^i z_t^j, \\
\tfrac{1}{2}\Big(\mathrm{E}[y_t^2] - \sum_i \mathrm{E}[{z_t^i}^2]\Big) &= \sum_{i \neq j} \mathrm{E}[z_t^i z_t^j] \\
\tfrac{1}{2}\Big(\mathrm{E}[y_t^2] - \sum_i \mathrm{E}[{z_t^i}^2]\Big) &= \sum_{i \neq j} \mathrm{E}[z_t^i z_t^j] \\
\tfrac{1}{2}\Big(\mathrm{E}[y_t^2] - \tfrac{N}{2}]\Big) &= \frac{N(N-1)}{2}\rho \\
\rho &= \Big(\mathrm{E}[y_t^2] - N\Big)/[N(N-1)]
\end{aligned}
$$

Then, the natural estimator for the correlation coefficient is

$$\hat{\rho} = \Big(\tfrac{1}{T}\sum_{t=1}^{T} y_t^2 - N)/[N(N-1)]\Big) \tag{2.3.19}$$

Example: For an index such as the Dow Jones Industrial Average containing 30 constituents this implies,

$$y_t = \sum_i^{30} z_{i,t},$$

and then the average constituent equicorrelation can be computed as

$$\hat{\rho}_{\text{equi.}} = \Big(\frac{1}{T}\sum_{t=1}^{T} y_t^2 - 30\Big)/(2 \times 435),$$

where 30 corresponds to the diagonal covariance matrix entries with two times 435 off–diagonal entries, which then yields a crude estimate of the (average) constituent equicorrelation based on standardized residuals.

Asymmetric Correlations We can then further disentangle the correlation estimator by conditioning on lagged aggregate market return innovations being either negative or positive. And, going further, by making the conditioning indicator variable a threshold function of the magnitude of standardized aggregate market residuals. Then, we define the up and down threshold correlation in the following way,

$$\hat{\rho}^{(\text{up})}_{\text{equi.},c} \mid I_{t-1} = \Big(\frac{1}{\sum 1}\sum_{t=1}^{T} y_t^2 1_{z_{t-1}^x > c} - N\Big)/(2 \times (N-1)), \tag{2.3.20}$$

$$\hat{\rho}^{(\text{down})}_{\text{equi.},c} \mid I_{t-1} = \Big(\frac{1}{\sum 1}\sum_{t=1}^{T} y_t^2 1_{z_{t-1}^x < c} - N\Big)/(2 \times (N-1)), \tag{2.3.21}$$

where the indicator function $1_{z^x_{t-1<c}}$ is one when $z^x_{t-1<c}$ is true and zero otherwise. Note that z^x_{t-1} can be either one of following: the standardized aggregate market return residuals obtained from a univariate index return series, the mean or median of the standardized residuals across constituent stocks with the exceedance level c being 0, 0.125, 0.25, 0.5.

Maximization in a likelihood framework

The success of ARCH–type models can probably be traced back to the relatively easy to implement estimation procedure through the likelihood function L determined by the product of conditional densities,

$$L^i(\theta^i) = f(r^i_2, \ldots, r^i_T | r^i_1, \theta) = \prod_{t=2}^{T} f(r^i_t | r^i_{t-1}, \theta^i), \tag{2.3.22}$$

more generally,

$$L(\theta^i) = \prod_{t=2}^{T} f(r^i_t | I_{t-1}, \theta^i), \tag{2.3.23}$$

where I is the information set. This is also the main feature in differentiating observation–driven ARCH–type from parameter–driven SV models in which the predictable component of the density of future returns is specified implicitly through the dynamics of the model, i.e. the variance can be modelled separately leaving much flexibility to superimpose almost any (often affine) dependence structure. The log likelihood (l) is straightforward to obtain, here the case of normally distributed (white noise) innovations underlying the conditional densities,

$$\begin{aligned} \max l^i(\theta^i) &= \textstyle\sum_{t=2}^{T} \Big[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h^i_t(\theta^i)) - \frac{(r_t - \mu^i_t(\theta^i))^2}{2h^i_t(\theta^i)} \Big], \\ &= -\frac{1}{2}\Big((T-1)\ln(2\pi) + \textstyle\sum_{t=2}^{T} \Big[\ln(h^i_t(\theta^i)) + \frac{(r^i_t - \mu^i_t(\theta^i))^2}{2h^i_t(\theta^i)} \Big] \Big), \\ &\propto -\frac{1}{2}\textstyle\sum_{t=2}^{T} \Big[\ln(h^i_t(\theta^i)) + {z^i_t}^2(\theta^i) \Big], \end{aligned} \tag{2.3.24}$$

The degrees of freedom parameter ν^i is included in θ^i for asset i. Then h^i_t and z^i_t are functions of some subset θ^{i*} of θ^i and the density function of z_t is determined by the remaining element ν^i of θ^i. The conditional density of observation t of asset i is consequently

$$f(r^i_t | I_{t-1}, \theta^i) = \frac{f(z_t(\theta^*)|\nu^i)}{\sqrt{h_t(\theta^*)}} \tag{2.3.25}$$

with the log likelihood function being proportional to

$$\propto \sum_{t=2}^{T} -\frac{1}{2}\ln(h^i_t(\theta^{i*})) + \ln(f(z^i_t(\theta^{i*})|\nu^i)) \tag{2.3.26}$$

whose maximization requires numerical methods in an iterative approach. The first variance estimate of h_1^i is set equal to the unconditional variance estimate $\widehat{\text{Var}}_{\text{unc.}}$ and after the optimization converges is included in the parameter vector θ^i which is to be re–estimated.

It can immediately be seen that a period of high volatility increases the likelihood estimate $l(\hat{\theta})$ compared to a period of low volatility without yielding any suggestions about the model's implied adequacy of the data generating process (DGP) when evaluating non– or overlapping time spans.

2.4 Empirical Analysis

2.4.1 Data Description

In the subsequent analysis we investigate into the dynamics and interrelation of the Dow Jones Industrial Average, henceforth DJIA, with its constituent stocks representing approx. 27% of the float–adjusted market capitalization of the Dow Jones US TSM Index, which in turn provides near complete coverage of the U.S. stock market (Dow Jones Index Product Description; Bloomberg).

Index history and historical constituent changes

Index inception. Charles H. Dow introduced his twelve–stock average on May 26$^{\text{th}}$ 1896. In 1916 the number was raised to 20 stocks and on October 1$^{\text{st}}$ 1928 it was raised again to 30 stocks, where it remains. Between 1928 and 1936 the index was subject to 23 constituent changes with a median time of the replacement constituents being index member of only 654 days incl. non–trading days, see Table 2.A.2,—rendering an observational driven model analysis unsuitable.

Time Period Covered. We estimate ARCH–type models over rolling observation windows of a fixed time span, restarting the estimation procedure at the 1st trading day of each new calendar month for all constituents and aggregate market returns proxies; here weighted combinations of index constituent stock returns. The CRSP database has recordings dating back to 1925 covering only few of the largest stock traded at that time. The analysis here starts from 1936 onwards to assured a long enough history of data for each constituent stock considered and the index composition being more stable in terms

of fewer constituent changes, see Table 2.A.3, 2.A.4.

Rolling Mechanism. Each likelihood optimization for index construct/ reconstructs generally aims to cover ten years of daily data. When a constituents stock does not have a ten year window of data I only keep expanding the time series (to the right) and start rolling once we have reached the full ten year time window. This only occurs during the early years in which we jointly assign an eight year window to all stocks. When a name change, merger or acquisition takes place we follow the usual convention and maintain the 'survivor time series' where we could have also arguably started a new time series of observations. Consequently, we are left with 64 time series of constituents since 1936 with an average (median) time of being constituents of 38 (29) years, which contains new entrants such that the conditional number upon having left the index is much higher at 47 (55) years.

Data cleaning, aggregation, index reconstruction and benchmarking

Return definition. For daily inference, we use the arithmetic one period return excluding dividends due to better aggregation properties as opposed to log returns (the log of a sum is not equal to the sum of the log), then

$$r_t^{arithm.}(\theta_0) = \frac{p_t - d_t}{p_{t-1}} - 1 \qquad t = 1, 2, \ldots, \tag{2.4.1}$$

where d_t are any dividends paid out on day t. For weekly inference, we convert the daily arithmetic returns into daily continuously compounded log returns,

$$r_t^{\ln}(\theta_0) = \ln(1 + r_t^{arithm.}) \qquad t = 1, 2, \ldots, \tag{2.4.2}$$

as then multi–period returns are the sums of single–period returns. For the portfolio aggregation we convert the weekly log return back into arithmetic returns. However, it would be surprising if any result would be heavily influenced by the choice of the return measure as can be seen from the first order approximation

$$r_t^{arithm.}(\theta_0) = exp(r_t^{\ln}(\theta_0)) - 1 = r_t^{\ln}(\theta_0) + \frac{1}{2}[r_t^{\ln}(\theta_0)]^2 + \ldots \qquad t = 1, 2, \ldots. \tag{2.4.3}$$

Trading days. We retrieve all days for which the CME group company Dow Jones Indexes publishes an index level and then disregard all days for which we cannot jointly

retrieve any constituent data at all from the CRSP database.[14] Note that in most of these cases, compared to the previous day, the index level remains unchanged, which confirms our procedure. The DJIA index and constituent data consequently covers the 19,805 trading days between 02/01/1936 and 30/11/2011 inclusively, which corresponds to on average 260.59 yearly trading days over the 76 year period due to the fact that trading took place on six days a week in earlier decades. Thus, we are on the conservation side with our cleaning procedure and have rather deleted too few than too many observations.

Index reconstruction. The Dow Jones Industrial Average is a price-weighted average measuring the return (excluding dividends) on a portfolio that holds one share of each stock adjusted by a divisor for stock splits, stock dividends of more than 10% and when constituents are replaced by another. We reconstruct the price–weighted index and manually check all days with absolute deviations larger than 0.3% (30bps) for pricing errors. Adjustments are made based on a reconciliation of the index divisor (construct from the information file regarding divisor updates on the Dow Jones Indexes website), and the CRSP database variables price (PRC), shares outstanding in thousands (SHROUT), cumulative adjustment factor current price (CAFCRP) and cumulative adjustment factor current shares (CAFSHR). Consequently, we make adjustments on days when either stock splits or stock dividends are reflected in the CRSP database but not yet in the index or vice versa.[15] The remaining days with large mispricings remain only partly unreconciled as the mispricing tends to be reversed the next trading day.[16] From these magnitudes in the footnote we further notice that up until the 1990s the reconstructed DJIA is more negative and therefore less positive the next or a following trading day and a reverse effect is observed for the last two recent observations. Potential reconciliation and thus pricing

[14] Surprisingly only on the following days between 1994 and 2001: 04/09/1994, 05/09/1995, 23/11/1995, 25/12/1995, 01/01/1996, 29/02/1996, 05/04/1996, 27/05/1996, 04/07/1996, 02/09/1996, 28/11/1996, 25/12/1996, 01/01/1997, 17/02/1997, 28/03/1997, 01/09/1997, 19/01/1998, 16/02/1998, 10/04/1998, 03/07/1998, 26/11/1998, 15/02/1999, 02/04/1999, 31/05/1999, 06/09/1999, 25/11/1999, 24/12/1999, 17/01/2000, 21/02/2000, 29/05/2000, 23/11/2000, 01/01/2001, 15/01/2001, 19/02/2001, 13/04/2001, 28/05/2001.

[15] e.g. on all 31 trading days between 04/01/1984 and 15/02/1984 for which American Telephone & Telegraph name changed into AT&T with a large divestment distributed as cash dividend resulting in a CAFCRP down from 5.258352 to 1.374932 which has not been reflected in the CRSP database (but in the index).

[16] 06/12/1937 (-0.0032), 07/12/1937 (0.0034), 30/04/1938 (-0.0035), 02/05/1938 (0.0032), 29/09/1938 (-0.0035), 30/09/1938 (0.0035), 04/06/1942 (-0.0035), 05/06/1942 (0.0035), 02/11/1946 (-0.003), 04/11/1946 (0.003), 14/11/1949 (-0.0105), 15/11/1949 (0.0106), 18/06/1952 (-0.0074), 19/06/1952 (0.0076), 11/07/1980 (-0.0045), 14/07/1980 (0.0044), 14/10/1987 (-0.0032), 15/10/1987 (0.0031), 19/10/1987 (-0.0118), 21/10/1987 (0.0177), 23/01/1990 (0.0048), 24/01/1990 (-0.0044), 03/08/1990 (-0.004), 06/08/1990 (0.0041), 09/08/1996 (0.0057), 12/08/1996 (-0.0059), 20/10/1999 (0.0035), 21/10/1999 (-0.0032)

errors could be due to missing closing prices in the CRSP database, which we reconstruct by using the mid price of the last traded ask and bid price without matching the price grid in earlier days. If these are not available either, i.e. there was no recorded trade on that day, we set the price equal to the closing price of the previous trading day.

Reconciliation error. Figure 2.B.1 shows these residual return series pricing errors (DJIA Original minus DJIA Reconstructed) from which we can infer distinct periods of higher disturbances (1977–2001) and periods in which the pricing error is virtually zero (1952–1977 and after 2001), see Figure 2.B.1. Our reconstructed DJIA index return series exhibits a root mean squared error (RMSE) of only 4.62bps (basis points, where 100bps = 1%). On 1,358 (1,378) days we note positive (negative) deviations larger than 5bps in magnitude, i.e. on 86.19% of trading days (17,069 thereof) deviations are smaller than 5bps and thus inside the red 5bps threshold lines in Figure 2.B.1.

Alternative index constructs and other benchmarks. For robustness checks, we also perform the analysis for a series of alternative benchmark indexes as the original DJIA, a price weighted index, overweights higher priced shares in determining the performance of the index. Consequently, we create a synthetic DJIA market value weighted index, which we denote DJIA MV, and a synthetic equally weighted DJIA index, denoted DJIA EW.

Figure 2.B.2 and 2.B.3 in the Appendix visualize the strong performance impact on the different index calculation methodologies. Compared to the price weighted original DJIA index level series, an (artificial) equally weighted construct, which requires daily rebalancing, based on the same underlyings appears to outperform the original index in booms and to underperform (reimburses its performance gains) during recessions. It corresponds to an investment strategy of holding a portfolio of shares in portion to their market capitalisation (excluding cash dividends) and thus provides a better proxy of the dynamics of the underlying economy than a price–weighted index. The DJIA Original return series correlations with its constructs are as follows: DJIA Reconstructed (0.999), DJIA MV (0.967), DJIA EW (0.985), and DJIA MV with DJIA EW (0.957).

For comparative purposes we also investigate into the following three market–value–weighted indices, listed in Table 2.A.1, which are plotted along in Figure 2.B.2 and Figure 2.B.3: One, the standard benchmark S&P500; two, the Wilshire 5,000 index, which despite its name comprises of about 4,000 - 7,000 stocks of all NYSE, AMEX and actively traded

NASDQ stocks; and three, the Wilshire 4,500, which corresponds to the Wilshire 5,000 excluding the S&P500 constituent companies.

2.4.2 Aggregate market & constituent returns summary statistics

Graph Legend. In the following analysis, the Original price weighted DJIA Index is shown as DARKRED, the price weighted reconstructed DJIA Index is shown as RED, the reconstructed market value weighted DJIA Index is shown as BLUE, and the reconstructed equally weighted DJIA Index is shown as ORANGE. The GREY lines and areas refer to underlying constituent stocks of the index at each point in time on a monthly rolling forward basis: The DARKGREY line refers to the average mean constituent value of the respective parameter or statistic, the GREY area around the mean refers to the inter–quartile area of the constituent parameter or statistic value and the LIGHTGREY, even wider area spans the area between the first and last decile respective value.

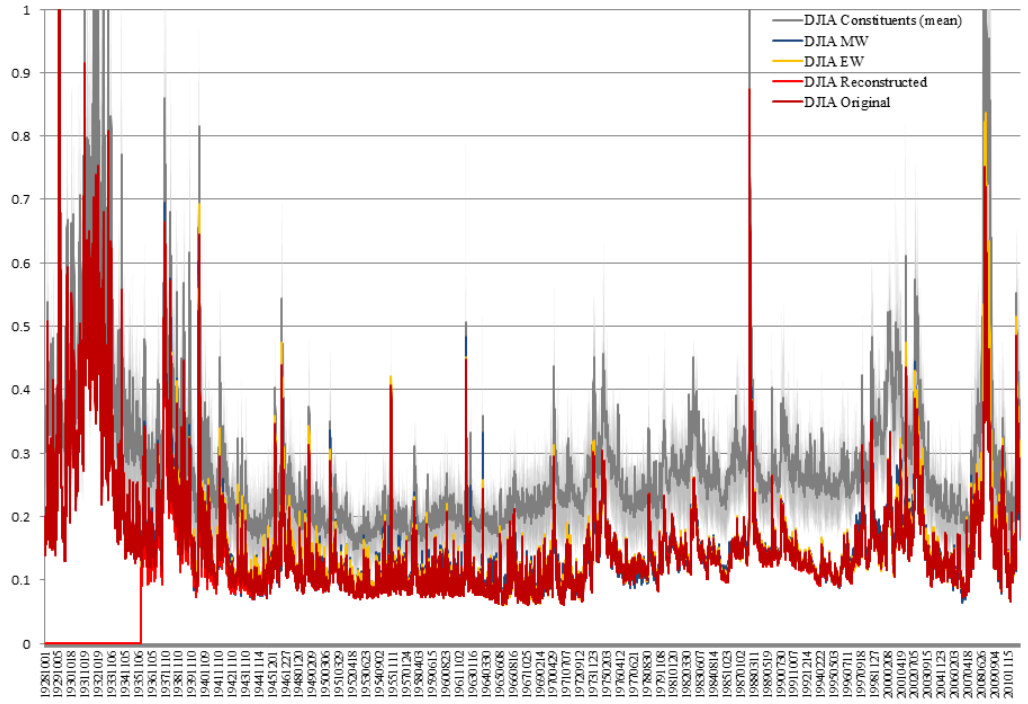
Volatility and Diversification. In order to grasp the volatility dynamics since the 1920s, we show in Figure 2.1 the daily conditional volatility plots for aggregate market return indices (for different weighting functions of constituent returns) along with the mean, interquartile range and 1st to 9th decile estimates of the constituents underlying the index at each point in time—estimated from one month rolling ten year periods and unconditional model implied volatility estimates, which is a function of the models' parameters in Figure **??**. The use of rolling observation windows enables insight into the evolution of parameter estimates in the subsequent subsections and here in particular it mitigates the well known violation of stationarity assumptions of longer observation periods.

Overall, the benefits from diversification are clearly visible. However, from Figure 2.1 it is also appears to be the case that diversification benefits decrease when overall volatility increases. This is even more apparent when comparing weekly model based conditional volatility estimates for aggregate market return to those of constituent stocks, refer to Figure 2.B.4 in the appendix.

A standard error of the model based unconditional variance estimates is not easily obtainable as it is a non–linear function of the model parameters. The model implied unconditional variances for constituents and the index are very high before the 1940s (mid 1940s when a t–distribution for the error density is imposed) and there might be non–trivial estimation error in the the ω^i which amplifies the estimation error in unconditional

Figure 2.1: Daily model based conditional volatility estimates; 1928–2011

Daily conditional volatility estimates (p.a.) based on the $EGARCH(1,1,1)$-$MA(1)$-N model, where N denotes normally distributed error density. Estimates shown are obtained from estimating the model over one month forward rolling ten year time windows.



variances.

Market dynamics are clearly different for the volatile latest period versus the earliest period in the overall sample considered. Further, we notice a persistent upward shift in the long run unconditional market variance from the 1970s onwards. Also note that the equally and market value weighted aggregates are more risky in an unconditional model based sense compared to the price weighted market aggregate which should help in explaining its higher realized returns as shown in Figures 2.B.2 and 2.B.3, both in the appendix. The appendix also contains the equivalent figures for weekly model based conditional volatility of returns, Figure 2.B.4 and model implied unconditional volatility estimates for the $GJR(1,1,1)$-$MA(1)$-t model, Figure 2.B.5. Other figures are available upon request from the author.

Sample skewness statistics. As previously elaborated upon, there is not much difference between a unit expectation of a symmetric and an asymmetric ARCH–type model as the error density is assumed symmetric. The conditional expectation based on a negative or positive return residual realization is well described by an asymmetric model which has important implication for multi period returns. Negative return volatility asymmetry then

leads to negatively skewed multi period return densities, see, e.g. Taylor (2005).

Figure 2.2: Sample skewness of weekly index and constituent returns; 1928–2011

The figures show the sample coefficients of skewness of weekly index and constituent plain returns from 1928 to 2011 over one month forward rolling ten year time windows.

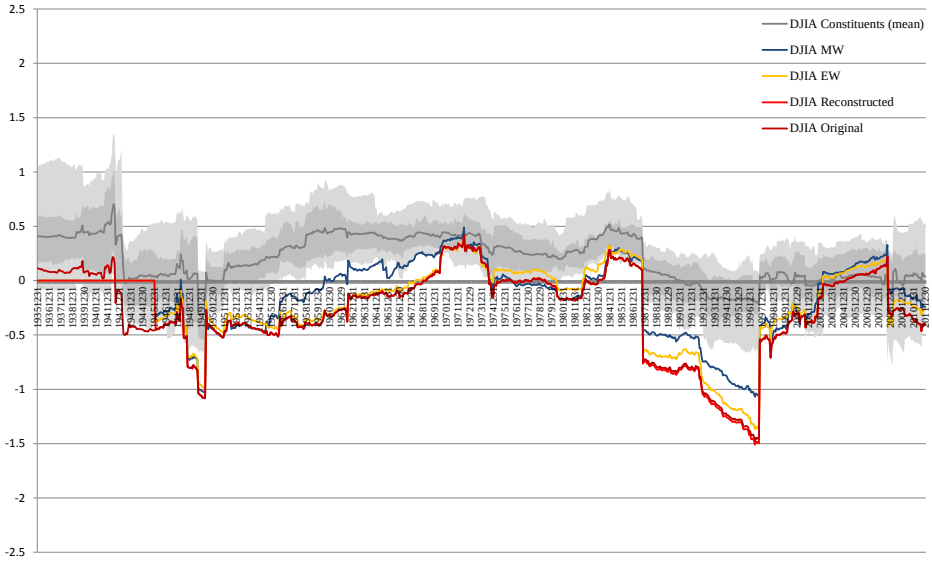


Figure 2.3: Sample skewness of standardized weekly return residuals; 1935–2011

We standardize weekly index and constituent returns by their respective $EGARCH(1,1,1)$-$MA(1)$-N model based conditional volatility estimates. Respective summary statistics for the GJR(1,1,1)-MA(1)-N and $-t$ models are very similar and thus not shown for clarity of exposition. Available upon request.

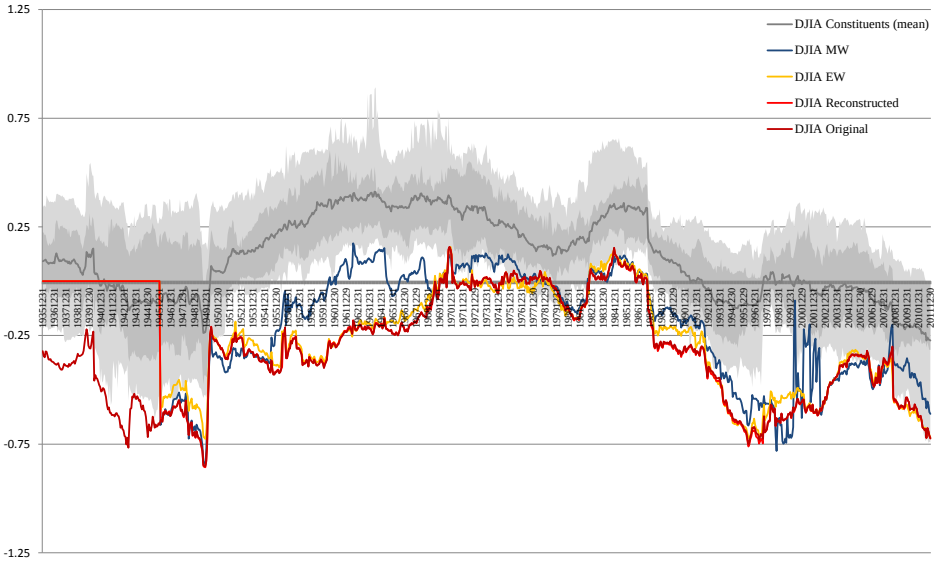


Figure 2.2 shows the sample skewness coefficients of plain weekly index and constituent returns from 1928 to 2011 and the following Figure 2.3 visualizes the sample skewness coefficients of standardized weekly return residuals obtained from estimating rolling $EGARCH(1,1,1)$-$MA(1)$-N models. Skewness patterns in plain and standardized weekly returns largely remain and appear to be driven by large outliers. The standard error of the standard sample skewness coefficient estimator depends on higher moments (Kendal, 1953) which might not exist for the DGP. Apparent are the long period of positive constituent skewness in return residuals between the 1950s and 1990s (1987 for plain returns), positive plain return skewness before 1941 with negative index return residuals. Since 2006, both plain and standardized weekly index return residuals are negatively skewed with the mean of constituents plain (standardized) weekly return skewness coefficients being centred around zero (negative).

2.4.3 Asymmetric dynamics in index and constituent return volatility

We investigate into the asymmetric return volatility dynamics by examining the asymmetry measures implied by the $GJR(1,1,1)$-$MA(1)$-N and $-t$ and $EGARCH(1,1,1)$-$MA(1)$-N model at both daily and weekly sampling frequency. The $EGARCH(1,1,1)$-$MA(1)$-t model suffers from parameter instability over time and appears to have many local maxima; we will consequently not present or report on these findings as findings are qualitatively albeit more noisy.[17] We will then further disentangle the asymmetry measures into the contribution from positive and negative return–related parameters and align these coefficients with their corresponding β coefficients and thus the persistency of the conditional variance.

The time series of asymmetry measure and parameter estimates of different models with N– or t–distributed residual densities at the daily and weekly level show very similar patterns with clearly visible common trends. With the GJR-t and $EGARCH$ model serving as robustness checks, we mostly defer those plots to the appendix.

Index and constituent asymmetry and impact asymmetry measures

Index. In the following we examine the time series of model implied asymmetry measures calculated based on Eq. (2.3.15) from the respective α and α^- coefficients of the GJR–

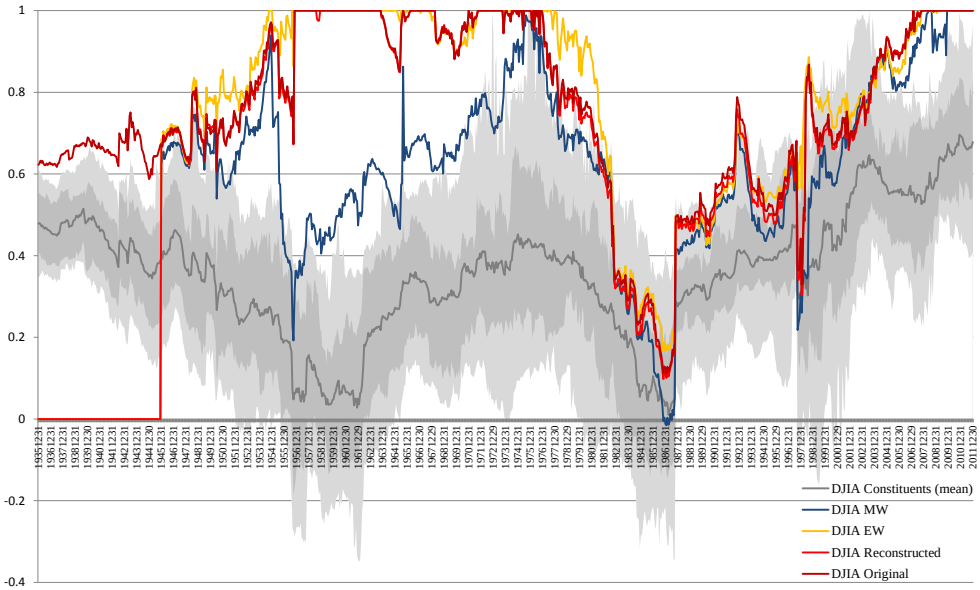
[17] Material available from the author upon request. Further note that the asymmetry measure for EGARCH–type models is not bounded.

N–distributed error structure model; with corresponding asymmetry measures for the GJR–t–distributed error structure (Fig. 2.B.6) and EGARCH model (Fig. 2.B.7) in the appendix.

Contrary to common believe, in Figure 2.6 we show that the large and increasingly more pronounced index asymmetry observed recently is NOT a new phenomenon. While the degree of asymmetry has been increasing over the last three decades, there were equally high asymmetric periods between the mid 1950s and mid 1970s.[18]

Figure 2.4: Time series dynamics of asymmetric return volatility measures; 1935–2011

The figure shows the time series of forward rolling asymmetry measures, Eq. 2.3.13, for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows based on daily data.



Within the next decade thereafter, until the mid 1980s the index asymmetry measure reverts from 1.0 (fully asymmetric) to 0.1 and the MV–weighted index even turns slightly below zero (which holds for the GJR-t model shown in the appendix) before the October 1987 crash, i.e. positive returns increase volatility more than negative ones–which might be interpreted as a fear indicator. When equity prices are too high relative to fundamentals, good news have the same impact on risk (volatility) as bad news.

The incorporation of the crash day pushes the asymmetry measure up to around 0.5 for the index (0.3 for the constituents) with a large spike up to 0.8 for the index which

[18]For clarification, note that the mid 1950s to mid 1970s parameter series extract covers an observation period of daily data for the optimisation procedure ranging from the mid 1940s until the mid 1970s.

is when the crash is approximately in the middle of the rolling time window. The spike reverts back to almost half its magnitude over the course of the following year.

We further observe some parameter instability following a spike in 1992 which might reflect certain return realizations leaving or entering the observation windows. Overall, we observe a strong upward trend in index asymmetry since the late 1980s until 2003 when it reaches 1.0 and where it approximately remains. The GJR–t model in the appendix shows a less pronounced asymmetry measure around the crash event compared to the same model with normally distributed error density but with even stronger increase in asymmetry thereafter.

We further observe a much larger magnitude of negative skewness in GJR–t standardized residuals over the the same crash period and with the variance of residuals being larger than one. While the skewness of plain daily returns appears to be inconclusive over the last recent period, standardized asymmetric ARCH–type model based daily return residuals are negatively skewed for both aggregate market returns and constituent stocks; compare Figure 2.3 and which holds for both daily and weekly return residuals.

The time series of $EGARCH$ asymmetry measures, Figure 2.B.7, is very similar to that computed from the GJR–N model. However, the decrease in asymmetry starting around 1974 is more steady and not as abrupt. The increase is asymmetry is almost proportional to the GJR–n asymmetry but less extreme in magnitude, roughly 0.1 units less pronounced. Also note that while both other series are fully asymmetric, the EGARCH model reverts in its degree of asymmetry beginning from the end of 2010.

It is further interesting to note a less asymmetric return volatility pattern of the market value weighted return index versus other weighting schemes for data covering the 1930s to the early 1980s— which has not repeated itself to the same extend in later periods.

Constituents. Note that the constituent stocks are far less asymmetric compared to the level of asymmetry exhibited by aggregate market returns, especially during the mid 1940s to mid 1980s period with constituent asymmetry even decreasing during the 1945 to 1962 period with index asymmetry increasing.

In line with index asymmetry also constituent asymmetry exhibits a steep and steady upward trend since approximately the crash day in Oct. 1987 but not as steep an increase as with regard to index asymmetry. Especially starting in the mid 1990s index asymmetry just takes off with a widening gap to average constituent asymmetry measures.

For the GJR–N model the overall long–lasting increase in index asymmetry since the Oct. 1987 crash is less extreme in terms of the index as it already starts on a high level (recall that the n–asymmetry is almost twice that of the t–asymmetry measure when the crash is included) but only little changed for constituents.

Impact Asymmetry Measures As the proposed asymmetry measures only describe relative weight given to positive and negative realized return innovations, it is further interesting to set these in relation to the actual magnitude of coefficients, as described in Eq. (2.3.15). The thus computed impact asymmetry measures are somewhat smoother as opposed to the rougher plain asymmetry measure. While GJR–N and $EGARCH$ model address, for instance, the 1987 crash by an increase in the magnitude in their ARCH parameters, the GJR–t model appears to absorb the shock differently with the impact asymmetry measure being almost symmetrical. By the end of 2011 the magnitude of GJR–N and $-t$ impact asymmetry coincides. For brevity, we choose to not report the statistics but opt to exhibit the remaining parameters individually instead.

The inner workings of asymmetric ARCH–type models revealed

In the following section we elaborate upon the interaction of parameters over time. These aforementioned differences can be explained when examining the α^i and $\alpha^{i,-}$ coefficients exhibited in the following figures in the appendix. With regard to the GJR–N model, following the 1987 crash the $\alpha^{i,-}$ coefficients strongly increase from virtually zero to 0.08–0.10 with α^i coefficients also increasing but to a lesser degree from 0.03 to 0.04.

This is the way in which the models with normally distributed return residual densities addresses different volatility persistencies, which are most noticeable from our inference results following the 1987 crash; particularly apparent in terms of the β^i coefficients, persistency and variance half–life plots.[19] As the variance is not as persistent as before the crash, the average β^i coefficients drops from 0.95 to approx. 0.88 when the crash is included in the observation window.

Then α^i and $\alpha^{i,-}$ coefficients have to explain more of the variance as also the half–life of variance shocks is reduced to a low of approx. twelve days only.

After the crash leaves the observation window the α^i coefficient drops sharply as does the $\alpha^{i,-}$ coefficient both to approximately 0.02, however, within weeks the $\alpha^{i,-}$ coefficient

[19] Not shown for brevity. Available from the author upon request.

almost jumps back towards 0.12, reverts back to 0.07, from where it steadily increases ever since while the α^i coefficients basically steadily decrease in magnitude which is puzzling.

Figure 2.5: Time series dynamics of $\alpha^{\mathbf{i}}$ coefficients; 1935–2011

The figure shows the time series of forward rolling α^i coefficients for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows.

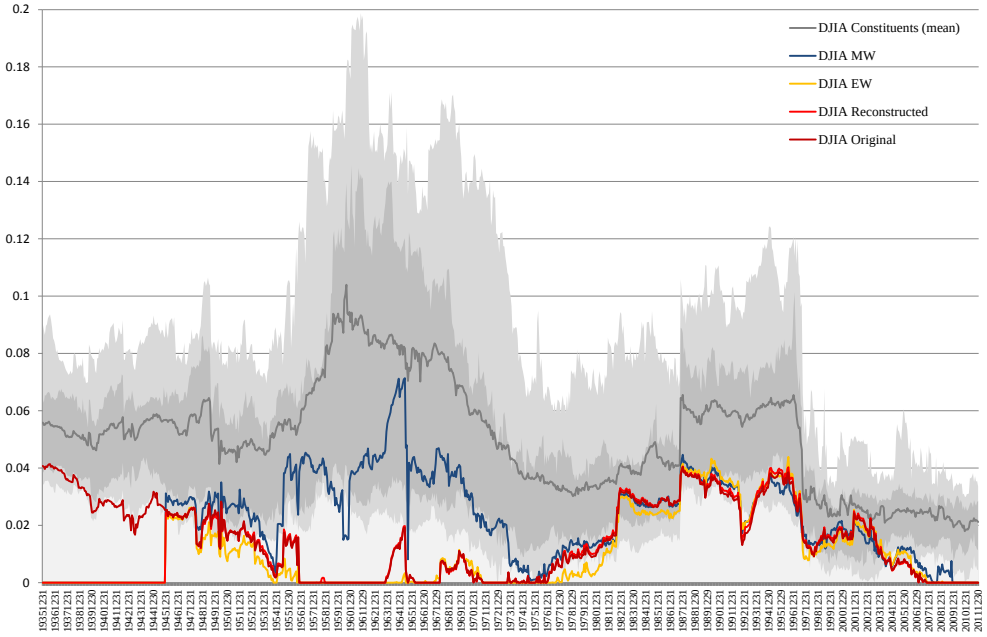
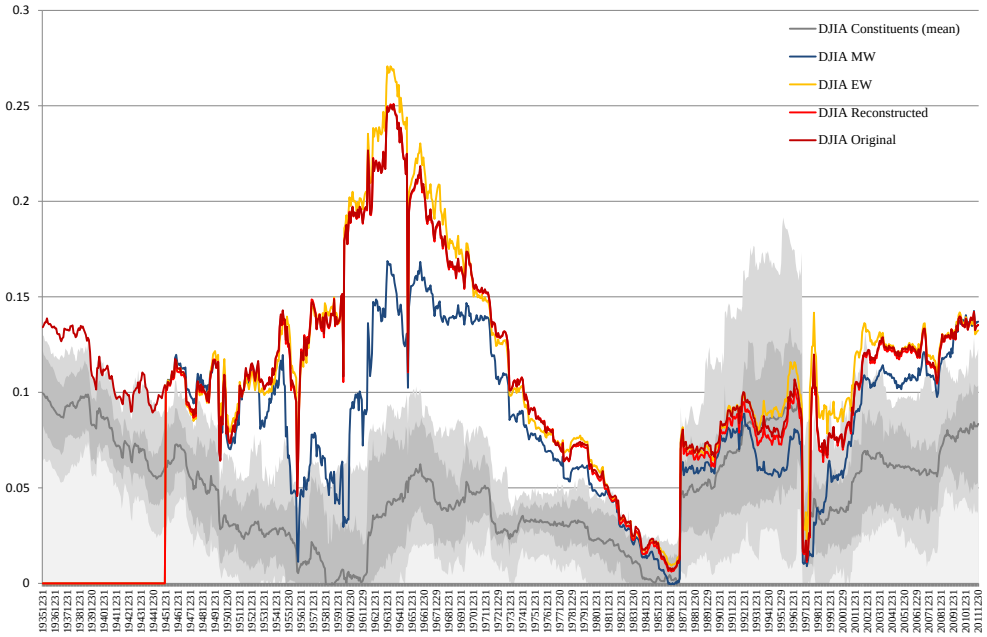


Figure 2.6: Time series dynamics of $\alpha^{\mathbf{i},-}$ coefficients; 1935–2011

The figure shows the time series of forward rolling $\alpha^{i,-}$ coefficients for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows.



In terms of constituents it is further interesting to notice that generally individual stock variance is highly sensitive to positive news and their α^i coefficients are always larger in magnitude than corresponding index α^i coefficients but the excess α^i of individual stocks versus index α^i seems to have decreased. Further, the $\alpha^{i,-}$ coefficients for the period when the 1987 crash is included are similar for index and constituents. Consequently, then only excess α^i and lower β^i coefficients of constituents drive the asymmetry over that period.

The aforementioned results are qualitatively the same with regard to the GJR–t model, see Figures 2.B.8 and 2.B.9 in the appendix, except for the fact that the asymmetry in recent years seems to be stronger in this case due to higher $\alpha^{i,-}$ coefficients and the impact of large shocks are not that straight forward as the degrees of freedom parameters ν^i for the t–distribution are allowed to vary across each time window. The β^i coefficients fluctuate only little around the value of 0.95 since the mid 1970s and the persistencies are roughly the same, close to 0.98 since 1998, for both t– and n–distributed errors for the index and constituents on average (not shown for brevity). The half–life for t–distributed errors is then also two to three times as high compared to the N–distributed error structure when the 1987 crash is included in the observation window while it is only slightly higher in the period thereafter (not shown for brevity).

The EGARCH model parameter decomposition, in specific the slope of the EGARCH news impact curve, shows it close relation to the GJR–N model. The corresponding EGARCH parameters (slopes) to α and α^- follow almost identical patterns except that the positive and negative ARCH–type parameters are proportionally higher in magnitude—in an unconstrained EGARCH model. But it is not straightforward due to the different functional form of the conditional variance process. The constituent vs. index deviation is mainly driven by the ν^i parameter, see Figure 2.B.10 in the appendix which comprehensively aggregates all information regarding the estimated EGARCH models for daily data.

To conduct a robustness check, we compare the asymmetry measures of the DJIA index which contains only large blue–chip stocks with those estimated for the S&P500, the Wilshire 5,000 and the Wilshire 4,500 index; only for the n–distributed error structure model for brevity. The S&P500 asymmetry measures are roughly in line with those for

the DJIA index except that the asymmetry is more pronounced for the S&P500 since approximately 2001. The asymmetry of the 'universe' of US stocks represented by the Wilshire 5,000 lies in between the two former ones. However, the asymmetry of the Wilshire 4,500 is roughly in line with the average large cap constituent asymmetry and only takes off in the last year (2011) to be fully asymmetric unlike constituents. This can be explained by the much higher α coefficients of the Wilshire 4,500. The persistency of the Wilshire 4,500 is relatively far below the estimates of the other three index series (approx. 0.02–0.04 units). Moreover it is unclear what happens to the persistency and α coefficients of the Wilshire 4,500 in 1998; see Figure 2.B.11 and 2.B.12 in the appendix. Overall, the drastic increase in asymmetry measures holds for even the Wilshire 4,500 comprising the US universe of mid and small cap stocks.

Asymmetric effects at the weekly horizon. A comparison.

Asymmetric effects appear to be much stronger at the weekly horizon with asymmetry measures being much pronounced for all models considered, see exemplarily Figure 2.7 below for the GJR-N model. Contrary to estimates based on daily data, the GJR–t model now also exhibits a sharply pronounced asymmetry ratio when the large 1987 crash gets included in the rolling observation period, see Figure 2.B.13 in the appendix.

Also for weekly we can confirm that α^i coefficients are usually higher for the firms than for the index, $\alpha^{i,-}$ are usually higher for the index than for firms, thus the index has more asymmetry than the firms with asymmetry measures having increased recently for both the index and the median firm, see Figure 2.8.

As the defined asymmetry measure for the EGARCH model is not bounded, recall the slope of the positive news impact curve can be negative, the corresponding figures for EGARCH model appear very rough with parameters bouncing up and down a lot in particular before 1982 after which it is relative smooth again. The interpretation of the real impact of the individual EGARCH coefficients is less clear as ω^i can be large and negative in unconstrained optimization when we do not impose the variance stationarity condition which is often violated. In contrast to unconstrained GJR model estimations which almost never violate the stationarity conditions.

Figure 2.7: Time series dynamics of asymmetric return volatility measures based on weekly data; 1935–2011

The figure shows the time series of forward rolling asymmetry measures, Eq. 2.3.13, for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows based on weekly data.

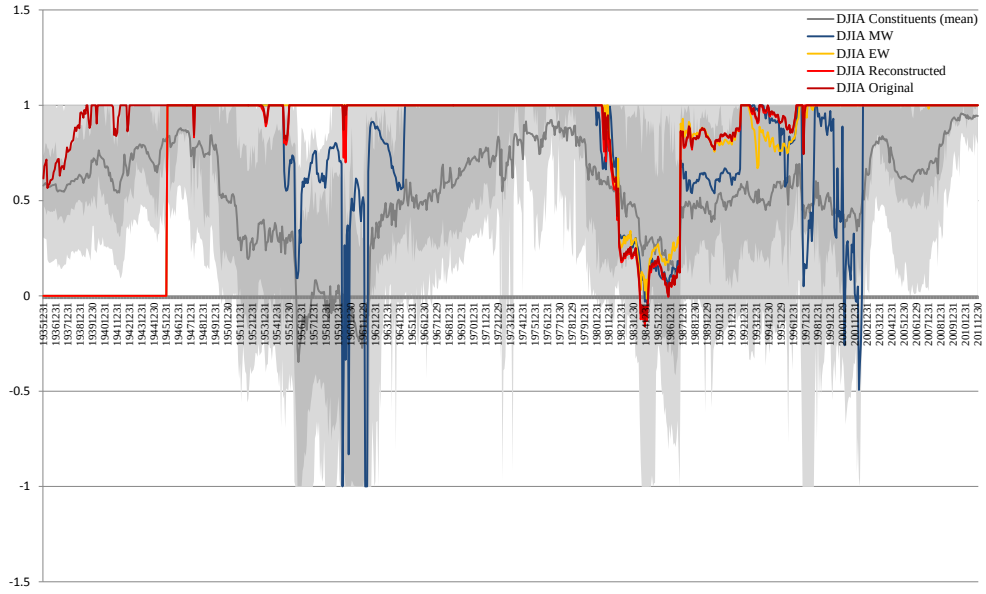
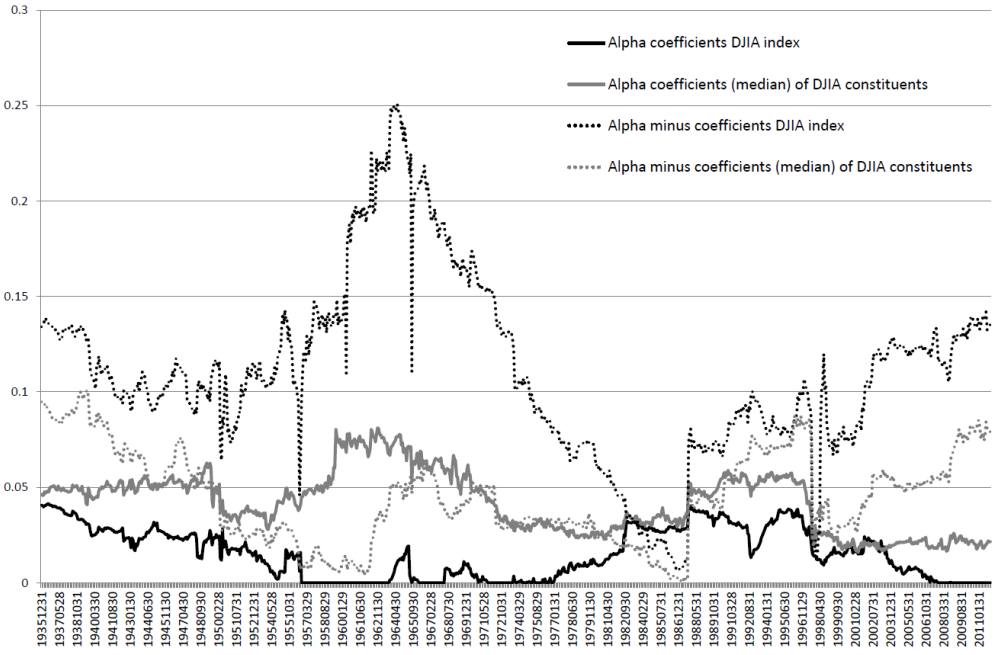


Figure 2.8: Time series dynamics of $\alpha^{\mathbf{i}}$ and $\alpha^{\mathbf{i},-}$ coefficients based on weekly data; 1935–2011

The figure shows the time series of forward rolling α^{i} and $\alpha^{i,-}$ coefficients for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows based on weekly data.



2.4.4 Analysing asymmetric correlations

Constituent return series are standardized following the procedure as outlined in Section 2.3.4. It is important to note that the following analysis is conditional upon $t-1$ information. We examine the threshold equicorrelation structure on the next day given a positive or negative innovation on the previous day.

In untabulated results we show that the 64 standardized constituent return series for both the GJR–N and GJR–t model exhibits negative albeit insignificant sample means which are more pronounced for the last ten stocks entering during or just before the financial crises. The average (median) variance of standardised residuals is with 0.987384 (0.991653) for the GJR–N model and 1.001124 (0.9999408) for the GJR–t model very close to one which benefits our correlation analysis. The difference in variances is driven by the higher variance of the last 20 and especially last ten new entrants for the GJR–t model and lower variance regarding the first ten entering stocks for the GJR–N model. Although there are empirical skewness deviations from symmetry for each standardized return series as elaborated upon earlier, the average (median) skewness is only 0.005 (0.027) and -0.028 (0.007) for the GJR–t model across time. The mean (median) excess kurtosis of standardized residuals is 1.68 (0.34) and as expected with 3.19 (1.92) slightly higher for the GJR–t model.

Note we refrain from contributing recent increases in index asymmetry measures to increases in the magnitude of negative skewness in standardized constituent return residuals as there are equally high aggregate market asymmetric return volatility periods throughout the 1950s to 1980s during which stocks exhibited positive return residual skewness; compare Figure 2.6 2.3.

In Table 2.A.5 in the appendix we present unconditional equicorrelations which are the same when returns are standardized using either the GJR–N or GJR–t model at approx. (0.306). Note this is the mean estimate over the entire period from 1928–2011. In terms of the correlation analysis conditional upon information at $t-1$, down–correlations in the range of 0.325 to 0.343 are between 0.043 to 0.069 units higher than up–correlation which are in the range of 0.275 to 0.284 depending on which information set is used. This corresponds to an average increase in down– vs. upside–correlation of 20.8%; see Table 2.A.5.

The observed relative spread in up– and down–correlations widens when the threshold level is raised and is more pronounced when returns are standardized by the GJR–t model estimated volatility series. More importantly, upside correlations remain approximately unchanged in the range of 0.267 to 0.278 independent from the threshold level while downside correlations increase to up to 0.413, i.e. the increase in the correlation spread stems from higher downside correlation only and upside correlations remain unchanged.

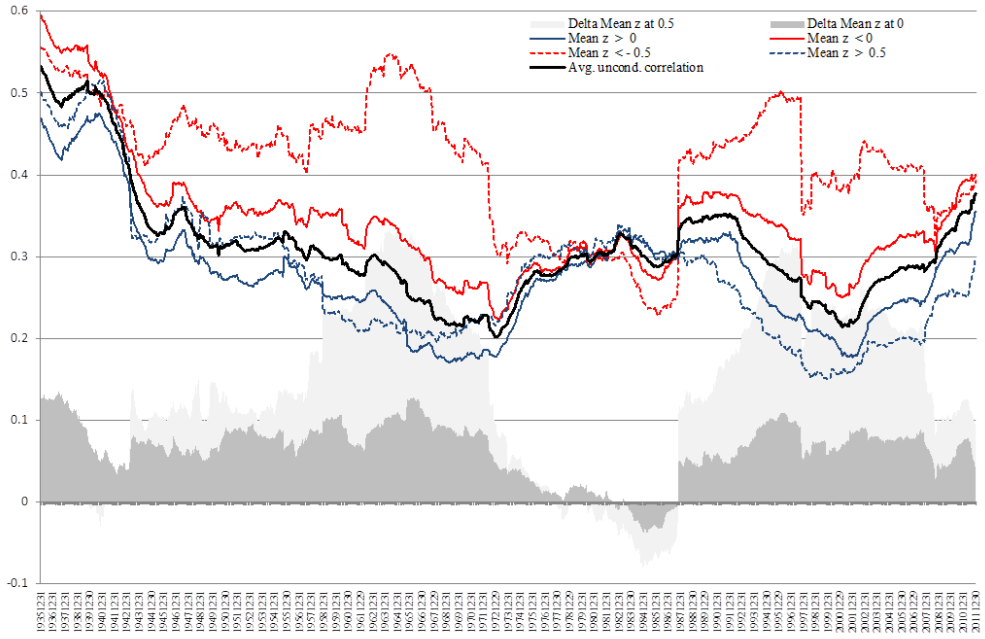
Also note that the average correlation among large cap stocks given a large positive innovation in the right tail of the standardized market residual density is lower than it is for positive market innovations of medium magnitude. However, for left tail events in the standardized market return distribution the average correlation among large cap stocks is strictly increasing towards unity. Loosely speaking, it appears that when the market is up by a large magnitude some stocks do not anticipate the increase in the market and firm specific news outweigh. However, when the market goes down stocks follow regardless of idiosyncratic news. More specifically, Table 2.A.5 also reveals that there is no apparent effect when condition on the median firm residual being either positive or negative unlike for the mean firm residual. This finding in line with Albuquerque (2012) who shows in a theoretical model that cross–sectional heterogeneity in firm announcement events can lead to conditional asymmetric stock return correlations and negative skewness in aggregate returns. Note that the negative skewness in aggregate market returns is consistent with our asymmetric return volatility story as skewness builds up over time based on the model dynamics.

Next we employ the one month forward rolling ten year observation period methodology to gain insights into the time series dynamics of the asymmetric threshold correlation analysis.

It is clearly visible that during the period 1976 until 1986 threshold equicorrelations are merely symmetric, i.e. an asymmetric term was not needed, and immediately before the 1987 crash downside correlations were less strong than upward correlations. While especially large downside correlations ($z < -0.5$) surge after the 1987 crash and still increase further when the crash event moves towards the beginning of the rolling observation window, all other correlation estimates steadily decline after initial (crash) increase. Until approx. the year 2000 correlations generally declined further with the exception being the higher downside threshold correlation and then steadily and monotonically increase

Figure 2.9: Threshold constituent equicorrelations conditional on I_{t-1} based on daily standardized residuals; 1935–2011

The figure shows the time series of forward rolling threshold constituent equicorrelations conditional on I_{t-1} based on daily standardized residuals; Eq. 2.3.13, for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows based on weekly data.



by roughly 0.15 units over the next decade. However, as the time series figures show correlations have been much stronger in the past starting from around 0.5 in 1935 and reverting towards approx. 0.2 in 1972. GJR–N and GJR–t and $EGARCH$ (not shown here for brevity) correlations are almost identical except for the fact that the GJR–t model produces a slightly higher (lower) downside (upside) correlation.

Figure 2.B.14 in the appendix shows the corresponding correlation estimates along with differences in correlation due to model choice (which are very small, note the scale) for the weekly return horizon with roughly very similar patterns.

Overall, correlations are much higher, however, while there is a large decrease in average correlations following 1935, there is no clear trend in the period 1956 until 1986. Noteworthy is the surge in upside threshold correlation ($z > 0.5$) after 2008 and the last recent burst in downside correlation at the end of 2011.

In further research, we would like to develop tests of the asymmetric volatility effect uncovered here. These could be of the form

$$h^*_{m,t} = 1/900 \sum \left[h_{i,t} + \rho (h_{i,t} h_{j,t})^{1/2} \right], \tag{2.4.4}$$

where we estimate

$$r_{m,t} = \left(h^*_{m,t} \right)^{1/2} z_{m,t}, \tag{2.4.5}$$

and then test ρ^- for significance when we replace ρ by ρ_t with

$$\rho_t = \rho + \rho^- S_{t-1} \tag{2.4.6}$$

where S denotes the lagged negativity indicator function.

2.5 Concluding remarks

There is substantial parameter variation when daily ARCH–type models are rolled over by as little as one month of data, i.e. 21 daily returns, even when the time frame is ten years along, i.e. 2,600 daily returns. In spite of the fact that stock market indices can be calculated based on different methodologies (market value, price or equally weighted) resulting in huge performance differences, ARCH–type model parameters vary together and the co–movements are particularly strong between the latter two index calculation methodologies. With the exception of a short period when the 1987 crash is included in the rolling series of ten year observation windows, index returns are always substantially more

asymmetric than their constituents. The asymmetry appears to be so strong in magnitude that it cannot be reconciled when constituent parameters are weighted or averaged in one way or another giving rise to the many asymmetry explanations put forward in the literature review.

In estimating the variance series $GJR(1,1,1)$–$MA(1)$–N models appears to be more stable in parameter estimates than the GJR–$GARCH(1,1)$–$MA(1)$–t model. $EGARCH$ models are even more unstable in terms of coefficient estimates but the inferred news impact slopes are very close to the GJR–N model. We neglect the $EGARCH$-t model as in unconstrained optimization procedures the parameter boundary conditions are often violated and several very different local maxima exist. While the 1987 crash catches the reader's eyes immediately, the inner workings of the model in terms of its parameter interaction has been disentangled and documented. We attached a great deal of attention to the high degree of asymmetry during the last decade—in what we refer to as fully asymmetric when our measure equals exactly one and all updating the variance process comes from past negative return residuals only. It seems to have been the continuing trend of an almost linear increase in asymmetry which was built up over the last two decades. Blair, Poon, and Taylor (2002) show that the minimum negative correlation between return and volatility innovation in the GJR GARCH model is bounded at around 0.71 which might explain the limitations of the model when correlations are indeed higher. Asymmetries in conditional volatility carry over to the weekly return horizon and are then even more pronounced.

In terms of correlations we suggest that correlations the next day are higher following a large negative shock on the previous day while they remain unchanged given a large positive shock, which gives rise to asymmetric correlations. This suggest that the strong index return–volatility asymmetry is attributable to the asymmetric return correlation of constituents. Time–varying threshold correlations obtained from rolling observation periods show that there is a strong upward trend in correlations regardless of the return horizon or the model choice.

Overall, we document that asymmetric effects are found in U.S. indices throughout the 1900s and early 21st century. The magnitude of the effect is time–varying and it has become stronger in recent years. The effects are weaker for individual firms than for indices. Correlations between firm returns increase when the market falls, which explains the higher index asymmetry. An older opinion is that debt/equity ratios increase in a

falling market and hence equity is more risky and so more volatile.

Appendix

2.A Tables for Chapter 2

Table 2.A.1: Benchmark Indices

The following table lists alternative benchmarks incorporated as robustness check.

#	Index Series	1st Available	Last available	Days
71	DJIA Original return series	01/10/1928	31/12/2011	30406
72	DJIA Reconstructed return series	02/01/1936	31/12/2011	27757
73	DJIA MV return series	02/01/1936	31/12/2011	27757
74	DJIA EW return series	02/01/1936	31/12/2011	27757
76	S&P500 return series	02/01/1964	31/12/2011	17530
78	Wilshire5000 return series	03/12/1979	31/12/2011	11716
80	Wilshire 4500 return series	02/02/1987	31/12/2011	9098

Table 2.A.2: DJIA constituent changes 1928–1936

The following table summarizes the Dow Jones Industrial Average constituent stock changes between 1928 and 1936. The leave/last day return is included in our analysis to mitigate the effects of survivor-ship bias in a broader sense.

#	Company Name	Enter	Leave/Last	Days
1	American Sugar Refining Company	01/10/1928	17/07/1930	654
2	American Tobacco Brands	01/10/1928	17/07/1930	654
3	Atlantic Refining	01/10/1928	17/07/1930	654
4	General Railway Signal	01/10/1928	17/07/1930	654
5	Goodrich	01/10/1928	17/07/1930	654
6	Mack Trucks Inc.	01/10/1928	25/05/1932	1332
7	International Shoe	26/05/1932	14/08/1933	445
8	North American	01/10/1928	28/01/1930	484
9	Paramount Publix	01/10/1928	25/05/1932	1332
10	Radio Corporation	01/10/1928	25/05/1932	1332
11	Texas Gulf Sulphur	01/10/1928	25/05/1932	1332
12	Victor Talking Machine	01/10/1928	07/01/1929	98
13	National Cash Register	08/01/1929	25/05/1932	1233
14	Wright Aeronautical	01/10/1928	13/09/1929	347
15	Curtiss-Wright	14/09/1929	17/07/1930	306
16	Borden	18/07/1930	19/11/1935	1950
17	Liggett & Myers	18/07/1930	25/05/1932	677
18	United Air Transport	18/07/1930	25/05/1932	677
19	United Aircraft	15/08/1933	12/08/1934	362
20	Hudson Motor	18/07/1930	25/05/1932	677
21	Drug Incorporated	26/05/1932	14/08/1933	445
22	Nash Motors	01/10/1928	17/07/1930	654
23	Coca-Cola Company	26/05/1932	19/11/1935	1272

Table 2.A.3: DJIA constituents 1936–2011

The following table continues to list the DJIA constituent changes from Table 2.A.2 where we keep track of name changes, ticker changes, mergers and acquisitions.

#	Company Name	Enter	Leave/Last	Days
1	Allied Chemical (Allied Corp)	01/10/1928	18/02/2008	28994
	Allied -Signal Incorp. (on 19/09/1985)			
	AlliedSignal Incorp. (on 26/04/1993)			
	Honeywell International Incorp. (on 02/12/1999)			
2	American Can	01/10/1928	05/05/1991	22861
	Primerica Corp (on 29/04/1987)			
	Commercial Credit Group Inc. (merger on 16/12/1988)			
3	American Smelting and Refining Com	01/10/1928	31/05/1959	11199
4	American Tobacco Brands Inc.	26/05/1932	29/10/1985	19514
5	Standard Oil of California	18/07/1930	31/10/1999	25307
	Chevron (on 02/07/1984)			
6	Chevron Corporation	19/02/2008	30/12/2011	1410
7	Bethlehem Steel	01/10/1928	16/03/1997	25003
8	Chrysler	01/10/1928	28/06/1979	18532
9	General Electric Company	01/10/1928	30/12/2011	30405
10	General Motors Company	01/10/1928	07/06/2009	29469
11	Goodyear	18/07/1930	31/10/1999	25307
12	International Harvester	01/10/1928	05/05/1991	22861
	Navistar International Corp. (on 20/02/1986)			
13	International Nickel	01/10/1928	11/03/1987	21345
	Inco (on 21/04/1976)			
14	John-Manville (Manville Corp.)	29/01/1930	29/08/1982	19205
15	Postum Incorporated	01/10/1928	29/10/1985	20847
	General Foods (on 24/07/1929)			
16	Sears Roebuck & Company	01/10/1928	31/10/1999	25962
17	Standard Oil (N.J.)	01/10/1928	30/12/2011	30405
	Exxon (on 01/11/1972)			
	Exxon Mobil Company (on 01/12/1999)			
18	Texas Company	01/10/1928	16/03/1997	25003
	Texaco Incorporated (on 22/04/1959)			
19	Union Carbide	01/10/1928	31/10/1999	25962
20	US Steel	01/10/1928	05/05/1991	22861
	USX Corp. (on 08/07/1986)			
21	Westinghouse Electric	01/10/1928	16/03/1997	25003
22	Woolworth	01/10/1928	16/03/1997	25003
23	Eastman Kodak Company	18/07/1930	07/04/2004	26927
24	United Aircraft	14/03/1939	30/12/2011	26589
	United Technologies Corp (on 01/05/1975)			
25	Procter & Gamble Company	26/05/1932	30/12/2011	29072
26	Loew's	26/05/1932	02/07/1956	8803
27	Nash Motors	26/05/1932	13/03/1939	2482
	Nash Kelvinator (on 08/01/1937)			
28	International Business Machines	26/05/1932	13/03/1939	2482
29	International Business Machines	29/06/1979	30/12/2011	11872
30	Coca-Cola Company	12/03/1987	30/12/2011	9059
31	Corn Products Refining	15/08/1933	31/05/1959	9420
32	National Distillers	13/08/1934	31/05/1959	9057
33	Du Pont	20/11/1935	30/12/2011	27799
34	National Steel	20/11/1935	31/05/1959	8593

Table 2.A.4: DJIA constituents 1936–2011 (cont.'d)

The following table continues to list the DJIA constituent changes from Table 2.A.3 where we keep track of name changes, ticker changes, mergers and acquisitions.

#	Company Name	Enter	Leave/Last	Days
35	American Telephone & Telegraph AT&T (on 04/01/1984) AT&T Corporation (on 20/04/1994)	14/03/1939	07/04/2004	23766
36	SBC Communications AT&T Incorporated (merger on 21/11/2005)	01/11/1999	30/12/2011	4442
37	International Paper	03/07/1956	07/04/2004	17445
38	Anaconda Copper	01/06/1959	08/08/1976	6278
39	Swift & Company Esmark (on 30/05/1973)	01/06/1959	28/06/1979	7332
40	Aluminum Company of America Alcoa Incorp. (on 04/01/1999)	01/06/1959	30/12/2011	19205
41	Owens-Illinois Glass	01/06/1959	11/03/1987	10145
42	Minnesota Mining & Manufacturing 3M Company (on 08/04/2002)	09/08/1976	30/12/2011	12926
43	Merck	29/06/1979	30/12/2011	11872
44	American Express Company	30/08/1982	30/12/2011	10714
45	Philip Morris Companies Altria Group, Inc. (on 27/01/2003)	30/10/1985	18/02/2008	8146
46	McDonald's Corporation	30/10/1985	30/12/2011	9557
47	Boeing Company	12/03/1987	30/12/2011	9059
48	Caterpillar Inc.	06/05/1991	30/12/2011	7543
49	Walt Disney Company	06/05/1991	30/12/2011	7543
50	J.P. Morgan & Company J.P. Morgan Chase (merger on 02/01/2001)	06/05/1991	30/12/2011	7543
51	Travelers Group Citigroup Inc. (on 08/10/1998)	17/03/1997	07/06/2009	4465
52	The Travelers Companies, Inc.	08/06/2009	30/12/2011	935
53	Hewlett-Packard Company	17/03/1997	30/12/2011	5401
54	Johnson & Johnson	17/03/1997	30/12/2011	5401
55	Wal-Mart Stores Incorp.	17/03/1997	30/12/2011	5401
56	Microsoft Corporation	01/11/1999	30/12/2011	4442
57	Intel Corporation	01/11/1999	30/12/2011	4442
58	Home Depot Incorporated	01/11/1999	30/12/2011	4442
59	American International Group Incorporated	08/04/2004	21/09/2008	1627
60	Pfizer Incorporated	08/04/2004	30/12/2011	2822
61	Verizon Communications Incorporated	08/04/2004	30/12/2011	2822
62	Bank of America Corporation	19/02/2008	30/12/2011	1410
63	Kraft Foods Inc.	22/09/2008	30/12/2011	1194
64	Cisco Systems, Inc.	08/06/2009	30/12/2011	935

Table 2.A.5: Threshold constituent equicorrelations conditional on I_{t-1} based on daily standardized residuals, 1928-2011

Correlation I z/r_t-1	**GJR-N**					**GJR-T**				
Unconditional	**0.306**					**0.307**				
Conditional	Mean z	Median z	DJIA z	DJIA r	DJIA MV r	Mean z	Median z	DJIA z	DJIA r	DJIA MV r
z or r < 0	0.337	0.325	0.337	0.341	0.335	0.339	0.327	0.339	0.343	0.337
(n)	(9770)	(10835)	(9733)	(9355)	(9463)	(9774)	(10829)	(9733)	(9355)	(9463)
z or r > 0	0.276	0.284	0.276	0.275	0.280	0.276	0.284	0.276	0.275	0.280
(n)	(10034)	(8969)	(10071)	(10449)	(10341)	(10030)	(8975)	(10071)	(10449)	(10341)
Delta in z or r corr	**0.061**	**0.041**	**0.061**	**0.066**	**0.054**	**0.063**	**0.043**	**0.063**	**0.069**	**0.056**
Conditional Threshold	Mean z	Median z	DJIA z			Mean z	Median z	DJIA z		
z < 0.125	0.349	0.347	0.340			0.352	0.349	0.342		
(n)	(7708)	(7611)	(8547)			(7712)	(7610)	(8547)		
z > 0.125	0.274	0.278	0.275			0.272	0.278	0.275		
(n)	(7905)	(7343)	(8945)			(7910)	(7352)	(8945)		
Delta in z corr	**0.075**	**0.068**	**0.064**			**0.079**	**0.071**	**0.067**		
z < 0.25	0.368	0.368	0.344			0.370	0.371	0.347		
(n)	(5844)	(6055)	(7367)			(5845)	(6068)	(7367)		
z > 0.25	0.268	0.277	0.268			0.267	0.276	0.268		
(n)	(6003)	(5549)	(7732)			(6008)	(5562)	(7732)		
Delta in z corr	**0.100**	**0.091**	**0.076**			**0.104**	**0.095**	**0.079**		
z < 0.5	0.409	0.398	0.360			0.413	0.408	0.364		
(n)	(3247)	(3226)	(5410)			(3244)	(3222)	(5410)		
z > 0.5	0.274	0.278	0.275			0.272	0.278	0.275		
(n)	(3163)	(2832)	(5644)			(3166)	(2836)	(5644)		
Delta in z corr	**0.135**	**0.120**	**0.084**			**0.141**	**0.130**	**0.089**		

2.B Figures for Chapter 2

Graph Legend Explanation. The Original price weighted DJIA Index is shown as DARKRED, the price weighted reconstructed DJIA Index is shown as RED, the reconstructed market value weighted DJIA Index is shown as BLUE, and the reconstructed equally weighted DJIA Index is shown as ORANGE. The GREY lines and areas refer to underlying constituent stocks of the index at each point in time on a monthly rolling forward basis: The DARKGREY line refers to the average mean constituent value of the respective parameter or statistic, the GREY area around the mean refers to the inter–quartile area of the constituent parameter or statistic value and the LIGHTGREY, even wider area spans the area between the first and last decile respective value.

Figure 2.B.1: Reconciliation error: DJIA Original minus DJIA Reconstructed

The figure shows the reconciliation error between the original DJIA and our reconstructed index.

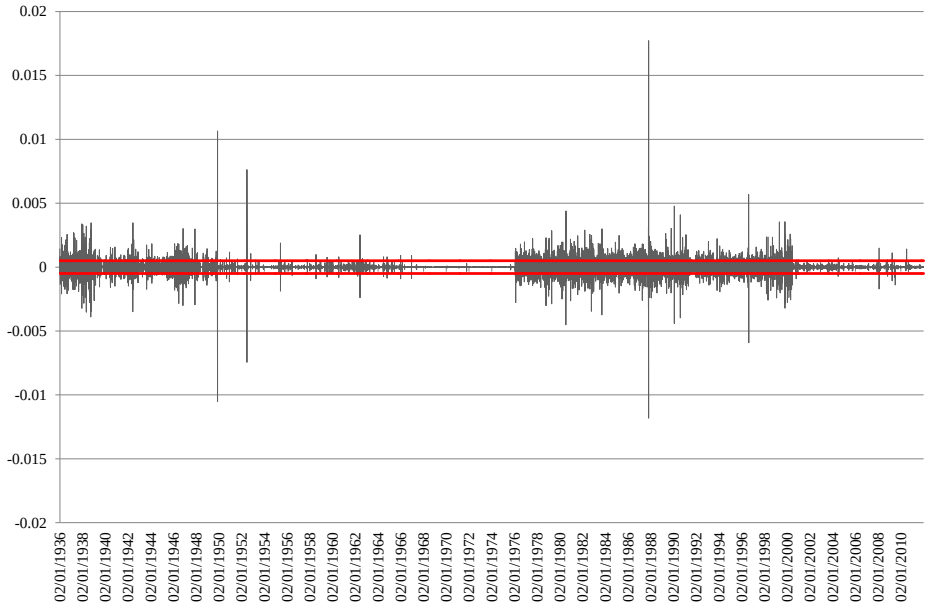


Figure 2.B.2: Index level comparison of DJIA Original vs. DJIA constructs

The figure shows the DJIA Original price history along with our constructs (equally weighed and market value weighted). We further show the Wilshire 4500, 5000 and the S&P500 index which are indexed as shown in the legend for comparative purposes.

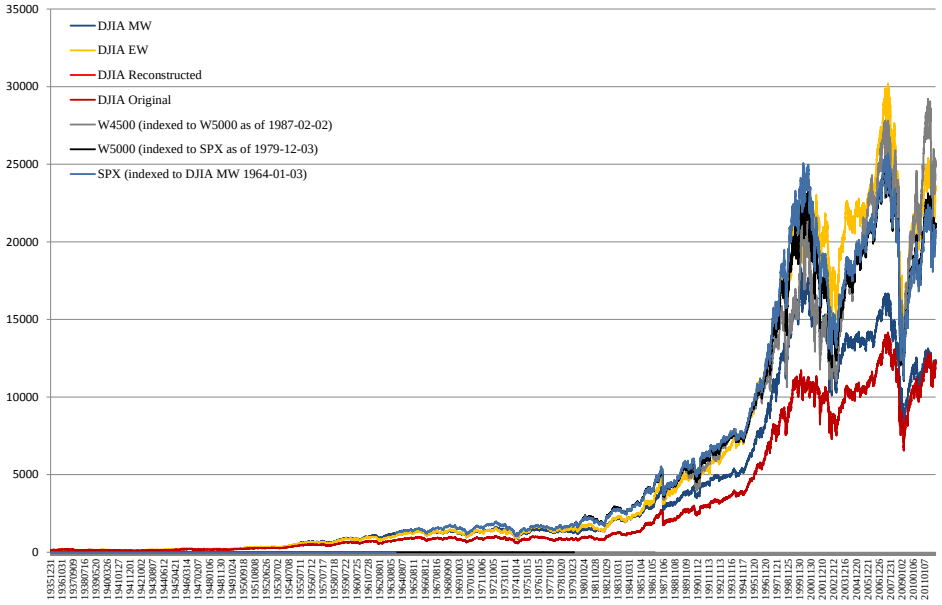


Figure 2.B.3: Index LOG level comparison of DJIA Original vs. DJIA constructs

The graphs shown corresponds to those in Figure 2.B.2 on a log scale.

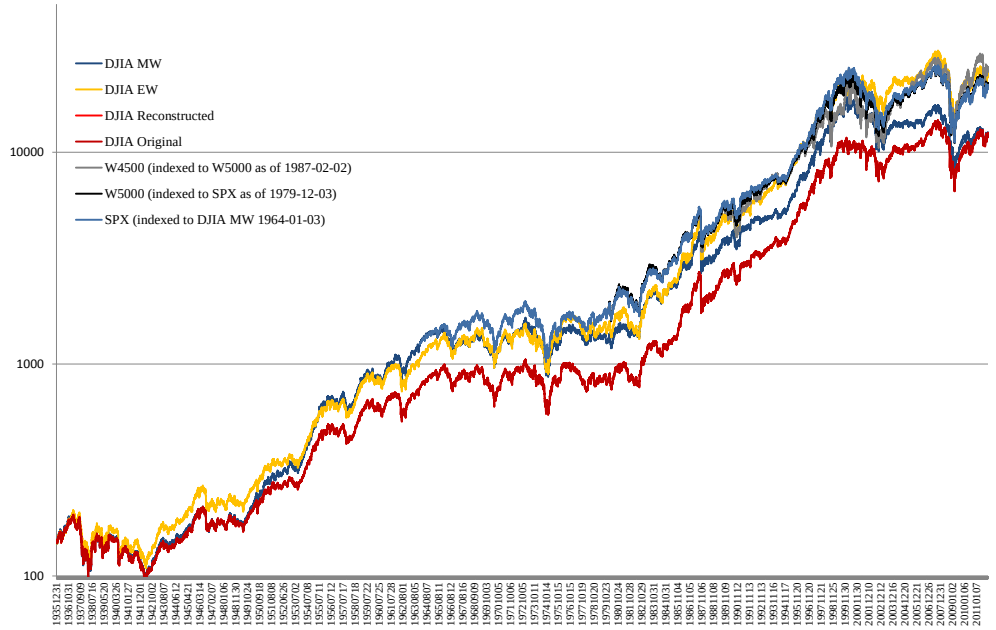


Figure 2.B.4: Weekly model based conditional volatility estimates; 1928–2011

Weekly conditional volatility estimates (p.a.) based on the $EGARCH(1,1,1)$-$MA(1)$-N model, where N denotes N–distributed error density. Estimates shown are obtained from estimating the model over one month forward rolling ten year time windows.

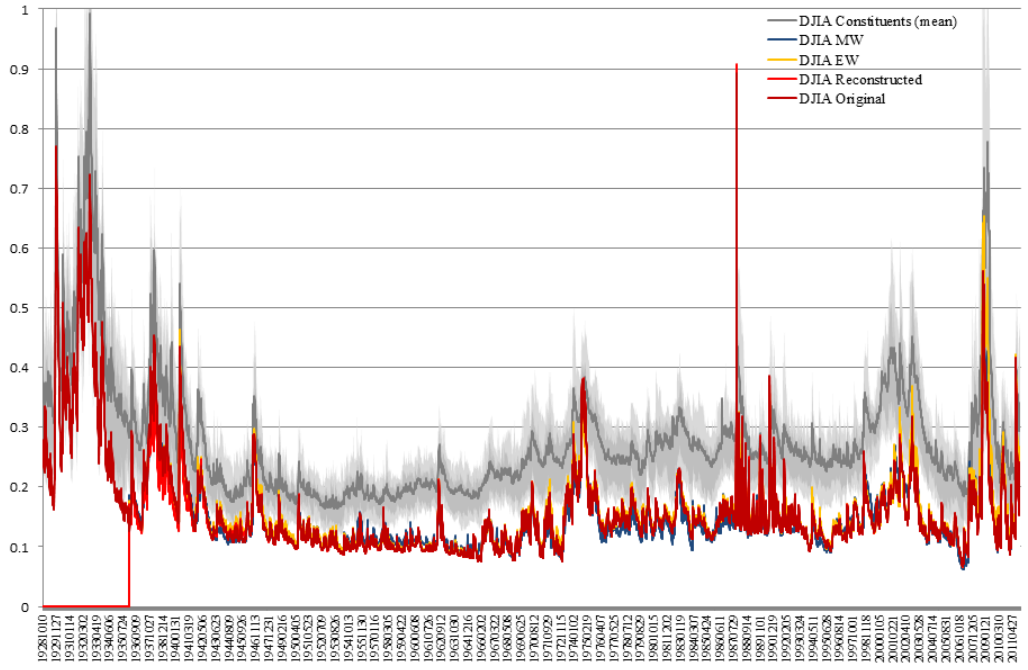


Figure 2.B.5: Unconditional GJR-t model implied volatility estimates; 1928–2011

Daily unconditional volatility estimates (p.a.) based on the $GJR(1,1,1)$-$MA(1)$-t model, where t denotes normally distributed error density. Estimates are functions of the model parameters which are estimated over one month forward rolling ten year time windows.

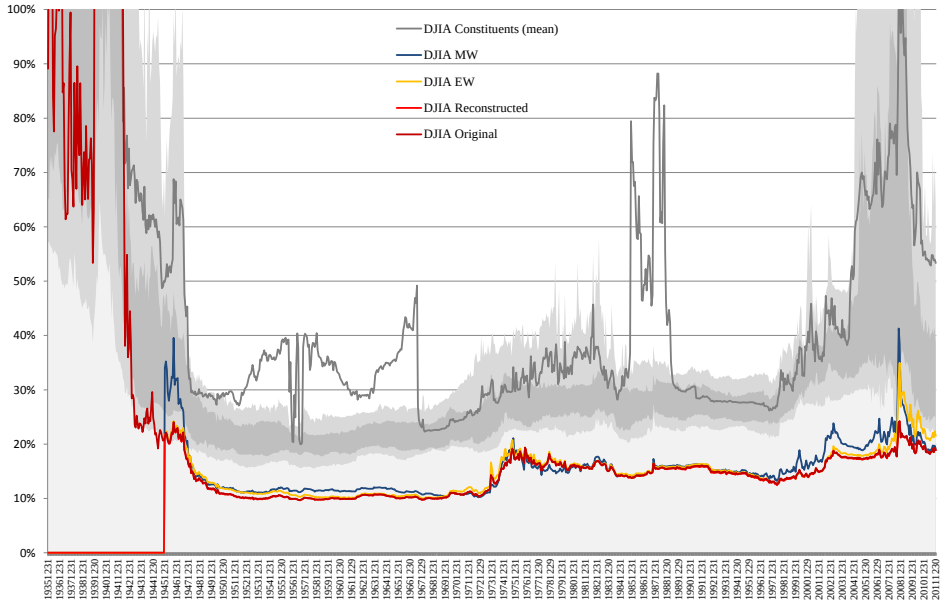


Figure 2.B.6: Time series dynamics of asymmetric return volatility measures (GJR-t); 1935–2011

The figure shows the time series of forward rolling asymmetry measures, Eq. 2.3.13, for the $GJR(1,1,1)$-$MA(1)$-t model, estimated over one month forward rolling ten year time windows.

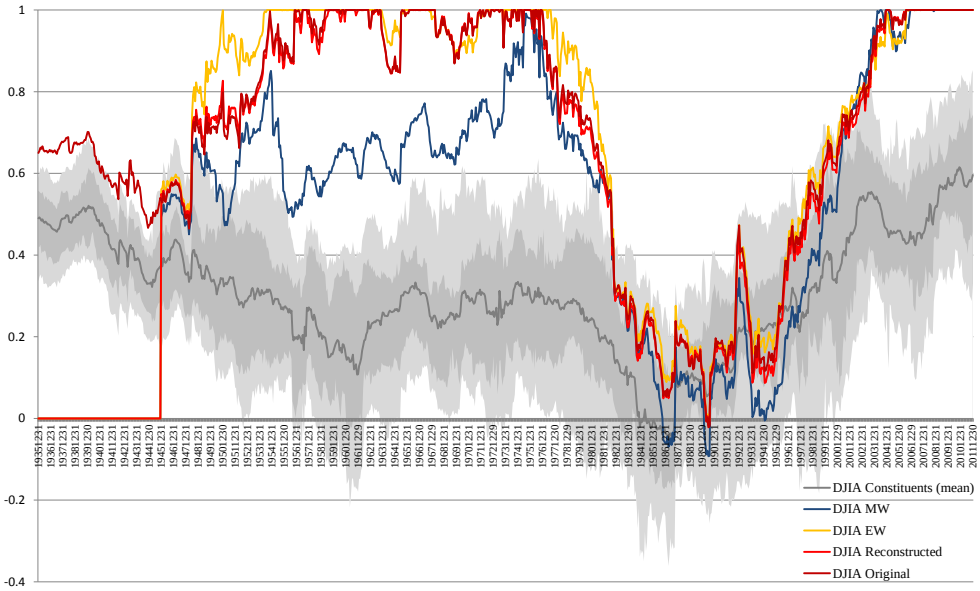


Figure 2.B.7: Time series dynamics of asymmetric return volatility measures (EGARCH); 1935–2011

The figure shows the time series of forward rolling asymmetry measures, Eq. 2.3.13, for the $EGARCH(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows.

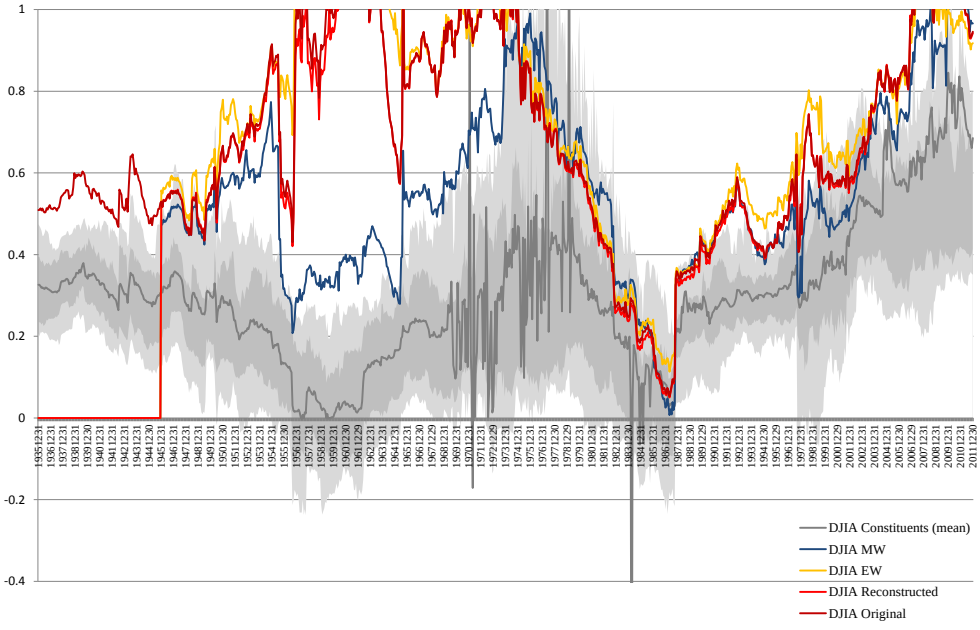


Figure 2.B.8: Time series dynamics of α^{i} coefficients (GJR-t); 1935–2011

The figure shows the time series of forward rolling asymmetry measures, Eq. (2.3.13), for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows.

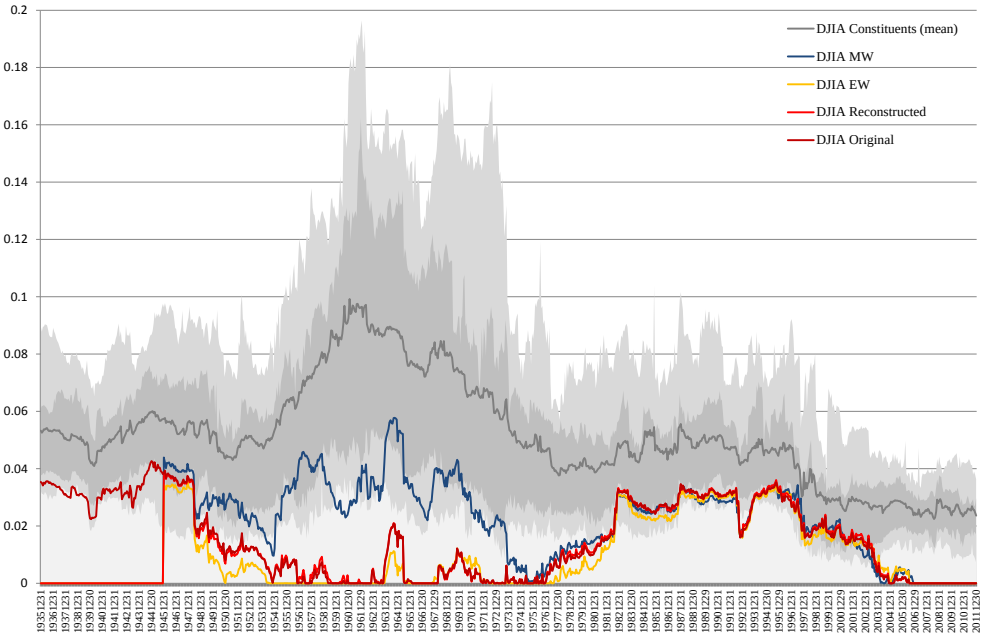


Figure 2.B.9: Time series dynamics of $\alpha^{\mathrm{i},-}$ coefficients (GJR-t); 1935–2011

The figure shows the time series of forward rolling asymmetry measures, Eq. (2.3.13), for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows.

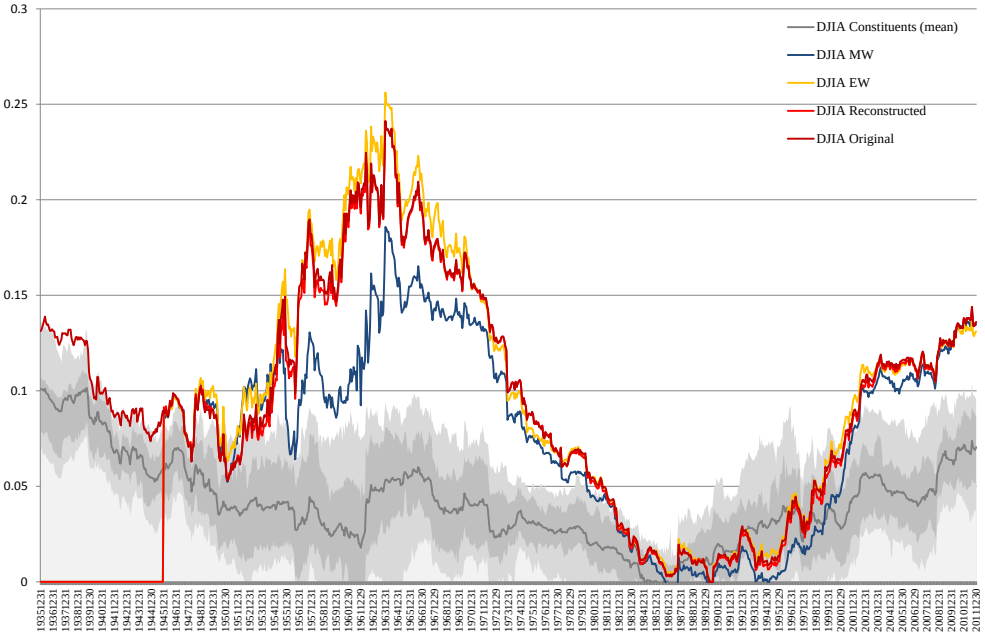


Figure 2.B.10: Time series dynamics and interaction of EGARCH parameters; 1935–2011

The figure reveals the time series dynamics of EGARCH parameters over time, estimated over one month forward rolling ten year time windows.

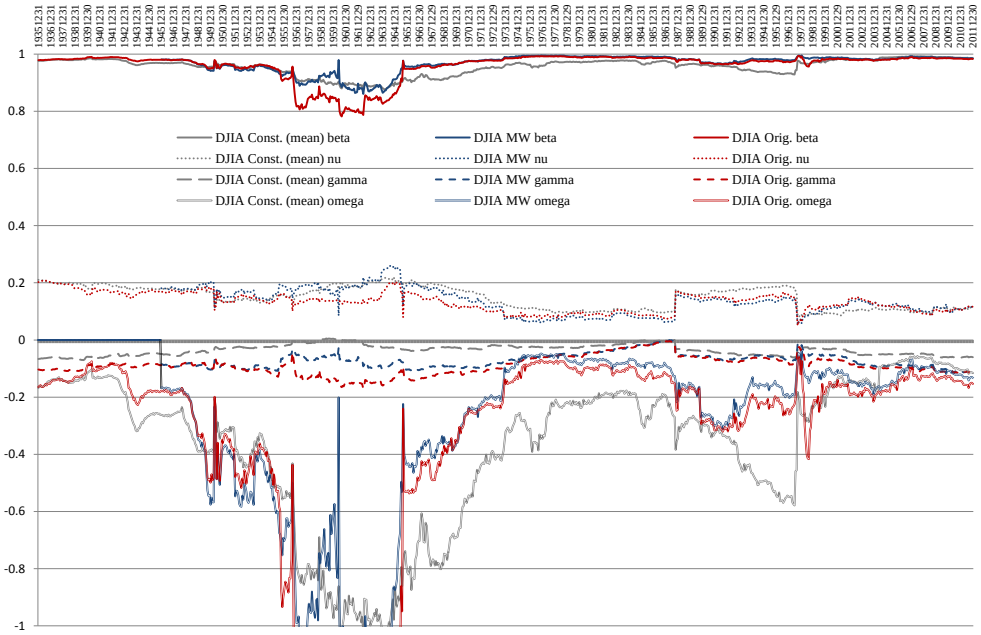


Figure 2.B.11: Time series dynamics of asymmetric return volatility measures of alternative benchmark indices; 1974–2011

The figure shows the time series of forward rolling asymmetry measures, Eq. (2.3.13), for the $GJR(1,1,1)$-$MA(1)$-n models, estimated over one month forward rolling ten year time windows.

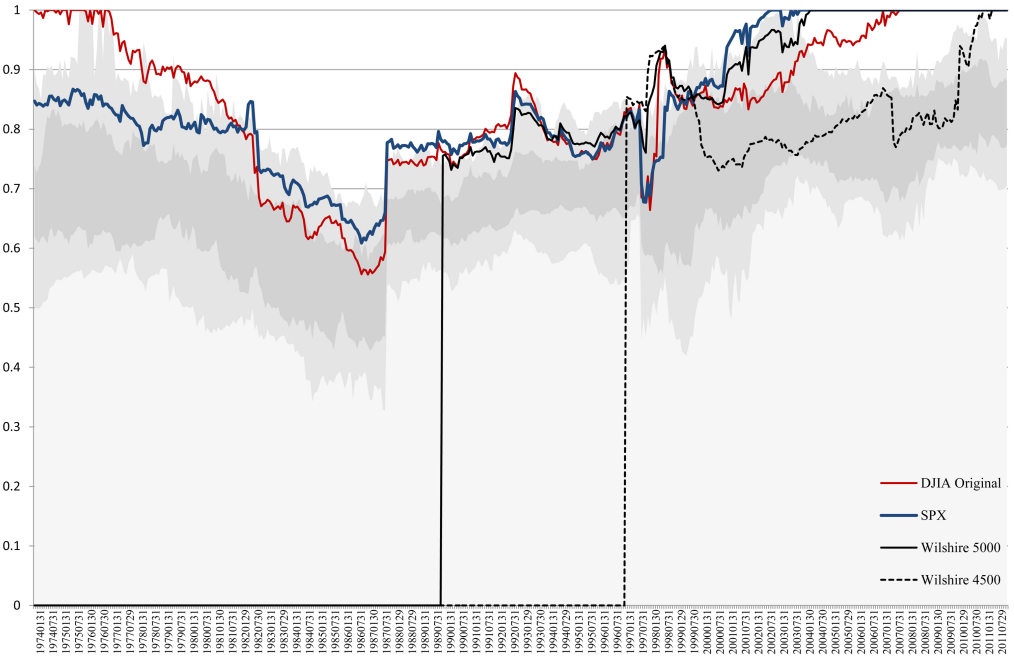


Figure 2.B.12: Time series of persistency estimates of alternative benchmark indices; 1974–2011

The figure shows the time series of forward rolling persistency estimates, Eq. (2.3.16), for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows.

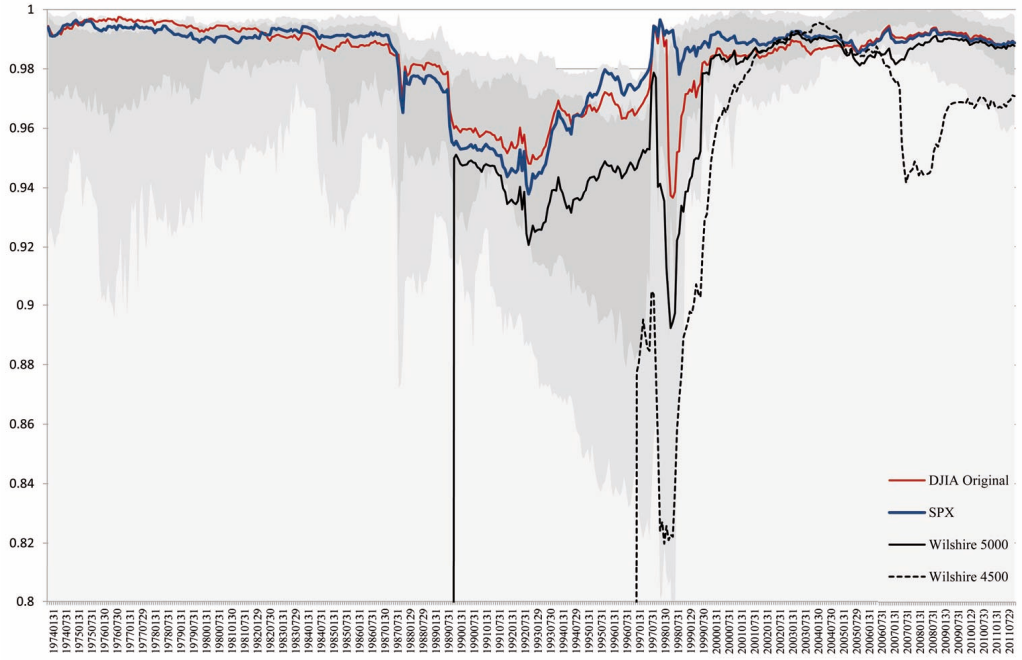


Figure 2.B.13: Time series dynamics of asymmetric return volatility measures based on weekly data (GJR-t); 1935–2011

The figure shows the time series of forward rolling asymmetry measures, Eq. (2.3.13), for the $GJR(1,1,1)$-$MA(1)$-t model, estimated over one month forward rolling ten year time windows based on weekly data.

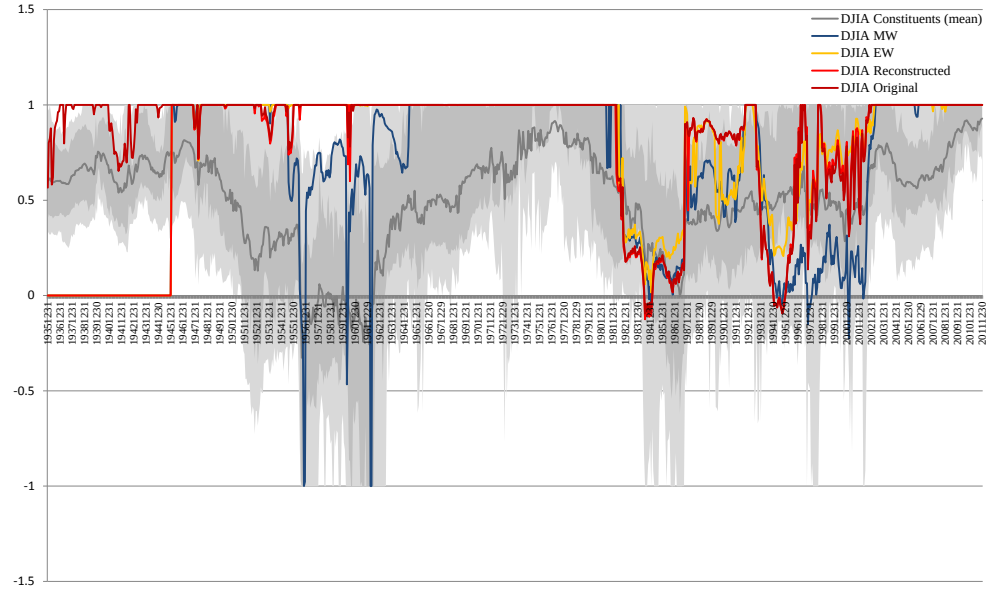
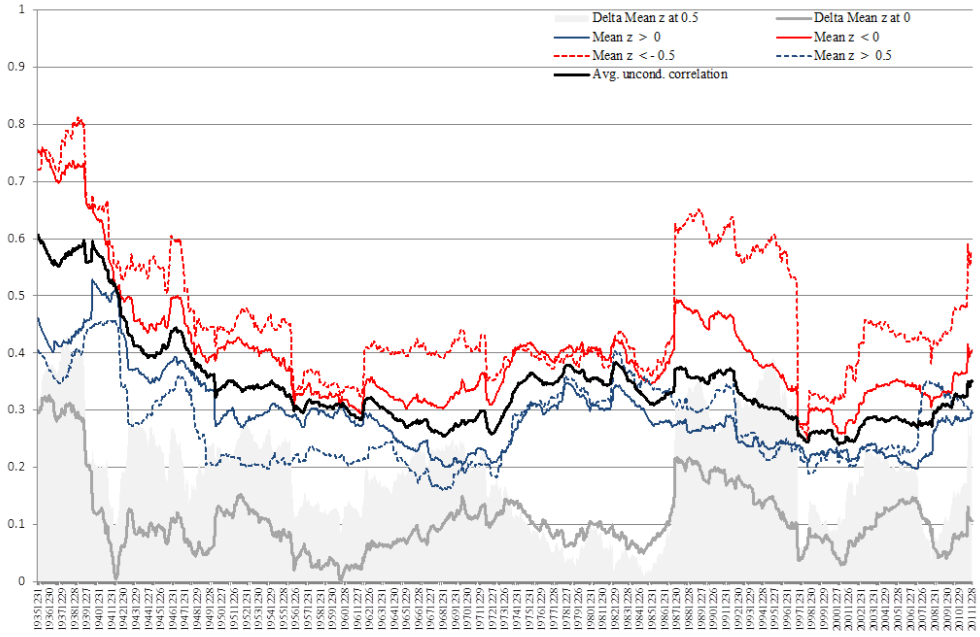


Figure 2.B.14: Threshold constituent equicorrelations conditional on I_{t-1} based on weekly standardized residuals; 1935–2011

The figure shows the time series of forward rolling threshold constituent equicorrelations conditional on I_{t-1} based on weekly standardized residuals; Eq. (2.3.13), for the $GJR(1,1,1)$-$MA(1)$-N model, estimated over one month forward rolling ten year time windows based on weekly data.



Chapter 3

The Role of Noise in the Estimation of Betas

Co–Author: Cisil Sarisoy[1]

Abstract

It has been an interest to use highly frequent data in estimating assets' sensitivities to systemic risks, β, in order to benefit from statistical precision gains. However, when the prices are contaminated by microstructure noise, we show that the estimates of the betas are biased and derive a closed–form expression for it. The estimates are biased towards zero, the magnitude of the bias rises with sampling more frequently. We conduct a simulation analysis to examine the properties of the estimates of the betas. We demonstrate that there is a trade-off caused by bias and the variance and lead to an optimal sampling frequency (in the RMSE sense) among a wide range of frequencies examined.

Keywords: (microstructure) noise, realized beta, MLE, LS

[1]Cisil Sarisoy is a post–doctoral fellow at the Department of Finance at Kellogg School of Management, Northwestern University, IL–Evanston, U.S.

3.1 Introduction

The linear discrete-time factor models stand as a corner-stone of the academic finance literature since the development of the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) and serve as the crucial parametric form in decisions regarding portfolio allocation and risk management. These linear factor models imply that the expected returns across assets should be linear in the loading(s) of the systematic risk factors, β. There is an enormous literature examining these relations.

The factor loading, β, of an asset with respect to a risk factor is of course not observable, and is to be estimated. With the availability of high–frequency data, the literature provided new econometric techniques to estimate the β of an asset, by employing the high frequency data. In specific, Andersen et al. (2005) and Barndorff-Nielsen and Shephard (2004a) explored the estimators of β based on high frequency data - which are inferred as realized betas ($R\beta$ henceforth). Moreover, Bollerslev and Zhang (2003) and Morana (2009) examines the cross–sectional pricing relations by inferring the factor loading via $R\beta$s.

The high frequency data is beneficial statistically: sampling the returns as often as possible would lead to higher precision in the estimates of beta. The rule is simple - use as much data as possible and exploit the information in the data as much as possible and gain from precision. However, there is a caveat: the observed prices are different than the true price process under which such benefits are documented typically - because of the microstructure noise contaminating the prices. Market microstructure noise is an umbrella term referring to the bid-ask bounces, the discretization errors or the liquidity issues. Although it may sound like a micro issue - maybe because of the name attached to it, we show that it leads to serious issues in the estimates of β based on high frequency data.

In this paper, we examine the implications of the market microstructure noise on the econometric inference about the betas. We start by asking - what happens to the estimates when the prices of the assets and the factor (the market henceforth) are both contaminated by market microstructure noise but ignored? We, first, show that there exists a bias in the estimates of β and provide a closed form formula for it when the market microstructure noise is present in the data but ignored. The bias is a multiple of the true β and related to the total market variance caused by the market microstructure

noise in the prices of market. The bias goes up with sampling returns more frequently: using very high frequency data leads to large biases in the estimates of β and the estimates are pushed towards zero.

On the opposite of this effect, sampling more frequently leads to higher precision in the estimates. This leads to the need for an answer to the question for an empirical researcher: which sampling frequency to choose to balance these two effects? The trade-off is clear intuitively: should we sample very coarsely, and save the estimates from the bias caused by microstructure noise with the cost of throwing away a large portion of data and loose the information in it?

We clearly document in a Monte Carlo study with 10,000 simulated paths for an asset and the market that there is a trade–off caused by these two effects, bias and the variance, in the presence of market microstructure noise which is not there when there is no market microstructure noise. By examining the issue for a wide range of sampling frequencies from 1min to 1 week (which are often used in the literature), we find that these effects counter balance each other lead to an optimal sampling frequency - with smallest RMSE among the alternative frequencies considered. Moreover, we illustrate that the sampling frequency yielding the smallest RMSE depends on the magnitude of the market microstructure contamination in the market variances.

With these results, we aim to emphasize a caveat: one should be careful in making econometric inferences about the betas using high frequency data. One may obtain economically very small estimates of the loadings β based on very high-frequency data although they may be large in reality. There may be a different optimal sampling frequency depending on how contaminated is the factor that is used by the market microstructure noise.

The rest of paper is organized as follows. Section 3.2 introduces our continuous time data generating process. In Section 3.2.1, we present the theoretical analyses regarding finite sample moments of the estimators when there is no microstructure noise. We, then, present in Section 3.2.2 our theoretical findings when the prices of both the asset and the market are contaminated by microstructure noise bu it is ignored when making econometric inference. A detailed simulation results are provided in Section 3.3.1. The final section is spared for the conclusions.

3.2 Model Specification

We start with a simple set up and assume that, the log–price of a security (a), that is $p_t^{(a)}$, and log–price of a proxy for the market (m) such as an exchange traded fund, $p_t^{(m)}$ have the following dynamics (on the time-interval $[0,T]$ and on some filtered probability space $(\Omega, F, (F)_{t\leq 0}, P)$)

$$dp_t^{(m)} = \mu\left(p_t^{(m)}, \theta\right) dt + \sigma^{(m)} dW_t^{(m)} \tag{3.2.1}$$

$$dp_t^{(a)} = \mu\left(dp_t^{(a)}, \theta\right) dt + \beta^{(a)} dp_t^{(m)} + \sigma^{(a)} dW_t^{(a)} \tag{3.2.2}$$

where $p_t^{(a)}(0) = 0$, and $p_t^{(m)} = 0$, $W_t^{(a)}$ and $W_t^{(m)}$ are standard Brownian Motions orthogonal to each other with time homogenous volatilty components $\sigma^{(a)} > 0$ and $\sigma^{(m)} > 0$. We assume that the loading on asset (a), i.e. $\beta^{(a)}$, which is associated with the market risk is constant over time. It has been widely documented in empirical studies that including the drift term leads to less precise estimates of variances and covariances based on high–frequency data. This is intuitive given the large standard errors in the estimates of drifts; see, e.g., Merton (1980). Accordingly, we simplify the setting by assuming $\mu(p_t^{(m)}, \theta) = 0$ and $\mu(dp_t^{(a)}, \theta) = 0$ for the sake of obtaining higher precision in our high frequency based estimates.

The log–prices of the securities and the market are assumed to be observed at discrete times $t_0 = 0, t_2, ..., t_N = T$, with equal distance in between, i.e. $t_i - t_{i-1} = \delta$ for all i. So, there are $N+1 = T/\delta$ observations over the period $[0,T]$. For example, if T stands for one month, then δ may stand for 5 minutes, 10 minutes or days over the course of the month.

Under the set up above, we analyze the role of microstructure noise in the estimation of $\beta^{(a)}$ based on equally spaced observations sampled from the continuous time process (3.2.1) and (3.2.2). We then precede as follows: For a fixed sampling frequency δ we analyse the finite sample properties of the estimators based on the sampled observations once in the absence of noise and then we examine the implications of the presence of the noise for the properties of the estimates of $\beta^{(a)}$.

3.2.1 What happens in the absence of noise?

We first examine the case under which there is no noise being present. Under our sampling scheme we fix δ and the observation times $t_0 = 0, t_2, ..., t_N = T$. Then, under the set up given in the previous section the returns on the market (m) and and the asset (a) implied by the continuous time model are as follows

$$\begin{aligned} R_{t_i}^{(m)} &= p_{t_i}^{(a)} - p_{t_{i-1}}^{(a)} \\ &= \sigma^{(m)}(W_{t_i}^{(m)} - W_{t_{i-1}}^{(m)}), \qquad (3.2.3) \\ R_{t_i}^{(a)} &= p_{t_i}^{(a)} - p_{t_{i-1}}^{(a)} \\ &= \beta^{(a)}\sigma^{(m)}(W_{t_i}^{(m)} - W_{t_{i-1}}^{(m)}) + \sigma^{(a)}(W_{t_j}^{(a)} - W_{t_{i-1}}^{(a)}) \\ &= \beta^{(a)}R_t^{(m)} + \sigma^{(a)}(W_{t_i}^{(a)} - W_{t_{i-1}}^{(a)}), \qquad (3.2.4) \end{aligned}$$

where $R_{t_i}^{(m)}$ is the return at time t_i on the market portfolio (m) and $R_{t_i}^{(a)}$ is the return on asset (a). Both $R_{t_i}^{(m)}$ and $R_{t_i}^{(a)}$ are i.i.d. with joint dynamics as follows

$$\begin{pmatrix} R_{t_i}^{(a)} \\ R_{t_i}^{(m)} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \left(\beta^{(a)2}\sigma^{(m)2} + \sigma^{(a)2}\right)\Delta & \beta^{(a)}\sigma^{(m)2}\delta \\ \beta^{(a)}\sigma^{(m)2}\delta & \sigma^{(m)2}\delta \end{pmatrix}\right),$$

for all t_i.

In order to estimate $\beta^{(a)}$ via maximum likelihood, one needs to maximize the logarithm of the following likelihood function

$$\mathcal{L}(\beta^a, \sigma^{(a)}, \sigma^{(m)}) = \prod_{i=1}^{N} f\Big(R_{t_i}^{(a)} | R_{t,i}^{(m)} \beta^{(a)}, \sigma^{(a)}, \sigma^{(m)}\Big).$$

Then,

$$\log\left(\mathcal{L}(\theta)\right) = -\frac{N}{2} ln(2\sigma^{(a)2}\delta) - \frac{\sum_{i=1}^{N}\left(R_{t_i}^{(a)} - \beta^{(a)}R_{t_i}^{(m)}\right)^2}{2\sigma^{(a)2}\delta}. \qquad (3.2.5)$$

The estimator $\hat{\beta}^{(a)}$ is consequently defined as

$$\begin{aligned} \hat{\beta}^{(a)} &= \underset{\beta^{(a)}}{\text{argmax}} \ \log\left(\mathcal{L}(\theta)\right) \\ &= \left(\sum_{i=1}^{N} R_{t_i}^{(m)2}\right)^{-1} \left(\sum_{i=1}^{N} R_{t_i}^{(m)} R_{t_i}^{(a)}\right). \qquad (3.2.6) \end{aligned}$$

Note that the maximum likelihood estimator (MLE) of $\beta^{(a)}$ is equivalent to the discrete time estimators of $\frac{\text{QCOV}_T}{\text{QV}_T}$, where QCOV_T is the quadratic covariation and $\text{Q}V_T$ is the quadratic variation of the process, Eq. (3.2.2) and Eq. (3.2.1). The discrete time estimators $\frac{\text{QCOV}_T}{\text{QV}_T}$ are often examined in the literature within the context of high–frequency data and referred as the realized beta, Rβ, see e.g. Andersen, Bollerslev, Diebold, and Wu (2006), Todorov and Bollerslev (2010), Reiss, Todorov, and Tauchen (2015) among others.

Theorem 3.2.1. *Fix δ. Assume (3.2.3), (3.2.4) holds with the dynamics (3.2.1). Then, the least squares estimator of $\beta^{(a)}$ coincides with the maximum likelihood estimator (3.2.6).*

Proof. Estimate $\beta^{(a)}$ via least squares technique

$$\hat{\beta}^{(a)} = \underset{\beta^{(a)}}{\text{argmin}} \sum_{i=1}^{N} (R_{t_i}^{(a)} - \beta^{(\alpha)} R_{t_i}^{(m)})^2.$$

Then,

$$2 \sum_{i=1}^{N} \left(R_{t_i}^{(a)} - \beta^{(a)} R_{t_i}^{(m)} \right) \times -R_{t_i}^{(m)} = 0.$$

Solving for $\beta^{(a)}$ yields

$$\hat{\beta}_{LS}^{(a)} = \left(\sum_{i=1}^{N} R_{t_i}^{(m)} R_{t_i}^{(m)} \right)^{-1} \left(\sum_{i=1}^{N} R_{t_i}^{(m)} R_{t_i}^{(a)} \right). \tag{3.2.7}$$

■

In the following, we provide the exact finite sample moments of the maximum likelihood, also the least squares estimators of $\beta^{(a)}$.

Finite Sample Moments of $\beta^{(a)}$ estimators

1.) Considering the first moment, we obtain the following theorem:

Theorem 3.2.2.

$$E[\hat{\beta}^{(a)}] = \beta^{(a)}.$$

Proof. Let $R_t^{(m)} = (R_{t_1}^{(m)}, R^{(m)}{}_{t_2}, ..., R_{t_N}^{(m)})'$ and $R_t^{(a)} = (R_{t_1}^{(a)}, R_{t_2}^{(a)}, ..., R_{t_N}^{(a)})'$. Moreover, let $u_t^{(a)} = (u_{t_1}^{(a)}, u_{t_2}^{(a)}, ..., u_{t_N}^{(a)})'$, with $u_{t_i} = (W_{t_i}^{(a)} - W_{t_{i-1}}^{(a)})$.

Then, we have

$$\begin{aligned}
\mathrm{E}[\hat{\beta}^{(a)}] &= \mathrm{E}\Big[\Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1} {R_t^{(m)}}^T R_t^{(a)}\Big] \\
&= \mathrm{E}\Big[\Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1} {R_t^{(m)}}^T \big(\beta^{(a)} R_t^{(m)} + u_t^{(a)}\big)\Big] \\
&= \beta^{(a)} + \mathrm{E}\Big[\Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1} {R_t^{(m)}}^T u_t^{(a)}\Big] \\
&= \beta^{(a)} + \mathrm{E}\Big[\Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1} {R_t^{(m)}}^T\Big] \underbrace{\mathrm{E}\Big[u_t^{(a)}\Big]}_{0} \\
&= \beta^{(a)},
\end{aligned}$$

where the second last equality follows from the fact that $R_t^{(m)}$ and $u_t^{(a)}$ are independent. ■

The theorem above implies the following: The maximum likelihood estimator (and least squares estimator), $\hat{\beta}^{(a)}$, based on the return observations on the asset and the market, $R_{t_i}^{(a)}$ and $R_{t_i}^{(m)}$, over [0,T] sampled from the continuous time process (3.2.1 and 3.2.2) is an unbiased estimator of $\beta^{(a)}$. Note that this result does not depend on δ: sampling returns very frequently or less frequently in this environment does not matter for the first sample moment of the estimator $\hat{\beta^{(a)}}$. In particular, estimating $\beta^{(a)}$ using five minute returns or daily returns will lead to same exact finite sample first moment of $\hat{\beta}^{(a)}$.

2.) The finite sample variance of the $\hat{\beta}^{(a)}$ is documented in the following theorem.

Theorem 3.2.3.

$$Var\Big[\hat{\beta}^{(a)}\Big] = \frac{{\sigma^{(a)}}^2 \times \delta}{{\sigma^{(m)}}^2 \times T}.$$

Proof.

$$\begin{aligned}
\mathrm{Var}\Big[\hat{\beta}^{(a)}\Big] &= \mathrm{E}\Big[\big(\hat{\beta^{(a)}} - \beta^{(a)}\big)\big(\hat{\beta^{(a)}} - \beta^{(a)}\big)^T\Big] \\
&= \mathrm{E}\Big[\Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1} {R_t^{(m)}}^T u_t^{(a)} {u_t^{(a)}}^T R_t^{(m)} \Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1}\Big] \\
&= \mathrm{E}\Big[\Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1} {R_t^{(m)}}^T \underbrace{\mathrm{E}\big[u_t^{(a)} {u_t^{(a)}}^T | {R_t^{(m)}}^T\big]}_{{\sigma^{(a)}}^2 \times \delta \times I} R_t^{(m)} \Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1}\Big] \\
&= {\sigma^{(a)}}^2 \times \delta \times \underbrace{\mathrm{E}\Big[\Big({R_t^{(m)}}^T R_t^{(m)}\Big)^{-1}\Big]}_{\Big(N \times {\sigma^{(m)}}^2 \times \delta\Big)^{-1}} \\
&= \frac{{\sigma^{(a)}}^2 \times \delta}{{\sigma^{(m)}}^2 \times T}.
\end{aligned}$$

■

Note that the finite sample variance, $\text{Var}\left[\hat{\beta}^{(a)}\right]$, is dependent on δ: The precision of the estimates of $\hat{\beta}^{(a)}$ is positively related to δ. In particular, sampling more frequently, i.e. δ being smaller, leads to higher precision of the estimates of $\beta^{(a)}$. This result is in line with the high frequency asymptotics provided in Barndorff-Nielsen, Graversen, and Shephard (2004), which are derived as $\delta \to 0$.

Moreover, the asymptotic variance for a given δ as $T \to \infty$

$$\sqrt{T}\big(\hat{\beta}^{(a)} - \beta^{(a)}\big) = \Big(\frac{1}{T/\delta}\sum_{i=1}^{T/\delta} R_{t-i}^{(m)} R_{t-i}^{(m)^T}\Big)^{-1}\Big(\frac{\sqrt{T}}{T/\delta}\sum_{i=1}^{T/\delta} R_{t_i}^{(m)} u_t^{(a)}\Big). \tag{3.2.8}$$

By making use of Slutsky's theorem and the weak law of large numbers, the asymptotic variance is

$$\sqrt{T}\Big(\hat{\beta}^{(a)} - \beta\Big) \xrightarrow{T\to\infty} N\Big(0, \underbrace{(\sigma^{(m)^2} \times \delta)^{-2} \times \sigma^{(a)^2}\sigma^{(m)^2}\delta^3}_{\frac{\sigma^{(a)^2}}{\sigma^{(m)^2}}\times\delta}\Big) \tag{3.2.9}$$

Hence, to summarize, for the asymptotic variance, we have

$$\text{AVAR}[\hat{\beta}^{(a)}] = \frac{\sigma^{(a)^2}}{\sigma^{(m)^2}} \times \delta, \tag{3.2.10}$$

with the approximation to the finite sample variance being

$$\text{Var}[\hat{\beta}^{(a)}] = \frac{\sigma^{(a)^2}\delta}{\sigma^{(m)^2}T} \tag{3.2.11}$$

As δ gets smaller, we obtain more precise estimates of $\beta^{(a)}$, both asymptotically and in finite samples.

3.2.2 Inference in the presence of market (microstructure) noise

In the following section we analyse the implications of noise for the properties of the obtained estimates. In the presence of noise, the econometrician is unable to observe $p_{t_i}^m$ and $p_{t_i}^{(a)}$ directly but instead observes the prices contaminated with the noise, $\tilde{p}_{t_i}^{(m)}$ and

$\tilde{p}_{t_i}^{(a)}$ as in

$$\tilde{p}_{t_i}^{(a)} = p_{t_i}^{(a)} + \varepsilon_{t_i}^{(a)},$$
$$\tilde{p}_{t_i}^{(m)} = p_{t_i}^{(m)} + \varepsilon_{t_i}^{(m)},$$

where $\varepsilon_{t_i}^{(a)} \sim$ i.i.d. $N(0, {s^{(a)}}^2)$ and $\varepsilon_{t_i}^{(m)} \sim$ i.i.d. $N(0, {s^{(m)}}^2)$ and $\varepsilon_{t_i}^{(a)}$ and $\varepsilon_{t_i}^{(m)}$ are independent of the $W^{(a)}$ and $W^{(m)}$ processes. The unobserved variables $p_{t_i}^{(m)}$ and $p_{t_i}^{(a)}$ are latent variables in this setting. The researcher computes returns by using the observed prices $\tilde{p}_{t_i}^{(a)}$ and $\tilde{p}_{t_i}^{(m)}$ rather than the latent efficient prices such that

$$\begin{aligned} \tilde{R}_{t_i}^{(a)} = \tilde{p}_{t_i}^{(a)} - \tilde{p}_{t_{j-1}}^{(a)} &= \left(p_{t_i}^{(a)} - p_{t_{j-1}}^{(a)}\right) + \left(\varepsilon_{t_i}^{(a)} - \varepsilon_{t_{j-1}}^{(a)}\right) \\ &= R_{t_i}^{(a)} + \eta_{t_i}^{(a)}, \end{aligned} \tag{3.2.12}$$

$$\begin{aligned} \tilde{R}_{t_i}^{(m)} = \tilde{p}_{t_i}^{(m)} - \tilde{p}_{t_{j-1}}^{(m)} &= \left(p_{t_i}^{(m)} - p_{t_{j-1}}^{(m)}\right) + \left(\varepsilon_{t_i}^{(m)} - \varepsilon_{t_{j-1}}^{(m)}\right) \\ &= R_{t_i}^{(m)} + \eta_{t_i}^{(m)}, \end{aligned} \tag{3.2.13}$$

In the following section, we aim to provide an answer to the question of what happens to the estimates of $\beta^{(a)}$ if the researcher ignores the presence of noise in prices and obtains the estimates via the standard maximum likelihood estimation as in the previous section?

It is obvious that the likelihood function will be misspecified because of the ignored noise components. So, if the estimation of the parameter $\beta^{(a)}$ is based on such a misspecified likelihood function, what would be the effect on the estimates of $\beta^{(a)}$?

The model (3.2.4) and (3.2.3) together with the dynamics of $\varepsilon_{t_i}^{(a)}$ and $\varepsilon_{t_i}^{(m)}$ together yield the following dynamics for $\tilde{R}_{t_i}^{(m)}$ and $\tilde{R}_{t_i}^{(a)}$

$$\tilde{R}_{t_i}^{(m)} = \sigma^{(m)}(W_{t_i}^{(m)} - W_{t_{j-1}}^{(m)}) + \eta_{t_i}^{(m)}, \tag{3.2.14}$$
$$\tilde{R}_{t_i}^{(a)} = \beta^{(a)}\sigma^{(m)}(W_{t_i}^{(m)} - W_{t_{j-1}}^{(m)}) + \sigma^{(a)}(W_{t_i}^{(a)} - W_{t_{j-1}}^{(a)}) + \eta_{t_i}^{(a)}, \tag{3.2.15}$$

where the $\eta_{t_i}^{(m)}$ are independent of $R_{t_i}^{(m)}$, and $\eta_{t_i}^{(a)}$ independent of $R_{t_i}^{(a)}$. Before switching to the estimation results, we first provide the variance and autocorrelation at lag one between the observed returns. These are

(i) $\mathrm{Var}(\tilde{R}_{t_i}^{(m)}) = {\sigma^{(m)}}^2\delta + 2{s^{(m)}}^2$,

(ii) $\mathrm{Var}(\tilde{R}_{t_i}^{(a)}) = {\sigma^{(a)}}^2\delta + {\beta^{(a)}}^2{\sigma^{(m)}}^2\delta + 2{s^{(a)}}^2$,

(iii) $\mathrm{Cov}(\tilde{R}_{t_i}^{(m)}, \tilde{R}_{t_{i-1}}^{(m)}) = -{s^{(m)}}^2$,

(iv) $\text{Cov}(\tilde{R}^{(a)}_{t_i}, \tilde{R}^{(a)}_{t_{i-1}}) = -{s^{(a)}}^2$.

The observed returns have the following joint dynamics

$$\begin{pmatrix} \tilde{R}^{(a)}_{t_i} \\ \tilde{R}^{(m)}_{t_i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \left({\beta^{(a)}}^2 {\sigma^{(m)}}^2 + {\sigma^{(a)}}^2\right)\delta + 2{s^{(a)}}^2 & \beta^{(a)}{\sigma^{(m)}}^2\delta \\ \beta^{(a)}{\sigma^{(m)}}^2\delta & {\sigma^{(m)}}^2\delta + 2{s^{(m)}}^2 \end{pmatrix}\right). \tag{3.2.16}$$

There are some interesting properties of the log–returns documented in (3.2.16)regarding the role of noise. First, the presence of noise leads to an increase in the variance of both the market's log-returns and the asset (a)'s log–returns, with an additive term of $2{s^{(m)}}^2$ and $2{s^{(a)}}^2$, for the market and the asset respectively. This finding is in line with Aït–Sahalia, Zhang and Jacod (2005).

The degree of contamination of the variances of observed returns through the presence of noise can be assessed by the proportion of variance induced by market (microstructure) noise. The proportion of the total variance caused by market (microstructure) noise is

$$Prop^{(a)} = \frac{2{s^{(a)}}^2}{\left({\beta^{(a)}}^2{\sigma^{(m)}}^2 + {\sigma^{(a)}}^2\right)\delta + 2{s^{(a)}}^2}, \tag{3.2.17}$$

$$Prop^{(m)} = \frac{2{s^{(m)}}^2}{{\sigma^{(m)}}^2\delta + 2{s^{(m)}}^2}. \tag{3.2.18}$$

Both $Prop^{(a)}$ and $Prop^{(m)}$ get smaller as δ gets larger whereas they get close to 1 as δ gets smaller. In particular, sampling returns more frequently, i.e. δ being small, leads to larger proportions of return variances to be driven by noise. On the other hand, sampling returns less frequently, i.e. δ being large, leads to smaller proportions or return variances to be driven by noise. Note that this result is in line with the empirical literature on realized estimators based on high frequency data, advising less frequent sampling to deal with arising market microstructure noise by use of e.g. volatility signature plots. Moreover, note that the log–return variations, for both the market and the asset (a), get larger with less frequent sampling, i.e. δ being larger, within the current setting.

From here, we move forward to investigate the properties of the estimators of $\beta^{(a)}$. (Microstructure) noise is present in the data but ignored and the estimation is based on the maximum likelihood estimation. The likelihood function is misspecified because the noise components are ignored.

The maximum likelihood estimation with $\tilde{R}_t^{(m)^T}$ and $\tilde{R}_t^{(a)}$ yields the following estimator

$$\begin{aligned}\hat{\beta}^{(mis)} &= \left(\tilde{R}_t^{(m)^T}\tilde{R}_t^{(m)}\right)^{-1}\tilde{R}_t^{(m)^T}\tilde{R}_t^{(a)} \\ &= \Big(\sum_{j=1}^{N}\tilde{R}_{t_i}^{(m)^2}\Big)^{-1}\tilde{R}_t^{(m)^T}\tilde{R}_{t_i}^{(a)}. \end{aligned} \tag{3.2.19}$$

The bias of the estimator, $\hat{\beta}^{(mis)}$, in finite samples (T being finite), is documented in the following theorem

Theorem 3.2.4. *Fix the sampling frequency δ. In finite samples (T finite):*
The first moment of $\hat{\beta}^{(mis)}$ is

$$E[\hat{\beta}^{(mis)}] = \beta^{(a)}\frac{{\sigma^{(m)}}^2\delta}{{\sigma^{(m)}}^2\delta + 2{s^{(m)}}^2}. \tag{3.2.20}$$

The bias is

$$-\beta^{(a)}\frac{2{s^{(m)}}^2}{{\sigma^{(m)}}^2\delta + 2{s^{(m)}}^2}. \tag{3.2.21}$$

Proof.

$$\begin{aligned}\mathrm{E}[\hat{\beta}^{(mis)}] &= \mathrm{E}\Big[\Big(\sum_{j=1}^{N}\tilde{R}_{t_i}^{(m)^2}\Big)^{-1}\Big(\sum_{j=1}^{N}\tilde{R}_{t_i}^{(m)^2}\tilde{R}_{t_i}^{(a)}\Big)\Big] \\ &= \big({\sigma^{(m)}}^2\delta + 2{s^{(m)}}^2\big)^{-1}\times\beta^{(a)}{\sigma^{(m)}}^2\delta \\ &= \beta^{(a)}\frac{{\sigma^{(m)}}^2\delta}{{\sigma^{(m)}}^2\delta + 2{s^{(m)}}^2}. \end{aligned}$$

given the bivariate normal distribution of $\tilde{R}_{t_i}^{(a)}$ and $\tilde{R}_{t_i}^{(m)}$. ■

The theorem above highlights important properties of the MLE estimator $\hat{\beta}^{(mis)}$. First, the MLE estimator, $\beta^{(mis)}$, based on the log–returns contaminated by noise is biased with respect to the correct population parameter value, $\beta^{(a)}$. In other words, estimating the loading by using the observed log–returns of the market and the asset (a) that are contaminated by noise, rather than the unobserved log–returns leads to a biased estimate compared to $\beta^{(a)}$. Second, the MLE estimator, $\hat{\beta}^{(mis)}$, is biased towards zero. The bias in $\hat{\beta}^{(mis)}$ as an estimator of $\beta^{(a)}$ is in a multiplicative form and

$$\text{Bias coeff.} = \frac{{\sigma^{(m)}}^2\delta}{{\sigma^{(m)}}^2\delta + 2{s^{(m)}}^2} \leq 1. \tag{3.2.22}$$

The quality of the estimators in the presence of the microstructure noise is dependent on the Bias coeff.. When Bias coeff. is close to 1, there exists less bias in the estimate $\hat{\beta}^{(mis)}$. Moreover, note when there is no (microstructure) noise, i.e. when ${s^{(m)}}^2 = 0$, the

Bias coeff. is equal to one and the MLE estimator $\hat{\beta}^{(mis)}$ is unbiased with respect to the population parameter value $\beta^{(a)}$.

Note that the sign of $\beta^{(a)} \frac{{\sigma^{(m)}}^2 \delta}{{\sigma^{(m)}}^2 \delta + 2{s^{(m)}}^2}$ is the same as the sign of $\beta^{(a)}$ and

$$\frac{{\sigma^{(m)}}^2 \delta}{{\sigma^{(m)}}^2 \delta + 2{s^{(m)}}^2} |\beta^{(a)}| \leq |\beta^{(a)}|. \tag{3.2.23}$$

where the equality holds when ${s^{(m)}}^2 = 0$. In other words, $\hat{\beta}^{(mis)}$ is downward biased in absolute value and always smaller than $\beta^{(a)}$ in absolute terms.

One important highlight of the Theorem (3.2.4) is the following: The bias in the estimate $\hat{\beta}^{(mis)}$ gets smaller as δ gets larger. Observe that when δ gets larger, the Bias coeff. gets closer to one such that the bias is reduced. That means, if the researcher samples log–returns less frequently, i.e. every ten rather than five minutes or daily rather than ten minutes, the estimate $\hat{\beta}^{(mis)}$ is less biased with respect to the population parameter $\beta^{(a)}$. These results are in spirit similar to the arguments in the empirical literature analysing realized estimators of integrated volatility—suggesting sampling coarser to decrease the arising bias in the realized volatility estimators; see e.g. Aït–Sahalia and Jacod (2012).

3.3 Monte Carlo Evidence

3.3.1 The set–up

In this section, we study the finite-sample properties of the estimators for $\beta^{(a)}$ when the prices of the market and the asset **are** and **are not** contaminated by market (microstructure) noise. We assume empirically realistic parameter values for our data generating process, $\sigma^{(a)} = 35\%$ per year and $\sigma^{(m)} = 25\%$ per year. We further assume a $\beta^{(a)}$ of unity. In order to see how the results change with the variation in the noise terms, we use various parameter values for ${s^{(m)}}^2$ such that ${s^{(m)}}^2 = 0.10\%$, ${s^{(m)}}^2 = 0.20\%$ and ${s^{(m)}}^2 = 0.30\%$. These parameter values are in line with the empirical market microstructure literature; see, e.g., Hasbrouk and Sofianos (1993), Madhavan, Richardson and Roomans (1997) and Aït–Sahalia, Zhang and Jacod (2005).

We generate S=10,000 log–price paths for asset (a) and the market, $p^{(a)}$ and $p^{(m)}$, based on the data generating process (3.2.1) and (3.2.2) with the sample length of $T = 1$ (one year) for each scenario considered. We then obtain the uncontaminated log-returns on the market and asset, i.e. $R^{(m)}$ and $R^{(a)}$ respectively. In order to impose (microstructure) noise contaminated prices, we then simply add the noise term, $\varepsilon^{(m)}$ and $\varepsilon^{(a)}$ which yields $\tilde{p}^{(m)}$

and $\tilde{p}^{(a)}$. We assume that both noise terms are normally distributed with $s^{(m)} = 0.20\%$. We generate S=10,000 simulations for each specification of $s^{(m)}$, i.e. ${s^{(m)}}^2 = 0.10\%$, ${s^{(m)}}^2 = 0.20\%$ and ${s^{(m)}}^2 = 0.30\%$.

The asset pricing literature has so far mainly focussed on using sparsely sampled returns at equidistant intervals such as 5–minute returns in an unbalanced panel (Amaya et al., 2015), 22.5–minute returns (Bollerslev and Todorov , 2010), 25–minute returns counting the overnight period as the first of 16 consecutive increments (Patton and Verardo, 2012), and 75–minute returns with distinct overnight period returns (Bollerslev et al., forthcoming).

In order to examine to role of sampling frequency δ, we here hypothetically consider the open market hours from 9.30–16.00 EST during which the stocks underlying the index are actively traded, generating observations over 390 minutes for a day (252×390 a year). We examine a wide range of sampling frequencies, i.e. we sample log–returns with sampling frequencies δ ranging from from every minute to every week. In particular, we sample every minute, five minutes, 15–minutes, 30–minutes, 78–minutes (five observations per day), 195–minutes (two observations per day), once day and one week, which approximately covers the aforementioned most popular choices in the literature for a broad range of different liquidity scenarios.

For each set of simulation results, we provide the theoretical finite sample (of size T) mean, $E[\hat{\beta}^{(a)}]$, then average finite sample (of size T) mean over S=10,000 simulations, *Estimated f.s. mean*. Moreover, we provide theoretical finite sample standard deviation,$Var[\hat{\beta}^{(a)}]^{1/2}$ for the case of no microstructure noise based on Theorem (**??**), and the corresponding finite sample standard deviation, *Estimated f.s. standard deviation*, computed from the sample of S=10000 simulated paths.

We also present the $Percentage\ Error = 100 \times \frac{\hat{\beta}_s^{(a)} - \beta^{(a)}}{\beta^{(a)}}$, and the root mean–square error, $RMSE = \left(\frac{1}{S}\Sigma_{s=1}^{S}\left(\hat{\beta}^{(a)} - \beta^{(a)}\right)\right)^{1/2}$, with S=10,000. Lastly, we present the mean absolute error $MAE = \frac{1}{S}\Sigma_{s=1}^{S}\left|\hat{\beta}_s^{(a)} - \beta^{(a)}\right|$. Note in cases where we do not have an analytic form for the sample moments, we instead use a large number of simulated paths, i.e. S=10,000 in order to approximate the finite sample moments. In that aspect, the mean squared error, approximates the finite sample second moment of the random variable, $\hat{\beta}^{(a)}$, plus the squared bias. The MAE is interesting as to assess the impact of large deviations in magnitude of the estimators from simulated paths from $\beta^{(a)} = 1$.

3.3.2 Results

We start by presenting our results for the case when there is no (microstructure) noise being present.

Table 3.A.1 refers to the simulation results when returns are not contaminated by (microstructure) noise. We note the following: Firstly, the theoretical mean provided in Theorem 3.2.2 closely matches the estimated finite sample mean. The same holds for the square–root of the theoretical variance provided in Theorem 3.2.3 and the estimated finite sample standard deviation. Moreover, the bias is very small across all sampling frequencies. The result is consistent with Theorem 3.2.2 as the theoretical finite sample mean is not dependent on the length of the sampling interval, i.e. δ.

We also observe a clear pattern between the finite sample variance and the length of sampling interval δ: The finite sample variance increases as the sampling frequency decreases. In particular, the finite sample standard deviation of the estimator $\hat{\beta}^{(a)}$ is 0.0045 when sampling every minute, whereas it increases to 0.1980 when sampling takes place once a week. Hence, sampling more and more frequently leads to precision gains in the absence of noise.

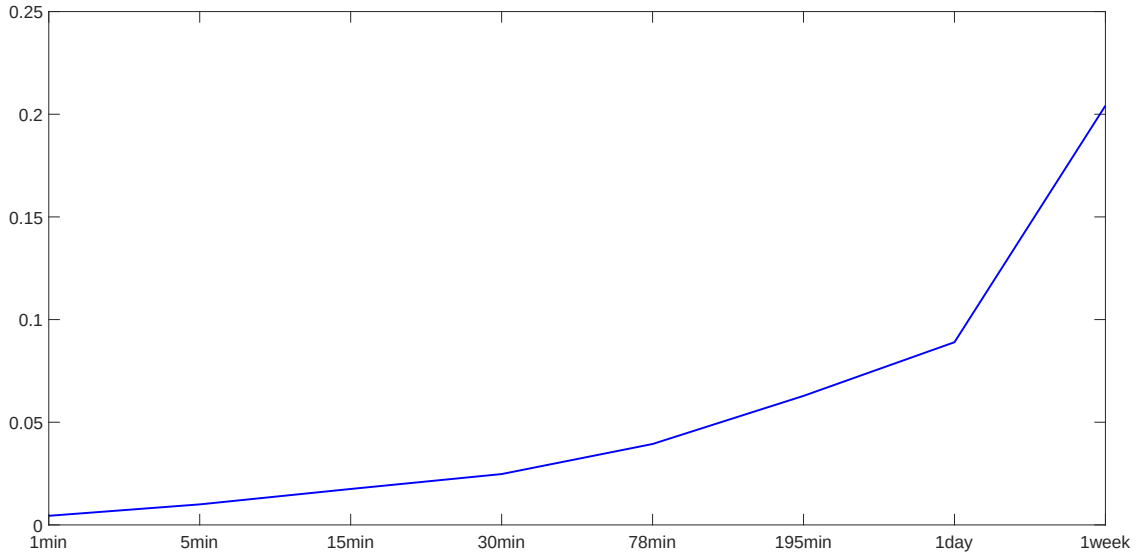
Note that this result is in line with precision gains documented in the high frequency literature based on the in–fill asymptotic analyses—based on $\delta \to 0$ as provided in Barndorff-Nielsen, Graversen, and Shephard (2004). When prices are not contaminated by (microstructure) noise there is almost no bias present in our estimates with respect to the true parameter value and there are large precision gains from sampling more frequently; see Figure 3.1.

However, the results are considerably different when (microstructure) noise is present in the data. In Table 3.A.2 and Table 3.A.3 we document the simulation results when observed prices are contaminated by market (microstructure) noise. Note that in order to assess the sensitivity of the results to the variation of the noise terms $\epsilon^{(m)}$, we show results for three different sensible values of ${s^{(m)}}^2$, i.e. ${s^{(m)}}^2 = 0.10/100$, ${s^{(m)}}^2 = 0.20/100$ and ${s^{(m)}}^2 = 0.30/100$, which are presented in Panel 1, Panel 2 and Panel 3 in each table respectively.

For each panel, Table 3.A.2 provides the theoretical mean, $\mathrm{E}[\hat{\beta}^{mis}]$ calculated from Eq. (3.2.20) in Theorem 3.2.4, followed by the average of the estimated $\hat{\beta}^{mis}$ across simulations. There are several important observations to be highlighted about the finite sample means of the estimates $\hat{\beta}^{mis}$. In the third row of each panel the theoretical bias, $\mathrm{E}[\hat{\beta}^{mis}]$,

Figure 3.1: RMSE in the no noise case

The figure shows the RMSE in the no noise case as a function of sampling time.



calculated from Eqn. 3.2.21 in Theorem 3.2.4 is presented, followed by the bias calculated from S=10,000 simulation paths. Table 3.A.3 presents the theoretical mean, E$[\hat{\beta}^{mis}]$, followed by Percentage Error (as explained previously). We also document, in the third row of each panel, the finite sample standard deviation based on simulations, followed by their RMSE and MAE (as both explained previously).

We would like to highlighted in particular the following about the finite sample mean of $\hat{\beta}^{mis}$. First of all, there is a clear downward bias in the estimates with respect to $\beta^{(a)} = 1$ and it tends to increase significantly with sampling returns more and more often. In particular, for $s^{(m)} = 0.10/100$, the bias is strikingly large and negative with percentage error of 76% when returns are sampled every minute. Sampling less often leads to a reduced bias in our estimates, with percentage error declining already to 17% when sampling every 15 minutes, to 3% when sampling takes place every 78 minutes. On the other extreme, the bias almost disappears if the returns are sampled every week. In fact, this decline in the pattern is observed for all specifications of the variation in noise terms, i.e. $s^{(m)} = 0.10/100$, $s^{(m)} = 0.20/100$ and $s^{(m)} = 0.30/100$. We also note the results of the simulations are in line with the analytical forms derived and presented in Theorem 3.2.4.

Next, we examine the finite sample standard deviations of the estimates in the presence of market microstructure noise. In contrast to first moments, the standard deviations tend to decline when sampling returns more frequently and the same pattern is observed for

all three values of $s^{(m)}$. For example, when $s^{(m)} = 0.10\%$ with returns sampled once a week, the estimated standard deviation is 0.3085. On the other hand, if the returns are sampled every minute, the standard deviation drops significantly to 0.1310. Sampling once every half an hour as well as five times daily (78 minutes) produces fairly close standard deviations, still for $s^{(m)} = 0.10\%$.

Consider now what happens with regard to the RMSEs when there is (microstructure) noise present in the data. It is very interesting that the RMSE is very large when returns are sampled very frequently, such as every minute. As returns are sampled at coarser and coarser frequencies, the RMSE declines up to a certain sampling frequency, i.e. 78 minutes for $s^{(m)} = 0.10\%$, and starts increasing again monotonically as the sampling interval becomes coarser. This clear convex shape indicates the existence of a sampling frequency which is optimal. This result is striking and very useful because the literature has so far been mostly silent with regard to which sampling frequency to choose when the interest lies is in estimating $\beta^{(a)}$ under contamination of (microstructure) noise in the data.

Thus, it is important to further analyse the role of the noise parameter $s^{(m)}$ in our results. The bias in estimates increases considerably when $s^{(m)}$ is larger for all sampling frequencies. Observe that percentage errors in Panel 1 are smaller than the percentage errors in Panel 2 for the corresponding sampling frequencies, and both are smaller than the corresponding percentage errors in Panel 3. Therefore, increase in $s^{(m)}$ leads to larger bias in the estimates. This is in line with the Eq. (3.2.21) in Theorem 3.2.4 where we show the bias in analytical form—the bias coefficient, $Biascoeff.$ increases with $s^{(m)}$. The first moment (mean) estimates are pushed downwards, even becoming close to zero as $s^{(m)}$ becomes larger. For example, when returns are sampled every minute, the mean estimate across simulations is 0.2427 (with Percentage Error -75%) for $s^{(m)} = 0.10\%$, and it decreases down to only 0.0344 (with Percentage Error -96%) for $s^{(m)} = 0.30\%$—recalling that the population $\beta^{(a)}$ equals one. For weekly sampling, the same statistic is 0.9976 (with Percentage Error -0.2%) for $s^{(m)} = 0.10\%$, and it decreases to 0.9884 (with Percentage Error -1.1% for $s^{(m)} = 0.30\%$.

Considering the standard deviations, estimates decline as $s^{(m)}$ gets larger for every sampling frequency. We observe that the standard deviations in Panel 1 are larger than the standard deviations in Panel 2 for corresponding sampling frequencies. The standard deviations in Panel 3 are smaller than the corresponding statistics both in Panel 1 and

Panel 2. For example, when the returns are sampled every day, the finite sample standard deviation is 0.2435 for $s^{(m)} = 0.10\%$, whereas it is 0.2431 for $s^{(m)} = 0.20\%$ and 0.2394 $s^{(m)} = 0.30\%$. The decline in standard deviations seems to be sharper when returns are sampled more often such as one minute, five or 15 minutes. For instance, when returns are sampled on a minute–by–minute basis, the standard deviation is 0.1310 for $s^{(m)} = 0.10\%$ and goes down to 0.0513 for $s^{(m)} = 0.10\%$. The same statistic is 0.1965 and 0.2146 when returns are sampled every five and 15 minutes respectively for $s^{(m)} = 0.10\%$ and they decline to 0.1057 and 0.1560 for $s^{(m)} = 0.30\%$

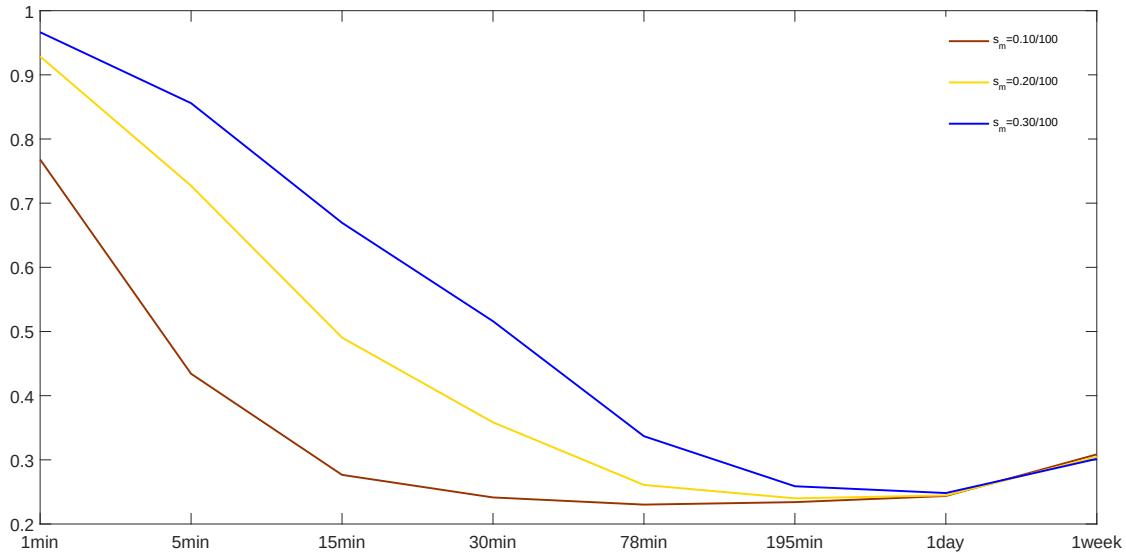
Lastly, we examine the effect of $s^{(m)}$ on the results regarding the RMSE. For the four sampling frequencies (one, five, 15, 78 minutes) there are large increases in the RMSEs as $s^{(m)}$ grows. For example, when returns are sampled every minute, the RMSE is 0.7682 for $s^{(m)} = 0.10\%$, which rises considerably up to 0.9664 when $s^{(m)} = 0.30\%$. Moreover, the RMSE is 0.4341 for $s^{(m)} = 0.10\%$ and rises up to 0.8558 for $s^{(m)} = 0.30\%$ when returns are sampled every five minutes. Interestingly, if the returns are sampled very coarsely such as on a daily or weekly basis, the RMSEs are very close to each other as $s^{(m)}$ increases from 0.10% to 0.30%. Moreover, the first decreasing and then increasing pattern in RMSE exists for all three values of $s^{(m)}$ analyzed. We plot the RMSEs for the various $s^{(m)}$ values in Figure 3.2. The first declining and then increasing pattern is clearly observable for all values of $s^{(m)}$. Observe that the RMSEs are considerably close to each other for daily and weekly returns.

However, there is an interesting observation to be highlighted: The sampling frequency that yields the smallest RMSE is different for different values $s^{(m)}$. For example, when $s^{(m)} = 0.10\%$, sampling every 78 minutes leads to the smallest RMSE among the other frequencies considered. However, sampling every 195 minutes yields the smallest RMSE when $s^{(m)} = 0.20\%$. Last, when $s^{(m)} = 0.30\%$, daily sampling leads to the smallest RMSE. All in all, these results indicate the following: The optimal sampling frequency appear to depend on the value of $s^{(m)}$.

3.4 Concluding remarks

β carries a crucial role for discrete time factor model literature and for portfolio allocation as well as risk management decisions. Using high frequency data brings precision gains in the estimates. However, it also leads to a bias in the estimates when the prices are

Figure 3.2: RMSE Noise case



contaminated by the market microstructure noise but this fact is not accounted for. These effects compensate each other and lead to an optimal sampling frequency (in smallest RMSE sense) - among the alternatives considered. Such trade–off does not exist when there is no microstructure noise present in the data.

One has to be careful in making inference with very highly frequent data: sampling very frequently may push the estimates toward zero - which may cause us to make quick conclusion that the asset does not load on the market if we estimate them by using data with very high frequencies. This may not be true in reality. We show that these economically small estimates of the beta may be caused by the bias caused by the contamination in the prices of the market.

Appendix

3.A Tables for Chapter 3

Table 3.A.1: Properties of $\hat{\beta}^{(a)}$–estimators when there is no microstructure noise

This table reports the theoretical mean ($\mathrm{E}[\hat{\beta}^{(a)}]$), theoretical standard deviation ($\mathrm{Var}(\hat{\beta})^{1/2}$) and average estimate, estimated standard deviation, the percentage error, root–mean–square–error (RMSE) and mean–absolute–error (MAE) over 10,000 simulated data sets. The theoretical mean ($\mathrm{E}[\hat{\beta}^{(a)}]$) is obtained from Eqn. (**??**) and the theoretical standard variation ($\mathrm{Var}(\hat{\beta})^{1/2}$) is calculated from Eqn. (**??**). The data–generating process for asset (a) and market is given by Eqn.(3.2.2) and Eqn.(3.2.1). The true parameter values to generate the log-prices for the asset (a) and the market are the following: $\beta^{(a)} = 1$, $\sigma^{(a)} = 35\%$, $\sigma^{(m)} = 25\%$. The length of each sample $T = 1$ year. Unit of time is assumed to be one year. The results are presented for the log-returns sampled with the following frequencies: 1min, 5min, 15min, 30min, 78min, 195min, 1day and 1week.

No Microstrcuture Noise	**1min**	**5min**	**15 min**	**30min**	**78min**	**195min**	**1day**	**1week**
$\boldsymbol{E[\hat{\beta}]}$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Estimated f.s. mean	1.0000	1.0001	1.0002	1.0003	1.0006	0.9998	1.0004	1.0000
$\boldsymbol{Var(\hat{\beta})^{1/2}}$	0.0045	0.0100	0.0174	0.0246	0.0396	0.0626	0.0885	0.1980
Estimated f.s. std.	0.0044	0.0100	0.0175	0.0248	0.0394	0.0628	0.0890	0.2042
Estimated f.s. mean	1.0000	1.0001	1.0002	1.0003	1.0006	0.9998	1.0004	1.0000
Percentage error	0.0002	0.0126	0.0159	0.0297	0.0582	-0.0193	0.0430	0.0032
Estimated f.s. std.	0.0044	0.0100	0.0175	0.0248	0.0394	0.0628	0.0890	0.2042
RMSE	0.0044	0.0100	0.0175	0.0248	0.0394	0.0628	0.0890	0.2042
MAE	0.0035	0.0080	0.0140	0.0198	0.0314	0.0502	0.0710	0.1622

Table 3.A.2: Properties of $\hat{\beta}^{mis}$–estimators when there is microstructure noise is present but ignored - 1

This table reports the theoretical mean ($\mathrm{E}[\hat{\beta}^{mis}]$), theoretical bias ($\mathrm{E}[\hat{\beta}^{mis}] - \beta^{(a)}$) and average estimate, estimated bias over S=10,000 simulated data sets. The theoretical mean ($\mathrm{E}[\hat{\beta}^{mis}]$) is obtained from Eqn. (**??**). The data–generating process for asset (a) and market is given by Eqn.(3.2.2) and Eqn.(3.2.1). The true parameter values to generate the log-prices for the asset (a) and the market are the following: $\beta^{(a)} = 1$, $\sigma^{(a)} = 35\%$, $\sigma^{(m)} = 25\%$ and $s^{(a)} = 0.20\%$. The length of each sample $T = 1$ year. Unit of time is assumed to be one year. The results are presented for the log-returns sampled with the following frequencies: 1min, 5min, 15min, 30min, 78min, 195min, 1day and 1week. Panel 1 presents the simulation results for $s^{(m)} = 0.10\%$, Panel 2 presents the simulation results for $s^{(m)} = 0.20\%$ and Panel 3 presents the simulation results for $s^{(m)} = 0.30\%$.

Panel 1-s_M=**0.10/100**	**1min**	**5min**	**15 min**	**30min**	**78min**	**195min**	**1day**	**1week**
$\boldsymbol{E[\hat{\beta}^{mis}]}$	0.2427	0.6158	0.8278	0.9058	0.9615	0.9843	0.9921	0.9984
Estimated f.s. mean	0.2431	0.6129	0.8255	0.9041	0.9606	0.9831	0.9918	0.9976
***Theoretical bias*= $\boldsymbol{E[\hat{\beta}^{mis}] - \beta^{(a)}}$**	-0.7573	-0.3842	-0.1722	-0.0942	-0.0385	-0.0157	-0.0079	-0.0016
Estimated bias	-0.7569	-0.3871	-0.1745	-0.0959	-0.0394	-0.0169	-0.0082	-0.0024
Panel 2-s_M=**0.20/100**								
$\boldsymbol{E[\hat{\beta}^{mis}]}$	0.0742	0.2860	0.5459	0.7062	0.8621	0.9398	0.9690	0.9936
Estimated f.s. mean	0.0744	0.2868	0.5470	0.7061	0.8637	0.9436	0.9732	0.9955
***Theoretical bias*= $\boldsymbol{E[\hat{\beta}^{mis}] - \beta^{(a)}}$**	-0.9258	-0.7140	-0.4541	-0.2938	-0.1379	-0.0602	-0.0310	-0.0064
Estimated bias	-0.9256	-0.7132	-0.4530	-0.2939	-0.1363	-0.0564	-0.0268	-0.0045
Panel 3-s_M=**0.30/100**								
$\boldsymbol{E[\hat{\beta}^{mis}]}$	0.0344	0.1511	0.3482	0.5165	0.7353	0.8741	0.9328	0.9858
Estimated f.s. mean	0.0349	0.1507	0.3490	0.5181	0.7349	0.8755	0.9341	0.9884
***Theoretical bias*= $\boldsymbol{E[\hat{\beta}^{mis}] - \beta^{(a)}}$**	-0.9656	-0.8489	-0.6518	-0.4835	-0.2647	-0.1259	-0.0672	-0.0142
Estimated bias	-0.9651	-0.8493	-0.6510	-0.4819	-0.2651	-0.1245	-0.0659	-0.0116

Table 3.A.3: Properties of $\hat{\beta}^{mis}$–estimators when there is microstructure noise is present but ignored - 2

This table reports average estimate, estimated standard deviation, the percentage error, root–mean–square–error (RMSE) and mean–absolute–error (MAE) over 10,000 simulated data sets. The data–generating process for asset (a) and market is given by Eqn.(3.2.2) and Eqn.(3.2.1). The true parameter values to generate the log-prices for the asset (a) and the market are the following: $\beta^{(a)} = 1$, $\sigma^{(a)} = 35\%$, $\sigma^{(m)} = 25\%$ and $s^{(a)} = 0.20\%$. The length of each sample $T = 1$ year. Unit of time is assumed to be one year. The results are presented for the log-returns sampled with the following frequencies: 1min, 5min, 15min, 30min, 78min, 195min, 1day and 1week. Panel 1 presents the simulation results for $s^{(m)} = 0.10\%$, Panel 2 presents the simulation results for $s^{(m)} = 0.20\%$ and Panel 3 presents the simulation results for $s^{(m)} = 0.30\%$.

Panel 1 s_M=0.10/100	1min	5min	15 min	30min	78min	195min	1day	1week
Estimated f.s. mean	0.2431	0.6129	0.8255	0.9041	0.9606	0.9831	0.9918	0.9976
Percentage error %	-75.6905	-38.7121	-17.4502	-9.5947	-3.9437	-1.6861	-0.8209	-0.2424
Estimated f.s. std.	0.1310	0.1965	0.2146	0.2215	0.2268	0.2335	0.2435	0.3085
RMSE	0.7682	0.4341	0.2766	0.2414	0.2302	0.2341	0.2436	0.3085
MAE	0.7569	0.3907	0.2246	0.1929	0.1833	0.1863	0.1939	0.2455
Panel 2 s_M=0.20/100								
Estimated f.s. mean	0.0744	0.2868	0.5470	0.7061	0.8637	0.9436	0.9732	0.9955
Percentage error %	-92.5648	-71.3178	-45.3031	-29.3912	-13.6286	-5.6397	-2.6756	-0.4544
Estimated f.s. std.	0.0745	0.1420	0.1880	0.2052	0.2225	0.2333	0.2431	0.3046
RMSE	0.9286	0.7272	0.4905	0.3584	0.2609	0.2400	0.2446	0.3046
MAE	0.9256	0.7132	0.4540	0.3074	0.2107	0.1914	0.1942	0.2423
Panel 3 s_M=0.30/100								
Estimated f.s. mean	0.0349	0.1507	0.3490	0.5181	0.7349	0.8755	0.9341	0.9884
Percentage error %	-96.5054	-84.9286	-65.0974	-48.1921	-26.5100	-12.4458	-6.5935	-1.1639
Estimated f.s. std	0.0513	0.1057	0.1560	0.1846	0.2078	0.2270	0.2394	0.3013
RMSE	0.9664	0.8558	0.6694	0.5161	0.3368	0.2588	0.2483	0.3015
MAE	0.9651	0.8493	0.6510	0.4823	0.2850	0.2084	0.1992	0.2381

Concluding Remarks

This thesis concentrates on establishing a better understanding of asset price dynamics which are important inputs into modelling capital markets, portfolio construction and optimization, as well as forecasting exercises. We present a thorough literature review that can serve as a financial econometric textbook for scholars to learn about the different angles and viewpoints of financial econometrics across different models and methodologies.

In the first chapter, we design a new class of estimators leveraging the realized methodology and apply it to the third moment, which we disentangle into its components, and present theoretical as well as empirical aspects which lead the way to do further research and apply derived estimators to financial economic research questions such as those encountered in asset pricing introduced in the introductory literature review. In the second chapter, we uncovered large, time–varying magnitude of the asymmetric effects with a focus on asymmetric constituents correlation which may be applied to unconditional portfolio construction methodologies for long term investors. In the last chapter, we focus on the variance and bias trade–off when sampling betas ever more frequently from a theoretical perspective.

Bibliography

Aït-Sahalia, Y., Fan, J., and Li, Y. (2013). The leverage effect puzzle: Disentangling sources of bias at high frequency. *Journal of Financial Economics* 109, 224-49. → 64

Aït-Sahalia, Y., Zhang, L., and Jacod, J. (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies* 18, 351–416. → 252, 254

Aït-Sahalia, Y., and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature* 50, 1007–1050. → 60, 254

Albuquerque, R. (2012). Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns. *The Review of Financial Studies* 25,5 1630–1673. → 51, 61, 62, 223

Alizadeh, S., Brandt, M., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57(3):1047–1091. → 34

Andersen, T. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance* 4, 75–102. → 11

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross–section of equity returns. *Journal of Financial Economics* 118,1 135-167. → 63, 255

Andersen, T. and Bollerslev, T. (1997). Intraday periodicity and volatility peristence in financial markets. *Journal of Empirical Finance* 4, 43–76. → 91

Andersen, T. and Bollerslev, T. (1998). DM–dollar volatility: Intraday activity patterns, macroeconomic announcements and longer run dependencies. *The Journal of Finance* 53, 219–65. → 21, 22, 67, 77

Andersen, T., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* 89, 4 701–20. → 24, 27, 30, 31, 77

Andersen, T., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76. → 22, 67, 79

Andersen, T., Bollerslev, T., Diebold, F. X., and Labys, P. (2000). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55. → 22, 79

Andersen, T., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 529–626. → 5, 68

Andersen, T., Bollerslev, T., Diebold, F. X., and Vega, C. (2007). Real–time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics* 73, 251–77. → 92

Andersen, T.G., Bollerslev, T., and Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics* 160, 176-189. → 31

Andersen, T., Bollerslev, T., Diebold, F. X., and Vega, C. (2007). Realized Beta: Persistence and Predictability. *Advances in Econometrics edited by Thomas B. Fomby, Dek Terrell. Emerald Group Publishing.* 20B, 1–39. → 248

Andersen, T., Bondarenko, O., González–Pérez, M. T. (2015). Exploring return dynamics via corridor implied volatility. *The Review of Financial Studies* 28, 10 2902–2945 → 101

Andersen, T., Bondarenko, O., Kyle, A., and Obizhaeva, A. (2015). Intraday trading invariance in the e–mini S&P 500. *Working Paper* → 11

Andersen, T., Chung, H.-Y., and Sørensen, B. (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics* 91, 61–87. →

Andersen, T., Dobrev, D., and Schaumburg, E. (2012). Jump–robust volatility estimation using nearest neighborhood truncation. *Journal of Econometrics* 169, 75–93. → 73, 91

Andersen, T., Dobrev, D., and Schaumburg, E. (2014). A robust neighborhood truncation approach to estimation of integrated quarticity. *Econometric Theory* 30, 3–59. →

Andersen, T., Fusari, N., and Todorov, V. (2015). Parametric inference and dynamic state recovery from option panels. *Econometrica* 83, 3 1081–1145. →

Andersen, T. and Sørensen, B. (1996). GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study. *Journal of Business Economic Statistics* 14,3, 328–52. →

Ane, T. and Geman, H. (2000). Order flow, transaction clock, and normality of asset returns. *The Journal of Finance* 55, 2259–2289. → 11

Areal, N. M. P. C. and Taylor, S. J. (2002). The realized volatility of FTSE–100 futures prices. *Journal of Futures Markets* 22, 627–48. → 22, 79

Avramov, D., Chordia, T., and Goyal, A. (2006). The impact of trades on daily volatility. *Review of Financial Studies* 19, 4 1241–77. → 198, 199

Aydemir, A., Gallmeyer, M., and Hollifield, B. (2006). Financial leverage does not cause the leverage effect. *Working Paper. Texas AM and Carnegie Mellon Universities.* → 193

Bae, J., Kim, C. J., and Nelson, C. (2007). Why are stock returns and volatility negatively correlated? *Journal of Empirical Finance* 14,1 41–58. → 197

Bakshi, G. and Kapadia, N. (2003). Volatility Risk Premiums Embedded in Individual Equity Options Some New Insights. *The Journal of Derivatives* 11,1 45–54. →

Bakshi, G., Kapadia, N. and Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies.* 16,1 101–43. → 65

Bandi, F. M. and Renò, R. (2012). Time-varying leverage effects. *Journal of Econometrics* 169,1 94–113. → 65

Barndorff-Nielsen, O.E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B* 63, 167–241. → 14

Barndorff-Nielsen, O.E., Graversen, S.E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of the Royal Statistical Society, Series B* 64 253–280. → 21, 23, 67

Barndorff-Nielsen and O.E., Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2,1 1–37. → 25, 27

Barndorff-Nielsen, O.E., Graverseon, S.E., and Shephard, N. (2004). Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica* 72, 885–9225. → 250, 256

Barndorff-Nielsen, O.E., Shephard, N., and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and Their Applications* 116, 796–806. → 73

Barndorff–Nielsen, O.E., Hansen, P.E., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the prescence of noise. *Econometrica* 76, 6, 1481–536. → 74

Barndorff–Nielsen, O.E., Hansen, P.E., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal* 12, 3, 1–32. → 11, 74

Barndorff–Nielsen, O.E., Hansen, P.E., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi–definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162, 149–69. → 26

Bauwens, L., Pohlmeier, W., and Veredas, D. (2008). *High Frequency Financial Econometrics. Recent Developments.* Heidelberg: Physica–Verlag. →

Bekaert, G. and Wu, G. (1996). Asymmetric volatility and risk in equity markets. *The Journal of Finance* 13,1 1–42. → 43, 197

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics* 3, 167–179. → 42

Black, F. (1976b). Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section* 1, 177–81. → 193, 194

Black, F. and Cox, J. (1976). Valuing corporate securities some effects of bond indenture provisions. *Journal of Finance* 31, 2. → 193

Black, F. and Scholes, M. (1972). The valuation of option contracts and a test of market efficiency. *The Journal of Finance* 27,2 399–417. → 10

Blair, B. J., Poon, S.-H., and Taylor, S. J. (2001a). Asymmetric and crash effects in stock volatility for the s&p100 index and its constituents. *Applied Financial Economics*, 12:319–329. →

Blair, B., Poon, S.H., and Taylor, S.J. (2002). Asymmetric and crash effects in stock volatility for the S&P 100 index and its constituents. *Applied Financial Economics* 12,5 319–29. → 226

Bochner, S. (1949). Diffusion equation and stochastic processes. *Proceedings of the National Academy of Science of the USA* 85,2 369-370. → 10

Bollerslev, T. (1987). Generalized autoregresive conditional heteroskedasticity. *Journal of Econometric* 31, 307–32. → 33

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of returns. *Review of Economics & Statistics* 69, 542–47. → 202

Bollerslev, T., Kretschmer, U., Pigorsch, C., and Tauchen, G. (2009). A discrete-time model for daily S&P500 returns and realized variations: jumps and leverage effectsp500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics*, 150:151–166. 31

Bollerslev, T., Litvinova, J., and Tauchen, G. (2006). Leverage and volatility feedback effects in high frequency data. *Journal of Financial Econometrics* 4,3 353–84. → 61, 64, 71, 91

Bollerslev, T., Li, S. Z., and Todorov, V. (2006). Roughing up beta: Continuous vs. discontinuous betas and the cross–section of expected stock returns. *Journal of Financial Economics* forthcoming. → 101, 255

Bollerslev, T. and Todorov, V. (2010). Jumps and betas: A new framework for disentangling and estimating systematic risks. *Journal of Econometrics* 157, 220-235. → 101, 255

Bollerslev, T. and Todorov, V. (2011). Estimation of jump tails. *Econometrica* 79,3 1727–83. → 63, 70

Boudt, K., Croux, C., and Laurent, S. (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance* 18, 353–67. → 77, 92, 93

Brys, G., Hubert, M., and Struyf, A. (2004). A robust measure of skewness. *Journal of Computational and Graphical Statistics* 13,4 996–1017. → 47, 52, 63

Carr, P. and Wu, L. (2009). Variance Risk Premia. *The Review of Financial Studies* 22 ,3 1311–1341. →

Corsi, F. and Reno, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economic Statistics* 30, 368380. → 31

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31,3 281–318. → 43, 195, 196

Clark, P. K. (1973). A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica* 41, 135–156. → 10, 15

Chen, X. and Ghysels, E. (2011). News—good or bad—and its impact on volatility predictions over multiple horizons. *The Review of Financial Studies* 24,1 46–81. → 61

Cheung, Y.-W. and Ng, L. (1992). Stock price dynamics and firm size: An empirical investigation. *Journal of Finance* 47,1985–97. → 194

Comte, F. and Renault, E. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics* 114, 576-99. → 31

Comte, F. and Renault, E. (1998). Long memory in continuous–time stochastic volatility models. *Mathematical Finance* 8,4 291–323. → 15, 21, 67

Christie, A. A. (1982) The stochastic behavior of common stock variances: value, leverage and interest rate effects. *Journal of Financial Economics* 10, 407–432. → 42, 193, 194, 195, 197

Das, S. R. and Sundaram, R. K. (1999). Of smiles and smirks: A term structure perspective. *Journal of Financial and Quantitative Analysis* 34, 211–39. → 13, 61

Daouk, H. and Ng, D. (2011). Is unlevered firm volatility asymmetric? *Journal of Empirical Finance* 18, 634–651. → 193

Diebold, F. X. and Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent factor arch model. *Journal of Applied Econometrics*, 4:1–21. → 16, 38

Drost, F. and Werker, B. (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics* 74, 31-57. → 6

Duffe, G. (1995). Stock returns and volatility: A firm–level analysis. *Journal of Financial Economics* 37, 399–420. → 43, 194

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump–diffusions. *Econometrica* 6,68 1343–76. → 14, 61, 71

Elandt, R. C. (1961). The folded normal distribution: two methods of estimating parameters from moments. *Technometrics* 3, 4 551–62. → 128

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(987–1007). → 33

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate garch models. *Journal of Business and Economic Statistics*, 20:339–350. → 35, 42

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11:122–150. → 34, 37

Fama, E. F. and Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics* 5, 115–43. → 193

Fan, R., Taylor, S. and Sandri, M.. (2017). Density forecast comparisons for stock prices, obtained from high-frequency returns and daily option prices. *Journal of Futures Market.* → 5

Fang, Y (1996). Volatility modeling and estimation of high–frequency data with Gaussian noise. *Doctoral Thesis, MIT, Sloan School of Management.* → 24, 101

Fiorentini, G., Sentana, E., and Shephard, N. (2004). Likelihood-based estimation of latent generalised arch structures. *Econometrica*, 72:1481–1517. → 16, 38

Figlewski, S. and Wang, X. (2001). Is the 'leverage effect' a leverage effect? *Working Paper, Department of Finance, New York University.* → 195

French, M.W. and Sichel, D.E.(1993). The price variability–volume relationship on speculative markets. *Journal of Business & Economic Statistics* 11,2 113–9. → 198

French, K. R., Schwert, G. W., and Stambaugh, R. (1987). Expected stock returns and volatility. *Journal of Financial Economics* 19, 3–29. → 193, 195, 196

Kim, D. and Kon, S. J. (1994). Expected stock returns and volatility. *Journal of Business* 67, 563–98. → 195

Gallant, A. R. (1987). *Nonlinear Statistical Models.* New York: Wiley. → 74

Ghysels, E., Santa–Clara, P., and Valkanov, R. (2005). There is a risk–return trade-off after all. *Journal of Financial Economics* 76, 509–48. → 87, 196

Ghysels, E. and Plazzi, A., and Valkanov, R. (2012). Conditional skewness in international stock markets: Portfolio implications and economic foundation. *Working Paper, Inquire Europe.* [Available online: http://ssrn.com/abstract=1761446] → 47, 63

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering.* New York: Springer–Verlag Inc. → 80, 116

Glosten, L. R., Jagannathan, R., and Runkle, D.(1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–801. → 7, 34, 194, 196, 201

Gilder, D., Shackleton, M., and Taylor, S. J. (2014). Cojumps in stock prices: empirical evidence. *Journal of Banking and Finance.* 40, 443–459. → 101

Hall, A. R. (2005). *Generalized Method of Moments (Advanced Texts in Econometrics).* Oxford: Oxford University Press. →

Hansen, L. P. (1982). Large sample properties of Generalized Method of Moments estimators. *Econometrica* 50, 1029-59. →

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review* 35, 5 705-30. → 55

Hansen, P. R. and Lunde, A. (2006). A realized variance for the whole day based on intermittent high–frequency data. *Journal of Financial Econometrics* 3,4 525-54. → 5

Hansen, P. R., Lunde, A., and Voev, V. (2014). Realized Beta Garch: A multivariate Garch Model with Realized Measures of Volatility. *Journal of Applied Econometrics* 29,5 774–99. → 7

Harvey, A. C. and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14, 429-34. → 61, 71, 81

Hasbrouck, J. and Sofianos (1993). The Trades of Market Makers: An Empirical Analysis of NYSE Specialists. *The Journal of Finance* 48,5 1565-93. → 254

Hasbrouck, J. (2003). Intraday price formation in U.S. equity index markets. *The Journal of Finance* 58, 2375–2400. → 89

Hong, Y., Tu, J., and Zhou, G. (2007). (1993). Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* 20, 1547–81. → 194

Heston, S. (1993). A closed–form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* 2,6 327–43. → 12, 13, 61

Hsieh, D. (1989). Modeling heteroskedasticity in daily foreign-exchange rates. *J. Business Econom. Statist.* 7 307–17. → 54

Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3,4 456-99. → 11, 27, 30, 31, 73, 80

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42, 281–300. → 12, 34

Jacod, J. (2008). Asymptotic properties of power variations and associated functionals of semimartingales. *Stochastic Processes and Their Applications* 118, 517–559. → 25

Jiang, G. J. and Oomen, R. C. A. (2008). Testing for jumps when asset prices are observed with noise—a 'swap variance' approach. *Journal of Econometrics* 144,2 352–70. → 30, 66

Kim, T.-H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1,1 56–73. → 47, 63

Kim, T.-H., Manganelli, S., and White, H. (2008). Modeling autoregressive conditional skewness and kurtosis with multi–quantile CAViaR *ECB Working Paper Series no 957. Reprinted in Volatility and Time Series Econometrics (2010), ed. by Bollerslev, T., Russell, J., and Watson, M.*. → 47, 63

Knuth, D. E. (1988). *The Art of Computer Programming. Volume II: Semi–numerical algorithms. 3rd Edition.* Boston: Addison-Wesley. → 97

Lahaye, J., Laurent, S. and Neely, C. J. (2011). Jumps, cojumps and macro announcements. *Journal of Applied Econometrics* 26, 893–921. → 77

Li, C. (2003). The skewness premium and the asymmetric volatility puzzle. *Working Paper, University of California Riverside.* 198

Liu, L. Y., Patton, A. and Sheppard, K. (2011). Does anything beat 5–minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics* 187, 293-311. → 101

Lucca, D. O. and Moench, E. (2015). The pre–FOMC announcement drift. *The Journal of Finance* 70, 1 329–71. → 93

Madhavan, A., M. Richardson, and M. Roomans. (1997). Why Do Security Prices Change? *TheReview of Financial Studies* 10, 1 1035–64. → 254

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dellIstituto Italiano degli Attuari* 64, 19–47. →

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* 36, 394–419. → 10

Mandelbrot, B. (1982). A commentary—the variation of certain speculative prices. *Weekly Citation Classic.* → 10

Mayfield, E. S. (2004). Estimating the market risk premium. *Journal of Financial Economics* 73, 3, 465–96. → 197

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics* 17, 479-508. → 21, 68

Meddahi, N. and Renault, E. (2004). Temporal aggregation of volatility models. *Journal of Econometrics* 119, 355–79. → 64, 84

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3, 125-144. → 10

Merton, R. (1980). On estimating the expected return on the market. *Journal of Financial Economics* 8, 323-61. → 246

Mykland, P. A. and Wang, D. C. (2014). The estimation of leverage effect with high frequency data. *American Statistical Association Journal* 109, 505, 197–215. → 64

Mykland, P.A., and L. Zhang (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* 77, 1403–45. → 65, 102

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–70. → 7, 34, 194, 196, 202

Nelson, D. B. and Cao, C. Q. (1992). Inequality constraints in the univariate garch model. *Journal of Business and Economic Statistics*, 10:229–235. → 34

Neuberger, A.(2012). Realized skewness. *The Review of Financial Studies* 25,11 3423–55. → 65

Neumann, M. and Skiadopoulos, G. (2013). Predictable dynamics in higher–order risk–neutral moments: Evidence from the S&P 500 options. *Journal of Financial and Quantitative Analysis* 48,03 947–77. → 66

Newey, W.K. and West, K. D. (1987). A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–8. →

Newey, W.K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–53. →

Nolte, I., Taylor S.J., Zhao, X. (2016). More accurate volatility estimation and forecasts using price durations. *Working paper, Lancaster University.* → 78

NYSE (2013). New York Stock Exchange special closings. *https://www.nyse.com* [Accessed: August 2014.] → 91

Ozdenoren, E. and Yuan, K. (2008). Feedback effects and asset prices. *The Journal of Finance* 63, 4 1939–75. → 198

Patton, A. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics.* 97,3 683–697. → 61

Patton, A. and Verardo, M. (2012). Does beta move with news? Firm–specific information flows and learning about profitability *The Review of Financial Studies.* 25,9 2789–2839. → 101, 255

Pindyck, R. S. (1984). Risk, inflation, and the stock market. *American Economic Review* 74, 334–353. → 195

Poterba, J. M. and Summers, L. H. (1986). The persistence of volatility and stock market fluctuations. *American Economic Review* 76, 1142–51. → 195

Reiss, M., Todorov, V. and Tauchen, G. (2015). Nonparametric Test for a Constant Beta between Ito Semimartingales based on High Frequency Data. *Stochastic Processes and Their Applications* 2955-88 → 248

RENAULT, E., SARISOY, Ç., AND WERKER, B. (2016). Efficient estimation of integrated volatility and related processes. *Econometric Theory* (Forthcoming) → 65

SARISOY, Ç.(2015). Monetary policy risk in the cross–section of expected returns. *Ph.D. Thesis No. 454, Chapter 3.* Tilburg: CentER, Tilburg University. → 93

SCHNEIDER, P.(2012). Fear trading. *Working Paper, University of Lugano* [Available online: http://ssrn.com/abstract=1994454] → 65

SCHWERT, G.S.(1989). Why Does Stock Market Volatility Change Over Time? *The Journal of Finance* 44,5 115-53. → 194, 198

SHEPPARD, K. (2006). Realized Covariance and Scrambling. *Oxford University Working Paper.* → 8, 9

SHIVELY, P. A. (2007). Asymmetric temporary and permanent stock–price innovations. *Journal of Empirical Finance* 14, 120–30. → 199

STEIN, E. M. AND STEIN, J. C. (1991). Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies* 4, 727–52. → 34

TALSEPP, T. AND RIEGER, M. O. (2010). Explaining asymmetric volatility around the world. *Journal of Empirical Finance* 17, 938–56. → 198

TAUCHEN, G. AND PITTS, M.(1983). The price variability–volume relationship on speculative markets. *Econometrica* 51,2 485–505. → 10

Tauchen, G.E. and Zhou, H. (2011). Realized jumps on financial markets and predicting credit spreads. *JJournal of Econometrics*, 160, 102118. 31

TAYLOR, S. J. (1980). Conjectured models for trends in financial prices, tests and forecasts. *Journal of the Royal Statistical Society, Series A.* 143, 338–362. → 11

TAYLOR, S. J. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices, 1962–79. In O.D. Anderseon (ed.), *Time Series Analysis: Theory and Practice* 1, 203–26 Amsterdam: North–Holland → 11, 80

TAYLOR, S. J. (1986). *Modelling Financial Time Series.* Chichester: John Wiley & Sons. → 33, 34, 80

TAYLOR, S. J. AND XU, X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317–40. → 28, 77

TAYLOR, S. J. (2005). *Asset price dynamics, volatility, and prediction.* New Jersey: Princeton University Press. → 24, 61, 65, 67, 213

TAYLOR, S. J. (2015). *Microstructure noise components of the S&P 500 index: variation, persistence and distributions.* Working pa per Lancaster University. → 89

THEODOSSIOU, P. (1998). Financial data and the skewed generalized t distribution. *Management Science* 44,12 1650–661. → 53

Theodossiou, P. and Savva. C. S (2015). Skewness and the relation between risk and return. *Management Science* Articles in Advance, 1-12. → 53, 54

Todorov, V. and Bollerslev, T. (2010). Jumps and Betas: A New Theoretical Framework for Disentangling and Estimating Systematic Risks. *Journal of Econometrics* 157, 220–35. → 248

Tse, Y. and Tsui, A. (2002). A multivariate garch model with time-varying correlations. *Journal of Business and Economic Statistics*, 20:351–362. → 35, 42

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177-188. → 12

Wu, G. (2001). The determinants of asymmetric volatility. *Review of Financial Studies* 14,3 837–59. → 43, 196, 197

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* 127, 165–78. → 81, 102

Zhang, L., Mykland, P., and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 492 1394–411. → 75, 91