



Department of Economics

Three Essays on the Econometric Analysis of High Frequency Financial Data

Roel C.A. OOMEN

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Florence, June 2003

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Three Essays on the Econometric Analysis
of High Frequency Financial Data

Roel C.A. Oomen

The Thesis Committee consists of:

Professor H. Peter Boswijk, University of Amsterdam

Professor Søren Johansen, University of Copenhagen (supervisor)

Professor Helmut Lütkepohl, EUI

Professor Stephen Taylor, Lancaster University

THREE ESSAYS ON THE ECONOMETRIC ANALYSIS OF HIGH FREQUENCY FINANCIAL DATA

Abstract

This thesis is motivated by the observation that the time series properties of financial security prices can vary fundamentally with their sampling frequency. Econometric models developed for low frequency data may thus be unsuitable for high frequency data and vice versa. For instance, while daily or weekly returns are generally well described by a martingale difference sequence, the dynamics of intra-daily, say, minute by minute, returns can be substantially more complex. Despite this apparent conflict between the behavior of high and low frequency data, it is clear that the two are intimately related and that high frequency data carries a wealth of information regarding the properties of the process, also at low frequency. The objective of this thesis is to deepen our understanding of the way in which high frequency data can be used in financial econometrics. In particular, we focus on (i) how to model high frequency security prices, and (ii) how to use high frequency data to estimate latent variables such as return volatility. One finding throughout the thesis is that the choice of sampling frequency is of fundamental importance as it determines both the dynamics and the information content of the data. A more detailed description of the chapters follows below.

Chapter one examines the impact of serial correlation in high frequency returns on the realized variance measure. In particular, it is shown that the realized variance measure yields a *biased* estimate of the conditional return variance when returns are serially correlated. Using 10 years of FTSE-100 minute by minute data we demonstrate that a careful choice of sampling frequency is crucial in avoiding substantial biases. Moreover, we find that the autocovariance structure (magnitude and rate of decay) of FTSE-100 returns at different sampling frequencies is consistent with that of an ARMA process under temporal aggregation. A simple autocovariance function based method is proposed for choosing the “optimal” sampling frequency, that is, the highest available frequency at which the serial correlation of returns has a negligible impact on the realized variance measure. We find that the logarithmic realized variance series of the FTSE-100 index, constructed using an optimal sampling frequency of 25 minutes, can be modelled as an ARFIMA process. Exogenous variables such as lagged returns and contemporaneous trading volume appear to be highly significant regressors and are able to explain a large portion of the variation in daily realized variance.

Chapter two (based on joint work with George Jiang) proposes an unbiased estimator of the latent variables within the Affine Jump Diffusion model. The estimator is model consistent, can be implemented based on high

frequency observations of the state variable, and we derive conditions under which it has minimum variance. In a simulation experiment we illustrate the performance of our estimator and show that its properties compare favorably to commonly used alternatives. Because our approach can, in principle, be applied to any latent variable model in the general Affine Jump Diffusion framework, it covers a wide range of models frequently studied in the finance literature including the stochastic volatility model. Based on the proposed estimator of latent variables, we outline a flexible GMM estimation procedure that relies on the matching of conditional moments or cumulants of both the observed and the unobserved state variables.

Chapter three studies two extensions of the compound Poisson process with iid Gaussian innovations which are able to characterize important features of high frequency security prices. The first model explicitly accounts for the presence of the bid/ask spread encountered in price-driven markets. This model can be viewed as a mixture of the compound Poisson process model by Press and the bid/ask bounce model by Roll. The second model generalizes the compound Poisson process to allow for an arbitrary dependence structure in its innovations so as to account for more complicated types of market microstructure. Based on the characteristic function, we provide a detailed analysis of the static and dynamic properties of the price process. Comparison with actual high frequency data suggests that the proposed models are sufficiently flexible to capture a number of salient features of financial return data including a skewed and fat tailed marginal distribution, serial correlation at high frequency, time variation in market activity both at high and low frequency. The current framework also allows for a comprehensive investigation of the “market-microstructure-induced bias” in the realized variance measure and we find that, for realistic parameter values, this bias can be substantial. We analyze the impact of the sampling frequency on the bias and find that for non-constant trade intensity, “business” time sampling maximizes the bias but achieves the lowest overall MSE.

JOINT RESEARCH STATEMENT CHAPTER 2

June, 2003

The following is a joint statement by George Jiang and Roel Oomen which aims at detailing each author's relative contribution to Chapter two of this thesis; "Latent Variable Estimation for the Affine Jump Diffusion." At the outset it should be stressed that the chapter, as it stands, is entirely written by Roel but draws heavily on a working version of the paper underlying this chapter. The work is in progress and the chapter in this thesis summarizes the results that were derived by December 2002, with minor modifications made afterwards (including some redrafting and updating of simulation results).

Both authors have contributed significantly to the theoretical *and* simulation results. However, George has primarily focussed on the development of the theoretical results while Roel has concentrated on the simulation design and interpretation of the results. More specifically, George has developed the unbiased estimator of the latent variables within the Affine Jump Diffusion framework and derived Lemma 2.2.1 and Corollary 2.2.2. Roel's contribution to the theory is Lemma 2.2.3. Regarding the simulation part, model specification and parameter choice is joint work. George has derived the solutions to the Ricatti equations for both the stochastic mean model and the stochastic volatility model while Roel has written the code for the simulations and proposed the comparison of the UMV filter to the Kalman filter (and particle filter, work in progress). The simulation results have been interpreted jointly leading to several intermediate revisions of model specification, parameter choice, and reporting style. The remaining part of the chapter is joint work.

George Jiang

Finance Department
Eller College of Business & Public Administration
University of Arizona, Tucson AZ, USA

Roel C.A. Oomen

Department of Economics
European University Institute
San Domenico di Fiesole, Italy

Acknowledgment

I owe a measure of gratitude to several people, without whom this thesis would not have been completed.

I have been very privileged to have Søren Johansen as the undoubtedly most intuitive and brilliant advisor anyone could ever ask for. Søren has an almost magical ability to get through to the essence of just about any problem in econometrics with a simple AR(1) process that I will always admire. Over the years I have learned so much from him and many of the results in this thesis would not have been derived if it wasn't for the constant flow of detailed and carefully thought out comments... even if they were scribbled in between the lines of a preliminary draft or on a napkin over lunch...

Bent Sørensen and George Jiang also deserve special mention. Bent was my advisor at Brown and introduced me to the fascinating field of financial econometrics and it is largely due to him that I decided to pursue a Ph.D. degree. George is since several years a dear friend of mine and co-author on several papers, including the second chapter of this thesis. His brilliant insights into finance continue to inspire me and working with him has been, without a doubt, productive, stimulating and a lot of fun!

I would also like to thank Peter Boswijk, Helmut Lutkepohl, and Stephen Taylor for serving on the thesis committee and giving me a lot of useful feedback. Coincidentally, I had the opportunity to discuss the very first paper I worked on in Florence with Peter during a visit to the Netherlands. Now, 5 years later, we meet again... His comments are much appreciated. Several other people have provided comments on parts of this thesis and I would like to thank Torben Andersen, Frank Diebold, Phil Dybvig, Marcelo Fernandes, Ilian Giorgiev, Mark Lauer, Jeff Russell, Juan Toro, Ilias Tsiakas, Bas Werker, and the participants of several conferences.

Throughout my studies, I was supported by the Netherlands Organization for Scientific Research (NWO) who have funded my fellowship at the EUI and awarded me several travel grants which allowed me to discuss parts of my thesis with researchers around the world and, most importantly, visit George on a regular basis. I am very grateful for this funding.

Then I would like to thank Roel in 't Veld for suggesting me to apply to the EUI in December 1996; Jacqueline Bourgonje for being the most helpful and efficient departmental administrator; Frank Diebold for a memorable dinner in Union Square Café back in 1999, stimulating discussions about research in high frequency data, and the many invitations to the NBER meetings; Juan Toro for keeping me busy in Chicago over summer; Jeroen Padt for giving me insights into investment banking at JP Morgan and finance in practice; Ilija Zovko and Doyme Farmer for an invitation to the Santa Fe Institute; Esther Follain for proofreading parts of this thesis; my patient colleagues at Warwick Business School for giving me the opportunity to complete my thesis; and the many friends who have come to visit me in Florence to, um... check-up on my academic progress... and who were always happy to find out that I wouldn't leave bella Italia just yet!

My time in Florence has been a great experience in many ways. The EUI is truly a unique place that is full of bright and diverse people and I feel very fortunate to have been part of that. My decision to leave the USA for Europe was a very difficult one but thanks to Søren it turned out to be a good one. Besides, being in Italy meant late-night dinners, great weather, sea and mountains, Rome and Rignano, amazing views, and the opportunity to spend countless hours exploring the Tuscan country side on a Ducati Monstro. . . ; things that don't take much to get used to.

The travelling I have done over the last couple of years has enabled me to see so many places and meet so many interesting people but it also meant that, ironically, my closest friends, girlfriend, and parents were thousands of kilometers away most of the time. I am very thankful to them for keeping these precious relationships alive by the many visits, e-mails, and telephone calls. In particular, I wish to thank my one and only Ozlem for her energy and devotion to bridge all those years of Atlantic divide between us.

My greatest thanks is reserved for my mother and late father to whom I owe everything good in my life. Although the last few years have been very difficult, thanks to their determination and optimism there have been so many good moments as well. It is sad to realize that my father will not be present for this moment in my life. He is being missed with admiration, a smile, and beautiful memories.

London, June 2003

Roel C.A. Oomen

*Dedicated to my late father Kees,
and my mother Antoinette.*

Contents

1	Modelling Realized Variance when Returns are Serially Correlated	1
1.1	Introduction	1
1.2	Realized Variance	3
1.2.1	Realized Variance in Practice	5
1.2.2	Serial Correlation, Time Aggregation & Sampling Frequency	8
1.3	Modelling Realized Variance	14
1.3.1	Fractional Integration & Realized Variance	15
1.3.2	Empirical Results	17
1.4	Conclusion	21
2	Latent Variable Estimation for the Affine Jump Diffusion	22
2.1	Introduction	22
2.2	The Affine Jump-Diffusion Model with Latent Variables	23
2.2.1	Unbiased Estimators of the Latent Variables	24
2.3	Performance of the Unbiased Minimum-Variance Latent Variable Estimator	29
2.3.1	The Mean-Reverting Stochastic Mean Model	29
2.3.2	The Square-Root Stochastic Volatility Model	37
2.4	Estimation of Latent Variable Models in Finance	47
2.4.1	GMM Estimation based on the Model-Consistent Unbiased Estimator of the Latent Variables	48
2.5	Conclusion	51
3	Statistical Models for High Frequency Security Prices	53
3.1	Introduction	53
3.2	The Bid/Ask Spread	55
3.3	General Return Dependence	61
3.3.1	Multiple Component Compound Poisson	65
3.3.2	Time Varying Trading Intensity	68
3.4	Realized Variance and Return Dependence	75
3.4.1	The “Covariance Bias Term”	76
3.4.2	Bias Reduction and Optimality of Sampling Schemes	77

3.5	Conclusion	80
A		94
A.1	Proofs	94
A.2	Stochastic Mean Model	94
A.2.1	State-Space Representation	94
A.2.2	Covariance Expressions	95
A.3	Stochastic Volatility Model	95
B		97
B.1	The Characteristic Function	97
B.2	The Intensity Process	97
B.3	Proofs	98

Chapter 1

Modelling Realized Variance when Returns are Serially Correlated

1.1 Introduction

A crucial element in the theory and practice of derivative pricing, asset allocation and financial risk management is the modelling of asset return variance. The Stochastic Volatility and the Autoregressive Conditional Heteroskedasticity class of models have become widely established and successful approaches to the modelling of the return variance process in both the theoretical and the empirical literature (see for example Bollerslev, Engle, and Nelson (1994) and Ghysels, Harvey, and Renault (1996)). Despite the enormous amount of research on return variance modelling carried out over the past two decades, complemented with overwhelming empirical evidence on the presence of heteroskedastic effects in virtually all financial time series, the variety of competing variance models highlights the disagreement on what the correct model specification should be. An alternative route to identifying the dynamics of the return variance process is to utilize the information contained in option prices. Yet, also here, several studies have documented a severe degree of model misspecification even for the more general option pricing formulas that incorporate stochastic volatility, interest rates and jumps (see for example Bakshi, Cao, and Chen (1997)). It is therefore not surprising that a growing number of researchers have turned their attention to the use of high frequency data which, under certain conditions, allow for an essentially non-parametric or model-free approach to the measurement of return variance. The objective of this paper is twofold. First, explore the extent to which the now widely available intra-day data on financial asset prices can be used to improve and facilitate the estimation and modelling of return variance. Special attention is given to the impact that market microstructure-induced serial correlations, present in returns sampled at high frequency, have on the resulting variance estimates. Second, analyze and model the time series of estimated (daily) return variance. Here the focus is on identifying a suitable model plus a set of exogenous variables that is able to characterize and explain variation in the return variance.

The idea of inferring the unobserved return variance from high frequency data is not new. In fact, it can be traced back to Merton (1980) who notes that the variance of a time-invariant Gaussian diffusion process (over a

fixed time-interval) can be estimated arbitrarily accurately as the sum of squared realizations, provided that the data are available at a sufficiently high sampling frequency. Empirical studies making use of this insight include French, Schwert, and Stambaugh (1987), who estimate monthly return variance as the sum of squared daily returns and Andersen and Bollerslev (1998), Hsieh (1991), and Taylor and Xu (1997) who estimate daily return variance as the sum of squared intra-day returns. More recent studies that apply and develop this idea further include Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2001, 2003), Areal and Taylor (2002), Barndorff-Nielsen and Shephard (2002, 2003), Blair, Poon, and Taylor (2001), Maheu and McCurdy (2002b), and Martens (2002).

One of the main attractions that has been put forward of estimating return variance by the sum of squared intra-period returns, a measure commonly referred to as “realized variance” (or “realized volatility” being the square root of realized variance), is that this approach does not require the specification of a potentially misspecified parametric model. In addition, when constructing the realized variance measure there is no need to take the widely documented and pronounced intra-day variance pattern of the return process into account. This feature contrasts sharply with parametric variance models which generally require the explicit modelling of intra-day regularities in return variance (see for example Engle (2000)). Finally, calculating realized variance is straightforward and can be expected to yield accurate variance estimates as it relies on large amounts of intra-day data. The theoretical justification for using the realized variance measure has been provided in a series of recent papers by Andersen, Bollerslev, Diebold, and Labys (2001, 2003, ABDL hereafter). In particular, ABDL have shown that when the return process follows a special semi-martingale, the Quadratic Variation (QV) process is the dominant determinant of the conditional return variance. By definition, QV can be approximated by the sum of squared returns at high sampling frequency, or in other words realized variance. Moreover, under certain restrictions on the conditional mean of the process, QV is the single determinant of the conditional return variance, thereby underlining the importance of the realized variance measure. In related work, Barndorff-Nielsen and Shephard (2003) derive the limiting distribution of realized power variation, that is the sum of absolute powers of increments (i.e. returns) of a process, for a wide class of SV models. It is important to note that, in contrast to conventional asymptotic theory, here, the limit distribution results rely on the concept of “in-fill” or “continuous-record” asymptotics, i.e. letting the number of observations tends to infinity while keeping the time interval fixed. In the context of (realized) variance estimation, this translates into cutting up, say, the daily return into a sequence of intra-day returns sampled at an increasingly high frequency (see for example Foster and Nelson (1996)).

The recently derived consistency and asymptotic normality of the realized variance measure greatly contribute to a better understanding of its properties and, in addition, provide a formal justification for its use in high frequency data based variance measurement. However, a major concern that has largely been ignored in the literature so far, is that in practice the applicability of these asymptotic results is severely limited for two reasons. First, the amount of data available over a fixed time interval is bounded by the number of transactions recorded. Second, the presence of market microstructure effects in high frequency data potentially invalidate the asymptotic results.

This paper studies the properties of the realized variance in the presence of market microstructure-induced serial correlation. In particular, we show that the realized variance measure is a biased estimator of the conditional return variance when returns are serially correlated. The return dependence at high sampling frequencies is

analyzed using a decade of minute by minute FTSE-100 index returns. We find that the autocovariance structure (magnitude and rate of decay) of returns at different sampling frequencies is consistent with that of an ARMA process under temporal aggregation. Based on this finding, an autocovariance based method is proposed to determine the “optimal” sampling frequency of returns, that is, the highest available frequency at which the market microstructure-induced serial correlations have a negligible impact on the realized variance measure¹.

Following the methodology outlined above, we find that the optimal sampling frequency for the FTSE-100 data set lies around 25 minutes. We construct a time series of daily realized variance, confirm several styled facts reported in earlier studies, and find that the logarithmic realized variance series can be modelled well using an ARFIMA specification. Exogenous variables such as lagged returns and contemporaneous trading volume appear to be highly significant regressors, explaining a large portion of the variation in daily realized variance. While the regression coefficients of lagged returns indicate the presence of Black’s leverage effect, there is no indication of reduced persistence in the return variance process upon inclusion of contemporaneous trading volume. This latter finding is in sharp contrast with the study by Lamoureux and Lastrapes (1990).

The remainder of this paper is organized as follows. Section 1.2 investigates the impact of serial correlation in returns on the realized variance measure. Here, results on temporal aggregation of an ARMA process are used to characterize the bias of the realized variance measure at different sampling frequencies. Section 1.3 reports the empirical findings based on the FTSE-100 data set while Section 1.4 concludes.

1.2 Realized Variance

The notion of realized variance, as introduced by ABDL, is typically discussed in a continuous time framework where logarithmic prices are characterized by a semi-martingale. More restrictive specifications have been considered by Barndorff-Nielsen and Shephard (2002, 2003). In this setting, the quadratic variation (QV) of the return process can be consistently estimated as the sum of squared intra-period returns. It is this measure that is commonly referred to as realized variance. Importantly, ABDL show that QV is the crucial determinant of the conditional return (co-) variance thereby establishing the relevance of the realized variance measure. In particular, when the conditional mean of the return process is deterministic or a function of variables contained in the information set, the QV is in fact equal to the conditional return variance which can thus be estimated consistently as the sum of squared returns. Notice that this case precludes random *intra*-period evolution of the instantaneous mean. However, it is argued by ABDL that such effects are likely to be trivial in magnitude and that the QV therefore remains the dominant determinant of the conditional return variance.

Below we analyze the impact of serial correlation in returns on the realized variance measure. As opposed to ABDL and Barndorff-Nielsen and Shephard (2002, 2003), a simple discrete time model for returns is used for the sole reason that it is sufficient to illustrate the main ideas. In what follows, the period of interest is set to one day.

Let $S_{t,j}$ ($j = 1, \dots, N$) denote the j^{th} intra day- t logarithmic price of security S . At sampling frequency

¹Independent work by Andersen, Bollerslev, Diebold, and Labys (2000b), Corsi, Zumbach, Müller, and Dacorogna (2001) have proposed a similar approach to determine the optimal sampling frequency. Other related studies include Aït-Sahalia and Mykland (2003), Andreou and Ghysels (2001), Bai, Russell, and Tiao (2001).

f , assuming equi-time spaced² observations, $N_f = \frac{N}{f}$ intra-day returns can be constructed as $R_{f,t,i} = S_{t,if} - S_{t,(i-1)f}$, for $i = 1, \dots, N_f$ and $S_{t,0} = S_{t-1,N}$. By the additive property of returns, it follows that the day- t return is given by:

$$R_t = \sum_{i=1}^{N_f} R_{t,f,i}.$$

We assume that the (excess) return follows a *martingale difference sequence* and that its conditional distribution, i.e. $R_{t,f,i} | \mathcal{F}_{t,f,(i-1)}$ where $\mathcal{F}_{t,f,j}$ denotes the information set available up to the j^{th} period of day t , is symmetric. The need for this symmetry assumption will become clear later on. While this specification allows for deterministic and stochastic fluctuations in the return variance, it also implies that returns are necessarily uncorrelated. Let $V_1 \equiv R_t^2$, i.e. the squared day- t return, and $V_2 \equiv \sum_{i=1}^{N_f} R_{t,f,i}^2$, i.e. the sum of squared *intra-day- t* returns sampled at frequency f . In the current context, V_2 is referred to as the realized variance measure. Since returns are serially uncorrelated at any given frequency f , it follows that:

$$V [R_t | \mathcal{F}_t] = E [R_t^2 | \mathcal{F}_t] = E \left[\sum_{i=1}^{N_f} R_{t,f,i}^2 | \mathcal{F}_t \right], \quad (1.1)$$

where \mathcal{F}_t denotes the information set available prior to the start of day t . Realized variance, like squared daily return, is therefore an *unbiased estimator of the conditional return variance*. However, it turns out that the variance of V_2 is strictly smaller than the variance of V_1 and is therefore the preferred estimator. To see this, it is sufficient to show that $E [V_2^2 | \mathcal{F}_t] < E [V_1^2 | \mathcal{F}_t]$:

$$E [V_1^2 | \mathcal{F}_t] = E \left[\sum_i \sum_j \sum_k \sum_m R_{t,f,i} R_{t,f,j} R_{t,f,k} R_{t,f,m} | \mathcal{F}_t \right] = E \left[\sum_i R_{t,f,i}^4 + 3 \sum_i \sum_{j \neq i} R_{t,f,i}^2 R_{t,f,j}^2 | \mathcal{F}_t \right],$$

because the cross product of returns is zero except when (i) $i = j = k = m$, (ii) $i = j \neq k = m$, (iii) $i = k \neq j = m$, (iv) $i = m \neq j = k$. Notice that $E [R_{t,f,i} R_{t,f,j}^3 | \mathcal{F}_t] = 0$ for $i > j$ by the martingale difference assumption and $E [R_{t,f,i} R_{t,f,j}^3 | \mathcal{F}_t] = 0$ for $i < j$ by symmetry of the conditional distribution of returns. On the other hand

$$E [V_2^2 | \mathcal{F}_t] = E \left[\sum_i R_{t,f,i}^4 + \sum_i \sum_{j \neq i} R_{t,f,i}^2 R_{t,f,j}^2 | \mathcal{F}_t \right],$$

from which it directly follows that

$$V [V_2 | \mathcal{F}_t] < V [V_1 | \mathcal{F}_t].$$

The conditional return variance over a fixed period can thus be estimated arbitrarily accurate by summing up squared intra-period returns sampled at increasingly high frequency. While this result does not depend on the choice of period (i.e. one day), it does crucially rely on the property that returns are serially uncorrelated at any sampling frequency. The additional symmetry assumption rules out any feedback effects from returns into the conditional third moment of returns but allows for skewness in the unconditional return distribution. Other than that, weak conditions are imposed on the return process. As mentioned above, the specification of the return

²This can straightforwardly be generalized to irregularly time spaced returns.

dynamics is sufficiently general so as to allow for deterministic and stochastic fluctuations in the return variance and, as a result, encompasses a wide class of variance models.

1.2.1 Realized Variance in Practice

The results above suggest that straightforward use of high frequency returns can reduce the measurement error in the return variance estimates provided that the return series is a martingale difference sequence (with a symmetric conditional return distribution). This section focuses on the implementation and potential pitfalls that may be encountered in practice. In particular, minute by minute FTSE-100 index level data³ are used to investigate whether the method of calculating the daily realized variance measure will yield satisfactory results. The additive property of returns allows us to decompose the squared daily return as:

$$R_t^2 = \left[\sum_{i=1}^{N_f} R_{f,t,i} \right]^2 = \sum_{i=1}^{N_f} R_{f,t,i}^2 + 2 \sum_{i=1}^{N_f-1} \sum_{j=i+1}^{N_f} R_{f,t,i} R_{f,t,j}. \quad (1.2)$$

It is clear that when the returns are serially uncorrelated at sampling frequency f , the second term on the right hand side of expression (1.2) is zero in expectation and the realized variance measure constitutes an unbiased estimator of the conditional return variance. However, when returns are serially correlated the cross product of returns may not vanish in expectation which, in turn, introduces a bias into the realized variance measure. In particular, when returns are positively (negatively) correlated⁴, the sum of squared intra-day returns will *under-estimate* (*over-estimate*) daily conditional return variance as the cross multiplication of returns will be positive (negative) in expectation.

At first sight, the practical relevance of this finding seems to be challenged by the efficient markets hypothesis which claims that the presence of significant serial correlation in returns, if any, is unlikely to persist for extended periods of time. It is important to note, however, that the efficient markets hypothesis concerns *economic* and not statistical significance of serial correlation. Therefore, due to the presence of market microstructure⁵ effects and transaction costs, a certain degree of serial correlation in returns does not necessarily conflict with market efficiency.

In the market microstructure literature, a prominent hypothesis that is able to rationalize serial correlation in stock index returns is non-synchronous trading. The basic idea is that when individual securities in an index do not trade simultaneously, the contemporaneous correlation among returns induces serial correlation in the index returns. Intuitively, when the index components non-synchronously incorporate shocks to a common factor that is driving their price, this will result in a sequence of correlated price changes at the aggregate or index price level.

³I thank *Logical Information Machines, Inc.* who kindly provided the data needed for the analysis. The data set contains minute by minute data on the FTSE-100 index level, starting May 1, 1990 and ending January 11, 2000. For each day, the data is available from 8:35 until 16:10 (except for the period from July 17, 1998 until September 17, 1999 during which the data is available from 9:00 until 16:10). The total number of observations exceeds one million.

⁴When returns exhibit both positive and negative serial correlation, the effect is not clear. The realized variance measure may be biased or unbiased depending on the relative magnitudes of the return autocovariance at different orders.

⁵For an in depth discussion of the relation between market microstructure and price dynamics see for instance Campbell, Lo, and MacKinlay (1997), Lequeux (1999), Madhavan (2000), O'Hara (1995), Wood (2000) and references therein.

As discussed by Lo and MacKinlay (1990), non-synchronous trading induces positive serial correlation in the index returns. On the other hand, the Roll (1984) bid/ask bounce hypothesis often applies to single asset returns which are typically found to exhibit negative serial correlation. Here the argument is as follows: when at a given point in time no new information arrives in a (dealer) market, the stock price is expected to bounce between the bid and the ask price whenever a trade occurs. Although this phenomenon may not be apparent at a daily or weekly frequency, it is likely to have a discernible impact on returns sampled at high (intra-day) frequency. Finally, transaction costs and feedback trading, in addition to non-synchronous trading and the bid-ask bounce, may also induce serial correlation in returns. For an empirical investigation of these issues see for example Säfvenblad (2000). Although this paper does not aim to analyze the various market microstructure effects in specific, we do want to highlight the presence of such effects and study their impact on the realized variance measure.

Several studies have encountered the impact of serial correlation in returns on the estimates of return variance. For example, French and Roll (1986) find that stock return variance is much lower when estimated using hourly instead of daily data, indicating the presence of positive serial correlation in their data set. Recognizing the presence of serial correlation, French, Schwert, and Stambaugh (1987) estimate monthly return variance as the sum of squared daily returns *plus* twice the sum of the products of adjacent returns. Froot and Perold (1995) also find significant positive serial correlation in 15 minute returns on S&P500 cash index from 1983-1989 and show that the annualized return variance estimates based on weekly data are significantly higher (about 20%) than the variance estimates based on 15-minute data. More recently, Andersen, Bollerslev, Diebold, and Labys (2000b) document the dependence of the realized variance measure on return serial correlation.

These findings offer an early recognition of the central idea of this paper: the results derived in the previous section, and the consistency and asymptotic results derived in ABDL and Barndorff-Nielsen and Shephard (2003), are not applicable to return data that exhibit a substantial degree of serial dependence. In particular, the conditional mean specification used in these studies does typically not allow for the random intra-day evolution of the conditional mean⁶. It is commonly argued that this flexibility is not required at low, say daily or weekly, frequencies. However, when moving to higher intra-day sampling frequencies, the characteristics of the data may change dramatically due to the presence of market microstructure which in turn, leads to substantial dependence in the conditional mean of the return process.

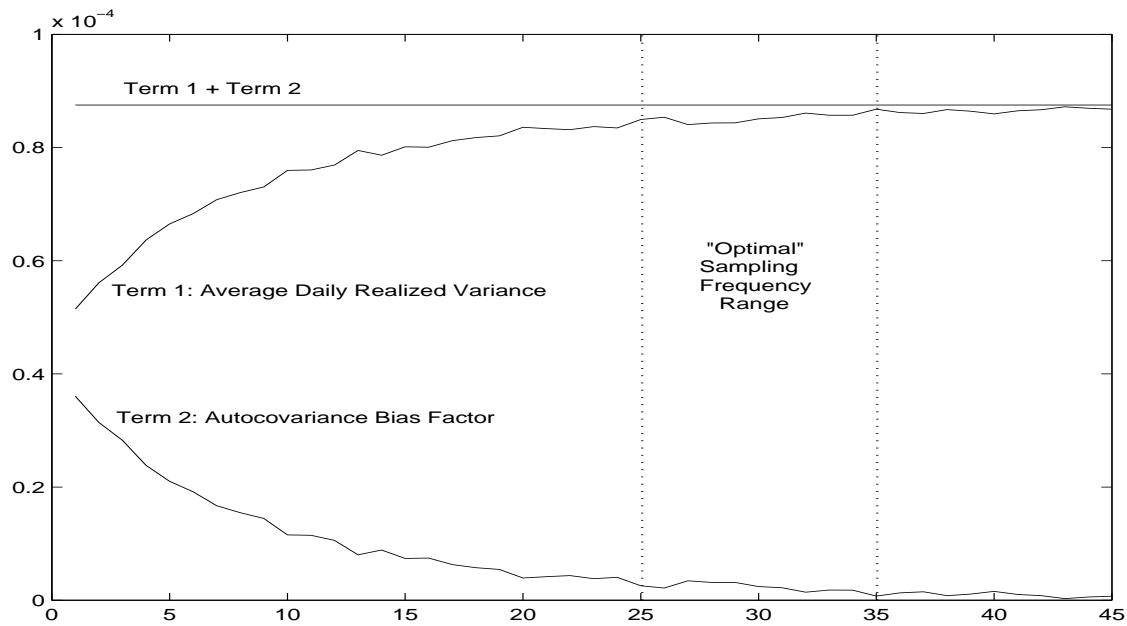
Because market microstructure effects are present in virtually all financial return series, the issue outlined above is central to the discussion of high frequency data based variance measurement. This is emphasized in the empirical analysis which is based on minute by minute returns on the FTSE-100 stock market index. Specifically, the 10 year average (1990-2000) of the two terms on the right hand side of expression (1.2) is computed for sampling frequencies between 1 and 45 minutes and the results are displayed in Figure 1.1. The implicit assumption we make here is that the return process is weakly stationary⁷ so that the averaging (over time) is justified and the estimates can be interpreted as (co)variance estimates.

It is clear that for FTSE-100 data the first term, the realized variance measure, increases with a decrease in

⁶An exception is the general model covered by Theorem 1 in Andersen, Bollerslev, Diebold, and Labys (2003) from which it is also clear that the realized variance measure yields a biased estimate of the conditional return variance.

⁷For the bootstrap analysis of Section 1.2.2 we need to impose strict stationarity and weak dependence on the return process.

FIGURE 1.1: REALIZED VARIANCE VERSUS SAMPLING FREQUENCY



Notes: The 1990-2000 average daily FTSE-100 realized variance (i.e. $T^{-1} \sum_{t=1}^T \sum_{i=1}^{N_f} R_{f,t,i}^2$) and autocovariance bias factor (i.e. $2T^{-1} \sum_{t=1}^T \sum_{i=1}^{N_f-1} \sum_{j=i+1}^{N_f} R_{f,t,i} R_{f,t,j}$) for sampling frequencies between 1 and 45 minutes.

sampling frequency while the second term, the summation of cross multiplied returns, decreases. The positivity of the second term indicates that the FTSE-100 returns are positively correlated, introducing a *downward bias* into the realized variance measure, while its decreasing pattern demonstrates that this dependence, and consequently the bias, diminishes when sampling is done less frequently. This term, which measures the bias that is introduced by the serial dependence of returns, is referred to as the “*autocovariance bias factor*” in the remainder of this paper. Figure 1.1 illustrates that an ad hoc choice of sampling frequency can lead to a substantial (downward) bias in the realized variance measure. In fact, at the highest available sampling frequency of 1 minute, the bias in the variance estimate is more than 35%! To stress the economic significance of this finding, we notice that in a Black Scholes world, a mere 10% under-estimation of the return variance leads to a 14.5% underpricing of a 3 month, 15% out of the money option. Also the option’s delta is 8.2% lower than its true value. Indeed, Figlewski (1998) finds that an accurate return variance estimate is of single most importance when hedging derivatives. When the return variance is stochastic, Jiang and Oomen (2002) also find that for the hedging of derivatives accurate estimation of the *level* of return variance is far more important than accurate estimation of the dynamic parameters of the variance process. Pricing and hedging options aside, it is easy to think of a number of other situations where accurate return variance estimates are of crucial importance. Risk managers often derive Value at Risk figures from the estimated return variance of a position. Also, in a multivariate setting, the covariance matrix of returns is the primary input for portfolio choice and asset allocation.

The above discussion naturally leads to the important question at which frequency the data should be sampled. Figure 1.1 plays a central role in answering this question by providing a graphical depiction of the trade-off one faces when constructing the realized variance measure: an increase in the sampling frequency yields a greater

amount of data, thereby attaining higher levels of efficiency (in theory), while at the same time a decrease in the sampling frequency mitigates the biases due to market microstructure effects surfacing at the highest sampling frequencies. A balance must be struck between these opposing effects and it is argued here that an autocovariance based method, such as the autocovariance bias factor of Figure 1.1, can be used to determine *the “optimal” sampling frequency as the highest available sampling frequency for which the autocovariance bias term is negligible*⁸. Clearly, deciding whether the bias term is “negligible”, and whether the sampling frequency is therefore “optimal”, may prove a difficult issue for at least two reasons. First, even though it may be possible to bootstrap confidence bounds around the autocovariance bias factor in order to determine the frequency at which the bias is statistically indistinguishable from zero (see Table 1.1 for some related results), for many applications economic significance, as opposed to statistical significance, may be the relevant metric with which to measure “negligibility”. The optimal sampling frequency may therefore very well depend on the particular application at hand. Second, when aggregating returns, a reduction in bias should generally be weighed against the loss in efficiency. In practice, however, both the loss or gain in bias and efficiency will often be difficult to quantify which, in turn, complicates the choice of optimal sampling frequency. It should be noted that for a general SV model, Barndorff-Nielsen and Shephard (2003) have shown that the realized variance measure converges to integrated variance at rate \sqrt{N} where N is the number of intra-period observations. Also, Oomen (2003) has derived an explicit characterization of the bias term as a function of the sampling frequency when the price process follows a compound Poisson process with correlated innovations. While the results in these studies may yield some valuable insights into the bias-efficiency trade-off, it is important to keep in mind that they are derived under potentially restrictive parametric specifications for the price process. As such, they should be interpreted cautiously when applied to high frequency data which, as we show below, are often contaminated by market microstructure effects. Without further going into this, it seems reasonable to expect that for the FTSE-100 data the optimal sampling frequency lies somewhere between 25 and 35 minutes, i.e. the range indicated in Figure 1.1.

1.2.2 Serial Correlation, Time Aggregation & Sampling Frequency

We now take a closer look at the autocovariance bias term and show how its shape is intimately related to the dynamic properties of intra-day returns at different sampling frequencies.

Table 1.1 reports some standard descriptive statistics for the FTSE-100 return data. Because it is well known that financial returns, and in particular high frequency returns, are not independently and identically distributed we bootstrap the confidence bounds around the statistics instead of deriving them from the well known asymptotic distributions that are valid under the iid null hypothesis. For the return volatility and the skewness and kurtosis coefficients we use the stationary bootstrap of Politis and Romano (1994) who show that this procedure is valid for strictly stationary, weakly dependent data. Let $\mathbf{x} = (x_1, \dots, x_N)$ denote the original data set (i.e. time series of returns at a given sampling frequency) and let $X_{i,k} \equiv (x_i, \dots, x_{i+k-1})$ where $i = 1, \dots, N$, $k = 1, 2, \dots$, and $x_j = x_{j \bmod N}$ for $j > N$. A bootstrap sample is constructed as $\mathbf{x}^* = (X_{i_1, k_1}, \dots, X_{i_b, k_b})$ where

⁸Independent work by Andersen, Bollerslev, Diebold, and Labys (2000b), Corsi, Zumbach, Müller, and Dacorogna (2001) have proposed a similar approach to determine the optimal sampling frequency.

TABLE 1.1: DESCRIPTIVE STATISTICS OF FTSE-100 RETURNS.

Frequency	No. Obs.	Volatility	Skewness	Kurtosis	BL[15]		
1 Min	1,046,862	11.5 (10.9;12.2)	1.61* (-14.9;17.0)	3305* (1209;5333)	26874* (312.4)		
5 Min	209,372	13.2 (12.6;14.0)	-0.43* (-6.82;5.18)	508.3* (177.5;886.2)	4844.2* (223.6)		
10 Min	104,686	14.2 (13.4;15.1)	-1.89* (-8.50;3.04)	344.8* (78.5;697.2)	1206.9* (138.0)		
30 Min	34,895	15.1 (14.2;16.4)	-1.11* (-4.61;1.51)	115.2* (27.1;234.8)	221.1* (181.6)		
60 Min	17,447	15.6 (14.3;17.3)	-0.32* (-2.71;1.86)	84.3* (14.3;161.5)	117.2* (146.2)		
1 Day	2,407	15.2 (14.0;16.6)	0.05 (-0.21;0.32)	5.43* (4.37;6.56)	44.34* (37.3)		
	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_{10}	ρ_{15}
1 Min	10.4* (9.06;11.7)	7.39* (6.22;8.55)	6.09* (5.08;7.12)	5.13* (3.89;6.68)	2.75* (1.77;3.68)	0.59* (-0.43;1.34)	0.95* (0.32;1.69)
5 Min	14.6* (12.4;16.7)	3.35* (1.92;4.75)	1.85* (0.70;3.05)	0.78* (-0.29;2.00)	0.96* (0.15;1.81)	-0.10 (-1.23;1.32)	-0.60* (-1.57;0.35)
10 Min	10.2* (8.40;12.1)	2.35* (0.94;3.83)	1.06* (0.11;2.07)	0.04 (-1.45;1.65)	-0.55 (-0.20;0.12)	0.17 (-0.71;1.13)	-0.68* (-2.72;0.82)
30 Min	6.39* (4.25;8.68)	-0.60 (-2.48;1.24)	1.28* (-0.58;3.19)	-0.88 (-2.21;0.50)	-0.10 (-3.57;2.58)	2.02* (-0.10;4.46)	-1.83* (-6.35;2.00)
60 Min	3.28* (0.12;6.41)	-0.79 (-4.59;2.58)	3.66* (1.16;6.04)	-0.00 (-2.40;2.41)	2.84* (-0.14;6.26)	-1.55* (-3.03;0.07)	-0.26 (-3.27;2.64)
1 Day	7.56* (3.64;11.2)	-3.62 (-8.45;1.64)	-3.70 (-8.76;1.63)	-0.46 (-5.69;4.73)	-2.44 (-8.20;3.20)	3.33 (-2.26;8.73)	1.84 (-3.23;6.75)

Notes: The upper panel reports the *annualized* return volatility in percentage points (“Volatility”), the skewness coefficient (“Skewness”), kurtosis coefficient (“Kurtosis”), and the Box-Ljung test statistic on the first 15 autocorrelations (“BL[15]”) for FTSE-100 returns sampled at frequencies between 1 minute and 1 day over the period 1990-2000. The lower panel reports the serial correlation coefficients in percentage points (ρ_k denotes the k^{th} order correlation coefficient). Bootstrapped 95% confidence bounds (and critical values for the Box-Ljung test) are reported in parentheses below. An asterisk indicates significance at 95% confidence level under the null hypothesis that returns are iid distributed; \sqrt{T} Skewness $\xrightarrow{a} \mathcal{N}(0, 6)$, \sqrt{T} Kurtosis $\xrightarrow{a} \mathcal{N}(0, 24)$, $\sqrt{T}\rho_k \xrightarrow{a} \mathcal{N}(0, 1)$, and $BL[K] \xrightarrow{a} \chi^2_K$.

$\sum_{j=1}^b k_j = N$, i has a discrete uniform distribution on $\{1, \dots, N\}$, and k has a geometric distribution, i.e. $P(k = m) = p(1 - p)^{m-1}$ for $m = 1, 2, \dots$. Based on this re-sampled time series, we then compute the relevant test statistics. By simulating a large number B of bootstrap samples we can approximate the true distribution of the test statistics by the empirical distribution of the B values of the associated statistics. The idea behind sampling blocks instead of single entries is that, when the block length is sufficiently large, the dependence structure of the original series will be preserved in the re-sampled series to a certain extent. Evidently, the correspondence between the distribution of the original and the re-sampled series will be closer the weaker the dependence and the longer the block length. To choose p , or equivalently the expected block length $E[k] = 1/p$, we have experimented with a number of different values but find, in line with several other studies (Horowitz, Lobato, Nankervis, and Savin 2002, Romano and Thombs 1996), that the results are rather insensitive to the choice of p . The results reported in Table 1.1 are based on $p = 1/15$ (i.e. $E[k] = 15$) and $B = 5,000$.

The confidence intervals for the correlation coefficients and the critical value the Box-Ljung test statistic are obtained by the “blocks-of-blocks” bootstrap. Instead of sampling a $1 \times k$ block, as is done in the stationary bootstrap, we now sample an $h \times k$ block $X_{i,k,h} = (x_i^h, \dots, x_{i+k-1}^h)$ where $x_i^h = (x_i, \dots, x_{i+h-1})'$ and $h - 1$ matches the maximum order of correlation coefficient to be computed. Analogous to the procedure described above, an $h \times N$ bootstrap sample is constructed as $\mathbf{x}^* = (X_{i_1,k_1,h}, \dots, X_{i_b,k_b,h})$ from which the k^{th} order correlation coefficients can be computed as

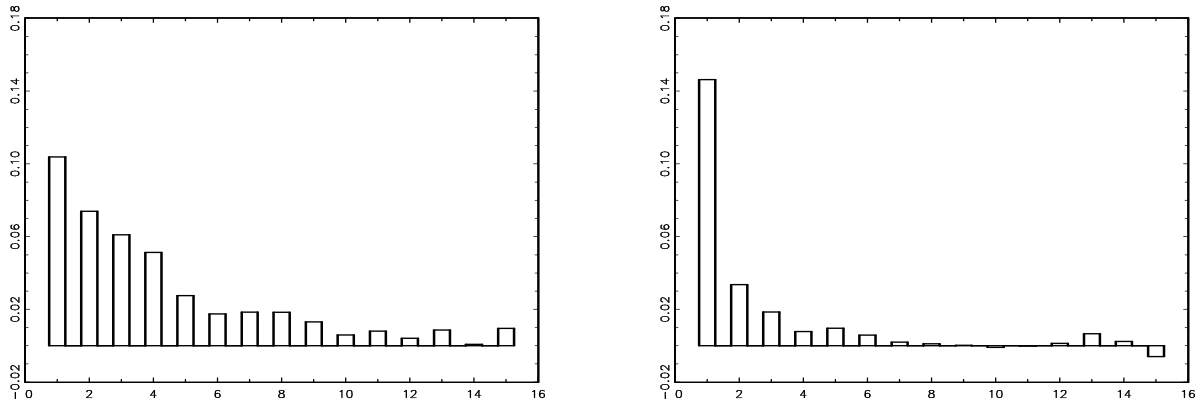
$$\hat{\rho}_k = \frac{\sum_{i=1}^N (\mathbf{x}_{1,i}^* - \bar{\mathbf{x}}_{1,\cdot}^*) (\mathbf{x}_{k+1,i}^* - \bar{\mathbf{x}}_{k+1,\cdot}^*)}{\left[\sum_{i=1}^N (\mathbf{x}_{1,i}^* - \bar{\mathbf{x}}_{1,\cdot}^*)^2 \sum_{i=1}^N (\mathbf{x}_{k+1,i}^* - \bar{\mathbf{x}}_{k+1,\cdot}^*)^2 \right]^{1/2}}$$

where $\bar{\mathbf{x}}_{i,\cdot}^* = N^{-1} \sum_{j=1}^N \mathbf{x}_{i,j}^*$. Because the null-hypothesis for the Box-Ljung statistic is uncorrelatedness, we first pre-whitened the data using an AR(15)⁹ and implement the bootstrap procedure on the residuals. As above, the geometric parameter and the number of bootstrap replications are set as $p = 1/15$ and $B = 5,000$. For more details on how to approximate the sampling distribution of the correlation coefficients and the Box-Ljung statistics using the (blocks-of-blocks) bootstrap see for example Davison and Hinkley (1997), Horowitz, Lobato, Nankervis, and Savin (2002) and Romano and Thombs (1996).

Based on the above bootstrap procedures we construct 95% confidence bounds for the descriptive statistics under the null that returns are weakly dependent and report them in parentheses in Table 1.1. The statistics that are significant are printed in bold. For comparison purposes, an asterisk indicates 95% significance under the alternative null hypothesis that returns are independently and identically distributed. For this case, it is well known that the square root of the sample size times the k^{th} order serial correlation, skewness, and kurtosis coefficients of returns are asymptotically distributed as normal with variance 1, 6, and 24 respectively. The Box-Ljung statistic on the first K autocorrelations, $BL[K]$, is asymptotically distributed as chi-square with K degrees of freedom. Turning to the results in Table 1.1, we find that there is substantial excess kurtosis and serial correlation in high frequency returns. At the minute frequency, most of the serial correlation coefficients up to order 15 are significant and the kurtosis coefficient indicates the presence of an extremely fat tailed marginal return distribution. However, aggregation of returns brings the distribution of returns closer to normal and reduces both the order and magnitude of the serial correlation (see also Figure 1.2). At the daily frequency, the excess kurtosis has come down from around 3000 to about 2.5, and the serial correlation coefficients of order higher than one are all insignificantly different from zero. Consistent with the autocovariance bias term above (Figure 1.1), we also see that the (annualized) return volatility increases with a decrease of the sampling frequency. Interestingly, the 95% confidence bounds for frequencies lower than 30 minutes (i.e. 1, 5, and 10 minutes) do not include the point estimate of the annualized return variance based on daily data. This suggests that the autocovariance bias term at these frequencies is statistically different from zero which, in turn, corroborates our choice of “optimal” sampling frequency range on statistical grounds.

⁹We note that the choice of AR-order is relatively ad hoc, and could arguably be lowered with a decrease in sampling frequency. However, with the amount of data we work with here, it can be expected that the efficiency loss associated with the potentially redundant AR-terms is minimal. Hence, for simplicity, we keep the AR-order fixed across the different sampling frequencies.

FIGURE 1.2: CORRELOGRAM OF 1 MINUTE AND 5 MINUTE FTSE-100 RETURNS



Notes: Correlogram of 1 minute (left panel) and 5 minute (right panel) FTSE-100 index returns for the period 1990-2000

It is also clear from Table 1.1 that the bootstrapped confidence bounds deviate substantially from their iid-asymptotic counterparts. As a result, a number of statistics that are significant under the (invalid) iid null hypothesis, turn out to be insignificant based on the bootstrapped confidence bounds which allows for weak dependence in the return data. For example, while the skewness of intra-daily returns is significant under the iid hypothesis, none of the skewness coefficients are significant under the alternative null-hypothesis. Also, the maximum order of the significant correlation coefficients is generally lower for the bootstrapped critical values than for the iid-asymptotic values. For example, at frequencies between 5 and 30 minutes, ρ_{15} is found significant under the iid-hypothesis but insignificant under the alternative hypothesis. These findings emphasize the inadequacy of the “iid-” asymptotic distributions for this data and illustrate the value of the bootstrap method.

Turning to the specification of the return process, we notice that the overwhelming significance of the serial correlation coefficients reported in Table 1.1 and Figure 1.2 suggests that the characteristics of *intra*-day returns are not consistent with those of a martingale difference sequence. Instead, modelling intra-day returns as an ARMA¹⁰ process is a natural and, as it turns out, successful approach for it is well suited to account for the serial dependence of returns at various sampling frequencies. From a market microstructure point of view, the AR part will arguably be able to capture any autocorrelation induced by non-synchronous trading while the MA part will account for potential negative first order autocorrelation induced by the bid-ask bounce. Further, the decreasing order and magnitude of serial correlation with the sampling frequency is, as it turns out, a consequence of temporal aggregation of the return process.

Suppose that returns at the highest sampling frequency, R_1 (the t subscript is momentarily dropped for notational convenience), can be described as an ARMA(p,q) process:

$$\alpha(L) R_{1,i} = \beta(L) \varepsilon_{1,i},$$

¹⁰More generally, one could specify an ARFIMA model for returns, thereby allowing for a hyperbolic decay of serial correlation. However, market microstructure and efficiency considerations aside, casual inspection of Table 1.1 and Figure 1.2 suggests that an ARMA process is sufficiently flexible to capture the dynamics of the returns process at high frequency.

where $\alpha(L)$ and $\beta(L)$ are lag polynomials of lengths p and q respectively. As before, we also assume that the return process is weakly stationary which justifies expression 1.4 and the analysis below. Consider the case where all the reciprocals of the roots of $\alpha(L) = 0$, denoted by $\theta_1, \dots, \theta_p$, lie inside the unit circle. The model through which the returns at an arbitrary sampling (or aggregation) frequency can be represented is derived using the results of Wei (1981) on temporal aggregation¹¹. In particular, if R_1 follows an ARMA(p,q) process, the returns sampled at frequency f , denoted by R_f , can be represented by an ARMA(p,r) process:

$$\prod_{j=1}^p (1 - \theta_j^f L^f) R_{f,i} = \prod_{j=1}^p \frac{1 - \theta_j^f L^f}{1 - \theta_j L} \frac{1 - L^f}{1 - L} \beta(L) \varepsilon_{f,i},$$

where r equals the integer part of $p + \frac{q-p}{f}$ and $\varepsilon_{f,i} = \sum_{j=0}^{f-1} \varepsilon_{1,f i - j}$. Due to the invertibility of the AR polynomial, the above model can be rewritten in MA(∞) form with parameters $\{\psi_j\}_{j=0}^{\infty}$ and $\psi_0 = 1$. Let φ_h^f denote the h^{th} autocovariance of the temporally aggregated returns at frequency f :

$$\varphi_h^f = E[R_{f,i} R_{f,i-h}] \propto \sum_{j=0}^{\infty} \left[\left(\sum_{i=\max(0, j-f+1)}^j \psi_i \right) \left(\sum_{i=j+1+f(h-1)}^{j+fh} \psi_i \right) \right]. \quad (1.3)$$

It can be shown that the ψ_j coefficients decay exponentially fast in terms of j and, as a result, the autocovariances disappear under temporal aggregation. To see this, let $|\psi_j| < w\delta^j$ for $|\delta| < 1$ and w some positive constant and notice that:

$$\varphi_h^f \propto \sum_{j=0}^{\infty} \left[\sum_{i=0}^j w\delta^i \sum_{i=j+f(h-1)}^{j+fh} w\delta^i \right] < \frac{w^2}{(1-\delta)^3} \delta^{f(h-1)},$$

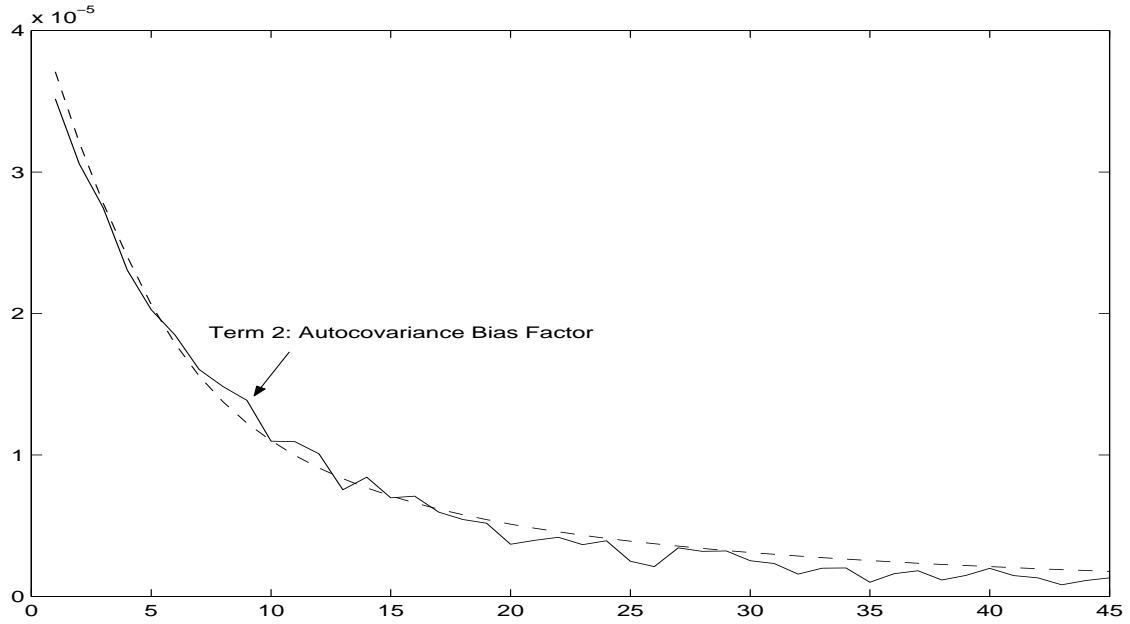
from which it can be seen that the autocovariances of order higher than two disappear when either the sampling frequency, f , or the displacement, h , increases. While it does not follow from the above that the first order autocovariance term also disappears, Wei (1981) has shown that the limit model of an ARMA(p,q) process under temporal aggregation is indeed an ARMA(0,0) or equivalently white noise.

It is important to emphasize that these theoretical properties of the ARMA process appear very much in accordance with the empirical properties of the return process as reported in Table 1.1. In particular, at high sampling frequencies the ARMA model can account for the observed serial dependence while at lower sampling frequencies these dependencies die off as a consequence of temporal aggregation of the return process. In addition, as the ARMA(p,q) model converges to an ARMA(0,0) under temporal aggregation, the model specification for returns at high frequency does not necessarily conflict with a model for returns at low frequency.

Relating the above aggregation results to the discussion of the previous section, we note that the expression for the autocovariance function of the ARMA process can be used to check the consistency of the model with the properties of the data by comparing the temporal aggregation implied decay of the autocovariance bias term with the empirically observed one. To this end, we estimate various ARMA models using the minute by minute

¹¹Temporal aggregation for ARMA models is discussed in Brewer (1973), Tiao (1972), Wei (1981), Weiss (1984) and the VARFIMA in Marcellino (1999).

FIGURE 1.3: THE “AUTOCOVARANCE BIAS FACTOR”



Notes: The empirical autocovariance bias factor (solid line, see also Figure 1.1) and the superimposed aggregation implied autocovariance bias factor (dotted line, see also expressions (1.3) and (1.4)) for sampling frequencies between 1 and 45 minutes.

returns on the FTSE-100 index and find that an ARMA(6,0) model yields satisfactory results¹². Although the residuals are highly heteroskedastic, the OLS parameter estimates remain consistent (Amemiya 1985). Moreover, the efficiency loss due to the non-normality of the errors is unimportant given the large amount of data. Based on the *single* set of ARMA(6,0) parameters associated with the 1-minute data, the autocovariances for the estimated return process at various sampling frequencies can be *deduced* using expression (1.3). It is noted that:

$$E \left[\sum_{i=1}^{N_f-1} \sum_{j=i+1}^{N_f} R_{f,t,i} R_{f,t,j} \right] = \sum_{h=1}^{N_f-1} (N_f - h) \varphi_h^f. \quad (1.4)$$

Hence, the “aggregation implied” autocovariance estimates can be used to calculate the “aggregation implied” autocovariance bias term as in expression (1.4). In particular, a single set of ARMA(6,0) parameters for the 1-minute data are used to imply the autocovariance bias factor at sampling frequencies between 1 and 45 minutes. Figure 1.3 demonstrates that the empirical and theoretically implied curves are remarkably close.

The above results illustrate that the ARMA model is a good description of the return data at different sampling frequencies. In fact, the decay of the (market microstructure-induced) serial dependencies in high frequency returns is consistent with the decay of an ARMA process under temporal aggregation. Also, it can be shown, based on expression 1.4, that the autocovariance bias term decays at an hyperbolic rate under temporal aggregation (i.e. $\sum_{h=1}^{N_f-1} (N_f - h) \varphi_h^f < \frac{N_f^2}{f^2}$). Finally, we notice that it is possible to trace out the entire autocovariance bias factor curve, and hence determine the optimal frequency, using solely a *single* set of ARMA parameters.

¹²Some of the higher order AR terms could arguably be replaced by low order MA terms. However, the AR specification has the advantage that inference is straightforward from a numerical point of view, as opposed to an MA specification. Since the AR and MA specification are largely equivalent preference is given here to the AR specification.

In summary, we have shown that the conditional return variance can be estimated consistently by the realized variance measure, provided that the intra-day returns are serially uncorrelated. When the intra-day returns are serially correlated, realized variance will either overestimate (with negative correlation) or underestimate (with positive correlation) the conditional return variance. Correcting for the bias term by adding up the cross products of intra-day returns, they are known after all, is not desirable as this is equivalent to using the squared daily return to estimate daily realized variance. Here we suggest that when the available high frequency return data are serially correlated, one approach¹³ is to aggregate the returns down to a frequency at which the correlation has disappeared, thereby avoiding (potentially) large biases in the realized variance measure. Plotting the sum of squared intra-day returns or the autocovariance bias factor versus the sampling frequency, as is done in Figure 1.1 proves a very helpful and easily implementable strategy to determine the frequency at which the correlation has died off. Further analysis suggests that the decay of the autocovariance bias factor is consistent with an ARMA process under temporal aggregation. This finding provides an alternative, yet closely related, parametric approach to determining the optimal sampling frequency.

1.3 Modelling Realized Variance

A number of studies¹⁴ have analyzed high frequency data for a variety of financial securities. Regarding the properties of the realized variance measure, several studies find that (i) the marginal distribution of realized variance is distinctly non-normal and extremely right skewed, whereas the marginal distribution of logarithmic realized variance is close to Gaussian, (ii) logarithmic realized variance displays a high degree of (positive) serial correlation which dies out very slowly (iii) logarithmic realized variance does not seem to have a unit root, but there is clear evidence of fractional integration¹⁵, roughly of order 0.40 and (iv) daily returns standardized by realized volatility¹⁶, i.e. the square root of realized variance, are close to Gaussian.

Based on the analysis in Sections 1.2.1 and 1.2.2, which indicates that the daily conditional return variance of the FTSE-100 can be estimated unbiasedly as the sum of squared intra-day returns sampled at a frequency of 25 minutes, a time series of (logarithmic) realized variance is constructed and is displayed in the left panel of Figure 1.4. Table 1.2 reports some descriptive statistics of the time series of realized variance and returns.

We find that our results are very much in line with the findings described above. In particular, the unconditional distribution of the realized variance appears significantly skewed and exhibits severe kurtosis, while the unconditional distribution of logarithmic realized variance is much less skewed and displays significantly reduced

¹³An alternative approach would be to utilize all of the observations by explicitly modelling the high-frequency market microstructure. However, as noted by Andersen, Bollerslev, Diebold, and Ebens (2001), that approach is much more complicated and subject to numerous pitfalls of its own.

¹⁴See for example Andersen, Bollerslev, Diebold, and Labys (2000b, 2000b), Blair, Poon, and Taylor (2001), Dacorogna, Gençay, Müller, Olsen, and Pictet (2001), Froot and Perold (1995), Goodhart and O'Hara (1997), Hsieh (1991), Lequeux (1999), Stoll and Whaley (1990), Zhou (1996).

¹⁵See for example Baillie (1996), Baillie, Bollerslev, and Mikkelsen (1996), Breidt, Crato, and de Lima (1998), Comte and Renault (1998), Henry and Payne (1998), Liu (2000), Lo (1991).

¹⁶In a multivariate setting it is found that the distribution of correlations between realized variance is close to normal with positive mean, and that the autocorrelations of realized correlation decays extremely slow.

TABLE 1.2: DESCRIPTIVE STATISTICS OF REALIZED VARIANCE AND RETURNS.

	Mean	Volatility	Skewness	Kurtosis	ADF[5]
Realized Variance	8.5e-5	2.6e-4	21.21	596	-16.2
Log Realized Variance	-9.98	0.962	0.558	4.11	-8.83
Daily Returns	4.6e-4	0.009	0.063	5.29	-21.8
Standardized Daily Return	0.091	1.091	0.036	2.23	-22.3

Notes: Descriptive statistics based on the FTSE-100 data set for the period 1990-2000. The augmented Dickey Fuller test (“ADF[5]”) includes a constant and 5 lags and has a 5% (1%) critical value of -2.865 (-3.439).

kurtosis (Table 1.2). Furthermore, the correlogram for the realized variance measure decays only very slowly but the Augmented Dickey Fuller test strongly rejects the null hypothesis of a unit root (Table 1.2 and right panel of Figure 1.4). This finding indicates that the (logarithmic) realized variance series may exhibit long memory, a feature that will be discussed below. Finally, daily returns standardized by realized variance are close to normal (Table 1.2). This indicates that the empirical findings obtained by Andersen, Bollerslev, Diebold, and Labys (2000a) on exchange rate data can be extended to the FTSE-100 stock market index data.

1.3.1 Fractional Integration & Realized Variance

A time series, X_t is said to be fractionally integrated of order d if after applying the difference operator $(1 - L)^d$ it follows a stationary ARMA(p, q) process where p and q are finite nonnegative integers. This concept has been developed by Granger (1980), Granger (1981), and Granger and Joyeux (1980). For values of d between 0 and 0.5, the fractionally integrated process¹⁷ exhibits “long memory” which has the property that the effect of a shock to the process is highly persistent but decays over time. This is in contrast to $I(1)$ processes, where a shock has infinite persistence, or at the other extreme $I(0)$ processes, where the effect of a shock decays exponentially fast. The ARFIMA(p, d, q) model can be written as

$$\alpha(L)(1 - L)^d X_t = \beta(L)\varepsilon_t, \quad (1.5)$$

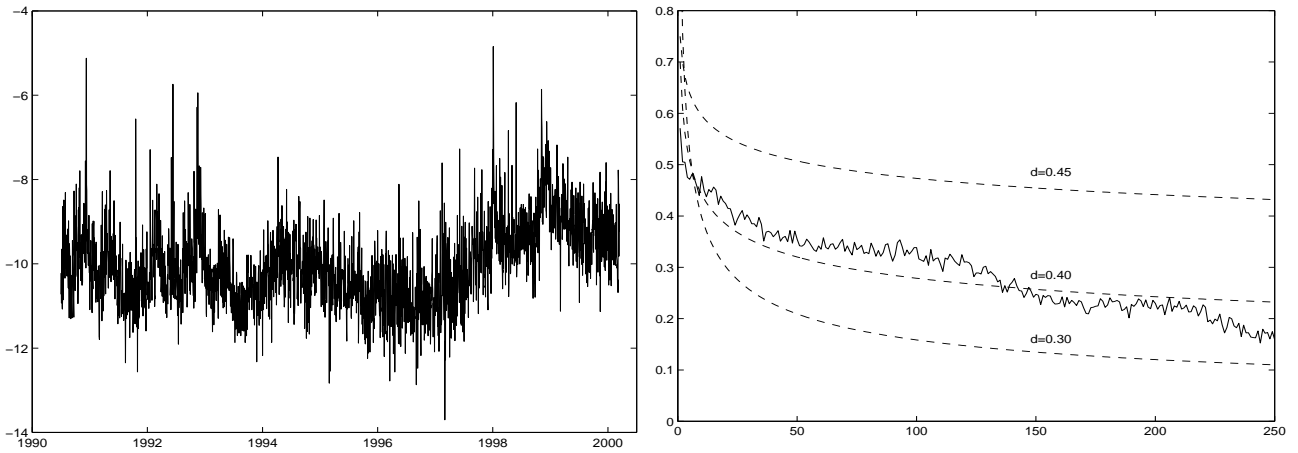
where $\alpha(L)$ and $\beta(L)$ are lag polynomials of order p and q respectively. For $d < \frac{1}{2}$ and $d \neq 0$, it can be shown that the decay of the correlogram is *hyperbolic*, i.e.

$$\varphi_h = \text{corr}(X_t, X_{t-h}) = \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(h+d)}{\Gamma(h+1-d)} \underset{h \text{ large}}{\propto} h^{2d-1}. \quad (1.6)$$

Regarding the estimation of d , Geweke and Porter-Hudak (1983, GPH hereafter) propose the use of a log periodogram regression. In particular, for given $\{X_t\}_{t=1}^T$, the fractional parameter d can be estimated as the slope coefficient in a linear regression of $I(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t e^{i\lambda_j t} \right|^2$, the log periodogram at harmonic

¹⁷The process is stationary with long memory for $0 < d < 0.5$ but stationary with intermediate memory for $-0.5 < d < 0$. For $d \geq 0.5$, the process is non-stationary.

FIGURE 1.4: LOGARITHMIC REALIZED VOLATILITY



Notes: Time series (left panel) and correlogram (right panel) of FTSE-100 daily logarithmic realized variance constructed at a sampling frequency of 25 minutes over the period 1990-2000. The superimposed dotted lines in the right panel represent the correlogram of a fractional process for values of d equal to 0.30, 0.40, and 0.45.

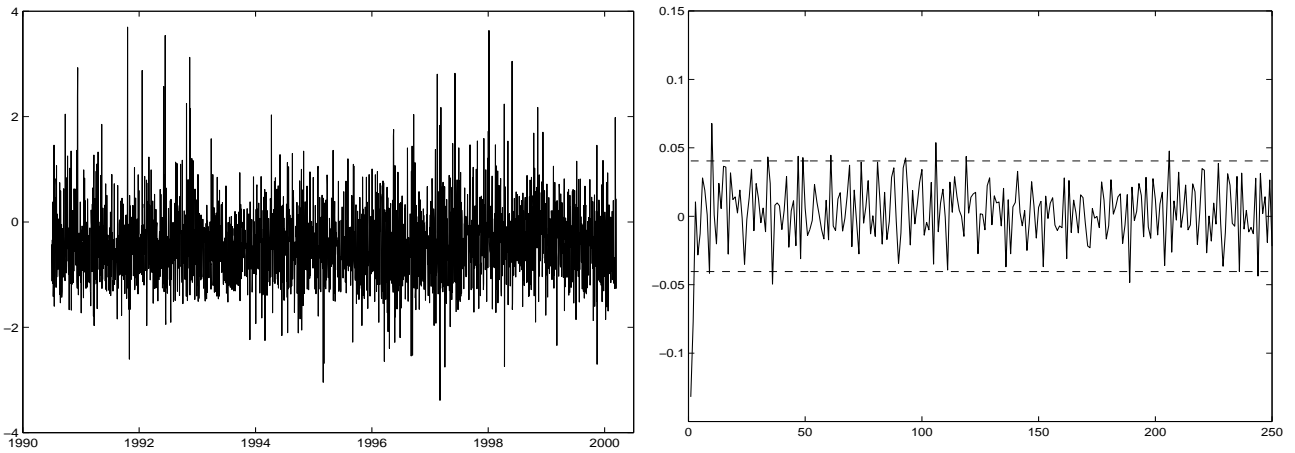
frequency $\lambda_j = \frac{2\pi j}{T}$, on a constant and $\ln [4 \sin^2(\lambda_j/2)]$ for $j = 1, \dots, m \ll T$. The “bandwidth” parameter m is required to increase at a slower rate than the sample size T and in many applications m is set to equal to the square root of the sample size T . Robinson (1995a, 1995b) derives an alternative estimator for d , which is shown to be asymptotically more efficient than the GPH estimator, and is given by the value of d that minimizes the following objective function:

$$Q(c, d) = \frac{1}{m} \sum_{j=1}^m \left[\ln \left(c \lambda_j^{-2d} \right) + \frac{\lambda_j^{2d}}{c} I(\lambda_j) \right],$$

where $c > 0$ and $-\frac{1}{2} < d < \frac{1}{2}$.

Turning to the FTSE-100 realized variance series, it is clear that long memory features are very much present. The right panel of Figure 1.4 displays the correlogram of the log realized variance series while the right panel of Figure 1.5 displays the correlogram of the *fractionally differenced* log realized variance series based on an ad hoc parameter value of $d = 0.40$. The serial correlations of the log realized variance series decay at a hyperbolic rate and the resemblance between the sample correlogram and the superimposed correlograms of a fractionally integrated process for various values of d is remarkable. In sharp contrast, the fractionally differenced series is virtually uncorrelated. A supplementary diagnostic check for the presence of long memory is based on expression (1.6) above. In particular, when the realized variance series exhibits long memory, its log autocorrelation function should yield a linear relationship in terms of log displacement, i.e. $\ln \varphi_h \propto (2d - 1) \ln h$. Figure 1.6 (left panel) indicates the required linear relationship between $\ln \varphi_h$ and $\ln h$ for values of h up to 100. An OLS regression can be used to determine the slope. Based on the entire sample ($h = 250$) the results suggest a value for d of around 0.37. Ignoring the last 150 autocorrelations ($h = 100$) raises d to about 0.43. Finally, the GPH and Robinson estimators, described above, are implemented. The bandwidth parameter m (controlling the range of

FIGURE 1.5: FRACTIONALLY DIFFERENCED LOGARITHMIC REALIZED VOLATILITY



Notes: Time series (left panel) and correlogram (right panel) of FTSE-100 *fractionally differenced* daily logarithmic realized variance constructed at a sampling frequency of 25 minutes over the period 1990-2000. The dotted lines in the right panel are the 95% confidence bounds calculated as $\pm 2N^{-1/2}$ where N denotes the number of observations.

periodic frequencies used), is set equal to a range of values between¹⁸ 25 and 275. The results of this estimation are summarized in Figure 1.6 (right panel) where the GPH and Robinson estimates are plotted as a function of m . For small m , the two alternative estimates both fall into the non-stationary region while for large m (above 150) they are both below 0.5. Although it is clear from this that the value for d will be close to 0.5, it is difficult to judge on the stationarity of the process as the choice of m is relatively arbitrary. In summary, all of the test results reported above suggest that the FTSE-100 log realized variance series is fractionally integrated and appear roughly consistent with Andersen, Bollerslev, Diebold, and Ebens (2001) who find that for their data set d is around 0.40.

1.3.2 Empirical Results

Motivated by the preliminary tests discussed above, the focus of our modelling approach will center around the ARFIMA specification. We consider the following model:

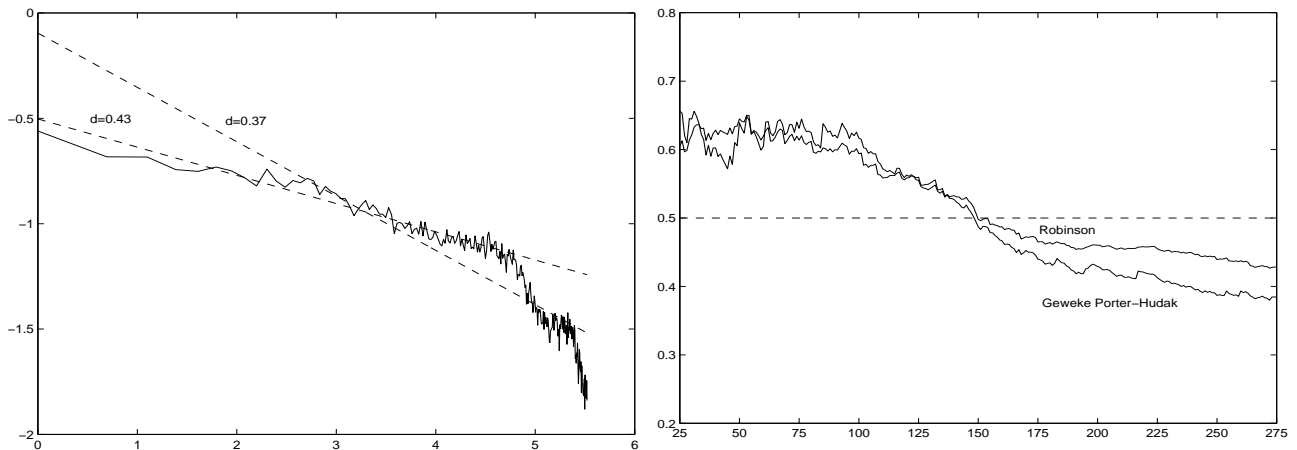
$$\alpha(L)(1-L)^d [\ln \hat{\sigma}_{25,t}^2 - \pi' X_t] = \beta(L)\varepsilon_t, \quad (1.7)$$

where $\hat{\sigma}_{25,t}^2$ denotes the day- t realized variance measure constructed based on 25 minute intra-day returns, $\alpha(L)$ is a lag polynomial of order p , $\beta(L)$ a lag polynomial of order q , and ε_t is a residual error term. The $k \times 1$ vector X_t allows for the inclusion of exogenous variables and deterministic terms such as a constant and time trend. Here, we consider the following specification:

$$\pi' X_t = \omega + \sum_{j=1}^k (\zeta_j R_{t-j} + \bar{\zeta}_j |R_{t-j}|) + \sum_{j=0}^m \lambda_j \ln VOL_{t-j} + \sum_{j=0}^n \delta_j (IR_{t-j} - IR_{t-j-1}) \quad (1.8)$$

¹⁸The sample size is 2445 and hence the range of m is between $T^{0.40}$ and $T^{0.70}$. This is in line with e.g. Bollerslev, Cai, and Song (2000) which set $m = T^{0.50}$ or Dittmann and Granger (2002) which set $m = T^{0.8}$.

FIGURE 1.6: TWO TESTS FOR FRACTIONAL INTEGRATION



Notes: Two tests for fractional integration. Linearity of $\ln \varphi_h$ versus $\ln h$ (left panel) and the Geweke Porter-Hudak and Robinson estimate for d as a function of the bandwidth m (right panel).

where IR_t and VOL_t denote the day- t short term interest rate (1 month UK Interbank rate) and daily trading volume respectively. The inclusion of lagged returns and lagged absolute returns mirrors the EGARCH specification of Nelson (1991) and is, in part, motivated by the well documented Black's leverage effect or the asymmetric impact that lagged returns have on the return variance. In particular, Black (1976) argues that one should expect negative returns to have a larger impact on future variance than positive returns. In the above specification we can test whether such a leverage effect is present at horizon h by testing whether ζ_h is significantly less than zero¹⁹.

Next, trading volume is included because it is often argued that it is intimately related to the return variance. A model which can rationalize such a relationship has been proposed by Clark (1973) where prices follow a subordinated process with information flow (proxied by trading volume or number of trades) being the subordinator. A number of papers have addressed the relationship from an empirical point of view (e.g. Karpoff (1987), Gallant, Rossi, and Tauchen (1992) and more recently Ané and Geman (2000)) and invariably report positive correlation between return variance and trading volume. In addition, an influential paper by Lamoureux and Lastrapes (1990) finds that the *persistence* of return variance decreases (or even disappears) when trading volume is accounted for. Finally, the inclusion of (changes in) the short term interest rate is motivated by Glosten, Jagannathan, and Runkle (1993) who find that it has a significant positive effect on stock market volatility.

Before moving on to the estimation results, we point out that the above specification does not allow us to study the *causal* relation between return volatility and trading volume. In particular, it could well be that, in addition to trading volume causing return volatility, return volatility also has a feedback effect onto subsequent trade activity. Whether such dynamics can be identified at a daily frequency is questionable but are clearly of interest. The

¹⁹Suppressing subscripts momentarily, define $R^+ = R$ when $R > 0$ and $R^+ = 0$ when $R \leq 0$. Similarly, define $R^- = -R$ when $R \leq 0$ and $R^- = 0$ when $R > 0$. Hence, $R = R^+ - R^-$ and $|R| = R^+ + R^-$. It is now straightforward to show that $\zeta R + \bar{\zeta} |R| = \zeta^+ R^+ + \zeta^- R^-$ where $\zeta^+ = \zeta + \bar{\zeta}$ and $\zeta^- = \bar{\zeta} - \zeta$. For the leverage effect to be present, it is required that $\zeta^- > \zeta^+ \Leftrightarrow \zeta^- - \zeta^+ > 0 \Leftrightarrow \zeta < 0$.

theoretical market microstructure has studied such relationships extensively. However, the primary focus has been on the impact of trade duration on the price process and results are mixed (see for example, Admati and Pfleiderer (1988), Diamond and Verrecchia (1987), Easley and O'Hara (1992), and Glosten and Milgrom (1985)). Engle (2000) has also focussed on the impact of trade durations on the price process. Using IBM high frequency data, he finds that low trading activity leads to a reduction in future return volatility (supporting the implications of the Easley and O'Hara (1992) model). A related study by Renault and Werker (2002) investigates the *instantaneous* causality relation between transaction durations and prices and finds that about two-thirds of return volatility can be attributed to instantaneous durations - in other words - transaction times cause transaction prices. Under the assumption that trade durations are inversely proportional to trade volume, the model we have specified in (1.7) and (1.8) is directly in line with the above mentioned work, although it should be kept in mind that we work with data at a daily frequency as opposed to transaction level data. The feedback effect of return volatility on trade durations - or trade volume - is, although of interest, not studied here.

Under the assumptions that (i) the roots of $\alpha(L)$ are simple and lie outside the unit circle, (ii) the residuals are i.i.d. Gaussian, and (iii) $d < \frac{1}{2}$, the ARFIMA model, specified by (1.7) and (1.8) above, can be estimated²⁰ using the maximum likelihood procedure of Sowell (1992). Alternatively, the model could have been estimated using a two-step procedure in which the fractional parameter is estimated in the first step (e.g. with the GPH or Robinson estimator), while the remaining ARMA coefficients are estimated in the second step based on the fractionally differenced data using ordinary least squares. However, as documented by Smith, Taylor, and Yadav (1997), such an approach may well lead to inaccurate or biased ARMA coefficient estimates. The Sowell procedure, allowing for the simultaneous estimation of the model parameters, is therefore preferred.

We first estimate the model without any exogenous variables and then subsequently add returns, volume, and the short rate. To address the concern that long memory may be induced by infrequent structural breaks²¹, we re-estimate the model on various subsamples of the data set. Table 1.3, summarizes the estimation results²² for two different samples and $p = q = 1$, $k = 4$, and $m = n = 2$. The first sample is the full sample while the second sample covers the period May 1, 1990 until June 15, 1997. As the point estimates for the fractional parameter remain within a tight range (with one exception, all estimates are between 0.44 and 0.48) and turn out to be highly significant irrespective of the sample period or the model specification, we argue that the realized variance series clearly exhibits a long memory feature that is not caused by structural breaks. Based on the t -statistic²³,

²⁰We have used the ARFIMA package in PcGive version 10.0. See Doornik and Ooms (1999) and Doornik (2002) for documentation.

²¹See for example Diebold and Inoue (2001), Engle and Smith (1999), Granger (1999), and Granger and Hyung (1999). A simple and representative model that can cause long memory is the stochastic break model, which takes the form: $y_t = u_t + \varepsilon_t$, where $u_t = u_{t-1} + q_{t-1}\eta_t$, $\varepsilon_t \sim iid\mathcal{N}(0, \sigma_\varepsilon^2)$, $\eta_t \sim iid\mathcal{N}(0, \sigma_\eta^2)$ and q_t equals 0 with probability p and 1 with probability $1 - p$. Diebold and Inoue (2001) note that in order to achieve a slowly declining autocorrelation function, whatever the model may be, the key idea is to let p decrease with the sample size so that regardless of the sample size, realizations of the process tend to have just a few breaks.

²²Based on the likelihood ratio test and the AIC criterion we find that an ARFIMA(1,d,1) model provides a parsimonious specification. The choice of k , m , and n is guided by the significance of the parameters. $\zeta_4, \bar{\zeta}_4, \lambda_1$, and δ_1 are included for completeness.

²³The validity of the t -statistics crucially relies on whether the residuals are IID Gaussian. The diagnostic tests reported in Table 1.3 indicate that even though the residuals appear uncorrelated some skewness, kurtosis and heteroskedasticity is present. Fortunately, these effects diminish to some extent when lagged returns and trading volume are included and we will therefore work under the assumption that the t -statistics - in particular for the full model - are reasonably accurate.

TABLE 1.3: ARFIMA ESTIMATION RESULTS

Par	Full Sample (1990-2000)				Sub Sample (1990-1997)			
	ARFIMA	+ Returns	+ Volume	+ Interest	ARFIMA	+ Returns	+ Volume	+ Interest
d	0.483 (22.6)	0.476 (16.6)	0.484 (24.0)	0.484 (23.8)	0.441 (8.74)	0.391 (6.73)	0.476 (16.5)	0.475 (16.2)
α_1	0.356 (5.76)	0.337 (7.27)	0.337 (7.94)	0.335 (7.87)	0.437 (5.70)	0.325 (4.69)	0.321 (6.42)	0.318 (6.30)
β_1	-0.602 (10.3)	-0.678 (15.2)	-0.695 (18.9)	-0.695 (18.8)	-0.635 (7.75)	-0.599 (6.99)	-0.680 (14.8)	-0.678 (14.5)
ζ_1	-	-3.377 (5.03)	-3.863 (5.99)	-3.999 (6.18)	-	-2.362 (2.60)	-3.708 (4.25)	-3.958 (4.51)
ζ_2	-	-2.652 (3.96)	-3.086 (4.79)	-3.107 (4.81)	-	-1.906 (2.10)	-2.834 (3.25)	-2.827 (3.22)
ζ_3	-	-1.439 (2.14)	-1.472 (2.28)	-1.449 (2.25)	-	-2.448 (2.69)	-2.312 (2.65)	-2.290 (2.63)
ζ_4	-	-1.206 (1.79)	-1.498 (2.32)	-1.527 (2.37)	-	-0.925 (1.02)	-1.393 (1.61)	-1.404 (1.62)
$\bar{\zeta}_1$	-	30.36 (28.0)	27.01 (25.3)	27.15 (25.4)	-	37.22 (25.4)	32.73 (22.8)	32.99 (23.0)
$\bar{\zeta}_2$	-	12.85 (11.7)	11.30 (10.5)	11.29 (10.5)	-	16.99 (11.5)	14.89 (10.3)	14.80 (10.2)
$\bar{\zeta}_3$	-	6.468 (5.91)	5.817 (5.52)	5.729 (5.44)	-	6.288 (4.27)	5.812 (4.11)	5.610 (3.97)
$\bar{\zeta}_4$	-	4.541 (4.19)	4.413 (4.24)	4.355 (4.19)	-	3.953 (2.70)	3.926 (2.81)	3.851 (2.76)
λ_0	-	-	0.338 (14.0)	0.335 (13.9)	-	-	0.370 (13.1)	0.365 (12.9)
λ_1	-	-	-0.007 (0.28)	-0.007 (0.30)	-	-	-0.011 (0.39)	-0.010 (0.37)
δ_0	-	-	-	-0.179 (2.29)	-	-	-	-0.198 (2.41)
δ_1	-	-	-	0.031 (0.39)	-	-	-	0.068 (0.83)
-LogL	977.3	607.7	504.5	501.6	710.5	414.0	328.8	325.4
AIC/T	0.805	0.509	0.426	0.425	0.795	0.475	0.382	0.381
No. Par	5	13	15	17	5	13	15	17
Skew	0.675	0.371	0.385	0.380	0.721	0.330	0.345	0.345
Kurt	5.680	4.291	4.184	4.144	5.850	3.924	3.785	3.734
PM[5]	3.888 (0.143)	2.822 (0.244)	1.945 (0.378)	2.095 (0.351)	5.391 (0.068)	4.906 (0.086)	4.365 (0.113)	4.516 (0.105)
ARCH[5]	4.443 (0.001)	4.143 (0.001)	2.662 (0.021)	2.028 (0.072)	3.402 (0.005)	5.389 (0.000)	4.718 (0.000)	3.406 (0.005)

Notes: ARFIMA(1,d,1) estimation results for the full sample (2 May 1990 - 11 January 2000; 2445 observations) and the sub sample (2 May 1990 - 15 June 1997; 1803 observations). The full model specification is given by expressions (1.7) and (1.8). The table reports all parameter estimates (except ω) with absolute t-statistics in parenthesis below. The residual test statistics include skewness ("Skew"), kurtosis ("Kurt"), and the Portmanteau ("PM[5]", χ^2_5) and ARCH ("ARCH[5]", $F(5, 1775)$ for sub-sample and $F(5, 2419)$ for full sample) statistics including 5 lags. p-values are reported in parenthesis below PM[5] and ARCH[5].

however, we cannot reject that $d > 0.5$ at a 95% confidence level, i.e. the realized variance series is potentially non-stationary. Turning to the exogenous variables, we notice a dramatic increase in log likelihood - accompanied by a substantial decrease in AIC criterion - upon inclusion of lagged (absolute) returns. In particular, for $k = 4$, the number of parameters increases by 8 to a total of 13 while the log likelihood increases by almost 370! As a result, the AIC criterion drops from 0.80 to 0.50. Further, the sign and significance of the ζ parameters suggest that Black's leverage effect is present at horizons up to 3 or 4 days. This finding provides support for the GJR-GARCH (Glosten, Jaganathan, and Runkle 1993) and EGARCH (Nelson 1991) specifications which explicitly account for this asymmetric effect that returns have on future variance. Regarding trading volume, we find that contemporaneous values further improve the fit of the model. Consistent with Clark's model, we find that the sign of λ_0 is positive and highly significant. However, in contrast to the findings of Lamoureux and Lastrapes, it appears that the *persistence* of the variance process (as measured by d) remains largely unchanged when trading volume is conditioned upon. Finally, the estimate for δ_0 suggests that an interest rate cut is accompanied by higher volatility than an interest rate hike. It must be said, however, that this effect is marginally significant and that the associated likelihood increase minimal.

1.4 Conclusion

Under certain assumptions on the return process, a number of recent papers have shown that realized variance is a consistent and virtually measurement error-free estimator of the conditional return variance. In this paper we show that realized variance measure constitutes a *biased* estimate of the return variance when (excess) returns are serially correlated. 10 years of FTSE-100 minute by minute data are used to illustrate that a careful choice of sampling frequency is crucial in avoiding a substantial bias. The relation between the sampling frequency and the presence of serial correlation is analyzed in detail and demonstrates that serial correlation in returns disappears under temporal aggregation at a rate of decay that is consistent with that one of an ARMA process. An autocovariance function based method is proposed for choosing the optimal sampling frequency, that is, the highest available sampling frequency for which the autocovariance bias term is negligible. Many alternative approaches to deal with this issue can be considered though. One route is to use all available data by explicitly modelling the market microstructure effects. Another is to "correct" for the bias by dividing the biased realized variance estimate by an appropriate constant (or any sort of function that achieves unbiasedness of the estimator). A third approach, which we may explore in future research, is to use a Newey-West type covariance estimator in order to take into account the serial correlation in the data. The advantage here is that it is potentially more efficient than the aggregation approach outlines in this paper as it makes use of all available data while the non-parametric nature of the estimator avoids the need to explicitly model the market microstructure.

Regarding the FTSE-100 data set, we find that the realized variance series can be modelled as an ARFIMA process. Exogenous variables such as lagged returns and contemporaneous trading volume appear to be highly significant regressors and are able to explain a large portion of the variation in realized variance. Also, statistical tests suggest that Black's leverage effect is significant at three or four days. Regarding contemporaneous trading volume we find that, despite its significance, the persistence of the variance process remains largely unchanged.

Chapter 2

Latent Variable Estimation for the Affine Jump Diffusion

Based on joint work with George Jiang

2.1 Introduction

In recent years, there has been a remarkable pace in the development of continuous-time asset return models, as the literature strives to find a satisfactory model for the representation of the data generating process of asset returns. Recent specifications of continuous-time asset return models have extended from univariate stochastic volatility component to multivariate stochastic volatility components, from random jumps in asset return only to random jumps in both asset return and asset return volatility, as well as from constant jump intensity to time-varying or state dependent jump intensity¹. While asset prices are inherently observed, and can thus be modelled as such, this is certainly not the case for a number of other important variables such as the jump term and the instantaneous mean and variance of the return process. Usually, these components are treated as latent², at the cost of substantially complicating statistical inference. The difficulty arises due to the fact that the latent variables have to be integrated out of the likelihood. The dimension of such integrals is typically very high and sophisticated numerical algorithms need to be employed in order to evaluate the likelihood function. An alternative approach is to construct proxies of the latent variables using observed information. For instance, in the finance literature the squared asset returns are often used as a proxy for time-varying stochastic volatility. More often than not, these proxies are not only very noisy but also inconsistent with the latent variables under specific model specification. Consequently, this can lead to inconsistent parameter estimators and invalid statistical inference as they are, in essence, based on a misspecified model.

¹See for example Andersen, Benzoni, and Lund (2002), Bakshi, Cao, and Chen (1997), Bates (1996, 2000), Bollerslev and Zhou (2002), Duffie, Pan, and Singleton (2000), Chernov, Gallant, Ghysels, and Tauchen (2002), Eraker, Johannes, and Polson (2002), Pan (2002), Singleton (2001).

²An exception is the ARCH class of model which specifies the conditional return variance as a function of (past) observables. See Bollerslev, Engle, and Nelson (1994) or Bollerslev, Chou, and Kroner (1992) for a review of this literature.

Motivated by the observation that many of the relevant continuous time latent variable models are nested in the affine jump diffusion (AJD) model of Duffie, Pan, and Singleton (2000), in this paper we fully exploit the analytical tractability of the AJD to propose a *model-consistent unbiased minimum-variance estimator of the latent variables*. More specifically, based on the closed analytic form of the conditional characteristic function and an uncovered dynamic relation between the latent variables and the cumulants of the observed state variables, we derive an unbiased minimum-variance estimator for the latent variables in terms of the model parameters and observations of the state variables. An important feature of the proposed estimator, is that it is *model-consistent* and can be implemented using *high frequency* data which, in turn, *may* lead to substantial efficiency gains. It is worth emphasizing that our approach can, in principle, be applied to any latent variable model in the general AJD framework and therefore covers a wide range of models frequently studied in the finance literature.

Further, within the AJD model framework, we outline a GMM estimation approach³ that is based on the conditional characteristic function (to derive exact moments) in conjunction with high frequency observations of the state variables and the proposed measurement of the latent variables. The basic idea is to match conditional moments or cumulants of both the observed and the unobserved state variable, where the latter is evaluated at its point estimate. In contrast to simulation based methods, such as simulated method of moments, efficient method of moments, or markov chain monte carlo, our approach does not involve simulations or discretization error of the continuous time model.

The remainder of this paper is organized as follows. In Section 2.2, we review the class of AJD models following Duffie, Pan, and Singleton (2000), and develop the model-consistent unbiased minimum-variance estimator of the latent variable based on high frequency observations on the observed state variable. In Section 2.3, we perform a simulation experiment to study the properties of the proposed latent variable estimator for the stochastic mean and square root diffusion model. Section 2.4 reviews the literature on the estimation of latent variable models in finance and outlines an alternative GMM estimation approach that is based on the proposed estimator of the latent variables. Section 2.5 concludes.

2.2 The Affine Jump-Diffusion Model with Latent Variables

Let $X_t \in \mathcal{R}^n$, $t \geq 0$, be the n -dimensional vector of state variables. Without loss of generality, we partition the whole vector of X_t into two sub-vectors, i.e. $X_t = (S_t', V_t')$, where $S_t \in \mathcal{R}^m$, $n > m > 0$, is the vector of *observed* state variables and $V_t \in \mathcal{R}^{n-m}$ is the vector of *unobserved or latent* variables. For instance, in financial models S_t can be observed asset prices, interest rates, and exchange rates, etc., and V_t can be unobserved instantaneous volatility, instantaneous mean, information flow, etc. Here, we consider a general continuous time affine jump-diffusion (AJD) model as defined in Duffie, Pan, and Singleton (2000) for the state variable X_t . Using the same notation as in Duffie, Pan, and Singleton (2000), we fix a probability space (Ω, \mathcal{F}, P) and an information filtration $(\mathcal{F}_t) = \{\mathcal{F}_t : t \geq 0\}$, and suppose that X_t is a Markov process in some state space $\mathcal{D} \in \mathbb{R}^n$, following

³A closely related study by Bollerslev and Zhou (2002) uses the realized volatility as a proxy for the integrated variance of the return process and propose a GMM estimation based on conditional returns and variance moments of the process.

the stochastic differential equation (SDE):

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dZ_t \quad (2.1)$$

where W_t is an (\mathcal{F}_t) -standard Brownian motion in \mathbb{R}^n , $\mu(\cdot) : \mathcal{D} \rightarrow \mathbb{R}^n$ and $\sigma(\cdot) : \mathcal{D} \rightarrow \mathbb{R}^{n \times n}$ are respectively the drift function and diffusion function, and Z is a pure jump process whose jumps have a fixed probability distribution \mathcal{J} on \mathbb{R}^n and arrive with intensity $\{\lambda(X_t) : t \geq 0\}$, for some $\lambda(\cdot) : \mathcal{D} \rightarrow [0, \infty)$. The initial value of the stochastic process X_0 is assumed to follow a trivial distribution. For X_t to be a well-defined Markov process, regularity conditions on the filtration $(\mathcal{F}_t) = \{\mathcal{F}_t : t \geq 0\}$ and restrictions on the state space as well as on the coefficient functions of the stochastic process, namely $(\mathcal{D}, \mu(\cdot), \sigma(\cdot), \lambda(\cdot), \mathcal{J})$, are required. For technical details, see e.g. Ethier and Kurtz (1986), Duffie and Kan (1996), Duffie, Pan, and Singleton (2000), and Dai and Singleton (2000).

The jump-diffusion (JD) process defined in (2.1) consists of three components. Namely, the drift term $\mu(\cdot)$ representing the instantaneous time trend of the process, the variance term $\sigma(\cdot)\sigma(\cdot)'$ representing the instantaneous variance of the process when no jump occurs, and the jump term Z_t capturing the discontinuous change of the sampling path with both random arrival of jumps and random jump sizes. Moreover we suppose, as in Duffie, Pan, and Singleton (2000), that conditional on the path of X_t , the jump times of Z_t are the jump times of a Poisson process with time varying intensity $\{\lambda(X_s) : 0 \leq s < t\}$, and the size of the jump of Z_t at a jump time τ is independent of $\{X_s : 0 \leq s < \tau\}$ and follows the probability distribution \mathcal{J} .

For convenience and tractability, many financial models impose an ‘‘affine’’ structure on the coefficient functions $\mu(\cdot)$, $\sigma(\cdot)\sigma(\cdot)'$, and $\lambda(\cdot)$, i.e. all of these functions are assumed to be affine on \mathcal{D} . Using the notation in Duffie, Pan, and Singleton (2000), we have

$$\begin{aligned} \mu(X_t) &= K_0 + K_1 X_t, \\ [\sigma(X_t)\sigma(X_t)']_{ij} &= [H_0]_{ij} + [H_1]_{ij} X_t, \\ \lambda(X_t) &= l_0 + l_1 X_t \end{aligned} \quad (2.2)$$

where $K = (K_0, K_1) \in \mathbb{R}^n \times \mathbb{R}^{n \times n}$, $H = (H_0, H_1) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n \times n}$, $l = (l_0, l_1) \in \mathbb{R}^n \times \mathbb{R}^{n \times n}$. Let $g(c) = \int_{\mathbb{R}^n} \exp\{c \cdot z\} d\mathcal{J}(z)$ be the jump transform whenever the integral is well defined, where $c \in \mathbb{C}^n$ the set of n -tuples of complex numbers, $g(\cdot)$ determines the jump size distribution. It is obvious that the set of ‘‘coefficients’’ or parameters (K, H, l, g) completely specifies the AJD process and determines its statistical properties, given the initial condition X_0 . When the jump intensity is set as zero, i.e. $\lambda(\cdot) = 0$, the process is referred to as an affine diffusion (AD) process.

2.2.1 Unbiased Estimators of the Latent Variables

Statistical inference of continuous-time models has presented a great challenge to statisticians and econometricians as it requires the knowledge of the dynamic properties or the transition density of the process. However, the transition density functions of the diffusion and jump-diffusion process are in general not available in a closed analytic form. For instance, in the simplest univariate pure-diffusion case, i.e. $n = 1$, $\lambda(\cdot) = 0$, the Brownian

motion with drift process, the Ornstein-Uhlenbeck process and the square-root diffusion process (also known as the Feller or CIR process) are three well known processes with explicit transition density functions. For these processes, both the drift function $\mu(\cdot)$ and the squared diffusion function $\sigma^2(\cdot)$ have an affine structure. The functional forms of the transition densities corresponding to specifications with non-affine coefficient structure are not known explicitly (Wong 1964). In the presence of latent or unobserved variables, as considered in the multivariate JD model framework, statistical inference presents an even greater challenge. Because of the latent variables in the conditional information set, any inference procedure must either integrate out the latent state variables from the model or rely on some proxies of the latent variables. As we shall see in Section 2.4, various estimation methods have been proposed recently in the literature for dynamic models with latent variables in general and stochastic volatility asset return models in particular. However, most of these methods are simulation based and involve path simulation, which may lead to discretization error, and numerically intensive integration of the latent variables. An alternative estimation approach relies on a proxy of the latent variable. In the finance literature, for instance, squared daily return is often used as proxy of daily volatility in the estimation of asset return models with stochastic volatility. As we will see later on, such proxies are not only very noisy but are generally inconsistent with the volatility measure within a specific model framework. As a result the statistical inference becomes invalid as it is in essence based on a misspecified model.

Below, we develop an *unbiased minimum-variance estimator of latent variable* that can be implemented with high frequency observations of observed state variables. The basic idea is to exploit the fact that under the affine continuous time model framework, the conditional characteristic function often has closed analytical form. Since there is an exact one-to-one correspondence between the characteristic function and the distribution function, the characteristic function contains the same information as the distribution function. Consequently, the dynamic properties can also be investigated based on the conditional characteristic function. Using the relationships derived between the conditional cumulants of observed state variables and the latent variables, we derive the estimator.

Duffie, Pan, and Singleton (2000) showed that under the affine structure, the conditional characteristic function of the jump-diffusion process as defined in (2.1) has a semi-closed form given by:

$$\psi(u; X_{t+\tau}, t, \tau | X_t) \equiv E[\exp\{iu'X_{t+\tau}\} | X_t] = \exp\{C(\tau, u) + D(\tau, u)'X_t\} \quad (2.3)$$

where $D(\cdot)$ and $C(\cdot)$ are the solutions of complex-valued Ricatti equations:

$$\begin{aligned} \frac{\partial D(\tau, u)}{\partial \tau} &= K_1' D(\tau, u) + \frac{1}{2} D(\tau, u)' H_1 D(\tau, u) + l_1 (g(D(\tau, u)) - \iota) \\ \frac{\partial C(\tau, u)}{\partial \tau} &= K_0' D(\tau, u) + \frac{1}{2} D(\tau, u)' H_0 D(\tau, u) + l_0 (g(D(\tau, u)) - \iota) \end{aligned}$$

with boundary conditions $D(0, u) = iu$ and $C(0, u) = 0$ and ι is a vector of ones. With certain specifications of the coefficient function (K, H, l, g) , explicit solutions of $D(\cdot)$ and $C(\cdot)$ can be found. In other cases, as noted in Duffie, Pan, and Singleton (2000), the solution would have to be found numerically.

In this paper, we assume that “high frequency” observations of the state variables S_t are available at fixed sampling interval δ . In particular, the series $\{S_{t+k\delta}\}_{k=0}^N$ for $N = 1/\delta$ is observed over the time period $[t, t + 1]$ and $t = 0, 1, \dots, T - 1$. It is noted that our approach also works for the case with irregular sampling intervals, with the

only difference being the more cumbersome notation. Specializing expression (2.3), and explicitly distinguishing between observed and unobserved state variables, we obtain:

$$\begin{aligned}\psi(u_1, u_2; S_{t+\tau}, V_{t+\tau}, t, \tau | S_t, V_t) &= E [\exp \{iu'_1 S_{t+\tau} + iu'_2 V_{t+\tau}\} | S_t, V_t] \\ &= \exp \{C(\tau, u_1, u_2) + D1(\tau, u_1, u_2)' S_t + D2(\tau, u_1, u_2)' V_t\}\end{aligned}$$

where $C(\tau, u_1, u_2) = C(\tau, u)$, $(D1(\tau, u_1, u_2)', D2(\tau, u_1, u_2)') = D(\tau, u)'$ with $u = (u'_1, u'_2)'$. Further, we note

$$\begin{aligned}\psi(u; S_{t+\tau}, t, \tau | S_t, V_t) &= E [\exp \{iu' S_{t+\tau}\} | S_t, V_t] \\ &= \exp \{C(\tau; u, 0) + D1(\tau; u, 0)' S_t + D2(\tau; u, 0)' V_t\}\end{aligned}\quad (2.4)$$

Lemma 2.2.1 *Let $X_t = (S'_t, V'_t)'$ be the affine jump-diffusion process as defined in (2.1). Given a high frequency sampling scheme of S_t with sampling interval δ , the l^{th} cumulant of $\Delta S_{t+k\delta} = S_{t+k\delta} - S_{t+(k-1)\delta}$, $k \geq 1$, conditional on \mathcal{F}_t is given by*

$$K^l(\Delta S_{t+k\delta} | \mathcal{F}_t) = c^l(k) + d_1^l(k)' S_t + d_2^l(k)' V_t, \quad l = 1, 2, \dots \quad (2.5)$$

where

$$\begin{aligned}c^l(k) &= \frac{\partial^l}{i^l \partial u^l} \{C(\delta, u, 0) + C[(k-1)\delta, -iD1(\delta, u, 0) - u, -iD2(\delta, u, 0)]\} |_{u=0}, \\ d_1^l(k) &= \frac{\partial^l}{i^l \partial u^l} \{D1[(k-1)\delta, -iD1(\delta, u, 0) - u, -iD2(\delta, u, 0)]\} |_{u=0}, \\ d_2^l(k) &= \frac{\partial^l}{i^l \partial u^l} \{D2[(k-1)\delta, -iD1(\delta, u, 0) - u, -iD2(\delta, u, 0)]\} |_{u=0}.\end{aligned}$$

In particular, we have

$$d_2^1(k)' V_t = E[\Delta S_{t+k\delta} | \mathcal{F}_t] - c^1(k) - d_1^1(k)' S_t \quad (2.6)$$

and

$$d_2^2(k)' V_t = Var[\Delta S_{t+k\delta} | \mathcal{F}_t] - c^2(k) - d_1^2(k)' S_t \quad (2.7)$$

Proof See Appendix A.1

Moreover, given the first L conditional cumulants of $\Delta S_{t+k\delta}$, from (2.5) we have explicitly

$$\begin{aligned}d_2^1(k)' V_t &= K^1(\Delta S_{t+k\delta}) - c^1(k) - d_1^1(k)' S_t \\ &\vdots \\ d_2^L(k)' V_t &= K^L(\Delta S_{t+k\delta}) - c^L(k) - d_1^L(k)' S_t\end{aligned}\quad (2.8)$$

based on which it is straightforward to have the following corollary:

Corollary 2.2.2 (to Lemma 2.2.1) *Given the first L conditional cumulants of $\Delta S_{t+k\delta}$, from relation (2.5) or (2.8), the latent variable V_t is exactly (over-, under-) identified from $\{K^l(\Delta S_{t+k\delta})\}_{l=1}^L$ if $\text{rank}(d_2(k)' d_2(k)) = (>, <) \dim(V_t)$, where $d_2(k) = (d_2^1(k), \dots, d_2^L(k))'$.*

Proof See Appendix A.1

It is noted from Lemma 2.2.1 that the latent variables are related to the conditional cumulants of the observed state variables. In particular, expression (2.6) relates the latent variable V_t to the conditional mean of the observed state variable $S_{t+k\delta}$, while expression (2.7) relates the latent variable V_t to the conditional variance of the observed state variable $S_{t+k\delta}$. Equation (2.8) extends the results in Lemma 2.2.1, i.e. linear combinations of latent variables are related to the conditional cumulants of certain order of the observed state variables.

Given a set of observations of the state variables, an unbiased estimator of the cumulant can be obtained from the k -statistic of the same order (see Kenney and Keeping (1951, 1962)). For example, for a given sample size n , the first four k -statics are given by

$$k_1 = m_1, \quad k_2 = \frac{n}{n-1}m_2, \quad k_3 = \frac{n^2}{(n-1)(n-2)}m_3, \quad k_4 = \frac{n^2[(n+1)m_4 - 3(n-1)m_2^2]}{(n-1)(n-2)(n-3)}$$

where m_1 is the sample mean, m_2 is the sample variance, and m_i is the i th sample central moment ($i \geq 3$). In other words, the unbiased estimator of cumulants can be directly calculated from sampling observations.

The uncovered relationship between the observed and unobserved variables through the conditional cumulants, as detailed in expression (2.5), points to the possibility of constructing an unbiased estimator of the latent variables from observations of the observed state variables. In particular, as both the theoretical and simulation results will show, the properties of such unbiased estimators can be improved with the use of high frequency observations on the observed state variables. However, the results in Corollary 2.2.2 indicate that when the dimension of V_t is greater than 1, the identification of the latent variables becomes more complicated. Obviously, when $\text{rank}(d_2(k)'d_2(k)) < \dim(V_t)$, some of the latent variables can not be identified. For notational convenience, we only present the case where the dimensions of S_t and V_t are both equal to 1. Following expressions (2.6) and (2.7), define the $(N \times T) \times 1$ vectors ϑ_t and ξ_t whose i^{th} entries are equal to:

$$\begin{aligned} \vartheta_t(i) &= [d_2^1(i)]^{-1} [\Delta S_{t+i\delta} - c^1(i) - d_1^1(i)' S_t] \\ \xi_t(i) &= [d_2^2(i)]^{-1} [(\Delta S_{t+i\delta})^2 - (c^1(i) + d_1^1(i)' S_t)^2 - c^2(i) - d_1^2(i) S_t] \end{aligned}$$

for $i = 1, \dots, N \times T$. Recall that N denotes the number of intra-period observations while T denotes the number of periods.

Lemma 2.2.3 (Unbiased Minimum-Variance Estimator) *For a given sequence of the observed state variable, $\{S_{t+i\delta}\}_{i=0}^{N \times T}$ and $d_2^1(k) \neq 0, \forall k \geq 1$, an unbiased estimator of the latent variable V_t is given by*

$$\hat{V}_t = W' \vartheta_t \quad (2.9)$$

for any $(N \times T) \times 1$ weighting vector W that satisfies $\iota' W = 1$ where ι is an $(N \times T) \times 1$ vector of ones. When $d_2^1(k) = 0$ but $d_2^2(k) \neq 0, \forall k \geq 1$, an unbiased estimator of the latent variable V_t is given by

$$\hat{V}_t = W' \xi_t \quad (2.10)$$

In addition, the following weighting vector minimizes the variance of the estimator

$$W = \frac{\Sigma_t^{-1} \iota}{\iota' \Sigma_t^{-1} \iota} \quad (2.11)$$

where $\Sigma_t = E[\vartheta_t \vartheta_t' | \mathcal{F}_t] - E[\vartheta_t | \mathcal{F}_t] E[\vartheta_t | \mathcal{F}_t]'$ for the estimator in expression (2.9) and $\Sigma_t = E[\xi_t \xi_t' | \mathcal{F}_t] - E[\xi_t | \mathcal{F}_t] E[\xi_t | \mathcal{F}_t]'$ for the estimator in expression (2.10).

Proof See Appendix

For the general case, diagonal elements of Σ_t can be obtained directly from the \mathcal{F}_t -conditional characteristic function of $\Delta S_{t+k\delta}$. Expressions for the off-diagonal elements can be derived using the joint \mathcal{F}_t -conditional characteristic function of $\Delta S_{t+k\delta}$ and $\Delta S_{t+j\delta}$ for $j \neq k > 0$. As illustrated in the next section using the stochastic mean and square-root stochastic volatility model, closed-form expressions for the optimal weight functions can be derived from various conditional cumulants.

To conclude, we note that there are two versions of asymptotics for the unbiased estimator of the latent variables. Namely, two different sampling schemes can lead to infinite sample size. One is to increase the sampling horizon T while fixing the sampling interval δ (or N). The other is to increase the sampling frequency N or reduce the sampling interval δ while fixing the sampling horizon. The first scheme gives rise to a discrete sample over infinite time horizon as mostly studied in statistics and econometrics, while the second leads to a continuous sampling path which is unique for continuous time models. Of course the combination of the above two sampling scheme will lead to continuous sampling over infinite time horizon. As is clear from the above, the unbiasedness property of the estimator does not depend on the specific sampling scheme. Unfortunately, this is not the case for the consistency of the estimator. Preliminary results have been derived (but are not included here) and suggest that for the pure diffusion process as defined in (2.1) with $\lambda(\cdot) = 0$ we have that (i) the estimator given by expression 2.9 is unbiased and consistent when $T \rightarrow \infty$ but unbiased and inconsistent when $\delta \rightarrow 0$ and (ii) the estimator given by expression 2.10 is both unbiased and consistent under either sampling scheme. In other words, as the sampling horizon $T \rightarrow \infty$ the estimators of latent variables are not only unbiased but also consistent. However, as the sampling interval $\delta \rightarrow 0$ (i.e. $N \rightarrow \infty$) with fixed sampling horizon, the estimators are unbiased but may be inconsistent. The consistency depends on the particular model specification. For example, the estimator of the latent conditional mean (variance) in the stochastic mean (volatility) model discussed below, is inconsistent (consistent) when $\delta \rightarrow 0$. The results further illustrate the advantage of our estimators as it allows for a flexible sampling scheme. In practice, the continuous sampling of asset returns or the ultra high frequency return observations are often plagued by the market microstructure related noises, such as the inherent discreteness of price quotes, time-of-day effect, bid-ask bounce, etc. The above property suggests that this drawback can be easily avoided by using the discrete sampling with extended sampling period.

2.3 Performance of the Unbiased Minimum-Variance Latent Variable Estimator

In continuous-time asset price models the observed state variable typically include the (first difference) logarithmic asset price, while the latent or unobserved variable can be the stochastic instantaneous mean or variance of the process. Below we investigate the properties of the unbiased minimum-variance (“UMV”) estimator of the conditional mean for the stochastic mean (SM) model and that of the conditional volatility for the square-root stochastic volatility (SV) model respectively. We do this by studying the relative performance of the “optimal” latent variable filter compared to alternative, potentially “sub-optimal” or inconsistent, filters. For the SM model we consider a simple moving average, the Kalman filter, and the Kalman smoother while for the SV model, we consider the realized variance measure, the Nelson and Foster (1994) ARCH filter, and the exponentially weighted moving average (“EWMA”) filter. We emphasize that all estimators are implemented using the true model parameter values. Estimation of these parameters will be discussed below in section 2.4.

2.3.1 The Mean-Reverting Stochastic Mean Model

Consider the following process with stochastic linear mean-reverting drift for the (de-trended) asset price:

$$\begin{aligned} dS_t &= \kappa(X_t - S_t)dt + \sigma dW_t^s \\ dX_t &= -\beta X_t dt + \sigma_x dW_t^x, \quad t \in [0, T] \end{aligned} \quad (2.12)$$

where $\kappa \neq \beta$ and W^s and W^x are independent. The model exhibits linear mean reversion to a stochastic conditional mean which itself follows a mean-reverting process. Both the price process, S_t , and its associated stochastic conditional mean, X_t , follow an Ornstein-Uhlenbeck (OU) process. It can be shown that:

$$X_{t+\tau} = e^{-\beta\tau} X_t + \int_t^{t+\tau} e^{-\beta(t+\tau-u)} \sigma_x dW_u^x$$

and

$$\begin{aligned} S_{t+\tau} &= e^{-\kappa\tau} S_t + \frac{\kappa}{\beta - \kappa} (e^{-\kappa\tau} - e^{-\beta\tau}) X_t \\ &+ \int_t^{t+\tau} e^{-\kappa(t+\tau-u)} \sigma dW_u^s + \frac{\kappa}{\beta - \kappa} \int_t^{t+\tau} (e^{-\kappa(t+\tau-u)} - e^{-\beta(t+\tau-u)}) \sigma_x dW_u^x \end{aligned}$$

Therefore, both S_t and X_t are normal with unconditional distributions given by $S_t \sim \mathcal{N}(0, \frac{\sigma^2}{2\kappa} + \frac{\kappa\sigma_x^2}{2\beta(\beta+\kappa)})$ and $X_t \sim \mathcal{N}(0, \frac{\sigma_x^2}{2\beta})$. Conditional mean and variance expressions follow directly from the SDE solutions above. From earlier discussion of the general model, the joint characteristic function of $(S_{t+\tau}, X_{t+\tau})$ conditional on \mathcal{F}_t can be written as:

$$\begin{aligned} \psi(u_1, u_2; S_{t+\tau}, X_{t+\tau}, t, \tau | S_t, X_t) &= E[\exp\{iu_1 S_{t+\tau} + iu_2 X_{t+\tau}\} | S_t, X_t] \\ &= \exp\{C(\tau; u_1, u_2) + D1(\tau; u_1, u_2)' S_t + D2(\tau; u_1, u_2)' X_t\}, \end{aligned}$$

where $C(\cdot)$, $D1(\cdot)$ and $D2(\cdot)$ can be solved from the Riccati equations. For the above SM specification, the solution is given by:

$$\begin{aligned} C(\tau; u_1, u_2) &= \frac{1}{4\kappa} u_1^2 (\sigma^2 + \frac{\kappa^2}{(\beta - \kappa)^2} \sigma_x^2) (e^{-2\kappa\tau} - 1) + \frac{1}{4\beta} (u_2 - u_1 \frac{\kappa}{\beta - \kappa})^2 \sigma_x^2 (e^{-2\beta\tau} - 1) \\ &\quad + u_1 (u_2 - u_1 \frac{\kappa}{\beta - \kappa}) \frac{\kappa}{\beta^2 - \kappa^2} \sigma_x^2 (e^{-(\beta+\kappa)\tau} - 1) \\ D1(\tau; u_1, u_2) &= i u_1 e^{-\kappa\tau} \\ D2(\tau; u_1, u_2) &= i (u_2 - u_1 \frac{\kappa}{\beta - \kappa}) (e^{(\beta-\kappa)\tau} - 1) e^{-\beta\tau} \end{aligned}$$

Analytic expressions of the conditional cumulants of various orders can be derived for both the observed and unobserved state variables. In particular, the first order conditional cumulant for $S_{t+\delta j}$ is derived as:

$$E[S_{t+j\delta} | \mathcal{F}_t] = a(j) S_t + c(j) X_t$$

where $a(j) = e^{-j\kappa\delta}$ and $c(j) = \kappa \frac{e^{-j\kappa\delta} - e^{-j\beta\delta}}{\beta - \kappa}$. Based on the above expression, we can construct an unbiased estimator of the instantaneous mean X_t in terms of the model parameters and an unbiased estimator of the first order conditional moment of the observed state variable. Define

$$\vartheta_t(j) = \frac{S_{t+j\delta} - a(j) S_t}{c(j)} \quad (2.13)$$

Any weighted sum of $\vartheta_t(j)$ for $j = 1, \dots, N \times T$, with weights summing to one, will yield an unbiased estimator of X_t , i.e.:

$$\widehat{X}_t = W' \vartheta_t$$

where W and ϑ_t are $(N \times T) \times 1$ vectors and $\iota' W = 1$. Following Lemma 2.2.3, the optimal choice of W for an unbiased minimum-variance estimator of the instantaneous volatility is given by:

$$W = \frac{\Sigma_t^{-1} \iota}{\iota' \Sigma_t^{-1} \iota}. \quad (2.14)$$

where Σ_t is the conditional variance-covariance matrix of the vector $[\vartheta_t(j)]_{j=1, \dots, N \times T}$. Based on the conditional characteristic function we can derive a closed form expression for Σ_t . In particular, for $j > k$ the off-diagonal (j, k) and (k, j) elements of Σ_t are given by

$$Cov[\xi_t(j), \xi_t(k) | \mathcal{F}_t] = \frac{1}{c(j)c(k)} Cov[S_{t+j\delta}, S_{t+k\delta} | \mathcal{F}_t]$$

while the diagonal (j, j) elements of Σ_t are given by

$$Var[\xi_t(j) | \mathcal{F}_t] = \frac{1}{c(j)^2} Var[S_{t+j\delta} | \mathcal{F}_t].$$

The relevant variance and co-variance expressions are given in Appendix A.2.2. It is noted from there that none of the state variables enter into the covariance expression and, as a result, the optimal weights are constant.

Simulation Design. In order to investigate the performance of the above estimator, we perform a simulation experiment. In particular, based on the SM model in (2.12), we simulate a time series of high frequency observations on both the observed and unobserved state variables, i.e. $\{S_{t+(k-1)\delta}, X_{t+(k-1)\delta}\}$ for $k = 1, \dots, 1/\delta$, and $t = 1, \dots, T + TT - 1$. The parameters δ , TT , and T can be interpreted as the sampling frequency, sample size, and sampling horizon respectively. In the remainder of this paper we assume without loss of generality that the variables T and TT are measured in days. Due to the analytic tractability of the OU process, the sample path can be simulated exact, without discretization error. From this simulated sample, we record the value of the latent variable at the beginning of each period, i.e. $\{X_t\}_{t=1}^{TT}$, and use the high frequency realizations of the observed state variable to construct the following four filters:

1. **UMV FILTER.** Based on the results in Lemma 2.2.3, we compute the SM model-consistent estimator with optimal weights using high frequency observations of the state variable over a *sampling horizon* of length T , i.e. $\{S_{t+(k-1)\delta}\}_{k=1}^{T/\delta}$. In particular,

$$\hat{X}_t^{sm} = W' \vartheta_t \quad (2.15)$$

where the j^{th} element of the $N \times T$ vector ϑ , and the optimal weighting matrix W , are given by expressions (2.13) and (2.14) respectively.

2. **MOVING AVERAGE.** At a sampling horizon of T days, we compute the average of the observed state variable

$$\hat{X}_t^{ma} = \frac{1}{T} \sum_{k=0}^{T/\delta-1} S_{t+k\delta} \quad (2.16)$$

3. **KALMAN FILTER.** As is shown in Appendix A.2.1, we can reformulate the SM model in (2.12) into linear state space form after which the Kalman filter (see for example Harvey (1989, 1993), Koopman and Harvey (2003)) can be applied to filter out the latent state variable. Here, we denote the Kalman filter estimates by \hat{X}_t^{kf} .
4. **KALMAN SMOOTHER.** The linear state space form can also be exploited to obtain Kalman smoother estimates of the latent state variable. Here, we denote the Kalman smoother estimates by \hat{X}_t^{ks} .

Notice that in order to construct an estimate of the latent variable, the UMV and Moving Average filter use contemporaneous and future information, the Kalman filter uses past information, and the Kalman smoother uses past, present, and future information.

Regarding the parameter choice, for the benchmark case (“Par I”, see Table 2.1) we fix $\kappa = 0.5$ and $\beta = 0.05$ and adjust σ and σ_x so as to achieve an unconditional volatility of the latent instantaneous mean of 10% annually and an unconditional volatility of the observed variable of 25% annually. Because the optimal weights, and possibly the performance of the various filters, may depend on the specific choice of model parameters we consider two alternative sets of parameters. The first set (“Par II”) increases the mean reversion of the observed process to its latent conditional mean while leaving the speed of mean reversion of the latent process itself unchanged. The second set (“Par III”) increases the persistence of the latent process. Both parameter configurations are expected

to facilitate inference relative to the benchmark case. To investigate the impact of sampling frequency and sample size we fix $T = 1$ and vary $\delta = \left\{ \frac{5}{60 \times 8}, \frac{30}{60 \times 8}, \frac{120}{60 \times 8}, \frac{480}{60 \times 8} \right\}$, corresponding to 5 minutes, 30 minutes, 2 hours, and 1 day data based on a trading day of 8 hours. We set $TT = 2520$, corresponding to a high frequency sample path with a length of 10 years.

The performance of all four filters is measured by regressing⁴ the measurement error, i.e. $X_t - \hat{X}_t^{sm}$, $X_t - \hat{X}_t^{ma}$, $X_t - \hat{X}_t^{kf}$, and $X_t - \hat{X}_t^{ks}$, on a constant and the actual realization of the latent variable, X_t , for $t = 1, \dots, TT$. The filter is unbiased when the intercept and slope coefficients are both insignificantly different from zero. The efficiency of the estimators is measured by the MSE statistic while the R^2 measures whether any systematic component remains in the measurement error that can be explained by the realizations of the latent variable. Ideally, we would want both the MSE and R^2 statistics to be as close to zero as possible. In anticipation of the simulation results, we use the ‘‘AR(1) plus noise model’’ to provide some intuition for the behavior of the slope coefficient in the regression suggested above. In particular, let the latent process X_t be specified as:

$$X_t = \rho X_{t-1} + \varepsilon_t = X_0 + \sum_{i=0}^{t-1} \rho^i \varepsilon_i$$

where $\varepsilon \sim \text{iid } N(0, \sigma_\varepsilon^2)$. The process is stationary when $|\rho| < 1$ in which case it has mean 0 and variance $\sigma_\varepsilon^2 / (1 - \rho^2)$. Further suppose that the *observed* process is equal to:

$$S_t = X_t + \eta_t$$

where $\eta \sim \text{iid } N(0, \sigma_\eta^2)$. Notice that the dynamic specification of the latent state variable, X , corresponds to the (discretized) process in expression 2.12. However, the observed process is different (and simpler) but will still serve its illustrative purpose here. Next, consider a *naive two-sided moving average estimator* for the latent state:

$$\hat{X}_t = \lambda_w^{-1} \sum_{i=-N_P}^{N_F} \lambda^{|i|} S_{t+i}$$

where

$$\lambda_w = 1 + \sum_{i=1}^{N_F} \lambda^i + \sum_{i=1}^{N_P} \lambda^i = \begin{cases} 1 - \frac{2 - \lambda^{N_F} - \lambda^{N_P}}{\lambda - 1} \lambda & \text{if } 0 < \lambda < 1 \\ 1 + N_F + N_P & \text{if } \lambda = 1 \end{cases}$$

The Kalman filter (smoother) is closely related to the above estimator when $N_P > 0$ and $N_F = 0$ ($N_P > 0$ and $N_F > 0$). Also notice that the moving average estimator is unbiased when $N_F = N_P = 0$. We are now interested in the slope coefficient of the regression discussed above, i.e. the measurement error on the actual realization of the latent variable:

$$X_t - \hat{X}_t = \alpha + \beta X_t + \xi_t$$

It can be shown that β will tend to

$$\beta = \frac{\lambda_w - 1}{\lambda_w} + \lambda \rho \frac{2 - (\lambda \rho)^{N_F} - (\lambda \rho)^{N_P}}{\lambda_w (1 - \lambda \rho)} > 0$$

⁴We point out that alternative regression formulations could have been considered. For example, we could have ran a regression of X on a constant and \hat{X} but this is likely to deliver a biased slope coefficient due to measurement error. The regression of \hat{X} on a constant and X circumvents this problem but has the drawback that the R^2 is difficult to interpret.

TABLE 2.1: SIMULATION PARAMETERS FOR “STOCHASTIC MEAN MODEL II”

	κ	β	σ	σ_x
Par I	0.50	0.05	14.6	1.99
Par II	2.50	0.05	32.3	1.99
Par III	0.50	0.01	14.5	0.89

Notes: All volatility coefficients (i.e. σ and σ_x) are scaled up by 1,000 in the table. The parameters are determined by fixing κ and β , while adjusting σ and σ_x so as to maintain $V[X] = (10\%)^2$ and $V[S] = (25\%)^2$ annually. μ and ρ are set equal to zero in which case $V[X] = (2\beta)^{-1}\sigma_x^2$ and $V[S] = (2\kappa)^{-1}\sigma^2 + (2\beta(\beta + \kappa))^{-1}\kappa\sigma_x^2$

when the sample size, T , tends to infinity. From this it is clear that when $N_F > 0$ and / or $N_P > 0$ the filter is “over-smoothing” and the slope coefficient in the above regression will be (i) positive⁵ and (ii) increasing in with the order (degree of smoothing) of the moving average filter. The R^2 statistic will then indicate how strong this correlation between the measurement error and the latent variable is.

To reduce variation in the simulation results, Table 2.2 reports the *average* slope coefficient (plus associated average absolute t-statistics), R^2 , and *RMSE* obtained from 50 independent simulation runs. It turns out that for some filters, sampling frequencies, and / or parameter configurations, the measurement error can be heteroskedastic and serially correlated. Table 2.2 therefore reports heteroskedastic and autocorrelation (HAC) consistent t-statistics (Newey and West 1987) computed using 15 lags. Although the choice of lag length is relatively ad hoc, and no data-driven lag selection procedure (Newey and West 1994) is implemented, inspection of the measurement error time series properties indicates that the order of serial correlation does not extent beyond 15 for the simulations we consider.

Simulation Results. The results reported in Table 2.2 illustrate that irrespective of the sampling frequency or parameter configuration, the SM model-consistent UMV filter yields unbiased estimates of the conditional mean process. Based on the heteroskedastic and autocorrelation consistent t-statistics, all intercept coefficients are insignificantly different from zero (not reported) and all slope coefficients are insignificantly different from one. At first sight, this also appears to hold true for the naive Moving Average filter. However, when the sampling horizon extends beyond 16 days, a substantial bias can be detected. Similarly, both the Kalman filter and smoother constitute biased estimators of the conditional mean. Further, we find that an increase in the speed of mean reversion of the observed process to its latent conditional mean (Par II) leads to a better performance of all filters as measured by the MSE. An increase in the persistence of the latent process itself (Par III) has little impact on the performance of the UMV and Moving Average filter while the performance of the Kalman filter and smoother do improve substantially. For the UMV filter and Par II, we find that a sampling frequency of about 2 hours is optimal in the sense that at this frequency the MSE minimized. For the other parameter configurations (Par I and III), the performance of the filter is best at the lowest sampling frequency considered. It could well be that by

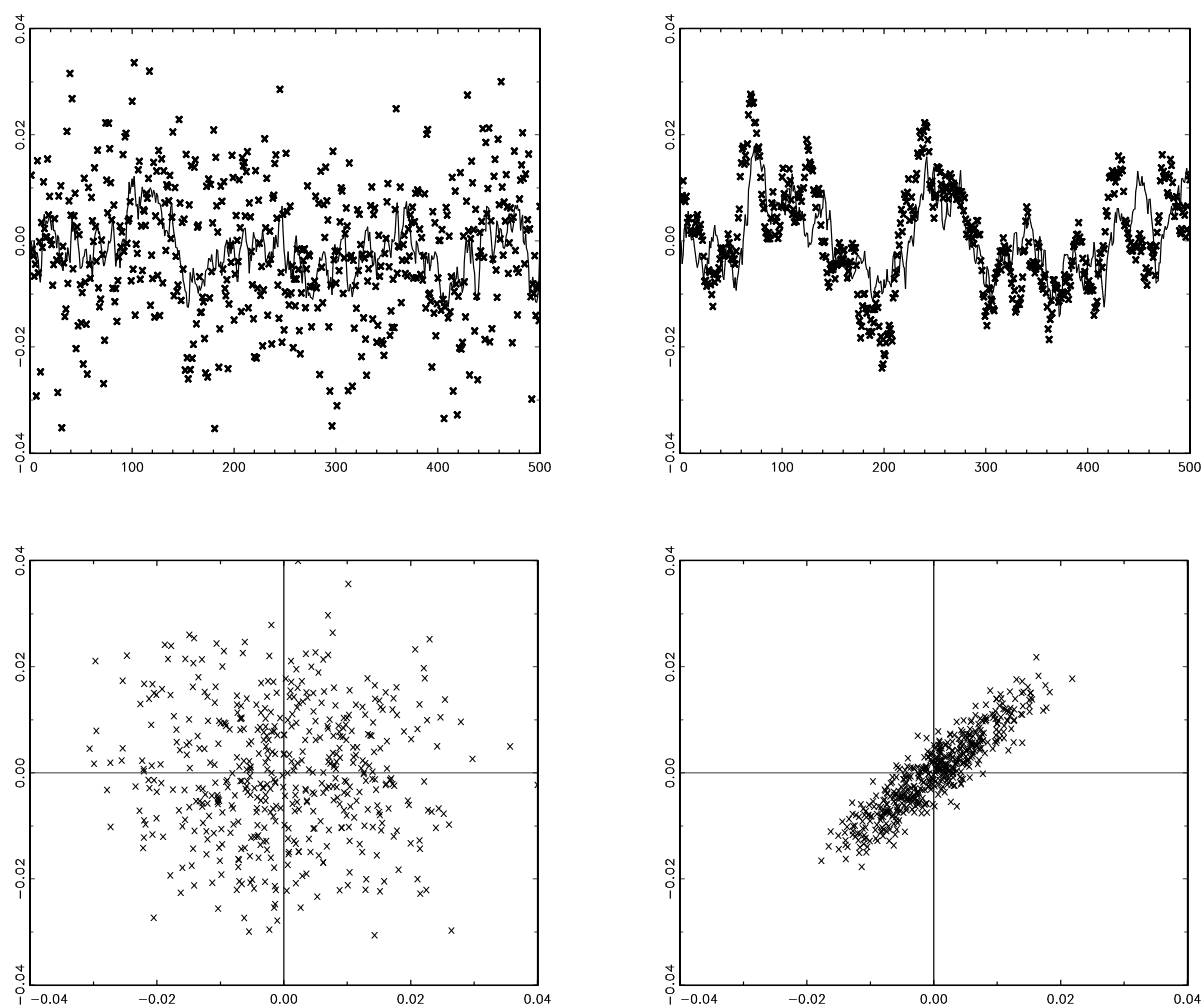
⁵For example, when $\lambda = 1$, $N_F = 1$, $N_P = 0$, then $\beta = \frac{1}{2}(1 - \rho)$. Thus, with a autoregressive coefficient of 0.8, the slope coefficient will be 0.10.

TABLE 2.2: RELATIVE PERFORMANCE OF FILTERS FOR THE STOCHASTIC MEAN MODEL

Par	δ (min)	T (day)	T/δ (obs)	UMV Filter			Moving Average			Kalman Filter			Kalman Smoother		
				π_1	R^2	MSE	π_1	R^2	MSE	π_1	R^2	MSE	π_1	R^2	MSE
I	5	1	96	0.030 (0.78)	0.04%	29.8	0.110 (1.39)	0.41%	13.6	0.754 (29.0)	51.6%	6.94	0.607 (19.3)	59.1%	4.93
	30	1	16	0.032 (0.77)	0.04%	29.0	0.110 (1.40)	0.41%	13.6	0.752 (28.9)	68.5%	5.70	0.607 (19.3)	59.1%	4.93
	120	1	4	0.049 (0.86)	0.06%	26.1	0.113 (1.41)	0.42%	13.7	0.748 (28.6)	71.3%	5.53	0.607 (19.3)	59.2%	4.93
	480	1	1	0.101 (1.25)	0.31%	15.1	0.125 (1.51)	0.41%	14.8	0.733 (27.6)	71.1%	5.44	0.611 (19.4)	59.6%	4.95
	480	4	4	0.010 (0.82)	0.19%	14.4	0.107 (1.40)	0.57%	11.4	0.733 (27.6)	71.1%	5.44	0.611 (19.4)	59.6%	4.95
II	480	16	16	0.002 (0.97)	0.50%	10.5	0.276 (3.94)	5.50%	7.64	0.733 (27.6)	71.1%	5.44	0.611 (19.4)	59.6%	4.95
	480	96	96	0.007 (1.07)	0.60%	10.1	0.793 (17.4)	56.5%	6.61	0.733 (27.6)	71.1%	5.44	0.611 (19.4)	59.6%	4.95
	5	1	96	0.013 (0.77)	0.04%	13.3	0.033 (1.01)	0.10%	10.4	0.484 (18.8)	40.7%	4.80	0.323 (12.0)	32.5%	3.54
	30	1	16	0.011 (0.75)	0.04%	13.0	0.032 (1.00)	0.10%	10.4	0.481 (18.6)	46.5%	4.40	0.323 (12.0)	32.5%	3.54
III	120	1	4	0.008 (0.71)	0.04%	12.3	0.031 (0.96)	0.09%	10.7	0.476 (18.4)	46.7%	4.34	0.327 (12.0)	32.8%	3.56
	480	1	1	-0.003 (0.75)	0.04%	15.1	0.031 (0.90)	0.06%	14.6	0.501 (18.5)	49.1%	4.46	0.380 (13.1)	37.7%	3.86
	5	1	96	0.032 (0.71)	0.04%	29.0	0.053 (0.88)	0.21%	13.4	0.514 (14.3)	30.2%	6.11	0.362 (11.2)	36.5%	3.60
	30	1	16	0.030 (0.71)	0.04%	28.3	0.053 (0.88)	0.21%	13.4	0.514 (14.4)	42.9%	4.71	0.362 (11.2)	36.5%	3.60
III	120	1	4	0.043 (0.84)	0.06%	25.6	0.054 (0.89)	0.21%	13.5	0.512 (14.4)	45.6%	4.53	0.363 (11.2)	36.6%	3.60
	480	1	1	0.051 (0.86)	0.18%	14.6	0.056 (0.88)	0.19%	14.6	0.507 (14.1)	46.1%	4.45	0.364 (11.1)	36.4%	3.62

Notes: Slope coefficients of the regressions $X_t - X_t^{filter} = \pi_0 + \pi_1 X_t + \varepsilon_t^{filter}$ for the UMV filter, Moving Average filter, Kalman filter, and Kalman smoother as described in section 2.3.1 above. The results are based on 50 independent simulation runs, each with a length of 10 years. The model parameters are listed in Table 2.1. Relative to Par I, Par II implies higher mean reversion of the observed process to the latent process while Par III implies higher mean reversion of the latent process. Absolute value of the associated HAC[15]-consistent t-statistics are in parenthesis below where $H_0 : \pi_0 = 0$ and $\pi_1 = 0$ (π_0 not reported). The column " R^2 " reports the coefficient of determination while the column "MSE" reports the root mean squared error ($\times 1,000$).

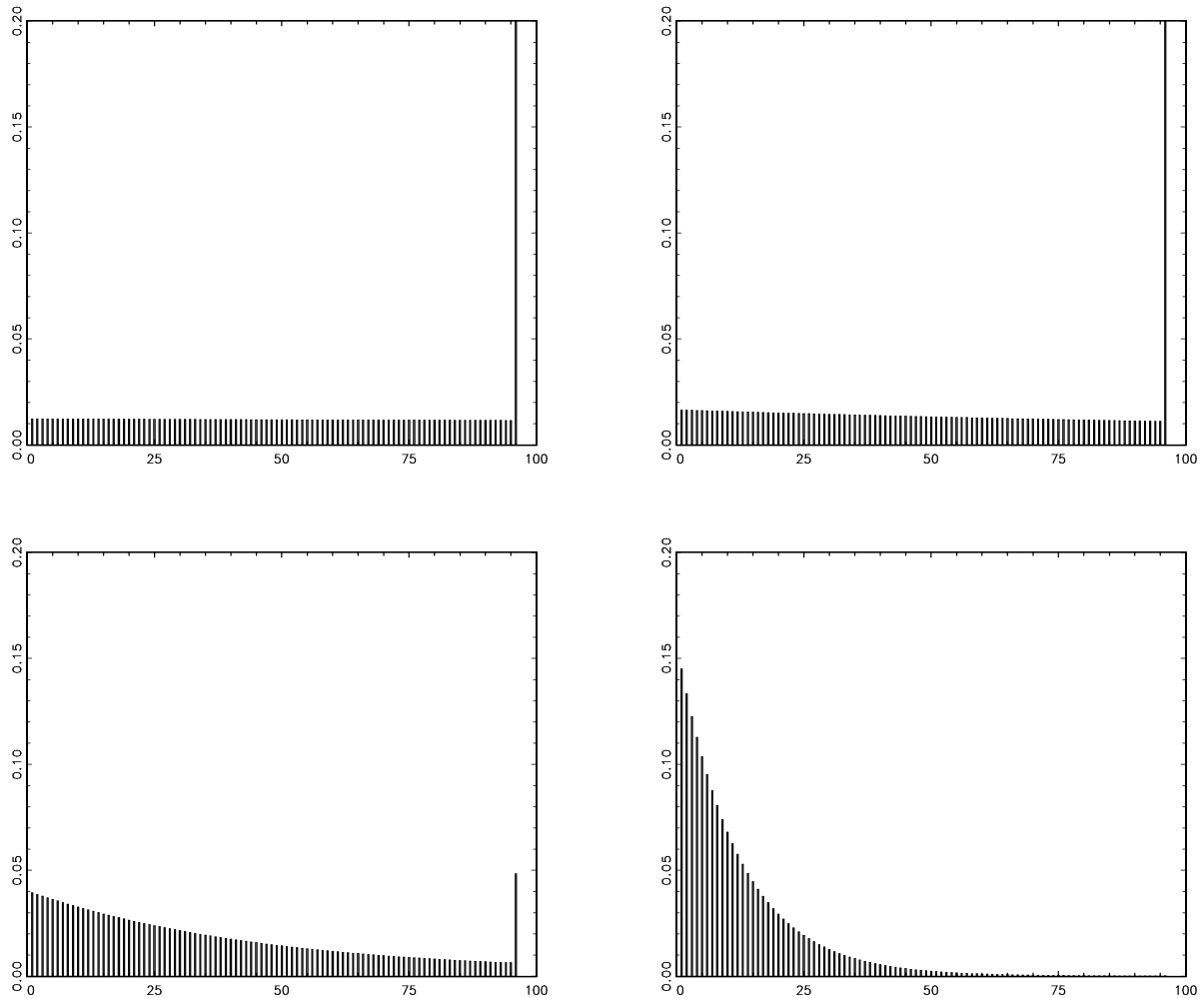
FIGURE 2.1: UMV FILTER FOR ALTERNATIVE “FREQUENCY - HORIZON” SAMPLING SCHEMES



Notes: The top left panel plots a time series of UMV estimates of the conditional mean in the SM model (crosses) together with the actual latent state variable (solid line) based on a 5 minute sampling frequency over a 1 day horizon. The bottom left panel plots the measurement error against its first lagged value. The right panel plots analogous results for a 1 day sampling frequency and a 96 day horizon. The SM parameters are set equal to Par II (see Table 2.1).

further decreasing the sampling frequency, the performance of the filter can be improved. Consistent with the theory, we find that the Kalman filter and smoother achieve the lowest overall MSE among all filters. However, the slope coefficients indicate that a substantial bias is present which is most severe for the Kalman filter. Also, we find that the performance of the Kalman filter varies with the sampling frequency in a non-linear fashion while an increase in the sampling frequency always leads to an improvement of the performance of the Kalman smoother. In particular, for Par II, the Kalman filter achieves the lowest MSE at a sampling frequency of 2 hours (same for UMV filter) as opposed to 5 minutes for the Kalman smoother. For Par I and III, the Kalman filter performs best at the lowest sampling frequency considered while this is exactly the opposite for the Kalman smoother. Finally, it should be noted that the lower MSE statistics for the Kalman filter and smoother are - in part - due to the fact that

FIGURE 2.2: OPTIMAL UMV FILTER WEIGHTS FOR THE SM MODEL



Notes: Optimal UMV filter weights for a given sampling horizon and frequency. Left top: 1 day at 5 minutes. Right top: 6 days at 30 minutes. Left bottom: 24 days at 2 hours. Right bottom: 96 days at 1 day. The SM parameters are set equal to Par I (see Table 2.1).

these filters use more data. In particular, the UMV and Moving Average filter only use one day's worth of data while the Kalman smoother uses the entire data set, from start to finish. Hence, the results in the table are slightly misleading and one should compare the performance of the Kalman smoother (filter) at say a one day frequency to the UMV or Moving Average filter performance at the one day frequency and a 96 day horizon. It is clear from Figure 2.2 that for this case, the horizon is sufficiently long for the UMV filter to achieve its optimal performance.

As mentioned above, an alternative to varying the sampling *frequency* is to vary the sampling *horizon*. For example, a sampling scheme where $T = 1$ day and $\delta = 30$ minutes generates the same number of observations as a sampling scheme where $T = 16$ days and $\delta = 1$ day. Nevertheless, the performance of the UMV filter is not expected to be invariant to the specific sampling scheme and we will therefore briefly study its impact based on some simulations. In particular, for three out of four sampling schemes discussed above, i.e. $T = 1$ day and $\delta = 5, 30, 120$ minutes, we consider the following three alternative sampling schemes, i.e. $\delta = 1$ day

and $T = 96, 16, 4$ days respectively, which generate the same number of observations at a *lower frequency and over a longer horizon*. Based on the sampled data, $\{S_{t+(k-1)\delta}\}_{k=1}^{T/1\text{day}}$, we again compute the UMV filter and the naive Moving Average filter and compare it to the actual realization of the latent variable. To conserve space, we only report the results for Par I in Table 2.2 (between the dotted lines). The results for Par II and III are qualitatively the same. Based on the MSE statistic it is clear that, for a fixed sampling frequency, an increase in the sampling horizon leads to better performance of the UMV filter. In particular, at a 1 days sampling frequency, an increase in the horizon from 1 day to 96 days leads to a reduction in MSE of about 33%. The more relevant comparison, however, is the one described above. So when sampling 4, 16, and 96 data points, the gain in MSE associated with a “low-frequency-long-horizon” sampling scheme – relative to a “high-frequency-short-horizon” sampling scheme – is about 45%, 64%, and 66% respectively. Hence, it is clear that for the estimation of the latent instantaneous mean, low frequency data over long horizons is preferred to high frequency data over short horizons. This point is also illustrated in Figure 2.1. Similar results hold for the moving average filter except that they are less pronounced and that at the long horizon a substantial bias kicks in.

2.3.2 The Square-Root Stochastic Volatility Model

Consider the following asset return process with stochastic conditional volatility for the logarithmic asset price $S_t = \ln P_t$

$$\begin{aligned} dS_t &= \mu dt + \sqrt{V_t} dW_t^s \\ dV_t &= \beta(\alpha - V_t) dt + \sigma \sqrt{V_t} dW_t^v \\ dW_t^s dW_t^v &= \rho dt, \quad t \in [0, T] \end{aligned} \tag{2.17}$$

This continuous-time SV model has been widely used in the finance literature for asset return dynamics, in part because it has an associated closed-form expression for European option prices (Heston 1993). Following Singleton (2001), the drift term of the asset return process is specified as a constant. It is noted that when the drift term of the asset return process is specified as a linear function of the state variable V_t , both the European option prices and the conditional characteristic function of the asset return will still yield closed forms. The specification of the instantaneous volatility process in the above model guarantees the non-negativeness of the volatility (as long as $2\beta\alpha \geq \sigma^2$ as shown by Cox, Ingersoll, and Ross (1985)). The solution of the square-root process in (2.17) can be written as:

$$V_{t+\tau} = \alpha + e^{-\beta\tau}(V_t - \alpha) + \sigma \int_t^{t+\tau} e^{-\beta(t+\tau-u)} \sqrt{V_u} dW_u^v$$

which is of an AR(1) form, where $\sigma \int_t^{t+\tau} e^{-\beta(t+\tau-u)} \sqrt{V_u} dW_u^v$ is a martingale. The variance process can thus be viewed as an autoregressive process of order one with heteroskedasticity in the innovation term. The parameter β measures the inter-temporal persistence of the volatility process, while the correlation between dW^s and dW^v measures the level of asymmetry of the conditional volatility. In particular, when $\rho < 0$ we have the so-called “leverage effect”, see Black (1976).

Again, from earlier discussion of the general model, the joint characteristic function of $(S_{t+\tau}, V_{t+\tau})$ conditional on \mathcal{F}_t can be written as:

$$\begin{aligned}\psi(u_1, u_2; S_{t+\tau}, V_{t+\tau}, t, \tau | S_t, V_t) &= E \left[\exp \{ iu'_1 S_{t+\tau} + iu'_2 V_{t+\tau} \} | S_t, V_t \right] \\ &= \exp \{ C(\tau; u_1, u_2) + D1(\tau; u_1, u_2)' S_t + D2(\tau; u_1, u_2)' V_t \},\end{aligned}$$

where $C(\cdot)$, $D1(\cdot)$ and $D2(\cdot)$ can be solved from the Riccati equations. For the above SV specification, the solution is given by:

$$\begin{aligned}C(\tau; u_1, u_2) &= (iu_1\mu + i\beta\alpha u_2)\tau + \frac{\alpha\beta}{\sigma^2} \left[(b-h)\tau - 2 \ln \left(\frac{1 - ge^{-h\tau}}{1-g} \right) \right] \\ D1(\tau; u_1, u_2) &= iu_1 \\ D2(\tau; u_1, u_2) &= iu_2 + \frac{b-h}{\sigma^2} \frac{1 - e^{-h\tau}}{1 - ge^{-h\tau}}\end{aligned}$$

with $h(u_1, u_2) = [b^2 + \sigma^2(u_1^2 + 2\rho\sigma u_1 u_2 + \gamma^2 u_2^2 + 2i\beta u_2)]^{1/2}$, $b = \beta - \rho\sigma i u_1 - \sigma^2 u_2 i$, $g(u_1, u_2) = (b-h)/(b+h)$. Based on the characteristic function above, analytic expressions for the conditional cumulants of any order can be derived for both the observed and the latent variables. In particular, the second order conditional return cumulant for the observed variable is derived as:

$$K[(\Delta S_{t+j\delta})^2 | \mathcal{F}_t] = a(j) + c(j) V_t$$

where

$$a(j) = \alpha\delta + \alpha \frac{1 - e^{\beta\delta}}{\beta e^{j\beta\delta}} \quad \text{and} \quad c(j) = \frac{e^{\beta\delta} - 1}{\beta e^{j\beta\delta}}.$$

Based on the above expression, we can construct an unbiased estimator of the instantaneous return variance V_t in terms of the model parameters and an unbiased estimator of the second order conditional return cumulant. Define

$$\xi_t(j) = \frac{(\Delta S_{t+j\delta})^2 - (\mu\delta)^2 - a(j)}{c(j)} \quad (2.18)$$

Any weighted sum of $\xi_t(j)$ for $j = 1, \dots, N \times T$, with weights summing to one, will yield an unbiased estimator of V_t , i.e.:

$$\widehat{V}_t = W' \xi_t$$

where W and ξ_t are $(N \times T) \times 1$ vector and $1'W = 1$. Following Lemma 2.2.3, the optimal choice of W for an unbiased minimum-variance (“UMV”) estimator of the instantaneous volatility is given by:

$$W = \frac{\Sigma_t^{-1} \mathbf{1}}{\mathbf{1}' \Sigma_t^{-1} \mathbf{1}}. \quad (2.19)$$

Based on the conditional characteristic function we can derive a closed form expression for Σ_t . In particular, for $j > k$ the (j, k) and (k, j) elements of Σ_t are given by

$$Cov[\xi_t(j), \xi_t(k) | \mathcal{F}_t] = \frac{1}{c(j)c(k)} Cov[(\Delta S_{t+j\delta})^2, (\Delta S_{t+k\delta})^2 | \mathcal{F}_t]$$

while the diagonal elements of Σ_t are given as

$$\text{Var} [\xi_t(j) | \mathcal{F}_t] = \frac{1}{c(j)^2} \text{Var} [(\Delta S_{t+j\delta})^2 | \mathcal{F}_t]$$

The relevant expressions for the conditional variance and co-variance are given in Appendix A.3

An interesting case arises when $\beta \rightarrow 0$. From expression (2.18) it directly follows that:

$$\lim_{\beta \rightarrow 0} \xi_t(j) = \frac{(\Delta S_{t+j\delta})^2 - (\mu\delta)^2}{\delta}$$

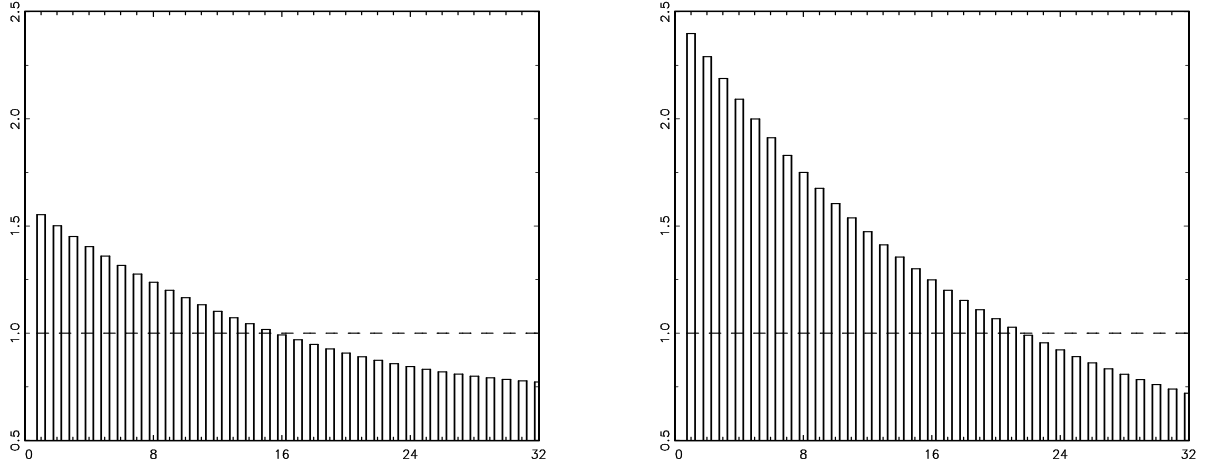
which basically says that the sample variance and any weighted sum (with weights summing to 1) will yield an unbiased estimator of V_t . When $\beta = 0$ or the drift of the volatility process becomes zero, the return variance process is a martingale and $E[V_{t+\tau} | \mathcal{F}_t] = V_t$ for $\tau > 0$. By iterative expectation, we have $E[\frac{(\Delta S_{t+(j+1)\delta})^2 - (\mu\delta)^2}{\delta} | \mathcal{F}_t] = E[E[\frac{(\Delta S_{t+(j+1)\delta})^2 - (\mu\delta)^2}{\delta} | \mathcal{F}_{t+j\delta}] | \mathcal{F}_t] = E[V_{t+j\delta} | \mathcal{F}_t] = V_t$. It is quite intuitive from this result that, the weighted sample variance of asset returns over time period $[t, t + T]$ constitutes an unbiased estimator of the instantaneous variance V_t . Furthermore, when $\delta \rightarrow 0$, we have

$$\lim_{\delta \rightarrow 0} \lim_{\beta \rightarrow 0} \frac{(\Delta S_{t+j\delta})^2}{\delta} - \xi_t(j) = 0$$

That is, when the sampling frequency goes to infinity or sampling interval goes to zero, the mean return term becomes negligible. When all weights are set equal, the estimator of the instantaneous variance coincides with the realized variance measure that is commonly used in financial econometrics, namely the sum of squared intra-period returns. In other words, when $\beta = 0$ and $\delta \rightarrow 0$, i.e. volatility is a first difference martingale and the asset returns are sampled almost continuously, our estimator is equivalent to the model-free realized volatility measure (see expression (2.21) below) except that our estimator provides optimal weights to the sum of squared returns.

To conclude, we include a brief illustration of the behavior of the weighting vector for the SV model. From the moment expressions in Appendix A.3, it can be seen that, unlike for the SM model, the shape of the optimal weighting vector depends on both the model parameters and the instantaneous value of the latent variable. Hence, for given parameters and value of the latent variable we can trace out the shape of the optimal weights. Figure 2.3 visualizes this dependence of the optimal weights on the level of the persistence parameter. In particular, for $\beta = 0.1$ and $\beta = 0.8$ we plot the weights that multiply the squared returns in the UMV estimator, i.e. $W(j)/c(j)$ for $j = 1, \dots, 1/\delta$. Notice that these re-scaled weights measure the structural difference between our proposed estimator and the realized variance estimator. The remaining SV parameters are set equal to $\mu = 0$, $\sqrt{\alpha} = 0.025$, $\sigma_v = 0.0075$, and $\rho = -0.25$ and the sampling frequency is fixed at 15 minutes based on 8 hours of trading per day (i.e. 32 intra-day observations). The downward sloping weights indicate that future observations on the state variable become less informative with an increase in the displacement (i.e. j). However, when the persistence increases, the slope of the weighting scheme decreases. In the limit, when $\beta \rightarrow 0$, the weighting function is flat and future observations are equally informative. Also, for the range of chosen parameter values, the weights deviate substantially from their naive realized variance counterparts (i.e. constant at 1). The performance of the proposed estimator can therefore be expected to be quite different from that of the realized variance measure.

FIGURE 2.3: OPTIMAL UMV FILTER WEIGHTS FOR THE SV MODEL



Notes: The re-scaled optimal weights, $W(j)/c(j)$ for $j = 1, \dots, 1/\delta$, that multiply squared returns, $(\Delta S_{t+j\delta})^2$, for $\beta = 0.1$ (high persistence, left panel) and $\beta = 0.8$ (low persistence, right panel) and a sampling frequency of 15 minutes. The remaining SV model parameters are set as $\mu = 0$, $\sqrt{\alpha} = 0.025$, $\sigma_v = 0.0075$, $\rho = -0.25$ and the instantaneous variance is set equal to its unconditional mean in both cases, i.e. $V_t = \alpha$.

Simulation Design. In order to investigate the performance of the above estimator, we perform a simulation experiment. The notation and much of the simulation design is similar to that of the SM model discussed above. Based on the SV model in (2.17), we simulate⁶ a time series of high frequency observations on both the observed and unobserved state variables, i.e. $\{S_{t+(k-1)\delta}, V_{t+(k-1)\delta}\}$ for $k = 1, \dots, 1/\delta$, and $t = 1, \dots, T + TT - 1$. From this simulated sample, we record the value of the latent variable at the beginning of each period, i.e. $\{V_t\}_{t=1}^{TT}$, and use the high frequency realizations of the observed state variable to construct the following four estimators.

1. **UMV FILTER.** Based on the results in Lemma 2.2.3, we compute the SV model-consistent estimator with optimal weights⁷ using high frequency observations of the state variable over a *sampling horizon* of length T , i.e. $\{S_{t+(k-1)\delta}\}_{k=1}^{T/\delta}$. In particular,

$$\hat{V}_t^{sv} = W' \xi_t \tag{2.20}$$

where the j^{th} element of the $N \times T$ vector ξ , and the optimal weighting matrix W , are given by expressions (2.18) and (2.19) respectively.

2. **REALIZED VARIANCE.** At a sampling horizon of 1 day (i.e. $T = 1$), we compute the realized variance

⁶As opposed to the SM model, the exact sampling path is unavailable for the SV model. To reduce simulation error to a minimum, we use an Euler scheme with 10 discretization steps per minute. Starting values, S_1 and V_1 , are random draws from their respective marginal distribution.

⁷Because the optimal weights are a function of the latent variable to be estimated, a circularity occurs. For a discussion on how to construct the optimal weights we refer to Section 2.4 below.

measure, denoted by \hat{V}_t^{rv} as the sum of squared intra-daily returns, i.e.

$$\hat{V}_t^{rv} = \sum_{k=1}^{1/\delta} (\Delta S_{t+k\delta})^2. \quad (2.21)$$

Notice that \hat{V}_t^{rv} corresponds to \hat{V}_t^{sv} when $\mu = 0$, the variance process has infinite persistence ($\beta = 0$), the sampling horizon coincides ($T = 1$), and the weighting vector is flat ($W(j) = \delta$).

3. ARCH FILTER. Nelson and Foster (1994) show that for various diffusion processes, the difference between the true conditional variance and the variance estimate produced by a *misspecified* ARCH model vanishes in the continuous time limit. Based on the asymptotic distribution of the measurement error, Nelson and Foster (1994) derive the asymptotically “optimal” ARCH filter which, for the square-root SV model in (2.17), takes the following form:

$$\hat{y}_{t+\delta} = \hat{y}_t + \delta \left[e^{-\hat{y}_t} (\beta\alpha - \sigma^2/2) - \beta \right] + \sqrt{\delta}\sigma e^{-\hat{y}_t/2} \left[\rho \bar{R}_{t+\delta} e^{-\hat{y}_t/2} + \left(\bar{R}_{t+\delta}^2 e^{-\hat{y}_t} - 1 \right) \sqrt{(1 - \rho^2)/2} \right]$$

where $y_t \equiv \ln V_t$ and $\bar{R}_{t+\delta} \equiv [S_{t+\delta} - S_t] \delta^{-1/2}$. For our purpose, we define the Nelson-Foster filter as

$$\hat{V}_t^{nf} = \exp(\hat{y}_t). \quad (2.22)$$

It is noted that the filter is closely related to the EGARCH specification proposed by Nelson (1991) in that it is specified in terms of logarithmic variance, includes both returns as well as squared returns, and has a first order autoregressive structure. A noticeable difference with typical ARCH specifications, however, is that the period- t (logarithmic) variance is a function of period- t returns instead of lagged returns and that lagged variance impacts in a highly non-linear fashion.

4. EWMA FILTER. The exponentially weighted moving average (EWMA) filter is given by

$$\hat{V}_t^{ma} = \lambda \hat{V}_{t-\delta}^{ma} + (1 - \lambda) \bar{R}_{t-\delta}^2 \quad (2.23)$$

where $\bar{R}_{t+\delta} \equiv [S_{t+\delta} - S_t] \delta^{-1/2}$ and $0 \leq \lambda < 1$. The parameter λ controls the persistence and is typically chosen close to one. The EWMA filter is widely used in practice; see for example Hull (2003) or the RiskMetrics Technical Document (Morgan Guaranty Trust Company 1996).

It is emphasized that while the estimator developed in this paper is aimed at estimating the *instantaneous* variance, the realized variance measure is closer related to the integrated variance of the return process, namely:

$$IV_{t-1}^t \equiv \int_{t-1}^t V_\tau d\tau \approx \frac{1}{N} \sum_{k=1}^N V_{t+(k-1)\delta}$$

The relation between this notion of integrated volatility and instantaneous volatility defined in the continuous time literature is clear as $d(IV_s^t) = V_t dt$ for any constant $s < t$. As pointed out in Andersen, Bollerslev, Diebold, and Labys (2001, 2003), when the returns are sampled at sufficiently high frequency, the ex post realized variance measure, namely the sum of intra day squared return $\sum_{k=1}^{1/\delta} (\Delta S_{t+k\delta})^2$, approximates the integrated, and not the

instantaneous, volatility arbitrarily well under certain conditions on the underlying process. Therefore, unlike \hat{V}_t^{sv} , we do not expect \hat{V}_t^{rv} to yield an unbiased estimate of V_t although the difference between the instantaneous variance and the (scaled) integrated variance will diminish when the persistence of the variance process increases. The Nelson-Foster ARCH filter and the EWMA filter⁸ are directly comparable to the instantaneous variance and are therefore also included in this study.

The SV model parameters are set as $\mu = 0$, $\sqrt{\alpha} = 0.025$, $\sigma_v = 0.0075$, $\rho = -0.25$ corresponding to an annualized return volatility of close to 40%. Because the optimal weights, and possibly the performance of the various filters, may depend on the specific choice of model parameters we distinguish among three cases where the persistence parameter of the variance process is varied, i.e. $\beta = 0.1$ (high persistence), $\beta = 0.4$ (intermediate persistence), and $\beta = 0.8$ (low persistence). To investigate the impact of sampling frequency and sample size we fix $T = 1$ and vary $\delta = \left\{ \frac{1}{60 \times 8}, \frac{5}{60 \times 8}, \frac{15}{60 \times 8}, \frac{30}{60 \times 8} \right\}$, corresponding to 1, 5, 15, and 30 minute data based on a trading day of 8 hours. We set $TT = 2520$, corresponding to a high frequency sample path with a length of 10 years.

As for the SM model, we assess the relative performance of the four competing filters by regressing the measurement error, i.e. $\hat{V}_t^{sv} - V_t$, $\hat{V}_t^{rv} - V_t$, $\hat{V}_t^{nf} - V_t$, and $\hat{V}_t^{ma} - V_t$ on a constant and the actual realizations of the latent variable, V_t , for $t = 1, \dots, TT$. The filter is unbiased when the intercept and slope coefficients are both insignificantly different from zero. The efficiency of the filter can be gauged by the R^2 and $RMSE$ (see the discussion on page 32 above for more details).

To reduce variation in the simulation results, Table 2.3 reports the *average* slope coefficient (plus associated average absolute t-statistics), R^2 , and $RMSE$ obtained from 50 independent simulation runs. In order to implement the EWMA filter, the persistence parameter λ should be set to a “reasonable” value. Since the EWMA is effectively an integrated GARCH(1,1) model, one approach is to use the simulated return data to estimate λ . Unfortunately, it turns out that this only yields reasonable results when the persistence of the variance process is relatively high (i.e. $\beta < 0.25$). For the simulations with $\beta = 0.4$ and $\beta = 0.8$ we find that the EWMA persistence parameter jumps to the boundary (i.e. $\lambda = 1$) which can be explained as follows. In the extreme case when persistence is infinite, and the conditional variance constant, λ should equal one (with appropriate initial conditions on \hat{V}_0^{ma}) as this implies constant EWMA variance. It is this tendency of λ to converge to one when the persistence of the variance process diminishes that explains this finding. Hence, an alternative approach is implemented to determine λ , namely in each simulation run we set λ equal to the value that maximizes the R^2 of the regression of \hat{V}_t^{ma} on \hat{V}_t . For $\beta = 0.1$, this approach generates virtually indistinguishable results as compared to the IGARCH estimation approach. For $\beta = 0.4$ and 0.8 , maximizing the R^2 yields much better results. It should be stressed that, especially when persistence is low, the EWMA severely misspecifies the volatility dynamics and the IGARCH restriction is clearly not appropriate. Finally - depending on the filter, sampling frequency, and parameters - the measurement error can be heteroskedastic and serially correlated. Table 2.3 therefore reports heteroskedastic and autocorrelation (HAC) consistent t-statistics (Newey and West 1987) computed using 15 lags. Although the choice of lag length is relatively ad hoc, and no data-driven lag selection procedure (Newey and

⁸The EWMA filter is effectively the integrated GARCH model of Bollerslev and Engle (1986). Hence, its persistence parameter λ can be estimated by specifying a GARCH(1,1) process for returns.

West 1994) is implemented, inspection of the measurement error time series properties indicates that the order of serial correlation does not extend beyond 15 for the simulations we consider.

Simulation Results. The second column of Table 2.3 (“UMV Filter”) illustrates that irrespective of the sampling frequency or mean reversion parameter, the SV model-consistent estimator of the instantaneous variance is unbiased. Based on the heteroskedastic and autocorrelation consistent t-statistics, both the intercept (not reported) and slope coefficients are insignificantly different from zero. As expected, the coefficient of determination is close to zero indicating that there is no systematic component in the measurement error that can be explained by realizations of the latent variable. From the MSE statistic it is evident that (i) the efficiency of the estimator increases with an increase in the sampling frequency and (ii) the latent variance component is best identified when the process is persistent.

Turning to the alternative filters, we find that each and every one constitutes a biased estimator of the instantaneous variance. As expected, this bias is most pronounced when mean reversion is high and persistence low. Regarding the realized variance measure, we have noted that when persistence is low, the “optimal” weights deviate quite substantially from the implicit realized variance weights (Figure 2.3). The reported difference in performance can therefore be expected, even more so because the realized variance measure is known to estimate the integrated and not the instantaneous return variance. This argument is, however, not valid for the ARCH and EWMA filters. Nevertheless, the slope coefficients reported in the fourth and the fifth column of Table 2.3 are significantly different from zero irrespective of the model parameters or sampling frequency suggesting that also these filters deliver biased estimates. Moreover, the R^2 coefficients indicate that a substantial degree of structure remains in the measurement error which, as for the SM model, could suggest that these filters are “over-smoothing”. To investigate this a little further, Figure 2.5 plots the measurement error of the Kalman filter for the SM model, the Nelson-Foster ARCH filter for the SV model and the UMV filter for both the SM and the SV model, against the actual realizations of the latent variable (stochastic mean and volatility). It is clear that the measurement error of the UMV filter has no structure left. This is not the case for either the Kalman filter or the ARCH filter. Both these filters appear “over-smooth”. In particular, a realization of the latent variable below (above) the unconditional mean of the process is likely to induce a negative (positive) measurement error indicating that the estimate is too high (low) relative to the true value. Although in a different context, this finding is in line with the discussion on page 32 above. Another interesting pattern which arises is that the MSE for the UMV filter is often higher than for the realized variance measure and the EWMA filter, and in all cases higher than the MSE of the ARCH filter. We argue that this is a consequence of the weighting scheme of the UMV filter. The UMV weights are aimed at minimizing the variance of the estimator *under the restriction that the estimator is unbiased*. The flat weights for the realized variance measure, for instance, may lead to a lower RMSE but render the estimator biased. Regarding the ARCH filter, we find that it achieves the lowest MSE among all competing filters. This finding is in line with the work by Nelson and Foster (1994) who show it is asymptotically optimal. The ranking of the remaining filters is less obvious and varies with the model parameters and sampling frequency.

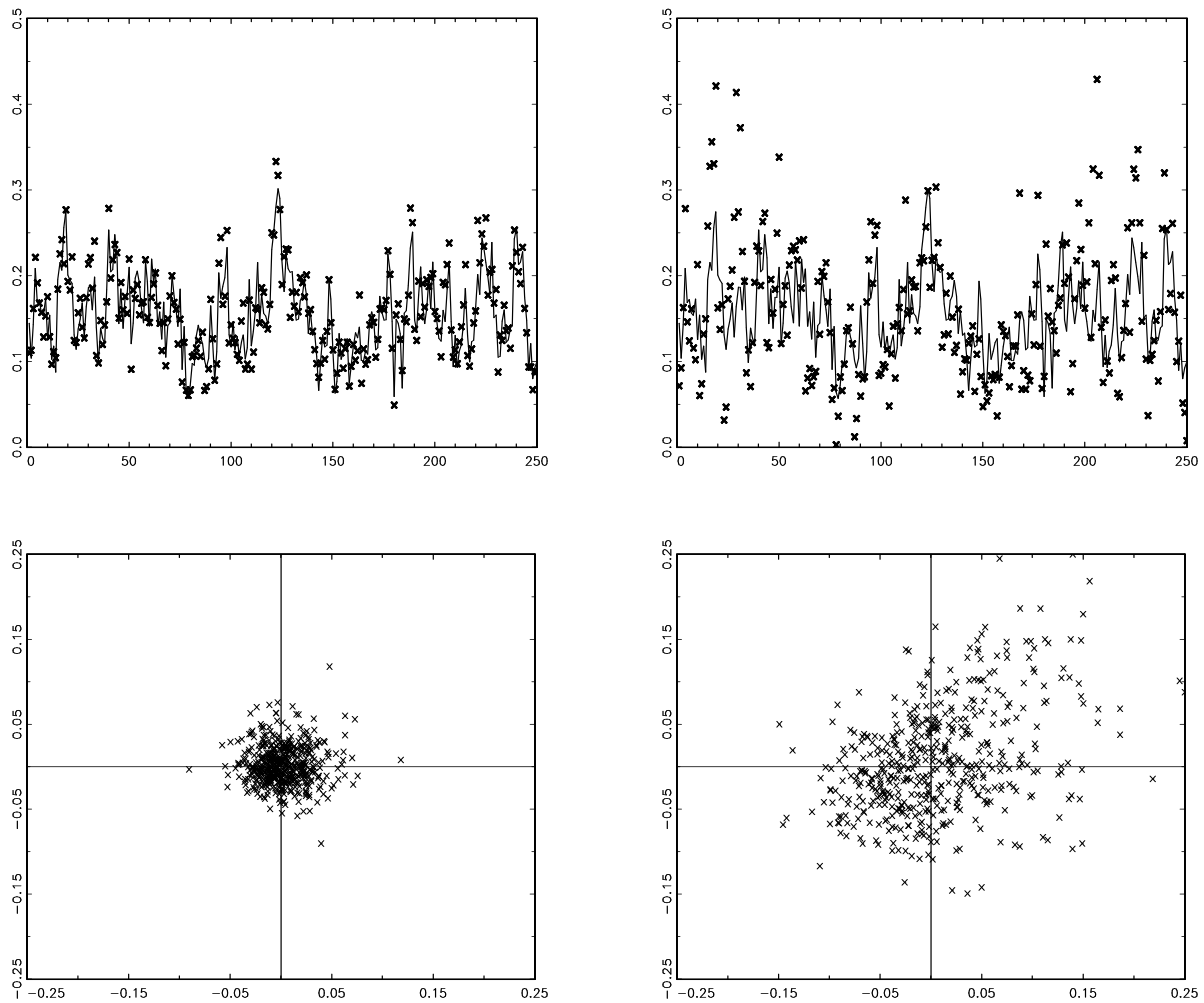
As mentioned above, an alternative to varying the sampling *frequency* is to vary the sampling *horizon*. For example, a sampling scheme where $T = 1$ day and $\delta = 5$ minutes generates the same number of observations as

TABLE 2.3: RELATIVE PERFORMANCE OF FILTERS FOR THE STOCHASTIC VOLATILITY MODEL

Par	δ (min)	T (day)	T/δ (obs)	UMV Filter			Realized Variance			ARCH Filter			EWMA(λ) Filter				
				π_1	R^2	MSE	π_1	R^2	MSE	π_1	R^2	MSE	π_1	R^2	MSE	λ	
I	1	1	480	0.001 (0.96)	0.14%	0.94	-0.049 (6.67)	3.02%	1.16	-0.045 (7.11)	4.17%	0.90	-0.026 (3.84)	1.40%	0.94	0.99	
	5	1	96	0.000 (0.85)	0.10%	1.48	-0.049 (4.64)	1.84%	1.50	-0.090 (8.88)	7.75%	1.33	-0.055 (4.75)	2.73%	1.38	0.98	
	15	1	32	0.003 (0.82)	0.13%	2.19	-0.046 (2.78)	0.89%	2.12	-0.137 (9.10)	10.3%	1.77	-0.086 (4.81)	4.03%	1.79	0.97	
	30	1	16	0.005 (0.90)	0.14%	2.95	-0.043 (2.06)	0.52%	2.82	-0.175 (8.92)	10.7%	2.22	-0.115 (5.05)	5.22%	2.09	0.96	
	30	2	32	0.003 (0.92)	0.24%	2.49	-0.043 (2.06)	0.52%	2.82	-0.175 (8.92)	10.7%	2.22	-0.115 (5.05)	5.22%	2.09	0.96	
	30	6	96	0.001 (1.00)	0.38%	2.35	-0.043 (2.06)	0.52%	2.82	-0.175 (8.92)	10.7%	2.22	-0.115 (5.05)	5.22%	2.09	0.96	
	30	30	480	0.002 (0.99)	0.42%	2.35	-0.043 (2.06)	0.52%	2.82	-0.175 (8.92)	10.7%	2.22	-0.115 (5.05)	5.22%	2.09	0.96	
	II	1	1	480	0.001 (1.00)	0.09%	0.93	-0.176 (16.7)	11.4%	1.09	-0.142 (17.1)	12.5%	0.84	-0.082 (8.66)	3.92%	0.87	0.99
		5	1	96	0.001 (0.85)	0.06%	1.50	-0.176 (12.1)	7.03%	1.38	-0.259 (21.5)	20.2%	1.20	-0.164 (11.0)	7.27%	1.27	0.98
		15	1	32	0.003 (0.84)	0.06%	2.23	-0.173 (8.18)	3.56%	1.92	-0.357 (22.4)	23.9%	1.52	-0.238 (11.4)	9.24%	1.63	0.96
		30	1	16	0.004 (0.79)	0.06%	3.01	-0.171 (6.01)	2.01%	2.53	-0.416 (21.3)	23.8%	1.77	-0.286 (11.0)	9.70%	1.91	0.94
	III	1	1	480	0.002 (0.88)	0.06%	0.96	-0.311 (24.3)	20.1%	1.03	-0.242 (23.8)	20.1%	0.80	-0.146 (12.1)	6.50%	0.85	0.99
5		1	96	0.002 (0.81)	0.05%	1.62	-0.311 (17.5)	12.1%	1.32	-0.402 (29.4)	28.8%	1.10	-0.264 (14.8)	10.2%	1.23	0.98	
15		1	32	0.002 (0.79)	0.05%	2.49	-0.308 (11.8)	6.04%	1.85	-0.512 (29.9)	31.2%	1.35	-0.341 (13.8)	9.89%	1.61	0.96	
30		1	16	0.002 (0.79)	0.04%	3.39	-0.305 (8.71)	3.40%	2.45	-0.573 (29.1)	30.5%	1.54	-0.374 (12.1)	8.11%	1.95	0.93	

Notes: Slope coefficients of the regressions $\hat{V}_t^{filter} - V_t = \pi_0 + \pi_1 V_t + \varepsilon_t^{filter}$ for the filters given by expressions (2.20), (2.21), (2.22), and (2.23). The simulations are based on 50 simulated times series with a length of 10 years. The persistence parameter is set equal to $\beta = 0.1$ ("Par I", high persistence), $\beta = 0.4$ ("Par II", intermediate persistence), and $\beta = 0.8$ ("Par III", low persistence). The remaining SV model parameters are set as $\mu = 0$, $\sqrt{\alpha} = 0.025$, $\sigma_v = 0.0075$, $\rho = -0.25$. Absolute value of the associated HAC[15]-consistent t-statistics are in parenthesis below where $H_0 : \pi_1 = 0$. The column " R^2 ", reports the coefficient of determination while the column "MSE" reports the root mean squared error ($\times 10, 000$). The persistence parameter λ for the EWMA scheme is optimized over the R^2 of the regression.

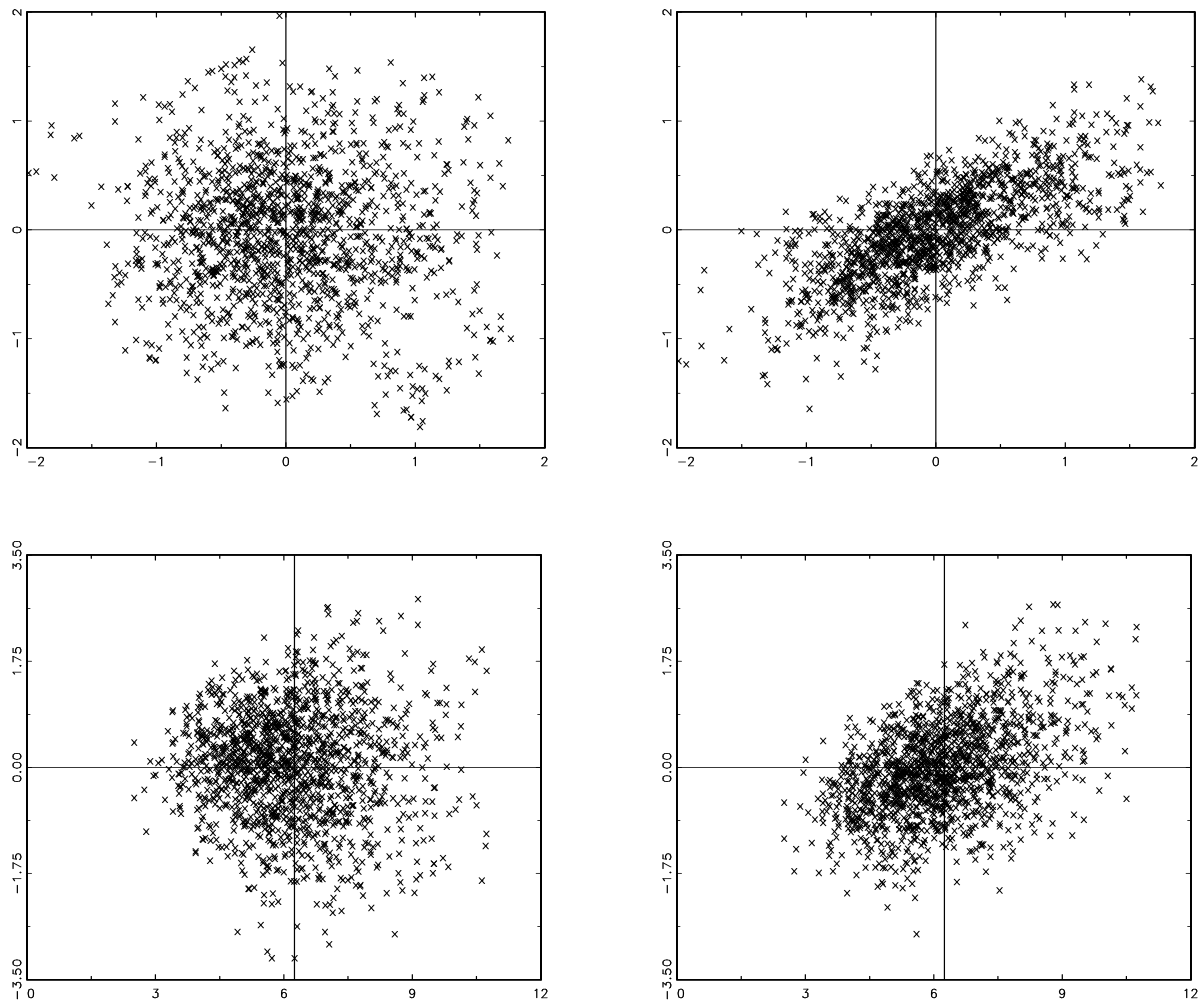
FIGURE 2.4: UMV FILTER FOR ALTERNATIVE “FREQUENCY - HORIZON” SAMPLING SCHEMES



Notes: The top left panel plots a time series of UMV filter estimates (crosses) together with the actual latent state variable (solid line) based on a 1 minute sampling frequency and a 1 day horizon. The bottom left panel plots the measurement error against its first lagged value. The right panel plots analogous results for a 30 minute sampling frequency over a 30 day horizon. The SV parameters are set as $\mu = 0$, $\sqrt{\alpha} = 0.025$, $\beta = 0.4$, $\sigma_v = 0.0075$, $\rho = -0.25$.

a sampling scheme where $T = 6$ days and $\delta = 30$ minutes. Nevertheless, the performance of the UMV filter is not expected to be invariant to the specific sampling scheme and we will therefore briefly study its impact based on some simulations. In particular, for three out of four sampling schemes discussed before, i.e. $T = 1$ day and $\delta = 1, 5, 15$ minutes, we consider the following three alternative sampling schemes, i.e. $\delta = 30$ minutes and $T = 2, 6, 30$ days, which generate the same number of observations at a lower frequency and over a longer horizon. Based on the sampled data, $\{S_{t+(k-1)\delta}\}_{k=1}^{T/30\text{Min}}$, we again compute the SV model-consistent estimator and compare it to the actual realization of the latent variable. To conserve space, we only report the results for Par I in Table 2.3 (between the dotted lines). The results for Par II and III are qualitatively the same. Based on the MSE statistic it is clear that, for a fixed sampling frequency, an increase in the sampling horizon leads to

FIGURE 2.5: MEASUREMENT ERROR AGAINST LATENT VARIABLE



Notes: The top panel plots the measurement error ($\times 10,000$) of the latent mean against the actual realization of the latent mean for the UMV filter (left) and the Kalman filter (right). The sampling frequency is set to 1 day, the sampling horizon equal to 96 days and the SM parameters are set equal to Par II (see Table 2.1). The bottom panel plots the measurement error ($\times 100$) of the latent volatility against the actual realization of the latent volatility for the UMV filter (left) and the Nelson-Foster ARCH filter (right). The sampling frequency is set to 5 minutes, the sampling horizon equal to 1 day and the SV parameters are set equal to Par III (i.e. $\beta = 0.8$).

better performance of the UMV filter. In particular, at a 30 minute sampling frequency, an increase in the horizon from 1 day to 30 days leads to a reduction in MSE of about 20%. The more relevant comparison, however, is the one described above. So when sampling 32, 96, and 480 data points, the gain in MSE associated with a “high-frequency-short-horizon” sampling scheme – relative to a “low-frequency-long-horizon” sampling scheme – is about 12%, 37%, and 60% respectively. Thus, as opposed to the SM model, it is clear that for the estimation of the instantaneous variance, high frequency data over short horizons is preferred to low frequency data over long horizons. This point is also illustrated in Figure 2.4.

2.4 Estimation of Latent Variable Models in Finance

Continuous time latent variable models have attracted a great deal of attention in the finance literature because they provide a useful framework for the modelling of return dynamics. Prominent examples include the baseline stochastic volatility (SV) model (Harvey, Ruiz, and Shephard 1994) or one of its many generalizations which incorporate multiple volatility components and random jumps. The major appeal of these type of models is that they are able to capture a number of salient features of financial asset returns including the presence of jumps in asset prices, time varying return volatility, and a skewed and fat-tailed marginal return distribution. In addition, many of the latent variable models have associated closed form or semi-closed form expressions for the arbitrage free price of important financial securities such as bonds, options, futures, and volatility derivatives. It is therefore not surprising that much effort has been spent on finding the model specification which is not only able to capture the time series dynamics of asset returns, but also accurately prices derivative securities.

As mentioned above, the estimation of these model is far from trivial and a wide variety of distinct inference procedures for nonlinear latent variable models in general, and SV models in particular, have been proposed. For instance, generalized method of moments (GMM) based estimation has been proposed by Andersen and Sørensen (1996), Ho, Perraudin, and Sørensen (1996), Melino and Turnbull (1990), and Taylor (1986). Quasi Maximum Likelihood (QML), which relies on a state-space form transformation and the Kalman filter, has been proposed by Harvey, Ruiz, and Shephard (1994). Although both GMM and QML are straightforward to implement, there have been indications that the small sample properties of these methods are poor (Jacquier, Polson, and Rossi 1994). Alternatively, a wide range of simulation based approaches have been developed, including the simulated method of moments (SMM) proposed by Duffie and Singleton (1993), simulated maximum likelihood (SML) proposed by Daniélsson (1994), the efficient methods of moments (EMM) proposed by Gallant and Tauchen (1998), the Monte Carlo maximum likelihood (MCL) proposed by Sandmann and Koopman (1998), direct maximum likelihood estimation through recursive numerical integration by Fridman and Harris (1998), and Markov Chain Monte Carlo (MCMC) first implemented by Jacquier, Polson, and Rossi (1994) and further developed by Kim, Shephard, and Chib (1998) and Elerian, Chib, and Shephard (2001). A major drawback of the simulation based methods is that they are computationally intensive and involve discretization when applied to the continuous-time processes. An approach which circumvents these difficulties has been proposed by Carrasco, Chernov, Florens, and Ghysels (2001), Chacko and Viceira (1999), Jiang and Knight (2002), and Singleton (2001) who develop an estimation methodology based on the empirical characteristic function (ECF), and Meddahi (2001) who exploits the eigenfunction expansion of the latent volatility process.

Based on this wide range of estimation methodologies, a number of distinct latent variable models have been estimated. For example, the discrete time SV model has been estimated by Andersen and Sørensen (1996) using GMM, by Harvey and Shephard (1996) using QML, by Gallant, Hsieh, and Tauchen (1997) using EMM, by Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998) and Chib, Nardari, and Shephard (2002) using MCMC, by Liesenfeld and Jung (2000) using SML, and by Sandmann and Koopman (1998) using MCL. “Simple” continuous time SV models have been estimated by Ho, Perraudin, and Sørensen (1996) using GMM, by Andersen and Lund (1997) using EMM, Jiang and Knight (2002) using the ECF, and by Chacko and Viceira

(1999) using an ECF-related spectral GMM approach. Finally, more complicated affine and non-affine SV models with or without multi-factor volatility components and state-dependent jump components in returns and volatility have been estimated by Andersen, Benzoni, and Lund (2002) and Chernov, Gallant, Ghysels, and Tauchen (2002) using EMM, by Eraker, Johannes, and Polson (2002) and Jones (1998) using MCMC, and by Pan (2002) using “implied-state” GMM where the latent volatility component is inverted from observed option prices.

2.4.1 GMM Estimation based on the Model-Consistent Unbiased Estimator of the Latent Variables

In this section we outline yet another approach to the statistical inference of continuous time asset return models with latent state variables. The basic idea is to implement the GMM with conditional moment restrictions on both the observed and the unobserved variables which are then evaluated at the unbiased estimate of the latent variables. Due to the analytic tractability of the AJD model, closed form expressions for the conditional moments or cumulants are often available and based on the unbiased measurement of the latent variable, as described above, the moment restrictions for the unobserved variables can be evaluated in a straightforward fashion. This approach is distinct from the above mentioned approaches in important ways. First and foremost, our approach exploits the availability of *high frequency* observations of state variables and through the estimator of the latent variables. Because it is derived under the exact parametric specification of the model, the estimator is consistent with the volatility measure under the specific model. Further, unlike SMM, EMM, and MCMC, the GMM estimation approach we propose does not require simulation and is based on *exact* moments or cumulants in the sense that they correspond to the continuous-time DGP without any discretization or approximation error.

In a closely related study, Bollerslev and Zhou (2002) use realized volatility as a proxy for the integrated variance process and propose a similar GMM approach estimation that is based on conditional return and variance moments of the process. The distinguishing element of our study is that we base our estimation procedure on an estimator of the instantaneous value of the latent variable and is not limited to SV models but can, in principle, be applied to any latent variable model in the AJD class.

For completeness we will briefly outline the general GMM procedure, after which we specialize the discussion around the estimation of the AJD model based on the proposed latent variable estimator. GMM estimation of a $p \times 1$ parameter vector θ_0 requires the specification of an $r \times 1$ ($r \geq p$) moment restriction vector, $f_t(\theta)$, which has expectation zero when evaluated at the true population parameter θ_0 , i.e. $E[f_t(\theta_0)] = 0$. The GMM procedure then minimizes a quadratic form of the moment restrictions over the admissible parameter space, Θ , i.e.

$$\hat{\theta}_T = \arg \min_{\theta \in \Theta} J_T(\theta) \quad \text{where} \quad J_T(\theta) \equiv g_T(\theta)' \Lambda g_T(\theta),$$

where $g_T(\theta) = \frac{1}{T} \sum_{t=1}^T f_t(\theta)$, and Λ is an $r \times r$ positive semi-definite weighting matrix. When $r = p$ the parameters are exactly identified. The parameter estimates are independent of the particular choice of weighting matrix and the objective function $J_T(\theta)$ will be zero at the minimum. Theory suggests, however, that the efficiency of the estimator may be improved by increasing the dimension of $g_T(\theta)$, i.e. using more moment restrictions than parameters to be estimated ($r > p$). In this case, the parameter estimates do depend on the specific choice of

weighting matrix. In fact Hansen (1982) has shown that choosing $\Lambda = \Sigma(\theta)^{-1}$ where $\Sigma(\theta) = E[f_t(\theta)f_t(\theta)']$ results in the GMM estimator of θ with the lowest asymptotic covariance. This “efficient” GMM estimator is typically implemented using a two-step procedure: in the first step the parameter vector is estimated using an arbitrary psd weighting matrix, resulting in θ_1 , while in the second step the parameter vector is re-estimated using $\Lambda = \widehat{\Sigma}_1(\theta)^{-1}$. Because GMM estimation allows for serial correlation in the moment restrictions, a Newey-West type estimator can be used to obtain $\widehat{\Sigma}(\theta)$. Regarding the choice of r , Andersen and Sørensen (1996) find that the efficiency gain resulting from additional moment restrictions is countered by the deterioration of the $\frac{r(r+1)}{2}$ estimated coefficients needed to construct the (optimal) weighting matrix Λ . For further discussion of (optimal) moment selection see also Gallant and Tauchen (1996) and Pan (2002). Under regularity conditions, as specified in Hansen (1982), the estimator $\widehat{\theta}_T$ is consistent and asymptotically normal with covariance matrix given by:

$$\frac{1}{T} \left[D'(\widehat{\theta}_T) \Sigma^{-1}(\widehat{\theta}_T) D'(\widehat{\theta}_T) \right]^{-1}$$

where $D(\theta)$ is the Jacobian of $g_T(\theta)$ with respect to θ . Due to the form of Λ , the objective function $J_T(\theta)$ is a measure of the distance between $g_T(\theta)$ and zero. The objective function at the minimum can therefore be regarded as a goodness-of-fit test for the model, i.e. a high value of this test suggests that the model is misspecified. In particular, it can be shown that $TJ_T(\widehat{\theta}_T) \stackrel{a}{\sim} \chi_{r-p}^2$ under the null hypothesis that the model is correctly specified.

In setting up the GMM estimation of the AJD model, we exploit the fact that in many cases the joint characteristic function of $S_{t+\delta}$ or $\Delta S_{t+\delta}$ and $V_{t+\delta}$ is known in closed form. As we have seen above, this allows us to derive closed form conditional moment or cumulant expressions for both the observed and the unobserved state variables in terms of the model parameters and possibly the instantaneous value of (some) state variables. By the definition of the characteristic function we have that the conditional cumulant of order (l_s, l_v) can be derived as:

$$K \left(\Delta S_{t+\delta}^{l_s} V_{t+\delta}^{l_v} | \mathcal{F}_t \right) = \frac{\partial^{l_s+l_v} \ln \psi(u_1, u_2, \Delta S_{t+\delta}, V_{t+\delta}, t, \tau | \theta_0, S_t, V_t)}{i^{l_s+l_v} \partial u_1^{l_s} \partial u_2^{l_v}} \Bigg|_{u=0}$$

for $l_s, l_v \in \{0, 1, \dots\}$. Based on this results, we can construct moment restrictions⁹ as follows:

$$f_t(\theta) = \left\{ \begin{array}{l} K \left(\Delta S_{t+\delta}^{l_{s,1}} V_{t+\delta}^{l_{v,1}} \right) - \frac{\partial^{l_{s,1}+l_{v,1}} \ln \psi(u_1, u_2, \Delta S_{t+\delta}, V_{t+\delta}, t, \tau | \theta, S_t, V_t)}{i^{l_{s,1}+l_{v,1}} \partial u_1^{l_{s,1}} \partial u_2^{l_{v,1}}} \Bigg|_{u=0} \\ \vdots \\ K \left(\Delta S_{t+\delta}^{l_{s,r}} V_{t+\delta}^{l_{v,r}} \right) - \frac{\partial^{l_{s,r}+l_{v,r}} \ln \psi(u_1, u_2, \Delta S_{t+\delta}, V_{t+\delta}, t, \tau | \theta, S_t, V_t)}{i^{l_{s,r}+l_{v,r}} \partial u_1^{l_{s,r}} \partial u_2^{l_{v,r}}} \Bigg|_{u=0} \end{array} \right\}$$

where $l_{s,1}, l_{v,1}, \dots, l_{s,r}, l_{v,r} \in \{0, 1, \dots\}$ and $t \in \{0, 1, \dots\}$. The first term of each entry in $f_t(\theta)$ is the sample cumulant¹⁰ while the second term is the conditional cumulant expression derived from the model in terms of the parameter vector θ and state variable S_t and V_t . Notice, however, that because the moment restriction above is conditional on the realization of the latent variable it cannot be evaluated. The solution we propose is to replace V_t by its estimated counterpart, \widehat{V}_t . This way, the moment restriction can be evaluated while consistency of the

⁹Instrumental variables, i.e. $z_t \in \mathcal{F}_t$, that are uncorrelated with $f_t(\theta_0)$, can be used to generate additional moment restrictions because $E[f_t(\theta_0) \otimes (1, z_t)] = 0$.

¹⁰The sample cumulant is estimated using $\Delta S_{t+\delta}$ and $V_{t+\delta}$ only. For example, $K(\Delta S_{t+\delta} V_{t+\delta}) = \Delta S_{t+\delta} V_{t+\delta}$

GMM estimator is maintained due to the consistency of \widehat{V}_t . It is noted that the estimation approach is based on the cumulants and not the moments. However, cumulants and moments are identical for orders lower than or equal to two. For orders higher than two, Kendall (1958) outlines the one-to-one correspondence between moments and cumulants. For reference related to cumulants based estimation, see Press (1967), Beckers (1981), and Knight and Satchell (1997).

Although the GMM procedure outlined above stands out as being quite straightforward, there are a number of pitfalls that one should be aware of. We will briefly discuss these below.

Error in Variable and Moment Restrictions The most important drawback of the above method is that in finite sample, the estimator of the latent variable will inevitably contain measurement error, i.e. $\widehat{V}_t = V_t + \varepsilon_t$ where ε has some distribution. Because the latent variable estimator is unbiased, the mean of ε will be zero. Further statements about the distribution of the error term are more difficult to make. Now consider the following moment restriction:

$$\widehat{V}_{t+\delta} - E \left[V_{t+\delta} | \theta, S_t, \widehat{V}_{t+\delta} \right]$$

Provided that the moment restriction is a linear function of the state variable, it will have zero expectation and is thus a valid moment restriction. However, when a non-linear function of the latent state variable enters into the moment restriction it invalidates the restriction because it will not have zero expectation. For instance:

$$E_t \left[\widehat{V}_{t+\delta}^2 - E(V_{t+\delta}^2 | \theta, S_t, \widehat{V}_{t+\delta}) \right] \neq 0$$

In a closely related setting, Bollerslev and Zhou (2002) propose to solve this problem by including a nuisance parameter in the above moment restriction which can absorb the contribution of the measurement error, i.e.

$$E_t \left[\widehat{V}_{t+\delta}^2 + \gamma - E(V_{t+\delta}^2 | \theta, S_t, \widehat{V}_{t+\delta}) \right] \stackrel{?}{=} 0$$

Although the results in their work seem to suggest that this approach works quite well, a disadvantage is that it relies on the IID'ness of ε , which, in many situations may not be justified. An alternative approach to the issue may be to further study the distribution of the measurement error and possibly derive the relevant moments for the *estimated* latent variable instead of the actual realization of the latent variable, i.e.

$$E_t \left[\widehat{V}_{t+\delta}^2 - E(\widehat{V}_{t+\delta}^2 | \theta, S_t, \widehat{V}_{t+\delta}) \right] = 0.$$

Clearly, the feasibility and desirability of the alternative approaches will ultimately depend on the model specification and the properties of the data.

Construction of Optimal Weights It is noted from the definition of the optimal weighting vector (expression 2.11) and the conditional moments for the SV model (Appendix A.3) that the optimal weighting vector may depend directly on both the model parameters *and* the value of the latent and observed variables at time t . For the SM model discussed above, the latent state variable does not enter into the optimal weights which, in this case, depend on the model parameter only. In contrast, the optimal weights for the SV model do depend both on the

model parameters *and* the time- t instantaneous variance. In general, when latent state variables, V , enter into the expression for the optimal weighting vector, its construction is not entirely trivial because in order to get a measurement of the latent variable one would first need to specify a weighting vector. In other words, $\widehat{V}(W, \theta)$ and $W(V, \theta)$ where θ denotes the model parameters and W the *optimal* weighting matrix. Fortunately, an iterative procedure which parallels the construction of the GMM optimal weighting matrix (Hansen 1982) provides the solution here. Specifically, given an initial estimate of the latent variable estimate, V_0 , (based on a naive proxy for example) and an arbitrary weighting matrix, W_0 , the iteration takes the following form (set $r = 0$):

- Step 1: Estimate \widehat{V}_{r+1} based on W_r and given θ
- Step 2: Compute W_{r+1} based on \widehat{V}_{r+1} and given θ
- Step 3: If $|W_{r+1} - W_r| < \eta$ quit, else $r = r + 1$ and goto Step 1

The first step clearly relies on the unbiasedness of the latent variable estimator for arbitrary weights. The initial choice of V_0 is therefore not very crucial but may, if chosen well, speed up convergence substantially. Our experience is that two to four steps are sufficient to obtain reasonable convergence in the weighting vector. When the model parameter vector θ is unknown, the approach would be to maximize the relevant likelihood function over the admissible parameter space with intermediate optimization steps for the construction of the optimal weighting vector. For long time series, this procedure may turn out to be computationally intensity as in every iteration over θ , one would need to calculate an entire sequence of weighting vectors associated with each observation.

Market Microstructure-Effects and Sampling Frequency While the estimator of the latent variable *can* be implemented based on high frequency data, the question remains whether this should in fact be done. The asymptotics indicate that there are substantial benefits from using high frequency data, but an obvious concern is whether the AJD model provides a good description of the data across a range of sampling frequencies. In fact, the nice properties of the proposed estimator are derived under the assumption that the AJD model is the data generating process and is hence “aggregationally consistent” with the observed data. A number of studies have shown that the SV model provides a reasonable description of returns at the daily frequency. However, at higher frequencies, return data seem to contain large amounts of noises which are mainly due to reporting errors and market microstructure effects such as bid/ask bounce, price discreteness, stale trading, etc. Without getting into more detail here, it is safe to say that the noise to signal ratio tends to increase with a decrease in the sampling interval. Importantly, this implies that the aggregational consistency of the SV model must break down at some stage when moving from low frequency to high frequency data. This observation should be taken into account when deciding on the frequency that the latent variable estimator will be based on.

2.5 Conclusion

The closed analytic form of the conditional characteristic function for the AJD class (Duffie, Pan, and Singleton 2000) allows us to derive a dynamic correspondence between the latent variables and the cumulants of the observed state variables. Based on this relation, we derive an unbiased minimum-variance estimator for the latent

variables in terms of the model parameters and observations of the state variables. An important feature of the proposed estimator, is that it is *model-consistent* and can be implemented using *high frequency* data. We also derive conditions under which the estimator has minimum variance. In a simulation study we investigate the properties of several competing latent variable estimators for the stochastic mean and volatility in the SM and SV models. The results illustrate that (i) the UMV estimator delivers unbiased estimates irrespective of sample size, sampling frequency, sampling horizon, or model parameters (ii) for the SM (SV) model the Kalman (ARCH) filter is optimal in a MSE sense (iii) for the estimation of the conditional mean low-frequency-long-horizon sampling is optimal and (iv) for the estimation of the conditional variance high-frequency-short-horizon sampling is optimal. Based on the proposed estimator of the latent variable, we outline a flexible GMM estimation procedure that relies on the matching of conditional moments or cumulants of both the observed and the unobserved state variables. The major advantage of this approach is that it can be implemented using high frequency data, does not require discretization of the continuous time process, and does not involve computationally expensive simulations. A couple of issues are left for future research, including the treatment of the error-in-variable issue for higher order moment restrictions, and the impact of market microstructure effects in high frequency data on the performance of the estimator.

Chapter 3

Statistical Models for High Frequency Security Prices

3.1 Introduction

The distributional properties of financial asset returns are of central interest to financial economics because they have wide ranging implications for issues such as market efficiency, asset pricing, volatility modelling, and risk management. Although the conditional and unconditional distribution of returns at the daily and weekly frequencies have been extensively studied and are typically well understood, this is certainly not the case for returns observed at higher frequencies. Intra-daily patterns in market activity plus numerous market microstructure effects¹ substantially complicate the analysis of so-called “high frequency” data and often render conventional return models inappropriate.

Much of modern finance theory builds on the martingale property of risk-adjusted asset prices, as originally laid out in Cox and Ross (1976) and Harrison and Kreps (1979). The development of econometric models for asset prices has progressed hand in hand and is, as a result, directed to models that are consistent with the martingale hypothesis. A prominent example is the geometric Brownian motion from which the celebrated Black and Scholes option pricing formula has been derived. To capture commonly observed characteristics of daily return data, such as skewness, fat tails and heteroscedasticity, this model has been extended in a number of directions to include for instance random jumps and the stochastic evolution of return variance². Although less suited for derivative pricing, an attractive alternative to the diffusion process is the compound Poisson process. Despite its long tradition in the statistics literature³, the model has received only moderate attention in finance⁴ after it has been introduced by

¹Market microstructure effects include bid/ask spreads, non-synchronous trading, stale prices, and price discreteness. See for example Campbell, Lo, and MacKinlay (1997), Madhavan (2000), O’Hara (1995), Wood (2000).

²See for example Bakshi, Cao, and Chen (1997), Bakshi and Madan (2000), Bates (1996, 2000), Bollerslev and Zhou (2002), Heston (1993), and Scott (1997).

³The Poisson process, often viewed as a special case of a renewal process, has been used extensively in for instance queue theory, ruin and risk theory, inventory theory, evolutionary theory, and bio-statistics. See Andersen, Borgan, Gill, and Keiding (1993), Karlin and Taylor (1981, 1997) and references therein.

⁴For some recent applications of the compound Poisson process in economics, finance, insurance mathematics and risk management

Press (1967, 1968). In its simplest form, the compound Poisson process with iid Gaussian increments is given by:

$$F(t) = F(0) + \sum_{j=1}^{M_I(t)} \varepsilon_j, \quad (3.1)$$

where $F(t)$ denotes the time- t logarithmic asset price, $\varepsilon_j \sim \text{iid } \mathcal{N}(\mu_I, \sigma_I^2)$ and $M_I(t)$ is a homogeneous Poisson process with intensity parameter $\lambda_I > 0$. Press (1967) has shown that the analytical characteristics of this model agree with the empirically observed properties of (low frequency) returns, namely a skewed and leptokurtic marginal return distribution. An appealing interpretation can be given to the Poisson process, $M_I(t)$, as counting the units of information flow that induce a random change in the asset's price. The model is therefore intimately related to time deformation models (Clark 1973) which have found renewed interest in high frequency data research⁵. Further, it is important to note that, like many of the diffusion processes used in finance, the (compensated) compound Poisson process embodies the martingale property.

While the compound Poisson process, and many of the diffusion processes in particular, have been shown to fit low frequency data relatively well, this is certainly not the case at the high frequency where market microstructure effects have been shown to have a decided, but often complex, impact on the properties of the price process. Roll (1984) demonstrates that the existence of a bid/ask spread can lead to spurious first order negative serial correlation in returns. Lo and MacKinlay (1990) study the impact of non-synchronous trading on the dynamic properties of returns and find that it induces contemporaneous cross-correlation among assets and serial correlation in returns. By and large, it is widely recognized that the various market microstructure effects distort the distributional properties of high frequency returns and typically induce a substantial degree of serial correlation. Any process that is consistent with the martingale hypothesis of (risk adjusted) asset prices, will therefore be inconsistent with much of the theoretical market microstructure literature and, more importantly, with many of the observed characteristics of high frequency data.

In this paper, we argue that the continuous time diffusion processes studied in the finance literature, valuable as they are, seem to lack the flexibility required for the modelling of high frequency security prices. We propose two distinct statistical models that we believe are capable of capturing many important features of high frequency returns. The first model generalizes the standard compound Poisson process, as given in expression (3.1), to account for the presence of a bid/ask spread. The second model allows for a general form of serial dependence in returns. We also study the case where there is both deterministic and stochastic time variation in the trading intensity and show that this can be used to capture (i) deterministic patterns in market activity, (ii) serial dependence in trade durations at high frequency (i.e. "ACD-effects") and (iii) persistence in the conditional return variance at low frequency (i.e. "ARCH-effects"). Based on the characteristic function, we analyze the static and dynamic properties of the price process in detail. Comparison with actual high frequency data suggests that the proposed models are sufficiently flexible to capture a number of salient features of financial return data including a skewed and fat tailed marginal distribution, serial correlation at high frequency, time variation in market activity both at

see for example Chan and Maheu (2002), Embrechts, Klüppelberg, and Mikosch (1997), Madan and Seneta (1984), Maheu and McCurdy (2002a), Mürmann (2001), Rogers and Zane (1998), Rolski, Schmidli, Schmidt, and Teugels (1999), Rydberg and Shephard (2003).

⁵See for example Andersen (1996), Ané and Geman (2000), Carr, Geman, Madan, and Yor (2002, 2003).

high and low frequency. A common feature of both models is that even though the martingale property is lost at high frequency, it can be retained under temporal aggregation. Motivated by this observation, we seek to address two issues that are relevant to the measurement of return volatility. Firstly, within the context of our models, we investigate the impact of serial correlation in returns on the recently proposed realized variance measure as discussed in Andersen, Bollerslev, Diebold, and Labys (2001, 2003) and Barndorff-Nielsen and Shephard (2002, 2003). We show that serial correlation in returns can induce a substantial bias in the variance estimate and characterize its decay under temporal aggregation of returns. Secondly, we discuss a set of sampling strategies which aim at minimizing this bias. Here, the key result is that the magnitude of the bias can be altered by a deformation of the time scale. Importantly, we find that when the trade arrival intensity is non-constant, “business” time sampling maximizes the bias for a given sampling frequency while it achieves the lowest overall MSE relative to calendar time sampling. Moreover, for both sampling schemes, the “optimal” sampling frequency which minimizes the MSE is much higher than the one which minimizes the bias.

In the present context, it is also important to emphasize a fundamental difference between the compound Poisson process and the diffusion process, namely, the former is a *finite* variation process while the latter is an *infinite* variation process. By taking a microscopic view at the data, it is evident that variation in high frequency returns is inherently finite because the number of price-change-inducing trades is finite. Diffusion processes are, by construction, not able to capture this prominent feature of the data. In contrast, the finite variation property of the compound Poisson process appears ideally suited for the modelling of asset price both at high and low frequency.

The remainder of this paper is organized as follows. In Section 2, we generalize the compound Poisson process for the presence of a bid/ask spread, derive the characteristic function of the price process, and analyze the properties of the price process. Section 3 contains analogous results for the compound Poisson process with correlated innovations. Section 4 derives additional results for when the trading intensity process is allowed to vary both deterministically and stochastically through time. Section 5 discusses the impact of serial correlation in returns on the realized variance measure. Section 6 concludes.

3.2 The Bid/Ask Spread

Financial market design distinguishes between two types of trading mechanisms, namely, price-driven markets and order-driven markets. In a price-driven market, all trades take place through a market maker (also referred to as a specialist or dealer) which serves as an intermediary between buyers and sellers. The market maker posts a bid (ask) price at which he is willing to buy (sell), thereby providing immediacy to the traders. Because the market maker is exposed to inventory risk and insider trading⁶ he requires a compensation that is equal to the disparity between the ask and the bid price, i.e. the “spread”. Examples of price-driven markets include the NASDAQ and FOREX. In an order-driven market, on the other hand, traders submit their orders to an electronic order book

⁶References of inventory and asymmetric information models include Admati and Pfleiderer (1988), Demsetz (1968), Easley, Kiefer, and O’Hara (1997), Easley and O’Hara (1992), Glosten and Milgrom (1985), Ho and Stoll (1983), Huang and Stoll (1997), Kyle (1985), O’Hara (1995) Stoll (1978).

which automatically matches orders based on price and time prioritization. In this trading mechanism, traders are exposed to execution risk due to the absence of a market maker. Examples of order-driven markets include the Paris Bourse and the LSE. Hybrid structures, combining both trading mechanisms, are adopted by the NYSE and Deutsche Börse.

The first model we discuss is designed to account for the presence of a bid/ask spread encountered in price-driven markets. For illustrative purposes, Figure 3.1 displays a time-series of 250 transaction prices of the German Bund Futures contract on August 24, 2000. The presence of the bid/ask spread is apparent. It is also clear that

the infinite variation processes, such as the popular diffusion models widely used in finance, are not well suited to characterize this type of price evolution. To investigate the serial correlation of returns, we distinguish between two sampling schemes, namely “business time” sampling and “calendar time” sampling. Sampling in calendar time amounts to recording the (most recent) price at equi-distant *time* intervals, e.g. annual, weekly, hourly etc. On the other hand, sampling in business time, amounts to recording the price whenever a trade (or a certain amount of trades) has occurred. Clearly, when the duration between trades

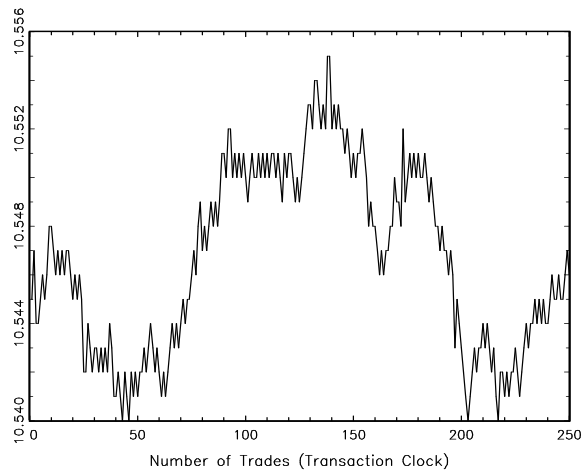


Figure 3.1: Transaction Prices of the German Bund Future

is non-constant, the two sampling schemes will differ. However, the impact of this on the distributional properties returns is non-trivial and will be discussed below in the context of our model. Based on all data for August 24 (over 2000 transaction prices), we find a highly significant first order serial correlation coefficient of -0.447 for returns sampled in business time (trade by trade) and -0.133 for returns sampled in calendar time (minute by minute). These results are in line with Roll (1984). Second order serial correlation is substantially reduced in magnitude and significantly different from zero only for the “trade by trade” returns. Higher order serial correlation is insignificant for both sampling schemes. All in all, it is clear that the price process violates the martingale property, at least when sampled at high frequency. The model we propose below aims to capture the presence of the bid/ask spread and allows us to analyze its impact on the distributional properties of returns.

In what follows, we decompose the observed transaction price into the unobserved mid-price (the average of the bid and ask) plus a spread component. The transaction price is thus equal to the mid-price plus or minus half the bid/ask spread depending on whether a trade is buy-side or sell-side initiated. We assume that the logarithmic *mid-price*, $F(t)$, evolves according to the standard compound Poisson process given in expression (3.1). More general specifications are avoided because the focus is on isolating the impact of the bid/ask spread. The process of the logarithmic *transaction price*, $Q(t)$, inherits the properties of the mid-price process and we assume that its dynamics are governed by:

$$Q(t) = Q(t^-) \underbrace{[1 - dM_{IBS}(t)]}_{\text{NO TRADE}} + F(t) \underbrace{dM_{IBS}(t)}_{\text{ANY TRADE}} + \delta \underbrace{[dM_B(t)]}_{\text{BUY}} - \underbrace{dM_S(t)}_{\text{SELL}}, \quad (3.2)$$

where $M_B(t)$ and $M_S(t)$ denote Poisson⁷ processes with intensity parameters $\lambda_B > 0$ and $\lambda_S > 0$, $dM_{IBS}(t) = dM_I(t) + dM_B(t) + dM_S(t)$, and δ is a positive constant. The intensity parameter of the “combined” Poisson process M_{IBS} is equal to $\bar{\lambda} = \lambda_I + \lambda_B + \lambda_S$.

In the absence of consistent mispricing, the mid-price process reflects the true or fundamental value of the asset. Only the arrival of new information will cause this price to change. In a trading environment, it is reasonable to assume that information is disseminated through order flow and one can thus think of M_I as a process counting the number of “informative” trades which randomly move the asset’s fundamental value (and the transaction price by necessity). Notice that the term ε_j in expression (3.1) represents the innovation to the mid-price process *net* of the bid/ask spread. A second source of randomness in the price process comes through “uninformative” trades. One can think of these as hedge or liquidity motivated trades that are non-speculative of nature and do not contain any (price sensitive) information. Uninformative trades leave the fundamental value of the asset unchanged, but they have the potential to move the transaction price process up or down as they are executed at the mid-price plus or minus a proportional spread δ , depending on whether the trade was buy-side or sell-side initiated. Notice from expression (3.2) that a sequence of uninformative buy orders will only move the transaction price once at the start. Similarly for a sequence of uninformative sell orders. The dynamics of the processes counting the number of uninformative buy- and sell-side initiated trades are governed by M_B and M_S respectively. The combined Poisson process, $M_{IBS}(t)$, therefore counts the total number of trades that occurred up to and including time t . Before moving on, we point out that based on the above interpretation of the model it follows that informative trades are transacted at zero bid-ask spread and that at such instances the mid-price is effectively observable. This contradicts both with what we observe in reality and with the statement that the mid-price is latent. What does remain valid, is that the transaction price associated with an informed trade serves as the mid-price for subsequent uninformative trades. However, it is important to stress that this inconsistency lies in the *interpretation* of the model and not in the model itself. In fact, we will see that the statistical properties of the specified price process are in close correspondence with the ones observed in practice. Because it is these statistical properties that are of primary interest at this stage we stick to the convenient interpretation of the model above although we recognize that alternative, and possibly more appropriate, interpretations can be assigned to the F and Q processes.

For the analysis in the remainder of this paper it proves useful to define a third process, $G(t) = Q(t) - F(t)$, which measures the difference between the transaction price and the mid-price. Because the Q process, as defined in (3.2), can be rewritten as:

$$dQ(t) = -Q(t^-) dM_{IBS}(t) + F(t^-) dM_{IBS}(t) + dF(t) + \delta [dM_B(t) - dM_S(t)].$$

it directly follows that the dynamics for G are given by:

$$dG(t) = -G(t^-) dM_{IBS}(t) + \delta [dM_B(t) - dM_S(t)]. \quad (3.3)$$

⁷The Poisson intensity parameters are defined such that $E[dM_B(t)] = \lambda_B dt$, $E[dM_S(t)] = \lambda_S dt$ and $E[dM_I(t)] = \lambda_I dt$. The sequence $\{\varepsilon_i\}$ is assumed to be independent of $\{M_I(t), t \geq 0\}$. Moreover, it is assumed that $\{M_I(t), t \geq 0\}$, $\{M_B(t), t \geq 0\}$, and $\{M_S(t), t \geq 0\}$ are independent which implies that $\Pr\{dM_B(t) dM_S(t') = 1\} = 0$, $\Pr\{dM_B(t) dM_I(t') = 1\} = 0$, and $\Pr\{dM_S(t) dM_I(t') = 1\} = 0$ for $t > 0, t' > 0$.

Expression (3.3) is known as the *Volterra equation* and the unique solution G is given by Theorem II.6.3 in Andersen, Borgan, Gill, and Keiding (1993):

$$G(t) = G(0) \prod_{[0,t]} [1 - dM_{IBS}(u)] + \delta \int_0^t [dM_B(u) - dM_S(u)] \prod_{(u,t]} [1 - dM_{IBS}(u)]. \quad (3.4)$$

Theorem 3.2.1 *The joint characteristic function of F and G , as defined by expressions (3.1) and (3.4), conditional on initial values is given by:*

$$\begin{aligned} \phi_{F,G}^*(\eta_1, \eta_2, \xi_1, \xi_2, t, m) &\equiv E_0 \left[e^{i\eta_1 F(t) + i\eta_2 F(t+m) + i\xi_1 G(t) + i\xi_2 G(t+m)} \right] \\ &= f(\eta_2, \xi_2) \phi_{F,G}(\eta_1 + \eta_2, \xi_1, t) \left(e^{m\lambda_I(\phi_\varepsilon(\eta_2) - 1)} - e^{-m\bar{\lambda}} \right) \\ &\quad + e^{-m\bar{\lambda}} \phi_{F,G}(\eta_1 + \eta_2, \xi_1 + \xi_2, t) \end{aligned} \quad (3.5)$$

where

$$\begin{aligned} \phi_{F,G}(\eta, \xi, t) &\equiv E_0 \left[e^{i\eta F(t) + i\xi G(t)} \right] \\ &= f(\eta, \xi) \left(\phi_F(\eta, t) - e^{i\eta F(0) - t\bar{\lambda}} \right) + e^{i\eta F(0) + i\xi G(0) - t\bar{\lambda}} \end{aligned} \quad (3.6)$$

for $m > 0$, $\phi_\varepsilon(\eta) = \exp(i\eta\mu_I - \frac{1}{2}\eta^2\sigma_I^2)$, $\phi_F(\eta, t) = \exp(i\eta F(0) + t\lambda_I(\phi_\varepsilon(\eta) - 1))$, and

$$f(\eta, \xi) = \frac{\lambda_I \phi_\varepsilon(\eta) + \lambda_B e^{i\xi\delta} + \lambda_S e^{-i\xi\delta}}{\lambda_I \phi_\varepsilon(\eta) + \lambda_B + \lambda_S}$$

Proof See Appendix B.3.

Based on expression (3.5), moments and cumulants of the mid-price process, F , and the transaction price process, Q , can be derived (see Appendix B.1 for details). In particular, the h^{th} order *conditional* moment of mid-price returns, i.e. $R_F(t|m) \equiv F(t) - F(t-m)$, and transaction price returns, i.e. $R_Q(t|m) \equiv Q(t) - Q(t-m)$, can be derived as:

$$i^{-h} \frac{\partial^h \phi_{F,G}^*(-\gamma, \gamma, 0, 0, t, m)}{\partial \gamma^h} \Big|_{\gamma=0} \quad \text{and} \quad i^{-h} \frac{\partial^h \phi_{F,G}^*(-\gamma, \gamma, -\gamma, \gamma, t, m)}{\partial \gamma^h} \Big|_{\gamma=0}$$

Unconditional moments are obtained by letting t tend to infinity. For completeness, we will briefly discuss the properties of the mid-price process below. More details can be found in Press (1967, 1968).

When $\mu_I \neq 0$, the unconditional *mean* and *variance* of $R_F(t|m)$, are equal to $m\lambda_I\mu_I$ and $m\lambda_I(\mu_I^2 + \sigma_I^2)$ respectively. The *third moment* takes the form:

$$m\lambda_I\mu_I^3 (1 + 3m\lambda_I + m^2\lambda_I^2) + 3m\lambda_I\mu_I\sigma_I^2(1 + m\lambda_I)$$

A non-zero mean of the innovation term therefore induces skewness in returns which increases under temporal aggregation of returns. In contrast, the distribution of returns on the de-trended price process is normal and thus symmetric. The *fourth moment* of returns is equal to:

$$m\lambda_I\mu_I^4 (1 + 7m\lambda_I + 6m^2\lambda_I^2 + m^3\lambda_I^3) + 6m\lambda_I\mu_I^2\sigma_I^2 (1 + 3m\lambda_I + m^2\lambda_I^2) + 3m\lambda_I\sigma_I^4(1 + m\lambda_I)$$

As is the case for skewness, when $\mu_I \neq 0$ return kurtosis increases under temporal aggregation of returns. The expression for the kurtosis simplifies to $3 + 3/(m\lambda_I)$ when $\mu_I = 0$. In this case, temporal aggregation of returns leads to a decrease in kurtosis. Also note that m and λ_I enter multiplicatively in all moment expressions. The impact of a change in either m or λ_I is thus identical.

We now turn to the properties of the transaction price process. Except for the first moment, we will state the moment expressions for the case where $\mu_I = 0$. Although it is straightforward to derive conditional and unconditional return moments when $\mu_I \neq 0$, it needlessly complicates notation and is therefore avoided. The conditional *first moment* of returns is given by:

$$E_0[R_Q(t|m)] = m\lambda_I\mu_I + \frac{e^{-t\bar{\lambda}}(1 - e^{-m\bar{\lambda}}) (\delta(\lambda_B - \lambda_S) - \bar{\lambda}G(0))}{\bar{\lambda}}$$

The above expression points out an interesting feature of the model: even when $\mu_I = 0$ it follows that $E_0[R_Q(m|m)] = E_0[Q(m)] - Q(0) \neq 0$ as long as $\lambda_B \neq \lambda_S$ and / or $G(0) \neq 0$. This directly implies that the logarithmic transaction price process is *not* a martingale. However, the compensated process, i.e. $Q(m) - m\lambda_I\mu$, looks more and more like a martingale when $m \rightarrow \infty$. Because the innovations to the mid-price are iid, this property of the transaction price process is exclusively due to the presence of the bid/ask spread. Taking t (and m) $\rightarrow \infty$ yields the unconditional *mean* of returns which equals $m\lambda_I\mu$ and thus corresponds to the mean of returns on F . For $\mu_I = 0$, the *second moment*, or equivalently the *variance*, of returns is given by:

$$m\lambda_I\sigma_I^2 + 2\delta^2(1 - e^{-m\bar{\lambda}}) \frac{\lambda_I\lambda_S + 4\lambda_S\lambda_B + \lambda_I\lambda_B}{\bar{\lambda}^2}$$

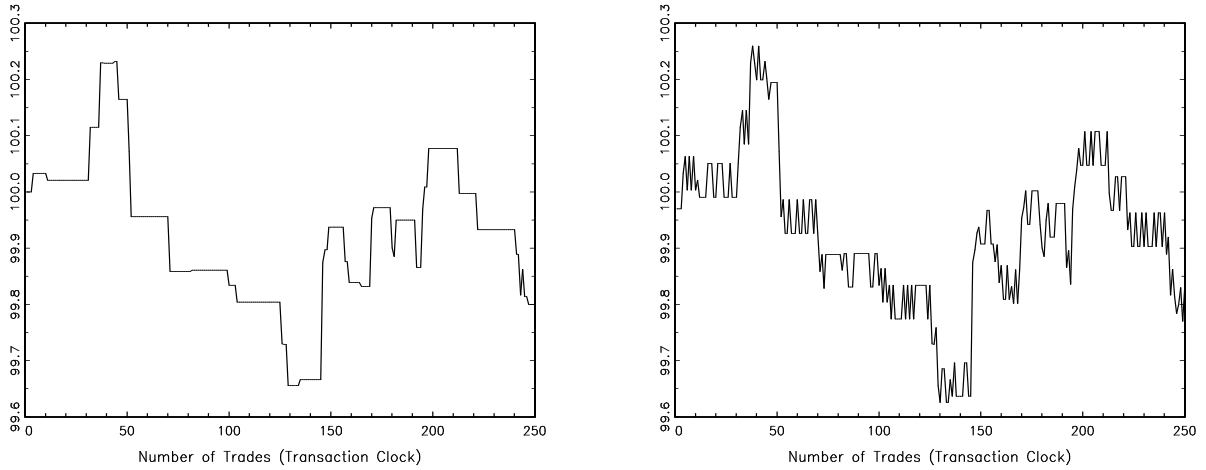
We can decompose the variance into two components, namely the return variance of the mid-price process (left hand side) plus a contribution of the bid/ask spread to the total return variance of the transaction price process (right hand side). Because δ , m , and the intensity parameters are strictly positive, the variance of returns on Q always exceeds the variance of returns on F . However, the relative difference, i.e. $(V[R_Q] - V[R_F])/V[R_F]$, decreases with (i) a decrease in the spread δ , (ii) an increase in the return horizon m , (iii) an increase in the arrival rate of informed trades λ_I , and (iv) a decrease in the arrival rate of uninformed trades λ_B and λ_S . The unconditional *third moment* of returns is given by:

$$\frac{3\lambda_I\delta\sigma_I^2(\lambda_B - \lambda_S)(1 - e^{-m\bar{\lambda}})}{\bar{\lambda}^2}$$

Even though $\mu_I = 0$, the return distribution may be skewed depending on λ_B and λ_S , i.e. when $\lambda_B > \lambda_S$ ($\lambda_B < \lambda_S$), there is positive (negative) skewness while the distribution of returns is symmetric when the arrival rates of uninformed buy-side and sell-side initiated trades are equal. Notice that $\lambda_B \neq \lambda_S$ does not necessarily imply that the market maker builds up or drains his inventory, as the informed trades may off-set the buy/sell imbalance of uninformed traders. The unconditional *fourth moment* of returns is given by the lengthy expression below:

$$3m\lambda_I(1 + m\lambda_I)\sigma_I^4 + 6\lambda_I\sigma_I^2\delta^2(1 - e^{-m\bar{\lambda}}) \frac{\lambda_B^2 + \lambda_S^2 - \lambda_I\lambda_S - \lambda_B\lambda_I - 6\lambda_B\lambda_S}{\bar{\lambda}^3} \\ + 12m\lambda_I\sigma_I^2\delta^2 \frac{\lambda_I\lambda_S + 4\lambda_B\lambda_S + \lambda_B\lambda_I}{\bar{\lambda}^2} + 2\delta^4(1 - e^{-m\bar{\lambda}}) \frac{\lambda_I\lambda_S + 16\lambda_B\lambda_S + \lambda_B\lambda_I}{\bar{\lambda}^2}$$

FIGURE 3.2: MID-PRICE AND TRANSACTION PRICE FOR “BID-ASK SPREAD MODEL”



Notes: Simulated time series of 250 mid-prices (F; left panel) based on expression (3.1) and transaction prices (Q; right panel) based on expression (3.2). The model parameters are set equal to $\sigma_I^2 = 5.16e - 7$, $\lambda_I = 1/\text{minute}$, $\lambda_S = \lambda_B = 2.5/\text{minute}$, and $\delta = 0.0003$.

The relation between the fourth moment or kurtosis and the model parameters is substantially more complicated than for the lower order moments. A few things can be said though. As for the mid-price process, when the return horizon, m , tends to 0 (∞), the kurtosis tends to ∞ (3). When the spread, δ , or the uninformed intensity parameters, λ_B and λ_S , tend to ∞ , the kurtosis tends to a strictly positive constant which can be either smaller, equal or larger than 3 depending on the model parameters. Negative excess kurtosis can thus be induced by the bid/ask spread although this seems to require unrealistic values for either the spread or the intensity parameters. Finally, the return *covariance*, at displacement $k > 0$, can be derived⁸ as:

$$E [R_Q(t|m) R_Q(t - m - k|m)] = -\omega(k, m, \bar{\lambda}) \delta^2 \frac{\lambda_I \lambda_S + 4\lambda_S \lambda_B + \lambda_I \lambda_B}{\bar{\lambda}^2}$$

where $\omega(k, m, \lambda) = e^{-k\lambda} (1 - e^{-m\lambda})^2$. Interestingly, it is noted that the auto-covariance function above corresponds to that of an ARMA(1, 1) process⁹. Because $\omega(k, m, \lambda) > 0$ the bid-ask bounce induces negative serial correlation in returns which disappears under temporal aggregation (increasing m) or increasing arrival frequency of informative trades (increasing λ_I). Roll (1984) finds that the “effective” bid-ask spread, i.e. 2δ , can be measured by 2 times the square root of the negative of the first order serial covariance of returns. The model discussed here, is consistent with Roll’s finding for the degenerate case where $\lambda_I = 0$, $\lambda_B = \lambda_S$, $k = 0$ (first order covariance) and m is large (long horizon returns, e.g. daily / weekly).

To illustrate a possible price path realization of the model, we simulate a time series of 250 mid-prices and associated transaction prices. The model parameters are set equal to $\sigma_I^2 = 5.16e - 7$, $\lambda_I = 1/\text{minute}$, $\lambda_S =$

⁸Using that $E_0 [Q(t+m)Q(t)] = F(0)^2 + t\lambda_I\sigma_I^2 + \frac{2\delta F(0)(\lambda_B - \lambda_S)}{\lambda} + \frac{\delta^2(\lambda_B - \lambda_S)^2}{\lambda^2} + e^{-m\bar{\lambda}}\delta^2 \frac{\lambda_I\lambda_S + 4\lambda_S\lambda_B + \lambda_I\lambda_B}{\bar{\lambda}^2}$.

⁹Recall that the auto-covariance function of an ARMA(1, 1) process with zero mean, i.e. $x_t = \alpha x_{t-1} + \varepsilon_t + \beta\varepsilon_{t-1}$ for $|a| < 1$ and $\varepsilon \sim IIDN(0, \sigma^2)$, is given by $E[x_t x_{t-k}] = \alpha^k \frac{(\alpha + \beta)(1 + \alpha\beta)}{\alpha(1 - \alpha^2)} \sigma^2$ for $j = 1, 2, \dots$. Setting $\alpha = e^{-\lambda}$ ensures the same rate of decay while β and σ^2 can be chosen so as to match the first order covariance term.

$\lambda_B = 2.5/\text{minute}$, and $\delta = 0.0003$ which corresponds to an annualized return volatility¹⁰ of 25% (28.4%) for minute by minute mid-price (transaction price) returns, an arrival rate of 60 informed trades per hour, an arrival rate of 150 uninformed buy-side and sell-side initiated trades, and a spread of 3 basis points. At first sight the resemblance between the actual Bund futures data (Figure 3.1) and the simulated data (Figure 3.2) seems striking. The *ad hoc* parameter values used in the simulation imply a first order serial correlation of minute by minute returns of -0.112 . Increasing the spread to $\delta = 0.0005$ increases the annualized transaction return variance to 33.6% and decreases the first order serial correlation to -0.222 . Returns aggregated over 5-minute intervals, have a theoretical first order serial correlation coefficient of -0.027 for $\delta = 0.0003$ and -0.069 for $\delta = 0.0005$.

The discussion above illustrates the ability of the model to capture a number of salient features of high frequency transaction data. The presence of a bid/ask spread is explicitly accounted for and the magnitude of serial correlation implied by the model is in the right ball park for realistic parameter values. Moreover, it is noted that our model can be viewed as a mixture of the bid/ask bounce model of Roll (1984) and the compound Poisson process model of Press (1967). Specifically, when $\delta = 0$, our model coincides with Press'. When $\lambda_I = 0$ and $\lambda_B = \lambda_S$ our model is closely related to Roll's.

To conclude, we point out a possible weakness of the model. A number of studies have reported a substantial degree of time variation in the bid/ask spread. Demsetz (1968), as one of the first to look into this issue, finds that most of the variation in the spread can be explained by changes in (i) market capitalization, (ii) the inverse of the price, (iii) return volatility, and (iv) market activity. Cross-sectional variation due to changes in market capitalization is clearly not relevant in the current context. Moreover, the proportionality of the spread can arguably capture most of the time variation that is induced by changes in the reciprocal of the price. However, variation of the spread due to changes in market volatility, or market activity, is something that our model clearly cannot account for. Because the arrival intensity parameters are constant, both market activity and return volatility are also constant. In addition δ is not allowed to depend on time or other exogenous variables such as $M_{IBS}(t)$. Unfortunately, it is not easy to resolve this shortcoming of the model because time variation in δ precludes a closed form solution for the characteristic function of $Q(t)$. Although the properties of the model can still be analyzed numerically, the need to choose specific parameter values would narrow the scope of the discussion substantially and is therefore not attempted here. We emphasize, however, that while the properties of the transaction return process will undoubtedly be more complex in such a case, we do not anticipate the qualitative features of the model to change much, i.e. the bid/ask spread is still expected to induce negative serial correlation which disappears under temporal aggregation as is observed in practice.

3.3 General Return Dependence

The bid/ask spread is arguably the most apparent and dominant market microstructure component in the price process of a price-driven market and can, as shown above, be modelled explicitly. However, a host of other market microstructure effects exist which are, as opposed to the bid/ask spread, more concealed or complex in nature. It is therefore not possible to individually address each and every one of these effects. The model we

¹⁰Based on 8 trading hours per day, 252 trading days per year.

propose below, exploits the view that no matter what the nature of the market microstructure effect is, its impact on the return distribution will likely be revealed through the autocorrelation function of returns. We thus study the return dependence structure without explicitly identifying its source. For example, high frequency index returns may be subject to non-synchronous trading, non-trading periods, temporary mispricing, and recording delays. While each and every attribute may be difficult to model, it seems reasonable to anticipate some sort of serial correlation in the first moment of returns, be it negative or positive, of high or low order, transient or persistent. This observation motivates us to generalize the compound Poisson process to allow for a general form of serial correlation in returns. In particular, we assume that the innovations of the logarithmic price, F , follow an MA(q)-process¹¹:

$$F(t) = F(0) + \sum_{j=1}^{M(t)} \varepsilon_j \quad \text{where} \quad \varepsilon_j = \rho_0 \nu_j + \rho_1 \nu_{j-1} + \dots + \rho_q \nu_{j-q}, \quad (3.7)$$

$\nu_j \sim \text{iid } \mathcal{N}(\mu_\nu, \sigma_\nu^2)$, $\rho_q \neq 0$ and $M(t)$ is a homogeneous Poisson process with intensity parameter $\lambda > 0$. No restrictions on ρ_0, \dots, ρ_q need to be imposed in order to ensure stationarity of the innovation process. Regarding the MA structure, it is important to emphasize that it is imposed on the innovation process in *transaction* time. Interestingly, the results below indicate that the autocovariance of returns, sampled at equi-distant *calendar* time intervals, decays exponentially similar to that of an ARMA process. Finally, we note that the price process F is, as opposed to the previous section, assumed to be observable and the single object of interest.

Theorem 3.3.1 *For the price process defined by expression (3.7) and $M(t) \gg q$, the joint characteristic function of $F(t)$ and $F(t+m)$, conditional on initial values, is accurately approximated by:*

$$\phi_F^*(\xi_1, \xi_2, t, m) = E_0 \left[e^{i\xi_1 F(t) + i\xi_2 F(t+m)} \right] = a(\bar{\xi}) \phi_S^*(\xi_1, \xi_2, t, m) \quad (3.8)$$

where

$$\begin{aligned} \phi_S^*(\xi_1, \xi_2, t, m) &= b(\bar{\xi}, t) e^{\bar{\xi}^2 \sigma_\nu^2 \rho(q,q)} \sum_{h=0}^{q-1} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} h \sigma_\nu^2 \xi_2^2 \bar{\rho}^2} \left(e^{-\xi_1 \xi_2 \sigma_\nu^2 \rho(q,h)} - e^{-\xi_1 \xi_2 \sigma_\nu^2 \rho(q,q)} \right) \frac{(m\lambda)^h}{h! e^{m\lambda}} \\ &\quad + b(\bar{\xi}, t) b(\xi_2, m) e^{(\bar{\xi}^2 - \xi_1 \xi_2) \sigma_\nu^2 \rho(q,q)} \end{aligned}$$

for $\bar{\xi} = \xi_1 + \xi_2$, $\bar{\rho} = \sum_{j=0}^q \rho_j$, $a(\xi) = \exp(i\xi F(0))$, $b(\xi, t) = \exp \left[t\lambda \left(e^{i\xi \bar{\rho} \mu_\nu - \frac{1}{2} \xi^2 \sigma_\nu^2 \bar{\rho}^2} - 1 \right) \right]$, and

$$\rho(q, p) = \begin{cases} \sum_{h=1}^{\min(q,p)} \sum_{j=h}^q h \rho_j \rho_{j-h} & \text{for } q \geq 1, p \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

For $t \rightarrow \infty$, the above expression of the characteristic function is exact.

Proof See Appendix B.3.

¹¹In principle it is also possible to impose an AR(q) structure on the price innovations. However, the expression for the characteristic function turns out to be substantially more complicated as it involves an infinite summation of the form $\sum_{n=0}^{\infty} \exp(\rho^n)$ which cannot be simplified.

The characteristic function, given by expression (3.8) above, can be used to derive *exact* unconditional moments of the price and return process as this requires t - and thus $M(t)$ - to tend to ∞ . Expressions for the conditional moments will be arbitrarily accurate when $M(t)$ exceeds the order of the MA process, q , by a sufficiently large amount. When $M(t)$ is small the above characteristic function cannot be used to derive conditional moments. For this case, however, it is possible to derive exact expressions at the cost of cumbersome notation. Because the focus of this paper lies elsewhere, we do not go into this (see footnote 1 in Appendix B.3 for more details on the source of this approximation error).

Below we discuss the properties of the compound Poisson process for $q = 1$ for it is sufficient to illustrate the main features of the model. The case for $q > 1$ adds to the notational complexity without providing much additional insight into the workings of the model. In practice, of course, the increased flexibility that comes with the higher order return dependence may be necessary to model the data and this case therefore remains of great interest. To simplify notation further, we set $\rho_0 = 1$ and $\rho_1 = \rho$. As mentioned above, no restrictions are imposed on the coefficients, although $\rho = -1$ is a degenerate case in the sense that all innovations to the price process cancel out with the exception of the first and last one. Analogous to the previous section, the unconditional return moments can be derived based on the characteristic function¹² given by expression (3.8). When $\mu_\nu \neq 0$ the unconditional *first moment* of returns equals $m\lambda\mu_\nu(1 + \rho)$ while its *variance* is given by:

$$m\lambda(\mu_\nu^2 + \sigma_\nu^2)(1 + \rho)^2 - 2\sigma_\nu^2\rho(1 - e^{-m\lambda}) \quad (3.9)$$

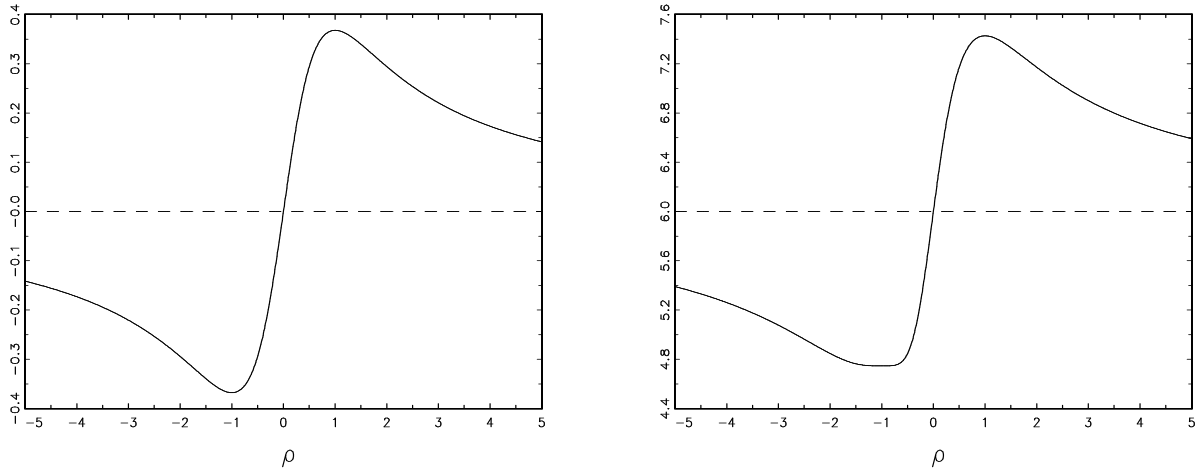
Because the impact of the innovation mean is trivial we set $\mu_\nu = 0$ and focus on the remaining model parameters. As expected, the contribution of the right hand side term in expression (3.9) diminishes relative to the left hand side term when m increases. In other words, the serial correlation of the innovations introduces a transient component into the return variance which disappears under temporal aggregation. To study the impact of ρ on the return variance it is important to take into account that a change in ρ , *ceteris paribus*, will change the return variance because $\sigma_\varepsilon^2 \equiv V[\varepsilon_j] = (1 + \rho)\sigma_\nu^2$. We therefore consider two cases, namely (i) vary ρ while $\sigma_\varepsilon^2 = (1 + \rho^2)\sigma_\nu^2$ and (ii) vary ρ while keeping σ_ε^2 fixed at $\bar{\sigma}^2$. Furthermore, in order to isolate the impact of a change in ρ we choose the MA(0) model with a return variance of $m\lambda\sigma_\varepsilon^2$ as a benchmark.

For the first case, MA(1) innovations inflate the return variance by $2\rho\sigma_\nu^2(e^{-m\lambda} + m\lambda - 1)$ relative to the benchmark case. Serial correlation increases the return variance when it is positive and decreases the return variance when it is negative. Intuitively, when serial correlation is negative (positive), innovations partly offset (reinforce) each other which leads to a decrease (increase) in the return variance. Moreover, notice that the contribution to the return variance consists of a component that only impacts the return variance at high frequency, i.e. $2\rho\sigma_\nu^2(e^{-m\lambda} - 1)$, and a component which impacts the return variance at any given sampling frequency, i.e. $2\rho\sigma_\nu^2 m\lambda$.

For the second case, the impact of a change in ρ is less obvious because it requires a simultaneous change in σ_ν^2 so as to keep σ_ε^2 constant. Here, the return variance exceeds the benchmark by $2\rho\bar{\sigma}^2(e^{-m\lambda} + m\lambda - 1)/(1 + \rho^2)$ which is similar as before but now includes the term $(1 + \rho^2)^{-1}$ and makes the relationship non-linear. To facilitate the discussion, the left panel of Figure 3.3 visualizes this expression as a function of ρ for $m\lambda = 1$ and

¹²Notice that $\xi_1 = -\xi_2$ implies that $a(\bar{\xi}) = b(\bar{\xi}, t) = 1$.

FIGURE 3.3: VARIANCE AND KURTOSIS AS A FUNCTION OF MA(1) COEFFICIENT



Notes: The left panel plots the return variance increase when the MA(1) parameter ρ is moved away from zero. We set $m\lambda = 1$ and $\sigma_\nu^2 = \bar{\sigma}^2/(1 + \rho^2) = 1$ in order to keep σ_ε^2 fixed at $\bar{\sigma}^2$. The left panel plots the return kurtosis as a function of the MA(1) parameter ρ (again $m\lambda = 1$).

$\sigma_\nu^2 = \bar{\sigma}^2/(1 + \rho^2) = 1$. While a negative (positive) return correlation decreases (increases) the return variance relative to the benchmark, the amount by which it does tends to zero when ρ grows in magnitude. Intuitively, an increase in ρ “shifts” variance from the contemporaneous innovation ν_j to the lagged innovation $\rho\nu_{j-1}$. When ρ is sufficiently large in magnitude, the variance of the lagged innovation will swamp that of the contemporaneous one and the process will effectively behave as if it was an MA(0) process.

As opposed to the bid/ask model, the *third moment* of returns is zero unless $\mu_\nu \neq 0$. The expression for this case is straightforward but sizeable and is therefore omitted. The unconditional *fourth moment* of returns for $\mu_\nu = 0$ is given by:

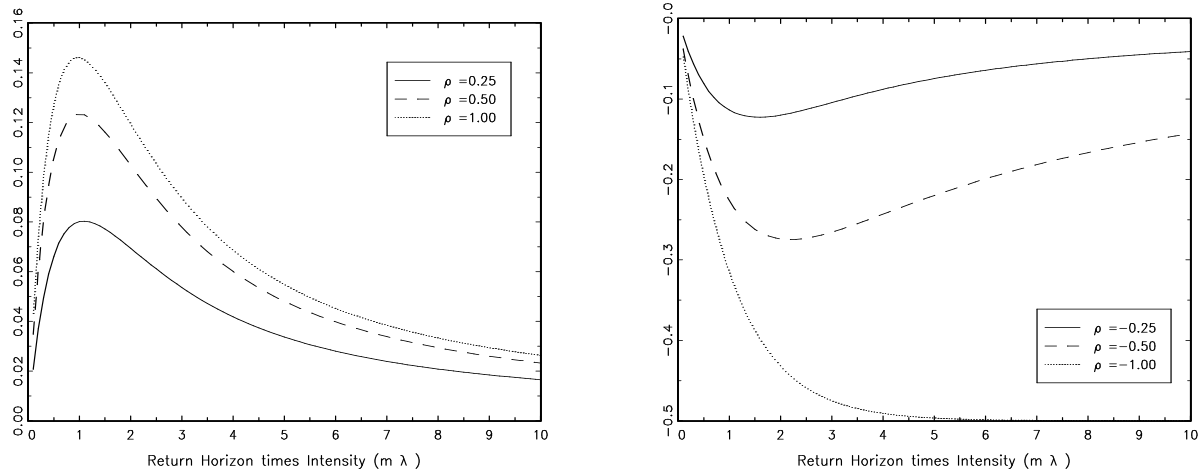
$$3m^2\lambda^2\sigma_\nu^4(1 + \rho)^4 + 3m\lambda\sigma_\nu^4(\rho^2 - 1)^2 - 12\sigma_\nu^4\rho^2(e^{-m\lambda} - 1)$$

It is clear from the expressions for the second and fourth moment, that the kurtosis of returns does not depend on σ_ν^2 . Also we note that the return horizon, m , and the arrival rate of trades, λ , enter multiplicatively into all expressions. The impact of an increase in m is therefore equivalent to the impact of an increase in λ . This simplifies matters substantially and to analyze the kurtosis, we only need to fix $m\lambda$ while varying ρ . The right panel of Figure 3.3 displays the return kurtosis as a function of ρ for $m\lambda = 1$. Here the MA(0) process serves as a benchmark with a kurtosis coefficient of $3 + 3/m\lambda = 6$. Positive (negative) serial correlation in the price innovations thus induces an increase (decrease) in kurtosis relative to the benchmark. The maximum (minimum) return kurtosis is attained by setting $\rho = 1$ ($\rho = -1$) and is equal to 7.43 (4.75) for the current parameter values. Finally, for $\mu_\nu = 0$, the *covariance* of non-overlapping returns can be derived¹³ as:

$$E[R_F(t|m)R_F(t-k-m|m)] = \sigma_\nu^2\rho\omega(k, m, \lambda),$$

¹³Using that $E_0[F(t+m)F(t)] = F(0)^2 + t\lambda\sigma_\nu^2(1 + \rho) - (e^{-m\lambda} + 1)\rho\sigma_\nu^2$.

FIGURE 3.4: SERIAL CORRELATION OF RETURNS AND THE SAMPLING FREQUENCY



Notes: First order (i.e. $k = 0$) serial correlation of returns for the MA(1) compound Poisson process, as a function of $m\lambda$, for positive (left panel) and negative (right panel) values of ρ .

where $m > 0$, $k \geq 0$ and $\omega(k, m, \lambda) = e^{-k\lambda} (1 - e^{-m\lambda})^2$. The discussion of the covariance is analogous to that of the variance. For fixed σ_v^2 , an increase (decrease) in ρ leads to an increase (decrease) of auto-covariance. For fixed σ_ε^2 , on the other hand, the expression is proportional to $\rho/(1 + \rho^2)$ and thus takes on the same form as the graph in the left panel of Figure 3.3. Based on the covariance and variance expression, the *serial correlation of returns* can be derived as:

$$\frac{\rho\omega(k, m, \lambda)}{m\lambda(1 + \rho)^2 - 2\rho(1 - e^{-m\lambda})}$$

As expected, an increase in k , the displacement between returns, leads to an exponential reduction in the magnitude of serial correlation and vice versa. The impact of a change in m , however, is less obvious¹⁴. Figure 3.4 displays the serial correlation of adjacent returns ($k = 0$) for return horizons between 0 and 10 (λ is kept fixed at 1). All curves are hump shaped, with the exception of the degenerate case where $\rho = -1$, implying that serial correlation may either increase or decrease under temporal aggregation depending on the value of m . At first sight this seems quite peculiar. However, when the return horizon (or sampling frequency) tends to zero, the time-series of sampled returns will contain an increasing number of entries that are equal to zero. This, in turn, causes the serial correlation to disappear in the limit. Importantly, this is *not* the case for the covariance.

3.3.1 Multiple Component Compound Poisson

Jumps in *low* frequency financial data are widely documented¹⁵. While transaction data are inherently discontinuous at any sampling frequency, the fact that some jumps can be identified even at low frequency indicates the

¹⁴Although the impact of a change in m is not equivalent to that of a change in λ , due to the term $e^{-k\lambda}$, it is very similar and will therefore not be discussed separately.

¹⁵See for example Andersen, Benzoni, and Lund (2002), Bates (1996, 2000), Duffie, Pan, and Singleton (2000), Eraker (2001), Jiang and Knight (2002), Pan (2002).

presence of jumps of different magnitude. While the jumps observable at high frequency are typically due to the bid/ask spread and price resolution, jumps observable at low frequency can be due to for example a market crash or certain macro-policy announcements. It therefore seems natural to extend the above model to a k -component compound Poisson process with MA(q) innovations:

$$F(t) = F(0) + \sum_{j=1}^{M_1(t)} \varepsilon_{1,j} + \dots + \sum_{j=1}^{M_k(t)} \varepsilon_{k,j}, \quad (3.10)$$

where

$$\varepsilon_{r,j} = \rho_{r,0}\nu_{r,j} + \rho_{r,1}\nu_{r,j-1} + \dots + \rho_{r,q}\nu_{r,j-q},$$

for $\nu_r \sim \text{iid } \mathcal{N}(\mu_{r,\nu}, \sigma_{r,\nu}^2)$ and $\{M_r(t)\}_{r=1}^k$ are independent homogenous Poisson processes with intensity parameters $\lambda_r > 0$ for $r = 1, \dots, k$. Notice that q denotes the *maximum* order of the MA(q) process driving the k components. Because $\nu_{r,j}$ and $M_r(t)$ are assumed to be independent, the present specification¹⁶ of the process does *not* allow for cross correlation among the components driving F . The derivation of the joint characteristic function of $F(t)$ and $F(t+m)$ is therefore analogous to the single component case.

Corollary 3.3.2 (to Theorem 3.3.1) *For the price process defined by expression (3.10) and $M_r(t) \gg q$, the joint characteristic function of $F(t)$ and $F(t+m)$, conditional on initial values, is accurately approximated by:*

$$\phi_F^*(\xi_1, \xi_2, t, m) \equiv E_0 \left[e^{i\xi_1 F(t) + i\xi_2 F(t+m)} \right] = a(\bar{\xi}) \prod_{r=1}^k \phi_{S,r}^*(\xi_1, \xi_2, t, m)$$

where

$$\begin{aligned} \phi_{S,r}^*(\xi_1, \xi_2, t, m) &= b_r(\bar{\xi}, t) e^{\bar{\xi}^2 \sigma_{r,\nu}^2 \rho_r(q,q)} \sum_{h=0}^{q-1} e^{i\xi_2 h \bar{\rho}_r \mu_{r,\nu} - \frac{1}{2} h \sigma_{r,\nu}^2 \xi_2^2 \bar{\rho}_r^2} \left(e^{-\xi_1 \xi_2 \sigma_{r,\nu}^2 \rho_r(q,h)} - e^{-\xi_1 \xi_2 \sigma_{r,\nu}^2 \rho_r(q,q)} \right) \frac{(m \lambda_r)^h}{h! e^{m \lambda_r}} \\ &\quad + b_r(\bar{\xi}, t) b_r(\xi_2, m) e^{(\bar{\xi}^2 - \xi_1 \xi_2) \sigma_{r,\nu}^2 \rho_r(q,q)} \end{aligned}$$

for $\bar{\rho}_r = \sum_{j=0}^q \rho_{r,j}$, $b_r(\xi, t) = \exp \left[t \lambda_r (e^{i\xi \bar{\rho}_r \mu_{r,\nu} - \frac{1}{2} \xi^2 \sigma_{r,\nu}^2 \bar{\rho}_r^2} - 1) \right]$, $\bar{\xi}$ and $a(\xi)$ as defined in Theorem 3.3.1, and

$$\rho_r(q, p) = \begin{cases} \sum_{h=1}^{\min(q,p)} \sum_{j=h}^q h \rho_{r,j} \rho_{r,j-h} & \text{for } q \geq 1, p \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

For $t \rightarrow \infty$, the above expression of the characteristic function is exact.

Proof See Appendix B.3.

For illustrative purposes we will now derive some properties for the 2-component compound Poisson process with MA(1) innovations, i.e. $k = 2$ and $q = 1$:

$$F(t) = F(0) + \underbrace{\sum_{j=1}^{M_1(t)} \varepsilon_{1,j}}_{\text{“DIFFUSION”}} + \underbrace{\sum_{j=1}^{M_2(t)} \varepsilon_{2,j}}_{\text{“JUMP”}}$$

¹⁶Allowing for cross dependence among components is likely to be unimportant for the applications we have in mind here and will therefore not be discussed.

where $\varepsilon_{1,j}$ and $\varepsilon_{2,j}$ follow an MA(1) process. For the analysis of the return moments, we set $\mu_{1,\nu} = \mu_{2,\nu} = 0$ and $\rho_{1,0} = \rho_{2,0} = 1$ for notational convenience. The *mean* is therefore zero while the return *variance* is given as:

$$m\lambda_1\sigma_1^2(1 + \rho_{1,1})^2 + m\lambda_2\sigma_2^2(1 + \rho_{2,1})^2 - 2(1 - e^{-m\lambda_1})\sigma_1^2\rho_{1,1} - 2(1 - e^{-m\lambda_2})\sigma_2^2\rho_{2,1}$$

and the *covariance* of returns can be derived¹⁷ as:

$$E[R(t|m)R(t-k-m|m)] = \sigma_1^2\rho_{1,1}\omega(k, m, \lambda_1) + \sigma_2^2\rho_{2,1}\omega(k, m, \lambda_2)$$

where $\omega(k, m, \lambda) = e^{-k\lambda}(1 - e^{-m\lambda})^2$ as before. Notice that the contribution of both individual components is clearly separated and each take the same form as in the single-component case. The *serial correlation* of returns can now be expressed as:

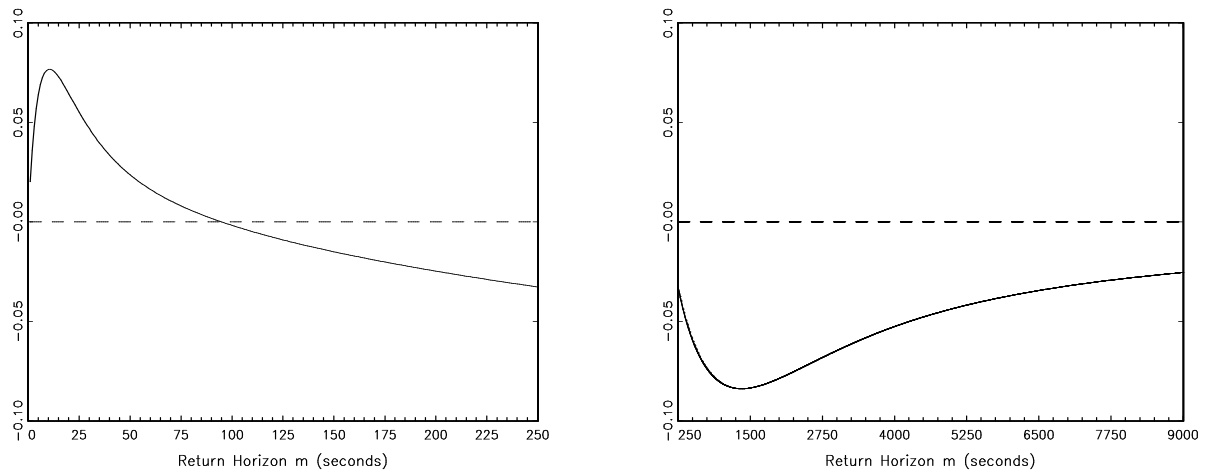
$$\frac{\rho_{1,1}\omega(k, m, \lambda_1) + \rho_{2,1}\omega(k, m, \lambda_2) \frac{\sigma_2^2}{\sigma_1^2}}{m\lambda_1(1 + \rho_{1,1})^2 - 2(1 - e^{-m\lambda_1})\rho_{1,1} + m\lambda_2(1 + \rho_{2,1})^2 \frac{\sigma_2^2}{\sigma_1^2} - 2\rho_{2,1}(1 - e^{-m\lambda_2}) \frac{\sigma_2^2}{\sigma_1^2}}$$

In contrast to the single component case, the innovation variance does not cancel out indicating that its relative magnitude is of interest. Because the return horizon m appears in the denominator, it follows that temporal aggregation of returns will lead to a reduction of serial correlation. A more distinctive feature of the model is that the multiple component structure may induce serial correlation in the price process which can be zero, negative *and* positive depending on the return horizon. This point is illustrated by Figure 3.5. We have set the parameter values to extreme, and empirically unrealistic values, so as to magnify the effect, i.e. $\lambda_1 = 6/\text{min}$, $\lambda_2 = 4/\text{hour}$, $\sigma_1^2 = 8e - 8$, $\sigma_2^2 = 8e - 6$, $\rho_1 = 0.8$, $\rho_2 = -0.8$. It appears that the first component generates positive serial correlation in returns at high frequency (up to approximately a 100 second return horizon). At lower frequencies the second component dominates and thereby induces negative return serial correlation. The location of the “turning” points in the correlogram is closely related to the value of λ_1 relative to λ_2 , although a closed form solution cannot be obtained.

An empirically interesting case is one where the parameters values are chosen such that $\lambda_1 \gg \lambda_2$ while $\sigma_1^2 \ll \sigma_2^2$. In particular, at low frequency, the sample path of the first component will be observationally equivalent to that of a standard diffusion process such as a Brownian Motion. However, for σ_2^2 sufficiently large, the second component will generate infrequent discontinuities or jumps in the path which are observable even at low sampling frequencies. This case is illustrated by Figure 3.6. The left panel displays minute by minute FTSE-100 prices for June 2, 1998. The right panel, contains simulated data based on the 2-component compound Poisson process with MA(1) innovations. The parameter values are chosen as $\lambda_1 = 4/\text{minute}$, $\lambda_2 = 2/\text{day}$, $\sigma_1^2 = 8e - 8$, $\sigma_2^2 = 8e - 5$, $\rho_{1,1} = 0.6$, $\rho_{2,1} = 0.1$ and correspond to an annualized return volatility of 38.5% and first order serial correlation of 4.4%. Although the parameter values are chosen ad hoc, the features of the actual and simulated data seem to agree. Clearly, more elaborate specifications can be considered. For instance, one may introduce a third component with an even lower arrival frequency and even higher variance so as to capture the impact of rare events such as the outbreak of a war or the occurrence of an earthquake. Because the discussion of the model is only illustrative at this point, we will not go further into the determination of the number of components or the estimation of the model parameters.

¹⁷Using that $E_0[F(t)F(t+m)] = F(0)^2 + t\lambda_1\sigma_1^2(1 + \rho_{1,1})^2 + t\lambda_2\sigma_2^2(1 + \rho_{2,1})^2 - (1 + e^{-m\lambda_1})\sigma_1^2\rho_{1,1} - (1 + e^{-m\lambda_2})\sigma_2^2\rho_{2,1}$.

FIGURE 3.5: SERIAL CORRELATION OF RETURNS AND THE SAMPLING FREQUENCY



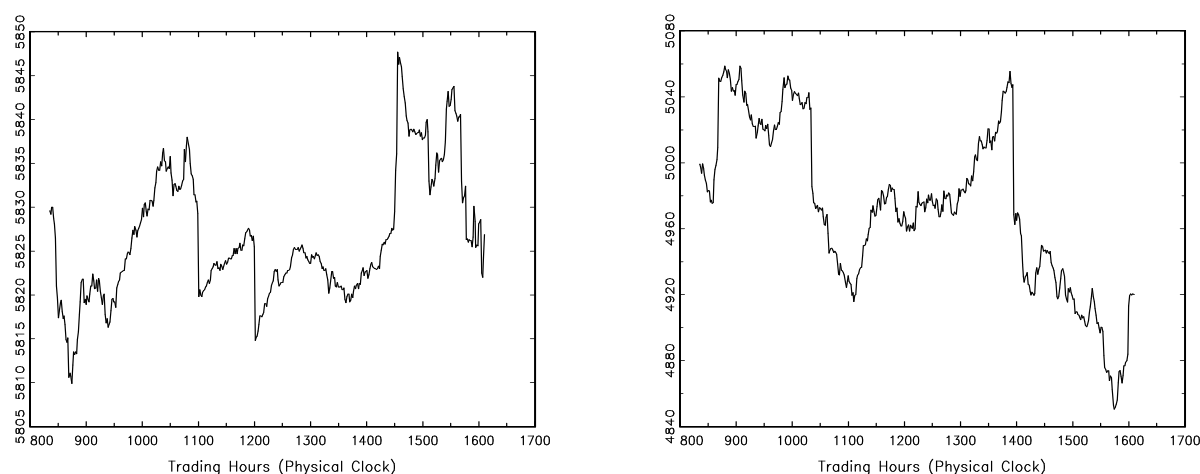
Notes: First order (i.e. $k = 0$) serial correlation of returns for the 2-component compound Poisson process with MA(1) innovations at horizons between 1 and 250 seconds (left panel) and between 251 second and 2.5 hours (right panel). The model parameters are set as $\lambda_1 = 6/\text{min}$, $\lambda_2 = 4/\text{hour}$, $\sigma_1^2 = 8e - 8$, $\sigma_2^2 = 8e - 6$, $\rho_1 = 0.8$, $\rho_2 = -0.8$.

3.3.2 Time Varying Trading Intensity

While the models discussed above are able to capture a variety of dependence structures in *returns*, the *durations* between successive trades are necessarily independent due to the “memory-less” property of the Poisson process (see Bauwens and Giot (2001) for a discussion). A number of empirical studies, however, find compelling evidence that trade durations exhibit a substantial degree of time variation and serial dependence. In this section, we will therefore generalize the model in such a way that it can account for this characteristic feature of high frequency transaction data.

In what follows, we assume that the intensity process, λ , can be decomposed into a *deterministic* component s , and a *stochastic* component $\hat{\lambda}$. Hence, we have $\lambda = \hat{\lambda} + s$ when the deterministic component is additive, and $\lambda = s\hat{\lambda}$ when the deterministic component is multiplicative. Examples of a deterministic component include the widely documented U-shaped pattern in intra-day market activity, day-of-the-week effects, time trends, and any other seasonalities that may be present (see for example Andersen, Bollerslev, and Das (2001), Dacorogna et al. (1993), Harris (1986)). The stochastic component, on the other hand, can account for serial dependencies in the deseasonalized trade intensity and duration. For example, Engle and Russell (1998) find strong evidence of autoregressive serial dependence in deseasonalized intra-day trade durations which motivates them to specify the Autoregressive Conditional Duration (ACD) model. Moreover, the extensive evidence of ARCH effects in low frequency (say daily / weekly) return data indicates that time variation in market activity is not only limited to intra-day frequencies, but extends forcefully to lower frequencies. At this level, the stochastic component typically dominates the deterministic one and, as a result, the time variation induced in low frequency return variance is predominantly stochastic. In this section we will discuss specifications for both components of the intensity process through which we seek to capture the following important stylized characteristics of return data

FIGURE 3.6: ACTUAL AND SIMULATED HIGH FREQUENCY DATA



Notes: Minute by minute FTSE-100 index data (left panel) for June 2, 1998. Simulated minute by minute data (right panel) using the 2-component compound Poisson process with MA(1) innovations. The model parameters are set as $\lambda_1 = 4/\text{minute}$, $\lambda_2 = 2/\text{day}$, $\sigma_1^2 = 8e - 8$, $\sigma_2^2 = 8e - 5$, $\rho_{1,1} = 0.6$, $\rho_{2,1} = 0.1$.

both at low and high frequency:

- (i) seasonality in trade durations and market activity
- (ii) serial dependence in deseasonalized trade duration
- (iii) persistence in return variance at low sampling frequencies

We refer to property (ii) as “ACD”-effects and to property (iii) as “ARCH”-effects, thereby alluding to the seminal work of Engle and Russell (1998), and Engle (1982) and Bollerslev (1986) respectively. Because the aim is to capture all of the above effects through the specification of the intensity process exclusively, a brief discussion of the relation between trading intensity, return variance, and trade duration is in order. Recall that for the standard compound Poisson process with unit innovation variance and (trade) intensity λ , the expected return variance over a unit time interval equals λ while the expected trade duration is equal to $1/\lambda$. Trading intensity is thus proportional to return variance and inversely proportional to trade durations. However, these relations may break down when we generalize the compound Poisson process. For example, when a bid/ask spread “contaminates” the data, we have shown that the return variance is equal to λ plus a non-linear correction term involving the spread. What’s more, when the trading intensity is a (non-degenerate) deterministic function of time, the return variance equals $\int \lambda(u) du$ even though the expected trade duration is *not* equal to $1/\int \lambda(u) du$. These cases are examples where the proportionality between trading intensity, return variance, and inverse of trade duration, is lost. However, it seems reasonable to expect that in many cases the proportionality will hold approximately. Clearly, the extent to which this is true depends on the model specification and also on the sampling frequency of the data (as we have shown that market microstructure effects vanish under temporal aggregation).

Corollary 3.3.3 (to Theorem 3.3.1) For the price process defined by expression (3.7), with a non-constant intensity process, $\lambda(\cdot)$, and $M(t) \gg q$, the joint characteristic function of $F(t)$ and $F(t+m)$, conditional on initial values, is accurately approximated by:

$$\phi_F^*(\xi_1, \xi_2, t, m) = E_0 \left[e^{i\xi_1 F(t) + i\xi_2 F(t+m)} \right] = a(\bar{\xi}) \phi_S^*(\xi_1, \xi_2, t, m) \quad (3.11)$$

where $\phi_S^*(\xi_1, \xi_2, t, m)$ equals:

$$e^{\bar{\xi}^2 \sigma_v^2 \rho(q,q)} \sum_{h=0}^{q-1} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} h \sigma_v^2 \xi_2^2 \bar{\rho}^2} \left(e^{-\xi_1 \xi_2 \sigma_v^2 \rho(q,h)} - e^{-\xi_1 \xi_2 \sigma_v^2 \rho(q,q)} \right) E_0 \left\{ b(\bar{\xi}, 0, t) \frac{(\lambda^*(t, m))^h}{h! e^{\lambda^*(t, m)}} \right\} \\ + e^{(\bar{\xi}^2 - \xi_1 \xi_2) \sigma_v^2 \rho(q,q)} E_0 \{ b(\bar{\xi}, 0, t) b(\xi_2, t, m) \}$$

$\lambda^*(t, \tau) \equiv \int_t^{t+\tau} \lambda(u) du$, $b(\xi, t, \tau) = \exp \left[\left(e^{i\xi \bar{\rho} \mu_\nu - \frac{1}{2} \xi^2 \sigma_v^2 \bar{\rho}^2} - 1 \right) \lambda^*(t, \tau) \right]$, and $\bar{\xi}$, $\bar{\rho}$, $a(\xi)$, $\rho(q, p)$ are as defined in Theorem 3.3.1.

For $t \rightarrow \infty$, the above expression of the characteristic function is exact.

Proof See Appendix B.3.

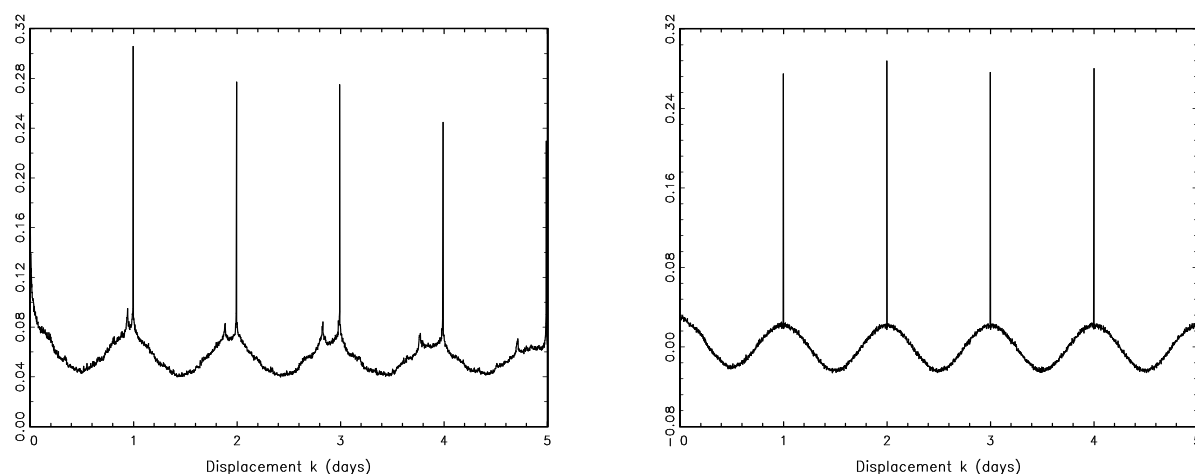
Allowing for time variation in the intensity process, leads to a modified characteristic function of the price process as can be seen by comparing expression (3.11) in Corollary 3.3.3 to expression (3.8) in Theorem 3.3.1. If time variation in the intensity process is entirely deterministic, or known at $t = 0$, the expectation operator vanishes in the expression for $\phi_S^*(\xi_1, \xi_2, t, m)$ and moments can be derived in the usual fashion. This holds true irrespective of the, potentially complex, functional form for $\lambda(\cdot)$. However, when time variation in the intensity process is (partly) stochastic, i.e. unknown at $t = 0$, the expectations operator remains because the integrated intensity process is now a random variable. Moments cannot be derived without explicit specification of the dynamics of the intensity process, and even then, closed form solutions will not be available in many cases.

Deterministic Intensity Process. We will now briefly illustrate the usefulness of allowing for deterministic variation in the intensity process. As mentioned above, one of the most prominent features of high frequency data in financial markets is the U-shaped pattern in intra-day market activity and return volatility. In particular, it is widely documented that market activity is substantially higher around the open and close of the market than around lunch time. Another important characteristic is that the overnight return typically accounts for a non-negligible fraction of the overall daily return variance. While trading in many securities is halted overnight, information flow is not. This in turn, leads to an accumulation of information which can only be incorporated into the price at the next open of the market. The overnight return may therefore reflect a disproportionately large amount of information relative to the subsequent intra-day returns. A highly stylized specification of the intensity process, that is consistent with the above observations, is the following:

$$\lambda(t) = a + b \cos(2\pi t) + c I_{\{t - [t] < \Delta\}} \quad (3.12)$$

where $a > b$, $c > 0$, $0 < \Delta \ll 1$, $[t]$ denotes the integer part of t , and I is an indicator function which equals 1 whenever $t - [t]$ is less than Δ and zero otherwise. Using the single component compound Poisson process with

FIGURE 3.7: ACTUAL AND SIMULATED HIGH FREQUENCY DATA



Notes: Correlogram of minute by minute absolute returns for the FTSE-100 index (left, period 1990-2000) and for simulated minute by minute data (right) using a single component compound Poisson process with MA(1) innovations and a deterministic intensity process given by expression (3.12). The parameter values are set as $a = 4/\text{minute}$, $b = 2.25$, $c = 120/\text{minute}$, $\Delta = 2/480$, $\rho = 0.3$, and $\sigma_v^2 = 7e - 8$

MA(1) innovations and an intensity process as specified above, we simulate 5 years of high frequency transaction prices using the following ad hoc parameter values; $a = 4/\text{minute}$, $b = 2.25$, $c = 120/\text{minute}$, $\Delta = 2/480$, $\rho = 0.3$, and $\sigma_v^2 = 7e - 8$. Based on 8 hours of trading per day, these parameters imply an average of 2160 trades per day, an annualized daily return volatility of 25.4%, and a more than 25 fold increase in market activity (relative to the daily average) during the first two minutes following the market open. The overnight return aside, trading intensity at open and close (mid-day) is 50% higher (lower) than the daily average.

The left-hand panel of Figure 3.7 plots the correlogram of minute by minute absolute returns on the FTSE-100 over the period 1990-2000. The displacement is up to 2400 lags, or equivalently, five trading days. The U-shaped pattern in market activity and the impact of the overnight return is apparent. Moreover, the magnitude of both effects underline the importance of allowing for a deterministic pattern in the intensity process. The right-hand panel of Figure 3.7 plots the correlogram for the simulated data sampled at minute intervals. The strong agreement among the correlograms of the actual and simulated data demonstrates that the naive and overly simplistic specification of the intensity process does capture important patterns in high frequency return data at least to some extent. However, a more detailed inspection of the graphs points to some important differences. For example, the correlogram for the FTSE-100 data indicates a peak in market activity during the afternoon trading session that is, most likely, associated with the open of the US markets. A more subtle difference in the correlogram for the actual data is that the correlations are strictly positive at any displacement and that there appears to be a slow decline in their magnitude. One possible explanation for this is that *stochastic* variation in market activity *across* days induces (positive) serial dependence in the return variance which comes to dominate the intra-daily seasonal pattern at longer horizons. Such dynamics are clearly absent in the above specification of

the intensity process and will be discussed next.

Stochastic Intensity Process. As can be seen from Corollary 3.3.3, when the intensity process is (partly) stochastic the expectation operator in the characteristic function remains. Hence, an expectation of the form $E_0 \left[\exp (a\lambda^* (0, t) + b\lambda^* (t, m)) \lambda^* (t, m)^h \right]$ for $h = 0, \dots, q - 1$ needs to be computed. If the joint Laplace transform for $\lambda^* (0, t)$ and $\lambda^* (t, m)$ is available, i.e. $\Phi (a, b) = E_0 [\exp (a\lambda^* (0, t) + b\lambda^* (t, m))]$, this expectation can be obtained as:

$$\frac{\partial^h \Phi (a, b)}{\partial b^h}$$

However, for many specifications the joint Laplace transform will not be available in closed form and moments need to be obtained by simulation. Below we will discuss a dynamic specification of the intensity process which is capable of generating both ACD and ARCH effects in the price process and for which the Laplace transform does exist in closed form (see Appendix B.2 for details). In spite of the models flexibility and analytic tractability, a major drawback of the specification is that there is nothing that prevents the intensity process from becoming negative. In practice this feature of the model is clearly undesirable. Here, however, this deficiency does not pose a problem to us as the discussion is purely illustrative and the intuition derived from this case is likely to remain in tact for alternative specifications.

ACD and ARCH effects are known to unveil themselves at different frequencies and we therefore decompose the stochastic intensity process into a high frequency and a low frequency component. In particular, ARCH effects are modelled through the low frequency component while ACD effects are modelled through the high frequency component. Market microstructure considerations are clearly of less importance for the low frequency component as they are for the high frequency component. It therefore seems reasonable to rely on proportionality between (integrated) intensity and (integrated) return variance when modelling the ARCH effects. For this case, the dependence structure of the intensity process will (closely) corresponds to that of the variance process and an appropriate specification for the low frequency component, α , is as follows:

$$d\alpha (t) = -\varphi (\alpha (t) - \mu) dt + \sigma_\alpha dW_\alpha (t), \quad (3.13)$$

where $\varphi \geq 0$, $\sigma_\alpha > 0$, and $W_\alpha (t)$ is a standard Brownian motion. The above process is known as the Ornstein-Uhlenbeck (OU) process and has the interesting property that it can be viewed as the continuous-time analogue of the Gaussian first order autoregression. One way to see this is to discretize the time scale as $t_i = i\Delta$ where $i = 1, \dots, T/\Delta$ so that Δ can be interpreted as the frequency at which the continuous time process is sampled while $T\Delta$ represents the total number of periods. The solution to the SDE in expression (3.13) can now be written as

$$\alpha (t_i) = \mu (1 - e^{-\varphi\Delta}) + e^{-\varphi\Delta} \alpha (t_{i-1}) + \varepsilon_{t_i}$$

where $\varepsilon_{t_i} \sim \text{i.i.d. } \mathcal{N} \left(0, \frac{1 - e^{-2\varphi\Delta}}{2\varphi} \sigma_\alpha^2 \right)$. The discretized sample path of α thus follows an autoregressive process of order one with autoregressive parameter equal to $e^{-\varphi\Delta}$. Its persistence therefore depends both on the parameter φ and the sampling frequency Δ . In particular, for fixed parameters φ and σ_α , the persistence of the process increases with an increase of the sampling frequency Δ , i.e. smaller Δ (see Boswijk (2002, Chapter 6) for more

details). Because ARCH effects are a low frequency phenomenon, we set φ and σ_α sufficiently small causing α to appear roughly constant at high frequency. However, at lower frequencies, the mean reversion will become more apparent, leading to an autoregressive dependence structure in return variance - ARCH effects.

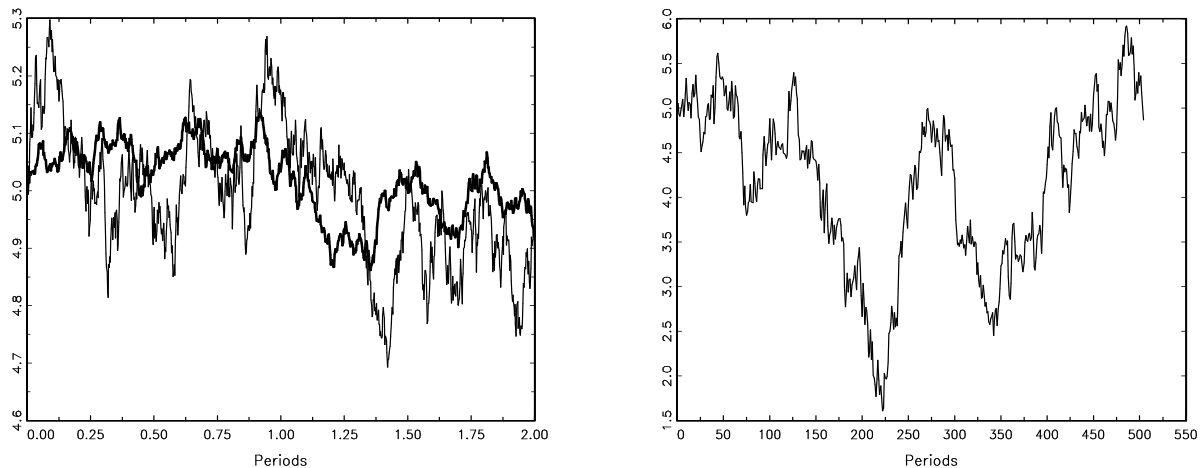
The modelling of ACD-effects is unfortunately more complicated. At high frequency, market microstructure effects and time variation in the intensity process can distort the proportionality between trade intensity and trade duration. In addition, we need to address the question what dependence structure should be imposed on the *intensity* process in order to generate ACD effects, i.e. autoregressive dependence in the *duration* process. Even in idealized situations, there is no clear answer to this question and we will proceed under the debatable assumption that ACD effects can be captured by means of an autoregressive component in the (deseasonalized) intensity process. With this in mind, we specify the high frequency component as follows:

$$d\hat{\lambda}(t) = -\kappa \left(\hat{\lambda}(t) - \alpha(t) \right) dt + \sigma_\lambda dW_\lambda(t) \quad (3.14)$$

where $\kappa \geq 0$, $\kappa \neq \varphi$, $\sigma_\lambda > 0$, and W_λ is a standard Brownian motion independent of W_α . The process given by expression (3.14) is a generalization of the standard Gaussian OU process. It has the property that $\hat{\lambda}$ mean-reverts towards the low frequency component, α , which itself varies stochastically through time. In the current context, the difference between $\hat{\lambda}$ and α constitutes the high frequency component of the intensity process. Quick mean reversion of $\hat{\lambda}$ towards the stochastic long run mean, α , can be expected to generate mean reversion in the duration process at high frequency, thereby leading to ACD effects. Hence, both ARCH and ACD effects can be generated when $\varphi \ll \kappa$ and $\sigma_\alpha \ll \sigma_\lambda$ and $\sigma_\alpha^2/\varphi \gg \sigma_\lambda^2/\kappa$. At high sampling frequencies, the process for $\hat{\lambda}$ will quickly “oscillate” around the stochastic long run mean α , which itself is roughly constant due to its extreme persistence and small innovation variance relative to $\hat{\lambda}$. The stochastic time variation of the intensity process over *short* time intervals will therefore be mainly driven by the OU process for $\hat{\lambda}$ whose mean reversion will lead to ACD effects. On the other hand, at low(er) sampling frequencies, the stochastic variation in the average (or integrated) intensity process arising from the OU process for $\hat{\lambda}$ will be minimal due to its quick mean reversion, and at some stage the stochasticity of the long run mean component will come to dominate. Slow mean reversion in α translates directly into slow mean reversion of trade intensity which, in turn, leads to ARCH effects. Another way to see this is by considering the intensity variance at low frequency which can be shown to equal $\frac{\sigma_\lambda^2}{2\kappa} + \frac{\kappa\sigma_\alpha^2}{2\varphi(\varphi+\kappa)}$ which is approximately equal to $\frac{\sigma_\lambda^2}{2\kappa} + \frac{\sigma_\alpha^2}{2\varphi}$ for $\kappa \gg \varphi$. Because, by assumption, the parameters are chosen such that $\sigma_\alpha^2/\varphi \gg \sigma_\lambda^2/\kappa$, it is clear that the stochastic long run mean dominates at low frequency. One can thus think of the OU process for $\hat{\lambda}$ as driving time variation in the intensity process at high sampling frequencies, while α has a “level-shifting” effect in the sense that it slowly moves the *level* at which $\hat{\lambda}$ operates.

In order to further illustrate this property of the model, we fix some *ad hoc* parameter values that satisfy the above criteria, i.e. $\kappa = 5$, $\sigma_\lambda = 0.25$, $\varphi = 0.0001$, $\sigma_\alpha = \sqrt{0.001}$, and $\mu = 5$ and simulate 2×252 periods of the intensity process with 480 discretization steps per period. The left panel of Figure 3.8 graphs a time series of intensity process $\hat{\lambda}$ over the first two periods of the simulated sample. The superimposed dashed line represents the corresponding long run mean component. It is clear that most of the variation in the intensity process at high frequency comes from the OU dynamics of $\hat{\lambda}$. The right panel of Figure 3.8 plots the period by period average (or integrated) intensity process which corresponds very closely to the low frequency component (not

FIGURE 3.8: TRADE INTENSITY PROCESS AT HIGH AND LOW FREQUENCY



Notes: Simulated intensity process (without deterministic component) based on the “double OU” process as defined by expressions (3.13) and (3.14). The left panel plots the intensity process at high frequency (thin line) for 2 periods together with its associated long run mean component (thick line). The right panel plots the average intensity process at lower frequency for the full simulated sample of 504 periods. The model parameters are set as $\kappa = 5$, $\sigma_\lambda = 0.25$, $\varphi = 0.0001$, $\sigma_\alpha = \sqrt{0.001}$, and $\mu = 5$.

displayed). At this frequency, α drives the overall variation in the intensity process, while the OU component for $\hat{\lambda}$ contributes little. Next, we estimate a simple Exponential-ACD(1,1) and GARCH(1,1) model on simulated trade durations and daily returns respectively and find that (results not reported) for appropriate parameter values, the autoregressive dependence structure in these (squared) variables can indeed be uncovered. However, as expected, the lagged duration (ACD) and lagged squared daily return (GARCH) terms enter insignificantly suggesting that our model is in fact more closely related to the Stochastic Volatility model (Harvey, Ruiz, and Shephard 1994) and the Stochastic Conditional Duration model (Bauwens and Veredas 2003).

In summary, stochastic variation in the high and low frequency component of the intensity process can lead to ACD and ARCH effects respectively. For the specification discussed above, closed form solutions for the intensity process are available (see Appendix B.2 for details). Because the integrated intensity process turns out to be conditionally normal, a closed form expression for the characteristic function in Corollary 3.3.3 is available as well. As mentioned above, a major flaw of the model is that there is nothing that prevents the intensity process from becoming negative. In the context of volatility modelling, Gupta and Subrahmanyam (2002), Stein and Stein (1991) have used a similar specification and justified this on the basis that for a wide range of relevant parameter values, the probability of actually reaching a negative value is so small as to be of no significant consequence. Also, at this point the discussion of the model is purely illustrative and the intuition derived from this case is likely to remain in tact for alternative specifications. Nevertheless, in practice it may clearly make sense to sacrifice analytic tractability in return for a more appropriate specification which ensures positivity of the intensity process. One approach is to specify the model in terms of logarithmic intensity or incorporate a state-dependent innovation variance as is done in the Feller or CIR process. Other models of potential interest are some of the non-Gaussian

OU processes discussed by Barndorff-Nielsen and Shephard (2001).

3.4 Realized Variance and Return Dependence

In the context of the models analyzed above, we now study the impact of - market microstructure induced - serial correlation in returns on the properties of the realized variance (RV) measure. Importantly, we show that serial correlation renders the RV a *biased* estimator of the conditional return variance. We derive closed form expressions for the bias term as a function of the sampling frequency and the model parameters and show that the magnitude of the bias decays under temporal aggregation of returns at a rate that is inversely proportional to the sampling frequency. We also discuss the optimality of alternative sampling schemes.

In an influential series of papers Andersen, Bollerslev, Diebold, and Labys (2001, 2003, ABDL hereafter) have shown that when the logarithmic price process follows a semi-martingale (i.e. a process which can be decomposed into a finite variation component and a martingale component), its associated quadratic variation (QV) process is a critical determinant of the conditional return variance. Importantly, the QV process can - by definition - be approximated as the sum of squared returns sampled at high frequency. It is this approximation of the QV process that is commonly referred to as realized variance or volatility. In full generality, the relation between the conditional return variance and the RV measure is not clear-cut. However, under certain (possibly restrictive) assumptions on the finite variation component of the semimartingale, ABDL show that *realized variance is an efficient and unbiased estimator of the conditional return variance*. ABDL also argue that a violation of the assumptions ensuring unbiasedness is likely to have a trivial impact on the properties of the RV measure, thereby establishing it as an unbiased, efficient, *and* robust estimator of the conditional return variance. In the notation established above, ABDL exploit the following equality:

$$E \left[\sum_{j=1}^{N/m} R(t + jm|m)^2 | \mathcal{F}_t \right] = E \left[R(t + N|N)^2 | \mathcal{F}_t \right]. \quad (3.15)$$

where R denotes excess returns, m denotes the sampling frequency, whereas N denotes the length of the period over which RV is calculated. It is clear from expression (3.15) that the unbiasedness of the RV measure crucially relies on the martingale property of logarithmic (risk adjusted) prices, or equivalently, the absence of serial correlation in excess returns. Nevertheless, a number of recent studies have implemented the RV measure without much concern for possible violations of the martingale assumption underlying the unbiasedness of this measure. It therefore seems appropriate to study the dependence structure of high frequency returns and its associated impact on the properties of the RV measure¹⁸. Although this is largely an empirical matter, and results can be expected to vary across securities and time, the models discussed in this paper seem to capture a number of salient features of high frequency returns particularly well and are therefore well suited to assess the properties of RV in a realistic, yet theoretical, setting.

¹⁸See Andreou and Ghysels (2001), Bai, Russell, and Tiao (2001), and Oomen (2002) for related work.

3.4.1 The ‘‘Covariance Bias Term’’

We investigate the properties of the RV measure for the single component compound Poisson process with MA(1) innovations. Because the results for the ‘‘bid/ask model’’ take the same form, we do not discuss this model separately. In the discussion below we distinguish between the case where trade intensity is constant and the case where it is time varying. To simplify notation we also set $\mu_\nu^2 = 0$.

Constant Trade Intensity. Due to the stationarity of the return process, the conditional and unconditional return variance coincide and can be expressed as

$$E [R(t + N|N)^2] = N\lambda\sigma_\nu^2(1 + \rho)^2 - 2\sigma_\nu^2\rho(1 - e^{-N\lambda}) \stackrel{N \text{ large}}{\approx} N\lambda\sigma_\nu^2(1 + \rho)^2 \quad (3.16)$$

On the other hand, the expectation of the RV measure is equal to:

$$E \left[\sum_{j=1}^{N/m} R(t + jm|m)^2 \right] = \underbrace{N\lambda\sigma_\nu^2(1 + \rho)^2}_{\text{RETURN VARIANCE}} - 2 \underbrace{(1 - e^{-m\lambda}) \frac{N\rho\sigma_\nu^2}{m}}_{\text{COVARIANCE BIAS TERM}} \quad (3.17)$$

In practice, N is typically large (e.g. a day or week) and the approximation error in expression (3.16) can therefore safely be ignored. In contrast, m is typically small (e.g. minute or hour) and the second term on the right hand side in expression (3.17) may therefore be substantial. This illustrates a crucial point: when high frequency (intra-period) returns are used to construct the RV measure, i.e. $m < N$, serial correlation of returns induces a bias that is characterized by the second term on the right hand side of expression (3.17). This bias can be either positive or negative depending on the sign of ρ . Moreover, the magnitude of the bias term decays at rate m^{-1} under temporal aggregation while it tends to $-2N\lambda\sigma_\nu^2\rho$ for $m \rightarrow 0$. It is emphasized that this result does not rely on the approximation in expression (3.16) and will hold true as long as intra-period return are used to construct the RV measure, i.e. $N > m$. Clearly, the magnitude of the bias will depend on specific parameter values and the sampling frequency.

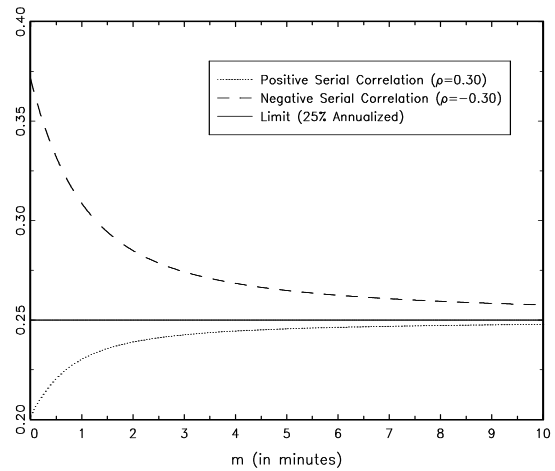


Figure 3.9: Covariance Bias Term for MA(1) with positive and negative correlation coefficients

This is illustrated in Figure 3.9. For¹⁹ $\rho = 0.3$ and $\rho = -0.3$, we plot the return variance (standardized by N) plus the bias component for return horizons up to 10 minutes. The parameter σ_ν^2 is adjusted so as to maintain an annualized return variance of 25%, i.e. for $\rho = 0.30$ ($\rho = -0.30$) we have $\sigma_\nu^2 = 1.529e - 7$ ($\sigma_\nu^2 = 5.272e - 7$). It turns out that for these parameter values the bias term is substantial, i.e. around 20% (12%) of the return variance

¹⁹Remember that the MA structure is imposed on returns in *transaction* time. For the bond futures data analyzed in Section 3.2 we found a first order serial correlation of about -0.45 . The chosen parameter values in the simulation are therefore reasonable from an empirical point of view.

when returns are negatively (positively) correlated and sampled at the 1 minute frequency. The magnitude of this bias can go up to 50% (20%) when sampled at even higher frequencies! These analytical results are in line with a recent study by Oomen (2002) which finds that for the FTSE-100 index over the period 1990-2000 (i) high frequency returns feature substantial serial dependence (for minute by minute data, the serial correlation is positive and significant up to high orders), (ii) the covariance bias term is around 40% for minute by minute returns and (iii) the magnitude of this bias term decays hyperbolically under temporal aggregation.

Time-Varying Trade Intensity For simplicity we focus on the case where the time variation in the trading intensity is a deterministic function of time only. Although more general results can be derived within the OU framework outlined above, the notation is complex and the stochastic case does not add much additional insight for the discussion below. In the deterministic setting, it directly follows from Corollary 3.3.3 that

$$E \left[R_F(t + N|N)^2 \right] = \sigma_v^2 (1 + \rho)^2 \lambda^*(t, N) - 2\rho\sigma_v^2 \left(1 - e^{-\lambda^*(t, N)} \right) \stackrel{N \text{ large}}{\approx} \sigma_v^2 (1 + \rho)^2 \lambda^*(t, N)$$

On the other hand, the conditional expectation of RV is:

$$E_t \left[\sum_{j=1}^{N/m} R(t + jm|m)^2 \right] = \underbrace{\sigma_v^2 (1 + \rho)^2 \lambda^*(t, N)}_{\text{RETURN VARIANCE}} - \underbrace{2\rho\sigma_v^2 \sum_{j=0}^{N/m-1} \left(1 - e^{-\lambda^*(t+jm, m)} \right)}_{\text{COVARIANCE BIAS TERM}} \quad (3.18)$$

Again, the bias term can be substantial depending on the sampling frequency and model parameters and similar results can be derived for this case as for the constant intensity case. A more interesting feature of the bias characterization for non-constant trade intensity, is that it allows us to analyze the performance of alternative sampling schemes to which we turn next.

3.4.2 Bias Reduction and Optimality of Sampling Schemes

As pointed out above, the presence of serial correlation in returns introduces a bias in the RV measure which can be substantial for realistic model parameter values. Because the efficiency of the RV measure crucially relies on the use of intra-period returns, one faces a trade off between the sampling returns at a high frequency, thereby minimizing the measurement error, and sampling returns at low frequency, thereby minimizing the bias term. This trade-off suggests the existence of an “optimal” sampling frequency, that is the highest available frequency at which the bias term is negligible. Alternatively, one could estimate the model parameters and correct for the bias term based on the expression derived above. In practice it is not clear which of these two approaches is preferable. While the bias correction method allows one to use all available data, it is clearly model dependent. The gain in efficiency may therefore be offset by the impact of model and parameter uncertainty. On the other hand, while specifying an “optimal” sampling frequency is essentially non-parametric or model *independent*, valuable information may be lost by the aggregation of returns.

A related issue that arises in this context is how to sample the data. Up to now we have only considered returns that are sampled at equidistant *time* intervals, i.e. $t + jm$ for $j = 1, \dots, N/m$. However, when transaction data is

available it is also possible to consider alternative sampling schemes. A particularly interesting one is where the price process is sampled at time points τ_j for $j = 1, \dots, N/m$ where $\tau_0 = t$, $\tau_{N/m} = t + N$ and

$$\int_{\tau_j}^{\tau_{j+1}} \lambda(u) du = \frac{m}{N} \int_{\tau_0}^{\tau_{N/m}} \lambda(u) du \equiv \lambda_m \quad (3.19)$$

The above sampling scheme effectively “deforms” the calendar time scale by compressing it when the arrival rate of trades is low and stretching it when the arrival rate of trades is high. In this case, one can think of returns being equally spaced on a “transaction” or “business” time scale as opposed to a calendar time scale. An attractive feature of this sampling scheme is that the statistical properties of returns sampled on this deformed time scale coincide with those of a *homogenous* compound Poisson process with intensity parameter equal to λ_m . Because the construction in expression (3.19) ensures that both sampling schemes generate the same number of intra-period returns (N/m), it is of interest to compare the bias term associated with each scheme. As can be seen from expression (3.18), for the calendar time sampling, the bias term is equal to:

$$2\rho\sigma_v^2 \sum_{j=0}^{N/m-1} \left(1 - e^{-\lambda^*(t+jm,m)}\right)$$

On the other hand, for the “business time” sampling, the bias is simply:

$$2\rho\sigma_v^2 \sum_{j=0}^{N/m-1} \left(1 - e^{-\lambda_m}\right)$$

Surprisingly, it turns out that the bias term associated with calendar time sampled returns is strictly *smaller* than the bias term associated with “business time” sampled returns. In order to show this it is sufficient to prove that

$$\sum_{j=0}^{N/m-1} e^{-\lambda^*(t+jm,m)} > \sum_{j=0}^{N/m-1} e^{-\lambda_m} \quad \text{or equivalently} \quad R \equiv \frac{m}{N} \sum_{j=0}^{N/m-1} e^{\lambda_m - \lambda^*(t+jm,m)} > 1$$

By the definition of λ_m and the convexity of the exponential function, the above inequality must hold as long as the intensity parameter is *non-constant*. Note that R measures the bias *reduction* associated with calendar time sampling relative to transaction sampling. This gain increases with an increase in the variability of $\lambda(\cdot)$. When the intensity parameter is constant, we have that $R = 1$, and both sampling schemes are equivalent.

Bias versus Mean Squared Error

The approach outlined above, classifies competing sampling schemes solely based on the relative magnitude of its associated bias. An alternative well known measure of performance is the mean squared error (MSE) which trades off a reduction in the bias against the loss of efficiency. While we have shown that calendar time sampling strictly dominates business time sampling when we use a bias-based ranking, it may very well be that this result is reversed when we use an MSE-based ranking which takes both bias and efficiency into account. Unfortunately, an analytic treatment of an MSE-based ranking of competing sampling schemes is not feasible because we do not have a closed form solution for the variance of the RV measure available. A small-scale simulation experiment

is therefore undertaken to gauge whether an MSE-based ranking will yield qualitatively different results than the bias-based ranking.

We focus on the single component compound Poisson process with MA(1) innovations and deterministic time variation of the intensity process, i.e. $\lambda(t) = s(t)$. The specification we use for $s(t)$ is similar to expression (3.12) with the indicator function left out. The parameter values are the same as discussed on page 70. Next, we simulate $T = 1000$ (disjoint) days of transaction prices. Let $F_t(u)$ denote the security price at time u during day t where $u \in [0, N]$ and $t = 1, \dots, T$. In addition, let $F_t(\tau_i)$ denote the security price associated with the i^{th} transaction on day t . The implementation of calendar time sampling is straightforward, i.e. for a given day t and a sampling frequency m , we sample N/m returns as

$$R_t^c(j|m) = F_t(jm) - F_t((j-1)m)$$

for $j = 1, \dots, N/m$. The corresponding business time sampling scheme, in contrast, samples the same amount of returns as follows:

$$R_t^b(j|k) = F_t(\tau_{jk}) - F_t(\tau_{(j-1)k})$$

for $j = 1, \dots, N/m$ and $k = mn_t/N$ where n_t denotes the total number of transactions for day t . Based on these return series we then compute the sample average of the realized variance measure under both sampling schemes:

$$CRV(m) = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{N/m} R_t^c(j|m)^2,$$

$$BRV(m) = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{N/m} R_t^b(j|k)^2,$$

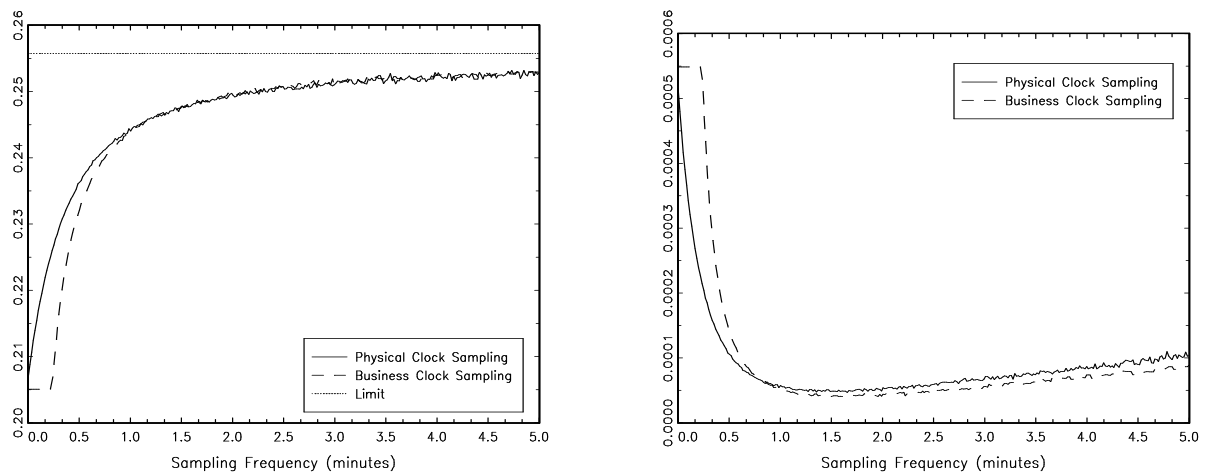
and the mean squared error under both sampling schemes:

$$CMSE(m) = \frac{1}{T} \sum_{t=1}^T \left\{ E_t [R_t^2] - \sum_{j=1}^{N/m} R_t^c(j|m)^2 \right\}^2,$$

$$BMSE(m) = \frac{1}{T} \sum_{t=1}^T \left\{ E_t [R_t^2] - \sum_{j=1}^{N/m} R_t^b(j|m)^2 \right\}^2.$$

Figure 3.10 displays all of the above statistics for sampling frequencies (m) between 1 second and 5 minutes. A number of interesting patterns arise. As expected, based on the bias-ranking, the calendar time scheme dominates. However, the difference in performance between both schemes rapidly shrinks as the sampling frequency decreases. At sampling frequencies lower than 1 minute, the difference is minimal which implies that the optimal sampling frequency will be the same for both schemes. In contrast, when the MSE is used to rank the sampling schemes, it appears that the business time sampling achieves the lowest overall MSE. Moreover, the sampling frequency which minimizes the MSE is substantially higher than the sampling frequency which minimizes the bias. Ignoring the efficiency loss associated with aggregation of returns, as is done for the bias-based ranking, clearly leads one to choose a much lower sampling frequency than if the MSE is taken as the relevant performance

FIGURE 3.10: REALIZED VARIANCE AND MSE FOR CALENDAR AND BUSINESS CLOCK SAMPLING



Notes: Sample average of realized variance measure (CRV and BRV , left panel) and mean squared error ($CMSE$ and $BMSE$, right panel) for “Calendar” clock (solid line) and “Business” clock” (dashed line) sampling schemes. The results are based on the single component compound Poisson process with MA(1) innovations and deterministic time variation in trade intensity, i.e. $\lambda(t) = s(t)$ where $s(t)$ is given by expression (3.12). The model parameters are set as $a = 4/$ minute, $b = 2.25$, $c = 0$, $\Delta = 0$, $\rho = 0.3$, and $\sigma_v^2 = 7e - 8$.

measure. Based on this simulation experiment we conclude that business time sampling dominates calendar time sampling when the objective is to either minimize the bias (in which case both schemes perform roughly equal) or minimize the MSE (in which case business time sampling dominates).

3.5 Conclusion

This article studies several extensions of the compound Poisson process which are able to capture important static and dynamic characteristics of high frequency security prices. In contrast to diffusion-based models, our framework is consistent with the finite variation property of high frequency returns and does not impose the usual martingale restriction on asset prices. By comparing the properties of simulated data to actual high frequency data we illustrate the flexibility of the model and its ability to capture important features of high frequency data including, (i) skewness, excess kurtosis and return serial correlation which diminishes under temporal aggregation, (ii) deterministic variation in trading activity such as the U-shaped intra-day pattern, day of the week effects, and the increased variance of the overnight return, and (iii) stochastic variation in trading activity leading to serial dependence in trade durations at high frequency (ACD-effects) and return volatility at low frequency (ARCH-effects). In addition, our models provide a useful context in which to investigate “market-microstructure-induced” serial correlation of returns at different sampling frequencies and its associated impact on the recently popularized realized volatility or variance measure. In particular, we show that for realistic parameter values the realized variance measure is a biased estimator of the integrated variance process and that the choice of sampling frequency proves crucial in minimizing this bias. Allowing for time variation in the trade intensity process yields interesting

insights into the properties of alternative, time-deformation-based, sampling schemes.

Throughout this paper we have illustrated the properties of the model for realistic parameter values but an issue which remains for future research is the specification and estimation of the various models using actual high frequency data. Specification and estimation of the bid-ask spread model is relatively straightforward because few parameters are involved and the availability of a closed form characteristic function suggests a variety of inference techniques that can be applied. For example, it is straightforward to specify a number of moment conditions (based on for example return variance, skewness, kurtosis, and serial correlation) and use GMM for estimation. Alternatively, it is possible to use a “continuum” of moment conditions by matching the theoretical characteristic function to its empirical counterpart (Carrasco, Chernov, Florens, and Ghysels 2002, Jiang and Knight 2002). However, the model that is potentially more relevant for the modelling of security prices both at low and high sampling frequency, is the multiple compound Poisson process with correlated innovations and time varying trade intensities. Specification and estimation of such a model is clearly more involved as it requires one to determine the MA-order, number of components, and dynamic specification of deseasonalized trade intensities. Also, the double OU process discussed above is clearly not well suited for practical purposes since it admits negative trade intensities and may cause numerical instability in the inference procedure. A host of alternative, and more appropriate, specification can be thought of including a simple log transformation or more elaborate non-Gaussian OU processes (Barndorff-Nielsen and Shephard 2001). Because analytic tractability is often lost for such models, moment conditions or characteristic function will have to be evaluated numerically, adding to the complexity of the inference procedure. Nevertheless, inference is still feasible and worthwhile pursuing as it may provide important insights into the dynamic properties of financial security prices across a range of sampling frequencies.

Bibliography

- Admati, A., and P. Pfleiderer, 1988, “A Theory of Intraday Patterns: Volume and Price Variability,” *Review of Financial Studies*, 1 (1), 3–40.
- Aït-Sahalia, Y., and P. A. Mykland, 2003, “How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise,” NBER working paper 9611.
- Amemiya, T., 1985, *Advanced Econometrics*. Harvard University Press, Boston, MA.
- Andersen, P. K., Ø. Borgan, R. Gill, and N. Keiding, 1993, *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- Andersen, T. G., 1996, “Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility,” *Journal of Finance*, 51 (1), 169–204.
- Andersen, T. G., L. Benzoni, and J. Lund, 2002, “An Empirical Investigation of Continuous-Time Equity Return Models,” *Journal of Finance*, 57(3), 1239–1284.
- Andersen, T. G., and T. Bollerslev, 1998, “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts,” *International Economic Review*, 39(4), 885–905.
- Andersen, T. G., T. Bollerslev, and A. Das, 2001, “Variance-Ratio Statistics and High Frequency Data: Testing for Changes in Intraday Volatility Patterns,” *Journal of Finance*, 56 (1), 305–327.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, “The Distribution of Stock Return Volatility,” *Journal of Financial Economics*, 61, 43–76.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2000a, “Exchange Rate Returns Standardized by Realized Volatility are (nearly) Gaussian,” *Multinational Finance Journal*, 4, 159–179.
- , 2000b, “Great Realizations,” *Risk*, pp. 105–108.
- , 2001, “The Distribution of Realized Exchange Rate Volatility,” *Journal of the American Statistical Association*, 96, 42–55.
- , 2003, “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71 (2), 579–625.

- Andersen, T. G., and J. Lund, 1997, “Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate,” *Journal of Econometrics*, 77(2), 343–377.
- Andersen, T. G., and B. E. Sørensen, 1996, “GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study,” *Journal of Business & Economic Statistics*, 14(3), 328–352.
- Andreu, E., and E. Ghysels, 2001, “When Does Microstructure Noise Affect the Estimation of Asset Pricing Models?,” Manuscript University of Manchester and North Carolina.
- Ané, T., and H. Geman, 2000, “Order Flow, Transaction Clock, and Normality of Asset Returns,” *Journal of Finance*, 55(5), 2259–2284.
- Areal, N. M., and S. J. Taylor, 2002, “The Realized Volatility of FTSE-100 Futures Prices,” *Journal of Futures Markets*, 22 (7), 627–648.
- Bai, X., J. R. Russell, and G. C. Tiao, 2001, “Beyond Merton’s Utopia (I): Effects of Non-Normality and Dependence on the Precision of Variance Estimates using High Frequency Data,” Manuscript University of Chicago.
- Baillie, R. T., 1996, “Long Memory Processes and Fractional Integration in Econometrics,” *Journal of Econometrics*, 73, 5–59.
- Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen, 1996, “Modelling and Pricing Long Memory in Stock Market Volatility,” *Journal of Econometrics*, 73, 151–184.
- Bakshi, G., C. Cao, and Z. Chen, 1997, “Empirical Performance of Alternative Option Pricing Model,” *Journal of Finance*, 52(5), 2003–2049.
- Bakshi, G., and D. Madan, 2000, “Spanning and Derivative-Security Valuation,” *Journal of Financial Economics*, 55(2), 205–238.
- Barndorff-Nielsen, O. E., and N. Shephard, 2001, “Non-Gaussian Ornstein-Uhlenbeck-Based Model and Some of Their Uses in Financial Economics,” *Journal of the Royal Statistical Society, Series B*, 63 (2), 167–241.
- Barndorff-Nielsen, O. E., and N. Shephard, 2002, “Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models,” *Journal of the Royal Statistical Society, Series B*, 64, 252–280.
- Barndorff-Nielsen, O. E., and N. Shephard, 2003, “Realised Power Variation and Stochastic Volatility Models,” forthcoming Bernoulli 9.
- Bates, D. S., 1996, “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options,” *Review of Financial Studies*, 9(1), 69–107.
- , 2000, “Post-87 Crash Fears in the S&P 500 Futures Option Market,” *Journal of Econometrics*, 94(1/2), 181–238.

- Bauwens, L., and P. Giot, 2001, *Econometric Modelling of Stock Market Intraday Activity*. Kluwer Academic Publishers, Boston.
- Bauwens, L., and D. Veredas, 2003, “The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations,” forthcoming *Journal of Econometrics*.
- Beckers, S., 1981, “A Note on Estimating the Parameters of the Diffusion-Jump Model of Stock Returns,” *Journal of Financial & Quantitative Analysis*, 16(1), 127–140.
- Black, F., 1976, “Studies of Stock Price Volatility Changes,” Proceeding from the American Statistical Association, Business and Economic Statistics Section.
- Blair, B. J., S.-H. Poon, and S. J. Taylor, 2001, “Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High-Frequency Index Returns,” *Journal of Econometrics*, 105, 5–26.
- Bollerslev, T., 1986, “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 51, 307–327.
- Bollerslev, T., J. Cai, and F. M. Song, 2000, “Intra-day Periodicity, Long Memory Volatility and Macroeconomic announcement effects in the US Treasury Bond Market,” *Journal of Empirical Finance*, 7, 37–55.
- Bollerslev, T., R. Y. Chou, and K. F. Kroner, 1992, “ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence,” *Journal of Econometrics*, 52, 5–59.
- Bollerslev, T., and R. F. Engle, 1986, “Modeling the Persistence of Conditional Variances,” *Econometric Reviews*, 5, 1–50.
- Bollerslev, T., R. F. Engle, and D. B. Nelson, 1994, “ARCH Models,” In: Engle, R.F., McFadden, D. (Eds.), *Handbook of Econometrics*, vol. 4, North-Holland, Amsterdam, pp. 2959–3038.
- Bollerslev, T., and H. Zhou, 2002, “Estimating Stochastic Volatility Diffusion using Conditional Moments of Integrated Volatility,” *Journal of Econometrics*, 109, 33–65.
- Boswijk, H. P., 2002, *Asymptotic Theory for Integrated Processes*. manuscript in preparation for Oxford University Press, Oxford.
- Breidt, F. J., N. Crato, and P. de Lima, 1998, “The Detection and Estimation of Long Memory in Stochastic Volatility,” *Journal of Econometrics*, 83, 325–348.
- Brewer, K., 1973, “Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX Models,” *Journal of Econometrics*, 1, 133–154.
- Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.

- Carr, P., H. Geman, D. B. Madan, and M. Yor, 2002, "The Fine Structure of Asset Returns: An Empirical Investigation," *Journal of Business*, 75 (2), 305–332.
- , 2003, "Stochastic Volatility for Levy Processes," *Mathematical Finance*, 13 (3), 345–382.
- Carrasco, M., M. Chernov, J.-P. Florens, and E. Ghysels, 2001, "Estimating Diffusions with a Continuum of Moment Conditions," Manuscript University of North Carolina.
- , 2002, "Efficient Estimation of Jump Diffusions and General Dynamic Models with a Continuum of Moment Conditions," Manuscript Columbia Business School.
- Chacko, G., and L. M. Viceira, 1999, "Spectral GMM Estimation of Continuous-Time Processes," Manuscript Graduate School of Business Administration, Harvard University.
- Chan, W. H., and J. M. Maheu, 2002, "Conditional Jump Dynamics in Stock Market Returns," *Journal of Business & Economic Statistics*, 20 (3), 377–389.
- Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen, 2002, "Alternative Models for Stock Price Dynamics," forthcoming *Journal of Econometrics*.
- Chib, S., F. Nardari, and N. Shephard, 2002, "Markov Chain Monte Carlo Methods for Stochastic Volatility Models," *Journal of Econometrics*, 108, 281–316.
- Clark, P. K., 1973, "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41(1), 135–155.
- Comte, F., and E. Renault, 1998, "Long Memory in Continuous Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291–323.
- Corsi, F., G. Zumbach, U. A. Müller, and M. Dacorogna, 2001, "Consistent High-Precision Volatility from High-Frequency Data," Olsen Group Working Paper.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "The Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.
- Cox, J. C., and S. A. Ross, 1976, "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics*, 3(1/2), 145–166.
- Dacorogna, M. M., R. Gençay, U. Müller, R. B. Olsen, and O. V. Pictet, 2001, *An Introduction to High-Frequency Finance*. Academic Press, London, UK.
- Dacorogna, M. M., U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet, 1993, "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market," *Journal of International Money and Finance*, 12, 413–438.

- Dai, Q., and K. J. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55(5), 1943–1978.
- Daniélsson, J., 1994, "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics*, 61, 375–400.
- Davison, A., and D. Hinkley, 1997, *Bootstrap Methods and their Applications*. Cambridge University Press, Cambridge, UK.
- Demsetz, H., 1968, "The Cost of Transacting," *Quarterly Journal of Economics*, 82, 33–53.
- Diamond, D., and R. Verrecchia, 1987, "Constraints on Short-Selling and Asset Price Adjustments to Private Information," *Journal of Financial Economics*, 18, 277–311.
- Diebold, F. X., and A. Inoue, 2001, "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131–159.
- Dittmann, I., and C. W. Granger, 2002, "Properties of Nonlinear Transformations of Fractionally Integrated Processes," *Journal of Econometrics*, 110 (2), 113–133.
- Doornik, J. A., 2002, *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, London, UK, 3 edn.
- Doornik, J. A., and M. Ooms, 1999, "A Package for Estimating, Forecasting and Simulating ARFIMA Models," manuscript Nuffield College, Oxford (www.nuff.ox.ac.uk/Users/Doornik).
- Duffie, D., and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6(4), 379–406.
- Duffie, D., J. Pan, and K. J. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68(6), 1343–1376.
- Duffie, D., and K. J. Singleton, 1993, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.
- Easley, D., N. Kiefer, and M. O'Hara, 1997, "One Day in the Life of a Very Common Stock," *Review of Financial Studies*, 10 (3), 805–835.
- Easley, D., and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47 (2), 577–605.
- Elerian, O., S. Chib, and N. Shephard, 2001, "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69 (4), 959–993.
- Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997, *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin and Heidelberg.

- Engle, R. F., 1982, "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation," *Econometrica*, 50, 987–1008.
- , 2000, "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68(1), 1–22.
- Engle, R. F., and J. R. Russell, 1998, "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66(5), 1127–1162.
- Engle, R. F., and D. Smith, 1999, "Stochastic Permanent Breaks," *Review of Economics & Statistics*, 81 (4), 553–574.
- Eraker, B., 2001, "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," Manuscript Duke University.
- Eraker, B., M. Johannes, and N. Polson, 2002, "The Impact of Jumps in Volatility and Returns," forthcoming *Journal of Finance*.
- Ethier, S., and T. Kurtz, 1986, *Markov Processes, Characterization and Convergence*. John Wiley & Sons, New York.
- Feller, W., 1968, *An Introduction to Probability Theory and Its Applications*, vol. 1. Wiley, New York, 3 edn.
- Figlewski, S., 1998, "Derivatives Risks, Old and New," *Brookings Wharton Papers on Financial Services*, pp. 159–238.
- Foster, D. P., and D. B. Nelson, 1996, "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64 (1), 139–174.
- French, K., and R. Roll, 1986, "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17, 5–26.
- French, K. R., G. W. Schwert, and R. Stambaugh, 1987, "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.
- Fridman, M., and L. Harris, 1998, "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 16(3), 284–291.
- Froot, K. A., and A. F. Perold, 1995, "New Trading Practices and Short-Run Market Efficiency," *Journal of Futures Markets*, 15(7), 731–765.
- Gallant, A., and G. Tauchen, 1998, "Reprojecting partially observed systems with application to interest rate diffusions," *Journal of the American Statistical Association*, 93, 10–24.
- Gallant, A. R., D. A. Hsieh, and G. E. Tauchen, 1997, "Estimation of Stochastic Volatility Models with Diagnostics," *Journal of Econometrics*, 81, 159–192.

- Gallant, A. R., P. E. Rossi, and G. Tauchen, 1992, "Stock Prices and Volume," *Review of Financial Studies*, 5, 199–242.
- Gallant, A. R., and G. Tauchen, 1996, "Which Moments to Match?," *Econometric Theory*, 12, 657–681.
- Geweke, J. F., and S. Porter-Hudak, 1983, "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4(4), 221–238.
- Ghysels, E., A. Harvey, and E. Renault, 1996, "Stochastic Volatility," In *Handbook of Statistics (14) Statistical Methods in Finance*, eds. G.S. Maddala and C.R. Rao, Elsevier Amsterdam, 119-191.
- Glosten, L., and P. R. Milgrom, 1985, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100.
- Glosten, L. R., R. Jagannathan, and D. E. Runkle, 1993, "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.
- Goodhart, C. A., and M. O'Hara, 1997, "High Frequency Data in Financial Markets: Issues and Applications," *Journal of Empirical Finance*, 4, 73–114.
- Granger, C. W., 1980, "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.
- , 1981, "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121–130.
- , 1999, "Aspects of Research Strategies for Time Series Analysis," Outline of presentation to the conference on New Developments in Time Series Economics at the Cowles Foundation, New Haven.
- Granger, C. W., and N. Hyung, 1999, "Occasional Structural Breaks and Long Memory," UCSD discussion paper 99-14.
- Granger, C. W., and R. Joyeux, 1980, "An Introduction to Long-Memory Time Series and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–30.
- Gupta, A., and M. G. Subrahmanyam, 2002, "Pricing and Hedging Interest Rate Options: Evidence from Cap-Floor Markets," manuscript Leonard N. Stern School of Business, New York University.
- Hansen, L. P., 1982, "Large Sample Properties of Generalized Method of Moments Estimator," *Econometrica*, 50(4), 1029–1054.
- Harris, L., 1986, "A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns," *Journal of Financial Economics*, 16, 99–117.
- Harrison, J. M., and D. M. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 2(3), 381–408.

- Harvey, A., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge UK.
- , 1993, *Time Series Models*. MIT Press, Cambridge, MA, 2nd edn.
- Harvey, A. C., E. Ruiz, and N. Shephard, 1994, “Multivariate Stochastic Volatility Models,” *Review of Economics & Statistics*, 61, 247–264.
- Harvey, A. C., and N. G. Shephard, 1996, “Estimation of an asymmetric stochastic volatility model for asset returns,” *Journal of Business & Economic Statistics*, 14, 429–434.
- Henry, M., and R. Payne, 1998, “An Investigation of Long Range Dependence in Intra-Day Foreign Exchange Rate Volatility,” Financial Markets Group discussion paper No. 264.
- Heston, S. L., 1993, “A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options,” *Review of Financial Studies*, 6, 327–344.
- Ho, M. S., W. R. Perraudin, and B. E. Sørensen, 1996, “A Continuous-Time Arbitrage-Pricing Model with Stochastic Volatility and Jumps,” *Journal of Business & Economic Statistics*, 14(1), 31–43.
- Ho, T., and H. Stoll, 1983, “The Dynamics of Dealer Markets under Competition,” *Journal of Finance*, 38, 1053–1074.
- Horowitz, J. L., I. Lobato, J. C. Nankervis, and N. Savin, 2002, “Bootstrapping the Box-Pierce Q Test: A Robust Test for Uncorrelatedness,” manuscript Northwestern University, Department of Economics.
- Hsieh, D. A., 1991, “Chaos and Nonlinear Dynamics: Application to Financial Markets,” *Journal of Finance*, 46(5), 1839–1877.
- Huang, R., and H. Stoll, 1997, “The Components of the Bid-Ask Spread: A General Approach,” *Review of Financial Studies*, 10 (4), 995–1034.
- Hull, J. C., 2003, *Options, Futures, and Other Derivatives*. Prentice Hall, Englewood Cliffs, New Jersey, 5 edn.
- Jacquier, E., N. G. Polson, and P. Rossi, 1994, “Bayesian Analysis of Stochastic Volatility Models,” *Journal of Business & Economic Statistics*, 12, 371–417, (with discussion).
- Jiang, G. J., and J. L. Knight, 2002, “Estimation of Continuous-Time Processes Via the Empirical Characteristic Function,” *Journal of Business & Economic Statistics*, 20 (2), 198–212.
- Jiang, G. J., and R. C. Oomen, 2002, “Hedging Derivatives Risks,” Manuscript University of Arizona and Warwick Business School.
- Jones, C. S., 1998, “Bayesian Estimation of Continuous-Time Finance Models,” manuscript University of Rochester.

- Karatzas, I., and S. E. Shreve, 1991, *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.
- Karlin, S., and H. M. Taylor, 1981, *A Second Course in Stochastic Processes*. Academic Press, New York.
- , 1997, *A First Course in Stochastic Processes*. Academic Press, New York.
- Karpoff, J., 1987, “The Relation Between Price Change and Trading Volume: A Survey,” *Journal of Financial & Quantitative Analysis*, 22, 109–126.
- Kendall, M., 1958, *The Advanced Theory of Statistics*, vol. I. Hafner Publishing Co., New York.
- Kenney, J., and E. Keeping, 1951, *Mathematics of Statistics*. Van Nostrand, Princeton, NJ, 2nd edn.
- , 1962, *Mathematics of Statistics* Van Nostrand, Princeton, NJ, chap. The k -Statistics, §7.9, pp. 99–100, 3rd edn.
- Kim, S., N. Shephard, and S. Chib, 1998, “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models,” *Review of Economics & Statistics*, 65, 361–393.
- Knight, J., and S. E. Satchell, 1997, “The Cumulant Generating Function Method Estimation, Implementation and Asymptotic Efficiency,” *Econometric Theory*, 13(2), 170–184.
- Koopman, S. J., and A. Harvey, 2003, “Computing Observation Weights for Signal Extraction and Filtering,” *Journal of Economic Dynamics & Control*, 27, 1317–1333.
- Kyle, A. S., 1985, “Continuous Auctions and Insider Trading,” *Econometrica*, 53 (6), 1315–1335.
- Lamoureux, C. G., and W. D. Lastrapes, 1990, “Heteroskedasticity in Stock Return Data: Volume versus GARCH effects,” *Journal of Finance*, 45(1), 221–229.
- Lequeux, P., 1999, *Financial Markets Tick by Tick: Insights in Financial Market Micro Structure*. John Wiley & Sons Ltd, England.
- Liesenfeld, R., and R. C. Jung, 2000, “Stochastic Volatility Models: Conditional Normality Versus Heavy-Tailed Distributions,” *Journal of Applied Econometrics*, 15, 137–160.
- Liu, M., 2000, “Modelling Long Memory in Stock Market Volatility,” *Journal of Econometrics*, 99, 139–171.
- Lo, A. W., 1991, “Long Memory in Stock Market Prices,” *Econometrica*, 59(5), 1279–1313.
- Lo, A. W., and A. C. MacKinlay, 1990, “An Econometric Analysis of Nonsynchronous-Trading,” *Journal of Econometrics*, 45, 181–212.
- Madan, D. B., and E. Seneta, 1984, “Compound Poisson Models for Economic Variable Movements,” *Sankhya Series B*, 46, 174–187.
- Madhavan, A. N., 2000, “Market Microstructure: A Survey,” *Journal of Financial Markets*, 3, 205–258.

- Maheu, J. M., and T. H. McCurdy, 2002a, “News Arrival, Jump Dynamics and Volatility Components for Individual Stock Returns,” Manuscript University of Toronto.
- , 2002b, “Nonlinear Features of Realized FX Volatility,” *Review of Economics & Statistics*, 84 (4), 668–681.
- Marcellino, M., 1999, “Some Consequences of Temporal Aggregation in Empirical Analysis,” *Journal of Business & Economic Statistics*, 17(1), 129–136.
- Martens, M., 2002, “Measuring and Forecasting S&P 500 Index-Futures Volatility Using High-Frequency Data,” *Journal of Futures Markets*, 22 (6), 497–518.
- Meddahi, N., 2001, “An Eigenfunction Approach for Volatility Modeling,” Working paper CRDE 29, University of Montreal.
- Melino, A., and S. M. Turnbull, 1990, “Pricing foreign currency options with stochastic volatility,” *Journal of Econometrics*, 45, 239–265.
- Merton, R. C., 1980, “On Estimating the Expected Return on the Market; An Exploratory Investigation,” *Journal of Financial Economics*, 8, 323–361.
- Morgan Guaranty Trust Company, 1996, *RiskMetrics Technical Document*. Morgan Guaranty Trust Company, New York, NY, 4 edn.
- Mürmann, A., 2001, “Pricing Catastrophe Insurance Derivatives,” Manuscript Insurance and Risk Management Department, The Wharton School.
- Nelson, D. B., 1991, “Conditional Heteroskedasticity in Asset Returns: A New Approach,” *Econometrica*, 59 (2), 347–370.
- Nelson, D. B., and D. P. Foster, 1994, “Asymptotic Filtering Theory for Univariate Arch Models,” *Econometrica*, 61 (1), 1–41.
- Newey, W., and K. West, 1987, “A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, 55, 703–708.
- Newey, W., and K. West, 1994, “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies*, 61, 631–653.
- O’Hara, M., 1995, *Market Microstructure Theory*. Blackwell Publishers Ltd, Oxford, UK.
- Oomen, R. C., 2002, “Modeling Realized Variance when Returns are Serially Correlated,” Manuscript Warwick Business School.
- , 2003, “Statistical Models for High Frequency Security Prices,” Manuscript Warwick Business School.

- Pan, J., 2002, "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3–50.
- Politis, D. N., and J. P. Romano, 1994, "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89 (428), 1303–1313.
- Press, S. J., 1967, "A Compound Events Model for Security Prices," *Journal of Business*, 40(3), 317–335.
- , 1968, "A Modified Compound Poisson Process with Normal Compounding," *Journal of the American Statistical Association*, 63, 607–613.
- Renault, E., and B. J. Werker, 2002, "Stochastic Volatility Models with Transaction Time Risk," manuscript Université de Montréal and Tilburg University.
- Robinson, P. M., 1995a, "Gaussian Semiparametric Estimation of Long Range Dependence," *Annals of Statistics*, 23, 1630–1661.
- , 1995b, "Log-Periodogram Regression of Time Series with Long Range Dependence," *Annals of Statistics*, 23(3), 1048–1072.
- Rogers, L., and O. Zane, 1998, "Designing and Estimating Models of High-Frequency Data," Manuscript University of Bath.
- Roll, R., 1984, "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39(4), 1127–1139.
- Rolski, T., H. Schmidli, V. Schmidt, and J. Teugels, 1999, *Stochastic Processes for Insurance and Finance*. John Wiley & Sons, Chichester.
- Romano, J. P., and L. A. Thombs, 1996, "Inference for Autocorrelations under Weak Assumptions," *Journal of the American Statistical Association*, 91 (434), 590–600.
- Rydberg, T. H., and N. Shephard, 2003, "Dynamics of Trade-by-Trade Price Movements: Decomposition and Models," forthcoming *Journal of Financial Econometrics*.
- Säfvenblad, P., 2000, "Trading Volume and Autocorrelation: Empirical Evidence from the Stockholm Stock Exchange," *Journal of Banking & Finance*, 24, 1275–1287.
- Sandmann, G., and S.-J. Koopman, 1998, "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood," *Journal of Econometrics*, 87(2), 271–301.
- Scott, L., 1997, "Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods," *Mathematical Finance*, 7(4), 413–426.
- Singleton, K. J., 2001, "Estimation of Affine Asset Pricing Models using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111–141.

- Smith, J., N. Taylor, and S. Yadav, 1997, "Comparing the Bias and Misspecification in ARFIMA models," *Journal of Time Series Analysis*, 18(5), 507–527.
- Sowell, F., 1992, "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics*, 53, 165–188.
- Stein, E. M., and J. C. Stein, 1991, "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," *Review of Financial Studies*, 4 (4), 727–752.
- Stoll, H., 1978, "The Pricing of Security Dealer Services: An Empirical Study of NASDAQ Stocks," *Journal of Finance*, 33 (4), 1153–1172.
- Stoll, H. R., and R. E. Whaley, 1990, "The Dynamics of Stock Index and Stock Index Futures Returns," *Journal of Financial & Quantitative Analysis*, 25(4), 441–468.
- Taylor, S. J., 1986, *Modelling Financial Time Series*. Wiley, Chichester.
- Taylor, S. J., and X. Xu, 1997, "The Incremental Volatility Information in One Million Foreign Exchange Quotations," *Journal of Empirical Finance*, 4, 317–340.
- Tiao, G., 1972, "Asymptotic Behavior of Temporal Aggregates of Time Series," *Biometrika*, 59, 525–531.
- Wei, W., 1981, "Effects of Systematic Sampling on ARIMA Models," *Communications in Statistical-Theoretical Mathematics*, A10, 2389–2398.
- Weiss, A., 1984, "Systematic Sampling and Temporal Aggregation in Time Series Models," *Journal of Econometrics*, pp. 271–281.
- Wong, E., 1964, "The Construction of a Class of Stationary Markov Processes," in *Stochastic Processes in Mathematical Physics and Engineering*, Proceedings of Symposia in Applied Mathematics, 16, ed. by R. Bellman. Providence, RI: American Mathematical Society, pp. 264-276.
- Wood, R., 2000, "Market Microstructure Research Databases: History and Projections," *Journal of Business & Economic Statistics*, 18(2), 140–145.
- Zhou, B., 1996, "High Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business & Economic Statistics*, 14(1), 45–52.

Appendix A

A.1 Proofs

Proof of Lemma 2.2.1 From (2.3) and (2.4), we have

$$\begin{aligned} & \psi(u; S_{t+k\delta} - S_{t+(k-1)\delta} \mid S_t, V_t) \\ = & E [\exp \{iu'(S_{t+k\delta} - S_{t+(k-1)\delta})\} \mid S_t, V_t] \\ = & E [E [\exp \{iu'S_{t+k\delta} - iu'S_{t+(k-1)\delta}\} \mid S_{t+(k-1)\delta}, V_{t+(k-1)\delta}] \mid S_t, V_t] \\ = & E [\exp \{C(\delta, u, 0) + (D1(\delta, u, 0)' - iu')S_{t+(k-1)\delta} + D2(\delta, u, 0)'V_{t+(k-1)\delta}\} \mid S_t, V_t] \\ = & \exp\{C(\delta, u, 0) + C[(k-1)\delta, -iD1(\delta, u, 0) - u, -iD2(\delta, u, 0)] \\ & + D1[(k-1)\delta, -iD1(\delta, u, 0) - u, -iD2(\delta, u, 0)]'S_t \\ & + D2[(k-1)\delta, -iD1(\delta, u, 0) - u, -iD2(\delta, u, 0)]'V_t\} \end{aligned}$$

The conditional cumulants follow by definition.■

Proof of Corollary 2.2.2 The results are follow directly from the equations in expression (2.8).■

A.2 Stochastic Mean Model

A.2.1 State-Space Representation

It is straightforward to show that the stochastic mean model in (2.12) can be written as:

$$\begin{aligned} R_{t+\tau} &= \frac{\kappa}{\beta - \kappa} (e^{-\kappa\tau} - e^{-\beta\tau}) X_t + \varepsilon_{t+\tau}^s \\ X_{t+\tau} &= e^{-\beta\tau} X_t + \varepsilon_{t+\tau}^x \end{aligned}$$

where $R_{t+\tau} \equiv S_{t+\tau} - e^{-\kappa\tau} S_t$, and ε_t^s and ε_t^x are iid jointly normal random variables with

$$\begin{aligned} V[\varepsilon_{t+\tau}^x] &= \frac{1 - e^{-2\tau\beta}}{2\beta} \sigma_x^2 \\ V[\varepsilon_t^s] &= \sigma^2 \frac{1 - e^{-2\tau\kappa}}{2\kappa} + \frac{\kappa\sigma_x^2 (1 - e^{-2\kappa\tau})}{2(\beta - \kappa)^2} + \frac{\kappa^2\sigma_x^2 (1 - e^{-2\beta\tau})}{2\beta(\beta - \kappa)^2} - 2 \frac{\kappa^2\sigma_x^2 (1 - e^{-(\beta+\kappa)\tau})}{(\beta + \kappa)(\beta - \kappa)^2} \\ E[\varepsilon_t^x \varepsilon_t^s] &= \frac{\kappa\sigma_x^2}{2\beta(\beta + \kappa)} + \frac{\kappa\sigma_x^2 e^{-2\tau\beta}}{2(\beta - \kappa)\beta} - \frac{\kappa\sigma_x^2 e^{-\tau(\beta+\kappa)}}{(\beta + \kappa)(\beta - \kappa)} \end{aligned}$$

Note that the marginal distribution of R (and S) $\sim \mathcal{N}\left(0, \frac{\sigma^2}{2\kappa} + \frac{\kappa\sigma_x^2}{2\beta(\beta+\kappa)}\right)$ and $X_t \sim \mathcal{N}\left(0, \frac{\sigma_x^2}{2\beta}\right)$. The above formulation is in (time-invariant) state-space form:

$$\begin{aligned} y_t &= Z\alpha_t + G\varepsilon_t && \text{(observation / measurement eqn)} \\ \alpha_{t+1} &= T\alpha_t + H\varepsilon_t && \text{(state / transition eqn)} \end{aligned}$$

where $\varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, I_2)$, $y_t = R_{t+\tau}$, $\alpha_t = X_t$, $Z = \kappa(\beta - \kappa)^{-1}(e^{-\kappa\tau} - e^{-\beta\tau})$, $T = e^{-\beta\tau}$, $G = (a; b)$, $H = (b; c)$ and

$$a = \frac{1 + \delta^2(\sigma_y^2 - \sigma_\alpha^2)}{2\delta}; \quad b = \delta\theta; \quad c = \frac{1 - \delta^2(\sigma_y^2 - \sigma_\alpha^2)}{2\delta}; \quad \delta = \left[\frac{\sigma_\alpha^2 + \sigma_y^2 + 2\sqrt{\sigma_\alpha^2\sigma_y^2 - \theta^2}}{4\theta^2 + (\sigma_y^2 - \sigma_\alpha^2)^2} \right]^{1/2}$$

for $\sigma_y^2 \equiv V[G\varepsilon_t] = V[\varepsilon_t^x] = a^2 + b^2$, $\sigma_\alpha^2 \equiv V[H\varepsilon_t] = V[\varepsilon_t^s] = b^2 + c^2$, and $\theta \equiv \text{Cov}[G\varepsilon_t, H\varepsilon_t] = E[\varepsilon_t^x \varepsilon_t^s] = ab + bc$.

A.2.2 Covariance Expressions

For the stochastic mean model in (2.12), the variance expression, $\text{Var}[S_{t+j\delta} | \mathcal{F}_t] = \text{Var}[S_{t+j\delta}]$, is given as:

$$-\frac{(-1 + e^{-2j\beta\delta})\sigma_x^2\kappa^2}{2\beta(\beta - \kappa)^2} + \frac{2(-1 + e^{-j\delta(\beta+\kappa)})\sigma_x^2\kappa^2}{(\beta - \kappa)^2(\beta + \kappa)} - \frac{(-1 + e^{-2j\delta\kappa})\left(\frac{\sigma_x^2\kappa^2}{(\beta - \kappa)^2} + \sigma^2\right)}{2\kappa}$$

The covariance expression, $\text{Cov}[S_{t+j\delta}, S_{t+k\delta} | \mathcal{F}_t] = \text{Cov}[S_{t+j\delta}, S_{t+k\delta}]$ for $j \neq k$, is given as:

$$\begin{aligned} &\frac{e^{-(j+k)\beta\delta}(-1 + e^{2k\beta\delta})\sigma_x^2\kappa^2}{2\beta(\beta - \kappa)^2} + \frac{e^{-(j+k)\delta\kappa}(-1 + e^{-k\delta(\beta+\kappa)})\sigma_x^2\kappa^2}{(\beta - \kappa)^2(\beta + \kappa)} \\ &-\frac{e^{-\delta(j\beta+k\kappa)}(-1 + e^{k\delta(\beta+\kappa)})\sigma_x^2\kappa^2}{(\beta - \kappa)^2(\beta + \kappa)} - \frac{e^{-(j+k)\delta\kappa}(-1 + e^{-2k\delta\kappa})\left(\frac{\sigma_x^2\kappa^2}{(\beta - \kappa)^2} + \sigma^2\right)}{2\kappa} \end{aligned}$$

A.3 Stochastic Volatility Model

For notational convenience we assume that $\mu = 0$. This implies that $E[x^2] = K[x^2]$ and $E[x^4] = K[x^4] + 3K[x^2]^2$. The results below can be generalized straightforwardly for a non-zero mean. The first conditional moment/cumulant of squared returns equals

$$E[(\Delta S_{t+j\delta})^2 | \mathcal{F}_t] = K[(\Delta S_{t+j\delta})^2 | \mathcal{F}_t] = a(j) + c(j)V_t$$

The cross moment of squared returns, $E[(\Delta S_{t+k\delta})^2 (\Delta S_{t+j\delta})^2 | \mathcal{F}_t]$, is given by:

$$\alpha^2 \delta^2 + (1 - e^{\beta\delta})^2 \frac{(\alpha - 2V_t) \sigma^2 + 2\beta(\alpha - V_t)^2}{2\beta^3 e^{(j+k)\beta\delta}} - \alpha \sigma^2 (e^{\beta\delta} - 1) \frac{4\rho^2 (1 + \beta\delta - e^{\beta\delta}) + 1 - e^{\beta\delta}}{2\beta^3 e^{(j-k+1)\beta\delta}} \\ + \delta (e^{\beta\delta} - 1) (V_t - \alpha) \frac{\beta\alpha + \sigma^2 \delta \rho^2 \beta + \sigma^2}{\beta^2 e^{j\beta\delta}} + \alpha \delta (e^{\beta\delta} - 1) \frac{V_t - \alpha}{\beta e^{k\beta\delta}}$$

for $j > k$. The fourth conditional cumulant of returns, $K[(\Delta S_{t+j\delta})^4 | \mathcal{F}_t]$, is given by

$$3\sigma^2 (\alpha - 2V_t) \frac{(1 - e^{\beta\delta})^2}{2\beta^3 e^{2j\beta\delta}} + 6\sigma^2 (\alpha - V_t) \frac{\delta^2 \rho^2 \beta^2 - (2\rho^2 + 1) (e^{\beta\delta} - \beta\delta - 1)}{\beta^3 e^{j\beta\delta}} \\ + 3\alpha \sigma^2 \frac{1 + e^{\beta\delta} (\beta\delta - 1) - 8\rho^2 (e^{\beta\delta} - 1) + 4\rho^2 \delta \beta (1 + e^{\beta\delta})}{\beta^3 e^{\beta\delta}}$$

from which we can derive the fourth moment using the following cumulant-moment relation:

$$E [(\Delta S_{t+j\delta})^4 | \mathcal{F}_t] = K [(\Delta S_{t+j\delta})^4 | \mathcal{F}_t] + 3K [(\Delta S_{t+j\delta})^2 | \mathcal{F}_t]^2$$

Appendix B

B.1 The Characteristic Function

Following Feller (1968), let X denote a random variable with probability measure μ . The characteristic function of X (or μ) is the function $\phi(\xi)$ defined for real ξ by:

$$\phi(\xi) \equiv E \left[e^{i\xi X} \right] = \int_{-\infty}^{\infty} e^{i\xi x} \mu(dx) = \int_{-\infty}^{\infty} (\cos(\xi x) + i \sin(\xi x)) \mu(dx).$$

The characteristic function of $aX + b$ equals $e^{ib\xi} \phi(a\xi)$. When X is Gaussian with zero mean and unit variance $\phi(\xi) = e^{-\frac{1}{2}\xi^2}$. Non-central moments (m_n) and cumulants (κ_n) of order n can be derived as:

$$m_n = i^{-n} \frac{\partial^n \phi(\xi)}{\partial \xi^n} \Big|_{\xi=0} \quad \text{and} \quad \kappa_n = i^{-n} \frac{\partial^n \ln \phi(\xi)}{\partial \xi^n} \Big|_{\xi=0}$$

There exists a one-to-one relationship between moments and cumulants of any order. For the first four orders they are as follows: $\kappa_1 = m_1$, $\kappa_2 = m_2 - m_1^2$, $\kappa_3 = m_3 - 3m_1m_2 + 2m_1^3$, and $\kappa_4 = m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4$. See Kendall (1958) for more details. The joint characteristic function of the set of random variables $\{X_j\}_{j=1}^k$ is given by:

$$\phi(\xi_1, \dots, \xi_k) \equiv E [\exp(i\xi_1 X_1 + i\xi_2 X_2 + \dots + i\xi_k X_k)],$$

which generates joint moments as follows:

$$E [X_1^{p_1} X_2^{p_2} \dots X_k^{p_k}] = i^{-\bar{p}} \frac{\partial^{\bar{p}} \phi(\xi_1, \dots, \xi_k)}{\partial \xi_1^{p_1} \partial \xi_2^{p_2} \dots \partial \xi_k^{p_k}} \Big|_{\xi=0},$$

where $\bar{p} = \sum_{i=1}^k p_i$.

B.2 The Intensity Process

The solution to the SDE in expression (3.14) directly follows from a general result on one-dimensional linear SDE as discussed by Karatzas and Shreve (1991, Section 5.6C):

$$\hat{\lambda}(t + \tau) = e^{-\kappa\tau} \hat{\lambda}(t) + \kappa \int_t^{t+\tau} e^{-\kappa(t+\tau-u)} \alpha(u) du + \sigma_\lambda \int_t^{t+\tau} e^{-\kappa(t+\tau-u)} dW_\lambda(u). \quad (\text{B.1})$$

for $\tau > 0$. The OU specification for α allows us to further specialize expression (B.1) above:

$$\begin{aligned} \widehat{\lambda}(t+\tau) &= \mu + e^{-\kappa\tau} \left(\widehat{\lambda}(t) - \mu \right) + \frac{\kappa(e^{-\kappa\tau} - e^{-\varphi\tau})}{\varphi - \kappa} (\alpha(t) - \mu) \\ &\quad + \kappa\sigma_\alpha \int_t^{t+\tau} \frac{e^{-\varphi(t+\tau-u)} - e^{-\kappa(t+\tau-u)}}{\kappa - \varphi} dW_\alpha(u) + \sigma_\lambda \int_t^{t+\tau} e^{-\kappa(t+\tau-u)} dW_\lambda(u) \end{aligned} \quad (\text{B.2})$$

using that

$$\int_t^{t+\tau} f(h) \left\{ \int_t^h g(h,u) dW(u) \right\} dh = \int_t^{t+\tau} \left\{ \int_u^{t+\tau} f(h) g(h,u) dh \right\} dW(u) \quad (\text{B.3})$$

where $f(h)$ and $g(h,u)$ are deterministic functions. Based on expression (B.2) above, it is straightforward to derive an expression for the integrated intensity process.

B.3 Proofs

Proof of Theorem 3.2.1 Let the characteristic function of innovations to the mid-price be given by

$$\phi_\varepsilon(\eta) = E_0 [e^{i\eta\varepsilon}] = e^{i\eta\mu_I - \frac{1}{2}\eta^2\sigma_I^2}$$

where $\varepsilon \sim \mathcal{N}(\mu_I, \sigma_I^2)$. Now derive the characteristic function of the mid-price process, i.e. $\phi_F(\eta, t) = E_0 [e^{i\eta F(t)}]$. Define $S(n) = \sum_{j=1}^n \varepsilon_j$ and notice that

$$\begin{aligned} \phi_F(\eta, t) &= \sum_{n=0}^{\infty} \frac{(t\lambda_I)^n e^{-t\lambda_I}}{n!} E_0 [e^{i\eta(F(0)+S(n))}] = e^{i\eta F(0) - t\lambda_I} \sum_{n=0}^{\infty} \frac{(t\lambda_I e^{i\eta\mu_I - \frac{1}{2}\eta^2\sigma_I^2})^n}{n!} \\ &= e^{i\eta F(0) - t\lambda_I} \sum_{n=0}^{\infty} \frac{[t\lambda_I \phi_\varepsilon(\eta)]^n}{n!} = e^{i\eta F(0) + t\lambda_I(\phi_\varepsilon(\eta) - 1)} \end{aligned}$$

using that $S(n) \sim \mathcal{N}(n\mu_I, n\sigma_I^2)$ and $\sum_{n=0}^{\infty} \frac{a^n}{n!} = e^a$. To derive the joint characteristic function of F and G , i.e. $\phi_{F,G}(\eta, \xi, t) = E_0 [e^{i\eta F(t) + i\xi G(t)}]$, use that:

$$\phi_{F,G}(\eta, \xi, t+h) - \phi_{F,G}(\eta, \xi, t) = E_0 \left(e^{i\eta F(t) + i\xi G(t)} E_t \left[e^{i\eta R_F(t+h|h) + i\xi R_G(t+h|h)} - 1 \right] \right).$$

Consider the random variable $e^{i\eta R_F(t+h|h) + i\xi R_G(t+h|h)}$ and notice that, for h sufficiently small, the memory-less property of the Poisson process implies:

$$\begin{aligned} \Pr \left[e^{i\eta R_F(t+h|h) + i\xi R_G(t+h|h)} = e^{-i\xi G(t) + i\eta \varepsilon_{M_I(t+h)}} \right] &= h\lambda_I, \\ \Pr \left[e^{i\eta R_F(t+h|h) + i\xi R_G(t+h|h)} = e^{i\xi(-\delta - G(t))} \right] &= h\lambda_S, \\ \Pr \left[e^{i\eta R_F(t+h|h) + i\xi R_G(t+h|h)} = e^{i\xi(\delta - G(t))} \right] &= h\lambda_B, \\ \Pr \left[e^{i\eta R_F(t+h|h) + i\xi R_G(t+h|h)} = 1 \right] &= 1 - h\bar{\lambda}, \end{aligned}$$

where $\varepsilon_{M_I(t+h)} \sim \mathcal{N}(\mu_I, \sigma_I^2)$. Therefore

$$E_t \left[e^{i\eta R_F(t+h|h) + i\xi R_G(t+h|h)} - 1 \right] = h\lambda_I e^{-i\xi G(t)} E_t e^{i\eta \varepsilon_{M_I(t+h)}} + h\lambda_B e^{i\xi(\delta - G(t))} + h\lambda_S e^{i\xi(-\delta - G(t))} - h\bar{\lambda}$$

Multiplying with $e^{i\eta F(t)+i\xi G(t)}$ yields:

$$E_t \left[e^{i\eta F(t+h)+i\xi G(t+h)} - e^{i\eta F(t)+i\xi G(t)} \right] = \left[h\lambda_I E_t e^{i\eta \varepsilon_{M_I(t+h)}} + h\lambda_B e^{i\xi\delta} + h\lambda_S e^{-i\xi\delta} \right] e^{i\eta F(t)} - h\bar{\lambda} e^{i\eta F(t)+i\xi G(t)}$$

Taking expectations of the above expression, dividing by h , and taking h to zero results in:

$$\begin{aligned} \frac{\partial \phi_{F,G}(\eta, \xi, t)}{\partial t} &= \lim_{h \rightarrow 0} \frac{\phi_{F,G}(\eta, \xi, t+h) - \phi_{F,G}(\eta, \xi, t)}{h} \\ &= \left[\lambda_I \phi_\varepsilon(\eta) + \lambda_B e^{i\xi\delta} + \lambda_S e^{-i\xi\delta} \right] \phi_F(\eta, t) - \bar{\lambda} \phi_{F,G}(\eta, \xi, t), \end{aligned} \quad (\text{B.4})$$

with the expressions for $\phi_\varepsilon(\eta, t)$ and $\phi_F(\eta, t)$ given above. Solving the differential equation in expression (B.4), subject to the boundary condition $\phi_{F,G}(\eta, \xi, 0) = e^{i\eta F(0)+i\xi G(0)}$, yields the joint characteristic function of F and G :

$$\phi_{F,G}(\eta, \xi, t) = f(\eta, \xi) \left(\phi_F(\eta, t) - e^{i\eta F(0)-t\bar{\lambda}} \right) + e^{i\eta F(0)+i\xi G(0)-t\bar{\lambda}}$$

where

$$f(\eta, \xi) = \frac{\lambda_I \phi_\varepsilon(\eta) + \lambda_B e^{i\xi\delta} + \lambda_S e^{-i\xi\delta}}{\lambda_I \phi_\varepsilon(\eta) + \lambda_B + \lambda_S}$$

This completes the proof of expression 3.6.

Now, based on the joint characteristic function of F and G , it is straightforward to derive that for $m > 0$:

$$\begin{aligned} \phi_{F,G}^*(\eta_1, \eta_2, \xi_1, \xi_2, t, m) &= E_0 \left[e^{i\eta_1 F(t)+i\eta_2 F(t+m)+i\xi_1 G(t)+i\xi_2 G(t+m)} \right] \\ &= E_0 \left[e^{i\eta_1 F(t)+i\xi_1 G(t)} E_t e^{i\eta_2 F(t+m)+i\xi_2 G(t+m)} \right] \\ &= E_0 \left[e^{i\eta_1 F(t)+i\xi_1 G(t)} a(\eta_2, \xi_2) \left(\phi_{F,t}(\eta_2, m) - e^{i\eta_2 F(t)-m\bar{\lambda}} \right) \right] \\ &\quad + E_0 \left[e^{i\eta_1 F(t)+i\xi_1 G(t)} e^{i\eta_2 F(t)+i\xi_2 G(t)-m\bar{\lambda}} \right] \\ &= f(\eta_2, \xi_2) \phi_{F,G}(\eta_1 + \eta_2, \xi_1, t) \left(e^{m\lambda_I(\phi_\varepsilon(\eta_2)-1)} - e^{-m\bar{\lambda}} \right) \\ &\quad + e^{-m\bar{\lambda}} \phi_{F,G}(\eta_1 + \eta_2, \xi_1 + \xi_2, t) \end{aligned}$$

which completes the proof of expression 3.5. ■

Proof of Theorem 3.3.1 Define the cumulative innovations $S(n) = \sum_{j=1}^n \varepsilon_j$ and notice that the joint characteristic function of $F(t)$ and $F(t+m)$ can be written as

$$E_0 \left[e^{i\xi_1 F(t)+i\xi_2 F(t+m)} \right] = a(\bar{\xi}) \phi_S^*(\xi_1, \xi_2, t, m)$$

where $\phi_S^*(\xi_1, \xi_2, t, m) \equiv E_0 \left[e^{i\xi_1 S(M_I(t))+i\xi_2 S(M_I(t+m))} \right]$, $\bar{\xi} = \xi_1 + \xi_2$ and $a(\xi) = \exp(i\xi F(0))$. The variance of $S(n)$ equals:

$$\Sigma_q(n) = n\sigma_\nu^2 \sum_{j=0}^q \rho_j^2 + 2\sigma_\nu^2 \sum_{h=1}^{\min(q,n)} \sum_{j=h}^q (n-h) \rho_j \rho_{j-h}, \quad (\text{B.5})$$

which, for $n \geq q$, simplifies to:

$$\Sigma_q(n) = n\sigma_\nu^2 \bar{\rho}^2 - 2\sigma_\nu^2 \rho(q, q),$$

where $\bar{\rho} = \sum_{j=0}^q \rho_j$ and

$$\rho(q, k) = \begin{cases} \sum_{h=1}^{\min(q,k)} \sum_{j=h}^q h \rho_j \rho_{j-h} & \text{for } q \geq 1, k \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

Note that $S(n) \sim \mathcal{N}(n\bar{\rho}\mu_\nu, \Sigma_q(n))$ and thus $E[e^{i\xi S(n)}] = e^{i\xi n\bar{\rho}\mu_\nu - \frac{1}{2}\xi^2 \Sigma_q(n)}$. The covariance of $S(n)$ and $S(n+h)$ equals:

$$\Sigma_q(n, h) = \Sigma_q(n) + \sigma_\nu^2 \rho(q, h),$$

and because $S(n)$ and $S(n+h)$ are jointly normal, their joint characteristic function can be derived as:

$$\begin{aligned} E_0 \left[e^{i\xi_1 S(n) + i\xi_2 S(n+h)} \right] &= e^{i\bar{\xi} n \bar{\rho} \mu_\nu + i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} [\xi_1^2 \Sigma_q(n) + \xi_2^2 \Sigma_q(n+h) + 2\xi_1 \xi_2 \Sigma_q(n, h)]} \\ &= e^{i\bar{\xi} n \bar{\rho} \mu_\nu + i\xi_2 h \bar{\rho} \mu_\nu} e^{-\frac{1}{2} [(\xi_1^2 + 2\xi_1 \xi_2) \Sigma_q(n) + \xi_2^2 \Sigma_q(n+h) + 2\xi_1 \xi_2 \sigma_\nu^2 \rho(q, h)]} \end{aligned}$$

Recall that

$$\phi_S^*(\xi_1, \xi_2, t, m) = E_0 \left[\sum_{h=0}^{\infty} \sum_{n=0}^{\infty} e^{i\xi_1 S(n) + i\xi_2 S(n+h)} \frac{(m\lambda)^h}{h! e^{m\lambda}} \frac{(t\lambda)^n}{n! e^{t\lambda}} \right] \quad (\text{B.6})$$

which, for t sufficiently large¹, can be approximated accurately by:

$$\begin{aligned} & e^{\bar{\xi}^2 \sigma_\nu^2 \rho(q, q)} \sum_{h=0}^{\infty} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} [h\xi_2^2 \sigma_\nu^2 \bar{\rho}^2 + 2\xi_1 \xi_2 \sigma_\nu^2 \rho(q, h)]} \frac{(m\lambda)^h}{h! e^{m\lambda}} \sum_{n=0}^{\infty} e^{i\bar{\xi} n \bar{\rho} \mu_\nu - \frac{1}{2} \bar{\xi}^2 n \sigma_\nu^2 \bar{\rho}^2} \frac{(t\lambda)^n}{n! e^{t\lambda}} \\ &= b(\bar{\xi}, t) e^{\bar{\xi}^2 \sigma_\nu^2 \rho(q, q)} \sum_{h=0}^{\infty} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} [h\xi_2^2 \sigma_\nu^2 \bar{\rho}^2 + 2\xi_1 \xi_2 \sigma_\nu^2 \rho(q, h)]} \frac{(m\lambda)^h}{h! e^{m\lambda}} \end{aligned}$$

where $b(\xi, t) = \exp \left[t\lambda \left(e^{i\bar{\xi} \bar{\rho} \mu_\nu - \frac{1}{2} \bar{\xi}^2 \sigma_\nu^2 \bar{\rho}^2} - 1 \right) \right]$. The summation over h can be rewritten as:

$$\sum_{h=0}^{q-1} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} [h\xi_2^2 \sigma_\nu^2 \bar{\rho}^2 + 2\xi_1 \xi_2 \sigma_\nu^2 \rho(q, h)]} \frac{(m\lambda)^h}{h! e^{m\lambda}} + e^{-\xi_1 \xi_2 \sigma_\nu^2 \rho(q, q)} \sum_{h=q}^{\infty} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} h\xi_2^2 \sigma_\nu^2 \bar{\rho}^2} \frac{(m\lambda)^h}{h! e^{m\lambda}}$$

where

$$\sum_{h=q}^{\infty} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} h\xi_2^2 \sigma_\nu^2 \bar{\rho}^2} \frac{(m\lambda)^h}{h! e^{m\lambda}} = b(\xi_2, m) - \sum_{h=0}^{q-1} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} h\xi_2^2 \sigma_\nu^2 \bar{\rho}^2} \frac{(m\lambda)^h}{h! e^{m\lambda}}$$

Collecting above expressions yields:

$$\begin{aligned} \phi_S^*(\xi_1, \xi_2, t, m) &= b(\bar{\xi}, t) e^{\bar{\xi}^2 \sigma_\nu^2 \rho(q, q)} \sum_{h=0}^{q-1} e^{i\xi_2 h \bar{\rho} \mu_\nu - \frac{1}{2} h\xi_2^2 \sigma_\nu^2 \bar{\rho}^2} \left(e^{-\xi_1 \xi_2 \sigma_\nu^2 \rho(q, h)} - e^{-\xi_1 \xi_2 \sigma_\nu^2 \rho(q, q)} \right) \frac{(m\lambda)^h}{h! e^{m\lambda}} \\ &\quad + b(\bar{\xi}, t) b(\xi_2, m) e^{\bar{\xi}^2 \sigma_\nu^2 \rho(q, q)} e^{-\xi_1 \xi_2 \sigma_\nu^2 \rho(q, q)} \end{aligned}$$

which completes the proof of expression 3.8. ■

¹Strictly speaking this is an approximation to the true characteristic function (which can be avoided at the cost of cumbersome notation) since $\Sigma_q(n)$ is approximated by $n\sigma_\nu^2 \bar{\rho}^2 - 2\sigma_\nu^2 \rho(q, q)$ for all $n \geq 0$ while this is only justified for $n \geq q$. However, q is typically small (say 1 or 2) and the contribution of the terms for which the variance expression is incorrect is negligible when t is large. Moreover, when calculating the unconditional moments, i.e. having $t \rightarrow \infty$, the approximation is exact.

Proof of Corollary 3.3.2 Define the cumulative innovations $S_r(n) = \sum_{j=1}^n \varepsilon_{r,j}$. The joint characteristic function of $S_r(n)$ and $S_r(n+k)$ is derived in the proof of Theorem 3.3.1. Because $Cov[S_h(n), S_j(n')] = 0$ for $h \neq j$ and $n, n' > 0$ it directly follows that:

$$E_0 \left[e^{i\xi_1 F(t) + i\xi_2 F(t+m)} \right] = a(\bar{\xi}) \prod_{r=1}^k \phi_{S_r}^*(\xi_1, \xi_2, t, m)$$

where $a(\xi) = \exp(i\xi F(0))$ and $\phi_{S_r}^*(\xi_1, \xi_2, t, m) = E_0 \left[e^{i\xi_1 S_r(M_r(t)) + i\xi_2 S_r(M_r(t+m))} \right]$. ■

Proof of Corollary 3.3.3 Follows directly from the proof of Theorem 3.3.1