STOCHASTIC FINANCE

STOCHASTIC FINANCE

Edited by

A.N. SHIRYAEV
Steklov Mathematical Institute, Moscow, Russia

M.R. GROSSINHO
ISEG, Universidade Técnica de Lisboa, Portugal

P.E. OLIVEIRA
Universidade de Coimbra, Portugal

M.L. ESQUÍVEL
FCT, Universidade Nova de Lisboa, Monte de Caparica, Portugal

Springer

Library of Congress Cataloging-in-Publication Data

Stochastic finance / edited by A.N. Shiryaev ... [et al.]/
p. cm.
"From 26–30 September 2004, the 'International Conference on Stochastic Finance 2004' took place at Instituto Superior de Economia e Gestão (ISEG) da Universidade Técnica de Lisboa, in Portugal"—Preface.
Includes bibliographical references.
ISBN-10: 0-387-28262-9 (alk. paper)
ISBN-10: 0-387-28359-5 (e-book)
ISBN-13: 978-0387-28262-6 (ISBN-13)
ISBN-13: 978-0387-28359-3 (e-ISBN-13)
1. Business mathematics—Congresses. 2. Finance—Congresses. 3. Stochastic analysis—Congresses. I. Shiriaev, Albert Nikolaevich.

HF5691.S68 2006
332'.01'51922—dc22

2005051577

AMS Subject Classifications: 91B70, 91B28, 60G70

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11367710

springeronline.com

Preface

From 26–30 September 2004, the "International Conference on *Stochastic Finance 2004*" took place at INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO (ISEG) da Universidade Técnica de Lisboa, in Portugal. The conference was one of the biggest international forums for scientists and practitioners working in financial mathematics and financial engineering.

Taking place just before the conference, on 20–24 September 2004 was the "Autumn School on Stochastic Finance 2004" hosted by the Universidade de Coimbra. The goal of this event was to present instances of the interaction of finance and mathematics by means of a coherent combination of five courses of introductory lectures, delivered by specialists, in order to stimulate and reinforce the understanding of the subject and to provide an opportunity for graduate students and researchers to develop some competence in financial mathematics and thereby simplify their participation in the conference.

At both meetings the organizing and scientific committees worked in close contact, which was crucial for inviting many leading specialists in financial mathematics and financial engineering — eleven plenary lecturers and eleven invited speakers. Besides these presentations, the conference included more than eighty contributed talks distributed among eight thematic sessions: *Mathematical Finance-Stochastic Models, Derivative Pricing, Interest Rate Term Structure Modelling, Portfolio Management, Integrated Risk Management, Mathematical Economics, Finance, and Quantitative and Computational Models and Methods.*

Stochastic financial mathematics is now one of the most rapidly developing fields of mathematics and applied mathematics. It has very close ties with economics and is oriented to the solution of problems appearing every day in real financial markets. We recall here an extract from the "Editorial" note presented in volume 1, issue 1 of the journal *Finance and Stochastics* that Springer-Verlag began publishing in 1997:

"Nearly a century ago, Louis Bachelier published his thesis "Théorie de la speculation", Ann. Sci. École Norm. Sup. 3 (1900), in which he invented Brownian motion as a tool for the analysis of financial markets. A.N. Kolmogorov, in his own landmark work "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung", Math. Annalen 104 (1931), pp.415-458, credits Bachelier with the first systematic study of stochastic processes in continuous time. But in addition, Bachelier's thesis marks the beginning of the theory of option pricing, now an integral part of modern finance. Thus the year 1900 may be considered as birth date of both Finance and Stochastics. For the first seven decades following Bachelier, finance and stochastics followed more or less independently. The theory of stochastic processes grew fast and incorporating classical calculus became a powerful mathematical tool — called stochastic calculus. Finance lay dormant until the middle of the twentieth century, and then was resurrected as an offshoot of general equilibrium theory in economics. With the work in the late 1960s and early 1970s of Black, Merton, Samuelson and Scholes, modelling stock prices as geometric Brownian motion and using this model to study equilibrium and arbitrage pricing, the two disciplines were reunited. Soon it was discovered how well suited stochastic calculus with its rich mathematical structure — martingale theory, Itô calculus, stochastic integration and PDE's — was for a rigorous analysis of contemporary finance, which would lead one to believe (erroneously) that also these tools were invented with the application to finance in mind. Since then the interplay of these two disciplines has become an ever growing research field with great impact both on the theory and practice of financial markets".

The aims formulated in this text were the leading ideas for our conference. Indeed, all talks had, first of all, financial meanings and interpretations. All talks used and developed stochastic methods or solutions for real problems. Such joint mutual collaboration was useful both for financial economics and stochastic theory, and it could bring the mathematical and financial communities together.

In the present volume the reader can find some papers based on the plenary and invited lectures and on some contributed talks selected for publication.

The editorial committee of these proceedings expresses its deep gratitude to those who contributed their work to this volume and those who kindly helped us in refereeing them.

It is our pleasure to express our thanks to the scientific committee of the conference, as well as to plenary and invited lecturers and all the participants of Stochastic Finance 2004; their presence and their work formed the main contribution to the success of the conference.

A special acknowledgement is due to the Governador do Banco de Portugal (Governor of the Portuguese Central Bank) for his sharp advice and sponsorship of the event.

We thank the financial supporters:

Arkimed Innovative Technologies,
Associacão Portuguesa de Seguradores,
Caixa Geral de Depósitos,
Comissão do Mercado de Valores Mobiliários,
Delta Cafés,
Fundação Luso-Americana,
Fundação Oriente,
Império Bonança,
Portugalia Airlines
Grupo Sumol.

Our gratitude goes to CIM (Centro Internacional de Matemática) for suggesting the organization of this event within its annual scientific planning, and to the academic institutions

CMUC (Centro de Matemática da Universidade de Coimbra),
FCT-UNL (Fac. de Ciências e Tecnologia da Univ. Nova de Lisboa),
UECE (Unidade de Estudos sobre a Complexidade na Economia),
FCT (Fundação para a Ciência e a Tecnologia),

for their support.

A special word of thanks is due to CEMAPRE (Centro de Matemática Aplicada à Previsão e Decisão Económica) whose support has been crucial for the viability of the event. We thank the staff that at different moments and in diverse tasks were key collaborators to the organizing procedure: Ana Sofia Nunes (computer support), Maria do Rosário Pato (secretary) and Maria Júlia Marmelada (public relations).

Thanks are due to Béatrice Huberty, the editorial secretary who prepared this volume, for her proficiency and dedicated work.

Our appreciation goes John Martindale and Robert Saley, editor and assistant editor of Springer, respectively, for their continuous support and active interest in the development of this project.

We sincerely hope that this volume will be an essential contribution to the literature in financial mathematics and financial engineering.

Lisbon,

Albert Shiryaev
Maria do Rosário Grossinho
Paulo Eduardo Oliveira
Manuel Leote Esquível

International Conference

Stochastic Finance 2004

Lisbon, 26-30 September

Chairpersons

- *Albert N. Shiryaev* - Scientific Committee
  Steklov Mathematical Institute, Moscow, Russia
- *Maria do Rosário Grossinho* - Organizing Committee
  ISEG, Universidade Técnica de Lisboa, Portugal

Scientific Committee

- *Ole Barndorff-Nielsen*, University of Aarhus, Denmark
- *Ernst Eberlein,* University of Freiburg, Germany
- *Albert N. Shiryaev,* Steklov Mathematical Institute, Moscow, Russia

Organizing Committee

- *Paulo Brito,* ISEG, Universidade Técnica de Lisboa, Portugal
- *Manuel L. Esquível*, FCT, Universidade Nova de Lisboa, Portugal
- *Maria do Rosário Grossinho,* ISEG, Univ. Técnica de Lisboa, Portugal
- *João Nicolau*, ISEG, Universidade Técnica de Lisboa, Portugal
- *Paulo Eduardo Oliveira,* Universidade de Coimbra, Portugal

Plenary Lecturers

- *Yacine Aït-Sahalia,* Princeton University, USA
- *Ole E. Barndorff-Nielsen*, University of Aarhus, Denmark
- *Tomas Björk,* Stockholm School of Economics, Sweden
- *Hans Föllmer,* Humboldt Universität zu Berlin, Germany
- *Paul Embrechts,* ETH Zurich, Switzerland
- *Ioannis Karatzas,* Columbia University, USA
- *Nicole El Karoui*, Ecole Polytechnique, Palaiseau, France
- *Stanley Pliska,* University of Illinois at Chicago, USA
- *Marek Rutkowski*, Warsaw University of Technology, Poland
- *Eckhard Platen,* University of Technology, Sydney, Australia
- *Albert Shiryaev,* Steklov Mathematical Institute, Russia

Invited Speakers

- *Bent Jesper Christensen*, University of Aarhus, Denmark
- *Ernst Eberlein,* University of Freiburg, Germany
- *Monique Jeanblanc,* Université d'Evry Val d'Essonne, France
- *Youri Kabanov,* Université de Franche-Comté, France
- *Claudia Klüeppelberg*, München University of Technology, Germany
- *Dmitry Kramkov,* Carnegie Mellon University, USA
- *Yoshio Miyahara,* Nagoya City University, Japan
- *Marek Musiela,* BNP Paribas Bank
- *Wolfgang Runggaldier,* Universitá di Padova, Italy
- *Walter Schachermayer,* Vienna University of Technology, Austria
- *Michael Søerensen*, University of Copenhagen, Denmark
- *Nizar Touzi,* CREST, Laboratoire de Finance et Assurance, France
- *Xun-Yu Zhou,* The Chinese University of Hong Kong

Contents

Part II Contributed Talks

List of Contributors

Albrecht Irle
Mathematisches Seminar
Universität Kiel
D-24098 Kiel, Germany
irle@math.uni-kiel.de

Alexander Lindner
Center for Mathematical Sciences
Munich University of Technology
D-85747 Garching, Germany
lindner@ma.tum.de

Claudia Klüppelberg
Center for Mathematical Sciences
Munich University of Technology
D-85747 Garching, Germany
cklu@ma.tum.de

Eckhard Platen
University of Technology Sydney,
School of Finance & Economics
and Department of Mathematical
Sciences
PO Box 123
Broadway, NSW, 2007, Australia
eckhard.platen@uts.edu.au

Ferry Jaya Permana
Delft University of Technology
Mekelweg 4
2628 CD Delft, The Netherlands
f.j.permana@ewi.tudelft.nl

Jeannette H.C. Woerner
Institut für Mathematische
Stochastik, Universität Göttingen
Maschmühlenweg 8-10
D-37073 Göttingen, Germany
woerner@math.uni-goettingen.de

Jörn Sass
RICAM, Austrian Academy of
Sciences
Altenberger Str. 69
A-4040 Linz, Austria
joern.sass@oeaw.ac.at

Lan Zhang
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213, USA
lzhang@stat.cmu.edu

Lian Yu
Department of Industrial Engineering and Operations Research
University of California
Berkeley, CA 94720
liany@ieor.berkeley.edu

Marek Rutkowski
School of Mathematics
University of New South Wales
Sydney, NSW 2052, Australia
marekr@maths.unsw.edu.au

Monique Jeanblanc
Département de Mathématiques
Université d'Évry Val d'Essonne
91025 Évry Cedex, France
monique.jeanblanc@univ-evry.fr

Neil Shephard
Nuffield College
Oxford OX1 1NF, U.K.
neil.shephard@nuf.ox.ac.uk

Ole E. Barndorff-Nielsen
Department of Mathematical
Sciences, University of Aarhus
Ny Munkegade
DK-8000 Aarhus C, Denmark
oebn@imf.au.dk

Per A. Mykland
Department of Statistics
The University of Chicago
Chicago, IL 60637, USA
mykland@uchicago.edu

Raquel M. Gaspar
Department of Finance, Stockholm
School of Economics
Dept. Finance, P.O. Box 6501
113 83 Stockholm, Sweden
raquel.gaspar@hhs.se

Sara Di Emidio
Dipartimento di Matematica Pura
ed Applicata
Universitá di Padova
Padova, Italy
sara.diemidio@gmail.com

Shuzhong Zhang
Department of Systems Engineering
and Engineering Management
The Chinese University of Hong
Kong Shatin
Hong Kong
zhang@se.cuhk.edu.hk

Stanley R. Pliska
Department of Finance, University
of Illinois at Chicago
601 S. Morgan Street
Chicago, IL 60607-7124 USA
srpliska@uic.edu

Svetlana Borovkova
Delft University of Technology
Mekelweg 4
2628 CD Delft, The Netherlands
s.a.borovkova@ewi.tudelft.nl

Tomasz R. Bielecki
Department of Applied Mathematics
Illinois Institute of Technology
Chicago, IL 60616, USA
bielecki@iit.edu

Vicky Fasen
Center for Mathematical Sciences
Munich University of Technology
D-85747 Garching, Germany
fasen@ma.tum.de

Wolfgang J.Runggaldier
Dipartimento di Matematica Pura
ed Applicata
Universitá di Padova
Padova, Italy
runggal@math.unipd.it

Xun Yu Zhou
Department of Systems Engineering
and Engineering Management
The Chinese University of Hong
Kong Shatin
Hong Kong
xyzhou@se.cuhk.edu.hk

Yacine Aït-Sahalia
Bendheim Center for Finance,
Princeton University,
Princeton, NJ 08540, USA
yacine@princeton.edu

Part I

Plenary and Invited Lectures

1

How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise*

Yacine Aït-Sahalia, Per A. Mykland, and Lan Zhang

[1] Bendheim Center for Finance, Princeton University, Princeton, NJ 08540 and NBER
yacine@princeton.edu
[2] Department of Statistics, The University of Chicago, Chicago, IL 60637
mykland@uchicago.edu
[3] Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213
lzhang@stat.cmu.edu

Note: Reprint of: Yacine Aït-Sahalia, Per A. Mykland and Lan Zhang, *How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise*, Review of Financial Studies, published in Volume 18 **2** (2005), pages 351–416, Oxford University Press (by permission of Oxford University Press).

Summary. In theory, the sum of squares of log returns sampled at high frequency estimates their variance. When market microstructure noise is present but unaccounted for, however, we show that the optimal sampling frequency is finite and derive its closed-form expression. But even with optimal sampling, using say five minute returns when transactions are recorded every second, a vast amount of data is discarded, in contradiction to basic statistical principles. We demonstrate that modelling the noise and using all the data is a better solution, even if one misspecifies the noise distribution. So the answer is: sample as often as possible.

Over the past few years, price data sampled at very high frequency have become increasingly available, in the form of the Olsen dataset of currency exchange rates or the TAQ database of NYSE stocks. If such data were not affected by market microstructure noise, the realized volatility of the process

* We are grateful for comments and suggestions from the editor, Maureen O'Hara, and two anonymous referees, as well as seminar participants at Berkeley, Harvard, NYU, MIT, Stanford, the Econometric Society and the Joint Statistical Meetings. Financial support from the NSF under grants SBR-0111140 (Aït-Sahalia), DMS-0204639 (Mykland and Zhang) and the NIH under grant RO1 AG023141-01 (Zhang) is also gratefully acknowledged.

(i.e., the average sum of squares of log-returns sampled at high frequency) would estimate the returns' variance, as is well known. In fact, sampling as often as possible would theoretically produce in the limit a perfect estimate of that variance.

We start by asking whether it remains optimal to sample the price process at very high frequency in the presence of market microstructure noise, consistently with the basic statistical principle that, ceteris paribus, more data is preferred to less. We first show that, if noise is present but unaccounted for, then the optimal sampling frequency is finite, and we derive a closed-form formula for it. The intuition for this result is as follows. The volatility of the underlying efficient price process and the market microstructure noise tend to behave differently at different frequencies. Thinking in terms of signal-to-noise ratio, a log-return observed from transaction prices over a tiny time interval is mostly composed of market microstructure noise and brings little information regarding the volatility of the price process since the latter is (at least in the Brownian case) proportional to the time interval separating successive observations. As the time interval separating the two prices in the log-return increases, the amount of market microstructure noise remains constant, since each price is measured with error, while the informational content of volatility increases. Hence very high frequency data are mostly composed of market microstructure noise, while the volatility of the price process is more apparent in longer horizon returns. Running counter to this effect is the basic statistical principle mentioned above: in an idealized setting where the data are observed without error, sampling more frequently cannot hurt. What is the right balance to strike? What we show is that these two effects compensate each other and result in a finite optimal sampling frequency (in the root mean squared error sense) so that some time aggregation of the returns data is advisable.

By providing a quantitative answer to the question of how often one should sample, we hope to reduce the arbitrariness of the choices that have been made in the empirical literature using high frequency data: for example, using essentially the same Olsen exchange rate series, these somewhat ad hoc choices range from 5 minute intervals (e.g., [5], [8] and [19]) to as long as 30 minutes (e.g., [6]). When calibrating our analysis to the amount of microstructure noise that has been reported in the literature, we demonstrate how the optimal sampling interval should be determined: for instance, depending upon the amount of microstructure noise relative to the variance of the underlying returns, the optimal sampling frequency varies from 4 minutes to 3 hours, if 1 day's worth of data is used at a time. If a longer time period is used in the analysis, then the optimal sampling frequency can be considerably longer than these values.

But even if one determines the sampling frequency optimally, it remains the case that the empirical researcher is not making use of the full data at his/her disposal. For instance, suppose that we have available transaction records on a liquid stock, traded once every second. Over a typical 6.5 hour day, we therefore start with $23,400$ observations. If one decides to sample once every 5 minutes, then – whether or not this is the optimal sampling frequency – this amounts to retaining only 78 observations. Said differently, one is throwing away 299 out of every 300 transactions. From a statistical perspective, this is unlikely to be the optimal solution, even though it is undoubtedly better than computing a volatility estimate using noisy squared log-returns sampled every second. Somehow, an optimal solution should make use of all the data, and this is where our analysis goes next.

So, if one decides to account for the presence of the noise, how should one go about doing it? We show that modelling the noise term explicitly restores the first order statistical effect that sampling as often as possible is optimal. This will involve an estimator different from the simple sum of squared log-returns. Since we work within a fully *parametric* framework, *likelihood* is the key word. Hence we construct the likelihood function for the observed log-returns, which include microstructure noise. To do so, we must postulate a model for the noise term. We assume that the noise is Gaussian. In light of what we know from the sophisticated theoretical microstructure literature, this is likely to be overly simplistic and one may well be concerned about the effect(s) of this assumption. Could it do more harm than good? Surprisingly, we demonstrate that our likelihood correction, based on Gaussianity of the noise, works *even if* one misspecifies the assumed distribution of the noise term. Specifically, if the econometrician assumes that the noise terms are normally distributed when in fact they are not, not only is it still optimal to sample as often as possible (unlike the result when no allowance is made for the presence of noise), but the estimator has the *same variance* as if the noise distribution had been correctly specified. This robustness result is, we think, a major argument in favor of incorporating the presence of the noise when estimating continuous time models with high frequency financial data, even if one is unsure about what is the true distribution of the noise term.

In other words, the answer to the question we pose in our title is "as often as possible", provided one accounts for the presence of the noise when designing the estimator (and we suggest maximum likelihood as a means of doing so). If one is unwilling to account for the noise, then the answer is to rely on the finite optimal sampling frequency we start our analysis with, but we stress that while it is optimal if one insists upon using sums of squares of log-returns, this is not the best possible approach to estimate volatility given the complete high frequency dataset at hand.

In a companion paper ([43]), we study the corresponding *nonparametric* problem, where the volatility of the underlying price is a stochastic process, and nothing else is known about it, in particular no parametric structure. In that case, the object of interest is the integrated volatility of the process over a fixed time interval, such as a day, and we show how to estimate it using again all the data available (instead of sparse sampling at an arbitrarily lower frequency of, say, 5 minutes). Since the model is nonparametric, we no longer use a likelihood approach but instead propose a solution based on subsampling and averaging, which involves estimators constructed on two different time scales, and demonstrate that this again dominates sampling at a lower frequency, whether arbitrary or optimally determined.

This paper is organized as follows. We start by describing in Section 1.1 our reduced form setup and the underlying structural models that support it. We then review in Section 1.2 the base case where no noise is present, before analyzing in Section 1.3 the situation where the presence of the noise is ignored. In Section 1.4, we examine the concrete implications of this result for empirical work with high frequency data. Next, we show in Section 1.5 that accounting for the presence of the noise through the likelihood restores the optimality of high frequency sampling. Our robustness results are presented in Section 1.6 and interpreted in Section 1.7. We study the same questions when the observations are sampled at random time intervals, which are an essential feature of transaction-level data, in Section 1.8. We then turn to various extensions and relaxation of our assumptions in Section 1.9: we add a drift term, then serially correlated and cross-correlated noise respectively. Section 1.10 concludes. All proofs are in the Appendix.

1.1 Setup

Our basic setup is as follows. We assume that the underlying process of interest, typically the log-price of a security, is a time-homogenous diffusion on the real line

$$dX_t = \mu(X_t; \theta)dt + \sigma dW_t , \tag{1.1}$$

where $X_0 = 0$, W_t is a Brownian motion, $\mu(.,.)$ is the drift function, σ^2 the diffusion coefficient and θ the drift parameters, $\theta \in \Theta$ and $\sigma > 0$. The parameter space is an open and bounded set. As usual, the restriction that σ is constant is without loss of generality since in the univariate case a one-to-one transformation can always reduce a known specification $\sigma(X_t)$ to that case. Also, as discussed in [4], the properties of parametric estimators in this model are quite different depending upon whether we estimate θ alone, σ^2 alone, or both parameters together. When the data are noisy, the main effects that we describe are already present in the simpler of these three cases, where σ^2 alone is estimated, and so we focus on that case. Moreover, in the high frequency context we have in mind, the diffusive component of (1.1) is of order $(dt)^{1/2}$ while the drift component is of order dt only, so the drift component is mathematically negligible at high frequencies. This is validated empirically: including a drift actually deteriorates the performance of variance estimates from high frequency data since the drift is estimated with a large standard error. Not centering the log returns for the purpose of variance estimation produces more accurate results (see [38]). So we simplify the analysis one step further by setting $\mu = 0$, which we do until Section 1.9.1, where we then show that adding a drift term does not alter our results. In Section 1.9.4, we discuss the situation where the instantaneous volatility σ is stochastic.

But for now,

$$X_t = \sigma W_t \,. \tag{1.2}$$

Until Section 1.8, we treat the case where the observations occur at equidistant time intervals Δ, in which case the parameter σ^2 is therefore estimated at time T on the basis of $N+1$ discrete observations recorded at times $\tau_0 = 0$, $\tau_1 = \Delta$,..., $\tau_N = N\Delta = T$. In Section 1.8, we let the sampling intervals be themselves random variables, since this feature is an essential characteristic of high frequency transaction data.

The notion that the observed transaction price in high frequency financial data is the unobservable efficient price plus some noise component due to the imperfections of the trading process is a well established concept in the market microstructure literature (see for instance [10]). So, where we depart from the inference setup previously studied in [4] is that we now assume that, instead of observing the process X at dates τ_i, we observe X with error:

$$\tilde{X}_{\tau_i} = X_{\tau_i} + U_{\tau_i} \,, \tag{1.3}$$

where the $U_{\tau_i}'s$ are i.i.d. noise with mean zero and variance a^2 and are independent of the W process. In that context, we view X as the efficient log-price, while the observed $\tilde{X}$ is the transaction log-price. In an efficient market, X_t is the log of the expectation of the final value of the security conditional on all publicly available information at time t. It corresponds to the log-price that would be in effect in a perfect market with no trading imperfections, frictions, or informational effects. The Brownian motion W is the process representing the arrival of new information, which in this idealized setting is immediately impounded in X.

By contrast, U_t summarizes the noise generated by the mechanics of the trading process. What we have in mind as the source of noise is a diverse array of market microstructure effects, either information or non-information related, such as the presence of a bid-ask spread and the corresponding bounces, the differences in trade sizes and the corresponding differences in representativeness of the prices, the different informational content of price changes due to informational asymmetries of traders, the gradual response of prices to a block trade, the strategic component of the order flow, inventory control effects, the discreteness of price changes in markets that are not decimalized, etc., all summarized into the term U. That these phenomena are real are important is an accepted fact in the market microstructure literature, both theoretical and empirical. One can in fact argue that these phenomena justify this literature.

We view (1.3) as the simplest possible reduced form of structural market microstructure models. The efficient price process X is typically modelled as a random walk, i.e., the discrete time equivalent of (1.2). Our specification coincides with that of [29], who discusses the theoretical market microstructure underpinnings of such a model and argues that the parameter a is a summary measure of market quality. Structural market microstructure models do generate (1.3). For instance, [39] proposes a model where U is due entirely to the bid-ask spread. [28] notes that in practice there are sources of noise other than just the bid-ask spread, and studies their effect on the Roll model and its estimators.

Indeed, a disturbance U can also be generated by adverse selection effects as in [20] and [21], where the spread has two components: one that is due to monopoly power, clearing costs, inventory carrying costs, etc., as previously, and a second one that arises because of adverse selection whereby the specialist is concerned that the investor on the other side of the transaction has superior information. When asymmetric information is involved, the disturbance U would typically no longer be uncorrelated with the W process and would exhibit autocorrelation at the first order, which would complicate our analysis without fundamentally altering it: see Sections 1.9.2 and 1.9.3 where we relax the assumptions that the $U's$ are serially uncorrelated and independent of the W process, respectively.

The situation where the measurement error is primarily due to the fact that transaction prices are multiples of a tick size (i.e., $\tilde{X}_{\tau_i} = m_i\kappa$ where κ is the tick size and m_i is the integer closest to X_{τ_i}/κ) can be modelled as a rounding off problem (see [14], [23] and [31]). The specification of the model in [27] combines both the rounding and bid-ask effects as the dual sources of the noise term U. Finally, structural models, such as that of [35], also give rise to reduced forms where the observed transaction price $\tilde{X}$ takes the form of an unobserved fundamental value plus error.

With (1.3) as our basic data generating process, we now turn to the questions we address in this paper: how often should one sample a continuous-time process when the data are subject to market microstructure noise, what are the implications of the noise for the estimation of the parameters of the X process, and how should one correct for the presence of the noise, allowing for the possibility that the econometrician misspecifies the assumed distribution of the noise term, and finally allowing for the sampling to occur at random points in time? We proceed from the simplest to the most complex situation by adding one extra layer of complexity at a time: Figure 1.1 shows the three sampling schemes we consider, starting with fixed sampling without market microstructure noise, then moving to fixed sampling with noise and concluding with an analysis of the situation where transaction prices are not only subject to microstructure noise but are also recorded at random time intervals.

1.2 The Baseline Case: No Microstructure Noise

We start by briefly reviewing what would happen in the absence of market microstructure noise, that is when $a = 0$. With X denoting the log-price, the first differences of the observations are the log-returns $Y_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}}$, $i = 1, ..., N$. The observations $Y_i = \sigma\left(W_{\tau_{i+1}} - W_{\tau_i}\right)$ are then i.i.d. $N(0, \sigma^2\Delta)$ so the likelihood function is

$$l(\sigma^2) = -N\ln(2\pi\sigma^2\Delta)/2 - (2\sigma^2\Delta)^{-1}Y'Y\,, \tag{1.4}$$

where $Y = (Y_1, ..., Y_N)'$.. The maximum-likelihood estimator of σ^2 coincides with the discrete approximation to the quadratic variation of the process

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{N} Y_i^2 \tag{1.5}$$

which has the following exact small sample moments:

$$E\left[\hat{\sigma}^2\right] = \frac{1}{T} \sum_{i=1}^{N} E\left[Y_i^2\right] = \frac{N\left(\sigma^2 \Delta\right)}{T} = \sigma^2,$$

$$\mathrm{Var}\left[\hat{\sigma}^2\right] = \frac{1}{T^2} \mathrm{Var}\left[\sum_{i=1}^{N} Y_i^2\right] = \frac{1}{T^2}\left(\sum_{i=1}^{N} \mathrm{Var}\left[Y_i^2\right]\right) = \frac{N}{T^2}\left(2\sigma^4\Delta^2\right) = \frac{2\sigma^4\Delta}{T}$$

and the following asymptotic distribution

$$T^{1/2}\left(\hat{\sigma}^2 - \sigma^2\right) \underset{T \longrightarrow \infty}{\longrightarrow} N(0, \omega), \tag{1.6}$$

where

$$\omega = \mathrm{AVAR}(\hat{\sigma}^2) = \Delta E\left[-\ddot{l}(\sigma^2)\right]^{-1} = 2\sigma^4\Delta. \tag{1.7}$$

Thus selecting Δ as small as possible is optimal for the purpose of estimating σ^2.

1.3 When the Observations Are Noisy But the Noise Is Ignored

Suppose now that market microstructure noise is present but the presence of the $U's$ is ignored when estimating σ^2. In other words, we use the log-likelihood (1.4) even though the true structure of the observed log-returns $Y_i's$ is given by an MA(1) process since

$$\begin{aligned} Y_i &= \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}} \\ &= X_{\tau_i} - X_{\tau_{i-1}} + U_{\tau_i} - U_{\tau_{i-1}} \\ &= \sigma\left(W_{\tau_i} - W_{\tau_{i-1}}\right) + U_{\tau_i} - U_{\tau_{i-1}} \\ &\equiv \varepsilon_i + \eta\varepsilon_{i-1}, \end{aligned} \tag{1.8}$$

where the $\varepsilon_i's$ are uncorrelated with mean zero and variance γ^2 (if the $U's$ are normally distributed, then the $\varepsilon_i's$ are i.i.d.). The relationship to the original parametrization (σ^2, a^2) is given by

$$\gamma^2(1+\eta^2) = \mathrm{Var}[Y_i] = \sigma^2\Delta + 2a^2 \tag{1.9}$$

$$\gamma^2\eta = \mathrm{Cov}(Y_i, Y_{i-1}) = -a^2. \tag{1.10}$$

Equivalently, the inverse change of variable is given by

$$\gamma^2 = \frac{1}{2}\left\{2a^2 + \sigma^2\Delta + \sqrt{\sigma^2\Delta\left(4a^2 + \sigma^2\Delta\right)}\right\} \tag{1.11}$$

$$\eta = \frac{1}{2a^2}\left\{-2a^2 - \sigma^2\Delta + \sqrt{\sigma^2\Delta\left(4a^2 + \sigma^2\Delta\right)}\right\}. \tag{1.12}$$

Two important properties of the log-returns Y_i's emerge from the two equations (1.9)-(1.10). First, it is clear from (1.9) that microstructure noise leads to spurious variance in observed log-returns, $\sigma^2\Delta + 2a^2$ vs. $\sigma^2\Delta$. This is consistent with the predictions of theoretical microstructure models. For instance, [16] develop a model linking the arrival of information, the timing of trades, and the resulting price process. In their model, the transaction price will be a biased representation of the efficient price process, with a variance that is both overstated and heteroskedastic as a result of the fact that transactions (hence the recording of an observation on the process $\tilde{X}$) occur at intervals that are time-varying. While our specification is too simple to capture the rich joint dynamics of price and sampling times predicted by their model, heteroskedasticity of the observed variance will also arise in our case once we allow for time variation of the sampling intervals (see Section 1.8 below).

In our model, the proportion of the total return variance that is market microstructure-induced is

$$\pi = \frac{2a^2}{\sigma^2\Delta + 2a^2} \tag{1.13}$$

at observation interval Δ. As Δ gets smaller, π gets closer to 1, so that a larger proportion of the variance in the observed log-return is driven by market microstructure frictions, and correspondingly a lesser fraction reflects the volatility of the underlying price process X.

Second, (1.10) implies that $-1 < \eta < 0$, so that log-returns are (negatively) autocorrelated with first order autocorrelation $-a^2/(\sigma^2\Delta + 2a^2) = -\pi/2$. It has been noted that market microstructure noise has the potential to explain the empirical autocorrelation of returns. For instance, in the simple Roll model, $U_t = (s/2)Q_t$ where s is the bid/ask spread and Q_t, the order flow indicator, is a binomial variable that takes the values $+1$ and -1 with equal probability. Therefore $\mathrm{Var}[U_t] = a^2 = s^2/4$. Since $\mathrm{Cov}(Y_i, Y_{i-1}) = -a^2$, the bid/ask spread can be recovered in this model as $s = 2\sqrt{-\rho}$ where $\rho = \gamma^2\eta$ is the first order autocorrelation of returns. [18] proposed to adjust variance estimates to control for such autocorrelation and [28] studied the resulting estimators. In [41], U arises because of the strategic trading of institutional investors which is then put forward as an explanation for the observed serial correlation of returns. [33] show that infrequent trading has implications for the variance and autocorrelations of returns. Other empirical patterns in high frequency financial data have been documented: leptokurtosis, deterministic patterns and volatility clustering.

Our first result shows that the optimal sampling frequency is finite when noise is present but unaccounted for. The estimator $\hat{\sigma}^2$ obtained from maximizing the misspecified log-likelihood (1.4) is quadratic in the $Y_i's$: see (1.5). In order to obtain its exact (i.e., small sample) variance, we therefore need to calculate the fourth order cumulants of the $Y_i's$ since

$$\mathrm{Cov}\left(Y_i^2, Y_j^2\right) = 2\,\mathrm{Cov}\left(Y_i, Y_j\right)^2 + \mathrm{Cum}(Y_i, Y_i, Y_j, Y_j) \tag{1.14}$$

(see e.g., Section 2.3 of [36] for definitions and properties of the cumulants). We have:

Lemma 1. *The fourth cumulants of the log-returns are given by*

$$\mathrm{Cum}(Y_i, Y_j, Y_k, Y_l) = \begin{cases} 2\ \mathrm{Cum}_4\left[U\right] & \text{if } i = j = k = l\,, \\ (-1)^{s(i,j,k,l)}\,\mathrm{Cum}_4\left[U\right], & \text{if } \max(i,j,k,l) = \min(i,j,k,l) + 1, \\ 0 & \text{otherwise}\,, \end{cases} \tag{1.15}$$

where $s(i,j,k,l)$ denotes the number of indices among (i,j,k,l) that are equal to $\min(i,j,k,l)$ and U denotes a generic random variable with the common distribution of the $U_{\tau_i}'s$. Its fourth cumulant is denoted $\mathrm{Cum}_4\left[U\right]$.

Now U has mean zero, so in terms of its moments

$$\mathrm{Cum}_4\left[U\right] = E\left[U^4\right] - 3\left(E\left[U^2\right]\right)^2. \tag{1.16}$$

In the special case where U is normally distributed, $\mathrm{Cum}_4\left[U\right] = 0$ and as a result of (1.14) the fourth cumulants of the log-returns are all 0 (since W is normal, the log-returns are also normal in that case). If the distribution of U is binomial as in the simple bid/ask model described above, then $\mathrm{Cum}_4\left[U\right] = -s^4/8$; since in general s will be a tiny percentage of the asset price, say $s = 0.05\%$, the resulting $\mathrm{Cum}_4\left[U\right]$ will be very small.

We can now characterize the root mean squared error

$$RMSE\left[\hat{\sigma}^2\right] = \left(\left(E\left[\hat{\sigma}^2\right] - \sigma^2\right)^2 + \mathrm{Var}\left[\hat{\sigma}^2\right]\right)^{1/2}$$

of the estimator:

Theorem 1. *In small samples (finite T), the bias and variance of the estimator $\hat{\sigma}^2$ are given by*

$$E\left[\hat{\sigma}^2\right] - \sigma^2 = \frac{2a^2}{\Delta}\,, \tag{1.17}$$

$$\operatorname{Var}\left[\hat{\sigma}^2\right] = \frac{2\left(\sigma^4 \Delta^2 + 4\sigma^2 \Delta a^2 + 6a^4 + 2\operatorname{Cum}_4[U]\right)}{T\Delta} - \frac{2\left(2a^4 + \operatorname{Cum}_4[U]\right)}{T^2}. \tag{1.18}$$

Its RMSE has a unique minimum in Δ which is reached at the optimal sampling interval

$$\Delta^* = \left(\frac{2a^4 T}{\sigma^4}\right)^{1/3} \left(\left(1 - \sqrt{1 - \frac{2\left(3a^4 + \operatorname{Cum}_4[U]\right)^3}{27\sigma^4 a^8 T^2}}\right)^{1/3} + \left(1 + \sqrt{1 - \frac{2\left(3a^4 + \operatorname{Cum}_4[U]\right)^3}{27\sigma^4 a^8 T^2}}\right)^{1/3} \right). \tag{1.19}$$

As T grows, we have

$$\Delta^* = \frac{2^{2/3} a^{4/3}}{\sigma^{4/3}} T^{1/3} + O\left(\frac{1}{T^{1/3}}\right). \tag{1.20}$$

The trade-off between bias and variance made explicit in (1.17)-(1.19) is not unlike the situation in nonparametric estimation with Δ^{-1} playing the role of the bandwidth h. A lower h reduces the bias but increases the variance, and the optimal choice of h balances the two effects.

Note that these are exact small sample expressions, valid for all T. Asymptotically in T, $\operatorname{Var}\left[\hat{\sigma}^2\right] \longrightarrow 0$, and hence the RMSE of the estimator is dominated by the bias term which is independent of T. And given the form of the bias (1.17), one would in fact want to select the largest Δ possible to minimize the bias (as opposed to the smallest one as in the no-noise case of Section 1.2). The rate at which Δ^* should increase with T is given by (1.20). Also, in the limit where the noise disappears ($a \longrightarrow 0$ and $\operatorname{Cum}_4[U] \longrightarrow 0$), the optimal sampling interval Δ^* tends to 0.

How does a small departure from a normal distribution of the microstructure noise affect the optimal sampling frequency? The answer is that a small positive (resp. negative) departure of $\operatorname{Cum} 4[U]$ starting from the normal value of 0 leads to an increase (resp. decrease) in Δ^*, since

$$\Delta^* = \Delta^*_{\text{normal}} + \frac{\left(\left(1 + \sqrt{1 - \frac{2a^4}{T^2\sigma^4}}\right)^{2/3} - \left(1 - \sqrt{1 - \frac{2a^4}{T^2\sigma^4}}\right)^{2/3}\right)}{3\, 2^{1/3} a^{4/3} T^{1/3} \sqrt{1 - \frac{2a^4}{T^2\sigma^4}}\, \sigma^{8/3}} \operatorname{Cum} 4[U] + O\left(\operatorname{Cum} 4[U]^2\right), \tag{1.21}$$

where Δ^*_{normal} is the value of Δ^* corresponding to $\text{Cum}_4[U] = 0$. And of course the full formula (1.20) can be used to get the exact answer for any departure from normality instead of the comparative static one.

Another interesting asymptotic situation occurs if one attempts to use higher and higher frequency data ($\Delta \to 0$, say sampled every minute) over a fixed time period (T fixed, say a day). Since the expressions in Theorem 1 are exact small sample ones, they can in particular be specialized to analyze this situation. With $n = T/\Delta$, it follows from (1.17)-(1.19) that

$$E\left[\hat{\sigma}^2\right] = \frac{2na^2}{T} + o(n) = \frac{2nE\left[U^2\right]}{T} + o(n), \tag{1.22}$$

$$\text{Var}\left[\hat{\sigma}^2\right] = \frac{2n\left(6a^4 + 2\,\text{Cum}_4[U]\right)}{T^2} + o(n) = \frac{4nE\left[U^4\right]}{T^2} + o(n), \tag{1.23}$$

so $(T/2n)\hat{\sigma}^2$ becomes an estimator of $E\left[U^2\right] = a^2$ whose asymptotic variance is $E\left[U^4\right]$. Note in particular that $\hat{\sigma}^2$ estimates the variance of the noise, which is essentially unrelated to the object of interest σ^2. This type of asymptotics is relevant in the stochastic volatility case we analyze in our companion paper [43].

Our results also have implications for the two parallel tracks that have developed in the recent financial econometrics literature dealing with discretely observed continuous-time processes. One strand of the literature has argued that estimation methods should be robust to the potential issues arising in the presence of high frequency data and, consequently, be asymptotically valid without requiring that the sampling interval Δ separating successive observations tend to zero (see, e.g., [2], [3] and [26]). Another strand of the literature has dispensed with that constraint, and the asymptotic validity of these methods requires that Δ tend to zero instead of or in addition to, an increasing length of time T over which these observations are recorded (see, e.g., [6], [7] and [8]).

The first strand of literature has been informally warning about the potential dangers of using high frequency financial data without accounting for their inherent noise (see e.g., page 529 of [2]), and we propose a formal modelization of that phenomenon. The implications of our analysis are most salient for the second strand of the literature, which is predicated on the use of high frequency data but does not account for the presence of market microstructure noise. Our results show that the properties of estimators based on the local sample path properties of the process (such as the quadratic variation to estimate σ^2) change dramatically in the presence of noise. Complementary to this are the results of [22] which show that the presence of even increasingly negligible noise is sufficient to adversely affect the identification of σ^2.

1.4 Concrete Implications for Empirical Work with High Frequency Data

The clear message of Theorem 1 for empirical researchers working with high frequency financial data is that it may be optimal to sample less frequently. As discussed in the Introduction, authors have reduced their sampling frequency below that of the actual record of observations in a somewhat ad hoc fashion, with typical choices 5 minutes and up. Our analysis provides not only a theoretical rationale for sampling less frequently, but also delivers a precise answer to the question of "how often one should sample?" For that purpose, we need to calibrate the parameters appearing in Theorem 1, namely σ, α, $\mathrm{Cum}_4[U]$, Δ and T. We assume in this calibration exercise that the noise is Gaussian, in which case $\mathrm{Cum}_4[U] = 0$.

1.4.1 Stocks

We use existing studies in empirical market microstructure to calibrate the parameters. One such study is [35], who estimated on the basis of a sample of 274 NYSE stocks that approximately 60% of the total variance of price changes is attributable to market microstructure effects (they report a range of values for π from 54% in the first half hour of trading to 65% in the last half hour, see their Table 4; they also decompose this total variance into components due to discreteness, asymmetric information, transaction costs and the interaction between these effects). Given that their sample contains an average of 15 transactions per hour (their Table 1), we have in our framework

$$\pi = 60\%, \ \Delta = 1/(15 \times 7 \times 252)\,. \tag{1.24}$$

These values imply from (1.13) that $a = 0.16\%$ if we assume a realistic value of $\sigma = 30\%$ per year. (We do not use their reported volatility number since they apparently averaged the variance of price changes over the 274 stocks instead of the variance of the returns. Since different stocks have different price levels, the price variances across stocks are not directly comparable. This does not affect the estimated fraction π however, since the price level scaling factor cancels out between the numerator and the denominator).

The magnitude of the effect is bound to vary by type of security, market and time period. [29] estimates the value of a to be 0.33%. Some authors have reported even larger effects. Using a sample of NASDAQ stocks, [32] estimate that about 50% of the daily variance of returns in due to the bid-ask effect. With $\sigma = 40\%$ (NASDAQ stocks have higher volatility), the values

$$\pi = 50\%, \ \Delta = 1/252$$

yield the value $a = 1.8\%$. Also on NASDAQ, [12] estimate that 11% of the variance of weekly returns (see their Table 4, middle portfolio) is due to bid-ask effects. The values

$$\pi = 11\%,\ \Delta = 1/52$$

imply that $a = 1.4\%$.

In Table 1.1, we compute the value of the optimal sampling interval Δ^* implied by different combinations of sample length (T) and noise magnitude (a). The volatility of the efficient price process is held fixed at $\sigma = 30\%$ in Panel A, which is a realistic value for stocks. The numbers in the table show that the optimal sampling frequency can be substantially affected by even relatively small quantities of microstructure noise. For instance, using the value $a = 0.15\%$ calibrated from [35], we find an optimal sampling interval of 22 minutes if the sampling length is 1 day; longer sample lengths lead to higher optimal sampling intervals. With the higher value of $a = 0.3\%$, approximating the estimate from [29], the optimal sampling interval is 57 minutes. A lower value of the magnitude of the noise translates into a higher frequency: for instance, $\Delta^* = 5$ minutes if $a = 0.05\%$ and $T = 1$ day. Figure 1.2 displays the RMSE of the estimator as a function of Δ and T, using parameter values $\sigma = 30\%$ and $a = 0.15\%$. The figure illustrates the fact that deviations from the optimal choice of Δ lead to a substantial increase in the RMSE: for example, with $T = 1$ month, the RMSE more than doubles if, instead of the optimal $\Delta^* = 1$ hour, one uses $\Delta = 15$ minutes.

1.4.2 Currencies

Looking now at foreign exchange markets, empirical market microstructure studies have quantified the magnitude of the bid-ask spread. For example, [9] computes the average bid/ask spread s in the wholesale market for different currencies and reports values of $s = 0.05\%$ for the German mark, and 0.06% for the Japanese yen (see Panel B of his Table 2). We calculated the corresponding numbers for the 1996-2002 period to be 0.04% for the mark (followed by the euro) and 0.06% for the yen. Emerging market currencies have higher spreads: for instance, $s = 0.12\%$ for Korea and 0.10% for Brazil. During the same period, the volatility of the exchange rate was $\sigma = 10\%$ for the German mark, 12% for the Japanese yen, 17% for Brazil and 18% for Korea. In Panel B of Table 1.1, we compute Δ^* with $\sigma = 10\%$, a realistic value for the euro and yen. As we noted above, if the sole source of the noise were a bid/ask spread of size s, then a should be set to $s/2$. Therefore Panel B reports the values of Δ^* for values of a ranging from 0.02% to 0.1%. For example, the dollar/euro or dollar/yen exchange rates (calibrated to $\sigma = 10\%$, $a = 0.02\%$) should be sampled every $\Delta^* = 23$ minutes if the overall sample length is $T = 1$ day, and every 1.1 hours if $T = 1$ year.

Furthermore, using the bid/ask spread alone as a proxy for all microstructure frictions will lead, except in unusual circumstances, to an understatement of the parameter a, since variances are additive. Thus, since Δ^* is increasing in a, one should interpret the value of Δ^* read off 1.1 on the row corresponding to $a = s/2$ as a lower bound for the optimal sampling interval.

1.4.3 Monte Carlo Evidence

To validate empirically these results, we perform Monte Carlo simulations. We simulate $M = 10,000$ samples of length $T = 1$ year of the process X, add microstructure noise U to generate the observations $\tilde{X}$ and then the log returns Y. We sample the log-returns at various intervals Δ ranging from 5 minutes to 1 week and calculate the bias and variance of the estimator $\hat{\sigma}^2$ over the M simulated paths. We then compare the results to the theoretical values given in (1.17)-(1.19) of Theorem 1. The noise distribution is Gaussian, $\sigma = 30\%$ and $a = 0.15\%$ – the values we calibrated to stock returns data above. Table 1.2 shows that the theoretical values are in close agreement with the results of the Monte Carlo simulations.

The table also illustrates the magnitude of the bias inherent in sampling at too high a frequency. While the value of σ^2 used to generate the data is 0.09, the expected value of the estimator when sampling every 5 minutes is 0.18, so on average the estimated quadratic variation is twice as big as it should be in this case.

1.5 Incorporating Market Microstructure Noise Explicitly

So far we have stuck to the sum of squares of log-returns as our estimator of volatility. We then showed that, for this estimator, the optimal sampling frequency is finite. But this implies that one is discarding a large proportion of the high frequency sample (299 out of every 300 observations in the example described in the Introduction), in order to mitigate the bias induced by market microstructure noise. Next, we show that if we explicitly incorporate the $U's$ into the likelihood function, then we are back in the situation where the optimal sampling scheme consists in sampling as often as possible – i.e., using all the data available.

Specifying the likelihood function of the log-returns, while recognizing that they incorporate noise, requires that we take a stand on the distribution of the noise term. Suppose for now that the microstructure noise is normally distributed, an assumption whose effect we will investigate below in Section 1.6. Under this assumption, the likelihood function for the $Y's$ is given by

$$l(\eta, \gamma^2) = -\ln\det(V)/2 - N\ln(2\pi\gamma^2)/2 - (2\gamma^2)^{-1}Y'V^{-1}Y\,, \qquad (1.25)$$

where the covariance matrix for the vector $Y = (Y_1, ..., Y_N)'$ is given by $\gamma^2 V$, where

$$V = [v_{ij}]_{i,j=1,\dots,N} = \begin{pmatrix} 1+\eta^2 & \eta & 0 & \cdots & 0 \\ \eta & 1+\eta^2 & \eta & \ddots & \vdots \\ 0 & \eta & 1+\eta^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \eta \\ 0 & \cdots & 0 & \eta & 1+\eta^2 \end{pmatrix}. \tag{1.26}$$

Further,

$$\det(V) = \frac{1-\eta^{2N+2}}{1-\eta^2}, \tag{1.27}$$

and, neglecting the end effects, an approximate inverse of V is the matrix $\Omega = [\omega_{ij}]_{i,j=1,\dots,N}$ where

$$\omega_{ij} = \left(1-\eta^2\right)^{-1}(-\eta)^{|i-j|}$$

(see [15]). The product $V\Omega$ differs from the identity matrix only on the first and last rows. The exact inverse is $V^{-1} = \left[v^{ij}\right]_{i,j=1,\dots,N}$ where

$$\begin{aligned} v^{ij} = {} & \left(1-\eta^2\right)^{-1}\left(1-\eta^{2N+2}\right)^{-1}\Big\{(-\eta)^{|i-j|} - (-\eta)^{i+j} - (-\eta)^{2N-i-j+2} - \\ & - (-\eta)^{2N+|i-j|+2} + (-\eta)^{2N+i-j+2} + (-\eta)^{2N-i+j+2}\Big\} \end{aligned} \tag{1.28}$$

(see [24] and [40]).

From the perspective of practical implementation, this estimator is nothing else than the MLE estimator of an MA(1) process with Gaussian errors: any existing computer routines for the MA(1) situation can therefore be applied (see e.g., Section 5.4 in [25]). In particular, the likelihood function can be expressed in a computationally efficient form by triangularizing the matrix V, yielding the equivalent expression:

$$l(\eta,\gamma^2) = -\frac{1}{2}\sum_{i=1}^{N}\ln(2\pi d_i) - \frac{1}{2}\sum_{i=1}^{N}\frac{\tilde{Y}_i^2}{d_i}, \tag{1.29}$$

where

$$d_i = \gamma^2\frac{1+\eta^2+\dots+\eta^{2i}}{1+\eta^2+\dots+\eta^{2(i-1)}},$$

and the $\tilde{Y}_i's$ are obtained recursively as $\tilde{Y}_1 = Y_1$ and for $i = 2,\dots,N$:

$$\tilde{Y}_i = Y_i - \frac{\eta\left(1+\eta^2+\dots+\eta^{2(i-2)}\right)}{1+\eta^2+\dots+\eta^{2(i-1)}}\tilde{Y}_{i-1}.$$

This latter form of the log-likelihood function involves only single sums as opposed to double sums if one were to compute $Y'V^{-1}Y$ by brute force using the expression of V^{-1} given above.

We now compute the distribution of the MLE estimators of σ^2 and a^2, which follows by the delta method from the classical result for the MA(1) estimators of γ and η :

Proposition 1. *When U is normally distributed, the MLE $(\hat{\sigma}^2, \hat{a}^2)$ is consistent and its asymptotic variance is given by*

$$\text{AVAR}_{normal}(\hat{\sigma}^2, \hat{a}^2) = \begin{pmatrix} 4\sqrt{\sigma^6 \Delta \left(4a^2 + \sigma^2 \Delta\right)} + 2\sigma^4 \Delta & -\sigma^2 \Delta h(\Delta, \sigma^2, a^2) \\ \bullet & \frac{\Delta}{2}\left(2a^2 + \sigma^2 \Delta\right) h(\Delta, \sigma^2, a^2) \end{pmatrix}$$

with

$$h(\Delta, \sigma^2, a^2) \equiv 2a^2 + \sqrt{\sigma^2 \Delta \left(4a^2 + \sigma^2 \Delta\right)} + \sigma^2 \Delta \,. \tag{1.30}$$

Since $\text{AVAR}_{\text{normal}}(\hat{\sigma}^2)$ is increasing in Δ, it is optimal to sample as often as possible. Further, since

$$\text{AVAR}_{\text{normal}}(\hat{\sigma}^2) = 8\sigma^3 a \Delta^{1/2} + 2\sigma^4 \Delta + o(\Delta) \,, \tag{1.31}$$

the loss of efficiency relative to the case where no market microstructure noise is present (and $\text{AVAR}(\hat{\sigma}^2) = 2\sigma^4 \Delta$ as given in (1.7) if $a^2 = 0$ is not estimated, or $\text{AVAR}(\hat{\sigma}^2) = 6\sigma^4 \Delta$ if $a^2 = 0$ is estimated) is at order $\Delta^{1/2}$. Figure 1.3 plots the asymptotic variances of $\hat{\sigma}^2$ as functions of Δ with and without noise (the parameter values are again $\sigma = 30\%$ and $a = 0.15\%$). Figure 1.4 reports histograms of the distributions of $\hat{\sigma}^2$ and $\hat{a}^2$ from 10,000 Monte Carlo simulations with the solid curve plotting the asymptotic distribution of the estimator from Proposition 1. The sample path is of length $T = 1$ year, the parameter values the same as above, and the process is sampled every 5 minutes – since we are now accounting explicitly for the presence of noise, there is no longer a reason to sample at lower frequencies. Indeed, the figure documents the absence of bias and the good agreement of the asymptotic distribution with the small sample one.

1.6 The Effect of Misspecifying the Distribution of the Microstructure Noise

We now study the situation where one attempts to incorporate the presence of the $U's$ into the analysis, as in Section 1.5, but mistakenly assumes a misspecified model for them. Specifically, we consider the case where the $U's$ are assumed to be normally distributed when in reality they have a different distribution. We still suppose that the $U's$ are i.i.d. with mean zero and variance a^2.

Since the econometrician assumes the $U's$ to have a normal distribution, inference is still done with the log-likelihood $l(\sigma^2, a^2)$, or equivalently $l(\eta, \gamma^2)$

given in (1.25), using (1.9)-(1.10). This means that the scores $\dot{l}_{\sigma^2}$ and $\dot{l}_{a^2}$, or equivalently (C.1) and (C.2), are used as moment functions (or "estimating equations"). Since the first order moments of the moment functions only depend on the second order moment structure of the log-returns $(Y_1, ..., Y_N)$, which is unchanged by the absence of normality, the moment functions are unbiased under the true distribution of the $U's$:

$$E_{\text{true}}[\dot{l}_\eta] = E_{\text{true}}[\dot{l}_{\gamma^2}] = 0\,, \tag{1.32}$$

and similarly for $\dot{l}_{\sigma^2}$ and $\dot{l}_{a^2}$. Hence the estimator $(\hat{\sigma}^2, \hat{a}^2)$ based on these moment functions is consistent and asymptotically unbiased (even though the likelihood function is misspecified.)

The effect of misspecification therefore lies in the asymptotic variance matrix. By using the cumulants of the distribution of U, we express the asymptotic variance of these estimators in terms of deviations from normality. But as far as computing the actual estimator, nothing has changed relative to Section 1.5: we are still calculating the MLE for an MA(1) process with Gaussian errors and can apply exactly the same computational routine.

However, since the error distribution is potentially misspecified, one could expect the asymptotic distribution of the estimator to be altered. This turns out not be the case, as far as $\hat{\sigma}^2$ is concerned:

Theorem 2. *The estimators $(\hat{\sigma}^2, \hat{a}^2)$ obtained by maximizing the possibly misspecified log-likelihood (1.25) are consistent and their asymptotic variance is given by*

$$\text{AVAR}_{true}(\hat{\sigma}^2, \hat{a}^2) = \text{AVAR}_{normal}(\hat{\sigma}^2, \hat{a}^2) + \text{Cum}_4\left[U\right] \begin{pmatrix} 0 & 0 \\ 0 & \Delta \end{pmatrix}, \tag{1.33}$$

where $\text{AVAR}_{normal}(\hat{\sigma}^2, \hat{a}^2)$ is the asymptotic variance in the case where the distribution of U is normal, that is, the expression given in Proposition 1.

In other words, the asymptotic variance of $\hat{\sigma}^2$ is identical to its expression if the $U's$ had been normal. Therefore the correction we proposed for the presence of market microstructure noise relying on the assumption that the noise is Gaussian is robust to misspecification of the error distribution.

Documenting the presence of the correction term through simulations presents a challenge. At the parameter values calibrated to be realistic, the order of magnitude of a is a few basis points, say $a = 0.10\% = 10^{-3}$. But if U is of order 10^{-3}, $\text{Cum}_4[U]$ which is of the same order as U^4, is of order 10^{-12}. In other words, with a typical noise distribution, the correction term in (1.33) will not be visible.

To nevertheless make it discernible, we use a distribution for U with the same calibrated standard deviation a as before, but a disproportionately large fourth cumulant. Such a distribution can be constructed by letting $U = \omega T_\nu$