FACULTEIT ECONOMIE
EN BEDRIJFSWETENSCHAPPEN

KATHOLIEKE
UNIVERSITEIT
LEUVEN

FINANCIAL TRANSACTION DATA AND VOLATILITY

Proefschrift voorgedragen
tot het behalen van de graad
van Doctor in de
Economische Wetenschappen

door

Olivier VERGOTE

Nummer 297 2008

Doctoral Committee

Advisor:	Prof. Dr. Geert Dhaene, K.U.Leuven
Committee:	Prof. Dr. Angelos Dassios, London School of Economics and Political Science
	Prof. Dr. Dick van Dijk, Erasmus University Rotterdam
	Prof. Dr. Hans Degryse, K.U.Leuven and Tilburg University
	Prof. Dr. Hans Dewachter, K.U.Leuven and Erasmus University Rotterdam

Acknowledgements

My gratitude first goes out to Geert for providing me this opportunity and guiding me patiently along this project. Our discussions put me on interesting tracks and benefited the thesis greatly. I would also like to thank my doctoral committee for their comments, which allowed me to improve parts of the thesis. I am grateful to Chris Werben for providing me information about NYSE procedures that were useful for Chapter 1.

I would like to thank my parents and family for their unconditional support. My thanks go to Simone for her support and encouragement. Thanks also to Els for her support and understanding. Furthermore, I keep good memories of the office times at Huis van 't Sestich, a refugee house for too many CES visitors to list. Some, however, managed to make it a more permanent base. Thanks Jorre, Katleen, Blanca and Matteo for the unforgettable moments. My gratitude also goes to the Econometrics Group for the welcoming atmosphere, the opportunities to present and discuss my work and the good times - also outside work. Thanks Anton, Bram, Catherine, Dirk, Famke, Frank, Frans, Helena, Koen, László, Luc, Peter and Wim. I also keep good memories of the lunch group with all too lively discussions spicing up the Alma food and picnicking in the park. Special thanks to the petanque team for all the fun and challenging le Patron. Thanks to the secretariat for your smiles and support for this PhD programme. Thanks to all the others who made the corridors and Leuven so much more colourful. Thanks Agnieszka, Alex, Annie, Bart, Bert, Cédric, Daniël, Eef, Filip, Filipa, George, Günther, Isabel, Jan, Javier, Julienne, Karla, Karolien, Koen, Kristien, Kristof, Laura, Lydia, Magdalena, Marco, Marijke, Mark, Romain, Kristof, Koen, Leonardo, Marie, Pablo, Paul, Pelin, Tom, Vera, Veerle, Vivien, Wim and Xu.

Thanks to the ECB and NBB people for the priceless experience gained and the good times at and outside work. While unrelated, it helped me to put things into perspective. Thanks to the Kolohania people for marvellous distraction operations across Europe and your honest distrust of anything beyond Master level - Cheers Bristol. The same pretty much holds for the Chiro people, who I also thank for dragging me to Houthulst city often enough.

Contents

General Introduction

Introduction

This Ph.D. thesis focuses on financial transaction data and volatility. I start by explaining the focus on transaction data, followed by the focus on volatility. Transaction data capture the characteristics of financial transactions (e.g. transaction time, transaction price, transaction volume, bid and ask price) as they take place on an exchange. Databases for these transaction data, also called tick data, only became publicly available in the 1990s. Before, financial research and analysis had been mainly based on daily data, i.e. daily averages, closing prices, etc. It soon became clear that this new type of data offered advantages and opened new research opportunities. For example, higher-frequency data allow more accurate measurement of volatility. The thesis contributes to the growing research on this issue. However, more is not always better. The new data have their own features such as unequally-spaced observations, non-synchronous trading, intra-day seasonal effects, measurement errors due to bid-ask spreads, reporting difficulties, etc., which brought new challenges. Only when these features are satisfactorily dealt with can the advantages of high-frequency data be fully exploited. The thesis contributes to the literature that seeks solutions to this type of problems.

A second focus of this thesis is asset market volatility, i.e. the degree to which financial prices tend to fluctuate. Volatility enters as an essential ingredient in many financial computations, like portfolio optimisation, option pricing and risk assessment. Despite its importance, volatility remains an ambiguous term for which there is no unique, universally accepted definition. The main approaches to compute volatilities are by historical indicators computed from daily squared returns, from econometric models such as GARCH, or by indirect computation from option prices based on a pricing model such as Black-Scholes'. Following the introduction of transaction databases, new estimators that exploit intradaily price dynamics have been proposed in the literature. The thesis also presents new estimators along this line. Before the attention switched to measuring volatility, the financial econometrics literature already contained a lot of

research on the modelling of volatility. Since the 1980s, starting from the observation of volatility clustering, i.e. periods of high volatility versus periods of low volatility, many models of volatility have been developed that produce and improve forecasts. To this end, Stochastic Volatility models were developed. These models treat volatility as unobserved, driven by a separate process. Their very nature makes them hard to estimate, however. The thesis points out that a simple estimation method (Generalized Method of Moments, GMM) should be reconsidered to estimate stochastic volatility models.

Outline

While the chapters of this thesis have a common theme, each chapter can be seen as a separate entity addressing different well-defined issues within financial econometrics.

Chapter 1 proposes a new procedure to determine the time of the prevailing quote relative to the time of the trade for New York Stock Exchange (NYSE) data. At the NYSE, trades and quotes are recorded separately, receiving their own time stamp. As a result, trades and quotes are subject to different and varying reporting lags, which makes it hard to reconstruct the sequence of trades and quotes. For market microstructure analysis that is based on trade and quote data at high frequency, it is important to be able to reconstruct this sequence, as mismatching potentially affects the analysis. The procedure put forward in chapter 1 tests whether the quote revision frequency around a trade is contaminated by quote revisions triggered by a trade, and then determines the smallest timing adjustment needed to eliminate this contamination. An application to various stocks and sample periods shows that the time difference between trade and quote reporting lags varies across stocks and time. The procedure takes this variation into account and hence offers a stock- and time-specific update to the Lee and Ready (1991) 5-second rule.

Chapter 2 contributes to the extensive literature on the estimation of stochastic volatility (SV) models. Due to the fact that in SV models the mean and the volatility are driven by separate stochastic processes (implying that volatility is unobservable), SV models are hard to estimate. This chapter presents analytical results that may be used to improve and assess the quality of GMM-based estimation of SV models. GMM, while not asymptotically efficient, is still the simplest estimation method for SV models currently available.

Chapter 3 proposes new estimators of volatility based on quantiles of the price series,

under the assumption that prices are observed without noise. It develops unbiased and consistent estimators of the diffusion coefficient based on quantiles of either the Brownian motion or the Brownian bridge. These estimators are shown to be much more efficient than the range-based estimators of Parkinson (1980) and Kunitomo (1992), where the range is the difference between the supremum and infimum. In particular, efficiency is improved by using more quantiles in the estimation. Moreover, two methods are presented that turn any of the unbiased estimators into consistent estimators. One way to obtain consistency is to apply the unbiased estimators to subintervals and then to average the subinterval estimators. This corresponds to a generalization of the realized range estimator of Christensen and Podolski (2005) and Martens and van Dijk (2007). Furthermore, a new type of consistent estimator based on permuted subintervals is presented. The quantile-based estimators provide an interesting alternative to the existing realized volatility and realized range estimators.

Chapter 4 deals with the time-discreteness bias and noise bias of quantile-based volatility estimators when applied to high-frequency data. The former bias is a result of the estimators being derived in continuous-time, but applied to discrete-time observations. Despite being derived in continuous time, quantile-based volatility estimators turn out to be fairly robust to the time-discreteness bias except if the estimator is based on price extrema or the number of observations is very small. Analytical and simulation-based bias corrections are presented to deal with the latter cases. Furthermore, attention is given to the bias introduced when the estimators are applied to a price series perturbed by noise. In practice, this noise is due to market microstructure effects, e.g. the transaction price bouncing between bid and ask prices, implying that the 'true' price is not observed. A simulation-based noise-bias correction is proposed that deals even with the case in which the noise distribution is unknown. The bias corrections allow the practitioner to exploit the efficiency gain of quantile-based volatility estimation at high frequency.

Chapter 1

How to Match Trades and Quotes for NYSE Stocks

1.1 Introduction

The Trade and Quote (TAQ) database managed by NYSE Euronext is a common source for tick data on NYSE stocks. The extraction programme of the TAQ database produces separate files for trades and quotes, each with its own time stamp. For market microstructure analysis that is based on variables from both sets at high frequency, e.g. analysis of transaction prices versus quoted spreads, one needs to construct the sequence of quotes and trades. The trade classification into buyer or seller initiated trades, computation of the effective spread and estimation of the information content of trades are important examples of such analysis. In principle, the job consists of merging both data sets and ranking their records chronologically. This would be a straightforward operation, were it not that trades and quotes can be subject to different reporting lags, which complicates the identification of the prevailing quotes at the time of a trade. If the matching of trades and quotes is not done appropriately, then this affects certain measures as those mentioned above and can potentially alter the conclusions of microstructure analysis. This problem was first reported by Lee and Ready (1991). The authors found the difference between the lag of trade and quote reports to be on average 5 seconds. Therefore, the solution suggested by these authors is to add five seconds to the reported times of quotes. So far, most studies followed this suggestion or did not adjust at all.

In this chapter it becomes clear, however, that the difference between trade and quote reporting lags has changed over time and can vary between stocks. This implies that the 5-second rule is too rigid. A procedure is proposed to identify the appropriate timing adjustment per stock, period and type of trade, which is flexible enough to deal

with the varying difference in lags between trade and quote reports. The procedure treats each stock individually and tests whether the quote revision frequency around a trade is contaminated by quote revisions triggered by a trade, and then determines the smallest timing adjustment needed to take this contamination into account. The procedure is applied to several stocks and sample periods between 1993 and 2003.

The chapter is structured as follows. In Section 1.2, I start with a description of the trade and quote reporting procedures at the NYSE, because these procedures are known to drive the lags. Section 1.3 discusses the 5-second rule and related reporting lags found in the literature. In Section 1.4, I discuss the data and have a preliminary look at the quote revision frequency around a trade. Section 1.5 presents a new procedure to determine the prevailing quote at the time of a trade. Section 1.6 concludes.

1.2 Trade and quote reporting procedures

Anyone who has noticed the rapid increase in market volume handled by the NYSE will find it natural that trade and quote reporting procedures have evolved over time. The subsequent sections show that changes to reporting procedures affected the difference between trade and quote reporting lags. For later reference, this section contains a short overview of the procedures, which is mainly based on Hasbrouck et al. (1993) and NYSE documents.

The NYSE records trades via the Consolidated Tape System (CTS) and revisions of the best quote via the Consolidated Quote System (CQS). The TAQ database is an extraction of these systems. The way that trades and quotes reach CTS and CQS has changed over the years.

After the 1987 market crash, electronic workstations were introduced to deal with high volumes of trades and quotes. Until June 1989 the procedure was as follows. The specialist calls out the details of trades and new quotes as they happen. These trades and quotes are recorded by the specialist assistant or by floor reporters. It is the specialist who determines whether a floor reporter is involved in the recording of trades and quotes. The specialist assistant controls the Display Book, an electronic workstation that keeps track of all limit orders and incoming market orders and assists in the recording and dissemination of trades and quotation changes. The floor reporter, employed by the exchange, records trades and quotes by filling in boxes on a mark-sense card and feeding it into an optical reader. Trade reports travel through the Post Support System (PSS) to the exchange's Market Data System (MDS). MDS performs

certain validation checks, before it sends the information to the CTS. Quote revisions travel through PSS to MDS and then to the CQS.

On 19 June 1989 the exchange began to abandon quote reporting by floor reporters. By September 1989, already 95% of quotes changed from the Display Book (Hasbrouck et al., 1993). By 2000, 99.9% of all quotes were updated by the Display Book, the exceptions being for trading halts and other related events (NYSE, 2000 and 2001).

Through the years 1987-2001, the floor reporter was also less frequently used for recording trades, as trades were more and more Display Book reported. In 1994, the mark-sense card system was abolished and the floor reporter began to use a hand-held device to report trades, which can be expected to speed up recording. The Display Book gained importance and by the year 2000, already 91% of all trades were Display Book reported (NYSE, 2000). On 24 July 2001, the floor reporter position was eliminated and since then all trade reporting has been done directly through the Display Book (NYSE, 2001). Prior to the elimination 99% of all trades were Display Book reported.

The best quote was not updated automatically when a trade affected the best quote until 27 May 2003, when the exchange introduced 'auto-quoting' for all stocks. This procedure implies that the NYSE automatically updates the NYSE's best bid or offer whenever a limit order is transmitted to the Display Book at a better price than the previous best bid or offer. When a trade occurs that involves the best bid or offer, the NYSE automatically updates the best bid or offer, and the associated depths, according to the specialist's book. Auto-quoting also includes adding size to the best quote as additional limit orders arrive and reducing size of the best quote as limit orders are executed or cancelled. Only in cases where the specialist trades for his own account are quotes not automatically updated. In other words, most quotes are automatically updated following a trade (NYSE, 2003a,b). In 2003, still only 5 percent of the quoting was performed manually on the Display Book (NYSE, 2003c).

1.3 Review of the literature

In an often-cited paper, Lee and Ready (1991) report a problem with the then existing reporting procedure to reconstruct the sequence of trades and quote revisions. As they point out, if the specialist assistant is faster in recording a quote revision than the floor reporter in recording a trade, the corresponding quote update can be recorded before the trade that triggered it. This is problematic if one seeks to determine the prevailing quote at the time of a trade. Lee and Ready (1991) investigated the lag of

trade reporting relative to quote reporting. Their results suggest using the prevailing quote at five seconds prior to the trade as the prevailing quote at the time of a trade. However, the dataset on which their analysis was based dates back to 1988 and the result was an average obtained for a cross-section of 150 stocks, while we will see below that the lags and required timing change can differ between stocks. Furthermore, the reporting procedure changes pointed out in Section 1.2 undoubtedly affected reporting lags and it seems unlikely that the 5-second rule is universal.

Although Lee and Ready (1991, footnote 10) realize that the delay can vary with the sample period, their 5-second rule has been used in many studies based on TAQ data of the nineties: see e.g. Ball and Chordia (2001), Busse and Green (2002), Chan et al. (2002), Chordia et al. (2001, 2002), Easley et al. (2001), Edelen and Gervais (2003), Engle and Patton (2004), Huang and Stoll (2001), Kryzanowski and Zhang (2002), Nyholm (2003), Schultz (2000), Stoll (2000) and Venkataraman (2001). This list is incomplete, but gives an idea of the popularity of the 5-second rule. Few studies check the robustness of their results with respect to the 5-second rule.

Delays of NYSE trade reports have already been studied. Blume and Goldstein (1997) report a median delay of sixteen seconds for NYSE trades between execution and reporting, for the period July 1994 - June 1995. Peterson and Sirri (2003) use a two-week sample of 1997 of the NYSE System Order Database Daily File (SOD file), which contains details of order entry and execution. This allows them to compare the execution time and the reporting time of trades. They report a median delay for trades of only 2 seconds for NYSE stocks. Piwowar and Wei (2003) study the impact of different trade and quote matching algorithms on estimates of the effective spread for Nasdaq and NYSE stocks. In order to determine an optimal matching algorithm, they search for the trade time adjustment that minimizes the rate of small trades occurring outside the prevailing spread. Small trades are defined as 1000 shares or less. Their results clearly show the sensitivity of the effective spread estimates to the algorithm and that the sensitivity has increased over time. However, their statistics are averages across several stocks. Henker and Wang (2006) computed the average adjustment rule for a cross-section of stocks in the style of Lee and Ready (1991) for more recent datasets between 1994 and 2002 and found a 1-second rule to be most appropriate. They also pointed out how timing specifications have a significant impact on the estimates of the adverse selection component of the spread. Bacidore et al. (2003) also report a series of lags belonging to the NYSE system, but do not compute statistics which can be used to account for reporting delays. Delays for Nasdaq stocks have been studied by

Bessembinder (2003), Ellis et al. (2000) and Piwowar and Wei (2003). In general, there is no consensus on how to deal with reporting delays.

1.4 Evidence on trade and quote reporting lags

It is intuitively clear that trade reporting lags depend on the way trades are reported. Hasbrouck et al. (1993) already pointed out that Display Book reported trades have a much smaller reporting delay than trades reported by floor reporters. They report 15% of trades to be Display Book reported for a sample of five days in November 1990 and this percentage to be increasing fast. As shown in Section 1.5, the increasing popularity of the Display Book, as e.g. described in NYSE (2003c), has decreased the overall trade reporting lag over time.

1.4.1 The data

I consider five 3-month samples, between 1993 and 2003, of NYSE trades and quotes from the TAQ database. The periods are April - June 1993, April - June 1997, April - June 2001, October - December 2001 and October - December 2003. In the discussion below, I refer to the different 3-month periods by the year only (1993, 1997, 2001a, 2001b and 2003). Most of the papers that make use of the Lee and Ready (1991) 5-second rule are based on samples that cover at least one of the first two sample periods. The third and fourth sample periods are just before and after the abolition of floor reporting. The last sample period is a period after auto-quoting was introduced.

I select five groups of five stocks based on the trading activity of the stocks, because a priori one may expect a link between reporting lags and trading activity. Ranking the stocks in ascending order according to dollar volume traded in 2001a, I first select three groups starting from the three stocks found at the 33%, 67% and 100% quantiles and each time moving down until I have five stocks that existed during the period 1993 - 2003. Secondly, I take two more groups of actively traded stocks, because floor reporters were especially used for actively traded stocks, which allows us to have a better look at non Display Book reported trades. More specifically, using the same procedure, I take two more groups starting from the two stocks found at the 90% and 98% quantile. The five groups of stocks are {GE, IBM, EMC, PFE, TYC}, {F, DD, BBY, HAL, ADI}, {NSC, BCR, MYL, LTD, JCI}, {KWD, NAB, NC, ACG, ESL} and {PYM, GTY, SAF, NNJ, EY}. See Table 1.1 for the corresponding dollar volume. In Section 1.5, it turns out that the choice of the groups based on the trading activity does not have any further

implications, because the results are not significantly different between the groups. For each stock only trades and quotes are selected that meet all of the following conditions:
- trades and quotes need to occur within the trading day: 9:30 - 16:00;
- trades need to be regular trades, which were not corrected, changed, or signified as cancel or error; this is indicated by a zero value of the correction indicator (CORR);
- trades need to be regular way or NYSE Direct+ trades; this is indicated by a blank or 'E' entry of the Condition indicator;
- quotes need to stem from normal trading conditions; this is indicated by the Mode indicator taking the value 1, 2, 3, 6, 10, 12 or 18.

Another feature that is used below is whether trades are Display Book reported. This is indicated by specific values of the G127 indicator; see the TAQ2 user's guide for more details.

Table 1.1: Dollar volume of selected stocks

Symbol	1997	2001a	2001b	2003	Symbol	1997	2001a	2001b	2003
GE	15735	52561	37078	28205	**LTD**	567	1558	1174	1825
IBM	25823	49375	45574	26646	**JCI**	535	1556	1798	2569
EMC	2753	31446	15246	9032	**KWD**	72	126	125	387
PFE	8735	27688	27540	28681	**NAB**	88	125	148	284
TYC	3109	26389	26189	10010	**NC**	84	123	60	75
F	5363	8029	4869	6501	**ACG**	17	121	272	107
DD	7733	7944	6776	7875	**ESL**	37	116	76	107
BBY	298	7683	11616	11661	**PYM**	9	9	10	22
HAL	3772	7657	7823	4457	**GTY**	13	9	79	67
ADI	1622	7606	8817	7676	**SAF**	18	9	6	30
NSC	986	1566	874	1673	**NNJ**	8	9	9	6
BCR	571	1565	1612	1529	**EY**	115	9	5	51
MYL	413	1563	2105	2973					

Note: in millions of dollars; unavailable for 1993.

1.4.2 Quote revision frequencies around trades

In order to accurately determine the prevailing quote at the time of a trade, it is necessary that each quote update triggered by a trade can be distinguished from other quote revisions. If the quote update is recorded before the trade, then one would take the prevailing quote at one second before the quote update as the prevailing quote at the time of the trade. Unfortunately, the data does not contain a field that links trades with the mechanic quote update they trigger, which would help to identify most quote revisions except for submissions triggered by a trade. Moreover, it is not always possible

to detect the link between a trade and its mechanic quote update from the trade and quote sequence, by comparing trade sizes with changes in quote depth and trade prices with quote prices. This is caused by active trading and cancellations of limit orders, which complicate the interpretation of the trade and quote sequence.

Although we cannot distinguish a quote update triggered by a trade from other quote revisions, we can compute the average timing of the quote updates relative to the trades. For this purpose, I compute the frequency of quote revisions at each second of a 30-second interval around a trade, for each stock and each 3-month period. If quote revisions are triggered by trades, then one would expect to find a peak in this frequency distribution at the time these quotes are reported. Quote revisions triggered by a trade refer to both the mechanical update of the order book due to a trade and new orders triggered by a trade. Other quote revisions are either the result of incoming orders or cancelled orders and their timing should be approximately independent of the timing of trades. The 30-second intervals $[-15, 15]$ centred on trades are allowed to overlap, which implies that some quote revisions are counted more than once. As a quote update triggered by a non Display Book reported trade is expected to occur further away from the trade, I look at 55-second intervals $[-35, 20]$ around this type of trades.

As an example, Figures 1.1a and 1.1b display the frequency distribution for BBY Display Book (DB) reported and non Display Book (NDB) reported trades of 2001a, respectively. The horizontal axis shows the number of seconds before and after the recording of a trade. The vertical bars represent the quote revision frequency around a trade, computed as the number of quote revisions divided by the number of trades. We notice a clear difference between both figures, which indicates that DB reported trades and NDB reported trades are subject to different lags. Both figures show a hump, which indicates the time relative to a trade at which quote revisions triggered by trades are recorded. The humps are situated at different times relative to the time the trade is recorded. In Figure 1.1a, the hump is steep and shows that quote revisions triggered by trades are recorded at the time the trade is recorded or up to a few seconds later. Hence, to avoid taking a quote update triggered by a trade as the prevailing quote at the time of that trade, it would seem optimal to take the prevailing quote *one second before the trade* (-1), i.e. just before the hump. The grey bar indicates the time of the prevailing quote that results from the test described in Section 1.5. For the NDB trades in Figure 1.1b, however, quote revisions triggered by a trade are also recorded before the trade. Again, to avoid taking a quote update triggered by a trade as the prevailing quote at the time of that trade, it appears to be a good choice to take the prevailing

quote about *ten seconds before the trade* as the prevailing quote at the time of the trade (-10). The marked difference between the two figures suggests that it is better to treat the two types of trades separately when determining the prevailing quote.

Unfortunately, the quote revision frequencies in the figures above are biased if overlapping intervals around trades are used, because it implies that quote revisions due to one trade can be counted as a quote revision around another trade. If the arrival of trades around a trade were uniformly distributed, then the use of overlapping intervals would have no side effects, because the quote revisions that the trades trigger would also be uniformly distributed. Figure 1.1c gives an example of the frequency distribution of trades around a DB trade. Across all stocks, this distribution is typically bimodal and approximately symmetric around zero, with a steep slope between zero and the mode at ± 3 to 4 seconds away from the trade, and after the mode the frequency decays (often to a higher level than that close to the trade).[1] As a consequence, the quote revisions that these trades trigger are not expected to be uniformly distributed, which causes a bias in the frequency of quote revisions around a trade.

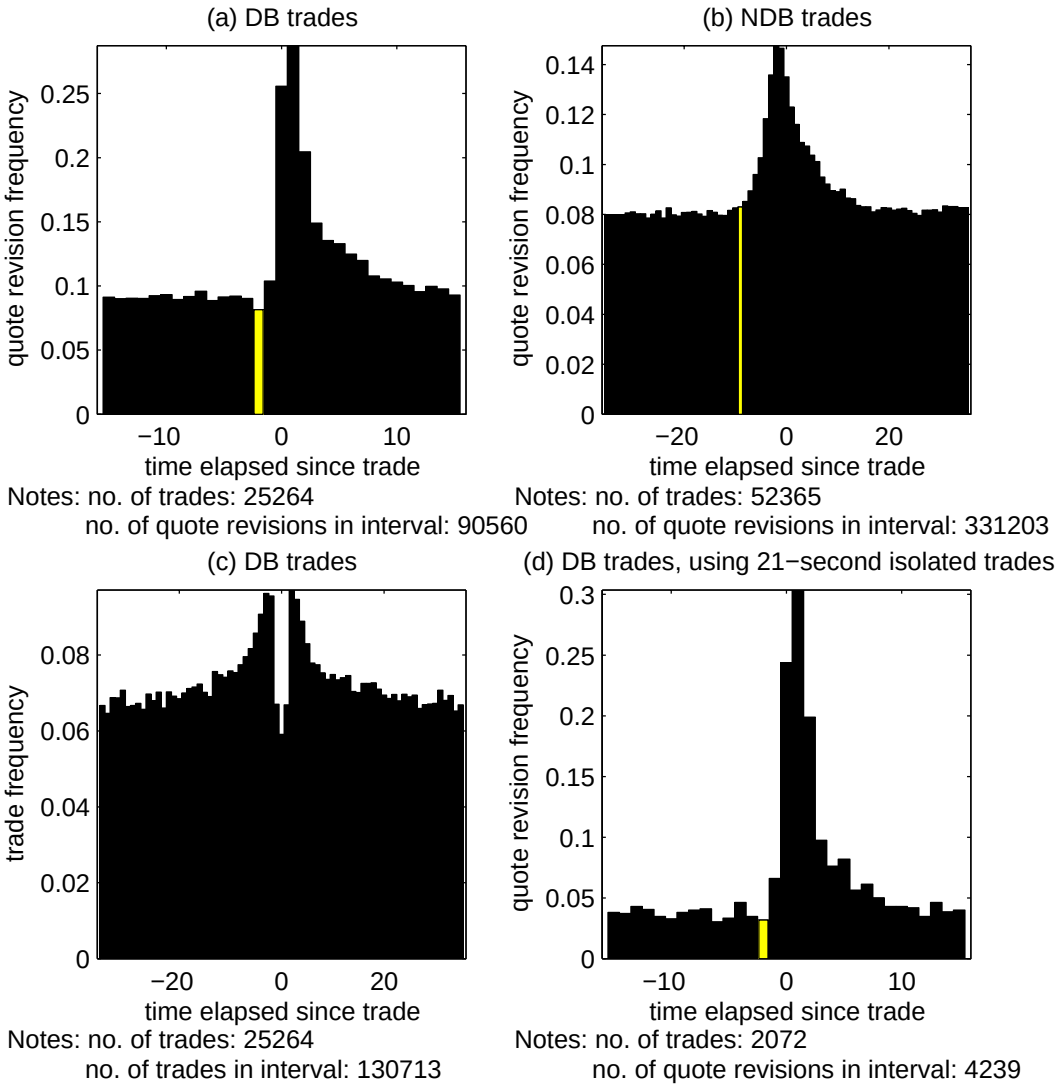
One way to deal with the bias is to use non-overlapping intervals, i.e. isolated trades. In order to avoid any overlap of the 20-second intervals for Display Book reported trades, one would need to select trades that are at least 21 seconds away from any other trade. For the non Display Book reported trades, trades isolated by about 56 seconds are more appropriate, because a high quote revision frequency is found between 35 seconds before the trade and 20 seconds after the trade. Compare e.g. Figure 1.1a with Figure 1.1d, where the isolated trades in the latter figure have a lower frequency before (and after) the trade compared to the frequency at the time of the trade. The problem with this alternative is that it implies far fewer observations and less smooth figures, especially for actively traded stocks.

The main results of this chapter, as summarized in Table 1.2 below, are based on a better method to tackle the bias in the number of quote revisions around a trade. That is to deconvolve the quote revision frequency around a trade. Let v be the perturbed quote revision frequency around a trade, u the trade frequency around a trade and q the 'true' quote revision frequency around a trade. It holds that v is the convolution of u and q, i.e. $v = u * q$. Let V, U and Q be the Fourier transformations of these densities, e.g. $V = \mathcal{F}(v)$, then it holds that $V = UQ$ and hence $Q = V/U$. Upon taking the inverse Fourier transform, $q = \mathcal{F}^{-1}(V/U)$, we obtain an estimate of the true quote

[1] For example order splitting leads to series of trades with short and similar durations, which can affect the trade distribution around a trade.

revision frequency around a trade that is based on all observations. In other words, if we deconvolve v given u, we obtain an estimate of q. In the appendix, I explain that this procedure meets certain difficulties due to the truncation of the densities at the borders of the intervals, but forgetting about practicalities, this represents the main idea about the method used below to compute the quote revision frequency around a trade. Figure 1.2 presents examples of the quote revision frequency around a trade computed in this way.

Figure 1.1: Quote revision and trade frequencies around BBY trades in 2001a.



1.5 Towards new adjustment rules

The distribution of quote revisions around a trade typically shows an increase in the frequency of quote revisions at some time before the trade. Thus, we seek a rule for

determining the start of this increase, and then we can take the prevailing quote at one second before the increase as the prevailing quote at the time of the trade.

I suggest the following rule. Let s be the elapsed time, in seconds, since a trade, e.g. $s = -15, ..., 15$. Let t be the second in this interval at which quote revisions triggered by a trade start to arrive. Then $t-1$ is the second at which we can find the prevailing quote. Assume that quote revisions that are not triggered by a trade arrive according to a Poisson process with rate λ_s at s. This assumption is realistic and does not contradict the literature on duration models that points to clustering in order flow, because order arrival around a trade is a different concept. The arrival of quote revisions not triggered by a trade can be treated as independent of the arrival of a trade. This can be seen from the empirical quote revision frequency around a trade, which converges to a constant level as we move away from the trade. Orders submitted in anticipation of a trade and trades systematically filling incoming orders could in theory also lead to higher quote revision frequencies before a trade. The fact that often hardly any distortion is observed before DB trades, however, indicates that this type of effect must be minor.

A test can be developed based on the idea that λ_s is expected to remain approximately constant over time, while estimates of λ_s will be biased upward from t onward due to the arrival of quote revisions triggered by trades. Let q_s be the number of quote revisions at s and n the total number of trades. Then, λ_s can be estimated by

$$\hat{\lambda}_s = \frac{q_s}{n}.$$

Let λ_s^p, $s = -11, ..., 15$, be the mean of λ_s, $s = -15, ..., 15$, over the past 5 seconds, for example

$$\lambda_{-5}^p = \frac{\lambda_{-9} + ... + \lambda_{-5}}{5},$$

and let $\hat{\lambda}_s^p$ be the mean of $\hat{\lambda}_s$ over the past five seconds. I will interpret a statistically significant increase in $\hat{\lambda}_s$ above $\hat{\lambda}_s^p$ as evidence that, at s, at least part of the quote revisions are triggered by a trade.

To test the null hypothesis that λ_s is not higher than λ_{s-1}^p, or that $s < t$, I use the test statistic

$$\frac{\hat{\lambda}_s - \hat{\lambda}_{s-1}^p}{s.e.} \overset{a}{\sim} N(0,1),$$

as $n \to \infty$. Under the null hypothesis ($s < t$), the standard error can be estimated as

$$s.e. = \left(\widehat{\text{Var}}\left(\hat{\lambda}_s\right) + \widehat{\text{Var}}\left(\hat{\lambda}_{s-1}^p\right)\right)^{1/2} = \left(\frac{\hat{\lambda}_s}{n} + \frac{1}{25n}\sum\nolimits_{j=1}^{5} \hat{\lambda}_{s-j}\right)^{1/2}$$

In the analysis below, I use critical values from the standard normal distribution. The true critical values are somewhat smaller than those from the standard normal distribution, because the finite sample distribution of the test statistic is left-skewed for small n and λ. However, when I obtain such critical values by simulation and use these in the test, then the resulting rule was different in only two of the cases considered in Table 1.2. Therefore, for simplicity one can apply the critical values of the normal distribution, while keeping in mind that the true p-value of the test is somewhat lower than those implied by the normal distribution.

The null hypothesis $H_0 : \lambda_s \leq \lambda_{s-1}^p$ is sequentially tested for $s = -10, ..., 15$ and a significance level of 2.5%. In addition, the null hypothesis $H_0 : \lambda_{s+1} \leq \lambda_{s-1}^p$ is sequentially tested for $s = -10, ..., 14$ and a significance level of 2.5%. The smallest s for which both null hypotheses can be rejected is inferred as t. The prevailing quote can be found at $t - 1$. The double test is motivated by the observation that a single null is sometimes rejected too early within the interval compared to the big hump further in the interval. The independence of λ_s and λ_{s+1} implies that the double test corresponds to the joint test with null hypothesis $H_0 : \lambda_s \leq \lambda_{s-1}^p$ and $\lambda_{s+1} \leq \lambda_{s-1}^p$, with an overall significance level of approximately 5%.

I apply this method to determine the adjustment rule for all 25 stocks in all 5 periods and for DB and NDB trades separately. In case of NDB trades, I let $s = -35, ..., 35$. Figure 1.2 presents examples of the prevailing quote that results from the procedure, which is indicated by a grey bar. Table 1.2 contains the derived time of the prevailing quote relative to the time of the trade (in seconds) for all cases. A blank indicates that there are no trades of that type, i.e. not even among the non-isolated trades. If there are less than 100 DB trades, less than 500 NDB trades or less than 200 quote revisions, then that case is ignored and is indicated by $\cdot$. For the last two periods, non Display Book reported trades are the exception and therefore they are not included in Table 1.2. The table also contains three cases indicated by $\times$ in which case the algorithm selected a positive rule, either due to a relatively low number of observations or due to a poor estimate of the quote frequency around a trade.

Table 1.2 shows that by 1993 the Display Book is well in use, because for quite a few stocks all trades are Display Book reported. This is confirmed by Table 1.3, which presents the percentage of DB trades. For the DB trades of 1993, depending on the stock, the adjustment rules in Table 1.2 vary moderately between 0 and -3 seconds; see e.g. Figure 1.2a for BCR. Contrary to DB trades, the adjustment rules for NDB trades of 1993 are much larger and go up to 20 seconds; see e.g. Figure 1.2b for LTD.

Table 1.2: Time of the prevailing quote relative to the time of the trade (in seconds)

	1993		1997		2001a		2001b	2003
	DB	NDB	DB	NDB	DB	NDB	DB	DB
GE	·	−9	·	−2	−6	−16	−1	−2
IBM	·	−3	0	−9	−1	−4	−5	−4
EMC	0	−20	−1		−1		−1	−4
PFE	·	−17	·	−6	−1	·	−1	−4
TYC	0		−1		−2		×	−2
F	×	×	−1	−7	−1		−1	−2
DD	−2	·	−2	·	−1		−2	−2
BBY	−2	·	−2	−11	−2	−8	−1	−4
HAL	·	−10	−1	−5	−1	−5	−1	−2
ADI	·	−14	−2	−26	−1		−1	−2
NSC	0		−2	·	−1		−1	−2
BCR	−1	−14	−2	−8	−2	·	−2	−2
MYL	0		−1		−1		−1	−2
LTD	·	−19	·	−11	−1		−1	−2
JCI	0	·	−1		−1		−2	−2
KWD	−1		−2		−1		−1	−2
NAB	·	·	−1		−1		−1	−2
NC	−1		−2		−2		−1	−2
ACG	−3		−3		−2		−2	−2
ESL	−2	·	−2		−1		−6	−2
PYM	·	·	−2		−4		−3	−3
GTY	−1		−1		0		−1	−2
SAF	−1		−2		−2		−3	−2
NNJ	·		−2		−1		−4	−5
EY	0		−2		−2		−8	−2

Notes: Blank = no trades; · = less than 100 DB trades, less than 500 NDB trades or less than 200 quote revisions around a trade; × = algorithm failed to determine a rule

IBM is an exception to this, with an adjustment rule of -3. The reason is that the distribution of quote revisions around NDB trades looks like the distribution around DB trades. Presumably it does concern Display Book reported trades that, however, have not been indicated as such in the TAQ database.

Table 1.3: Percentage of DB trades and average transaction volume

	DB %			Average Transaction Volume					
				DB			NDB		
	1993	**1997**	**2001a**	**1993**	**1997**	**2001a**	**1993**	**1997**	**2001a**
GE	0	0	9	248	2248	1833	1569	2089	5891
IBM	0	2	53	243	1906	1432	2248	2482	2436
EMC	67	100	100	2271	2860	4555	2369		
PFE	1	0	100	252	1305	2953	2333	1388	5007
TYC	100	100	100	1409	3002	2385			
F	12	3	100	1365	11672	2882	3344	4148	
DD	100	99	100	2260	1854	1339	5067	6337	
BBY	93	3	33	1581	2582	1333	2000	3061	1956
HAL	5	4	95	468	1448	1330	2537	1660	3199
ADI	2	94	100	662	3883	1349	2398	5538	
NSC	100	98	100	1639	916	1457		2099	
BCR	31	82	97	2464	2638	1317	1814	3120	5414
MYL	100	100	100	2292	3546	1864			
LTD	0	0	100	2014	1031	2093	4082	3219	
JCI	79	100	100	1323	1545	593	1612		
KWD	100	100	100	1531	1520	713			
NAB	15	100	100	540	1058	719	611		
NC	100	100	100	625	562	357			
ACG	100	100	100	1476	1708	1792	200		
ESL	100	100	100	886	842	752			
PYM	7	100	100	1632	1643	1311	1648		
GTY	100	100	100	1078	1557	314			
SAF	100	100	100	798	1485	1306			
NNJ	100	100	100	864	1130	1075			
EY	100	100	100	2149	3786	2381			

Note: In 2001b and 2003 trades are all DB reported.

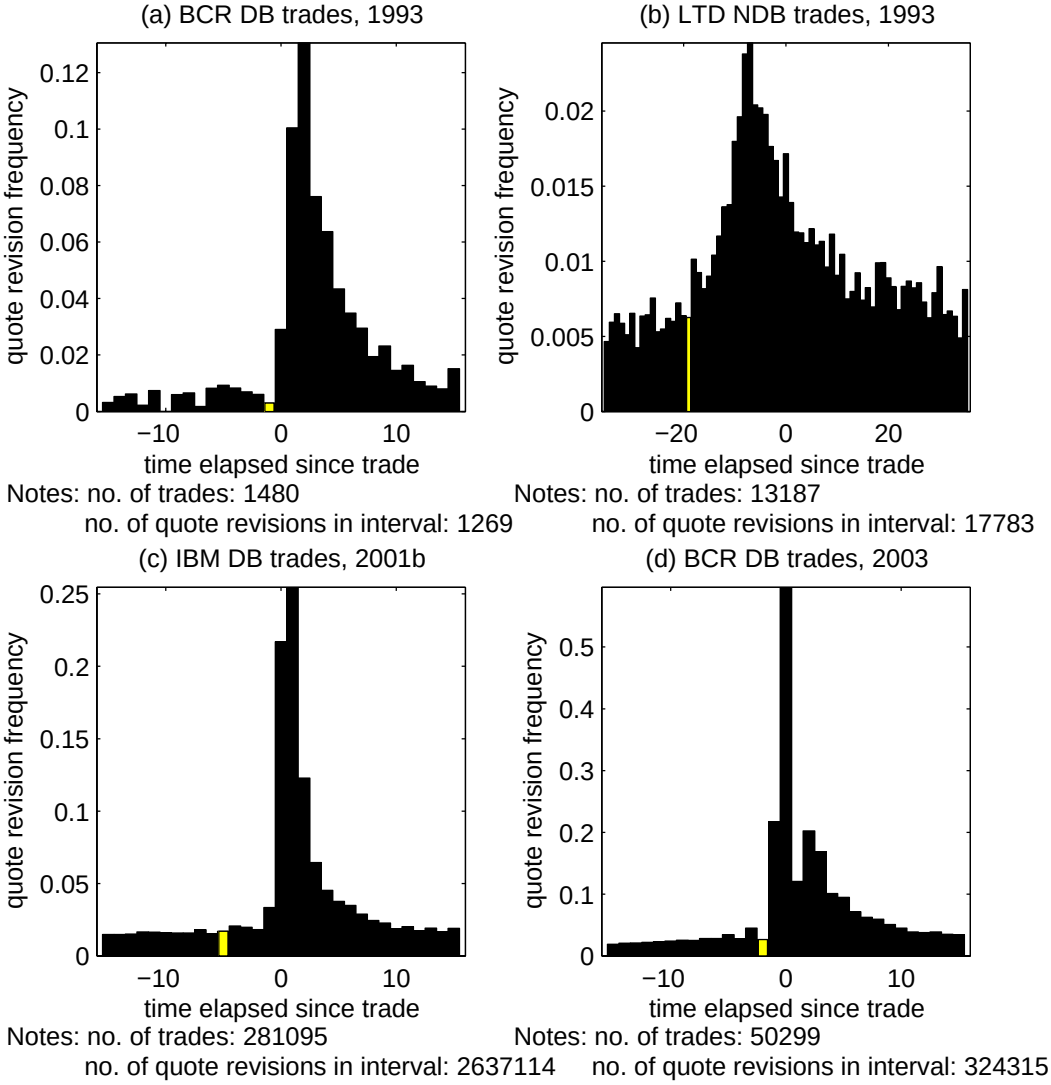
In general, it is clear from these figures that the problem of reporting lags is severe for NDB trades, and in fact accurately determining the prevailing quotes for these trades is not really feasible. If a trade report can be delayed up to 20 seconds compared to the associated quote report, then this demands a considerable adjustment. However, such a large adjustment implies that we can be far too conservative for trade reports that have no delay relative to the quote update report.

From Table 1.3 it can be seen that the average transaction volume of NDB trades is

typically larger compared to DB trades. Given that floor brokers typically walk large orders to the specialist post, apparently it are typically floor reporters who record the trades associated with these orders instead of the specialist assistant.

For Display Book reported trades of 1997, depending on the stock, I mainly find that the prevailing quote can be found 1 or 2 seconds before the trade. When we compare the results for NDB trades of the same stock between the 1993 and 1997 periods, it appears that the relative trade report delays have decreased, except in two cases. Apparently, the switch from the mark-sense card system to a hand-held device for the floor reporter decreased the trade reporting lags; see Section 1.2.

Figure 1.2: Quote revision frequency around trades and the timing adjustment



For 2001a, the trades are predominantly Display Book reported, except for GE, for which about 91% of trades are still non Display Book reported, see Table 1.3. All trades

are Display Book reported from mid-2001 on, although a few exceptions can be found in the TAQ data. For most stocks in 2001 and 2003, the prevailing quote for Display Book reported trades is found at 1 or 2 seconds before the trade. There are a few cases, however, where the algorithm stops early, e.g. at -6, because λ_s has already moved up significantly, which is interpreted as contamination by quote revisions triggered by a trade. Looking at these figure, see e.g. Figure 1.2c, one may want to ignore this small amount of contamination and adopt a -2 or -1 rule.

In the year 2003 autoquoting is introduced and, as expected, we notice more quote revisions at the time of the trade; see e.g. Figure 1.2d. It appears, however, that quote updates can still be reported before a trade is reported, because a substantial portion of the quote revisions is recorded one second before the trade is recorded. The reason is not clear, but a small discrepancy between the clocks of CTS and CQS can already have this effect. The optimal adjustment is therefore –2 seconds. Contrary to the other periods, the shape of the figure and the adjustment rule are very stable across stocks. The stabilising effect of autoquoting on the reporting of quotes makes determining the prevailing quote easier. This result suggests a new adjustment rule for NYSE stock data from mid-2003 onward. That is, to take the prevailing quote two seconds before the trade as the prevailing quote at the time of the trade. The cases where the algorithm stops earlier look like Figure 1.2c and the small contamination before -2 could be ignored.

The improvement of changing the adjustment rule from five - as suggested by Lee and Ready (1991) - to two seconds depends on whether there are actually changes in the quote between five and two seconds prior to a trade. In order to study this I compute the prevailing quote at the time of a trade based on the two rules separately, for each stock in the 2003 sample. The two rules yield the same ask quote between 69% and 98% of the time (depending on the stock), the same bid quote between 70% and 96% of the time, and the same spread between 62% and 94% of the time. Thus, the different rules do lead to different quotes in a substantial number of cases. More interesting is to know the potential consequences of such a difference in timing. One common measure that depends on the correct timing is the effective spread. Let m be the prevailing midquote at the time of a trade and p the transaction price, then the effective spread is computed as $2\,|m-p|$. One can easily verify that the effective spread will typically be overestimated if the absolute timing adjustment to determine the prevailing midquote is too large, because the larger the time interval between m and p, the more p can move away from m. For the cases in 2001b and 2003, the average effective spread at -5 was

between -0.5% lower and 19% higher than the spread at -2, and can thus be significant in size. If one would for example measure transaction costs of the stock exchange based on effective spreads, then timing adjustments are important. Henker and Wang (2006) discuss the impact of timing adjustments on a specific application more in depth.

1.6 Conclusion

This chapter proposed a new procedure to match trades and quotes of NYSE stocks. The procedure tests whether the quote revision frequency around a trade is contaminated by quote revisions triggered by a trade, and then determines the smallest timing adjustment needed to take this contamination into account. The procedure was applied to a sample of 25 stocks in 5 sample periods. The results show that the difference between trade reporting lags and quote reporting lags varies across stocks and time. The variation can be mainly explained by changes in the reporting procedures of the NYSE and the co-existence of two reporting systems for trades, i.e. trades are Display Book reported or not. The non Display Book reported trades usually require a larger adjustment to match trade and quote times.

In summary, given a sample of NYSE stocks from the TAQ database, it is recommended to take the following issues into account when applying the new procedure. For samples prior to mid-2001 it is better to treat Display Book reported trades separately. In case of large sample periods, it is better to split the sample in order to take further differences across time into account. To sum up, it is recommended to determine a separate adjustment rule for each trade type, each stock, and each sub-sample.

Finally, an interesting observation is that the system of auto-quoting, in which quotes are automatically updated since mid-2003, has a stabilising effect on the reporting of quotes. In particular, taking the prevailing quote at two seconds before the trade as the prevailing quote at the time of the trade appears to be an appropriate adjustment rule for all stocks.

Appendix

Deconvolution of the quote frequency around a trade

Let v be the perturbed quote revision frequency around a trade, u the trade frequency around a trade[2] and q the 'true' quote revision frequency around a trade. The direct

[2] When computing the trade frequency around a trade, this trade should be counted too such that the frequency is at least 1 at the time of the trade.

deconvolution of v given u as explained in the main text has to be extended with some rearranging of vector parts to obtain the correct deconvolution q. The reason is that u, v and q are truncated at the borders of the interval around a trade; see Figures 1.1 and 1.2. For example, assume that we observe both u, q and v on the interval $[-15, 15]$ around a trade. The vector of the convolution $v' = u * q$ will be of length $(2 \times 31 - 1) = 61$ and typically contains non-zero elements across the entire interval. Thus, the convolution v' is not equal to the observed v unless parts of the vector are rearranged, i.e. $v = [v'_2 + v'_4 \;\; v'_m \;\; v'_1 + v'_3]$, where v'_m is the element in the middle of the vector v' and v'_i, $i = 1, ..., 4$, is the i-th 15-element part of the vector v' excluding v'_m. Similarly, when deconvolving v by u, the estimate q' has to be rearranged.

In order to deconvolve v, first extend the vector on both sides with two vectors of 15 zeros to increase its length to 61 elements and extend the interval to [-30,30]. Let V, U and Q be the Fourier transformations of the densities, e.g. $V = \mathcal{F}(v)$. Then it holds that $V = UQ$ and hence $Q = V/U$. Upon taking the inverse Fourier transform, $q' = \mathcal{F}^{-1}(V/U)$, we obtain an estimate of the true quote revision frequency around a trade. The result q' has non-zero elements across the 61 elements interval. Its parts have to be rearranged to obtain the truncated version q'' of q' on the interval [-15,15] as we would observe it, i.e. $q'' = [q'_1 + q'_3 \;\; q'_a \;\; q'_2 + q'_4]$, where q'_a is the 16th element of q' plus the average of the 46th and 47th element and q'_i, $i = 1, ..., 4$, is the i-th 15-element part of the vector q' excluding its 16th element.

Quote revision frequency computation by iteration

The estimate q'' can still be a poor estimate of the true quote revision frequency around a trade, because we deconvolve the observed frequency v instead of the unknown v'. However, the latter can be obtained by iteration, which allows better estimates of q. Each iteration consists of the following steps:

1. Compute the convolution $d = u * q''$.

2. Replace the middle part of the vector d (of 31 elements) by the middle part of v minus the outer parts of d. That is, define $z = [d_1 \;\; v_2 - d_4 \;\; v_m \;\; v_3 - d_1 \;\; d_4]$, where v_m is the element in the middle of v plus the average of the first and last element of d.

3. Deconvolve z by u using Fourier transformation as explained above to obtain a new estimate q''.

Step 1 computes the convolution implied by the estimate q''. In Step 2, the tails of this convolution are imposed on the observed frequencies v. In step 3, this new vector of frequencies is deconvolved and a new estimate of q is obtained. By iterating these steps, z converges to v' and q'' to q.

Chapter 2

Asymptotic Results for GMM Estimators of Stochastic Volatility Models

2.1 Introduction

Over the last two decades there has been an increasing interest in stochastic volatility (SV), which was introduced by Clark (1973) and extended by Tauchen and Pitts (1983), as a framework for the analysis of time-varying volatility in financial markets. This interest is partly due to an important contribution by Hull and White (1987), where SV models arise as discrete time approximations to continuous time volatility diffusions used in option pricing. More generally, it is recognized that SV models constitute a valuable alternative to GARCH-type models for analysing financial time series (Ghysels, Harvey, and Renault (1996), Shephard (1996)).

Due to the fact that in SV models the mean and the volatility are driven by separate stochastic processes (implying that volatility is unobservable), SV models are much harder to estimate than GARCH models. This chapter presents analytical results that may be used to improve and assess the quality of GMM-based estimation of SV models. GMM, while not asymptotically efficient, is still the simplest estimation method for SV models currently available. It has been proposed by Taylor (1986) and Melino and Turnbull (1990), and its properties have been studied using Monte Carlo methods by Jacquier, Polson, and Rossi (1994), Andersen and Sørensen (1996, 1997), and Andersen, Chung, and Sørensen (1999). Other available estimation methods for SV models include quasi-maximum likelihood (Nelson (1988), Harvey, Ruiz, and Shephard (1994), Ruiz (1994)), simulated maximum likelihood (Danielsson and Richard (1993), Danielsson (1994)), simulation-based GMM (Duffie and Singleton (1993)), indirect inference (Gouriéroux, Monfort, and Renault (1993), Monfardini (1998)), Markov

chain Monte Carlo methods (Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998), Chib, Nardari, and Shephard (2002)), efficient method of moments (Gallant, Hsieh, and Tauchen (1997), Andersen, Chung, and Sørensen (1999)), Monte Carlo maximum likelihood (Sandmann and Koopman (1998)), and (approximate) maximum likelihood (Fridman and Harris (1998)). Apart from quasi-maximum likelihood, all of these methods are computationally more demanding, as they rely – often quite heavily – on numerical simulation and/or integration techniques both for obtaining point estimates and for assessing the accuracy of the latter. In view of its simplicity, we consider GMM estimation as a useful alternative to the more elaborate methods.

In this chapter we derive closed-form expressions for the optimal weighting matrix for GMM estimation of the basic SV model, and for the asymptotic covariance matrix of the optimal GMM estimator, for a large class of moment conditions. To date, applications of GMM in this context have typically relied on a nonparametrically estimated weighting matrix, because an expression for the optimal weighting matrix (as a function of the parameters) was not available. For the SV model, such weighting matrix estimates (and hence the estimates of the covariance matrix of the GMM estimator) can be very imprecise even for relatively large sample sizes (Andersen and Sørensen (1996)). Using the exact expressions eliminates this problem. Furthermore, the accuracy of the GMM estimator can now be assessed by analytical means, which is currently not possible for any of the other estimators.

The moment conditions that we consider fall into two categories. The first set of conditions is obtained by considering the first two moments and the autocovariances of any order of the log-squared observations. These conditions have recently been considered by Wright (1999), in connection with the fractionally integrated SV model. The second set of moment conditions are derived from the absolute observations and are more standard in this literature. We study moment conditions that involve the product of any number of absolute observations, each one raised to any positive real power and lagged any number of periods. This set considerably extends the set of moment conditions that have been employed so far. The results that we present pertain to any selection of moment conditions from these two sets.

The expressions for the asymptotic covariance matrix of the GMM estimator allow us to compare the relative efficiency of GMM estimators based on various sets of moment conditions, and to compare these efficiencies with those of other estimators, reported earlier in Monte Carlo studies. One of the conclusions from this comparison is that a judicious selection of a very small number of moment conditions yields a GMM estimator

with only a small efficiency loss compared to the MCMC estimator (which is known to be asymptotically efficient).

The analytical results regarding the optimal weighting matrix allow us to fastly and accurately assess the information content of any subset of moment conditions considered. This, in turn, permits the optimal selection of a small set of highly informative moment conditions from very large sets, and subsequent GMM estimates to be based on the optimal selection of moments. We propose a general four-step data-based procedure for the optimal selection of moment conditions, and apply it in the SV context.

In Section 2.2, we present the basic SV model and the moment conditions. Expressions for the optimal weighting matrix and the asymptotic covariance matrix of the GMM estimator are derived in Section 2.3. Section 2.4 presents some comparative evidence on the relative efficiencies of the GMM and other estimators (partly compiled from the literature). The moment selection procedure is given in Section 2.5. Section 2.6 compares the performance in moderate to large samples of the iterated GMM estimator using Bartlett weights to estimate the optimal weighting matrix versus using analytical expressions for the optimal weighting matrix. Section 2.7 concludes. Proofs are given in the Appendix.

2.2 Moment conditions for the SV model

The basic SV model is given by

$$y_t = \exp\left(h_t/2\right) u_t, \tag{2.1}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma\sqrt{1-\phi^2}v_t, \tag{2.2}$$

where y_t is observable, h_t is latent log-volatility, (u_t, v_t) is i.i.d. $N(0, I)$, and $\theta = (\mu, \phi, \sigma)'$ is a vector of parameters. The restriction $|\phi| < 1$ is imposed, ensuring that y_t is stationary and ergodic. While it is more common to parameterise the model in terms of $\lambda = (\alpha, \phi, \omega)'$, with $\alpha = \mu(1-\phi)$ and $\omega = \sigma\sqrt{1-\phi^2}$, we prefer the parameterisation in terms of θ for algebraic reasons and because of an invariance with respect to μ given below. For comparison with earlier studies, however, numerical standard errors will be presented in terms of λ.

From the point of view of inference, the fundamental problem with the SV model is the latent character of h_t, which makes it difficult to compute the values of the likelihood function and hence to estimate θ by maximum likelihood. It is easy, however, to derive moment conditions implied by the SV model and then to apply the Generalized Method

of Moments (Hansen, 1982). The moment conditions considered in this chapter relate either to the log-squared observations, $\log y_t^2$, or to the absolute observations, $|y_t|$. The latter class of moment conditions constitutes the standard approach to GMM estimation of SV models (Taylor (1986), Melino and Turnbull (1990), Jacquier, Polson, and Rossi (1994), Andersen and Sørensen (1996, 1997), Andersen, Chung, and Sørensen (1999)). The former class of moment conditions is suggested in passing by Jacquier, Polson, and Rossi (1994), and is effectively employed by Wright (1999) in the context of the fractionally integrated SV model.

Moment conditions related to $\log y_t^2$ are easily obtained. It follows from (2.1) that $\log y_t^2 = h_t + \log u_t^2$. The mean and variance of $\log u_t^2$ are known to be $c_1 = -\log 2 - \gamma = -1.270\,4$ and $c_2 = \frac{1}{2}\pi^2 = 4.934\,8$, respectively, where $\gamma = 0.577\,2$ is Euler's constant. Let

$$\begin{aligned} z_t &= \log y_t^2 - \mu - c_1 \\ &= h_t - \mu + \log u_t^2 - c_1. \end{aligned}$$

Since $h_t \sim N(\mu, \sigma^2)$, $\mathrm{Cov}(h_t, h_{t-i}) = \phi^{|i|}\sigma^2$, and u_t is i.i.d. and independent of h_t, it follows that

$$E[z_t] = 0, \tag{2.3}$$

$$E[z_t z_{t-i}] = \phi^i \sigma^2 + I_{(i=0)} c_2, \qquad i \geq 0, \tag{2.4}$$

where $I_{(\cdot)}$ is the indicator function. It can be shown that none of these conditions is redundant in the sense of Breusch et al. (1999).

We now derive the class of moment conditions generated by the expectation of $|y_{t_1}^{r_1} ... y_{t_p}^{r_p}|$, where $r_1, ..., r_p$ are positive real numbers and $t_1 > ... > t_p$. Let ν_r be the r-th absolute moment of a standard normal random variate, i.e.

$$\nu_r = E|u_t|^r = \frac{2^{r/2}}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right)$$

where $\Gamma(z)$ is the gamma function. Then, because $t_1 > ... > t_p$,

$$E \prod_{j=1}^{p} \left| \nu_{r_j}^{-1} u_{t_j}^{r_j} \right| = 1.$$

Furthermore, $\sum_{j=1}^{p} r_j h_{t_j}$ is normally distributed with mean $\mu \sum_{j=1}^{p} r_j$ and variance $\sigma^2 \sum_{j,j'=1}^{p} r_j r_{j'} \phi^{|t_j - t_{j'}|}$. So, by property that $E \exp(X) = \exp(a + \frac{1}{2}b^2)$ when $X \sim N(a, b^2)$, we have

$$E \exp\left(\frac{1}{2} \sum_{j=1}^{p} r_j h_{t_j} \right) = \exp\left(\delta_{t_1,...,t_p}^{r_1,...,r_p} \right),$$

where

$$\delta_{t_1,...,t_p}^{r_1,...,r_p} = \frac{\mu}{2}\sum_{j=1}^{p} r_j + \frac{\sigma^2}{8}\sum_{j,j'=1}^{p} r_j r_{j'}\phi^{|t_j - t_{j'}|}.$$

Hence, defining

$$\begin{aligned} Y_{t_1,...,t_p}^{r_1,...,r_p} &= \exp\left(-\delta_{t_1,...,t_p}^{r_1,...,r_p}\right)\prod_{j=1}^{p}\left|\nu_{r_j}^{-1} y_{t_j}^{r_j}\right| \\ &= \exp\left(-\delta_{t_1,...,t_p}^{r_1,...,r_p}\right)\exp\left(\frac{1}{2}\sum_{j=1}^{p} r_j h_{t_j}\right)\prod_{j=1}^{p}\left|\nu_{r_j}^{-1} u_{t_j}^{r_j}\right|, \end{aligned}$$

it follows that

$$E\left[Y_{t_1,...,t_p}^{r_1,...,r_p}\right] = 1, \qquad r_1,...,r_p > 0; \quad t_1 > ... > t_p. \tag{2.5}$$

It is obvious that adding the same integer to $t_1, ..., t_p$ yields the same moment condition. As far as we know, within the class of moment conditions defined by (2.5), only moment conditions where $p = 1$ or where $p = 2$ and $r_1 = r_2 \in \{1, 2\}$ have so far been considered in the literature.

2.3 Optimal GMM

Let $E(f_t) = f$ be a finite selection of the set of moment conditions given by (2.3)–(2.5) that identifies θ. Let $g_t = f_t - f$. By assumption, the observations on y_t permit us to calculate $g_1, ..., g_T$ as functions of θ. The optimal GMM estimator (Hansen (1982)) of θ based on this selection is $\hat{\theta} = \arg\min_\theta \bar{g}'\hat{V}^{-1}\bar{g}$, where $\bar{g} = T^{-1}\sum_{t=1}^{T} g_t$ and $\hat{V}$ consistently estimates V, where

$$V = \sum_{l=-\infty}^{\infty} E(g_t g_{t-l}') = \sum_{l=-\infty}^{\infty} \mathrm{Cov}(f_t, f_{t-l}').$$

The asymptotic covariance matrix of $\sqrt{T}(\hat{\theta} - \theta)$ is $(D'V^{-1}D)^{-1}$, where $D = E(\frac{\partial g_t}{\partial \theta'})$. Expressions for D and V, for an arbitrary selection of moment conditions, are presented below. These expressions make it possible to compute the optimal weighting matrix V^{-1} and the asymptotic covariance matrix $(D'V^{-1}D)^{-1}$ of the GMM estimator as functions of the parameter values. Substituting estimates for these parameter values yields estimates of V^{-1} and $(D'V^{-1}D)^{-1}$, which will generally be more precise than the nonparametric estimator based on Bartlett weights that is routinely used in a GMM context. The Monte Carlo results of Andersen and Sørensen (1996) show that the latter estimator of V may be imprecise even in samples of size $50,000$. Using the expressions presented here avoids such problems. Furthermore, the expression for V makes it also

possible to estimate θ by the continuous-updating GMM estimator of Hansen, Heaton, and Yaron (1996), that is, by solving $\min_\theta \bar{g}' V^{-1} \bar{g}$.

Some straightforward calculus shows that the rows of D are to be selected (according to the selection of moments) from

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -i\phi^{i-1}\sigma^2 & -2\phi^i\sigma \\ -\nabla_\mu \delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} & -\nabla_\phi \delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} & -\nabla_\sigma \delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} \end{pmatrix} \begin{array}{l} i \geq 0; \\ \\ r_1,\ldots,r_p > 0; \quad t_1 > \ldots > t_p; \end{array}$$

where

$$\begin{aligned} \nabla_\mu \delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} &= \frac{1}{2}\sum_{j=1}^{p} r_j, \\ \nabla_\phi \delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} &= \frac{\sigma^2}{8}\sum_{j,j'=1}^{p} r_j r_{j'} |t_j - t_{j'}| \phi^{|t_j - t_{j'}|-1}, \\ \nabla_\sigma \delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} &= \frac{\sigma}{4}\sum_{j,j'=1}^{p} r_j r_{j'} \phi^{|t_j - t_{j'}|}. \end{aligned}$$

The main result of this chapter is an expression for the elements of V, given in Theorem 1 below. Let $c_i = E(\log u_t^2 - c_1)^i$, $i = 3, 4$. It is shown in the Appendix that

$$\begin{aligned} c_3 &= -14\zeta(3) = -16.829, \\ c_4 &= \tfrac{7}{4}\pi^4 = 170.47, \end{aligned}$$

where $\zeta(z)$ is the Riemann zeta function. For $r > 0$, let

$$\begin{aligned} \kappa_r &= \log 2 + \psi\left(\frac{r+1}{2}\right) - c_1, \\ \xi_r &= \kappa_r^2 + \psi'\left(\frac{r+1}{2}\right) - c_2, \end{aligned}$$

where $\psi(z) = \frac{d}{dz}\log\Gamma(z)$, the digamma function, and $\psi'(z) = \frac{d}{dz}\psi(z)$, the trigamma function.

Theorem 1 *For any a_t and b_t, let $V(a_t, b_t) = \sum_{l=-\infty}^{\infty} \mathrm{Cov}(a_t, b_{t-l})$. Let $i, j \geq 0$. Then*

$$V(z_t, z_t) = \frac{1+\phi}{1-\phi}\sigma^2 + c_2, \tag{2.6}$$

$$V(z_t, z_t z_{t-j}) = I_{(j=0)} c_3, \tag{2.7}$$

$$V(z_t z_{t-i}, z_t z_{t-j}) = A_1\sigma^4 + A_2 c_2 \sigma^2 + I_{(i=j\neq 0)} c_2^2 + I_{(i=j=0)}(c_4 - c_2^2), \tag{2.8}$$

where

$$A_1 = |i-j|\,\phi^{|i-j|} + |i+j|\,\phi^{|i+j|} + \left(\phi^{|i-j|} + \phi^{|i+j|}\right)\frac{1+\phi^2}{1-\phi^2},$$
$$A_2 = 2\left(\phi^{|i-j|} + \phi^{|i+j|}\right).$$

Let $r_1, ..., r_{p+q} > 0$*;* $t_1 > ... > t_p$*;* $t_{p+1} > ... > t_{p+q}$*; and let*

$$L = \{l|\,\{t_1, ..., t_p\} \cap \{t_{p+1} - l, ..., t_{p+q} - l\} \neq \varnothing\}.$$

Then

$$V\left(Y^{r_1,...,r_p}_{t_1,...,t_p}, Y^{r_{p+1},...,r_{p+q}}_{t_{p+1},...,t_{p+q}}\right) = B + \sum_{l\in L}(B_l + 1)C_l, \tag{2.9}$$

where

$$B = \sum_{l=-\infty}^{\infty} B_l, \qquad B_l = \exp\left(\frac{\sigma^2}{4}\sum_{j=1}^{p}\sum_{j'=p+1}^{p+q} r_j r_{j'}\phi^{|t_j - t_{j'}+l|}\right) - 1,$$
$$C_l = \left(\prod_{j=1}^{p+q}\nu_{r_j}^{-1}\right) E\left(\prod_{j=1}^{p}\left|u^{r_j}_{t_j}\right|\prod_{j=p+1}^{p+q}\left|u^{r_j}_{t_j-l}\right|\right) - 1,$$

and

$$V\left(z_t, Y^{r_1,...,r_p}_{t_1,...,t_p}\right) = D_1\sigma^2 + D_2, \tag{2.10}$$
$$V\left(z_t z_{t-i}, Y^{r_1,...,r_p}_{t_1,...,t_p}\right) = D_1'\sigma^4 + D_2'\sigma^2 + D_3', \tag{2.11}$$

where

$$D_1 = \frac{1}{2}\left(\frac{1+\phi}{1-\phi}\right)\sum_{j=1}^{p} r_j, \qquad D_2 = \sum_{j=1}^{p}\kappa_{r_j},$$
$$D_1' = \frac{1}{4}\sum_{j,j'=1}^{p} r_j r_{j'}\phi^{\left|t_j - t_{j'}+i\right|}\left(|t_j - t_{j'} + i| + \frac{1+\phi^2}{1-\phi^2}\right),$$
$$D_2' = \frac{1}{2}\sum_{j,j'=1}^{p} r_j\kappa_{r_{j'}}\left(\phi^{\left|t_j - t_{j'}+i\right|} + \phi^{\left|t_j - t_{j'}-i\right|}\right),$$
$$D_3' = I_{(i=0)}\sum_{j=1}^{p}\xi_{r_j} + \sum_{j,j'=1}^{p} I_{\left(i=t_j-t_{j'}\neq 0\right)}\kappa_{r_j}\kappa_{r_{j'}}.$$

We see that, not unexpectedly, the optimal weighting matrix, V^{-1}, and the GMM asymptotic covariance matrix, $(D'V^{-1}D)^{-1}$, do not depend on μ. From a computational point of view, notice that L has at most pq elements, so computing $\sum_{l\in L}(B_l + 1)C_l$ requires a finite number of steps. Furthermore, B can be approximated by $B(I) = \sum_{l=-I}^{I} B_l$, where I is a positive integer. As the following lemma shows, the error of approximation $|B - B(I)|$ is bounded by an exponentially decaying function in I, and this bound can be inverted to determine I as a function of the desired accuracy of the approximation.

Lemma 2 *Let $r_1, ..., r_{p+q} > 0$; $t_1 > ... > t_p$; $t_{p+1} > ... > t_{p+q}$; and let I be a positive integer. Then*

$$|B - B(I)| \leq 2\frac{\exp(a|\phi|^I) - 1}{1 - |\phi|},$$

where

$$a = \frac{\sigma^2}{4} \sum_{j=1}^{p} \sum_{j'=p+1}^{p+q} r_j r_{j'} \, |\phi|^{-|t_j - t_{j'}|} .$$

If one is interested in λ rather than θ, one may apply the transformation $\theta \longmapsto \lambda(\theta)$ to yield $\hat{\lambda} = \lambda(\hat{\theta})$, the optimal GMM estimator of λ, which has asymptotic covariance matrix $(\frac{\partial \lambda}{\partial \theta'})(D'V^{-1}D)^{-1}(\frac{\partial \lambda}{\partial \theta'})'$, with

$$\frac{\partial \lambda}{\partial \theta'} = \begin{pmatrix} 1-\phi & -\mu & 0 \\ 0 & 1 & 0 \\ 0 & -\sigma\phi(1-\phi^2)^{-1/2} & (1-\phi^2)^{1/2} \end{pmatrix}.$$

2.4 Comparison of GMM and other estimators

In this section we first compare the relative efficiencies of GMM and other estimators, for two sets of values of λ, namely $(\alpha, \phi, \omega) = (-0.736, 0.90, 0.363)$ and $(\alpha, \phi, \omega) = (-0.1472, 0.98, 0.1657)$. These parameter values have been used in earlier Monte Carlo studies (Jacquier, Polson, and Rossi (1994), Andersen and Sørensen (1996), Fridman and Harris (1998), Sandmann and Koopman (1998), Andersen, Chung, and Sørensen (1999)). Tables 2.1 and 2.2 present the results. The asymptotic standard errors of the GMM estimators were computed using the expressions derived above. The moment conditions were selected from the set related to the log-squared observations, or from the set related to the absolute observations, or from both. For comparability with other studies, from (2.5) we only selected moment conditions for which $p = 1$ or for which $p = 2$ and $r_1 = r_2 \in \{1, 2\}$. The (finite sample) standard errors of the other estimators were taken from the aforementioned Monte Carlo studies and multiplied by $\sqrt{T}$. The relative asymptotic efficiency of the GMM estimators is seen to increase rapidly with the number of moments, at least when this number is small. Using a large number of moment conditions yields asymptotic standard errors slightly above those of the MCMC method, which is known to be asymptotically efficient. In this respect, it appears that some of the published standard errors regarding ML and Monte Carlo ML are not in line with those of the MCMC method.

Table 2.1: Standard errors of $\sqrt{T}\hat{\lambda}$

T	# Moments	Method of estimation	Standard Error of		
			$\sqrt{T}\hat{\alpha}$	$\sqrt{T}\hat{\phi}$	$\sqrt{T}\hat{\omega}$
∞	3	GMM (log-moments)[a]	127.52	17.31	32.66
∞	12	GMM (log-moments)[a]	12.04	1.63	3.80
∞	27	GMM (log-moments)[a]	10.06	1.36	3.22
∞	102	GMM (log-moments)[a]	10.04	1.36	3.22
∞	3	GMM (absolute moments)[b]	178.46	24.18	46.78
∞	15	GMM (absolute moments)[b]	11.34	1.53	2.96
∞	30	GMM (absolute moments)[b]	8.14	1.10	2.18
∞	75	GMM (absolute moments)[b]	7.55	1.02	2.03
∞	14	GMM (joint moments)[c]	16.92	2.29	4.27
∞	22	GMM (joint moments)[c]	11.30	1.53	2.92
∞	42	GMM (joint moments)[c]	8.12	1.10	2.14
∞	102	GMM (joint moments)[c]	7.53	1.02	1.99
10000	14	Infeasible GMM[d] (true weight)	11.4	1.6	3.1
4000	14	Infeasible GMM[d] (true weight)	10.6	1.5	3.1
4000	4	EMM: GARCH(1,1)[e]	9.51	1.2	3.1
4000	6	EMM: GARCH(1,1) - Kz(2)[e]	9.68	1.3	3.2
4000	8	EMM: GARCH(1,1) - Kz(4)[e]	8.28	1.1	2.1
2000	24	GMM[f]	18	3	3.8
2000	-	Quasi-ML[f]	20	3	4.8
2000	-	MCMC[f]	6.6	1	1.5
500	-	ML[g]	9.1	1	2
500	-	Monte Carlo ML[h]	0.5	2.2	2

Parameter values: $(\alpha, \phi, \omega) = (-0.736, 0.90, 0.363)$.
GMM conditions are selected from Eqs. (2.3)–(2.5), as indicated below. Most footnotes refer to multiple lines in the Table.
a. Eqs. (2.3)–(2.4) with i running from 0 to $1, 10, 25$, and 100, respectively.
b. Eqs. (2.5) with $p = 1$, r_1 running from 1 to $1, 5, 10$, and 25 respectively; and Eq. (2.5) with $p = 2$, $r_1 = r_2 \in \{1, 2\}$, $t_1 - t_2$ running from 1 to $1, 5, 10$, and 25 respectively.
c. Eqs. (2.3)–(2.4) with i running from 0 to $3, 5, 10$, and 25, respectively; Eq. (2.5) with $p = 1$, r_1 running from 1 to $3, 5, 10$, and 25, respectively; and Eq. (2.5) with $p = 2$, $r_1 = r_2 \in \{1, 2\}$, $t_1 - t_2$ running from 1 to $3, 5, 10$, and 25, respectively.
d. Andersen and Sørensen (1996), Table 3: Eq. (2.5) with $p = 1$, r_1 running from 1 to 4; Eq. (2.5) with $p = 2$, $r_1 = r_2 = 1$, $t_1 - t_2 \in \{6, 8, 10, 12, 14\}$; and Eq. (2.5) with $p = 2$, $r_1 = r_2 = 2$, $t_1 - t_2 \in \{15, 17, 19, 21, 23\}$. 'Infeasible GMM' uses a nonparametric estimate of the weighting matrix based on a large sample of simulated data using true parameter values.
e. Andersen, Chung, and Sørensen (1999), Table 3.
f. Jacquier, Polson, and Rossi (1994), Tables 5–7. For GMM: Eq. (2.5) with $p = 1$, r_1 running from 1 to 4; and Eq. (2.5) with $p = 2$, $r_1 = r_2 \in \{1, 2\}$, $t_1 - t_2$ running from 1 to 10.
g. Fridman and Harris (1998), Table 1.
h. Sandmann and Koopman (1998), Table 2.

Table 2.2: Standard errors of $\sqrt{T}\hat{\lambda}$

T	# Moments	Method of estimation	Standard Error of		
			$\sqrt{T}\hat{\alpha}$	$\sqrt{T}\hat{\phi}$	$\sqrt{T}\hat{\omega}$
∞	3	GMM (log-moments)[a]	136.37	18.53	77.30
∞	12	GMM (log-moments)[a]	6.67	0.90	4.00
∞	27	GMM (log-moments)[a]	2.96	0.40	1.71
∞	52	GMM (log-moments)[a]	2.51	0.34	1.39
∞	102	GMM (log-moments)[a]	2.49	0.34	1.37
∞	3	GMM (absolute moments)[b]	264.71	35.95	150.79
∞	15	GMM (absolute moments)[b]	8.49	1.15	4.79
∞	30	GMM (absolute moments)[b]	4.15	0.56	2.28
∞	75	GMM (absolute moments)[b]	2.48	0.34	1.23
∞	14	GMM (joint moments)[c]	14.95	2.03	8.43
∞	22	GMM (joint moments)[c]	8.45	1.15	4.76
∞	42	GMM (joint moments)[c]	4.12	0.56	2.26
∞	102	GMM (joint moments)[c]	2.44	0.33	1.20
4000	4	EMM: GARCH(1,1)[d]	2.8	0.37	1.3
4000	4	EMM: GARCH(1,1) - Kz(2)[d]	2.9	0.39	1.8
4000	6	EMM: GARCH(1,1) - Kz(4)[d]	2.7	0.36	1.0
500	24	GMM[e]	5.8	0.80	2
500	-	Quasi-ML[e]	12	2	3.1
500	-	MCMC[e]	2.7	0.4	1
500	-	ML[f]	0.4	0.30	0.8
500	-	Monte Carlo ML[g]	0.2	2	1

Parameter values: $(\alpha, \phi, \omega) = (-0.1472, 0.98, 0.1657)$.
GMM conditions are selected from Eqs. (2.3)–(2.5), as indicated below. Most footnotes refer to multiple lines in the Table.
a. Eqs. (2.3)–(2.4) with i running from 0 to $1, 10, 25, 50$, and 100, respectively.
b. Eq. (2.5) with $p = 1$, r_1 running from 1 to $1, 5, 10$, and 25 respectively; and Eq. (2.5) with $p = 2$, $r_1 = r_2 \in \{1, 2\}$, $t_1 - t_2$ running from 1 to $1, 5, 10$, and 25 respectively.
c. Eqs. (2.3)–(2.4) with i running from 0 to $3, 5, 10$, and 25, respectively; Eq. (2.5) with $p = 1$, r_1 running from 1 to $3, 5, 10$, and 25, respectively; and Eq. (2.5) with $p = 2$, $r_1 = r_2 \in \{1, 2\}$, $t_1 - t_2$ running from 1 to $3, 5, 10$, and 25, respectively.
d. Andersen, Chung, and Sørensen (1999), Table 3.
e. Jacquier, Polson, and Rossi (1994), Tables 5–7. For GMM: Eq. (2.5) with $p = 1$, r_1 running from 1 to 4; and Eq. (2.5) with $p = 2$, $r_1 = r_2 \in \{1, 2\}$, $t_1 - t_2$ running from 1 to 10.
f. Fridman and Harris (1998), Table 1.
g. Sandmann and Koopman (1998), Table 2.

2.5 Data-based selection of moment conditions

In Monte Carlo studies it is often found that the small sample bias of the GMM estimator grows with the number of moment conditions. Newey and Smith (2000, 2001) show that the number of terms of the second-order bias increases linearly with the number of moment conditions. Thus, rather than using a large number of moment conditions (relative to the sample size), it is in terms of bias often safer to select only a small number of them. It is important, then, to choose the moments judiciously, in the sense that they contain as much information as possible for the estimand. Several authors have addressed the question of how to select the moment conditions to estimate the SV model, essentially by resorting to Monte Carlo simulation of the accuracy of the GMM estimator for any given choice of moments. The results of the previous section provide a more precise and much faster tool to guide the choice of moments.

Table 2.3: Asymptotic standard errors of $\sqrt{T}\hat{\phi}$ for parsimoniously selected moments

Moment set	k	Selected moments	Standard Error of		
			$\sqrt{T}\hat{\alpha}$	$\sqrt{T}\hat{\phi}$	$\sqrt{T}\hat{\omega}$
M_L	3	$z_t; z_t z_{t-1}; z_t z_{t-11}$	18.31	2.49	5.41
M_A	3	$Y_t^2; Y_{t,t-7}^{1,2}; Y_{t,t-5,t-14}^{1,1,1}$	10.59	1.44	4.72
$M_L \cup M_A$	3	$z_t z_{t-10}; Y_t^2; Y_{t,t-7,t-15}^{1,1,1}$	10.08	1.37	4.07
M_L	4	$z_t; z_t z_{t-1}; z_t z_{t-10}; z_t z_{t-12}$	14.78	2.01	4.62
M_A	4	$Y_t^1; Y_t^2; Y_{t,t-10}^{1,1}; Y_{t,t-8,t-15}^{1,1,1}$	9.65	1.31	2.55
$M_L \cup M_A$	4	$z_t z_{t-10}; Y_t^2; Y_{t,t-5,t-14}^{1,1,1}; Y_{t,t-7,t-13}^{1,1,1}$	9.46	1.28	4.16
M_L	5	$z_t; z_t z_{t-1}; z_t z_{t-9}; z_t z_{t-11}; z_t z_{t-14}$	13.37	1.82	4.31
M_A	5	$Y_t^1; Y_t^2; Y_{t,t-7}^{1,1}; Y_{t,t-9}^{1,1}; Y_{t,t-13}^{1,1}$	9.07	1.23	2.47
$M_L \cup M_A$	5	$Y_t^1; Y_t^2; Y_{t,t-7}^{1,1}; Y_{t,t-9}^{1,1}; Y_{t,t-13}^{1,1}$	9.07	1.23	2.47

Parameter values: $(\alpha, \phi, \omega) = (-0.736, 0.90, 0.363)$.

We first discuss 'parameter-based' optimal selection of moments—which requires the parameters to be known—and then data-based optimal selection, which does not require the parameters to be known. The method has general applicability and the SV model can be viewed as an application here. Consider the sets M_L and M_A of

log-moment and absolute moment conditions, respectively, defined as

M_L : (2.3)–(2.4) with integers $i \leq 50$,

M_A : (2.5) with $\max_{j,j'} |t_j - t_{j'}| \leq 15$ and integers r_j with $\sum_{j=1}^{p} r_j \leq \begin{cases} 20 & \text{for } p = 1, \\ 4 & \text{for } p = 2, 3, 4. \end{cases}$

The sets M_L and M_A comprise 52 and 985 moment conditions, respectively. We performed a search for the set of k moment conditions, selected from either M_L, M_A, or $M_L \cup M_A$, that yield the smallest asymptotic standard error of $\hat{\phi}$. Global optimisation, by enumeration, was performed over M_L for $k = 3, 4, 5$, and over M_A and $M_L \cup M_A$ for $k = 3$. Global optimisation over M_A and $M_L \cup M_A$ for $k = 4$ and $k = 5$ turned out to be infeasible in terms of computation time, and in these cases we experimented with the Point Exchange algorithm (Fedorov (1972)). This algorithm does not necessarily yield the global optimum, and its output depends on the starting selection of moment conditions as input. By picking the starting selection at random and repeating this a couple of times, the algorithm was able to reproduce the global optimum in all cases where enumeration was possible. We therefore applied it in those cases where global optimisation was not feasible, without the guarantee of having found the globally optimal selection of moments from the specified sets. The parameter values were fixed at $(\alpha, \phi, \omega) = (-0.736, 0.90, 0.363)$, as in Table 2.1. Table 2.3 reports the selected moments and the asymptotic standard errors of the corresponding GMM estimators. Comparing Table 2.3 with Table 2.1 yields the following conclusions: (i) there is a dramatic increase in efficiency by selecting the moments in an optimal way; (ii) given that the efficiency bound for the asymptotic standard error of $\sqrt{T}\hat{\phi}$ (which is asymptotically attained by the MCMC estimator) appears to be around 1, the efficiency loss of the GMM estimator with optimal moment selection from $M_L \cup M_A$ is not excessively large, even in the just-identified case ($k = 3$); (iii) while M_A contains a richer (also a much larger) set of moment conditions than M_L – as is reflected by the smaller asymptotic standard errors – the combination of M_A and M_L may yield an improvement upon M_A, as is the case here for $k = 3$ and $k = 4$.

The optimal selection of moment conditions from any given set, as described above, depends on the parameter values. The following four-step data-based procedure gives, asymptotically, the optimal selection independently of the parameter values. In step 1, the parameters are consistently estimated, for example by GMM using a large set of moment conditions. In step 2, the optimal selection of moment conditions is made

for parameter values set equal to the estimates obtained. In step 3, the parameters are re-estimated by GMM using the moment conditions selected in step 2. In step 4, steps 2 and 3 are repeated once. Steps 1–3 and steps 1–4 are asymptotically equivalent procedures and the resulting GMM estimates based on the selected moments are asymptotically equivalent to the GMM estimates that use the optimal selection of moments. Step 4 is added to ensure that the final selection of moments is based on estimates that are relatively efficient while typically less biased than those obtained in step 1.

2.6 Finite sample properties

So far, we discussed the asymptotic properties of the GMM estimator based on the analytical results (3.1)–(2.11). In this section we report Monte Carlo results that compare, for fixed sets of moment conditions, the iterated GMM estimator based on the analytically derived optimal weighting matrix and the iterated GMM estimator based on an estimate, using Bartlett weights with lag parameter set to $T^{1/3}$, of the optimal weighting matrix. We also present results on the performance of the former estimator with optimally selected moments according to the four-step procedure given in Section 5. The results are used to double-check the asymptotic results, to assess their usefulness in improving the estimates, and to explore the finite sample properties of the moment selection procedure.

All simulation results are based on $1,000$ replications, with data generated by the SV model (2.1)–(2.2) and parameter values as in Table 2.1, i.e. $(\alpha, \phi, \omega) = (-0.736, 0.90, 0.363)$. Table 2.4 presents, for $T = 10,000$ and the same sets of moment conditions as in Table 2.1, the Monte Carlo average and standard deviation of the estimates along with the ratio of the standard deviation and the asymptotic standard error (std/ase). The estimates based on the analytical optimal weighting matrix have standard deviations that are in line with Table 2.1, with std/ase close to 1. For the two sets of 3 moment conditions the standard deviation is smaller than the analytically derived standard error. This is due to the fact that the boundary condition $|\phi| < 1$ is imposed on the estimates, which reduces std in finite samples compared to the asymptotic standard error. For large sets of moment conditions, the estimates are more efficient, hit the boundary much less often, and as a result std/ase is close to 1. As expected, the results for iterated GMM based on Bartlett weights show that estimating the optimal weighting matrix has an efficiency cost compared to knowing the optimal weighting matrix. Table 2.4 lacks results for Bartlett-based iterated GMM estimation with many

Table 2.4: iterated GMM estimates based on the analytical versus the estimated optimal weighting matrix in large samples ($T = 10,000$)

		analytical weights			Bartlett weights		
Moments		$\hat{\alpha}$	$\hat{\phi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\omega}$
3 log-moments[a]	mean	-0.834	0.887	0.366	-0.891	0.879	0.351
	$\sqrt{T}std$	59.0	7.99	13.6	82.3	11.2	21.1
	std/ase	0.463	0.462	0.418	0.645	0.644	0.647
12 log-moments[a]	mean	-0.742	0.899	0.363	-0.736	0.900	0.360
	$\sqrt{T}std$	12.3	1.67	3.90	12.4	1.68	3.95
	std/ase	1.02	1.02	1.02	1.03	1.03	1.04
27 log-moments[a]	mean	-0.748	0.898	0.365	-0.735	0.900	0.359
	$\sqrt{T}std$	9.99	1.35	3.24	9.93	1.35	3.27
	std/ase	0.993	0.995	1.01	0.987	0.988	1.01
102 log-moments[a]	mean	-0.748	0.898	0.365	-0.706	0.904	0.344
	$\sqrt{T}std$	9.96	1.35	3.23	10.0	1.35	3.32
	std/ase	0.992	0.993	1.01	1.00	0.995	1.03
3 absolute moments[b]	mean	-0.864	0.883	0.387	-0.663	0.910	0.308
	$\sqrt{T}std$	53.6	7.15	11.4	63.4	8.51	17.6
	std/ase	0.300	0.295	0.243	0.355	0.352	0.376
15 absolute moments[b]	mean	-0.719	0.902	0.361	-0.742	0.899	0.350
	$\sqrt{T}std$	11.5	1.56	3.02	17.96	2.34	91.23
	std/ase	1.02	1.02	1.02	1.58	1.52	30.8
30 absolute moments[b]	mean	-0711	0.903	0.362	-	-	-
	$\sqrt{T}std$	8.35	1.13	2.20	-	-	-
	std/ase	1.03	1.02	1.01	-	-	-
75 absolute moments[b]	mean	-0.676	0.907	0.359	-	-	-
	$\sqrt{T}std$	7.91	1.06	2.06	-	-	-
	std/ase	1.05	1.04	1.02	-	-	-
14 joint moments[c]	mean	-0.724	0.902	0.359	-0.780	0.894	0.355
	$\sqrt{T}std$	17.7	2.40	4.60	21.57	2.92	5.11
	std/ase	1.05	1.05	1.08	1.28	1.28	1.20
22 joint moments[c]	mean	-0.720	0.902	0.360	-0.711	0.903	0.313
	$\sqrt{T}std$	11.5	1.56	2.99	17.9	2.43	6.12
	std/ase	1.02	1.02	1.02	1.59	1.59	2.09
42 joint moments[c]	mean	-0.713	0.903	0.361	-	-	-
	$\sqrt{T}std$	8.30	1.12	2.16	-	-	-
	std/ase	1.02	1.02	1.01	-	-	-
102 joint moments[c]	mean	-0.682	0.906	0.359	-	-	-
	$\sqrt{T}std$	7.79	1.05	2.02	-	-	-
	std/ase	1.03	1.03	1.02	-	-	-

std = standard deviation; ase = asymptotic standard error.
Parameter values: $(\alpha, \phi, \omega) = (-0.736, 0.90, 0.363)$. Footnotes a, b and c are as in Table 2.1.

moment conditions, because then estimation is difficult due to numerical problems with the scaling and near-singularity of the weighting matrix. This problem is particularly severe in the case of absolute (or joint) moments. There may be ways to overcome the problem, but in any case we found the estimates based on the analytical weighting matrix, in addition to being more efficient, easier to compute.

The efficiency gain of iterated GMM based on the analytical optimal weighting matrix relative to iterated GMM based on an estimate of the optimal weighting matrix is larger in small samples. Table 2.5 presents results for the two estimators based on the parsimoniously selected set of 5 moments as in Table 2.3, i.e. Y_t^1, Y_t^2, $Y_{t,t-7}^{1,1}$, $Y_{t,t-9}^{1,1}$ and $Y_{t,t-13}^{1,1}$, for $T = 500$, 2000, 4000, and 10000. As required, the ratio std/ase shrinks to 1 as T grows, but it does so much faster for the estimator based on the analytical weight matrix. For this design and choice of moments, the latter estimator is much more efficient. In addition, the asymptotic standard errors are reliable when $T \geq 4,000$, while it requires $T \geq 10,000$ with estimates of the optimal weighting matrix. Note also that, when $T = 500$, the mean of the estimates points to a bias of both estimators, in particular for μ, which appears somewhat bigger when Bartlett weights are used.

Table 2.5: iterated GMM estimates based on the analytical versus the estimated optimal weighting matrix in small samples

		analytical weights			Bartlett weights		
T		$\hat{\alpha}$	$\hat{\phi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\omega}$
500	mean	-1.09	0.852	0.379	-1.375	0.814	0.394
	$\sqrt{T}std$	25.4	3.45	3.11	32.3	4.39	3.83
	std/ase	2.80	2.82	1.26	3.58	3.58	1.55
2,000	mean	-0.769	0.895	0.361	-0.853	0.884	0.368
	$\sqrt{T}std$	12.9	1.77	2.58	23.4	3.19	3.50
	std/ase	1.43	1.44	1.04	2.58	2.60	1.41
4,000	mean	-0.750	0.898	0.362	-0.773	0.895	0.363
	$\sqrt{T}std$	9.54	1.29	2.47	15.1	2.04	2.90
	std/ase	1.05	1.05	0.996	1.67	1.66	1.17
10,000	mean	-0.742	0.899	0.363	-0.748	0.898	0.362
	$\sqrt{T}std$	9.019	1.221	2.409	10.187	1.383	2.646
	std/ase	0.995	0.996	0.974	1.123	1.128	1.070

Parameter values: $(\alpha, \phi, \omega) = (-0.736, 0.90, 0.363)$.
GMM conditions are Y_t^1, Y_t^2, $Y_{t,t-7}^{1,1}$, $Y_{t,t-9}^{1,1}$ and $Y_{t,t-13}^{1,1}$ (see Table 2.3, bottom line).

Finally, we experimented with the data-based selection of moment conditions. Step 1 was implemented with 12 log-moments, trying to strike a balance between efficiency, bias, and ease of computation. In step 2, five moments were selected from the set of

102 joint moments considered in Table 2.1. Steps 3 and 4 were executed as described above. Throughout we used iterated GMM based on the analytical optimal weighting matrix. Table 2.6 reports results for $T = 500$, 2000, 4000, and 10000.

Table 2.6: iterated GMM estimation based on the analytical optimal weighting matrix with data-based selection of 5 moments

T		$\hat{\alpha}$	$\hat{\phi}$	$\hat{\omega}$
500	mean	-1.18	0.839	0.388
	$\sqrt{T}std$	27.1	3.69	4.52
	std/ase	2.99	3.01	1.83
2,000	mean	-0.796	0.892	0.366
	$\sqrt{T}std$	13.3	1.82	3.03
	std/ase	1.47	1.48	1.22
4,000	mean	-0.752	0.898	0.362
	$\sqrt{T}std$	9.94	1.35	2.69
	std/ase	1.10	1.10	1.09
10,000	mean	-0.743	0.899	0.363
	$\sqrt{T}std$	9.80	1.33	2.75
	std/ase	1.08	1.08	1.11

Parameter values: $(\alpha, \phi, \omega) = (-0.736, 0.90, 0.363)$.

If we compare Tables 2.5 and 2.6, then we notice that the data-based search for the optimal set of moment conditions comes at a relatively small efficiency cost compared to the case where the optimal set of moment conditions is known. The four-step procedure is also easy to implement and limits the bias via a parsimonious selection of moments. Moreover, it achieves high efficiency via the optimal selection of moments and its reliance on the analytically derived optimal weighting matrix.

2.7 Conclusion

The standard approach in the literature on GMM estimation of SV models has been to derive closed-form moment conditions from the expectations of $|y_t^r|$, $|y_{t_1} y_{t_2}|$ and $|y_{t_1}^2 y_{t_2}^2|$ for any r, t_1, and t_2. We have extended this class of conditions to include the expectation of $|y_{t_1}^{r_1} ... y_{t_p}^{r_p}|$ for arbitrary $r_1, ..., r_p$ and $t_1, ..., t_p$, and, following Wright (1999), the first two moments and the autocovariances of $\log y_t^2$. A closed-form expression for the optimal weighting matrix for any subset of those conditions has been derived and, as a by-product, an expression for the GMM asymptotic covariance matrix. These expressions can be used for improved GMM estimation of the SV model with AR(1) log-volatility and to compute GMM standard errors more accurately. It is also of interest to note that, upon redefining c_i, ν_r, κ_r, and ξ_r appropriately, all expressions are

generalised to SV models where the multiplicative shocks in the mean equation (1) are non-normal.

The comparison to other estimators showed the relatively small efficiency loss of the GMM estimator compared to the MCMC method. Monte Carlo results illustrate the efficiency gain of iterated GMM based on the analytical optimal weighting matrix compared to iterated GMM based on estimation of the optimal weighting matrix. The analytical results regarding the optimal weighting matrix allow us to fastly and accurately assess the information content of any subset of moment conditions considered. This, in turn, permits the optimal selection of a small set of highly informative moment conditions from very large sets, and subsequent GMM estimates to be based on the optimal selection of moments. We proposed a four-step data-based procedure for the optimal selection of moment conditions. It was found that the search for the optimal selection of moment conditions comes at a small efficiency cost compared to when this selection is known.

Appendix

Calculation of c_3 and c_4. For any positive integer n, upon substituting $t = x^2/2$,

$$\begin{aligned}\int_{-\infty}^{\infty}\left(\log\tfrac{x^2}{2}\right)^n\left(\tfrac{1}{\sqrt{2\pi}}\right)e^{-x^2/2}dx &= 2\int_0^{\infty}\left(\log\tfrac{x^2}{2}\right)^n\left(\tfrac{1}{\sqrt{2\pi}}\right)e^{-x^2/2}dx\\ &= \tfrac{1}{\sqrt{\pi}}\int_0^{\infty}(\log t)^n\, t^{-1/2}e^{-t}dt\\ &= \frac{\Gamma^{(n)}\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)},\end{aligned}$$

where $\Gamma^{(n)}(z)$ is the n-th derivative of $\Gamma(z)$. See Abramowitz and Stegun (1970) for properties and values of the gamma and related functions that are used below. Now, $c_1 = -\log 2 - \gamma = \psi\left(\frac{1}{2}\right) + \log 2$, where $\psi(z) = \frac{d}{dz}\log\Gamma(z) = \frac{\Gamma^{(1)}(z)}{\Gamma(z)}$. Hence, for $n = 3, 4$,

$$\begin{aligned}c_n &= \int_{-\infty}^{\infty}(\log x^2 - c_1)^n\left(\tfrac{1}{\sqrt{2\pi}}\right)e^{-x^2/2}dx\\ &= \int_{-\infty}^{\infty}\left(\log\tfrac{x^2}{2} - \psi\left(\tfrac{1}{2}\right)\right)^n\left(\tfrac{1}{\sqrt{2\pi}}\right)e^{-x^2/2}dx\\ &= g_n\left(\tfrac{1}{2}\right),\end{aligned}$$

where

$$g_n(z) = \sum_{i=0}^{n}\binom{n}{i}\frac{\Gamma^{(i)}(z)}{\Gamma(z)}\left(-\psi(z)\right)^{n-i}. \tag{2.12}$$

Taking successive derivatives of $\Gamma^{(1)}(z) = \Gamma(z)\psi(z)$ gives, upon rewriting,

$$\Gamma^{(2)}(z) = \Gamma(z)\left\{\psi'(z) + [\psi(z)]^2\right\},$$
$$\Gamma^{(3)}(z) = \Gamma(z)\left\{\psi''(z) + 3\psi'(z)\psi(z) + [\psi(z)]^3\right\},$$
$$\Gamma^{(4)}(z) = \Gamma(z)\left\{\psi'''(z) + 4\psi''(z) + 3\psi'(z)\left[\psi'(z) + 2[\psi(z)]^2\right] + [\psi(z)]^3\right\},$$

where primes denote derivatives. Substituting these expressions into (2.12) yields

$$g_3(z) = \psi''(z)$$

and

$$g_4(z) = \psi'''(z) + 3[\psi'(z)]^2.$$

Now, $\psi'\left(\frac{1}{2}\right) = \frac{\pi^2}{2}$, $\psi''\left(\frac{1}{2}\right) = -14\zeta(3)$, and $\psi'''\left(\frac{1}{2}\right) = \pi^4$, where $\zeta(3) = 1.202$. Hence $c_3 = -14\zeta(3) = -16.83$ and $c_4 = \frac{7}{4}\pi^4 = 170.5$.

The proof of Theorem 1 makes use of the following lemmas.

Lemma 3 *Let $X \sim N(0,1)$ and let a be a positive real number. Then*

$$\mathrm{Cov}(\log X^2, |X^a|) = \nu_a \kappa_a \tag{2.13}$$

and

$$\mathrm{Cov}((\log X^2 - c_1)^2, |X^a|) = \nu_a \xi_a. \tag{2.14}$$

Proof. Upon substituting $z = x^2/2$,

$$\begin{aligned}
\mathrm{Cov}(\log X^2, |X^a|) &= 2\int_0^\infty \left(\log x^2 - c_1\right) x^a \left(\tfrac{1}{\sqrt{2\pi}}\right) e^{-x^2/2} dx \\
&= \frac{2^{a/2}}{\sqrt{\pi}} \int_0^\infty (\log z + \log 2 - c_1)\, z^{(a-1)/2} e^{-z} dz \\
&= \frac{2^{a/2}}{\sqrt{\pi}} \Gamma\left(\frac{a+1}{2}\right)\left(\psi\left(\frac{a+1}{2}\right) + \log 2 - c_1\right) \\
&= \nu_a \kappa_a,
\end{aligned}$$

and, using $\frac{\Gamma''(z)}{\Gamma(z)} = \psi'(z) + (\psi(z))^2$ (with primes denoting derivatives),

$$\begin{aligned}
\mathrm{Cov}((\log X^2 - c_1)^2, |X^a|) &= 2\int_0^\infty \left[\left(\log x^2 - c_1\right)^2 - c_2\right] x^a \left(\tfrac{1}{\sqrt{2\pi}}\right) e^{-x^2/2} dx \\
&= \frac{2^{a/2}}{\sqrt{\pi}} \int_0^\infty \left[(\log z + \log 2 - c_1)^2 - c_2\right] z^{(a-1)/2} e^{-z} dz \\
&= \frac{2^{a/2}}{\sqrt{\pi}} \Gamma\left(\frac{a+1}{2}\right)\left(\frac{\Gamma''\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a+1}{2}\right)} + (\log 2 - c_1)^2 - c_2\right. \\
&\quad \left. + 2(\log 2 - c_1)\psi\left(\frac{a+1}{2}\right)\right) \\
&= \nu_a \xi_a.
\end{aligned}$$

Lemma 4 *Let X_1, X_2, and X_3 be jointly normal with $\mu_i = EX_i$ and $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$. Then*

$$\mathrm{Cov}(X_1, \exp X_3) = \sigma_{13} \exp\left(\mu_3 + \tfrac{1}{2}\sigma_{33}\right) \tag{2.15}$$

and

$$\mathrm{Cov}(X_1 X_2, \exp X_3) = (\sigma_{13}\sigma_{23} + \mu_1\sigma_{23} + \mu_2\sigma_{13}) \exp\left(\mu_3 + \tfrac{1}{2}\sigma_{33}\right). \tag{2.16}$$

Proof. Assume first that $\mu_i = 0$ and $\sigma_{ii} > 0$ for all i. Let $\mu_{i|j} = \sigma_{ij}\sigma_{jj}^{-1}X_j$ be the conditional mean of X_i, given X_j, and $\sigma_{ij|k} = \sigma_{ij} - \sigma_{ik}\sigma_{jk}\sigma_{kk}^{-1}$ the conditional covariance between X_i and X_j, given X_k. Then,

$$\begin{aligned}
\mathrm{Cov}(X_1, \exp X_3) &= \mathrm{Cov}(E(X_1|X_3), \exp X_3) \\
&= \sigma_{13}\sigma_{33}^{-1}\,\mathrm{Cov}(X_3, \exp X_3) \\
&= \sigma_{13} \exp\left(\tfrac{1}{2}\sigma_{33}\right), \\
\mathrm{Cov}(X_1 X_2, \exp X_3) &= \mathrm{Cov}(E(X_1 X_2|X_3), \exp X_3) \\
&= \mathrm{Cov}(\mu_{1|3}\mu_{2|3} + \sigma_{12|3}, \exp X_3) \\
&= \sigma_{13}\sigma_{23}\sigma_{33}^{-2}\,\mathrm{Cov}(X_3^2, \exp X_3) \\
&= \sigma_{13}\sigma_{23} \exp\left(\tfrac{1}{2}\sigma_{33}\right),
\end{aligned}$$

using the fact that, for a standard normal variate X,

$$\begin{aligned}
\mathrm{Cov}(X, \exp(bX)) &= \int_{-\infty}^{\infty} x e^{bx} \left(\tfrac{1}{\sqrt{2\pi}}\right) e^{-x^2/2} dx \\
&= \int_{-\infty}^{\infty} x \left(\tfrac{1}{\sqrt{2\pi}}\right) e^{-(x-b)^2/2 + b^2/2} dx \\
&= b \exp\left(\tfrac{1}{2}b^2\right)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}(X^2, \exp(bX)) &= \int_{-\infty}^{\infty} (x^2 - 1) e^{bx} \left(\tfrac{1}{\sqrt{2\pi}}\right) e^{-x^2/2} dx \\
&= \int_{-\infty}^{\infty} (x^2 - 1) \left(\tfrac{1}{\sqrt{2\pi}}\right) e^{-(x-b)^2/2 + b^2/2} dx \\
&= b^2 \exp\left(\tfrac{1}{2}b^2\right).
\end{aligned}$$

The extension to the case where $\mu_i \neq 0$ for some i is straightforward, and any degenerate case follows upon taking the appropriate limit in the non-degenerate case.

Proof of Theorem 1. Write $z_t = k_t + w_t$, where $k_t = h_t - \mu$ and $w_t = \log u_t^2 - c_1$. Then, w_t and k_t have zero mean and are independent, and, for any integers i, j, l, we have

$\mathrm{Cov}(k_t, k_{t-i}) = \phi^{|i|}\sigma^2$, $\mathrm{Cov}(k_t k_{t-i}, k_{t-j}) = 0$, and $\mathrm{Cov}(k_t k_{t-i}, k_{t-j}k_{t-l}) = (\phi^{|j|+|i-l|} + \phi^{|l|+|i-j|})\sigma^4$. Using these properties and the equalities

$$\sum_{l=-\infty}^{\infty} \phi^{|l|} = \frac{1+\phi}{1-\phi},$$
$$\sum_{l=-\infty}^{\infty} \phi^{|i+l|+|j+l|} = \phi^{|i-j|}\left(|i-j| + \frac{1+\phi^2}{1-\phi^2}\right),$$

we obtain, for $i, j \geq 0$,

$$\begin{aligned}
V(z_t, z_t) &= \sum_{l=-\infty}^{\infty} \left[\mathrm{Cov}(k_t, k_{t-l}) + \mathrm{Cov}(w_t, w_{t-l})\right] \\
&= \sum_{l=-\infty}^{\infty} \left(\phi^{|l|}\sigma^2 + I_{(l=0)}c_2\right) \\
&= \frac{1+\phi}{1-\phi}\sigma^2 + c_2, \\
V(z_t, z_t z_{t-j}) &= \sum_{l=-\infty}^{\infty} \mathrm{Cov}(w_t, w_{t-l}w_{t-j-l}) \\
&= \sum_{l=-\infty}^{\infty} I_{(l=j=0)}c_3 \\
&= I_{(j=0)}c_3,
\end{aligned}$$

and

$$\begin{aligned}
V(z_t z_{t-i}, z_t z_{t-j}) = \sum_{l=-\infty}^{\infty} &\left[\mathrm{Cov}(k_t k_{t-i}, k_{t-l}k_{t-j-l}) + \mathrm{Cov}(w_t k_{t-i}, k_{t-l}w_{t-j-l})\right. \\
&+ \mathrm{Cov}(w_t k_{t-i}, w_{t-l}k_{t-j-l}) + \mathrm{Cov}(k_t w_{t-i}, k_{t-l}w_{t-j-l}) \\
&\left.+ \mathrm{Cov}(k_t w_{t-i}, w_{t-l}k_{t-j-l}) + \mathrm{Cov}(w_t w_{t-i}, w_{t-l}w_{t-j-l})\right] \\
= \sum_{l=-\infty}^{\infty} &\left[\left(\phi^{|l|+|j-i+l|} + \phi^{|j+l|+|l-i|}\right)\sigma^4 + I_{(l=-j)}\phi^{|i+j|}c_2\sigma^2\right. \\
&+ I_{(l=0)}\phi^{|i-j|}c_2\sigma^2 + I_{(l=i-j)}\phi^{|i-j|}c_2\sigma^2 + I_{(l=i)}\phi^{|i+j|}c_2\sigma^2 \\
&\left.+ I_{(i=j\neq 0)}I_{(l=0)}c_2^2 + I_{(i=j=0)}I_{(l=0)}\left(c_4 - c_2^2\right)\right] \\
= &\left(|i-j|\,\phi^{|i-j|} + |i+j|\,\phi^{|i+j|} + \left(\phi^{|i-j|} + \phi^{|i+j|}\right)\frac{1+\phi^2}{1-\phi^2}\right)\sigma^4 \\
&+ 2\left(\phi^{|i-j|} + \phi^{|i+j|}\right)c_2\sigma^2 + I_{(i=j\neq 0)}c_2^2 + I_{(i=j=0)}(c_4 - c_2^2),
\end{aligned}$$

giving (3.1)–(3.3). To establish (2.9), recall the definition of $Y^{r_1,\ldots,r_p}_{t_1,\ldots,t_p}$, from which

$$\begin{aligned}
&\operatorname{Cov}\left(Y^{r_1,\ldots,r_p}_{t_1,\ldots,t_p}, Y^{r_{p+1},\ldots,r_{p+q}}_{t_{p+1}-l,\ldots,t_{p+q}-l}\right) \\
&\quad = E\left(Y^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} Y^{r_{p+1},\ldots,r_{p+q}}_{t_{p+1}-l,\ldots,t_{p+q}-l}\right) - 1 \\
&\quad = E\exp\left(-\delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} - \delta^{r_{p+1},\ldots,r_{p+q}}_{t_{p+1},\ldots,t_{p+q}} + \frac{1}{2}\sum_{j=1}^{p} r_j h_{t_j} + \frac{1}{2}\sum_{j=p+1}^{p+q} r_j h_{t_j-l}\right) \\
&\qquad \times E\left(\prod_{j=1}^{p}\left|\nu^{-1}_{r_j} u^{r_j}_{t_j}\right| \prod_{j=p+1}^{p+q}\left|\nu^{-1}_{r_j} u^{r_j}_{t_j-l}\right|\right) - 1.
\end{aligned}$$

Now,

$$\begin{aligned}
&E\exp\left(\frac{1}{2}\sum_{j=1}^{p} r_j h_{t_j} + \frac{1}{2}\sum_{j=p+1}^{p+q} r_j h_{t_j-l}\right) \\
&\quad = \exp\left(\delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} + \delta^{r_{p+1},\ldots,r_{p+q}}_{t_{p+1},\ldots,t_{p+q}} + \frac{\sigma^2}{4}\sum_{j=1}^{p}\sum_{j'=p+1}^{p+q} r_j r_{j'} \phi^{|t_j - t_{j'}+l|}\right),
\end{aligned}$$

so

$$\begin{aligned}
\operatorname{Cov}\left(Y^{r_1,\ldots,r_p}_{t_1,\ldots,t_p}, Y^{r_{p+1},\ldots,r_{p+q}}_{t_{p+1}-l,\ldots,t_{p+q}-l}\right) &= (B_l+1)(C_l+1) - 1 \\
&= B_l + (B_l+1)C_l.
\end{aligned}$$

Summing over l gives (2.9), because $C_l = 0$ whenever $l \notin L$. Furthermore, by (2.13) and (2.15),

$$\begin{aligned}
&\operatorname{Cov}\left(z_t, Y^{r_1,\ldots,r_p}_{t_1-l,\ldots,t_p-l}\right) \\
&\quad = E\left(k_t \exp\left(-\delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} + \frac{1}{2}\sum_{j=1}^{p} r_j h_{t_j-l}\right)\right) \\
&\qquad + \left(E\exp\left(-\delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} + \frac{1}{2}\sum_{j=1}^{p} r_j h_{t_j-l}\right)\right) E\left(w_t \prod_{j=1}^{p}\left|\nu^{-1}_{r_j} u^{r_j}_{t_j-l}\right|\right) \\
&\quad = \frac{\sigma^2}{2}\sum_{j=1}^{p} r_j \phi^{|t-t_j+l|} + \sum_{j=1}^{p} I_{(t=t_j-l)}\kappa_{r_j},
\end{aligned}$$

which, upon summing over l, gives (2.10). Finally,

$$\begin{aligned}
\operatorname{Cov}\left(z_t z_{t-i}, Y^{r_1,\ldots,r_p}_{t_1-l,\ldots,t_p-l}\right) &= \operatorname{Cov}\Bigg(k_t k_{t-i} + k_t w_{t-i} + w_t k_{t-i} + w_t w_{t-i}, \\
&\qquad \exp\left(-\delta^{r_1,\ldots,r_p}_{t_1,\ldots,t_p} + \frac{1}{2}\sum_{j=1}^{p} r_j h_{t_j-l}\right) \prod_{j=1}^{p}\left|\nu^{-1}_{r_j} u^{r_j}_{t_j-l}\right|\Bigg) \\
&= T_1 + T_2 + T_3 + T_4,
\end{aligned}$$

say, with, using Lemma 2 and Lemma 3,

$$
\begin{aligned}
T_1 &= \left(\frac{\sigma^2}{2}\sum_{j=1}^{p} r_j \phi^{|t-t_j+l|}\right)\left(\frac{\sigma^2}{2}\sum_{j=1}^{p} r_j \phi^{|t-i-t_j+l|}\right), \\
T_2 &= \left(\frac{\sigma^2}{2}\sum_{j=1}^{p} r_j \phi^{|t-t_j+l|}\right)\sum_{j=1}^{p} I_{(t-i=t_j-l)}\kappa_{r_j}, \\
T_3 &= \left(\frac{\sigma^2}{2}\sum_{j=1}^{p} r_j \phi^{|t-i-t_j+l|}\right)\sum_{j=1}^{p} I_{(t=t_j-l)}\kappa_{r_j}, \\
T_4 &= I_{(i=0)}\sum_{j=1}^{p} I_{(t=t_j-l)}\xi_{r_j} + I_{(i\neq 0)}\sum_{j,j'=1}^{p} I_{(t=t_j-l)}I_{(t-i=t_{j'}-l)}\kappa_{r_j}\kappa_{r_{j'}}.
\end{aligned}
$$

Summing over l gives (2.11), which concludes the proof.

Proof of Lemma 1. Since $|\phi|^{|t_j-t_{j'}+l|} \leq |\phi|^{-|t_j-t_{j'}|+|l|}$,

$$
B_l \leq \exp\left(\frac{\sigma^2}{4}\sum_{j=1}^{p}\sum_{j'=p+1}^{p+q} r_j r_{j'}|\phi|^{|t_j-t_{j'}+l|}\right) - 1 \leq \exp(a|\phi|^{|l|}) - 1.
$$

By an argument of symmetry,

$$
\exp(-a|\phi|^{|l|}) - 1 \leq B_l \leq \exp(a|\phi|^{|l|}) - 1
$$

and so, because $\exp(z) - 1 \geq 1 - \exp(-z)$ for any z,

$$
|B_l| \leq \exp(a|\phi|^{|l|}) - 1.
$$

Therefore,

$$
\begin{aligned}
|B - B(I)| &\leq \sum_{l=I+1}^{\infty}(|B_{-l}| + |B_l|) \leq 2\sum_{l=I}^{\infty}\left(\exp\left(a|\phi|^{l}\right) - 1\right) \\
&= 2\sum_{l=I}^{\infty}\sum_{k=1}^{\infty}\frac{a^k|\phi|^{kl}}{k!} = 2\sum_{k=1}^{\infty}\left(\frac{a^k}{k!}\right)\left(\frac{|\phi|^{kI}}{1-|\phi|^k}\right) \\
&< \frac{2}{1-|\phi|}\sum_{k=1}^{\infty}\frac{\left(a|\phi|^I\right)^k}{k!} = 2\frac{\exp(a|\phi|^I) - 1}{1-|\phi|}.
\end{aligned}
$$

Chapter 3

Quantile-Based Estimation of Volatility

3.1 Introduction

Let P_t be the price of an asset at time t. Assume that $p_t = \log(P_t)$ is a driftless Brownian motion with constant diffusion coefficient σ, hence following the stochastic differential equation

$$dp_t = \sigma dB_t,$$

where B_t is a standard Brownian motion. For $0 \leq t \leq 1$, let b_t be the Brownian bridge corresponding to p_t, that is,

$$b_t = p_t - tp_1 = \sigma B_t - t\sigma B_1 = \sigma \left(B_t - tB_1 \right).$$

This chapter develops unbiased estimators of σ and σ^2 based on quantiles of p_t and b_t, $0 \leq t \leq 1$.

If the price process is observed in continuous time, then quadratic variation offers a simple consistent estimator of σ^2, which quantile-based estimators cannot beat. However, if the 'true' price process is perturbed by noise and hence not observed, then quadratic variation is biased and the returns need to be sampled at a lower frequency to make the sum of squared returns more robust. Quantiles are more robust to noise and can still be based on all observations. Moreover, for any given time interval, quantile-based estimators are often more efficient. That is, the return measured over a given time interval is less efficient than e.g. the range over the same interval (Parkinson, 1980). Thus, in a setting with high-frequency observations perturbed by noise, it becomes worthwhile to consider quantile-based estimators. In this chapter, it is assumed that p_t is observed without noise. Noise is introduced in the next chapter.

So far, the only existing quantile-based estimators for the diffusion coefficient are the range-based estimators, where the range is the difference between the supremum and the infimum of either p_t or b_t. The distribution of the range of p_t was derived by Feller (1951) and was introduced as an asset price volatility estimator by Parkinson (1980). From then on range-based estimators received considerable attention in applications and several extensions have been proposed; see e.g. Garman and Klass (1980), Rogers and Satchell (1991) and Alizadeh et al. (2001). However, as shown in this chapter, other quantiles than the supremum and the infimum can also be used to estimate volatility.

Volatility estimators are receiving new attention since ultra-high frequency data became available, because volatility estimators that exploit the information present in these data achieve higher efficiency. The realized volatility estimator, the sum of squared high frequency returns, was proposed as the first consistent estimator of this kind; see e.g. Andersen et al. (2001). However, much of the literature that followed deals with correcting the realized volatility estimator for the bias that market microstructure noise introduces at high frequency. At the same time, the range has been found to be fairly robust to such noise (see e.g. Alizadeh et al. (2002)), a property that holds for all quantiles. This finding motivates the search for quantile-based estimators that exploit this property. One approach is the realized range estimator of Christensen and Podolski (2005) and Martens and van Dijk (2007), which replaces the squared returns of the realized volatility estimator by squared ranges. This estimator is consistent and more efficient than realized volatility for a given sampling scheme. The estimator is based on the range measured over very short intervals, at which quantiles are seriously biased by market microstructure noise. Therefore, the estimators require bias-corrections, which the authors also provide. In this chapter, I present a generalization of their consistent range-based estimator, by using more quantiles in the estimation. Furthermore, I present a new consistent estimator that is based on permutations of subintervals. This estimator promises to be fairly robust to market microstructure noise, because it uses quantiles measured over the unit interval.

This chapter presents the estimators assuming ideal conditions of continuous-time observations and in the absence of noise. An assesment of the new estimators in discrete time and in the presence of market microstructure noise is provided in the next chapter. The chapter is structured as follows. In Section 3.2, I develop unbiased estimators of σ and σ^2 based on quantiles of a Brownian motion. Section 3.3 does the same based on quantiles of a Brownian bridge. In Section 3.4, I show how consistent estimators are obtained from the unbiased estimators. Section 3.5 concludes.

3.2 Unbiased estimators based on quantiles of Brownian motion

3.2.1 Quantiles of Brownian motion

I start from the following distributional results for Brownian quantiles. For $\alpha \in [0,1]$ let

$$M(\alpha) = \begin{cases} \inf\limits_{0\le t\le 1} p_t & \alpha = 0 \\ \inf\left\{x : \int_0^1 1_{(p_t \le x)} dt \ge \alpha\right\} & 0 < \alpha \le 1 \end{cases}$$

be the α-quantile of p_t, $t \in [0,1]$. The density of $M(\alpha)$ was derived by Yor (1995) and for a Brownian motion with drift by Dassios (1995). Dassios (1995) also obtained the identity in law

$$M(\alpha) \overset{(law)}{=} \sup_{0\le t\le \alpha} p_t + \inf_{0\le t\le 1-\alpha} \tilde{p}_t,$$

where $\tilde{p}_t$ is an independent copy of p_t. Wendel (1960) already stated this identity in characteristic function form for the random walk case and it was extended with the time the quantile is reached by Port (1963). Due to the symmetry of Brownian motion it also holds that

$$\inf_{0\le t\le 1-\alpha} \tilde{p}_t \overset{(law)}{=} - \sup_{0\le t\le 1-\alpha} \tilde{p}_t.$$

Thus

$$M(\alpha) \overset{(law)}{=} \sup_{0\le t\le \alpha} p_t - \sup_{0\le t\le 1-\alpha} \tilde{p}_t. \tag{3.1}$$

Further, it is known that

$$\sup_{0\le t\le 1} p_t \sim \left|N\left(0,\sigma^2\right)\right|,$$

and so, by the scaling property of Brownian motion,

$$\sup_{0\le t\le \alpha} p_t \sim \left|N\left(0,\alpha\sigma^2\right)\right|.$$

Hence, for $i \ge 1$,

$$E\left(\sup_{0\le t\le \alpha} p_t\right)^i = \Gamma\left(\frac{i+1}{2}\right)\sqrt{\frac{(2\alpha\sigma^2)^i}{\pi}}, \tag{3.2}$$

where $\Gamma(\cdot)$ is the Gamma function. The moments of $M(\alpha)$ follow easily from (3.1) and (3.2), e.g.

$$E\left[M(\alpha)\right] = \sqrt{\frac{2}{\pi}}\sigma\left(\sqrt{\alpha}-\sqrt{1-\alpha}\right), \tag{3.3}$$

$$\mathrm{Var}(M(\alpha)) = \left(1-\frac{2}{\pi}\right)\sigma^2, \tag{3.4}$$

$$E\left[M(\alpha)^2\right] = \left(1-\frac{4\sqrt{\alpha(1-\alpha)}}{\pi}\right)\sigma^2, \tag{3.5}$$

$$\mathrm{Var}(M(\alpha)^2) = \left[3-\left(1+\frac{4\sqrt{\alpha(1-\alpha)}}{\pi}\right)^2\right]\sigma^4. \tag{3.6}$$

Invoking the scaling property of Brownian motion again, for $\alpha_1, \alpha_2 \in [0,1]$, it holds that

$$\mathrm{Cov}\left(M(\alpha_1), M(\alpha_2)\right) = C\left(\alpha_1, \alpha_2\right)\sigma^2, \tag{3.7}$$

where $C(\alpha_1, \alpha_2)$ is the covariance between the α_1-quantile and the α_2-quantile of a standard Brownian motion, i.e. with $\sigma = 1$. Obtaining $C(\alpha_1, \alpha_2)$ in closed form for $\alpha_1 \neq \alpha_2$ is difficult. A direct way to derive these covariances would be to use the joint density of two quantiles. This density has been studied by Fujita (2000), but it appears not to have a closed form. I have not been able to derive a closed-form expression for $C(\alpha_1, \alpha_2)$. However, the covariances can be obtained by simulation. I computed $C(\alpha_1, \alpha_2)$ by simulating 1 million standard Brownian motions, each with 1 million increments on the unit interval. Furthermore, we do have an expression for the variance of any quantile, i.e. (3.4), and we can derive the covariance between $M(1)$ and $M(0)$ from the variance of the range and the variance of a quantile. Parkinson (1980), based on Feller (1951), shows that the range,

$$R = M(1) - M(0),$$

has variance

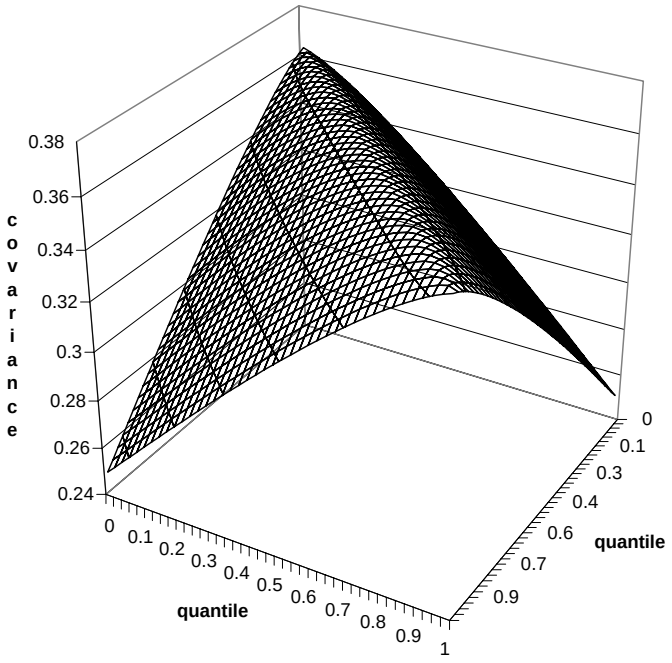
$$\mathrm{Var}(R) = 4\left(\log 2 - \frac{2}{\pi}\right)\sigma^2. \tag{3.8}$$

Hence, because $\mathrm{Var}(M(1)) = \mathrm{Var}(M(0))$, we get $\mathrm{Var}(R) = 2\,\mathrm{Var}(M(1)) - 2C(0,1)\sigma^2$ and so

$$\begin{aligned} C(0,1) &= \left(\mathrm{Var}(M(1)) - \mathrm{Var}(R)/2\right)\sigma^{-2} \\ &= 1 - \frac{2}{\pi} - 2\left(\log 2 - \frac{2}{\pi}\right) \\ &= 1 - 2\left(\log 2 - \frac{1}{\pi}\right). \end{aligned}$$

Moreover, by the symmetry of Brownian motion, $C(\alpha_i, \alpha_j) = C(1-\alpha_i, 1-\alpha_j) = C(\alpha_j, \alpha_i) = C(1-\alpha_j, 1-\alpha_i)$. In order to avoid small discrepancies between the simulation-based estimates of these covariances, I replace each by their average. The covariances between Brownian quantiles, thus obtained by simulation, are graphed in Figure 3.1. The symmetry also implies that for a quick estimate of the covariance matrix only one in four covariances needs to be computed.

Figure 3.1: Covariances between Brownian quantiles ($\sigma = 1$)



The following subsections present quantile-based unbiased estimators of σ and σ^2, respectively. The estimators are method of moment or generalized method of moment estimators, derived from (3.3), (3.5) and similar moment conditions.

3.2.2 Quantile-based unbiased estimators of σ

Unbiased estimators of σ can be developed from one or more Brownian quantiles. I start with an estimator based on one quantile and then present extensions.

One quantile

Except for the median, all quantiles are informative for σ. For any $\alpha \in [0,1]$ except $\alpha = 1/2$, an unbiased estimator of σ follows from (3.3) as

$$\tilde{\sigma}(\alpha) = \sqrt{\frac{\pi}{2}} \frac{M(\alpha)}{\sqrt{\alpha} - \sqrt{1-\alpha}}.$$

Using (3.4), it follows that

$$\mathrm{Var}(\tilde{\sigma}(\alpha)) = \frac{(\pi - 2)\,\sigma^2}{2\left(\sqrt{\alpha} - \sqrt{1-\alpha}\right)^2}.$$

This variance is minimized for $\alpha = 0$ and for $\alpha = 1$, yielding $\mathrm{Var}(\tilde{\sigma}(0)) = \mathrm{Var}(\tilde{\sigma}(1)) = (\pi/2 - 1)\,\sigma^2 = 0.571\sigma^2$. Thus $\tilde{\sigma}(0)$ and $\tilde{\sigma}(1)$ are the minimum variance unbiased estimators based on moment condition (3.3) for a single quantile.

Several quantiles

It is possible to construct unbiased estimators of σ based on more than one quantile. The best-known estimator of this type is the range, i.e. the difference between the 'high', $M(1) = \max_{0 \le t \le 1} p_t$, and the 'low', $M(0) = \min_{0 \le t \le 1} p_t$, over the trading day. This estimator has received considerable attention and 'high-low' figures are reported in business newspapers as an indication of volatility since long. So far, no attention has been paid to the fact that, next to the range, we can also consider any other combination of quantiles.

Fix a set of distinct values $\alpha_1, ..., \alpha_n \in [0,1]$, with corresponding quantiles $M(\alpha_1), ..., M(\alpha_n)$ and unbiased estimates $\tilde{\sigma}(\alpha_1), ..., \tilde{\sigma}(\alpha_n)$. Let $\tilde{s} = (\tilde{\sigma}(\alpha_1), ..., \tilde{\sigma}(\alpha_n))'$ and let ι be an $n \times 1$ vector of ones. Then, for any vector of weights w satisfying $\iota' w = 1$, the estimator

$$\hat{\sigma}(w) = w'\tilde{s}$$

is unbiased for σ. This estimator has variance $w'Vw$, where $V = \mathrm{Var}(\tilde{s})$ with (i,j)-th element given by

$$V_{ij} = \frac{\pi C(\alpha_i, \alpha_j)\sigma^2}{2\left(\sqrt{\alpha_i} - \sqrt{1-\alpha_i}\right)\left(\sqrt{\alpha_j} - \sqrt{1-\alpha_j}\right)}.$$

This variance is minimized by taking $w = (\iota' V^{-1} \iota)^{-1} V^{-1} \iota$, yielding

$$\hat{\sigma} = (\iota' V^{-1} \iota)^{-1} \iota' V^{-1} \tilde{s}$$

as the minimum-variance unbiased estimator of σ based on condition (3.3) for the chosen quantiles, with variance

$$\mathrm{Var}\,(\hat{\sigma}) = (\iota' V^{-1} \iota)^{-1}.$$

The optimal weights are scale invariant, which implies for applications that w can be computed without knowing σ.

Table 3.1 presents the variance of $\hat{\sigma}$ for different choices of $\alpha_1, ..., \alpha_n$. The elements α_i are each time chosen to be equally spaced over the interval $[0, 1]$, but with $1/2$ removed, because $M(1/2)$ is uninformative for σ. For example, in the case of 4 quantiles $\alpha_1, ..., \alpha_4 = 0, 0.25, 0.75, 1$. A comparison of the variances of the optimal one-quantile estimators, $\text{Var}(\tilde{\sigma}(0)) = \text{Var}(\tilde{\sigma}(1)) = 0.571\sigma^2$, with the variances reported in Table 3.1, shows that using more than one quantile dramatically improves the efficiency of the estimator.

Table 3.1: Variance of $\hat{\sigma}$

Number of quantiles		2	4	6	10	20	40
$\text{Var}(\hat{\sigma})$	$\sigma^2\times$	0.089	0.074	0.071	0.061	0.053	0.047

Notes: The quantiles are equally spaced, e.g. 2 = high and low, 4 = quartiles, etc. The median is not used, because it is uninformative.

As expected the efficiency increases with the number of quantiles at a decreasing rate, because the smaller the distance between the quantiles is, the less additional information they contain.

A special case: Interquantile ranges

Let $0 \leq \alpha < 1/2$ and consider $\tilde{\sigma}(\alpha)$ and $\tilde{\sigma}(1-\alpha)$. The optimal weights to combine the estimates $\tilde{\sigma}(\alpha)$ and $\tilde{\sigma}(1-\alpha)$ are

$$w = V^{-1}\iota(\iota' V^{-1}\iota)^{-1} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix},$$

because the diagonal elements of V are equal. Thus, the minimum-variance unbiased linear combination of $\tilde{\sigma}(\alpha)$ and $\tilde{\sigma}(1-\alpha)$ is the interquantile range (IQR) based estimator

$$\hat{\sigma}_{IQR}(\alpha) = \frac{\sqrt{\pi}\left[M(1-\alpha) - M(\alpha)\right]}{\sqrt{8}\left(\sqrt{1-\alpha} - \sqrt{\alpha}\right)}; \qquad \alpha \in [0, 0.5[,$$

and its variance can be derived using (3.4) and (3.7) as

$$\text{Var}\left(\hat{\sigma}_{IQR}(\alpha)\right) = \frac{\pi(1 - C(\alpha, 1-\alpha)) - 2}{4\left(1 - 2\sqrt{\alpha(1-\alpha)}\right)}\sigma^2.$$

A special case is Parkinson's (1980) range-based estimator

$$\hat{\sigma}_R = \sqrt{\frac{\pi}{8}}\left[M(1) - M(0)\right],$$

which, by (3.8), has variance

$$\mathrm{Var}(\hat{\sigma}_R) = \left(\frac{\pi}{2}\log 2 - 1\right)\sigma^2 = 0.0889\sigma^2,$$

which corresponds with the simulation result in Table 3.1.

Table 3.2 presents the variances of $\hat{\sigma}_{IQR}(\alpha)$ for various choices of α. It is clear that the further apart the quantiles, the higher the efficiency. The range (corresponding to $\alpha = 0$) provides the most efficient IQR-based estimator. Yet, as Table 3.1 shows, the range itself is less efficient than estimators that are based on many quantiles.

Table 3.2: Variance of $\hat{\sigma}_{IQR}(\alpha)$

α		0.000	0.025	0.050	0.100	0.125	0.150	0.175	0.200	0.225
$\mathrm{Var}\left(\hat{\sigma}_{IQR}(\alpha)\right)$	$\sigma^2\times$	0.089	0.122	0.139	0.170	0.185	0.200	0.217	0.234	0.253
α		0.250	0.275	0.300	0.350	0.375	0.400	0.425	0.450	0.475
$\mathrm{Var}\left(\hat{\sigma}_{IQR}(\alpha)\right)$	$\sigma^2\times$	0.274	0.297	0.322	0.385	0.427	0.478	0.547	0.645	0.817

Absolute quantiles

Instead of developing an estimator based on raw quantiles, one can also consider absolute quantiles. From the density of a quantile as presented in Yor (1995), it follows that

$$E\left[|M(\alpha)|\right] = \sqrt{\frac{2}{\pi}}\sigma\left(2 - \sqrt{\alpha} - \sqrt{1-\alpha}\right), \tag{3.9}$$

$$\mathrm{Var}(|M(\alpha)|) = \left(1 - \frac{2}{\pi} - \frac{8}{\pi}\left(1-\sqrt{\alpha}\right)\left(1-\sqrt{1-\alpha}\right)\right)\sigma^2. \tag{3.10}$$

For any $\alpha \in [0,1]$ an unbiased estimator of σ follows from (3.9) as

$$\breve{\sigma}(\alpha) = \sqrt{\frac{\pi}{2}}\frac{|M(\alpha)|}{\left(2 - \sqrt{\alpha} - \sqrt{1-\alpha}\right)},$$

and using (3.10) it follows that

$$\mathrm{Var}(\breve{\sigma}(\alpha)) = \left(\frac{\pi - 4\sqrt{\alpha(1-\alpha)}}{2\left(2 - \sqrt{\alpha} - \sqrt{1-\alpha}\right)^2} - 1\right)\sigma^2.$$

The variance is minimized for $\alpha = 0$ and $\alpha = 1$, and equal to the case when raw quantiles are used as the estimator is the same for the extrema. For any other quantile, the estimator based on one absolute quantile is more efficient than its raw quantile counterpart. Note that the median is now also informative. However, the estimator based on the optimal combination of several absolute quantiles performs worse than

its raw quantile counterpart. For example, the variance of the estimator based on 41 absolute quantiles (including the absolute median) has a variance of $0.062\sigma^2$ compared to $0.047\sigma^2$ in the case of raw quantiles (see Table 3.1). Therefore, this chapter focusses on raw quantiles instead of absolute quantiles.

3.2.3 Quantile-based unbiased estimators of σ^2

Unbiased estimators of σ^2 can be constructed in a similar manner as for σ. For $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in [0,1]$ it holds, by the scaling property of Brownian motion, that

$$\mathrm{Cov}\left(M(\alpha_1)M(\alpha_2), M(\alpha_3)M(\alpha_4)\right) = D(\alpha_1,\alpha_2,\alpha_3,\alpha_4)\sigma^4, \tag{3.11}$$

where $D(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ is the covariance between the product of the α_1- and α_2-quantiles and the product of the α_3- and α_4-quantiles of a standard Brownian motion. Obtaining $D(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ is difficult, except for $D(\alpha_1,\alpha_1,\alpha_1,\alpha_1) = \mathrm{Var}(M(\alpha_1)^2)$, which is given by (3.6). The other covariances can be obtained by simulation. I computed $D(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ by simulating 1 million standard Brownian motions, each with 1 million increments on the unit interval. Furthermore, due to the symmetry of Brownian motion,

$$D\left(\alpha_1,\alpha_2,\alpha_3,\alpha_4\right) = D(1-\alpha_1, 1-\alpha_2, 1-\alpha_3, 1-\alpha_4). \tag{3.12}$$

Therefore I replaced the estimates of both sides of (3.12) by their average to avoid small discrepancies.

The product of two quantiles

Let $\alpha_i, \alpha_j \in [0,1]$ and consider the quantiles $M(\alpha_i)$ and $M(\alpha_j)$. From (3.3) and (3.7) it follows that

$$\begin{aligned} E\left[M(\alpha_i)M(\alpha_j)\right] &= \mathrm{Cov}(M(\alpha_i), M(\alpha_j)) + E\left[M(\alpha_i)\right]E\left[M(\alpha_j)\right] \\ &= C(\alpha_i,\alpha_j)\sigma^2 + \frac{2}{\pi}\left(\sqrt{\alpha_i} - \sqrt{1-\alpha_i}\right)\left(\sqrt{\alpha_j} - \sqrt{1-\alpha_j}\right)\sigma^2. \end{aligned} \tag{3.13}$$

The expected value of the product of two quantiles is depicted in the left panel of Figure 3.2 for the case $\sigma = 1$.

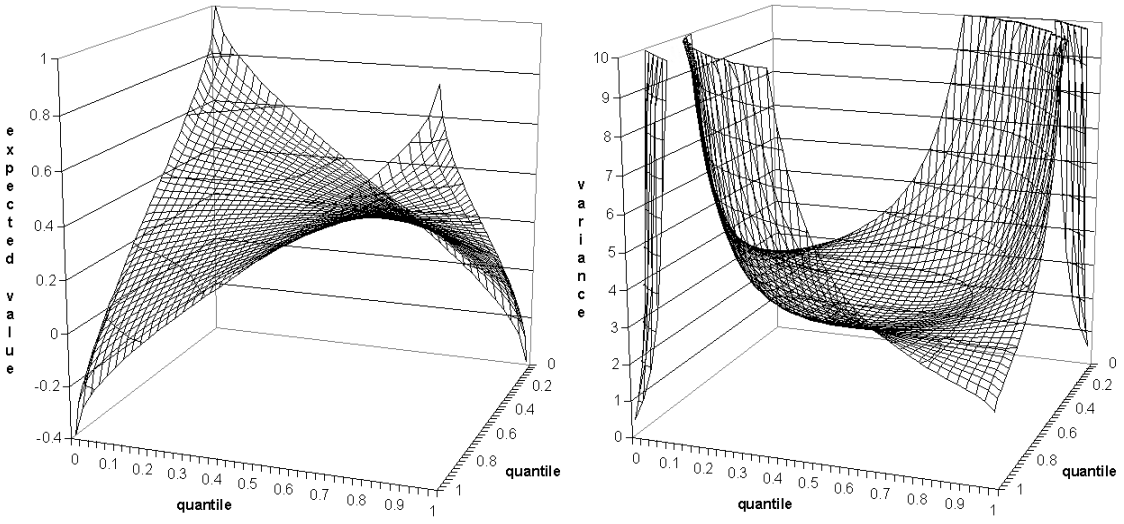
Provided α_i and α_j are such that $E\left[M(\alpha_i)M(\alpha_j)\right] \neq 0$, an unbiased estimator of σ^2 follows as

$$\tilde{\sigma}^2(\alpha_i,\alpha_j) = \frac{M(\alpha_i)M(\alpha_j)}{C(\alpha_i,\alpha_j) + \frac{2}{\pi}\left(\sqrt{\alpha_i} - \sqrt{1-\alpha_i}\right)\left(\sqrt{\alpha_j} - \sqrt{1-\alpha_j}\right)}, \tag{3.14}$$

with variance

$$\mathrm{Var}\left(\tilde{\sigma}^2(\alpha_i,\alpha_j)\right) = \frac{D(\alpha_i,\alpha_j,\alpha_i,\alpha_j)\sigma^4}{\left[C(\alpha_i,\alpha_j) + \frac{2}{\pi}\left(\sqrt{\alpha_i} - \sqrt{1-\alpha_i}\right)\left(\sqrt{\alpha_j} - \sqrt{1-\alpha_j}\right)\right]^2}.$$

Figure 3.2: Expected value of the product of two quantiles (left panel) and the variance of $\tilde{\sigma}^2(\alpha_i,\alpha_j)$ (right panel) when $\sigma = 1$



Combinations of α_i and α_j for which $E\left[M(\alpha_i)M(\alpha_j)\right]$ is zero or close to zero have to be avoided, because then $\tilde{\sigma}^2(\alpha_i,\alpha_j)$ is either not defined or has a very large variance. However, since $C(\alpha_i,\alpha_j)$ is not known analytically, neither is an expression describing these combinations. Thus one has to resort to numerical analysis.

The variance of $\tilde{\sigma}^2(\alpha_i,\alpha_j)$ can be computed using the numerical results for $C(\alpha_i,\alpha_j)$ and $D(\alpha_i,\alpha_j,\alpha_i,\alpha_j)$. For $\sigma = 1$, the right panel of Figure 3.2 shows the variances for different choices of $\alpha_i,\alpha_j \in [0,1]$. It turns out that $\tilde{\sigma}^2(0,1)$ is the minimum variance unbiased estimator, with variance $0.524\sigma^4$. Obviously, if $E\left[M(\alpha_i)M(\alpha_j)\right]$ is close to zero, then $\tilde{\sigma}^2(\alpha_i,\alpha_j)$ is very inefficient, e.g. $E\left[M(0)M(0.8)\right] = -0.0078\sigma^2$ and $\tilde{\sigma}^2(0,0.8)$ has variance $3401\sigma^4$.

For the one-quantile case ($\alpha_i = \alpha_j$), by (3.5) the estimator is

$$\tilde{\sigma}^2(\alpha_i,\alpha_i) = \frac{M(\alpha_i)^2}{1 - \frac{4\sqrt{\alpha_i(1-\alpha_i)}}{\pi}}. \tag{3.15}$$

Upon rewriting (3.6) as

$$\mathrm{Var}(M(\alpha)^2) = \left[3 - \frac{16\sqrt{\alpha(1-\alpha)}}{\pi} - \left(1 - \frac{4\sqrt{\alpha(1-\alpha)}}{\pi}\right)^2\right]\sigma^4,$$

it follows that

$$\text{Var}\left(\tilde{\sigma}^2(\alpha_i,\alpha_i)\right) = \left[\frac{3-\frac{16\sqrt{\alpha_i(1-\alpha_i)}}{\pi}}{\left(1-\frac{4\sqrt{\alpha_i(1-\alpha_i)}}{\pi}\right)^2}-1\right]\sigma^4.$$

This variance is minimized for $\alpha_i = 0$ and for $\alpha_i = 1$, yielding $\tilde{\sigma}^2(0,0) = M(0)^2$ and $\tilde{\sigma}^2(1,1) = M(1)^2$ as the minimum variance unbiased estimators of σ^2 (within the class defined by (3.15)), with $\text{Var}\left(\tilde{\sigma}^2(0,0)\right) = \text{Var}\left(\tilde{\sigma}^2(1,1)\right) = 2\sigma^4$.

Several products of two quantiles

Similar to Subsection 3.2.2, we can construct optimal unbiased estimators of σ^2 as linear combinations of estimators given by (3.14). Any set of distinct values $\alpha_1, ..., \alpha_n \in [0,1]$ corresponds with a set of $n(n+1)/2$ distinct estimators $\tilde{\sigma}^2(\alpha_i,\alpha_j)$, where $i = 1, ..., n$ and $j = 1, ..., i$. Let $\tilde{s}^2$ be the vector of these estimators. Then the minimum-variance unbiased estimator of σ^2 (within the class of linear combinations of those $\tilde{\sigma}^2(\alpha_i,\alpha_j)$) is

$$\hat{\sigma}^2 = (\iota' V^{-1}\iota)^{-1}\iota' V^{-1}\tilde{s}^2, \tag{3.16}$$

where $V = \text{Var}(\tilde{s}^2)$ has elements

$$\text{Cov}\left(\tilde{\sigma}^2(\alpha_i,\alpha_j),\tilde{\sigma}^2(\alpha_k,\alpha_l)\right) = d(\alpha_i,\alpha_j)D(\alpha_i,\alpha_j,\alpha_k,\alpha_l)d(\alpha_k,\alpha_l),$$

where

$$d(\alpha_i,\alpha_j) = \frac{1}{C(\alpha_i,\alpha_j)+\frac{2}{\pi}\left(\sqrt{\alpha_i}-\sqrt{1-\alpha_i}\right)\left(\sqrt{\alpha_j}-\sqrt{1-\alpha_j}\right)}.$$

Table 3.3 presents the variance of $\hat{\sigma}^2$ for different choices of $\alpha_1, ..., \alpha_n$. It shows how the efficiency increases as the number of quantiles involved increases.

Table 3.3: Variance of $\hat{\sigma}^2$

Number of quantiles	2	3	5	6	11	21	41
$\text{Var}(\hat{\sigma}^2)$ $\sigma^4\times$	0.341	0.330	0.267	0.253	0.214	0.183	0.159

Note: The quantiles are equally spaced.

Interquantile ranges

Unbiased IQR-based estimators of σ^2 are easily derived. Let $0 \leq \alpha < 1/2$. Then, by symmetry,

$$\begin{aligned} E\left[M(1-\alpha)-M(\alpha)\right]^2 &= 2E\left(M(\alpha)^2\right)-2E\left(M(\alpha)M(1-\alpha)\right) \\ &= \left(2+\frac{4}{\pi}-\frac{16}{\pi}\sqrt{\alpha(1-\alpha)}-2C(\alpha,1-\alpha)\right)\sigma^2, \end{aligned}$$

using (3.5) and (3.13), and an unbiased estimator of σ^2 follows as

$$\hat{\sigma}^2_{IQR}(\alpha) = \frac{[M(1-\alpha) - M(\alpha)]^2}{2 + \frac{4}{\pi} - \frac{16}{\pi}\sqrt{\alpha(1-\alpha)} - 2C(\alpha, 1-\alpha)}. \tag{3.17}$$

Its variance can be derived using (3.6), (3.11) and (3.12), yielding

$$\mathrm{Var}\left(\hat{\sigma}^2_{IQR}(\alpha)\right) = \frac{\left(\begin{array}{c} 3 - \left(1 + \frac{4\sqrt{\alpha(1-\alpha)}}{\pi}\right)^2 + D(1-\alpha, 1-\alpha, \alpha, \alpha) + \\ 2D(1-\alpha, \alpha, 1-\alpha, \alpha) - 4D(\alpha, \alpha, 1-\alpha, \alpha) \end{array}\right)\sigma^4}{2\left(1 + \frac{2}{\pi} - \frac{8}{\pi}\sqrt{\alpha(1-\alpha)} - C(\alpha, 1-\alpha)\right)^2}.$$

The unbiased range-based estimator of σ^2 is a special case of the IQR-based estimator (corresponding to $\alpha = 0$) and was derived by Parkinson (1980) as

$$\hat{\sigma}^2_R = \frac{R^2}{4\log(2)},$$

with $\mathrm{Var}\left(\hat{\sigma}^2_R\right) = 0.407\sigma^4$. Table 4 presents the variance of $\hat{\sigma}^2_{IQR}(\alpha)$ for different choices of α. From Table 3.4 it is clear that the range-based estimator is the most efficient estimator among the unbiased IQR-based estimators.

Table 3.4: Variance of $\hat{\sigma}^2_{IQR}(\alpha)$

α		0.000	0.025	0.050	0.100	0.125	0.150	0.175	0.200	0.225
$\mathrm{Var}\left(\hat{\sigma}^2_{IQR}(\alpha)\right)$	$\sigma^4\times$	0.408	0.562	0.648	0.806	0.888	0.976	1.071	1.177	1.295

α		0.250	0.275	0.300	0.350	0.375	0.400	0.425	0.450	0.475
$\mathrm{Var}\left(\hat{\sigma}^2_{IQR}(\alpha)\right)$	$\sigma^4\times$	1.431	1.592	1.788	2.351	2.814	3.546	5.273	11.451	81.463

The estimator $\hat{\sigma}^2_{IQR}(\alpha)$ is also a linear combination of the estimators $\tilde{\sigma}^2(\alpha, \alpha)$, $\tilde{\sigma}^2(\alpha, 1-\alpha)$ and $\tilde{\sigma}^2(1-\alpha, 1-\alpha)$ and it can be checked using (3.14) and (3.15) that the weights implicitly imposed by (3.17) are

$$\begin{aligned} w_1 &= \frac{1 - \frac{4\sqrt{\alpha(1-\alpha)}}{\pi}}{2 + \frac{4}{\pi} - \frac{16}{\pi}\sqrt{\alpha(1-\alpha)} - 2C(\alpha, 1-\alpha)} = w_3, \\ w_2 &= 1 - 2w_1. \end{aligned}$$

It is worth pointing out that these weights are not optimal. As a consequence, $\hat{\sigma}^2_{IQR}(\alpha)$ is not the minimum-variance unbiased estimator (based on those quantiles). For example, the range-based estimator, which has variance $0.407\sigma^4$, is less efficient than $\hat{\sigma}^2$ based on the same quantiles, which has variance $0.341\sigma^4$ (see Table 3.3). In particular, it can be checked that the weights imposed on the estimates $\left(\tilde{\sigma}^2(0,0), \tilde{\sigma}^2(0,1), \tilde{\sigma}^2(1,1)\right)'$ by the range-based estimator are $(0.361, 0.279, 0.361)'$, while the optimal weigts are $(0.225, 0.550, 0.225)'$, thus putting more weight on $\tilde{\sigma}^2(0,1)$.

3.3 Unbiased estimators based on quantiles of Brownian bridge

3.3.1 Quantiles of Brownian bridge

For $\alpha \in [0,1]$ let

$$M_b(\alpha) = \begin{cases} \inf\limits_{0\le t\le 1} b_t & \alpha = 0 \\ \inf\left\{x : \int_0^1 1_{(b_t \le x)} dt \ge \alpha\right\} & 0 < \alpha \end{cases}$$

be the α-quantile of the Brownian bridge b_t, $t \in [0,1]$. The low order moments of $M_b(\alpha)$ will be derived from the corresponding moments of $Q(\alpha)$, the α-quantile of the standard Brownian bridge $B_t - tB_1$, and the scaling property of Brownian bridge. The density function, $f_\alpha(\cdot)$, of $Q(\alpha)$ can be derived from $G(q,\cdot)$, the distribution function of the occupation time of b_t above the level q, using

$$\begin{aligned} f_\alpha(q) &= \frac{\partial}{\partial q} P\left(Q(\alpha) \le q\right) \\ &= \frac{\partial}{\partial q} P\left(\int_0^1 1_{(BB_t<q)} dt \ge \alpha\right) \\ &= \frac{\partial}{\partial q} P\left(\int_0^1 1_{(BB_t>q)} dt \le 1-\alpha\right) \\ &= \frac{\partial}{\partial q} G(q, 1-\alpha). \end{aligned}$$

$G(q,\cdot)$ was derived by Takács (1999), but Theorem 2.1 in Hooghiemstra (2002) provides a more convenient representation of $G(q,\cdot)$ to compute the moments from. In particular, for $0 < \alpha < 1$

$$G(q,1-\alpha) = \begin{cases} 1 - 2\frac{\sqrt{1-\alpha}}{\pi} \int_{1-\alpha}^1 \frac{\sqrt{u-(1-\alpha)}}{u^2\sqrt{1-u}} \exp\left(-\frac{2q^2}{1-u}\right) du & q \ge 0 \\ 2\frac{\sqrt{\alpha}}{\pi} \int_\alpha^1 \frac{\sqrt{u-\alpha}}{u^2\sqrt{1-u}} \exp\left(-\frac{2q^2}{1-u}\right) du & q < 0, \end{cases}$$

and has derivative with respect to q

$$f_\alpha(q) = \begin{cases} \frac{8q\sqrt{1-\alpha}}{\pi} \int_{1-\alpha}^1 \frac{\sqrt{u-(1-\alpha)}}{u^2(1-u)^{3/2}} \exp\left(-\frac{2q^2}{1-u}\right) du & q \ge 0 \\ \frac{-8q\sqrt{\alpha}}{\pi} \int_\alpha^1 \frac{\sqrt{u-\alpha}}{u^2(1-u)^{3/2}} \exp\left(-\frac{2q^2}{1-u}\right) du & q < 0. \end{cases} \tag{3.18}$$

Define, for $0 < \alpha < 1$,

$$\begin{aligned} G(\alpha) &= -\sqrt{\alpha} \int_\alpha^1 \frac{\sqrt{u-\alpha}}{u^2} du + \sqrt{1-\alpha} \int_{1-\alpha}^1 \frac{\sqrt{u-(1-\alpha)}}{u^2} du, \\ H(\alpha) &= \sqrt{\alpha} \int_\alpha^1 \frac{\sqrt{(u-\alpha)(1-u)}}{u^2} du + \sqrt{1-\alpha} \int_{1-\alpha}^1 \frac{\sqrt{(u-(1-\alpha))(1-u)}}{u^2} du, \\ J(\alpha) &= \sqrt{\alpha} \int_\alpha^1 \frac{\sqrt{(u-\alpha)(1-u)^3}}{u^2} du + \sqrt{1-\alpha} \int_{1-\alpha}^1 \frac{\sqrt{(u-(1-\alpha))(1-u)^3}}{u^2} du. \end{aligned}$$

The integrals can easily be computed numerically. It can be checked that the functions have the limits

$$-\lim_{\alpha\downarrow 0} G(\alpha) = \lim_{\alpha\uparrow 1} G(\alpha) = \lim_{\alpha\downarrow 0} H(\alpha) = \lim_{\alpha\uparrow 1} H(\alpha) = \lim_{\alpha\downarrow 0} J(\alpha) = \lim_{\alpha\uparrow 1} J(\alpha) = \frac{\pi}{2}, \quad (3.19)$$

which are used to extend the definition of G, H and J to the boundary of the domain. In the appendix the low order moments of $Q(\alpha)$ are derived using the density (3.18). Using those moments, it follows from the scaling property of Brownian bridge that, for $0 \leq \alpha \leq 1$,

$$\begin{aligned} E[M_b(\alpha)] &= \frac{G(\alpha)}{\sqrt{2\pi}}\sigma, & (3.20)\\ \mathrm{Var}\left(M_b(\alpha)\right) &= \left(\frac{H(\alpha)}{\pi} - \frac{G(\alpha)^2}{2\pi}\right)\sigma^2, & (3.21)\\ E[M_b(\alpha)^2] &= \frac{H(\alpha)}{\pi}\sigma^2, & (3.22)\\ \mathrm{Var}\left(M_b(\alpha)^2\right) &= \left(\frac{J(\alpha)}{\pi} - \frac{H(\alpha)^2}{\pi^2}\right)\sigma^4. & (3.23) \end{aligned}$$

In particular,

$$\begin{aligned} -E[M_b(0)] &= E[M_b(1)] = \frac{\sqrt{2\pi}}{4}\sigma, & (3.24)\\ \mathrm{Var}\left(M_b(0)\right) &= \mathrm{Var}\left(M_b(1)\right) = \frac{4-\pi}{8}\sigma^2, & (3.25)\\ E[M_b(0)^2] &= E[M_b(1)^2] = \frac{\sigma^2}{2}, & (3.26)\\ \mathrm{Var}\left(M_b(0)^2\right) &= \mathrm{Var}\left(M_b(1)^2\right) = \frac{\sigma^4}{4}. & (3.27) \end{aligned}$$

It may be remarked that an alternative representation of the density $f_\alpha(q)$ can be obtained based on results in Dassios (1996a), but this representation does not lead to closed-form expressions for the moments (3.20)-(3.23) either.

For $\alpha_1, \alpha_2 \in [0,1]$ it holds that

$$\mathrm{Cov}\left(M_b(\alpha_1), M_b(\alpha_2)\right) = C_Q\left(\alpha_1, \alpha_2\right)\sigma^2, \quad (3.28)$$

where $C_Q(\alpha_1, \alpha_2)$ is the covariance between the α_1-quantile and the α_2-quantile of a standard Brownian bridge. The latter can be obtained by simulation. I computed $C_Q(\alpha_1, \alpha_2)$ by simulating 1 million standard Brownian bridges, each with 1 million increments on the unit interval. Moreover, similar to the Brownian motion case we replace the estimates of $C_Q(\alpha_1, \alpha_2) = C_Q(1-\alpha_1, 1-\alpha_2) = C_Q(\alpha_1, \alpha_2) = C_Q(\alpha_2, \alpha_1) = C_Q(1-\alpha_2, 1-\alpha_1)$ by their average. Furthermore, we do have an expression for the

variance of any quantile, i.e. (3.21), and we can derive the covariance between $M_b(1)$ and $M_b(0)$ from the variance of the range, R_b, of b_t, $0 \leq t \leq 1$, and the variance of $M_b(1)$ given by (3.25). From Feller (1951),

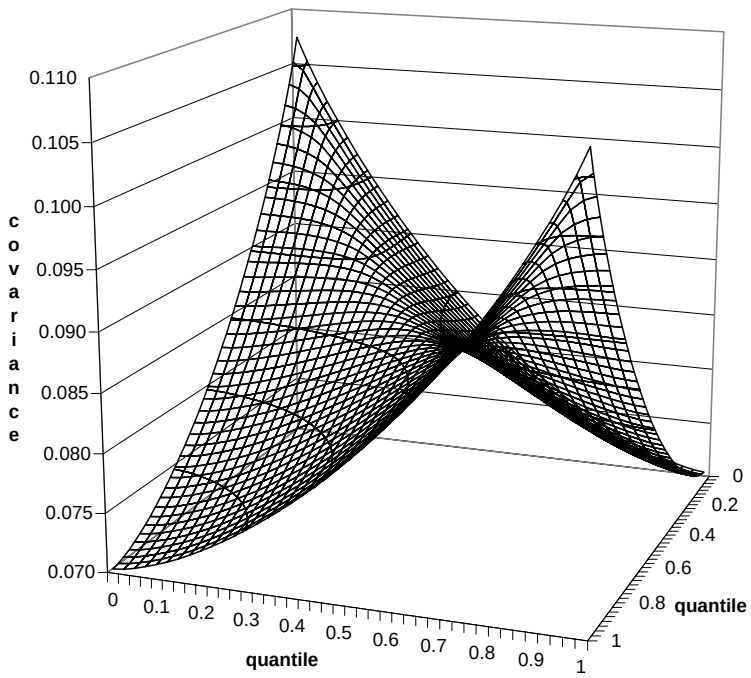
$$\mathrm{Var}\,(R_b) = \left(\frac{\pi^2}{6} - \frac{\pi}{2}\right)\sigma^2.$$

Hence, because $\mathrm{Var}(M_b(1)) = \mathrm{Var}(M_b(0))$,

$$\begin{aligned} C_Q(0,1) &= (\mathrm{Var}(M_b(1)) - \mathrm{Var}(R_b)/2)\,\sigma^{-2} \\ &= \frac{4-\pi}{8} - \frac{\pi^2}{12} + \frac{\pi}{4} \\ &= -\frac{\pi^2}{12} + \frac{\pi}{8} + \frac{1}{2}. \end{aligned}$$

The covariances between Brownian bridge quantiles are graphed in Figure 3.3.

Figure 3.3: Covariances between quantiles of a Brownian bridge ($\sigma = 1$)



3.3.2 Quantile-based unbiased estimators of σ

One quantile

Except for the median, all quantiles are informative for σ. Note, e.g., that $G(1/2) = 0$, so for $\alpha = 1/2$ the right-hand side of (3.20) vanishes. For any other α in $[0,1]$, an unbiased estimator of σ follows from (3.20) as

$$\tilde{\sigma}_b(\alpha) = \frac{\sqrt{2\pi}M_b(\alpha)}{G(\alpha)},$$

and by (3.21) its variance is

$$\text{Var}(\tilde{\sigma}_b(\alpha)) = \left(2\frac{H(\alpha)}{G(\alpha)^2} - 1\right)\sigma^2.$$

It can be checked numerically that this variance is minimized for $\alpha = 0$ and for $\alpha = 1$, yielding $\text{Var}(\tilde{\sigma}_b(0)) = \text{Var}(\tilde{\sigma}_b(1)) = \left(\frac{4}{\pi} - 1\right)\sigma^2 = 0.273\sigma^2$. Thus, $\tilde{\sigma}_b(0)$ and $\tilde{\sigma}_b(1)$ are the minimum variance unbiased estimators based on moment condition (3.20) for a single quantile. Note that these estimators are much more efficient than the Brownian motion version with $\text{Var}(\tilde{\sigma}(0)) = 0.571\sigma^2$.

Several quantiles

Unbiased estimators of σ can be constructed based on more than one quantile. Fix a set of distinct values $\alpha_1, ..., \alpha_n \in [0, 1]\backslash\{1/2\}$, with corresponding quantiles $M_b(\alpha_1), ...,$ $M(\alpha_n)$ and unbiased estimates $\tilde{\sigma}_b(\alpha_1), ..., \tilde{\sigma}_b(\alpha_n)$. Let $\tilde{s}_b = (\tilde{\sigma}_b(\alpha_1), ..., \tilde{\sigma}_b(\alpha_n))'$. Then the minimum-variance unbiased estimator of σ is

$$\hat{\sigma}_b = (\iota' V^{-1}\iota)^{-1}\iota' V^{-1}\tilde{s}_b,$$

with variance $(\iota' V^{-1}\iota)^{-1}$, where $V = \text{Var}(\tilde{s}_b)$. The (i, j)-th element of V is

$$V_{ij} = 2\pi\frac{C_Q(\alpha_i, \alpha_j)}{G(\alpha_i)G(\alpha_j)}\sigma^2.$$

Table 3.5 presents the variance of $\hat{\sigma}_b$ for different choices of $\alpha_1, ..., \alpha_n$. Again, using more than one quantile dramatically improves the efficiency of the estimator. Interestingly, when we compare these results to Table 3.1, estimators based on two quantiles of Brownian bridge, $M_b(0)$ and $M_b(1)$, are as efficient as estimators based on forty quantiles of Brownian motion.

Table 3.5: Variance of $\hat{\sigma}_b$

Number of quantiles		2	4	6	10	20	40
$\text{Var}(\hat{\sigma}_b)$	$\sigma^2\times$	0.047	0.041	0.039	0.035	0.032	0.029

Notes: The quantiles are equally spaced. The median is not used, because it is uninformative.

A special case: Interquantile ranges

For $0 \leq \alpha < 1/2$, the optimal weights to combine the estimates $\tilde{\sigma}_b(\alpha)$ and $\tilde{\sigma}_b(1-\alpha)$ are $w = (1/2, 1/2)'$. Consequently, the minimum-variance unbiased linear combination of $\tilde{\sigma}_b(\alpha)$ and $\tilde{\sigma}_b(1-\alpha)$ is the IQR-based estimator

$$\hat{\sigma}_{IQR_b}(\alpha) = \frac{\sqrt{\pi}\left[M_b(1-\alpha) - M_b(\alpha)\right]}{\sqrt{2}G(1-\alpha)},$$

(note that $G(1-\alpha) = -G(\alpha)$) and its variance can be derived using (3.21) and (3.28) as

$$\text{Var}\left(\hat{\sigma}_{IQR_b}(\alpha)\right) = \left(\frac{H(\alpha)}{G(\alpha)^2} - \frac{1}{2} - \frac{\pi C_Q(\alpha, 1-\alpha)}{G(\alpha)^2}\right)\sigma^2. \tag{3.29}$$

For $\alpha = 0$, we have Kunitomo's (1992) range-based estimator of σ,

$$\hat{\sigma}_{IQR_b}(0) = \sqrt{\frac{2}{\pi}}\left[M_b(1) - M_b(0)\right]$$

with variance $\left(\frac{\pi}{3} - 1\right)\sigma^2 = 0.047\sigma^2$, corresponding with the simulation result in Table 3.5. Table 3.6 presents the variances of $\hat{\sigma}_{IQR_b}(\alpha)$ for various choices of α. The range provides the most efficient IQR-based estimator. From (3.29) it follows that

$$\lim_{\alpha \to 0.5} \text{Var}\left(\hat{\sigma}_{IQR_b}(\alpha)\right) = \infty.$$

Table 3.6: Variance of $\hat{\sigma}_{IQR_b}(\alpha)$

α		0.000	0.025	0.050	0.100	0.125	0.150	0.175	0.200	0.225
$\text{Var}\left(\hat{\sigma}_{IQR_b}(\alpha)\right)$	$\sigma^2\times$	0.047	0.068	0.078	0.096	0.104	0.113	0.122	0.131	0.141
α		0.250	0.275	0.300	0.350	0.375	0.400	0.425	0.450	0.475
$\text{Var}\left(\hat{\sigma}_{IQR_b}(\alpha)\right)$	$\sigma^2\times$	0.151	0.163	0.176	0.207	0.227	0.251	0.280	0.321	0.383

3.3.3 Quantile-based unbiased estimators of σ^2

Unbiased estimators of σ^2 can be constructed from the product of two quantiles of a Brownian bridge. For $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in [0, 1]$ it holds, due to the scaling property of Brownian motion, that

$$\text{Cov}\left(M_b(\alpha_1)M_b(\alpha_2), M_b(\alpha_3)M_b(\alpha_4)\right) = D_Q(\alpha_1, \alpha_2, \alpha_3, \alpha_4)\sigma^4, \tag{3.30}$$

where $D_Q(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ is the covariance between the product of the α_1-quantile and the α_2-quantile and the product of the α_3-quantile and the α_4-quantile of a standard Brownian bridge. Apart from $D_Q(\alpha_1, \alpha_1, \alpha_1, \alpha_1) = \text{Var}(M_b(\alpha_1)^2)$, which is given by (3.23), the covariances need to be obtained by simulation. I computed $D_Q(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ by simulating 1 million standard Brownian motions, each with 1 million increments on the unit interval. Furthermore, since

$$D_Q\left(\alpha_1, \alpha_2, \alpha_3, \alpha_4\right) = D_Q(1-\alpha_1, 1-\alpha_2, 1-\alpha_3, 1-\alpha_4), \tag{3.31}$$

both simulated covariances can be replaced by their average to avoid small discrepancies.

The product of two quantiles

Let $\alpha_i, \alpha_j \in [0,1]$ and consider the quantiles $M_b(\alpha_i)$ and $M_b(\alpha_j)$. From (3.20) and (3.28) it follows that

$$\begin{aligned} E\left[M_b(\alpha_i)M_b(\alpha_j)\right] &= \mathrm{Cov}(M_b(\alpha_i), M_b(\alpha_j)) + E\left[M_b(\alpha_i)\right] E\left[M_b(\alpha_j)\right] \\ &= C_Q(\alpha_i,\alpha_j)\sigma^2 + \frac{G(\alpha_i)G(\alpha_j)}{2\pi}\sigma^2. \end{aligned} \tag{3.32}$$

The expected value of the product of two quantiles is depicted in the left panel of Figure 3.4 for the case $\sigma = 1$.

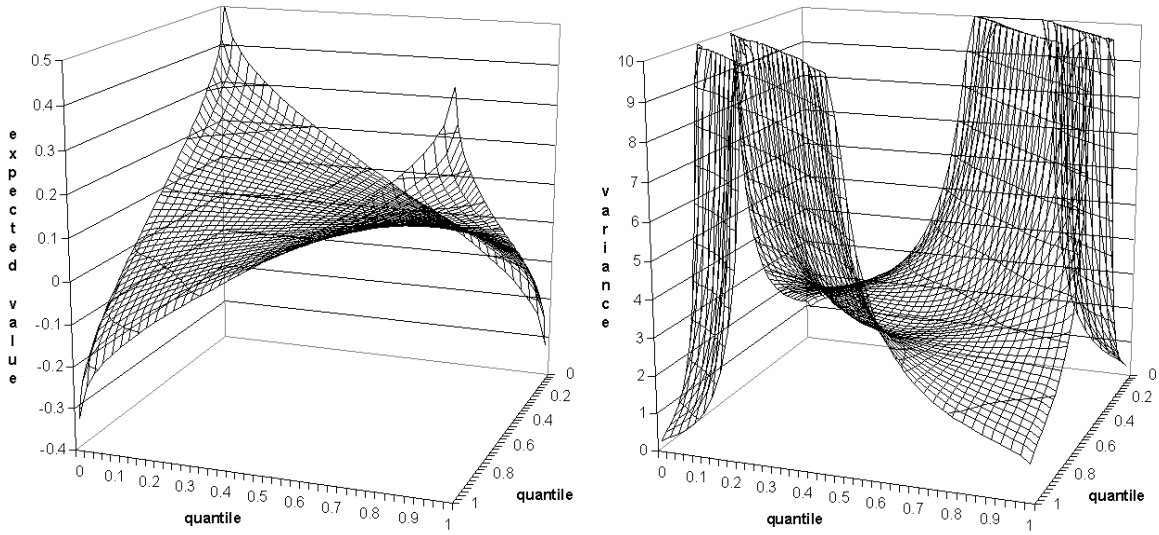
Provided α_i and α_j are such that $E\left[M_b(\alpha_i)M_b(\alpha_j)\right] \neq 0$, an unbiased estimator of σ^2 follows as

$$\tilde{\sigma}_b^2(\alpha_i,\alpha_j) = \frac{M_b(\alpha_i)M_b(\alpha_j)}{C_Q(\alpha_i,\alpha_j) + \frac{G(\alpha_i)G(\alpha_j)}{2\pi}\sigma^2}, \tag{3.33}$$

with variance

$$\mathrm{Var}\left(\tilde{\sigma}_b^2(\alpha_i,\alpha_j)\right) = \frac{D_Q(\alpha_i,\alpha_j,\alpha_i,\alpha_j)\sigma^4}{\left(C_Q(\alpha_i,\alpha_j) + \frac{G(\alpha_i)G(\alpha_j)}{2\pi}\sigma^2\right)^2}.$$

Figure 3.4: Expected value of the product of two quantiles (left panel) and the variance of $\tilde{\sigma}_b^2(\alpha_i,\alpha_j)$ (right panel) when $\sigma = 1$



The variance of $\tilde{\sigma}_b^2(\alpha_i,\alpha_j)$ can be computed using the numerical results for $C_Q(\alpha_i,\alpha_j)$ and $D_Q(\alpha_i,\alpha_j,\alpha_i,\alpha_j)$. For $\sigma = 1$, the right panel of Figure 3.4 shows the variances for different choices of $\alpha_i, \alpha_j \in [0,1]$. The minimum variance unbiased estimator is $\tilde{\sigma}_b^2(0,1)$, with variance $0.271\sigma^4$. If $E\left[M_b(\alpha_i)M_b(\alpha_j)\right]$ is close to zero, then $\tilde{\sigma}_b^2(\alpha_i,\alpha_j)$ is very inefficient, e.g. $E\left[M_b(0)M_b(0.65)\right] = 0.00026\sigma^2$ and $\tilde{\sigma}_b^2(0,0.65)$ has variance $562,019\sigma^4$.

For the one-quantile case ($\alpha_i = \alpha_j$), by (3.22) the estimator is

$$\tilde{\sigma}_b^2(\alpha_i, \alpha_i) = \pi \frac{M_b(\alpha_i)^2}{H(\alpha_i)}. \tag{3.34}$$

Its variance follows from (3.23) as

$$\text{Var}\left(\tilde{\sigma}^2(\alpha_i, \alpha_i)\right) = \left(\pi \frac{J(\alpha_i)}{H(\alpha_i)^2} - 1\right) \sigma^4.$$

This variance is minimized for $\alpha_i = 0$ and for $\alpha_i = 1$, yielding $\tilde{\sigma}_b^2(0,0) = 2M_b(0)^2$ and $\tilde{\sigma}_b^2(1,1) = 2M_b(1)^2$ as the minimum variance unbiased estimators, with $\text{Var}\left(\tilde{\sigma}_b^2(0,0)\right) = \text{Var}\left(\tilde{\sigma}_b^2(1,1)\right) = \sigma^4$.

Several products of two quantiles

Unbiased estimators of σ^2 can be constructed as linear combinations of estimators given by (3.33). In general, a set of distinct values $\alpha_1, ..., \alpha_n \in [0,1]$ corresponds with a set of $n(n+1)/2$ distinct estimators $\tilde{\sigma}_b^2(\alpha_i, \alpha_j)$, where $i = 1, ..., n$ and $j = 1, ..., i$. Let $\tilde{s}_b^2$ be the vector of these estimators. Then the minimum-variance unbiased estimator of σ^2 is

$$\hat{\sigma}_b^2 = (\iota' V^{-1} \iota)^{-1} \iota' V^{-1} \tilde{s}_b^2, \tag{3.35}$$

where $V = \text{Var}(\tilde{s}_b^2)$ has elements

$$\text{Cov}\left(\tilde{\sigma}_b^2(\alpha_i, \alpha_j), \tilde{\sigma}_b^2(\alpha_k, \alpha_l)\right) = d(\alpha_i, \alpha_j) D_Q(\alpha_i, \alpha_j, \alpha_k, \alpha_l) d(\alpha_k, \alpha_l),$$

where

$$d(\alpha_i, \alpha_j) = \frac{1}{C_Q(\alpha_i, \alpha_j) + \frac{G(\alpha_i)G(\alpha_j)}{2\pi}}.$$

Table 3.7 presents the variance of $\hat{\sigma}_b^2$ for different choices of $\alpha_1, ..., \alpha_n$. It shows that the variance of quantile-based estimators can be brought down to $0.110\sigma^4$. Again the estimator is much more efficient than the Brownian motion case.

Table 3.7: Variance of $\hat{\sigma}_b^2$

Number of quantiles		2	3	5	6	11	21	41
$\text{Var}(\hat{\sigma}_b^2)$	$\sigma^4\times$	0.197	0.195	0.167	0.160	0.140	0.123	0.110

Note: The quantiles are equally spaced.

A special case: Interquantile ranges

Unbiased IQR-based estimator of σ^2 can be derived. Let $0 \leq \alpha < 1/2$. Then, by symmetry,

$$\begin{aligned} E\left[M_b(1-\alpha) - M_b(\alpha)\right]^2 &= 2E\left(M_b(\alpha)^2\right) - 2E\left(M_b(\alpha)M_b(1-\alpha)\right) \\ &= \left(\frac{2H(\alpha)}{\pi} - 2C_Q(\alpha, 1-\alpha) + \frac{G(\alpha)^2}{\pi}\right)\sigma^2, \end{aligned}$$

using (3.22) and (3.32), and an unbiased estimator of σ^2 follows as

$$\hat{\sigma}^2_{IQR_b}(\alpha) = \frac{\left[M(1-\alpha) - M(\alpha)\right]^2}{\frac{2H(\alpha)}{\pi} - 2C_Q(\alpha, 1-\alpha) + \frac{G(\alpha)^2}{\pi}}. \tag{3.36}$$

Its variance can be derived using (3.23), (3.30) and (3.31), yielding

$$\mathrm{Var}\left(\hat{\sigma}^2_{IQR_b}(\alpha)\right) = \frac{\left(\begin{array}{c} \frac{2J(\alpha)}{\pi} - \frac{2H(\alpha)^2}{\pi^2} + 4D_Q(\alpha, 1-\alpha, \alpha, 1-\alpha) \\ -8D_Q(\alpha, \alpha, \alpha, 1-\alpha) + 2D_Q(\alpha, \alpha, 1-\alpha, 1-\alpha) \end{array}\right)\sigma^4}{\left(\frac{2H(\alpha)}{\pi} - 2C_Q(\alpha, 1-\alpha) + \frac{G(\alpha)^2}{\pi}\right)^2}.$$

For $\alpha = 0$, we obtain the unbiased range-based estimator of σ^2, which was derived by Kunitomo (1992) as

$$\hat{\sigma}^2_{R_b} = \frac{6R_b^2}{\pi^2},$$

with $\mathrm{Var}\left(\hat{\sigma}^2_{R_b}\right) = 0.2\sigma^4$. Table 3.8 presents the variance of $\hat{\sigma}^2_{IQR_b}(\alpha)$ for different choices of α. It shows that the range-based estimator is the most efficient estimator among the unbiased IQR-based estimators.

Table 3.8: Variance of $\hat{\sigma}^2_{IQR_b}(\alpha)$

α		0.000	0.025	0.050	0.100	0.125	0.150	0.175	0.200	0.225
$\mathrm{Var}\left(\hat{\sigma}^2_{IQR_b}(\alpha)\right)$	$\sigma^4\times$	0.200	0.291	0.339	0.423	0.466	0.510	0.559	0.611	0.669

α		0.250	0.275	0.300	0.350	0.375	0.400	0.425	0.450	0.475
$\mathrm{Var}\left(\hat{\sigma}^2_{IQR_b}(\alpha)\right)$	$\sigma^4\times$	0.734	0.810	0.899	1.14	1.31	1.55	2.00	3.74	37.6

The estimator $\hat{\sigma}^2_{IQR_b}(\alpha)$ is also a linear combination of the estimators $\tilde{\sigma}^2_b(\alpha, \alpha)$, $\tilde{\sigma}^2_b(\alpha, 1-\alpha)$ and $\tilde{\sigma}^2_b(1-\alpha, 1-\alpha)$ and it can be checked using (3.33) and (3.34) that the weights implicitly imposed by (3.36) are

$$\begin{aligned} w_1 &= w_3 = \frac{H(\alpha)}{2H(\alpha) - 2\pi C_Q(\alpha, 1-\alpha) + G(\alpha)^2}, \\ w_2 &= 1 - 2w_1. \end{aligned}$$

These weights are not optimal and $\hat{\sigma}^2_{IQR_b}(\alpha)$ is not the minimum-variance unbiased estimator (based on these quantiles). For example, it can be checked that the weights imposed on the estimates $\left(\tilde{\sigma}^2_b(0,0), \tilde{\sigma}^2_b(0,1), \tilde{\sigma}^2_b(1,1)\right)'$ by the range-based estimator are $(0.304, 0.392, 0.304)'$, while the optimal weigts are $(0.257, 0.486, 0.257)'$, thus putting more weight on $\tilde{\sigma}^2_b(0,1)$. However, the gain is minimal since the variance of the range-based estimator of $0.200\sigma^4$ is only slightly decreased to $0.197\sigma^4$ (see Table 3.7).

3.4 Consistent Estimators

The estimators I derived so far are not consistent despite the continuous-time observations. Even when we increase the number of quantiles used, the variance of the estimators converges to some (unknown) limit value greater than zero. Yet, as I show below, consistent estimators can be obtained.

So far, we have focused on estimation of volatility based on the unit interval as a whole. Alternatively, we can split the unit interval into subintervals, estimate volatility based on each of the subintervals and average these estimates to obtain a new estimate. The underlying idea is to extract information from the quantiles in each subinterval in order to attain higher efficiency of the volatility estimator. By yet another method we take the average of estimates based on the unit interval, but where we permute the subintervals each time.

There are several ways of defining and using subintervals and the efficiency of the estimator will differ depending on the set of subintervals used.

3.4.1 Averaging subinterval estimates

Let us divide the unit interval into n non-overlapping subintervals of length $1/n$ and define a new estimator of σ^2 as

$$\breve{\sigma}^2 = \sum_{i=1}^{n} \hat{\sigma}^2_i,$$

with $\hat{\sigma}^2_i$ as before, but applied to the i-th subinterval. Clearly, $\hat{\sigma}^2_i$, $i = 1, ..., n$, are $i.i.d.$ and $\hat{\sigma}^2_i$ has the same distribution as $\hat{\sigma}^2/n$. Therefore,

$$\text{Var}\left(\breve{\sigma}^2\right) = \frac{\text{Var}\left(\hat{\sigma}^2\right)}{n}, \tag{3.37}$$

and thus as $n \to \infty$

$$\text{Var}\left(\breve{\sigma}^2\right) \to 0.$$

The variance of the estimator can be estimated from the subinterval estimates $\hat{\sigma}_i^2$

$$\widehat{\mathrm{Var}}\left(\breve{\sigma}^2\right) = \frac{n}{n-1}\sum_{i=1}^{n}\left(\hat{\sigma}_i^2 - \overline{\hat{\sigma}_i^2}\right)^2.$$

The estimator $\hat{\sigma}^2$ is a linear combination of estimates based on the product of two quantiles $\tilde{\sigma}^2(\alpha_1, \alpha_2)$. This implies that the subinterval estimates $\tilde{\sigma}_i^2(\alpha_1, \alpha_2)$ and their covariances (with subinterval estimates based on other quantiles $\tilde{\sigma}_i^2(\alpha_3, \alpha_4)$) allow us to estimate the optimal weights w needed for the optimal linear combination. This estimation of the optimal weights provides an alternative to the procedure proposed above where the optimal weights are obtained based on simulation results for the covariances of quantiles $D(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

We can generalize this case to overlapping intervals. For example, we can consider a subinterval that moves through the sample as a moving window. For a given subinterval length, this raises the number of subintervals considerably compared to the non-overlapping case. Now, the estimator

$$\breve{\sigma}^2 = \frac{\sum_{i=1}^{m}\hat{\sigma}_i^2 n}{m},$$

with m the number of overlapping intervals considered, each of length $1/n$, has variance

$$\begin{aligned}\mathrm{Var}\left(\breve{\sigma}^2\right) &= \frac{n^2}{m^2}\mathrm{Var}\left(\sum_{i=1}^{m}\hat{\sigma}_i^2\right) = \frac{n^2}{m^2}\left(\sum_{i=1}^{m}\mathrm{Var}\left(\hat{\sigma}_i^2\right) + \sum_{\substack{i=1\\ i\neq j}}^{m}\sum_{j=1}^{m}\mathrm{Cov}\left(\hat{\sigma}_i^2, \hat{\sigma}_j^2\right)\right)\\ &= \frac{\mathrm{Var}\left(\hat{\sigma}^2\right)}{m} + \frac{n^2}{m^2}\sum_{\substack{i=1\\ i\neq j}}^{m}\sum_{j=1}^{m}\mathrm{Cov}\left(\hat{\sigma}_i^2, \hat{\sigma}_j^2\right).\end{aligned}$$

As $n < m$, the first term of the variance is smaller than (3.37), but since a lot of the covariances are bigger than zero the gain is reduced. However, the estimator is consistent, because as $n \to \infty$, with $n < m$, $\mathrm{Var}\left(\breve{\sigma}^2\right) \to 0$. We can estimate the variance of the estimator from the covariances between subinterval estimates

$$\begin{aligned}\widehat{\mathrm{Var}}\left(\breve{\sigma}^2\right) &= \frac{n^2}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\widehat{\mathrm{Cov}}\left(\hat{\sigma}_i^2, \hat{\sigma}_j^2\right),\\ \widehat{\mathrm{Cov}}\left(\hat{\sigma}_i^2, \hat{\sigma}_j^2\right) &= \mathrm{Var}\left(\hat{\sigma}_i^2\right) A_{ij} n,\end{aligned}$$

where $0 \leq A_{ij} \leq 1/n$ is the amount of overlap between the i-th and j-th subinterval.

3.4.2 Averaging estimates after permuting subintervals

An interesting case is when the time order of the subintervals is no longer fixed. Let us divide the unit interval into n subintervals of length $1/n$, take a permutation of these

subintervals and glue the pieces of the series together according to the new order to obtain a new series of unit length. The new series has the same start and end points as the original series, but the path differs. This implies that the new series has different quantiles and a new volatility estimate can be obtained. In particular, for n subsequent subintervals we can obtain $k = n!$ different permutations and estimates. Let $\hat{\sigma}_i^2$ be an estimator based on the i-th permutation of n subintervals of a Brownian motion, then a new estimator can be defined that takes the average of the estimates obtained from the k different permutations

$$\mathring{\sigma}^2 = \frac{\sum_{i=1}^{k} \hat{\sigma}_i^2}{k},$$

Consider the limit of the permuted Brownian motion, as $n \to \infty$. For each fixed n, and hence also in the limit, the permuted Brownian motion has the same end point as the original series. This end point causes permuted Brownian motions, their quantiles and volatility estimates from these quantiles to be correlated. For example, if the endpoint is a large positive number, then the maximum will be at least this number for any permutations taken from the original series. The correlation implies that as we increase the number of permutations used in the estimation, the variance of the estimator does not converge to zero and the estimator is not consistent. The estimator has a variance that converges to

$$\begin{aligned} \lim_{k\to\infty} \operatorname{Var}\left(\frac{\sum_{i=1}^{k} \hat{\sigma}_i^2}{k}\right) &= \lim_{k\to\infty} \operatorname{Var}\left(\hat{\sigma}_i^2\right)\left(\frac{1}{k} + \frac{k(k-1)}{k^2}\rho\right) \\ &= \rho \operatorname{Var}\left(\hat{\sigma}_i^2\right) > 0, \end{aligned}$$

where ρ is the correlation between the estimates of two permuted Brownian motions.

On the contrary, a Brownian bridge removes the trend from its corresponding Brownian motion and always ends at zero. If we again take the example of the maximum, then the maximum of a Brownian bridge can be any (positive) number after permuting. In other words, permutations of a Brownian motion that is transformed into a Brownian bridge, its quantiles and quantile-based estimates are independent in the limit, as $k \to \infty$. This implies that the estimator that takes the average of estimates based on the Brownian bridge version of permutations of a Brownian motion,

$$\mathring{\sigma}_b^2 = \frac{\sum_{i=1}^{k} \hat{\sigma}_{b,i}^2}{k},$$

is consistent, because as $k \to \infty$

$$\operatorname{Var}\left(\mathring{\sigma}_b^2\right) \to 0.$$

Let us now consider the case where the number of subintervals n is small. In that case, permuted series will be dependent (even in the Brownian bridge case). It can be checked by simulation, however, that permuted series quickly approach independence as n (and hence k) increases.

3.5 Conclusion

In this chapter I have developed unbiased estimators of the diffusion coefficient based on quantiles of either Brownian motion or Brownian bridge. Within each class of estimators I have looked for the minimum-variance unbiased estimator. It turns out that the estimators based on the quantiles of Brownian bridge are more efficient than their Brownian motion counterparts. Furthermore, we can conclude that the supremum and infimum are the quantiles that contain the most information, but when combined with other quantiles more efficient estimators can be obtained. If we consider the class of unbiased estimators based on only the supremum and the infimum, it turns out that the range-based estimator of σ^2 is not the minimum-variance unbiased estimator. This implies that the historic 'high-low' data can be used more efficiently than standard practice did so far.

I have also shown how consistent estimators are obtained by taking the average of subinterval estimates. This approach corresponds with the realized volatility literature where averages are taken of squared subinterval returns. A new approach is to take averages of estimates from permutations of subintervals of the Brownian bridge. The latter estimator is based on quantiles taken over the unit interval, which are fairly robust to market microstructure noise. In this respect, the estimator promises to be suited to estimate volatility from high frequency data.

Appendix

The i-th (non-central) moment of $Q(\alpha)$, for $0 < \alpha < 1$, is

$$E\left[Q(\alpha)^i\right] = \int_{-\infty}^{0} q^i f_\alpha(q)dq + \int_{0}^{\infty} q^i f_\alpha(q)dq,$$

with $f_\alpha(q)$ given by (3.18). For $0 < \alpha < 1$, $f_\alpha(q)$ is given by (3.18). The double integrals easily reduce to single integrals, for example

$$\begin{aligned}
\int_{-\infty}^{0} q f_\alpha(q)dq &= \int_{-\infty}^{0} q\frac{-8q\sqrt{\alpha}}{\pi}\int_\alpha^1 \frac{\sqrt{u-\alpha}}{u^2(1-u)^{3/2}}\exp\left(-\frac{2q^2}{1-u}\right)dudq, \\
&= \frac{-8\sqrt{\alpha}}{\pi}\int_\alpha^1 \frac{\sqrt{u-\alpha}}{u^2(1-u)^{3/2}}\int_{-\infty}^{0} q^2\exp\left(-\frac{2q^2}{1-u}\right)dqdu, \\
&= \frac{-8\sqrt{\alpha}}{\pi}\int_\alpha^1 \frac{\sqrt{u-\alpha}}{u^2(1-u)^{3/2}}\frac{\sqrt{2\pi}(1-u)^{3/2}}{16}du, \\
&= -\sqrt{\frac{\alpha}{2\pi}}\int_\alpha^1 \frac{\sqrt{u-\alpha}}{u^2}du.
\end{aligned}$$

Similarly, it can be derived that

$$\begin{aligned}
\int_{0}^{\infty} q f_\alpha(q)dq &= \sqrt{\frac{1-\alpha}{2\pi}}\int_{1-\alpha}^1 \frac{\sqrt{u-(1-\alpha)}}{u^2}du, \\
\int_{-\infty}^{0} q^2 f_\alpha(q)dq &= \frac{\sqrt{\alpha}}{\pi}\int_\alpha^1 \frac{\sqrt{(u-\alpha)(1-u)}}{u^2}du, \\
\int_{0}^{\infty} q^2 f_\alpha(q)dq &= \frac{\sqrt{1-\alpha}}{\pi}\int_{1-\alpha}^1 \frac{\sqrt{(u-(1-\alpha))(1-u)}}{u^2}du, \\
\int_{-\infty}^{0} q^4 f_\alpha(q)dq &= \frac{\sqrt{\alpha}}{\pi}\int_\alpha^1 \frac{\sqrt{(u-\alpha)(1-u)^3}}{u^2}du, \\
\int_{0}^{\infty} q^4 f_\alpha(q)dq &= \frac{\sqrt{1-\alpha}}{\pi}\int_{1-\alpha}^1 \frac{\sqrt{(u-(1-\alpha))(1-u)^3}}{u^2}du.
\end{aligned}$$

With the definitions of $G(\alpha)$, $H(\alpha)$ and $J(\alpha)$ one easily obtains

$$\begin{aligned}
E[Q(\alpha)] &= \frac{G(\alpha)}{\sqrt{2\pi}}, \\
E[Q(\alpha)^2] &= \frac{H(\alpha)}{\pi}, \\
E\left[Q(\alpha)^4\right] &= \frac{J(\alpha)}{\pi}, \\
\mathrm{Var}\left(Q(\alpha)\right) &= \frac{H(\alpha)}{\pi} - \frac{G(\alpha)^2}{2\pi}, \\
\mathrm{Var}\left(Q(\alpha)^2\right) &= \frac{J(\alpha)}{\pi} - \frac{H(\alpha)^2}{\pi^2}.
\end{aligned}$$

A continuity argument shows that these expressions are valid for $0 \leq \alpha \leq 1$. Higher moments of $Q(\alpha)$ can be obtained in a similar fashion.

Chapter 4

Quantile-Based Estimation of Volatility in Discrete Time and in the Presence of Market Microstructure Noise

4.1 Introduction

In the previous chapter volatility estimators were derived assuming we have continuous-time observations. In reality, prices are only observed in discrete time with the transaction level as the highest frequency attainable. It is known from the literature on range-based volatility estimators that time discreteness introduces a bias. In this chapter I investigate whether this holds for quantile-based estimators in general. Furthermore, the observed transaction price can differ from the "true" price due to market microstructure effects. It is known from the literature on realized volatility and realized range estimators that this introduces a bias. This chapter investigates how this issue affects the performance of the quantile-based estimators and proposes adjusted estimators if needed. In Section 4.2, the discreteness bias is discussed and bias-corrected estimators are presented. Section 4.3 discusses the noise bias and presents simulation-based noise-bias corrections.

4.2 Quantile-based estimators in discrete time

Quantiles derived from discrete-time data typically differ from their continuous-time counterparts. Therefore, we denote these quantiles differently. Let $t_1, ..., t_n$ be the sequence of sampling times of the continuous-time price process p_t, $0 \leq t \leq 1$. For

$\alpha \in [0,1]$, define the α-quantile of $\{p_{t_1}, ..., p_{t_n}\}$ as

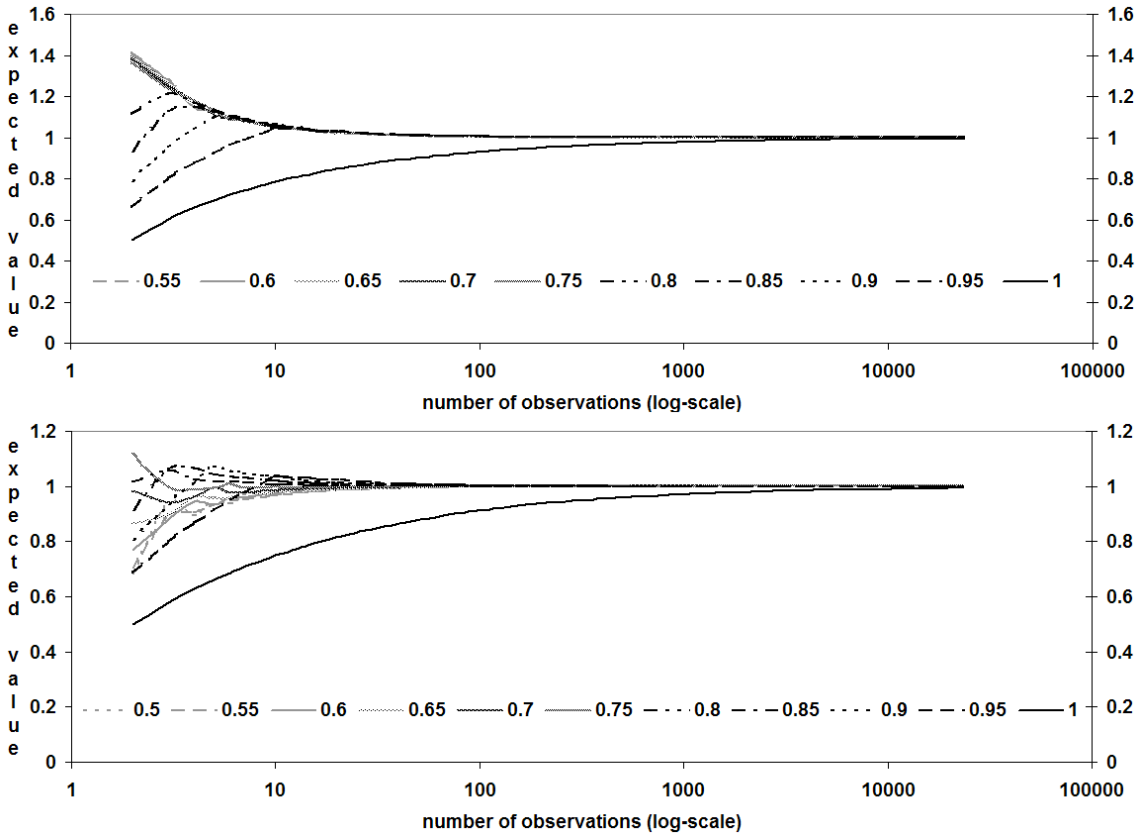
$$M(\alpha, n) = \begin{cases} \min\{p_{t_1}, ..., p_{t_n}\} & \alpha = 0 \\ \inf\left\{z : \frac{1}{1-t_1}\sum_{i=1}^{n}(t_{i+1} - t_i)\,(p_{t_i} \leq z) \geq \alpha\right\} & 0 < \alpha \leq 1 \end{cases},$$

where $t_{n+1} = 1$. According to this definition a price is weighted by the time it prevails and thereby prevents overweighting active trading periods with lots of price changes.

4.2.1 Discreteness bias

It is known that discreteness introduces a bias for quantile estimators, i.e. $E[M(\alpha, n)] \neq E[M(\alpha)]$ for $n < \infty$. As a result, continuous-time quantile-based volatility estimators applied to discrete-time data will be biased. In particular, the bias of range-based estimators has been studied starting with Garman and Klass (1980). The reason for the bias of the range is intuitively clear, knowing that the sample range can only decrease as n decreases. Consequently, the true range and volatility will be underestimated.

Figure 4.1: Expected value of 1-quantile based estimators of σ (top-panel) and σ^2 (bottom panel) as a function of the number of observations (log-scale)



I start with a simulation exercise to study the discreteness bias of quantile-based

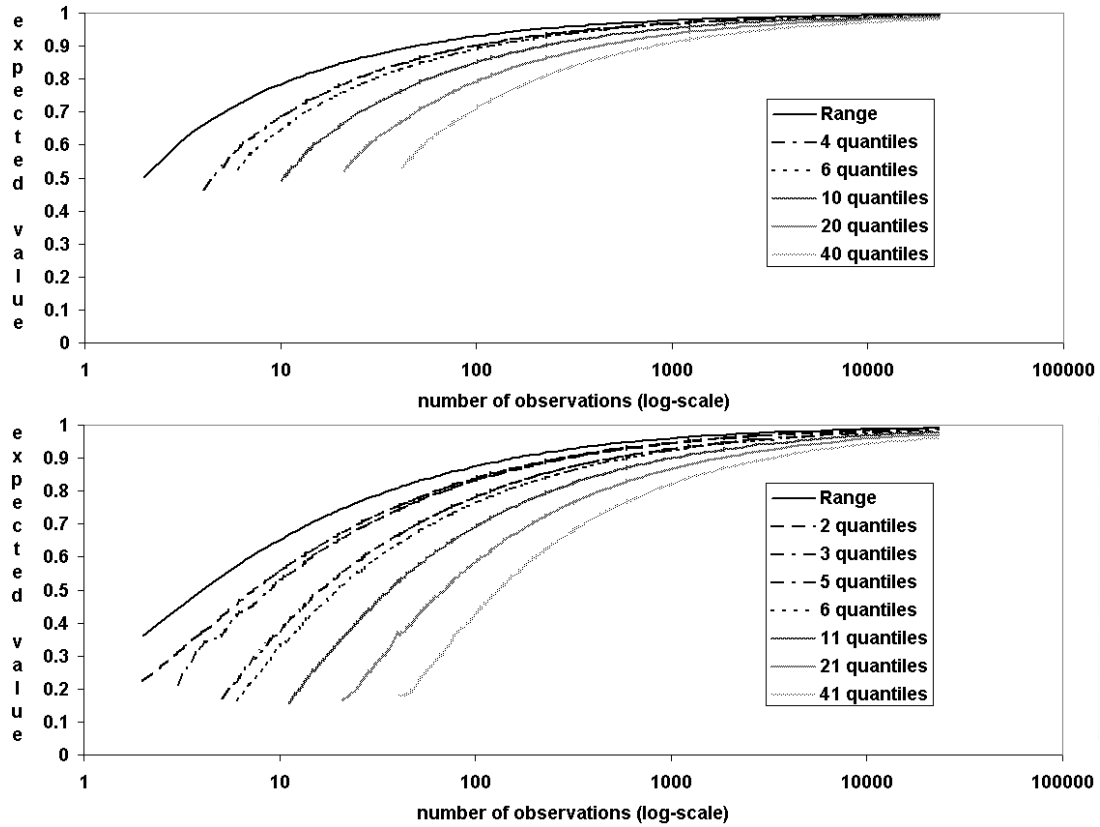
estimators. For simplicity, the simulations are based on equally spaced observations. I simulate 1 million standard random walks of $23,401$ observations each, representing price series observed at each second of a 6.5 hours trading day as is for example the case for the NYSE and Nasdaq. The expected values for estimators based on 1 quantile of Brownian motion, with $\alpha \geq 0.5$ and $\sigma = 1$, are presented in Figure 4.1 as a function of the number of observations, $n \leq 23,401$. The top panel shows the expected value of estimators of σ, and the bottom panel those of σ^2. It is clear that the estimator based on the maximum is biased most (and via symmetry also the one based on the minimum). The downward bias is already apparent for relatively high values of n. Estimators based on any other quantile are much less affected by discreteness. Only when the number of observations becomes very low things get out of hand. Note that the expected value obtained by simulation when the number of observations is very low is dependent on the algorithm used to compute quantiles. In particular, some programs use interpolation techniques, while others do not. This affects the bias and can lead to an upward bias. We can conclude that a number of observations of at least one hundred guarantees a robust estimator for all but the estimator based on extrema. Furthermore, time discreteness has hardly any effect on the variance of these estimators. The simulation results showed that only the variance of the extrema-based estimator of σ^2 decreases somewhat with the number of observations; e.g. it has a variance of $1.852\sigma^4$ with $23,401$ observations and $1.701\sigma^4$ with 101 observations.

The question now remains what happens if we combine the extrema-based estimators with other 1-quantile based estimators. Figure 4.2 presents the expected value of different minimum-variance quantile-based estimators and the range-based estimator. The expected value is not computed for cases in which the number of observations is smaller than the number of quantiles used, because the estimator does not make much sense in such cases.

The estimators are biased downwards and the log-scale clearly shows how the bias becomes severe (only) for small numbers of observations. It is remarkable that the bias increases with the number of quantiles used, while the range-based estimator is least affected. This may appear counterintuitive, because one would expect that by adding more and more quantiles that are less biased than the maximum and minimum the overall estimator will be less biased than the range-based estimator. This turns out not to be the case. The reason becomes apparent when we look at the optimal weights used in the estimation. For example, when σ is estimated with more quantiles, more weight is attached to the maximum and minimum (compensated by negative weights on other

quantiles). Similarly, for σ^2 more weight is attached to the product of the maximum and minimum. Furthermore, time discreteness again has little impact on the variance of the estimators. Only for the estimators of σ^2 the variance decreases with the number of observations. For example, the variance of the 41-quantile-based estimator is $0.154\sigma^4$ with $23,401$ observations, while $0.095\sigma^4$ with 101 observations. Given the observation above that only the variance of the extrema-based estimator is somewhat affected by time discreteness, the bias of the variance can be linked to the strong reliance of these estimators on extrema.

Figure 4.2: Expected value of quantile-based estimators of σ (top-panel) and σ^2 (bottom panel) as a function of the number of observations (log-scale)



The simulation results make it clear that if the estimators are applied to a day of observations of an actively traded stock, the bias is small independent of which quantiles are used. This also holds for the consistent estimators based on permuted Brownian bridge. The estimators that take the average of subinterval estimates are robust except if the extrema are among the quantiles used and receive a considerable weight or if the number of observations per subinterval is very low. The smaller the number of observations per subinterval, the higher the discreteness bias. If we do

not correct the bias, then there exists a trade-off between a bias that is increasing in the number of subintervals and an efficiency loss that is decreasing in the number of subintervals. Consequently, there exists an optimal subinterval size, which one could determine by minimizing the root mean squared error of the estimates with respect to the subinterval length. In the next subsection two bias corrections are proposed to deal with the nonrobust cases.

4.2.2 Unbiased estimators

There are several ways to take the bias into account. Following Garman and Klass (1980), one can simply divide volatility estimates by the corresponding simulated expected volatility of Figure 4.2 to correct the bias. These corrections are only suited for equally spaced observations, however, while transaction data are not equally spaced. The simulations can of course be adjusted to the sampling scheme.

Alternatively, one can derive bias-corrections for the quantiles analytically. Rogers and Satchell (1991) derived a bias-corrected version of the range-based estimator assuming equidistant prices. However, it does not appear to be easily generalized to estimators based on other quantiles. In this subsection, we use results of Dassios (1996b) for Poisson distributed observations to derive bias-corrected estimators.

From Dassios (1996b) we can derive that $Q(\alpha,\lambda)$, the α-quantile of a standard Brownian motion observed at a Poisson rate λ over the time interval $[0,1]$, satisfies

$$E\left[Q(\alpha,\lambda)\right] = \sqrt{\frac{2\alpha}{\pi}} - \sqrt{\frac{2(1-\alpha)}{\pi}} - \sqrt{\frac{2}{\lambda}}\left[\Phi\left(\sqrt{2\alpha\lambda}\right) - \Phi\left(\sqrt{2(1-\alpha)\lambda}\right)\right], \tag{4.1}$$

$$\begin{aligned} E\left[Q(\alpha,\lambda)^2\right] = {} & 1 - \alpha\exp\left(-\frac{\lambda\alpha}{2}\right)\left(I_0\left(\frac{\alpha\lambda}{2}\right) + I_1\left(\frac{\alpha\lambda}{2}\right)\right) \\ & - (1-\alpha)\exp\left(-\frac{\lambda(1-\alpha)}{2}\right)\left(I_0\left(\frac{(1-\alpha)\lambda}{2}\right) + I_1\left(\frac{(1-\alpha)\lambda}{2}\right)\right) \\ & -2\left(\sqrt{\frac{2\alpha}{\pi}} - \sqrt{\frac{2}{\lambda}}\Phi\left(\sqrt{2\alpha\lambda}\right)\right)\left(\sqrt{\frac{2(1-\alpha)}{\pi}} - \sqrt{\frac{2}{\lambda}}\Phi\left(\sqrt{2(1-\alpha)\lambda}\right)\right), \end{aligned} \tag{4.2}$$

$$E\left[Q\left(\alpha,\lambda\right)^4\right] = 3\alpha^2\left(1-F\left(\frac{1}{2},3,-\lambda\alpha\right)\right)+3(1-\alpha)^2\left(1-F\left(\frac{1}{2},3,-\lambda(1-\alpha)\right)\right) \tag{4.3}$$
$$-16\sqrt{\frac{\alpha^3}{2\pi}}\left(1-F\left(\frac{1}{2},\frac{5}{2},-\lambda\alpha\right)\right)\left(\sqrt{\frac{2(1-\alpha)}{\pi}}-\sqrt{\frac{2}{\lambda}}\Phi\left(\sqrt{2(1-\alpha)\lambda}\right)\right)$$
$$-16\sqrt{\frac{(1-\alpha)^3}{2\pi}}\left(1-F\left(\frac{1}{2},\frac{5}{2},-\lambda(1-\alpha)\right)\right)\left(\sqrt{\frac{2\alpha}{\pi}}-\sqrt{\frac{2}{\lambda}}\Phi\left(\sqrt{2\alpha\lambda}\right)\right)$$
$$+6\left(1-\alpha-(1-\alpha)\exp\left(-\tfrac{\lambda(1-\alpha)}{2}\right)\left(I_0\left(\tfrac{(1-\alpha)\lambda}{2}\right)+I_1\left(\tfrac{(1-\alpha)\lambda}{2}\right)\right)\right),$$
$$*\left(\alpha-\alpha\exp\left(-\tfrac{\lambda\alpha}{2}\right)\left(I_0\left(\tfrac{\alpha\lambda}{2}\right)+I_1\left(\tfrac{\alpha\lambda}{2}\right)\right)\right)$$

where $\Phi\left(x\right)$ is the standard normal cumulative density function, $I_r(x)$ is the modified Bessel function of the first kind and $F(n,d,x)$ is the generalized hypergeometric function. Since the expected value of $M\left(\alpha,\lambda\right)$, the α-quantile of a Brownian motion with diffusion coefficient σ observed at a Poisson rate λ, is linear in σ, an unbiased 1-quantile-based estimator of σ follows from (4.1) as

$$\tilde{\sigma}(\alpha,\lambda) = \frac{M(\alpha,\lambda)}{\sqrt{\frac{2\alpha}{\pi}}-\sqrt{\frac{2(1-\alpha)}{\pi}}-\sqrt{\frac{2}{\lambda}}\left[\Phi\left(\sqrt{2\alpha\lambda}\right)-\Phi\left(\sqrt{2(1-\alpha)\lambda}\right)\right]}, \tag{4.4}$$

with

$$\mathrm{Var}(\tilde{\sigma}(\alpha,\lambda)) = \frac{\mathrm{Var}(Q\left(\alpha,\lambda\right))\sigma^2}{\left(\sqrt{\frac{2\alpha}{\pi}}-\sqrt{\frac{2(1-\alpha)}{\pi}}-\sqrt{\frac{2}{\lambda}}\left[\Phi\left(\sqrt{2\alpha\lambda}\right)-\Phi\left(\sqrt{2(1-\alpha)\lambda}\right)\right]\right)^2},$$

where $\mathrm{Var}(Q\left(\alpha,\lambda\right))$ follows from (4.1) - (4.2). Linear combinations of the 1-quantile-based estimator will also be unbiased. The optimal weights, $w=(\iota' V^{-1}\iota)^{-1}V^{-1}\iota$, can be derived as before. Let $C(\alpha_i,\alpha_j;\lambda)$ be the covariance between the α_i- and α_j-quantile of a standard Brownian motion. Then the (i,j)-th element of V is

$$V_{ij} = \frac{C(\alpha_i,\alpha_j;\lambda)}{E[Q(\alpha_i,\lambda)]E[Q(\alpha_j,\lambda)]},$$

where $E[Q(\alpha_i,\lambda)]$ is given by (4.1). The covariance between two quantiles is still unknown and needs to be obtained by simulation.

Similarly, from the expected value of the product of two quantiles, we can derive unbiased estimators of σ^2, noting that

$$E[M\left(\alpha_i,\lambda\right)M\left(\alpha_j,\lambda\right)] = \left(C(\alpha_i,\alpha_j;\lambda)+E[Q(\alpha_i,\lambda)]E[Q(\alpha_j,\lambda)]\right)\sigma^2.$$

Provided α_i and α_j are such that $E[M\left(\alpha_i,\lambda\right)M\left(\alpha_j,\lambda\right)]\neq 0$, an unbiased estimator of σ^2 follows as

$$\tilde{\sigma}^2(\alpha_i,\alpha_j;\lambda) = \frac{M(\alpha_i,\lambda)M(\alpha_j,\lambda)}{C(\alpha_i,\alpha_j;\lambda)+E[Q(\alpha_i,\lambda)]E[Q(\alpha_j,\lambda)]}. \tag{4.5}$$

Let $D(\alpha_1, \alpha_2, \alpha_3, \alpha_4; \lambda)$ be the covariance between the product of the α_1- and α_2-quantiles and the product of the α_3-and α_4-quantiles of a standard Brownian motion observed at Poisson rate λ, then

$$\text{Var}(\tilde{\sigma}^2(\alpha_i, \alpha_j; \lambda)) = \frac{D(\alpha_i, \alpha_j, \alpha_i, \alpha_j; \lambda)\sigma^4}{\left(C(\alpha_i, \alpha_j; \lambda) + E[Q(\alpha_i, \lambda)]E[Q(\alpha_j, \lambda)]\right)^2}.$$

Again linear combinations can be taken of the estimators based on the product of two quantiles. The optimal weights for such estimators, $w = (\iota' V^{-1}\iota)^{-1}V^{-1}\iota$, can be derived as before and V has elements

$$\text{Cov}\left(\tilde{\sigma}^2(\alpha_i, \alpha_j; \lambda), \tilde{\sigma}^2(\alpha_k, \alpha_l; \lambda)\right) = d\left(\alpha_i, \alpha_j; \lambda\right) D(\alpha_i, \alpha_j, \alpha_k, \alpha_l; \lambda) d(\alpha_k \alpha_l; \lambda),$$

where

$$d\left(\alpha_i, \alpha_j; \lambda\right) = \left(C(\alpha_i, \alpha_j; \lambda) + E[Q(\alpha_i, \lambda)]E[Q(\alpha_j, \lambda)]\right)^{-1}.$$

The 1-quantile-based estimator of σ^2 ($\alpha_i = \alpha_j$) is easily found analytically. From (4.5) it follows that

$$\tilde{\sigma}^2(\alpha_i, \alpha_i; \lambda) = \frac{M(\alpha_i, \lambda)^2}{E\left[Q\left(\alpha_i, \lambda\right)^2\right]}, \tag{4.6}$$

with

$$\text{Var}\left(\tilde{\sigma}^2(\alpha_i, \alpha_i; \lambda)\right) = \left[\frac{E\left[Q\left(\alpha_i, \lambda\right)^4\right]}{E\left[Q\left(\alpha_i, \lambda\right)^2\right]^2} - 1\right]\sigma^4,$$

where the moments are given by (4.2) and (4.3).

In practice, an estimate of λ is required. The natural choice is the maximum likelihood estimator, $\hat{\lambda} = n$. The literature on duration between transactions points out that the arrival of transactions is typically not Poisson distributed, but that durations are time-varying, autocorrelated and follow an intradaily pattern. This may seem to invalidate the suggested bias correction. However, kernel densities of time-of-day adjusted trade durations for a number of NYSE stocks presented in Bauwens and Giot (2001) show, mostly, decreasing density functions and the exponential distribution would therefore provide a reasonable fit. If durations follow a seasonal pattern and λ varies, then $\widehat{\lambda}$ will be an estimate of the average λ, which is representative for the interval. For subintervals, λ can be re-estimated each time, thereby allowing variation of the Poisson rate. Note, however, that this makes the estimation more simulation intensive if (4.5) is used or optimal weights are computed, because both depend on the covariance between two quantiles, which needs to be simulated and depends on λ. In order to avoid the simulations needed to apply estimator (4.5), one can also use (4.6) and linear combinations of it. Finally, it should be remarked that the results in Dassios (1996b) allow

the derivation of estimators based on other trade arrival distributions than the Poisson distribution.

4.3 Quantile-based estimators in the presence of market microstructure noise

We need to deal with a second problem before we can apply quantile-based estimators to high frequency data. At high frequency, certain market microstructure effects have an impact on the price process. As a result, the observed price is said to differ from the "true" price. The difference is market microstructure noise. While these market microstructure effects are the matter of interest in much of the market microstructure literature, for the estimation of volatility it is a nuisance and is treated as noise. First, I discuss several sources of market microstructure noise. Secondly, I investigate the effect of the presence of noise on our estimators by simulation. Finally, I propose noise-bias corrections.

4.3.1 Sources of market microstructure noise

The transaction price and its variance heavily depend on the way trades are organized by the market. Here I describe some of the effects reported in the literature.

Apart from time discreteness there is also price discreteness. As is well known, prices can only change by a certain tick size. If we interpret the observed price as the true price rounded to the nearest tick, then we observe the price with some error. This measurement error would introduce a bias. Financial markets have decreased their tick sizes through time, however, and at e.g. the NYSE the tick size is as small as one cent. This type of discreteness should therefore loose importance, but is still relevant for low-priced stocks. Alternatively, one can argue that a transaction will only occur when both parties agree on the price such that deals can only occur on the price grid. In this sense, price discreteness can only withhold or postpone trades and hence is transformed into time discreteness.

A more important microstructure effect is the bid-ask bounce, which is caused by the sequence of buys and sells and implies that the observed prices bounce between bid and ask quotes (Roll, 1984). The latter causes negative autocorrelation in the returns and artificially increases the volatility of high-frequency returns. As a result, the realized volatility measure based on the sum of squared high-frequency returns has an upward bias. This bias increases with the sampling frequency, because the microstructure effect

becomes a dominant part of the observed return. As the data is sampled more finely, the change in 'true' returns becomes smaller while the microstructure noise remains of the same magnitude.

There are more effects that can cause measurement errors and dependence in returns. For example, think of the strategic behaviour of market participants (inventory control, informational asymmetries, etc.). Similarly, block trades and the process of 'working an order', i.e. traders who distribute their order over time, potentially create positive autocorrelation in returns (Hasbrouck and Ho, 1987). Negative autocorrelation in high frequency returns could also be explained by mean reversion of stock prices after they were disturbed by trading behaviour (Poterba and Summers, 1988). Furthermore, different trade and quote sizes imply differences in the representativeness of prices, which introduces measurement errors.

4.3.2 Noise bias

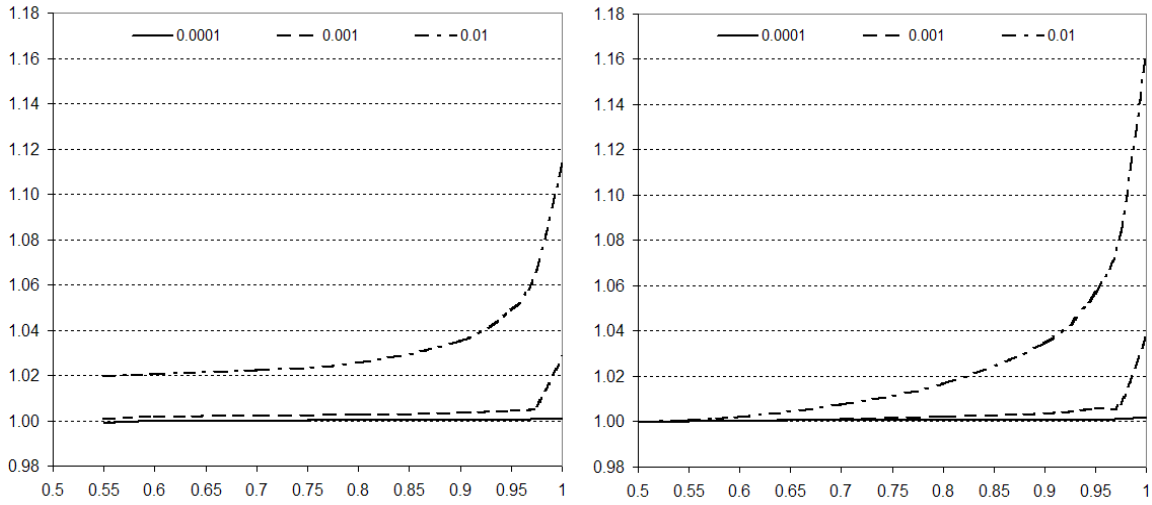
In the presence of market microstructure noise the observed price differs from the true price. Consequently, the observed quantiles also differ from the true quantiles. It is, for example, intuitively clear that when we add random noise to the true price series the observed maximum will be typically higher than the true maximum. The following simulation excercise gives an idea about the bias of different 1-quantile based estimators. Let $s_t = p_t + e_t$ be the observed price at time t, $0 \leq t \leq 1$, with p_t a Brownian motion with diffusion coefficient σ representing the true price and e_t the noise process with density g. A common assumption is that the noise is $i.i.d.$ with

$$E[e_t] = 0, \qquad E[e_t^2] = \omega^2, \qquad e \perp p. \tag{4.7}$$

We also need to assume that the support of the density g is bounded. Otherwise it implies that $\Pr[s_t > b] > 0$ for any b, hence implying an infinite expected maximum (and minimum) in continuous time. Interestingly, this suggests that the expected value of any other quantile is finite despite an unbounded support for g. Unfortunately, there are no analytical results available concerning quantiles in the presence of noise. Some densities used in the literature are to treat e_t as Bernoulli distributed, uniformly distributed or to ignore the bounded support condition and assume g is Gaussian. I simulate price series s_t assuming a standard random walk for p_t and a Bernoulli distribution for e_t with $\Pr[e_t = \omega] = \Pr[e_t = -\omega] = 1/2$. The size of the noise and the price signal are taken from the web-based appendices of Barndorff-Nielsen et al. (2007) that contain estimates of the daily noise variance ω^2, for the 30 Dow Jones Industrial

Average stocks, ranging between 0.0005 and 0.0197, while estimates of daily integrated volatility vary around $\sigma = 1$ percentage point. This implies that the noise is small relative to the daily volatility for actively traded stocks. I consider three sizes for the noise variance: $\omega^2 = 0.0001$, 0.001 and 0.01. Noise with variance $\omega^2 = 0.01$ should be considered as really big. Figure 4.3 shows simulation results for the expected value of 1-quantile based estimators, with α between 0.5 and 1, when the number of observations is large ($n = 23,401$). We notice that the estimator based on the maximum is biased most and as we move away from the maximum the bias drops sharply. As expected, the bias increases with the noise, but remains relatively small for normal levels of noise. Similarly, the simulation results in Alizadeh et al. (2002) show a small bias for the range-based estimator.

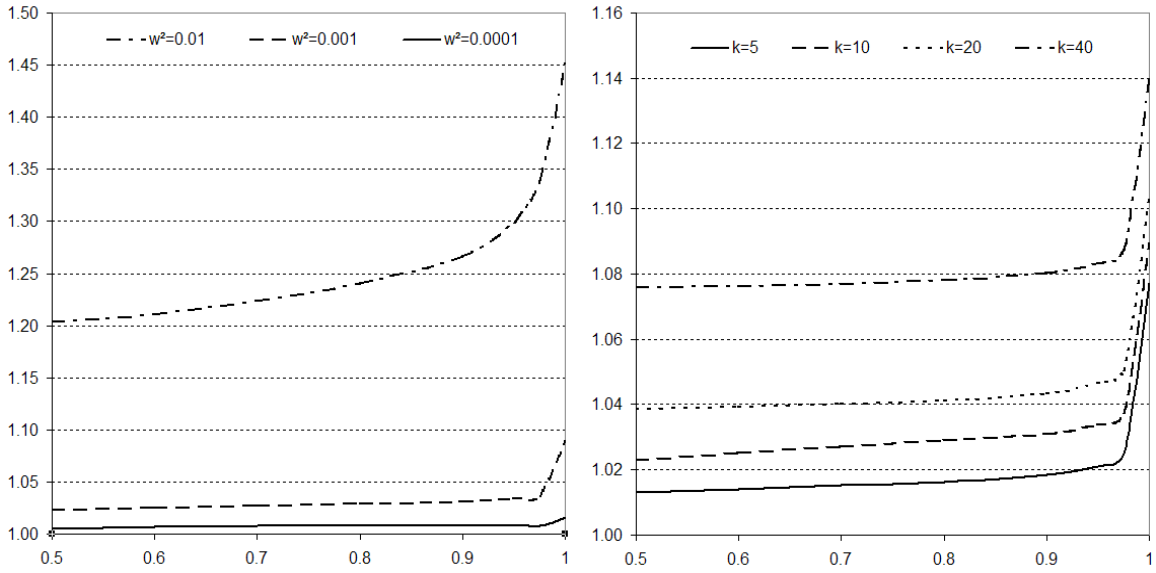
Figure 4.3: Expected value of 1-quantile based estimators of σ (left panel) and σ^2 (right panel) for 3 noise levels when $\sigma = 1$



Note, however, that the robustness only holds for estimators that use quantiles computed over a long interval like a day. In this case, the price signal dominates the noise and the bias remains relatively small. On the contrary, the estimator based on subintervals can be severely biased. The more subintervals we consider, the smaller the price signal in each subinterval, while the size of the noise remains equal. The dominance of the noise in each subinterval implies a severe bias for the estimator based on averaging subinterval estimates, which increases in the number of subintervals. Note that the discussion was based on simulation results for Brownian motion, but clearly that Brownian bridge is affected in much the same way. This implies that the higher efficiency achieved by estimation based on subintervals can only be exploited if we find

a bias correction. Bias corrections are proposed in the next subsection.

Figure 4.4: Expected value of permuted Brownian bridge estimator of σ^2 per quantile for different noise levels (left panel) and different numbers of subintervals (right panel)



The estimator based on permuted Brownian bridge is biased, because the observed start and end prices of the subintervals differ from the true prices. When the subintervals are permuted and glued together, then there will typically be a noise-induced jump in the new price series at the transition points. This increases the variance of the permuted series artificially and implies a bias for our volatility estimators. The simulation results in Figure 4.4 show how the bias increases with the variance of the noise (ω^2) and the number of subintervals (k). For three sizes of the noise, the left panel shows the expected value of the estimator, which takes the average of 10 random permutations of $k = 10$ subintervals. The right panel shows what happens to the expected value if the noise variance is kept at 0.001, but k varies. The rest of the simulation set-up is as before. The results clearly show how the estimator based on the maximum is again much more biased than those based on other quantiles. The bias also takes off for modest choices of k. If the extrema are avoided and k is kept small, however, then the estimator is reasonably robust to noise. For the other cases a bias correction is required.

4.3.3 Bias corrections

It is hard to derive properties of the quantiles in the presence of noise. Even the simplest type of noise renders the path of the observed price discontinuous. The key analytical results we built on to develop quantile-based estimators do not hold any

longer when noise is superimposed and the probability literature does not seem to contain the required extensions. As long as an analytically derived bias correction is not available, I propose a simulation-based bias correction. Bias corrections for the range-based estimator have been presented in the literature already and I discuss these first.

Bias-corrected range-based volatility estimators

Christensen et al. (2006) propose bias corrections for the realized range estimator for a few cases where the noise distribution has bounded support. For example, if the noise is Bernoulli distributed on $\{-\omega, \omega\}$, then the bias of the range equals 2ω. Further, under general noise distribution, we know from Zhang et al. (2005) and Oomen (2005) that consistent estimators of the variance of the noise, ω^2, are given by

$$\begin{aligned} \widehat{\omega}_N^2 &= \frac{1}{2N}\sum_{i=1}^{N} r_i^2 = \frac{RV}{2N} \xrightarrow{p} \omega^2, \\ \widetilde{\omega}_N^2 &= -\frac{1}{N-1}\sum_{i=1}^{N-1} r_i r_{i+1} \xrightarrow{p} \omega^2, \end{aligned} \tag{4.8}$$

as $N \to \infty$, with r a high-frequency return, RV the realized volatility estimator and N the number of returns. This allows one to obtain estimates of ω and to correct the observed range. In order to deal with the discreteness bias of the bias-corrected range-based estimator a scaling factor is obtained by simulation.

Martens and van Dijk (2007) propose another bias correction for the realized range estimator of daily volatility. They exploit the fact that the daily range is robust to noise relative to the range over shorter time intervals and define a scaling factor as the average of the estimator based on the daily range divided by the average of the realized range estimator. The average is taken over the previous q days. The estimator which is otherwise upward biased now is scaled down appropriately. The approach of Martens and van Dijk (2007) can be applied to other quantile-based estimators. The choice of q remains arbitrary, however. Furthermore, with noise that varies in size from day to day, a correction that depends on the past may not be reliable.

Simulation-based bias correction for quantile-based volatility estimators

The bias correction of the range proposed by Christensen et al. (2006) cannot be easily generalized to other quantiles. Given an estimate of ω, we do not know a priori how much it affects a certain quantile. Below I describe simulation-based bias corrections and make a distinction whether the type of noise distribution is known or not.

Both methods can be applied to both averaging-subinterval-estimates and permuted-Brownian-bridge estimators. I discuss the case of estimators of σ^2, but bias corrections for estimators of σ can be set up in a similar way.

The type of noise distribution is known

Let $\check{\sigma}_s^2$ be a quantile-based estimator of σ^2 obtained from n observed prices from the price process s_t, $0 \leq t \leq 1$, which is subject to noise. Then given the noise described above by (4.7), its bias is a function of ω, σ and the number of observations, n:

$$E[\check{\sigma}_s^2] = \sigma^2 + f_n(\omega, \sigma). \tag{4.9}$$

By an equivariance argument,

$$f_n(\omega, \sigma) = f_n(\omega/\sigma, 1)\,\sigma^2 = g_n(\omega/\sigma)\,\sigma^2,$$

say. Hence, given estimates $\check{\sigma}_s^2$ and $\tilde{\omega}$, a bias-corrected estimator $\hat{\sigma}_s^2$ is obtained by solving

$$\check{\sigma}_s^2 = \sigma^2 [1 + g_n(\tilde{\omega}/\sigma)]$$

for σ^2. The values of $g_n(\cdot)$ are obtained by simulation. Note that $g_n(0) \approx 0$ for large n, and that $g_n(\omega/\sigma)$ is increasing in ω/σ.

The type of noise distribution is unknown

The bias correction presented above is vulnerable to misspecification of the noise distribution. Here, I develop a bias correction that assumes noise to be symmetrically distributed about zero and to be independent over time, but does not require the noise to be known. Instead, the noise density is estimated. The fact that high-frequency returns of an actively traded stock are dominated by market microstructure noise allows us to get a fairly good idea about the noise distribution. In fact, if we ignore the price signal completely, then the return in period t is the sum of the noise in the periods $t-1$ and t, and the return density function (v) equals the self-convolution of the noise density function (u), i.e. $v = u * u$. This implies that by computing the (symmetric) self-deconvolution of the high-frequency return density an estimate of the noise density is obtained. In turn, this noise density estimate can be used to obtain values of $f_n(\cdot)$ by simulation, where $f_n(\sigma)$ is the bias of $\check{\sigma}_s^2$, i.e. $E[\check{\sigma}_s^2] = \sigma^2 + f_n(\sigma)$. Thus, given the estimate $\check{\sigma}_s^2$, a bias-corrected estimator $\hat{\sigma}_s^2$ is obtained by solving

$$\check{\sigma}_s^2 = \sigma^2 + f_n(\sigma) \tag{4.10}$$

for σ^2. A self-deconvolution method for density functions does, to my knowledge, not exist in the literature. As part of the simulation exercise in the next subsection, I apply a simple search algorithm that is described in appendix.

4.3.4 Simulation study

In this subsection, I analyse the performance of bias-corrected estimators when the noise distribution is unknown.

Simulation design

The study takes $1,000$ replications of a standard random walk to replicate price processes. Each random walk consists of $23,400$ intervals representing price series observed each second of a 6.5-hour trading day. A process with time-varying volatility, as sometimes used in the literature, would lead to similar results. As estimators of σ^2, I consider the estimator averaging 10-minute subinterval estimates ($n = 39$) and the permuted Brownian bridge estimator based on 10-minute subintervals ($k = 39$) and 10 permutations. Both estimators are based on 2 quantiles ($\alpha = 0.1, 0.9$) and let us denote the estimators by SUB and PERM. Of course many other specifications of these estimators could have been considered instead. The chosen estimators are based on a relatively large number of subintervals, which implies a big noise bias, even more so for the SUB estimator than the PERM estimator. The estimators are applied after perturbing the random walk by noise. As noise distributions the Bernoulli and Normal distributions are considered. In order to solve (4.10), Newton's method is applied and at each iteration $f_n(\cdot)$ is obtained by simulation based on $1,000$ random walks perturbed by noise drawn from the noise density estimate. The latter is obtained as the self-deconvolution of the return distribution according to the method described in appendix.

First, each estimator (without noise bias correction) is applied to the true price process to serve as a benchmark. Secondly, each noise-bias-corrected estimator is applied to the perturbed price series for the cases where the noise variance is $\omega^2 = 0.001$ or 0.01. This size of noise can be considered to be of respectively average and really big order compared to the variance of the true price $\sigma^2 = 1$. For each case, the average, the standard error and the root mean squared error of the estimates are computed and presented in Table 4.1.

Results

The simulation results are presented in Table 4.1. The first column shows the mean, standard deviation and root mean squared error of the SUB estimator and the PERM estimator when there is no noise. The other columns show the results for the bias-corrected versions of the estimators when the price is perturbed by Bernoulli or Normally distributed noise with variance (ω^2) 0.001 or 0.01.

Table 4.1: Simulation results for noise-bias-corrected quantile-based estimators

	Noise	**-**	**Bernoulli**		**Normal**	
	$\boldsymbol{\omega^2}$	**-**	**0.001**	**0.01**	**0.001**	**0.01**
SUB	mean	1.002	1.031	1.010	0.998	0.998
	st.dev.	0.133	0.139	0.158	0.134	0.143
	RMSE	0.133	0.142	0.158	0.134	0.143
PERM	mean	0.995	0.992	0.984	0.996	1.002
	st.dev.	0.269	0.284	0.459	0.289	0.492
	RMSE	0.269	0.285	0.459	0.289	0.492

Presented are the mean, standard deviation and the root mean squared error (RMSE) of estimators SUB and PERM. The estimator SUB is based on averaging 10-minute subinterval estimates ($n = 39$) and the estimator PERM is based on 10-minute subintervals ($k = 39$) and averaging estimates from 10 permutations. Both estimators are based on 2 quantiles ($\alpha = 0.1, 0.9$). The estimators are applied to standard random walks and to the same series perturbed by Bernoulli distributed noise or Gaussian noise. The variance of the noise (ω^2) takes values 0.001 or 0.01.

In the case without noise, we notice that the SUB estimator is more efficient than the PERM estimator, which purely depends on the specification of the estimator. Another specification of the PERM estimator based on more subintervals and permutations would increase the estimator's efficiency. However, our focus here is on how the noise-bias-corrected estimators perform in the presence of noise. The average of the estimates is very close to the expected value of 1 in all cases and hence the bias-corrected estimators are nearly unbiased. The bias corrections are able to deal with both noise that is Bernoulli and Normally distributed. The standard error of the bias-corrected SUB estimator increases little with the size of the noise, while the variance of the PERM estimator is considerably higher if the noise is big. The variance of the bias-corrected estimators could, to a certain extent, be decreased by using a more precise bias estimate, which could be obtained by increasing the number of simulations used to estimate the bias $f_n(\cdot)$.

4.4 Conclusion

This chapter tackled the two main problems encountered when applying quantile-based estimators, which are derived in continuous time, to noisy discrete-time data. First, the chapter dealt with the time-discreteness bias. It turns out that most quantile-based estimators are robust to discreteness and bias-corrections are not always needed in practice unless the maximum and/or minimum are used in the estimation or the number of observations is very small. In cases where a correction is needed, a scaling factor can be obtained by simulation. Moreover, unbiased quantile-based estimators in discrete time were derived in case the observations are Poisson distributed. Secondly, the bias introduced by market microstructure noise was dealt with. The fact that the observed price differs from the 'true' price due to market microstructure effects introduces a bias for volatility estimators that increases with the number of subintervals used in the estimation. Simulation-based bias corrections were presented both in case the noise distribution is known and unknown.

Appendix

Noise density estimation

Let $t_1, ..., t_n$ be the sampling times of the noise-perturbed continuous-time price process $s_t = p_t + e_t$, $0 \leq t \leq 1$, with p_t the true price process and e_t $i.i.d.$ noise with symmetric density u. Let $r_{t_2} = s_{t_2} - s_{t_1}, ..., r_{t_n} = s_{t_n} - s_{t_{n-1}}$ be the returns corresponding to the observed prices $s_{t_1}, ..., s_{t_n}$. The price signal decreases as the sampling frequency increases, while the noise remains of the same size. That is, at high frequency the noise dominates and a return mainly exists of two noise components, i.e. $r_{t_i} \approx e_{t_i} - e_{t_{i-1}}$, $i = 2, ..., n$. Therefore, the higher the sampling frequency, the closer the return density v is to the self-convolution of u, i.e. $v \approx u * u$. This implies that the noise density can be estimated as the self-deconvolution of the return density. Note that, while v is symmetric, the empirical return distribution is skewed (the skewness disappears as $\max_i |t_i - t_{i-1}| \to 0$). Therefore, the empirical return distribution is symmetrized by replacing the returns $r_{t_2}, ..., r_{t_n}$ by the symmetrized returns $\pm |r_{t_2}|, ..., |r_{t_n}|$. A standard kernel estimator is used to estimate the density of symmetrized returns, with the Normal density function as kernel smoother and a bandwidth parameter that is optimal for estimating Normal densities. The self-deconvolution is then applied to the empirical density v of the symmetrized returns.

Self-deconvolution

The self-deconvolution applied in the simulation exercise of the main text is based on a property of the Fourier transformation. Let $U = \mathcal{F}(u)$ and $V = \mathcal{F}(v)$ be the Fourier transforms of u and v, with $v = u * u$. It holds that $V = UU$ and $U = V^{1/2}$, where the operations are element-by-element, such that by the inverse Fourier transformation, $u = \mathcal{F}^{-1}(V^{1/2})$. The latter expression implies that we can estimate the noise density u by taking the empirical density v, taking its Fourier transform V, taking the square root of the elements in V and applying the inverse Fourier transformation. However, there is one problem. Each element in V has a positive and a negative square root. If m is the number of elements in V, this leaves us with 2^m possible sign combinations and except for small m it is unfeasible to try them all out.

To solve this problem approximately, a search algorithm is applied to find the sign combination of the elements in $V^{1/2}$ that offers the best estimate of u. The algorithm starts from the vector with positive roots, $Y = V^{1/2}$, runs through the vector Y element-by-element and each time decides on the sign of the element. The sign of an element Y_i, $i \in \{1, ..., z\}$, is changed if it decreases the criterion $\sum_{i=1}^{z} |v_i - (y * y)_i|$, where $y = \mathcal{F}^{-1}(Y)$. This criterion determines whether the convolution $y * y$ fits v better when the sign of Y_i is changed. The algorithm stops at the end of vector Y if the criterion has not been reduced by more than 10^{-5}, otherwise it continues with another run through Y.

The algorithm does not guarantee the best possible fit between $y * y$ and v and hence not the optimal estimate y of u. In order to increase the likelihood of obtaining a good fit I let the algorithm start from a second set of starting values, i.e. instead of the positive roots $V^{1/2}$ it starts from randomly signed roots and checks whether the criterion can be reduced further in this way.

Conclusions

This Ph.D. thesis presented research on financial transaction data and volatility. Chapter 1 concentrated on the data and offered a solution to a particular matching problem, while Chapter 2 focused on volatility, in particular the estimation of stochastic volatility models. Chapters 3 and 4 focused on both transaction data and volatility and developed volatility estimators that are meant to be applied to high-frequency data.

Chapter 1 dealt with the matching of trades and quotes of NYSE stocks. Inefficient matching could, depending on the application, affect the outcome of market microstructure analysis. A procedure was proposed that tests whether the quote revision frequency around a trade is contaminated by quote revisions triggered by a trade, and then determines the smallest timing adjustment needed to take this contamination into account. The procedure was applied to a sample of 25 stocks in 5 sample periods. The results showed that the difference between trade reporting lags and quote reporting lags varies across stocks and time. The variation could be mainly explained by changes in the reporting procedures of the NYSE and the co-existence of two reporting systems for trades, i.e. trades are Display Book reported or not. The non Display Book reported trades usually required a larger adjustment to match trade and quote times. The procedure can be applied to each trade type, stock and subsample to determine and implement the appropriate timing adjustment. Since mid-2003 the reporting of lags stabilized and, from then on, taking the prevailing quote at two seconds before the trade as the prevailing quote at the time of the trade appeared to be an appropriate adjustment rule for all stocks.

Chapter 2 contributed to the literature on the GMM estimation of stochastic volatility models. The class of moment conditions has been extended and a closed-form expression for the optimal weighting matrix for any subset of those conditions has been derived and, as a by-product, an expression for the GMM asymptotic covariance matrix. These expressions can be used for improved GMM estimation of the SV model with AR(1) log-volatility and to compute GMM standard errors more accurately. The comparison to other estimators showed the relatively small efficiency loss of the GMM

estimator compared to the asymptotically efficient MCMC method. Monte Carlo results illustrated the efficiency gain of iterated GMM based on the analytical optimal weighting matrix compared to iterated GMM based on estimation of the optimal weighting matrix. The analytical results regarding the optimal weighting matrix allow us to fastly and accurately assess the information content of any subset of moment conditions considered. This, in turn, permits the optimal selection of a small set of highly informative moment conditions from very large sets, and subsequent GMM estimates to be based on the optimal selection of moments. A four-step data-based procedure was proposed for the optimal selection of moment conditions. It was found that the search for the optimal selection of moment conditions comes at a small efficiency cost compared to the situation where the optimal selection is known a priori. In other words, Chapter 2 pointed out that the relatively simple GMM estimator deserves to be reconsidered in the SV context.

Chapter 3 developed quantile-based volatility estimators. It turned out that the estimators based on the quantiles of Brownian bridge are more efficient than their Brownian motion counterparts. Furthermore, we concluded that the supremum and infimum are the quantiles that contain the most information, but when combined with other quantiles more efficient estimators can be obtained. If we consider the class of unbiased estimators based on only the supremum and the infimum, it turned out that the range-based estimator is not (always) the minimum-variance unbiased estimator. This implies that the historic 'high-low' data can be used more efficiently than standard practice did so far. I have also shown how consistent estimators are obtained by taking the average of subinterval estimates or, alternatively, by taking averages of estimates from permutations of subintervals of the Brownian bridge. The quantile-based volatility estimators provide an interesting alternative to the existing realized volatility and realized range estimators.

The estimators of Chapter 3 are biased, however, when applied to high-frequency data. Chapter 4 tackled the underlying time-discreteness bias and noise bias. Time discreteness turned out to be a problem only if the maximum and/or minimum were used in the estimation or the number of observations was very small. In cases where a correction was needed, a scaling factor could be obtained by simulation. Moreover, unbiased quantile-based estimators in discrete time were derived in case the observations were Poisson distributed. The noise bias is due to market microstructure effects blurring the 'true' price signal. Simulation-based bias corrections were presented both in the case where the noise distribution is known and where it is unknown. The bias cor-

rections allow the practitioner to exploit the efficiency gain of quantile-based volatility estimation at high frequency.

Bibliography

Alizadeh, S., M.W. Brandt, and F.X. Diebold (2002): "Range-based estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047-1091.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43-76.

Andersen, T.G., H.-J. Chung, and B.E. Sørensen (1999): "Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study," *Journal of Econometrics*, 91, 61–87.

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, 328–352.

Andersen, T.G., and B.E. Sørensen (1997): "GMM and QML ssymptotic standard deviations in stochastic volatility models: Comments on Ruiz (1994)," *Journal of Econometrics*, 76, 397–403.

Bacidore, J., K. Ross, and G. Sofianos (2003): "Quantifying Market Order Execution Quality at the New York Stock Exchange," *Journal of Financial Markets*, 6, 281-307.

Ball, C.A., and T. Chordia (2001): "True Spreads and Equilibrium Prices," *Journal of Finance*, 56(5), 1801-1835.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard (2006): "Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," *Working Paper, Nuffield College, University of Oxford*, and appendices on the web.

Bauwens, L., and, P. Giot (2001): "Econometric Modelling of Stock Market Intraday Activity," *Advanced Studies in Theoretical and Applied Econometrics*, 38, Kluwer Academic Publishers, Boston.

Bessembinder, H. (2003): "Issues in Assessing Trade Execution Costs," *Journal of Financial Markets*, 6, 233-257.

Blume, M., and M. Goldstein (1997): "Quotes, Order Flow, and Price Discovery," *Journal of Finance*, 52, 221-244.

Breusch, T, H.L. Qian, P. Schmidt, and D. Wyhowski (1999): "Redundancy of moment conditions," *Journal of Econometrics*, 91, 89–111.

Busse, J.A., and T.C. Green (2002): "Market Efficiency in Real Time," *Journal of Financial Economics*, 65(3), 415-437.

Chan K., Y.P. Chung, and W.M. Fong (2002): "The Informational Role of Stock and Option Volume," *Review of Financial Studies*, 15(4), 1049-1075.

Chib, S., F. Nardari, and N. Shephard (2002): "Markov chain Monte Carlo methods for stochastic volatility models," *Journal of Econometrics*, 108, 281–316.

Chordia, T., R. Roll, and A. Subrahmanyam (2001): "Market Liquidity and Trading Activity," *Journal of Finance*, 56(2), 501-530.

Chordia, T., R. Roll, and A. Subrahmanyam (2002): "Order Imbalance, liquidity, and Market Returns," *Journal of Financial Economics*, 65(1), 111-130.

Christensen, K., and M. Podolski (2005): "Asymptotic Theory for Range-Based Estimation of Integrated Variance of a Continuous Semi-Martingale," Working Paper.

Christensen, K., M. Podolski, and M. Vetter (2006): "Bias-Correcting the Realized Range-Based Variance in the Presence of Market Microstructure Noise," Working Paper.

Clark, P.K. (1973): "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica*, 41, 135–155.

Danielsson, J. (1994): "Stochastic volatility in asset prices: Estimation with simulated maximum likelihood," *Journal of Econometrics*, 54, 375–400.

Danielsson, J. and J.F. Richard (1993): "Accelerated Gaussian importance sampler with application to dynamic latent variable models," *Journal of Applied Econometrics*, 8 (Suppl.), S153–S173.

Dassios, A. (1995): "The Distribution of the Quantile of a Brownian Motion with Drift and the Pricing of Related Path-Dependent Options," *The Annals of Probability,* 5(2), 389-398.

DASSIOS, A. (1996a): "Sample Quantiles of Stochastic Processes with Stationary and Independent Increments," *The Annals of Applied Probability*, 3, 1041-1043.

DASSIOS, A. (1996b): "Sample Quantiles of Additive Renewal Reward Processes," *Journal of Applied Probability*, 33(4), 1018-1032.

DUFFIE, D. AND K.J. SINGLETON (1993): "Simulated moments estimation of Markov models of asset prices," *Econometrica*, 61, 929–952.

EASLEY, D., M. O'HARA, AND G. SAAR (2001): "How Stock Splits Affect Trading: A Microstructure Approach," *Journal of Financial and Quantitative Analysis*, 36(1), 25-51.

EDELEN, R., AND S. GERVAIS (2003): "The Role of Trading Halts in Monitoring a Specialist Market," *Review of Financial Studies*, 16(1), 263-300.

ELLIS, K., R. MICHAELY, AND M. O'HARA (2000): "The Accuracy of Trade Classification Rules: Evidence from Nasdaq," *Journal of Financial and Quantitative Analysis*, 35(4), 529-552.

ENGLE, R.F., AND A.J. PATTON (2004): "Impacts of Trades in an Error-Correction Model of Quote Prices," *Journal of Financial Markets*, 7(1), 1-25.

FEDOROV, V.V. (1972): *Theory of optimal experiments*, Academic Press, New York.

FELLER, W. (1951): "The Asymptotic Distribution of the Range of Sums of Independent Random Variables," *Annals of Mathematical Statistics*, 22, 427-432.

FRIDMAN, M. AND L. HARRIS (1998): "A maximum likelihood approach for non-Gaussian stochastic volatility models," *Journal of Business & Economic Statistics*, 16, 284–291.

FUJITA, T. (2000): "A Note on the Joint Distribution of α, β -Percentiles and Its Application to the Option Pricing," *Asia-Pacific Financial Markets,* 7, 339-344.

GALLANT, A.R., D.A. HSIEH, AND G.E. TAUCHEN (1997): "Estimation of stochastic volatility models with diagnostics," *Journal of Econometrics*, 81, 159–192.

GARMAN, M.B., AND M.J. KLASS (1980): "On the Estimation of Security Price Volatilities from Historical Data," *Journal of Business*, 53, 67-78.

GHYSELS, E., A.C. HARVEY, AND E. RENAULT (1996): "Stochastic volatility," in: G.S. Maddala, ed., *Handbook of statistics: Statistical methods in finance*, Vol. 14, North-Holland, Amsterdam, 119–191.

GOURIÉROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect inference," *Journal of Applied Econometrics*, 8, S85–S118.

HANSEN, L.P. (1982): "Large sample properties of generalized methods of moments estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L.P., J. HEATON, AND A. YARON (1996): "Finite-sample properties of some alternative GMM estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

HARRIS, L. (2003): *Trading and Exchanges: Market Microstructure for Practitioners*, Oxford University Press.

HARVEY, A.C, E. RUIZ, AND N. SHEPHARD (1994): "Multivariate stochastic volatility models," *The Review of Economic Studies*, 61, 247–264.

HASBROUCK, J., AND T. HO (1987): "Order Arrival, Quote Behavior, and the Return-Generating Process," *Journal of Finance*, 42(4), 1035-1048.

HASBROUCK, J., G. SOFIANOS, AND D. SOSEBEE (1993): "New York Stock Exchange Systems and Trading Procedures," NYSE Working Paper 93-01.

HENKER, T., AND J.-X. WANG (2006): "On the Importance of Timing Specifications in Market Microstructure Research," *Journal of Financial Markets*, 9, 162-179.

HOOGHIEMSTRA, G. (2002): "On Explicit Occupation Time Distributions for Brownian Processes," *Statistics & Probability Letters*, 56, 405-417.

HUANG, R.D., AND H.R. STOLL (2001): "Exchange Rates and Firms' Liquidity: Evidence from ADRs," *Journal of International Money and Finance*, 20(3), 297-325.

HULL, J., AND A. WHITE (1987): "The pricing of options on assets with stochastic volatilities," *The Journal of Finance*, 17, 281–300.

JACQUIER, E., N.G. POLSON, AND P.E. ROSSI (1994): "Bayesian analysis of stochastic volatility models," *Journal of Business & Economic Statistics*, 12, 69–87.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic volatility: optimal likelihood inference and comparison with ARCH models," *The Review of Economic Studies*, 45, 361–393.

KRYZANOWSKI, L., AND H. ZHANG (2002): "Intraday Market Price Integration for Shares Cross-Listed Internationally," *Journal of Financial Quantitative Analysis*, 37(2), 243-369.

KUNITOMO, N. (1992): "Improving the Parkinson Method of Estimating Security Price Volatilities," *Journal of Business*, 65(2), 295-302.

LEE, C.M., AND M.J. READY (1991): "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 46, 733-746.

MARTENS, M., AND D. VAN DIJK (2007): "Measuring Volatility with the Realized Range," *Journal of Econometrics*, 138, 181-207.

MELINO, A., AND S.M. TURNBULL (1990): "Pricing foreign currency options with stochastic volatility," *Journal of Econometrics*, 45, 239–265.

MONFARDINI, C. (1998): "Estimating stochastic volatility models through indirect inference," *The Econometrics Journal*, 1, C113–C128.

NELSON, D.B. (1988): *Time series behavior of stock market volatility and returns*, Ph.D. dissertation, M.I.T..

NEWEY, W.K., AND R.J. SMITH (2000): "Asymptotic bias and equivalence of GMM and GEL estimators," Discussion paper No. 01/517, University of Bristol.

NEWEY, W.K., AND R.J. SMITH (2001): "Higher order properties of GMM and generalized empirical likelihood estimators," Discussion paper, M.I.T..

NYHOLM, K. (2003): "Inferring the Private Information Content of Trades: A Regime-Switching Approach," *Journal of Applied Econometrics*, 18(4), 457-470.

NYSE (2000): "Stock Market Activity," *2000 Factbook*.

NYSE (2001): "Stock Market Activity," *2001 Factbook*.

NYSE (2003): *2003 Annual Report*.

NYSE (2003a): *Information Memo*, 03-21, May 15.

NYSE (2003b): *Information Memo*, 03-27, June 11.

NYSE (2003c): "How to Compete with the New York Stock Exchange," *The Exchange*, 10(8).

OOMEN, R.A.C. (2005): "Properties of Bias-Corrected Realized Variance under Alternative Sampling Schemes," *Journal of Financial Econometrics*, 4(3), 555-577.

PARKINSON, M. (1980): "The Extreme Value Method for Estimating the Variance of the Rate of Return," *Journal of Business*, 53(1), 61-65.

PETERSON, M., AND E. SIRRI (2003): "Evaluation of the Biases in Execution Cost Estimation Using Trade and Quote Data," *Journal of Financial Markets*, 6, 259-280.

PIWOWAR, M.S., AND L. WEI (2003): "The Sensitivity of Effective Spread Estimates to Trade-Quote Matching Algorithms," NYSE Working Paper, 1-45.

PORT, S.C. (1963): "An Elementary Approach to Fluctuation Theory," *Journal of Mathematical Analysis and Applications*, 6(1), 109-151.

POTERBA, J.M., AND L.H. SUMMERS (1988): "Mean Reversion in Stock Prices," *Journal of Financial Economics*, 22(1), 27-59.

ROGERS, L.C.G., AND S.E. SATCHELL (1991): "Estimating Variance from High, Low, and Closing Prices," *Annals of Applied Probability*, 1, 504-512.

ROLL, R. (1984): "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127-1139.

RUIZ, E. (1994): "Quasi-maximum likelihood estimation of stochastic volatility models," *Journal of Econometrics*, 63, 289–306.

SANDMANN, G., AND S.J. KOOPMAN (1998): "Estimation of stochastic volatility models via Monte Carlo maximum likelihood," *Journal of Econometrics*, 87, 271–301.

SCHULTZ, P. (2000): "Stock Splits, Tick Size, and Sponsorship," *Journal of Finance*, 55(1), 429-450.

SHEPHARD, N. (1996): "Statistical aspects of ARCH and stochastic volatility," in: D.R. Cox, O.E. Barndorff-Nielsen, and D.V. Hinkley, eds., *Time series models in econometrics, finance and other fields*, Chapman & Hall, London, 1–67.

STOLL, H.R. (2000): "Friction," *Journal of Finance*, 55(4), 1479-1514.

TAKÁCS, L. (1999): "The Distribution of the Sojourn Time for the Brownian Excursion," *Methodology and Computing in Applied Probability*, 1, 7-28.

TAUCHEN, G.E., AND M. PITTS (1983): "The price variability-volume relationship on speculative markets," *Econometrica*, 51, 484–505.

TAYLOR, S. (1986): *Modelling financial time series*, Wiley, Chichester.

TEWELES, R.J., AND E.S. BRADLEY (1998): *The Stock Market,* John Wiley & Sons, 7th Edition.

VENKATARAMAN, K. (2001): "Automated versus Floor Trading: An Analysis of Execution Costs on the Paris and New York Exchanges," *Journal of Finance*, 56(4), 1445-1485.

WENDEL, J.G. (1960): "Order Statistics of Partial Sums," *The Annals of Mathematical Statistics*, 31, 1034-1044.

WRIGHT, J.H. (1999): "A new estimator of the fractionally integrated stochastic volatility model," *Economics Letters*, 63, 295–303.

YOR, M. (1995): "The Distribution of Brownian Quantiles," *Journal of Applied Probability*, 2, 405-416.

ZHANG, L., P.A. MYKLAND AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data, *Journal of the American Statistical Association*, 472, 1394-1411.

List of Doctoral Dissertations

Doctoral dissertations from the Faculty of Economic and Applied Economic Sciences from August 1, 1971.

1. GEPTS, Stefaan, Stability and Efficiency of Resource Allocation Processes in Discrete Commodity Spaces. Leuven, KUL, 1971. 86 pp.

2. PEETERS, Theo, Determinanten van de Internationale Handel in Fabrikaten. Leuven, Acco, 1971. 290 pp.

3. VAN LOOY, Wim, Personeelsopleiding: Een Onderzoek naar Investeringsaspekten van Opleiding. Hasselt, Vereniging voor wetenschappelijk onderzoek in Limburg, 1971. VII, 238 pp.

4. THARAKAN, Mathew, Indian Exports to the European Community: Problems and Prospects. Leuven, Faculty of economics and applied economics, 1972. X,343 pp.

5. HERROELEN, Willy, Heuristische Programmatie: Methodologische Benadering en Praktische Toepassing op Complexe Combinatorische Problemen. Leuven, Aurelia scientifica, 1972. X, 367 pp.

6. VANDENBULCKE, Jacques, De Studie en de Evaluatie van Data-organisatiemethodes en Data-zoekmethodes. Leuven, s.n., 1973. 3 V.

7. PENNYCUICK, Roy A., The Economics of the Ecological Syndrome. Leuven, Acco, 1973. XII, 177 pp.

8. KAWATA, T. Bualum, Formation du Capital d'Origine Belge, Dette Publique et Stratégie du Développement au Zaire. Leuven, KUL, 1973. V, 342 pp.

9. DONCKELS, Rik, Doelmatige Oriëntering van de Sectorale Subsidiepolitiek in België: Een Theoretisch Onderzoek met Empirische Toetsing. Leuven, K.U.Leuven, 1974. VII, 156 pp.

10. VERHELST, Maurice, Contribution to the Analysis of Organizational Information Systems and their Financial Benefits. Leuven, K.U.Leuven, 1974. 2 V.

11. CLEMEUR, Hugo, Enkele Verzekeringstechnische Vraagstukken in het Licht van de Nutstheorie. Leuven, Aurelia scientifica, 1974. 193 pp.

12. HEYVAERT, Edward, De ontwikkeling van de moderne bank- en krediettechniek tijdens de zestiende en zeventiende eeuw in Europa en te Amsterdam in het bijzonder. Leuven, K.U.Leuven, 1975. 186 pp.

13. VERTONGHEN, Robert, Investeringscriteria voor Publieke Investeringen: Het Uitwerken van een Operationele Theorie met een Toepassing op de Verkeersinfrastructuur. Leuven, Acco, 1975. 254 pp.

14. Niet toegekend.

15. VANOVERBEKE, Lieven, Microeconomisch Onderzoek van de Sectoriële Arbeidsmobiliteit. Leuven, Acco, 1975. 205 pp.

16. DAEMS, Herman, The Holding Company: Essays on Financial Intermediation, Concentration and Capital Market Imperfections in the Belgian Economy. Leuven, K.U.Leuven, 1975. XII, 268 pp.

17. VAN ROMPUY, Eric, Groot-Brittannië en de Europese Monetaire Integratie: Een Onderzoek naar de Gevolgen van de Britse Toetreding op de Geplande Europese Monetaire Unie. Leuven, Acco, 1975. XIII, 222 pp.

18. MOESEN, Wim, Het Beheer van de Staatsschuld en de Termijnstructuur van de Intrestvoeten met een Toepassing voor België. Leuven, Vander, 1975. XVI, 250 pp.

19. LAMBRECHT, Marc, Capacity Constrained Multi-facility Dynamic Lot-size Problem. Leuven, KUL, 1976. 165 pp.

20. RAYMAECKERS, Erik, De Mens in de Onderneming en de Theorie van het Producentengedrag: Een Bijdrage tot Transdisciplinaire Analyse. Leuven, Acco, 1976. XIII, 538 pp.

21. TEJANO, Albert, Econometric and Input-output Models in Development Planning: The Case of the Philippines. Leuven, KUL, 1976. XX, 297 pp.

22. MARTENS, Bernard, Prijsbeleid en Inflatie met een Toepassing op België. Leuven, KUL, 1977. IV, 253 pp.

23. VERHEIRSTRAETEN, Albert, Geld, Krediet en Intrest in de Belgische Financiële Sector. Leuven, Acco, 1977. XXII, 377 pp.

24. GHEYSSENS, Lieven, International Diversification through the Government Bond Market: A Risk-return Analysis. Leuven, s.n., 1977. 188 pp.

25. LEFEBVRE, Chris, Boekhoudkundige Verwerking en Financiële Verslaggeving van Huurkooptransacties en Verkopen op Afbetaling bij Ondernemingen die aan Consumenten Verkopen. Leuven, KUL, 1977. 228 pp.

26. KESENNE, Stefan, Tijdsallocatie en Vrijetijdsbesteding: Een Econometrisch Onderzoek. Leuven, s.n., 1978. 163 pp.

27. VAN HERCK, Gustaaf, Aspecten van Optimaal Bedrijfsbeleid volgens het Marktwaardecriterium: Een Risico-rendementsanalyse. Leuven, KUL, 1978. IV, 163 pp.

28. VAN POECK, Andre, World Price Trends and Price and Wage Development in Belgium: An Investigation into the Relevance of the Scandinavian Model of Inflation for Belgium. Leuven, s.n., 1979. XIV, 260 pp.

29. VOS, Herman, De industriële Technologieverwerving in Brazilië: Een Analyse. Leuven, s.n., 1978. onregelmatig gepagineerd.

30. DOMBRECHT, Michel, Financial Markets, Employment and Prices in Open Economies. Leuven, KUL, 1979. 182 pp.

31. DE PRIL, Nelson, Bijdrage tot de Actuariële Studie van het Bonus-malussysteem. Brussel, OAB, 1979. 112 pp.

32. CARRIN, Guy, Economic Aspects of Social Security: A Public Economics Approach. Leuven, KUL, 1979. onregelmatig gepagineerd.

33. REGIDOR, Baldomero, An Empirical Investigation of the Distribution of Stock-market Prices and Weak-form Efficiency of the Brussels Stock Exchange. Leuven, KUL, 1979. 214 pp.

34. DE GROOT, Roger, Ongelijkheden voor Stop Loss Premies gebaseerd op E.T. Systemen in het kader van de Veralgemeende Convexe Analyse. Leuven, KUL, 1979. 155 pp.

35. CEYSSENS, Martin, On the Peak Load Problem in the Presence of Rationizing by Waiting. Leuven, KUL, 1979. IX, 217 pp.

36. ABDUL RAZK ABDUL, Mixed Enterprise in Malaysia: the Case Study of Joint Venture between Malysian Public Corporations and Foreign Enterprises. Leuven, KUL, 1979. 324 pp.

37. DE BRUYNE, Guido, Coordination of Economic Policy: A Game-theoretic Approach. Leuven, KUL, 1980. 106 pp.

38. KELLES, Gerard, Demand, Supply, Price Change and Trading Volume on Financial Markets of the Matching-order Type. Leuven, KUL, 1980. 222 pp.

39. VAN EECKHOUDT, Marc, De Invloed van de Looptijd, de Coupon en de Verwachte Inflatie op het Opbrengstverloop van Vastrentende Finaciële Activa. Leuven, KUL, 1980. 294 pp.

40. SERCU, Piet, Mean-variance Asset Pricing with Deviations from Purchasing Power Parity. Leuven, s.n., 1981. XIV, 273 pp.

41. DEQUAE, Marie-Gemma, Inflatie, Belastingsysteem en Waarde van de Onderneming. Leuven, KUL, 1981. 436 pp.

42. BRENNAN, John, An Empirical Investigation of Belgian Price Regulation by Prior Notification: 1975 - 1979 - 1982. Leuven, KUL, 1982. XIII, 386 pp.

43. COLLA, Annie, Een Econometrische Analyse van Ziekenhuiszorgen. Leuven, KUL, 1982. 319 pp.

44. Niet toegekend.

45. SCHOKKAERT, Eric, Modelling Consumer Preference Formation. Leuven, KUL, 1982. VIII, 287 pp.

46. DEGADT, Jan, Specificatie van een Econometrisch Model voor Vervuilingsproblemen met Proeven van Toepassing op de Waterverontreiniging in België. Leuven, s.n., 1982. 2 V.

47. LANJONG, Mohammad Nasir, A Study of Market Efficiency and Risk-return Relationships in the Malaysian Capital Market. s.l., s.n., 1983. XVI, 287 pp.

48. PROOST, Stef, De Allocatie van Lokale Publieke Goederen in een Economie met een Centrale Overheid en Lokale Overheden. Leuven, s.n., 1983. onregelmatig gepagineerd.

49. VAN HULLE, Cynthia (08/83), Shareholders' Unanimity and Optimal Corporate Decision Making in Imperfect Capital Markets. s.l., s.n., 1983. 147 pp. + appendix.

50. VAN WOUWE, Martine (2/12/83), Ordening van Risico's met Toepassing op de Berekening van Ultieme Ruïnekansen. Leuven, s.n., 1983. 109 pp.

51. D'ALCANTARA, Gonzague (15/12/83), SERENA: A Macroeconomic Sectoral Regional and National Account Econometric Model for the Belgian Economy. Leuven, KUL, 1983. 595 pp.

52. D'HAVE, Piet (24/02/84), De Vraag naar Geld in België. Leuven, KUL, 1984. XI, 318 pp.

53. MAES, Ivo (16/03/84), The Contribution of J.R. Hicks to Macro-economic and Monetary Theory. Leuven, KUL, 1984. V, 224 pp.

54. SUBIANTO, Bambang (13/09/84), A Study of the Effects of Specific Taxes and Subsidies on a Firms' R&D Investment Plan. s.l., s.n., 1984. V, 284 pp.

55. SLEUWAEGEN, Leo (26/10/84), Location and Investment Decisions by Multinational Enterprises in Belgium and Europe. Leuven, KUL, 1984. XII, 247 pp.

56. GEYSKENS, Erik (27/03/85), Produktietheorie en Dualiteit. Leuven, s.n., 1985. VII, 392 pp.

57. COLE, Frank (26/06/85), Some Algorithms for Geometric Programming. Leuven, KUL, 1985. 166 pp.

58. STANDAERT, Stan (26/09/86), A Study in the Economics of Repressed Consumption. Leuven, KUL, 1986. X, 380 pp.

59. DELBEKE, Jos (03/11/86), Trendperioden in de Geldhoeveelheid van België 1877-1983: Een Theoretische en Empirische Analyse van de "Banking school" Hypothese. Leuven, KUL, 1986. XII, 430 pp.

60. VANTHIENEN, Jan (08/12/86), Automatiseringsaspecten van de Specificatie, Constructie en Manipulatie van Beslissingstabellen. Leuven, s.n., 1986. XIV, 378 pp.

61. LUYTEN, Robert (30/04/87), A Systems-based Approach for Multi-echelon Production/Inventory Systems. s.l., s.n., 1987. 3V.

62. MERCKEN, Roger (27/04/87), De Invloed van de Data Base Benadering op de Interne Controle. Leuven, s.n., 1987. XIII, 346 pp.

63. VAN CAYSEELE, Patrick (20/05/87), Regulation and International Innovative Activities in the Pharmaceutical Industry. s.l., s.n., 1987. XI, 169 pp.

64. FRANCOIS, Pierre (21/09/87), De Empirische Relevantie van de Independence from Irrelevant Alternatives. Assumptie Indiscrete Keuzemodellen. Leuven, s.n., 1987. IX, 379 pp.

65. DECOSTER, André (23/09/88), Family Size, Welfare and Public Policy. Leuven, KUL. Faculteit Economische en toegepaste economische wetenschappen, 1988. XIII, 444 pp.

66. HEIJNEN, Bart (09/09/88), Risicowijziging onder Invloed van Vrijstellingen en Herverzekeringen: Een Theoretische Analyse van Optimaliteit en Premiebepaling. Leuven, KUL. Faculteit Economische en toegepaste economische wetenschappen, 1988. onregelmatig gepagineerd.

67. GEEROMS, Hans (14/10/88), Belastingvermijding. Theoretische Analyse van de Determinanten van de Belastingontduiking en de Belastingontwijking met Empirische Verificaties. Leuven, s.n., 1988. XIII, 409, 5 pp.

68. PUT, Ferdi (19/12/88), Introducing Dynamic and Temporal Aspects in a Conceptual (Database) Schema. Leuven, KUL. Faculteit Economische en toegepaste economische wetenschappen, 1988. XVIII, 415 pp.

69. VAN ROMPUY, Guido (13/01/89), A Supply-side Approach to Tax Reform Programs. Theory and Empirical Evidence for Belgium. Leuven, KUL. Faculteit Economische en toegepaste economische wetenschappen, 1989. XVI, 189, 6 pp.

70. PEETERS, Ludo (19/06/89), Een Ruimtelijk Evenwichtsmodel van de Graanmarkten in de E.G.: Empirische Specificatie en Beleidstoepassingen. Leuven, K.U.Leuven. Faculteit Economische en toegepaste economische wetenschappen, 1989. XVI, 412 pp.

71. PACOLET, Jozef (10/11/89), Marktstructuur en Operationele Efficiëntie in de Belgische Financiële Sector. Leuven, K.U.Leuven. Faculteit Economische en toegepaste economische wetenschappen, 1989. XXII, 547 pp.

72. VANDEBROEK, Martina (13/12/89), Optimalisatie van Verzekeringscontracten en Premieberekeningsprincipes. Leuven, K.U.Leuven. Faculteit Economische en toegepaste economische wetenschappen, 1989. 95 pp.

73. WILLEKENS, Francois (), Determinance of Government Growth in Industrialized Countries with Applications to Belgium. Leuven, K.U.Leuven. Faculteit Economische en toegepaste economische wetenschappen, 1990. VI, 332 pp.

74. VEUGELERS, Reinhilde (02/04/90), Scope Decisions of Multinational Enterprises. Leuven, K.U.Leuven. Faculteit Economische en toegepaste economische wetenschappen, 1990. V, 221 pp.

75. KESTELOOT, Katrien (18/06/90), Essays on Performance Diagnosis and Tacit Cooperation in International Oligopolies. Leuven, K.U.Leuven. Faculteit Economische en toegepaste economische wetenschappen, 1990. 227 pp.

76. WU, Changqi (23/10/90), Strategic Aspects of Oligopolistic Vertical Integration. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1990. VIII, 222 pp.

77. ZHANG, Zhaoyong (08/07/91), A Disequilibrium Model of China's Foreign Trade Behaviour. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1991. XII, 256 pp.

78. DHAENE, Jan (25/11/91), Verdelingsfuncties, Benaderingen en Foutengrenzen van Stochastische Grootheden Geassocieerd aan Verzekeringspolissen en -portefeuilles. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1991. 146 pp.

79. BAUWELINCKX, Thierry (07/01/92), Hierarchical Credibility Techniques. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1992. 130 pp.

80. DEMEULEMEESTER, Erik (23/3/92), Optimal Algorithms for Various Classes of Multiple Resource-Constrained Project Scheduling Problems. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1992. 180 pp.

81. STEENACKERS, Anna (1/10/92), Risk Analysis with the Classical Actuarial Risk Model: Theoretical Extensions and Applications to Reinsurance. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1992. 139 pp.

82. COCKX, Bart (24/09/92), The Minimum Income Guarantee. Some Views from a Dynamic Perspective. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1992. XVII, 401 pp.

83. MEYERMANS, Eric (06/11/92), Econometric Allocation Systems for the Foreign Exchange Market: Specification, Estimation and Testing of Transmission Mechanisms under Currency Substitution. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1992. XVIII, 343 pp.

84. CHEN, Guoqing (04/12/92), Design of Fuzzy Relational Databases Based on Fuzzy Functional Dependency. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1992. 176 pp.

85. CLAEYS, Christel (18/02/93), Vertical and Horizontal Category Structures in Consumer Decision Making: The nature of product hierarchies and the effect of brand typicality. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1993. 348 pp.

86. CHEN, Shaoxiang (25/03/93), The Optimal Monitoring Policies for Some Stochastic and Dynamic Production Processes. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1993. 170 pp.

87. OVERWEG, Dirk (23/04/93), Approximate Parametric Analysis and Study of Cost Capacity Management of Computer Configurations. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1993. 270 pp.

88. DEWACHTER, Hans (22/06/93), Nonlinearities in Speculative Prices: The Existence and Persistence of Nonlinearity in Foreign Exchange Rates. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1993. 151 pp.

89. LIN, Liangqi (05/07/93), Economic Determinants of Voluntary Accounting Choices for R & D Expenditures in Belgium. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1993. 192 pp.

90. DHAENE, Geert (09/07/93), Encompassing: Formulation, Properties and Testing. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1993. 117 pp.

91. LAGAE, Wim (20/09/93), Marktconforme Verlichting van Soevereine Buitenlandse Schuld door Private Crediteuren: Een Neo-institutionele Analyse. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1993. 241 pp.

92. VAN DE GAER, Dirk (27/09/93), Equality of Opportunity and Investment in Human Capital. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1993. 172 pp.

93. SCHROYEN, Alfred (28/02/94), Essays on Redistributive Taxation when Monitoring is Costly. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1994. 203 pp. + V.

94. STEURS, Geert (15/07/94), Spillovers and Cooperation in Research and Development. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1994. 266 pp.

95. BARAS, Johan (15/09/94), The Small Sample Distribution of the Wald, Lagrange Multiplier and Likelihood Ratio Tests for Homogeneity and Symmetry in Demand Analysis: A Monte Carlo Study. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1994. 169 pp.

96. GAEREMYNCK, Ann (08/09/94), The Use of Depreciation in Accounting as a Signalling Device. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1994. 232 pp.

97. BETTENDORF, Leon (22/09/94), A Dynamic Applied General Equilibrium Model for a Small Open Economy. Leuven, K.U.Leuven, Faculteit Economische en toegepas-te economische wetenschappen, 1994. 149 pp.

98. TEUNEN, Marleen (10/11/94), Evaluation of Interest Randomness in Actuarial Quantities. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1994. 214 pp.

99. VAN OOTEGEM, Luc (17/01/95), An Economic Theory of Private Donations. Leuven. K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 236 pp.

100. DE SCHEPPER, Ann (20/03/95), Stochastic Interest Rates and the Probabilistic Behaviour of Actuarial Functions. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 211 pp.

101. LAUWERS, Luc (13/06/95), Social Choice with Infinite Populations. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 79 pp.

102. WU, Guang (27/06/95), A Systematic Approach to Object-oriented Business Modeling. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 248 pp.

103. WU, Xueping (21/08/95), Term Structures in the Belgian Market: Model Estimation and Pricing Error Analysis. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 133 pp.

104. PEPERMANS, Guido (30/08/95), Four Essays on Retirement from the Labor Force. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 128 pp.

105. ALGOED, Koen (11/09/95), Essays on Insurance: A View from a Dynamic Perspective. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 136 pp.

106. DEGRYSE, Hans (10/10/95), Essays on Financial Intermediation, Product Differentiation, and Market Structure. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 218 pp.

107. MEIR, Jos (05/12/95), Het Strategisch Groepsconcept Toegepast op de Belgische Financiële Sector. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1995. 257 pp.

108. WIJAYA, Miryam Lilian (08/01/96), Voluntary Reciprocity as an Informal Social Insurance Mechanism: a Game Theoretic Approach. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1996. 124 pp.

109. VANDAELE, Nico (12/02/96), The Impact of Lot Sizing on Queueing Delays: Multi Product, Multi Machine Models. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1996. 243 pp.

110. GIELENS, Geert (27/02/96), Some Essays on Discrete Time Target Zones and their Tails. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1996. 131 pp.

111. GUILLAUME, Dominique (20/03/96), Chaos, Randomness and Order in the Foreign Exchange Markets. Essays on the modelling of the markets. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1996. 171 pp.

112. DEWIT, Gerda (03/06/96), Essays on Export Insurance Subsidization. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1996. 186 pp.

113. VAN DEN ACKER, Carine (08/07/96), Belief-function Theory and its Application to the Modeling of Uncertainty in Financial Statement Auditing. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1996. 147 pp.

114. IMAM, Mahmood Osman (31/07/96), Choice of IPO Flotation Methods in Belgium in an Asymmetric Information Framework and Pricing of IPO's in the Long-Run. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1996. 221 pp.

115. NICAISE, Ides (06/09/96), Poverty and Human Capital. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1996. 209 pp.

116. EYCKMANS, Johan (18/09/97), On the Incentives of Nations to Join International Environmental Agreements. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1997. XV + 348 pp.

117. CRISOLOGO-MENDOZA, Lorelei (16/10/97), Essays on Decision Making in Rural Households: a study of three villages in the Cordillera Region of the Philippines. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1997. 256 pp.

118. DE REYCK, Bert (26/01/98), Scheduling Projects with Generalized Precedence Relations: Exact and Heuristic Procedures. Leuven, K.U.Leuven, Faculteit Econo-mische en toegepaste economische wetenschappen, 1998. XXIV + 337 pp.

119. VANDEMAELE Sigrid (30/04/98), Determinants of Issue Procedure Choice within the Context of the French IPO Market: Analysis within an Asymmetric Information Framework. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1998. 241 pp.

120. VERGAUWEN Filip (30/04/98), Firm Efficiency and Compensation Schemes for the Management of Innovative Activities and Knowledge Transfers. Leuven, K.U.Leuven,

Faculteit Economische en toegepaste economische wetenschappen, 1998. VIII + 175 pp.

121. LEEMANS Herlinde (29/05/98), The Two-Class Two-Server Queueing Model with Nonpreemptive Heterogeneous Priority Structures. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1998. 211 pp.

122. GEYSKENS Inge (4/09/98), Trust, Satisfaction, and Equity in Marketing Channel Relationships. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1998. 202 pp.

123. SWEENEY John (19/10/98), Why Hold a Job ? The Labour Market Choice of the Low-Skilled. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1998. 278 pp.

124. GOEDHUYS Micheline (17/03/99), Industrial Organisation in Developing Countries, Evidence from Côte d'Ivoire. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1999. 251 pp.

125. POELS Geert (16/04/99), On the Formal Aspects of the Measurement of Object-Oriented Software Specifications. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1999. 507 pp.

126. MAYERES Inge (25/05/99), The Control of Transport Externalities: A General Equilibrium Analysis. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1999. XIV + 294 pp.

127. LEMAHIEU Wilfried (5/07/99), Improved Navigation and Maintenance through an Object-Oriented Approach to Hypermedia Modelling. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1999. 284 pp.

128. VAN PUYENBROECK Tom (8/07/99), Informational Aspects of Fiscal Federalism. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1999. 192 pp.

129. VAN DEN POEL Dirk (5/08/99), Response Modeling for Database Marketing Using Binary Classification. Leuven, K.U.Leuven, Faculteit Economische en toegepas-te economische wetenschappen, 1999. 342 pp.

130. GIELENS Katrijn (27/08/99), International Entry Decisions in the Retailing Industry: Antecedents and Performance Consequences. Leuven, K.U.Leuven, Faculteit Economische en toegepas-te economische wetenschappen, 1999. 336 pp.

131. PEETERS Anneleen (16/12/99), Labour Turnover Costs, Employment and Temporary Work. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1999. 207 pp.

132. VANHOENACKER Jurgen (17/12/99), Formalizing a Knowledge Management Architecture Meta-Model for Integrated Business Process Management. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 1999. 252 pp.

133. NUNES Paulo (xx/03/2000), Contingent Valuation of the Benefits of Natural Areas and its Warglow Component. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2000. XXI + 282 pp.

134. VAN DEN CRUYCE Bart (7/04/2000), Statistische Discriminatie van Allochtonen op Jobmarkten met Rigide Lonen. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2000. XXIII + 441 pp.

135. REPKINE Alexandre (15/03/2000), Industrial Restructuring in Countries of Central and Eastern Europe: Combining Branch-, Firm- and Product-level Data for a Better Understanding of Enterprises' Behaviour during Transition towards Market Economy. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2000. VI + 147 pp.

136. AKSOY, Yunus (21/06/2000), Essays on International Price Rigidities and Exchange Rates. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2000. IX + 236 pp.

137. RIYANTO, Yohanes Eko (22/06/2000), Essays on the Internal and External Delegation of Authority in Firms. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2000. VIII + 280 pp.

138. HUYGHEBAERT, Nancy (20/12/2000), The Capital Structure of Business Start-ups. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2000. VIII + 332 pp.

139. FRANCKX Laurent (22/01/2001), Ambient Inspections and Commitment in Environmental Enforcement. Leuven, K.U.Leuven, Faculteit Economische en toege-paste economische wetenschappen, 2001 VIII + 286 pp.

140. VANDILLE Guy (16/02/2001), Essays on the Impact of Income Redistribution on Trade. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 VIII + 176 pp.

141. MARQUERING Wessel (27/04/2001), Modeling and Forecasting Stock Market Returns and Volatility. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 V + 267 pp.

142. FAGGIO Giulia (07/05/2001), Labor Market Adjustment and Enterprise Behavior in Transition. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 150 pp.

143. GOOS Peter (30/05/2001), The Optimal Design of Blocked and Split-plot Experiments. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 X + 224 pp.

144. LABRO Eva (01/06/2001), Total Cost of Ownership Supplier Selection based on Activity Based Costing and Mathematical Programming. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 217 pp.

145. VANHOUCKE Mario (07/06/2001), Exact Algorithms for Various Types of Project Scheduling Problems. Nonregular Objectives and time/cost Trade-offs. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 316 pp.

146. BILSEN Valentijn (28/08/2001), Entrepreneurship and Private Sector Development in Central European Transition Countries. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 XVI + 188 pp.

147. NIJS Vincent (10/08/2001), Essays on the dynamic Category-level Impact of Price promotions. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001

148. CHERCHYE Laurens (24/09/2001), Topics in Non-parametric Production and Efficiency Analysis. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 VII + 169 pp.

149. VAN DENDER Kurt (15/10/2001), Aspects of Congestion Pricing for Urban Transport. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001 VII + 203 pp.

150. CAPEAU Bart (26/10/2001), In Defence of the Excess Demand Approach to Poor Peasants' Economic Behaviour. Theory and Empirics of non-recursive agricultural household modelling. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001, XIII + 286 blz.

151. CALTHROP Edward (09/11/2001), Essays in Urban Transport Economics. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2001.

152. VANDER BAUWHEDE Heidi (03/12/2001), Earnings management in an Non-Anglo-Saxon environment. Leuven, K.U.Leuven, Faculteit Economische en toege-paste economische wetenschappen, 2001, 408 pp.

153. DE BACKER Koenraad (22/01/2002), Multinational Firms and Industry Dynamics in Host Countries : the case of Belgium. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002, VII + 165 pp.

154. BOUWEN Jan (08/02/2002), Transactive Memory in Operational Workgroups. Concept Elaboration and Case Study. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002, 319 pp. + appendix 102 pp.

155. VAN DEN BRANDE Inge (13/03/2002), The psychological contract between employer and employee : a survey among Flemish employees. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002, VIII + 470 pp.

156. VEESTRAETEN Dirk (19/04/2002), Asset Price Dynamics under Announced Policy Switching. Leuven, K.U.Leuven, Faculteit Economische en toegepas-te economische wetenschappen, 2002, 176 pp.

157. PEETERS Marc (16/05/2002), One Dimensional Cutting and Packing: New Problems and Algorithms. Leuven, K.U.Leuven, Faculteit Economische en toege-paste economische wetenschappen, 2002

158. SKUDELNY Frauke (21/05/2002), Essays on The Economic Consequences of the European Monetary Union. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002

159. DE WEERDT Joachim (07/06/2002), Social Networks, Transfers and Insurance in Developing countries. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002, VI + 129 pp.

160. TACK Lieven (25/06/2002), Optimal Run Orders in Design of Experiments. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002, XXXI + 344 pp.

161. POELMANS Stephan (10/07/2002), Making Workflow Systems work. An investigation into the Importance of Task-appropriation fit, End-user Support and other Technological Characteristics. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002, 237 pp.

162. JANS Raf (26/09/2002), Capacitated Lot Sizing Problems : New Applications, Formulations and Algorithms. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002.

163. VIAENE Stijn (25/10/2002), Learning to Detect Fraud from enriched Insurance Claims Data (Context, Theory and Applications). Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002, 315 pp.

164. AYALEW Tekabe (08/11/2002), Inequality and Capital Investment in a Subsistence Economy. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002. V + 148 pp.

165. MUES Christophe (12/11/2002), On the Use of Decision Tables and Diagrams in Knowledge Modeling and Verification. Leuven, K.U.Leuven, Faculteit Economi-sche en toegepaste economische wetenschappen, 2002. 222 pp.

166. BROCK Ellen (13/03/2003), The Impact of International Trade on European Labour Markets. K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002.

167. VERMEULEN Frederic (29/11/2002), Essays on the Collective Approach to Household Labour Supply. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2002. XIV + 203 pp.

168. CLUDTS Stephan (11/12/2002), Combining Participation in Decision-making with Financial Participation: Theoretical and Empirical Perspectives. Leuven, K.U.Leuven,

Faculteit Economische en toegepaste economische wetenschappen, 2002. XIV + 247 pp.

169. WARZYNSKI Frederic (09/12/2003), The Dynamic Effect of Competition on Price Cost Margins and Innovation. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

170. VERWIMP Philip (14/01/2003), Development and Genocide in Rwanda; a Political Economy Analysis of Peasants and Power under the Habyarimana Regime. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

171. BIGANO Andrea (25/02/2003). Environmental Regulation of the Electricity Sector in a European Market Framework. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003, xx + 310 pp.

172. MAES Konstantijn (24/03/2003). Modeling the Term Structure of Interest Rates Across Countries. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003, VII+246 pp.

173. VINAIMONT Tom (26/02/2003). The performance of One- versus Two-Factor Models of the Term Structure of Interest Rates. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

174. OOGHE Erwin (15/04/2003). Essays in Multi-dimensional Social Choice. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003, VIII+108 pp.

175. FORRIER Anneleen (25/04/2003). Temporary Employment, Employability and Training. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

176. CARDINAELS Eddy (28/04/2003). The Role of Cost System Accuracy in Managerial Decision Making. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003. 144 pp.

177. DE GOEIJ Peter (02/07/2003). Modeling Time-Varying Volatility and Interest Rates. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003 VII+225 pp.

178. LEUS Roel (19/09/2003). The Generation of Stable Project Plans - Complexity and Exact Algorithms. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

179. MARINHIERO Carlos (23/09/2003). EMU and Fiscal Stabilisation Policy: the Case of Small Countries. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

180. BAESENS Bart (24/09/2003). Developping Intelligent Systems for Credit Scoring Using Machien Learning Techniques. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

181. KÓCZY László Á. (18/09/2003). Solution Concepts and Outsider Behaviour in Coalition Formation Games. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

182. ALTOMONTE Carlo (25/09/2003). Essays on Foreign Direct Investment in Transition Countries: learning from the Evidence. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

183. DRIES Liesbeth (10/11/2003). Transition, Globalisation and Sectoral Restructuring: Theory and Evidence from the Polish Agro-Food Sector. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

184. DEVOOGHT Kurt (18/11/2003). Essays on Responsability-Sensitive Egalitarianism and the Measurement of Income Inequality. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

185. DELEERSNYDER Barbara (28/11/2003). Marketing in Turbulente Tijden. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

186. AYALEW Daniel Ali (19/12/2003). Essays on Household Consumption and Production Decisions under Uncertainty in Rural Ethiopia. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2003.

187. WILLEMS Bert (14/01/2004). Electricity Networks and Generation Market Power. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

188. JANSSEN Gust (30/01/2004). Advanced Modelling of Conditional Volatility and Correlation in Financial Markets. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

189. THOEN Vincent (19/01/2004). On the Valuation and Disclosure Practices Implemented by Venture Capital Providers. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

190. MARTENS Jurgen (16/02/2004). A Fuzzy Set and Stochastic System Theoretic Technique to Validate Simulation Models. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

191. ALTAVILLA Carlo (21/05/2004). Monetary Policy Implementation and Transmission Mechanisms in the Euro Area. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

192. DE BRUYNE Karolien (07/06/2004). Essays in the Location of Economic Activity. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

193. ADEM Jan (25/06/2004). Mathematical Programming Approaches for the Supervised Classification Problem. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

194. LEROUGE Davy (08/07/2004). Predicting Product Preferences: The Effect of Internal and External Cues. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

195. VANDENBROECK Katleen (16/07/2004). Essays on Output Growth, Social Learning and Land Allocation in Agriculture: Micro-evidence from Ethiopia and Tanzania. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

196. GRIMALDI Maria (03/09/2004). The Exchange Rate, Heterogeneity of Agents and Bounded Rationality. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

197. SMEDTS Kristien (26/10/2004). Financial Integration in EMU in the Framework of the No-Arbitrage Theory. Leuven, K.U.Leuven, Faculteit Economische en toegepaste economische wetenschappen, 2004.

198. KOEVOETS Wim (12/11/2004). Essays on Unions, Wages and Employment. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

199. CALLENS Marc (22/11/2004). Essays on multilevel logistic Regression. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

200. RUGGOO Arvind (13/12/2004). Two stage designs robust to model uncertainty. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

201. HOORELBEKE Dirk (28/01/2005). Bootstrap and Pivoting Techniques for Testing Multiple Hypotheses. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

202. ROUSSEAU Sandra (17/02/2005). Selecting Environmental Policy Instruments in the Presence of Incomplete Compiance. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

203. VAN DER MEULEN Sofie (17/02/2005).Quality of Financial Statements : Impact of the external auditor and applied accounting standards. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

204. DIMOVA Ralitza (21/02/2005). Winners and Losers during Structural Reform and Crisis : the Bulgarian Labour Market Perspective. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

205. DARKIEWICZ Grzegorz (28/02/2005). Value-at-risk in Insurance and Finance : the Comonotonicity Approach. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

206. DE MOOR Lieven (20/05/2005). The Structure of International Stock Returns : Size, Country and Sector Effects in Capital Asset Pricing. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

207. EVERAERT Greetje (27/06/2005). Soft Budget Constraints and Trade Policies : The Role of Institutional and External Constraints. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

208. SIMON Steven (06/07/2005). The Modeling and Valuation of complex Derivatives : The Impact of the Choice of the term structure model. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

209. MOONEN Linda (23/09/2005). Algorithms for some Graph-Theoretical Optimization Problems. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

210. COUCKE Kristien (21/09/2005). Firm and industry adjustment under de-industrialisation and globalization of the Belgian economy. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

211. DECAMPS Marc (21/10/2005). Some actuarial and financial applications of generalized diffusion processes. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

212. KIM Helena (29/11/2005). Escalation games: an instrument to analyze conflicts. The strategic approach to the bargaining problem. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

213. GERMENJI Etleva (06/01/2006). Essays on the Economics of Emigration from Albania. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

214. BELIEN Jeroen (18/01/2006). Exact and Heuristic Methodologies for Scheduling in Hospitals: Problems, Formulations and Algorithms. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

215. JOOSSENS Kristel (10/02/2006). Robust discriminant analysis. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

216. VRANKEN Liesbet (13/02/2006). Land markets and production efficiency in transition economies. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

217. VANSTEENKISTE Isabel (22/02/2006). Essays on non-linear modelling in international macroeconomics. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

218. WUYTS Gunther (31/03/2006). Essays on the liquidity of financial markets. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

219. DE BLANDER Rembert (28/04/2006). Essays on endogeneity and parameter heterogeneity in cross-section and panel data. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

220. DE LOECKER Jan (12/05/2006). Industry dynamics and productivity. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

221. LEMMENS Aurélie (12/05/2006). Advanced classification and time-series methods in marketing. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

222. VERPOORTEN Marijke (22/05/2006). Conflict and survival: an analysis of shocks, coping strategies and economic mobility in Rwanda, 1990-2002. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

223. BOSMANS Kristof (26/05/2006). Measuring economic inequality and inequality aversion. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

224. BRENKERS Randy (29/05/2006). Policy reform in a market with differentiated products: applications from the car market. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

225. BRUYNEEL Sabrina (02/06/2006). Self-control depletion: Mechanisms and its effects on consumer behavior. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

226. FAEMS Dries (09/06/2006). Collaboration for innovation: Processes of governance and learning. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

227. BRIERS Barbara (28/06/2006). Countering the scrooge in each of us: on the marketing of cooperative behavior. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

228. ZANONI Patrizia (04/07/2006). Beyond demography: Essays on diversity in organizations. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

229. VAN DEN ABBEELE Alexandra (11/09/2006). Management control of interfirm relations: the role of information. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

230. DEWAELHEYNS Nico (18/09/2006). Essays on internal capital markets, bankruptcy and bankruptcy reform. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

231. RINALDI Laura (19/09/2006). Essays on card payments and household debt. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

232. DUTORDOIR Marie (22/09/2006). Determinants and stock price effects of Western European convertible debt offerings: an empirical analysis. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

233. LYKOGIANNI Elissavet (20/09/2006). Essays on strategic decisions of multinational enterprises: R&D decentralization, technology transfers and modes of foreign entry. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

234. ZOU Jianglei (03/10/2006). Inter-firm ties, plant networks, and multinational firms: essays on FDI and trade by Japanse firms. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

235. GEYSKENS Kelly (12/10/2006). The ironic effects of food temptations on self-control performance. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

236. BRUYNSEELS Liesbeth (17/10/2006). Client strategic actions, going-concern audit opinions and audit reporting errors. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

237. KESSELS Roselinde (23/10/2006). Optimal designs for the measurement of consumer preferences. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

238. HUTCHINSON John (25/10/2006). The size distribution and growth of firms in transition countries. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

239. RENDERS Annelies (26/10/2006). Corporate governance in Europe: The relation with accounting standards choice, performance and benefits of control. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

240. DE WINNE Sophie (30/10/2006). Exploring terra incognita: human resource management and firm performance in small and medium-sized businesses. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

241. KADITI Eleni (10/11/2006). Foreign direct investments in transition economies. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

242. ANDRIES Petra (17/11/2006). Technology-based ventures in emerging industries: the quest for a viable business model. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

243. BOUTE Robert (04/12/2006). The impact of replenishment rules with endogenous lead times on supply chain performance. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

244. MAES Johan (20/12/2006). Corporate entrepreneurship: an integrative analysis of a resource-based model. Evidence from Flemish enterprises. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

245. GOOSSENS Dries (20/12/2006). Exact methods for combinatorial auctions. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

246. GOETHALS Frank (22/12/2006). Classifying and assessing extended enterprise integration approaches. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

247. VAN DE VONDER Stijn (22/12/2006). Proactive-reactive procedures for robust project scheduling. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

248. SAVEYN Bert (27/02/2007). Environmental policy in a federal state. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

249. CLEEREN Kathleen (13/03/2007). Essays on competitive structure and product-harm crises. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

250. THUYSBAERT Bram (27/04/2007). Econometric essays on the measurement of poverty. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

251. DE BACKER Manu (07/05/2007). The use of Petri net theory for business process verification. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

252. MILLET Kobe (15/05/2007). Prenatal testosterone, personality, and economic behavior. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

253. HUYSMANS Johan (13/06/2007). Comprehensible predictive models: New methods and insights. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

254. FRANCKEN Nathalie (26/06/2007). Mass Media, Government Policies and Economic Development: Evidence from Madagascar. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

255. SCHOUBBEN Frederiek (02/07/2007). The impact of a stock listing on the determinants of firm performance and investment policy. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

256. DELHAYE Eef (04/07/2007). Economic analysis of traffic safety. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

257. VAN ACHTER Mark (06/07/2007). Essays on the market microstructure of financial markets. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

258. GOUKENS Caroline (20/08/2007). Desire for variety: understanding consumers' preferences for variety seeking. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

259. KELCHTERMANS Stijn (12/09/2007). In pursuit of excellence: essays on the organization of higher education and research. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

260. HUSSINGER Katrin (14/09/2007). Essays on internationalization, innovation and firm performance. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

261. CUMPS Bjorn (04/10/2007). Business-ICT alignment and determinants. Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2007.

262. LYRIO Marco (02/11/2007). Modeling the yield curve with macro factors. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2007.

263. VANPEE Rosanne (16/11/2007). Home bias and the implicit costs of investing abroad. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2007.

264. LAMBRECHTS Olivier (27/11/2007). Robust project scheduling subject to resource breakdowns. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2007.

265. DE ROCK Bram (03/12/2007). Collective choice behaviour: non parametric characterization. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2007.

266. MARTENS David (08/01/2008). Building acceptable classification models for financial engineering applications. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

267. VAN KERCKHOVEN Johan (17/01/2008). Predictive modelling: variable selection and classification efficiencies. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

268. CIAIAN Pavel (12/02/2008). Land, EU accession and market imperfections. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

269. TRUYTS Tom (27/02/2008). Diamonds are a girl's best friend: five essays on the economics of social status. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

270. LEWIS Vivien (17/03/2008). Applications in dynamic general equilibrium macroeconomics. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

271. CAPPELLEN Tineke (04/04/2008). Worldwide coordination in a transnational environment: an inquiry into the work and careers of global managers. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

272. RODRIGUEZ Victor (18/04/2008). Material transfer agreements: research agenda choice, co-publication activity and visibility in biotechnology. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

273. QUAN Qi (14/04/2008). Privatization in China: Examining the endogeneity of the process and its implications for the performance of newly privatized firms. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

274. DELMOTTE Jeroen (30/04/2008). Evaluating the HR function: Empirical studies on HRM architecture and HRM system strength. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

275. ORSINI Kristian (05/05/2008). Making work pay: Insights from microsimulation and random utility models. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

276. HOUSSA Romain (13/05/2008). Macroeconomic fluctuations in developing countries. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

277. SCHAUMANS Catherine (20/05/2008). Entry, regulation and economic efficiency: essays on health professionals. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

278. CRABBE Karen (21/05/2008). Essays on corporate tax competition in Europe. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008

279. GELPER Sarah (30/05/2008). Economic time series analysis: Granger causality and robustness. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008

280. VAN HOVE Jan (20/06/2008). The impact of technological innovation and spillovers on the pattern and direction of international trade. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

281. DE VILLE DE GOYET Cédric (04/07/2008). Hedging with futures in agricultural commodity markets. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

282. FRANCK Tom (15/07/2008). Capital structure and product market interactions: evidence from business start-ups and private firms. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

283. ILBAS Pelin (15/09/2008). Optimal monetary policy design in dynamic macroeconomics. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

284. GOEDERTIER Stijn (16/09/2008). Declarative techniques for modeling and mining business processes. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

285. LAMEY Lien (22/09/2008). The private-label nightmare: can national brands ever wake up?. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

286. VANDEKERCKHOVE Jan (23/09/2008). Essays on research and development with spillovers. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

287. PERNOT Eli (25/09/2008). Management control system design for supplier relationships in manufacturing: Case study evidence from the automotive industry. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

288. AERTS Kris (16/10/2008). Essays on the economics of evaluation: public policy and corporate strategies in innovation. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

289. BOUDT Kris (27/11/2008). Estimation of financial risk under non-normal distributions. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

290. MARKIEWICZ Agnieszka (01.12.2008). Essays in exchange rate economics. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

291. GLADY Nicolas (08.12.2008). Customer profitability modeling. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

292. MIERZEJEWSKI Fernando (11.12.2008). Essays on liquidity-preference in markets with borrowing restrictions. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

293. VAN BEVEREN Ilke (15.12.2008). Globalization and firm dynamics. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

294. VERBRUGGEN Marijke (16.12.2008). The role of career counseling in the new career era. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

295. CORNIL Anneleen (15.12.2008). Essays on financial statement quality: the influence and the effects of accounting restatements. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

296. SOENEN Helke (16.12.2008). Between vulnerability and creativity: an anthropological economic analysis of undocumented migrants in Brussels. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.

297. VERGOTE Olivier (17.12.2008). Financial transaction data and volatility. Leuven, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen, 2008.