Nonparametric estimation of stochastic volatility models in the presence of market microstructure noise

Ferdinand R. Houwink
Student number: 260640

Rotterdam, The Netherlands
August 18, 2010

MASTER'S THESIS IN QUANTITATIVE FINANCE
ERASMUS SCHOOL OF ECONOMICS
ERASMUS UNIVERSITY ROTTERDAM

SUPERVISOR:

Prof. dr. D.J.C. van Dijk

Acknowledgements

My heartfelt gratitude goes out to all that have stood by me during what has become a long and interesting path towards completing this thesis.

Above all, I am deeply grateful for the unwavering emotional (and financial) support of my parents, who never stopped believing in me. Without them by my side none of this would have been possible. Together with my girlfriend, they have been an invaluable source of motivation and inspiration.

I would like to take this opportunity to give special thanks to my supervisor, prof. Van Dijk, for not losing his patience with me. His positive support and constructive feedback have been pivotal during the research, writing, and completion of this study.

Further thanks go out to my colleagues and friends for putting up with me over the last year. Among those, special thanks to my managers Hielke Hielkema and Sebastiaan van den Dries for their backing and understanding throughout the process.

Any errors or inadequacies that may remain in this thesis are entirely my own.

Abstract

Recently, kernel-based infinitesimal moment estimators have been proposed to nonparametrically estimate the drift and diffusion coefficients of stochastic volatility models by exploiting the availability of high frequency returns. The methodology uses nonparametric spot variance estimators based high frequency data, which are typically exposed to the effects of market microstructure noise, to construct low frequency estimates of the latent variance series driving the returns. With the low frequency spot variance estimates, the kernel-based estimates of the variance dynamics are constructed. Given the exposure of the spot variance estimates to market microstructure noise, this thesis studies its effects on the estimation of the variance dynamics of stochastic volatility models. By means of a simulation study which generates high frequency data from a GARCH and a Heston model, the effects of noise, in the form of bid-ask bounce and infrequent trading, on the kernel-based estimators are studied by using several bias-corrected and unadjusted estimators of integrated variance. Evidence is provided that controlling for the noise by bias-correction procedures on the estimators of spot variance is sufficient to obtain consistent estimates of the drift and diffusion coefficients of the underlying variance process.

Contents

1 Introduction

Modeling volatility has its roots in financial econometrics and mathematical finance. The importance of the research in this field is due to the paramount need for a better understanding of what drives the plethora of financial applications in the market. Originally asset price models considered the volatility of the asset price to be constant over time. Despite that homogeneous volatility was an over-simplification of reality, doing so enabled the development of many effective methods e.g. the Black-Scholes option pricing model (Hull (2008)), that now serve as the basis for much further research.

Nevertheless, characteristics of financial time series, such as volatility clustering and fat tails in the return distribution, are inadequately explained by the available univariate models. Further empirical evidence, such as that options with lower strike prices have higher implied volatility than options with higher strike prices, supports the notion that volatility is time-varying. These are mere examples of the complex behavior of financial instruments that is sought to be explained, giving rise to the research done on time-varying volatility when modeling its dynamics. The evolution and development of this research and the rise of (stochastic) volatility modeling in general is well-documented by Shephard and Andersen (2008), Andersen et al. (2002), and McAleer (2005) amongst others.

What is clear in the advancement of the research on volatility modeling is the division in approaches. On the one hand there are the parametric methodologies that include the popular autoregressive conditional heteroscedasticity (ARCH) models and the discrete- and continuous-time stochastic volatility models. On the other hand are the nonparametric methodologies that include such measures as the realized volatility, which is based on high-frequency data (McAleer and Medeiros (2008)).

The parametric methods require a distinction to be made between the ARCH type models and the stochastic volatility models. Specifically the discrete-time ARCH models fully define the one-step ahead conditional volatility of the returns series on the previously observed returns. As Shephard and Andersen (2008) state, this approach directly defines the likelihood function as the product of one-step ahead predictive densities, allowing for a convenient estimation procedure. Stochastic volatility models move away from this and model the volatility dynamics as an unobservable state variable, thereby implicitly modeling the predictive distribution, which as a result can usually only be evaluated by

means of numerical methods. This method inherently runs the risk of misspecification, therefore requiring great care and diligent evaluation when defining the model's dynamics.

The nonparametric methodologies have no need for restrictive assumptions on the functional form, yet allow for consistent estimation of the volatility. However these methods tend to lack the informative value with respect to the dynamics driving the volatility series. Nevertheless, estimators like the realized volatility, which can accurately estimate the volatility over infinitesimal time steps, have intuitive potential in the estimation of continuous-time stochastic volatility models.

A promising step in this regard is provided by Renò (2006) by using nonparametric spot volatility estimates for the estimation of stochastic volatility models. He provides simulation-based evidence on the performance of nonparametric estimators of the drift and diffusion functions for a general class of stochastic volatility (SV) models. The estimation procedures for the drift term in the (univariate) asset price process has been well-documented already, allowing Renò (2006) and onward to focus on the estimation of the dynamics of the variance process of bivariate models. The proposed methodology relies on using the realized variance estimator to create visibility of the underlying variance process, by exploiting high frequency data to obtain consistent spot variance estimates. The next step employs kernel-based estimators, akin to the Nadaraya-Watson kernel estimators developed for the univariate setting (Renò (2008, 2006)), to identify the dynamics of the variance process without imposing strict assumptions on its functional form.

As Martens and Van Dijk (2007) point out, in an ideal world, the realized variance provides unbiased and efficient estimates of the spot variance, converging to the true integrated variance as the time step between consecutive estimates decreases to zero. However, in practice estimators using high frequency data are often hindered by market microstructure noise, which becomes increasingly noticeable as the sample frequency increases, i.e. when the time step is reduced. Market microstructure noise such as bid-ask bounce and infrequent trading cripple the increases in efficiency associated with increasing the sample frequencies as estimates become inconsistent and biased, as shown by Martens and Van Dijk (2007) among others.

Given the impact of market microstructure noise on the estimates of the spot variance of the associated asset price process, the question arises how these frictions influence the methodology of Renò (2006). Specifically, does the estimation of the variance dynamics of bivariate models hinge on the obtaining the most efficient estimates of the spot variance. If so, is controlling for the effects of market microstructure noise by means of bias-corrected spot variance estimates enough to eliminate the noise-induced biases?

Consequently, this thesis primarily extends the work in Renò (2006) by studying the effects of market microstructure noise, in the form of bid-ask bounce and infrequent trading, on the effectiveness of the methodology. To study this effect and possible solutions, different estimators of integrated variance are introduced to complement the implementation by Renò (2006) of realized variance as an estimator of spot variance. For this purpose, the work of Martens and Van Dijk (2007) provide the reference material for a number of estimators, as they compare the realized variance, realized range, and several (bias-corrected) variations thereof based on high frequency data. Their simulation experiment uses a constant volatility model to generate data and considers the relative performance of the estimators of the spot variance at several sampling frequencies in the presence of bid-ask bounce and infrequent trading. By evaluating the performance of (most of) these estimators in the setting of stochastic volatility, it will be possible to identify if their results translate to this setting. Furthermore, it allows the selection of which estimators perform best in the context of this thesis and to investigate the impact of market microstructure noise on the overall estimation procedure.

The contribution of the research presented here is twofold. First of all, the simulation experiment of Martens and Van Dijk (2007) is extended to the setting of stochastic volatility. Evidence is provided that when estimating the daily volatility when dealing with time-varying volatility dynamics, that the results of Martens and Van Dijk (2007) translate almost directly to this setting. Deviations are found in the performance of their promising bias-correction procedure of the realized range, as here it has a sub-optimal performance with respect to theirs in the presence of bid-ask bounce. Nevertheless, the results indicate that this is due to the settings used in this thesis for the implementation of their bias-correction procedure.

Second, the methodology suggested by Renò (2006) is implemented using estimated series of daily variance for each of the aforementioned estimators. The comparison of the performance of these estimators starts in the basic setting without noise and is followed by the separate introduction of bid-ask bounce and infrequent trading followed by a combined scenario. The presence of noise has clear consequences for the efficacy of the estimation of the underlying dynamics, as the weaknesses and strengths of the daily variance estimators identified in the work of Martens and Van Dijk (2007) extend to this situation and are clearly visible in the resulting estimates of the drift and diffusion. In this the two-time-scales estimator of Zhang et al. (2005) is shown to have the most stable and reliable estimate of the true dynamics in the absence and presence of noise. The scaled realized range should not be dismissed, but needs more sophisticated (automated) methods for selecting its settings (i.e. number of lags used in its scaling factor) for it to be an effective estimator in this setting. The results indicate that if the noise is controlled for in the estimates of the spot variance, then the true dynamics can be consistently estimated.

A closing comment for this introduction is in order. As the work on this thesis has progressed, the author has become aware of the work by Bandi and Renò (2009), which expands on Renò (2006) by considering a larger class of models that include jump processes in both the log asset price and volatility dynamics. Their work controls for market microstructure noise through selection of robust estimators. Also, they develop the basic theory for the spot volatility estimation, however are not specifically focused on the effects of noise itself. Furthermore, no specific (simulation) evidence appears to be given in their work. Therefore this thesis could be seen as a supporting document between the two papers, at least partially abridging the simulation-based evidence of Renò (2006) and the work done by Bandi and Renò (2009).

The remainder of this thesis is organized as follows. The following section describes the general setting and methodology developed by Renò (2006) and Bandi and Renò (2009). Section 3 introduces the stochastic volatility models, along with various estimators for the spot volatility, and the practical implementation of the proposed estimators for the volatility dynamics. Subsequently, section 4 discusses the simulation experiment employed to illustrate the efficacy of the methodology when exposed to market microstructure noise. This section is naturally followed by the results of the various simulations and finally the conclusions which include recommendations for further research.

2 Econometric framework

This section presents the methodology employed in this thesis, developed in Renò (2006) and Bandi and Renò (2009), and introduces some notation and underlying concepts. What follows is a description of the general class of stochastic volatility models under consideration and subsequently the nonparametric estimators central to this thesis with the associated assumptions, concepts, and intuition.

2.1 General stochastic volatility model

Following Renò (2006) and borrowing some notation, the dynamics of the observable state variable X_t and its latent variance V_t on an interval $[0, T]$ are modeled with a stochastic volatility model of the following general form,

$$\begin{aligned} \mathrm{d}X_t &= \mu(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_{S,t}, \\ \mathrm{d}\sigma_t^2 &= m(\sigma_t^2)\mathrm{d}t + \Lambda(\sigma_t^2)\mathrm{d}W_{\sigma,t}, \end{aligned} \tag{1}$$

where $\mu(\cdot)$ and $m(\cdot)$ are real drift functions, and $\Lambda^2(\cdot)$ is a diffusion function. All functions have characteristics such that there exist strong solutions of the stochastic differential equations (SDEs) of the observable variable X_t and unobserved variable σ_t^2. Furthermore, the standard Brownian motions $\mathrm{d}W_{S,t}$ and $\mathrm{d}W_{\sigma,t}$ have (instantaneous) correlation ρ. Note that X_t is purposely left as a generic variable to retain the flexibility to model the asset price or log asset price directly.

Although the technical details on the drift and diffusion functions were omitted in Renò (2006), the necessary assumptions and requirements are partially developed in Renò (2008) and fully in Bandi and Renò (2009). Where the latter paper allows for a larger class of models by including the specification of jumps, allowing the general model considered in Renò (2006) and by extension this thesis, as a subclass (i.e. the case without jumps) with X_t modeled as the log asset price. The full technical details are left for the interested reader, however for intuitive purposes a (mostly) qualitative discussion is included in what follows.

2.2 Infinitesimal moment estimators

A feature of the (general) SV-model under consideration is that one has definitions for the conditional moments that fully describe the temporal development of the process over infinitely small time-steps (Bandi and Phillips (2002)). Therein lies the potential for translation to discrete-time analogues that allow for estimation with observable data. Renò (2006) sets out to exploit the existence of these infinitesimal conditional moment definitions.

The following exposition is restricted to the setting of this paper and therefore focusses on the drift and diffusion of the variance process, corresponding to the first and second infinitesimal conditional moments. As Bandi and Phillips (2002) state, they respectively describe the conditional expectation of the rate of change and the conditional rate of change of the volatility of the process at an arbitrary level x on its domain:

$$m(x) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}\left[\left(\sigma^2_{t+\Delta} - \sigma^2_t\right) | \sigma^2_t = x\right] \tag{2}$$

$$\Lambda^2(x) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{E}\left[\left(\sigma^2_{t+\Delta} - \sigma^2_t\right)^2 | \sigma^2_t = x\right]. \tag{3}$$

Borrowing from Bandi and Phillips (2002) to highlight the intuitive appeal of these functions, consider what follows. The drift at x in (2) can be estimated by constructing the (weighted) average of the first differences of the sample process, conditional on that σ^2_t is in the neighborhood of x. To obtain an asymptotically consistent estimate, one would like as many observations as possible in order to converge in the limit to the conditional moment of interest. To make this possible, over time the underlying process would have to visit x an infinite number of times, which occurs when recurrence is satisfied. Bandi and Renò (2009) state that under recurrence, the process revisits open sets in its range an infinite number of times, or more formally, for any $x \in \mathbb{R}^N$ and $\varepsilon > 0$,

$$\mathbb{P}_x\left(|\sigma^2_t - x| < \varepsilon \text{ for a sequence of times increasing to } \infty\right) = 1. \tag{4}$$

It is in fact this feature that allows for the methodology of Renò (2006) without the necessity for a time-invariant stationary distribution. As Bandi and Phillips (2002) point out, methodologies generally mean to identify the continuous structure of the underlying process(es) by assuming the existence of a time-invariant stationary distribution. Thus

if the joint probability distribution of the underlying process does not change over time, then the statistical properties (such as the mean and variance) are temporally constant. Given the existence of such a time-invariant distribution allows for more straightforward identification using the informational content of discrete observations on the continuous sample path.

However, Renò (2008) points out that while assuming stationarity is convenient, empirical results often indicate the presence of nonstationary behavior (such as persistence features in variance, Bandi and Renò (2009)), thus likely causing inaccurate identification of the continuous dynamics. This gives credence to the imposed relaxation of the assumptions on its distributional properties by requiring recurrence instead.

According to Renò (2008), it is the presence of recurrence that allows for the nonparametric estimation procedure to work as it makes consistent point-wise kernel estimation possible. Reconsider how one would estimate the drift at an arbitrary variance level x. Logically one would expect the 'observed' variations where σ_t^2 is closest to x to contain more information on the drift at x and should accordingly contribute to an estimate thereof the most. Thus instead of an equally-weighted estimator, it is intuitive to consider weighting functions that are centered at x and converge to 0 when the observations are far away, i.e. when $|\sigma_t^2 - x| \to \varepsilon$ the assigned weight approaches zero (Bandi and Phillips (2002)).

Consulting Renò (2008) to give a formal introduction of these type of weighting functions known as kernels. A kernel is considered to be a symmetric nonnegative continuously-differentiable, bounded function $\mathbf{K}(\cdot)$ with an absolutely integrable and bounded derivative $\mathbf{K}'(\cdot)$, and is such that $\int \mathbf{K}(s)\mathrm{d}s = 1$, $\int s^2\mathbf{K}(s)\mathrm{d}s < \infty$, and $\int \mathbf{K}^2(s)\mathrm{d}s < \infty$.

Thus, assuming the processes of the SV-model are recurrent and there is a $\mathbf{K}(\cdot)$ with the aforementioned properties, there remains a final hurdle: volatility is unobservable. To make the step to a practical estimator, Renò (2006) states that if one would be able to make observations on the variance process, then applying nonparametric estimation procedures for the univariate series would lead to reliable estimates of its dynamics. Exploiting this notion, the remaining step is to deal with the latency of the variance process. This can be accomplished by making the variance 'visible' using suitable estimates of the spot variance, $\tilde{\sigma}_t^2$, based on discrete high frequency observations on the associated (log) asset price process.

Combining the above concepts forms the following point-wise moment estimators for the drift and diffusion of the variance process.

Borrowing from Renò (2006) and Bandi and Renò (2009), assume there are $n+1$ equidistant observations on the state variable X_t in the interval $[0, T]$. This yields n intervals of size $\Delta_{n,T} = \frac{T}{n}$. Furthermore, exploiting the present day availability of high frequency data, it is assumed that for $i = 0, 1, ..., n$, the interval $[i\Delta_{n,T}, (i+1)\Delta_{n,T}]$ has a sub-interval of length $\phi_{n,T}$ containing p intra-period observations, which need not be equidistant.

Renò (2006) proposes the following estimators for the real functions $m(\cdot), \Lambda^2(\cdot)$ for the drift and diffusion of V_t respectively, and bases these solely on the (discrete) observations on X_t:

$$\hat{m}(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=0}^{n-1} K\left(\frac{\tilde{\sigma}^2_{iT/n} - x}{h_{n,T}}\right)\left(\tilde{\sigma}^2_{(i+1)T/n} - \tilde{\sigma}^2_{iT/n}\right)}{\sum_{i=0}^{n-1} K\left(\frac{\tilde{\sigma}^2_{iT/n} - x}{h_{n,T}}\right)}, \tag{5}$$

$$\hat{\Lambda}^2(x) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=0}^{n-1} K\left(\frac{\tilde{\sigma}^2_{iT/n} - x}{h_{n,T}}\right)\left(\tilde{\sigma}^2_{(i+1)T/n} - \tilde{\sigma}^2_{iT/n}\right)^2}{\sum_{i=0}^{n-1} K\left(\frac{\tilde{\sigma}^2_{iT/n} - x}{h_{n,T}}\right)}, \tag{6}$$

where $K(\cdot)$ is a kernel, $h_{n,T}$ is a bandwidth parameter, and the integrated variance $\tilde{\sigma}^2_{iT/n}$ is defined as:

$$\tilde{\sigma}^2_{iT/n} = \frac{1}{\phi_{n,T}} \int_{iT/n}^{iT/n+\phi_{n,T}} \sigma^2_s(d)s, \tag{7}$$

where $\Delta_{n,T} \to 0$ as $n, T \to \infty$. Perhaps redundant, but as stated before $[iT/n, iT/n + \phi_{n,T}] \subseteq [iT/n, (i+1)T/n] = [i\Delta_{n,T}, (i+1)\Delta_{n,T}]$ to make the order of the intervals specific. Clearly, this setup is created to allow for trading days shorter than the full 24 hours.

However, as stated before, the integrated variance $\tilde{\sigma}^2_{iT/n}$ is unobservable, thus it must be estimated by means of a suitable estimator $\hat{\sigma}^2_{iT/n}$. The methodology of Renò (2006, 2008) suggests using the p high frequency observations on the state variable X_t in the interval $[iT/n, iT/n + \phi_{n,T}]$ to estimate the integrated variance over that interval, thereby constructing variance estimates at a lower frequency. Replacing $\tilde{\sigma}^2_{iT/n}$ by $\hat{\sigma}^2_{iT/n}$ in (5) and (6) produces the feasible estimators. Thus for example, if there are observations on X_t at 5 minute intervals, then one could construct estimates of the daily integrated variance, with which one would then construct $\hat{m}(\cdot)$ and $\hat{\Lambda}^2(\cdot)$.

3 Implementation

This section discusses the implementation of the described methodology to data series that are to be generated with two models for time-varying volatility, being a generalized ARCH (GARCH) model and the Heston model. This is followed by a specification of the implemented estimators of the integrated variance. Subsequently, the implementation is described of the estimators for the drift and diffusion of the variance process.

3.1 Stochastic volatility models

To facilitate the extension on the results of Renò (2006), price paths (with equidistant prices) are simulated using the Heston model as well as a GARCH continuous time model. The choice of the simulated models stems in part from their popularity, as both models are widely used in financial applications ranging from portfolio management to option pricing.

On the one hand, the Heston model is a natural extension on the Black-Scholes model, essentially extending it to allow for the otherwise constant volatility to be modeled as a time-varying process, thus creating the bivariate set of SDEs described below in (9). The Heston model is widely used for option pricing applications as closed-form evaluation is readily available, which is not the case with most stochastic volatility models. The Heston model falls in the category of continuous-time stochastic volatility models and implicitly models the predictive distribution via the structure of the model (Shephard and Andersen (2008)).

On the other hand, the GARCH model (also) readily takes into account well known characteristics of financial time-series, being fat tails in the return distribution (i.e. excess kurtosis) and volatility clustering. The (originally) discrete-time specification allows for the GARCH model to be widely used in practice for applications such as describing the term structure of interest rates and the modeling of persistence and fat tails in exchange rate data. However wide and simple the application of the GARCH model is, the reliance on parametric specification makes it vulnerable to unanticipated conditions in the market. The continuous-time specification of the GARCH model that is used here, is employed in several works such as Drost and Werker (1996) and Alexander and Lazar (2005). This specification makes the implemented nonparametric approach interesting for GARCH models, as it allows for observation-based inference of the model while imposing a minimal structure on its continuous-time dynamics.

Continuous-time specification

To place the volatility dynamics of the respective models in terms of the basic continuous-time model presented in (1) of the previous section, both models exhibit linear drift,

$$m(\sigma^2) = \kappa(\theta - \sigma^2),$$

and the diffusions for Heston and GARCH are respectively:

$$\begin{aligned} \Lambda(\sigma^2) &= \omega\sqrt{\sigma^2} \\ \Lambda(\sigma^2) &= \omega\sigma^2, \end{aligned} \tag{8}$$

where κ is the speed of mean-reversion of the variance process, θ is the long-term mean level of variance, ω signifies the volatility of variance (or volatility depending on α), and are all nonnegative constants. Borrowing some notation used in Lord et al. (2009) for the general class of constant elasticity of volatility SV models (CEV-SV), we have:

$$\begin{aligned} \mathrm{d}X_t &= \mu X_t \mathrm{d}t + \sqrt{\sigma_t^2} X_t^{\beta} \mathrm{d}W_{X,t} \\ \mathrm{d}\sigma_t^2 &= \kappa(\theta - \sigma_t^2)\mathrm{d}t + \omega(\sigma_t^2)^{\alpha} \mathrm{d}W_{\sigma,t}, \end{aligned} \tag{9}$$

where β controls the elasticity of the variance and the (log) asset price, α allows for summarizing both implemented models in one general scheme. Clearly, $\alpha = 0.5$ for the Heston model and $\alpha = 1$ for the GARCH model.

While on the subject of this general class of model, Lord et al. (2009) list some properties of interest when used for simulation purposes. Limiting the discussion thereof to what is relevant to the Heston and GARCH models, consider the mean-reverting CEV process σ_t^2. Then 0 is an attainable boundary for the Heston model when $\omega^2 > 2\kappa\theta$, such that the process can touch zero, but will spend no time in it. For the GARCH model, 0 is unattainable as $\alpha > 1/2$. For both models, ∞ is unattainable (due to the mean-reversion). The list of properties extends itself to the asset price process X_t, commenting that it can reach 0 with positive probability. Clearly, these properties are of importance for the simulation of both models and must be taken into consideration when selecting parameter configurations.

Discretization

To allow for meaningful simulations, a sensible conversion of the continuous-time to a discrete-time specification of the implemented models must be made to allow for an honest assessment of the performance of the moment estimators. To accomplish this, it is convenient for the models to be discretized with an Euler-Maruyama scheme (Euler from here on).

However, as Lord et al. (2009) point out, despite the straightforward nature of Euler discretizations, there are some downsides to the methodology. The first being that the bias introduced by the selected time discretization is unknown, requiring numerous runs of the simulation to identify the accomplished level of accuracy. Conversely, given a required level of accuracy, one might need an excessively small time step to eliminate the bias associated with the discretization. Nevertheless, this also means that it can be shown that with increasingly small time steps the discretization bias will become negligible and consequently that the discretized process converges to the true continuous-time process (under certain conditions). In this thesis, high frequency data is simulated, thereby keeping the size of the time step between consecutive prices quite small, thus likely keeping the discretization bias to a minimum.

A more practical drawback is that, in the setting of the square root process of the Heston model, the discretized variance process has a strictly positive probability of becoming negative. As Lord et al. (2009) point out, this is directly related to the higher volatility of the variance in the stochastic volatility setting than the volatility in the interest rate setting. Furthermore, convergence of the discretized square root process is not guaranteed, especially around zero. Thus, there is a possibility of the discretized variance process to step out of the nonnegative domain of the square root process of the Heston model, causing the discrete-time process to become undefined, while the true process remains positive and close to zero.

Introducing some notation to denote the difference between the continuous-time series and discrete-time series, let V_t represent the discrete version of σ_t^2 and Δt the discretized time step. As stated in Lord et al. (2009), the probability of the discretization causing a negative

value at $V_{t+\Delta t}$ at when $V_t > 0$ is:

$$\mathbb{P}(V_{t+\Delta t} < 0) = \Phi\left(\frac{-\kappa(\theta - V_t)\Delta t - V_t}{\omega V_t^{\alpha}\sqrt{\Delta t}}\right) \tag{10}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and Δt reflects the time-step between observed high frequency data. Clearly, the probability decreases with the size of the time step and has a dependency on the degree of mean-reversion and volatility of variance.

To remedy this (potential) behavior, there are a few quick fixes described in Lord et al. (2009) that are often used. However, they unify these methods and introduce their own *full truncation* scheme to guarantee the nonnegativity of the discretized variance process. They show that in the setting of option pricing, the full truncation scheme yields the best results (smallest bias) when compared to several alternative euler discretizations. Furthermore, strong convergence is proven for the full truncation scheme, dealing with the earlier objection to the method.

As stated, the full truncation scheme outlined in Lord et al. (2009) is used on the discretized Heston and GARCH models to deal with possible negativity of the variance process. Applying the full truncation scheme yields the following discretized and augmented process:

$$\begin{aligned} V_{t+\Delta t} &= V_t + \kappa\Delta t(\theta - \max(V_t, 0)) + \omega(\max(V_t, 0))^{\alpha}\Delta W_{V,t} \\ \breve{V}_{t+\Delta t} &= \max(V_{t+\Delta t}, 0) \end{aligned} \tag{11}$$

With respect to using the full truncation scheme for the GARCH model, similar to the Heston model, the variance process under GARCH has a risk of becoming negative. This because the probability of $V_{t+\Delta t}$ becoming negative, as described in Lord et al. (2009), is independent of the choice for α and is an issue related to Euler discretizations rather than to the stochastic specifications. Regardless, the variance process is not hindered by the full truncation scheme if it never becomes negative as the scheme then collapses to the identity function, and if it does become negative it requires a scheme to adequately deal with that.

To finish up the considerations on the implemented models, there remains the specification of the asset price processes. Following the exposition in Lord et al. (2009) for the Heston model, X_t represents the asset price. To ensure that the price process in (9) stays positive,

it is log-transformed, Itô's lemma yields the following process:

$$\begin{aligned}\ln X_{t+\Delta t} = \ln X_t + \left(\mu - \frac{1}{2}\breve{V}_t X_t^{2(\beta-1)}\right)\Delta t \\ + \sqrt{\breve{V}_t} X_t^{\beta-1}\left(\rho \Delta W_{V,t} + \sqrt{1-\rho^2}\Delta Z_t\right),\end{aligned} \tag{12}$$

where the Brownian motions $\Delta W_{V,t}$ and ΔZ_t are independent and implemented as $\epsilon_{V,t}\sqrt{\Delta t}$ and $\epsilon_{S,t}\sqrt{\Delta t}$, where $\epsilon_{\cdot,t}$ are (semi-)random numbers from the standard normal distribution generated with Matlab's `randn` function. As noted in Lord et al. (2009), the relationship $\Delta W_{X,t} = \rho \Delta W_{V,t} + \sqrt{1-\rho^2}\Delta Z_t$ can be found by means of a Cholesky decomposition. Furthermore, note that if the parameters are annualized, then $\Delta t = 1/(D \cdot J)$, with D the number of annual trading days and J the number of data points per trading day (Martens and Van Dijk (2007)). If the parameters were daily parameters, then D would be equal to one. On a practical note, in the programming code $\ln X_t$ is actually X_t, meaning that $X_t^{2(\beta-1)}$ in the formula above is implemented as $\exp(X_t)^{2(\beta-1)}$.

For the implementation of the GARCH model, as in Renò (2006), this paper considers the continuous time model defined in the article by Drost and Werker (1996). With zero drift ($\mu = 0$), $\beta = 0$, and X_t modeled directly as the log asset price, the implemented GARCH model is consistent with the continuous time model of Drost and Werker (1996). Thus the log asset prices are modeled directly as:

$$\ln X_{t+\Delta t} = \ln X_t + \sqrt{\breve{V}_t}\left(\rho \Delta W_{V,t} + \sqrt{1-\rho^2}\Delta Z_t\right) \tag{13}$$

Drost and Werker (1996) assume zero correlation between the log asset price and variance processes, i.e. $\rho = 0$. However, to stay consistent with the presented framework and maintain (some) generality, no restriction on the correlation is made here.

Summarizing, there are two stochastic volatility models under consideration, the Heston and GARCH model, which are employed to generate high frequency log asset prices. Both are implemented by means of an Euler discretization and augmented with the full truncation scheme detailed in Lord et al. (2009) to ensure that the discretized models are defined throughout the simulations.

3.2 Spot variance estimators

To estimate the (intra-period) integrated variance for estimation of the functions, several spot variance estimators are implemented. These estimators are the realized range (RR), realized variance (RV), the two-time-scales realized variance (RV_{TTS} or TTS) estimator as well as the scaled realized range (RR_S or RRS) and scaled realized variance (RV_S or RVS). Using (mostly) the notation in Martens and Van Dijk (2007), we have the following definitions for the first three estimators for the variance on day t:

$$RV_t^{\Delta} = \sum_{i=1}^{I} r_{t-1+i\Delta,\Delta}^2 \tag{14}$$

$$RR_t^{\Delta} = \frac{1}{4 \ln 2} \sum_{i=1}^{I} (\sup_{(i-1)\Delta \le j \le i\Delta}(\ln S_{t-1+j}) - \inf_{(i-1)\Delta \le j \le i\Delta}(\ln S_{t-1+j}))^2 \tag{15}$$

$$RV_{\text{TTS},t}^{\Delta} = \frac{p}{p - \bar{I}} \left(RV_{\text{Subs},t}^{\Delta} - \frac{\bar{I}}{p} RV_{\text{Max},t} \right) \tag{16}$$

where $r_{t-1+i\Delta,\Delta} = \ln X_{t-1+i\Delta} - \ln X_{t-1+(i-1)\Delta}$. Following Martens and Van Dijk (2007) we observe $p+1$ equidistant prices per trading day, where p is an integer multiple of I, which denotes the number of intraday intervals of length Δ corresponding to a certain sampling frequency.

As described in Martens and Van Dijk (2007) a grid of intervals of length $\Delta = 1/I$ can be laid over the trading day in p/I different ways. Each of these configurations of the grid of Δ-length intervals forms a subsample for the trading day. Note however that each subsample need not contain the same number of return observations. This is captured by the average number of intraday returns per subsample $\bar{I} = (I-1) + I/p$, assuming each consecutive equidistant data point is used as a starting point for a subsample. In equation (16) the average realized variance over the subsamples is denoted by $RV_{\text{Subs},t}^{\Delta}$ and the realized variance over the highest frequency is denoted by $RV_{\text{Max},t}^{\Delta}$. Following Martens and Van Dijk (2007), the realized variances that miss part of the trading day are inflated proportionally.

To be specific on the intermediate steps for $RV_{\text{TTS},t}^{\Delta}$, the construction of $RV_{\text{Subs},t}^{\Delta}$ is as follows:

$$RV_{\text{Subs},t}^{\Delta} = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{i=1}^{I_k} (\ln X_{t-1+k\delta+i\Delta} - \ln X_{t-1+k\delta+(i-1)\Delta})^2$$

where K denotes the number of subsamples and δ denotes the shift between subsamples such that $K\delta = N/I$, and I_k denotes the number of intervals in subsample k. Note that the adjustment on the definition of the $RV^{\Delta}_{\text{Subs},t}$ allows for non-consecutive (but still equally-spaced) data points to serve as a starting point of a subsample. A further adjustment is to the formula for $\bar{I}$. When consecutive data points are used $K = p/I$ as $\delta = 1$, but otherwise $\bar{I}$ should be defined as $\bar{I}^* = (I-1) + (I\delta)/p$ or equivalently $\bar{I}^* = (I-1) + 1/K$. Clearly, when $K = 1$ and/or $\delta = 1$ as is the case for RV^{Δ}_t and $RV^{\Delta}_{\text{Max},t}$, this yields the same as the initial definition of RV^{Δ}_t in (14).

With respect to the proportional inflation for the realized variances that miss part of the trading day, this involves a 'per-subsample' adjustment by means of a multiplier of I/I_k. E.g. in case of a sample frequency of four times a day (90 minutes on a 6 hour trading day), all but one subsample miss a quarter of the trading day thus implying the multiplier for those shorter subsamples to be 4/3.

The intuition behind the TTS estimator is described by Martens and Van Dijk (2007) and Zhang et al. (2005) as that the lower frequency is adjusted for noise-induced bias by means of the highest available frequency. The highest sampling frequency is thought to estimate the variance of the noise, thereby allowing an adjustment on the lower frequency aimed to remove this noise variance in order to better estimate the actual volatility of the asset in question. The variance of the estimates is further reduced by means of subsampling, thereby better exploiting the availability of high frequency data. As far as this exploitation goes, it could be seen as an intermediate between the realized range, which uses all data in an interval, and the realized variance, which uses only the endpoints of an interval, for the construction of the spot variance estimate.

As previously stated, the scaled versions of the realized range and realized variance described in Martens and Van Dijk (2007) are also implemented. The scaling factor being the ratio between the lowest-frequency (e.g. daily) range and the realized range over the previous q trading days given a sampling frequency. So we have:

$$RR^{\Delta}_{S,t} = \left(\frac{\sum_{l=1}^{q} RR_{t-l}}{\sum_{l=1}^{q} RR^{\Delta}_{t-l}} \right) RR^{\Delta}_t \tag{17}$$

$$RV^{\Delta}_{S,t} = \left(\frac{\sum_{l=1}^{q} RV_{t-l}}{\sum_{l=1}^{q} RV^{\Delta}_{t-l}} \right) RV^{\Delta}_t \tag{18}$$

where RR_t^{Δ} is as defined in (15), thus including the $\frac{1}{4 \ln 2}$. The thought behind this is that the daily variance contains little contamination by microstructure noise and thus gives a good indication of the actual level of variance, and thus can be used to adjust the high-frequency estimates back towards the true value (Martens and Van Dijk (2007)).

3.3 Estimating the model functions

Using the high frequency estimators of the previous section, the next step is to estimate the drift and diffusion functions using equations (5) and (6) of section 2. For each of the (intra-period) paths of the simulation, we have an estimate $\hat{V}_{iT/n}$ of the spot variance with which the model dynamics can be estimated. To do so, further details are needed on the choice for the kernel function $K(\cdot)$ and for the bandwidth parameter $h_{T/n}$.

To stay in line with the proposed methodology, the Gaussian kernel is chosen for $K(\cdot)$. As common knowledge will tell, this entails:

$$K\left(\frac{\hat{V}_{iT/n}^2 - x}{h_{n,T}}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(\hat{V}_{iT/n}^2 - x)^2}{2h_{T/n}^2}\right), \tag{19}$$

There a number of alternative choices for the kernel function, such as the Epanechnikov and quartic kernels. Both of which assign higher values to observations that lie further away from the grid point.

However, the choice of the kernel function is not as crucial as the selection of the bandwidth parameter, as it indirectly controls the variance of the kernel. A larger bandwidth will allow for a good feel for the global characteristics of the estimator, but more subtle properties will be smoothed out. A smaller bandwidth might overaccentuate certain aspects due to undersmoothing. Ergo, the selection of the optimal bandwidth can be quite challenging.

Following Renò (2006), the bandwidth parameter is set by the following:

$$h_{T/n} = h_s \cdot \tilde{s} \cdot n^{-1/5} \tag{20}$$

where h_s is approximately equal to 1.06 (which is $(4/3)^{1/5}$), and $\tilde{s}$ is the sample standard deviation of the variance estimators. This choice for the bandwidth parameter stems from the rule-of-thumb developed in Scott (1992) (for the univariate case) and is generally adequate for most applications.

So prior to estimating the model functions, it will be necessary to determine the sample standard deviation of the intra-period variance estimators. Furthermore the set of values for x at which the model functions are to be evaluated must be determined. The values for x are a grid spanning the closed interval with a lower-bound and an upper-bound determined as the mean intra-period variance minus and plus 3 times the standard deviation. This choice allows for the (observed) domain of the estimated variances to determine what is displayed in the (graphical) results. The probability of reaching values beyond that interval will (effectively) be zero, causing the estimators in equations (5) and (6) to be undefined. If there are higher variances estimated, then these will likely not be well-represented and fail to give meaningful results at those grid points. Clearly, the lower-bound of the interval on x will not surpass zero onto the negative real number line.

Having set the interval, it remains to set the number of grid-points in that interval at which the moment estimators are evaluated. Clearly, a higher value for G will give a better read of the model function and how it approximates the actual variance dynamics, but of course the computational load must be taken into consideration.

4 Simulation setup

This section discusses the simulation setup employed to gain (asymptotic) insight in the effects of market microstructure noise on the estimation of the dynamics of the variance process. First the basic simulation setup is discussed, including the chosen parameter sets for the models and relevant practical information relating to the implementation of the simulations. Subsequently the implementation of the noise is discussed.

4.1 Basic simulation

To achieve some sense of the asymptotic behavior of the moment estimators in the presence of noise, 500 independent price paths are simulated for both the GARCH and Heston model. Although this is only half the number of paths employed in most studies, including Renò (2006), the used number should provide enough insight for indicative purposes. Each price path consists of a total of 6 000 trading days, which includes the $q = 1\,000$ trading days as a lead-in period for the scaled estimators. Each trading day sees 1 price per second for a 24-hour period. This yields 86 400 prices per day and thus a grand total of 432 000 000 in-sample prices over the 5 000 trading days in one price path (518 400 000 prices over the full 6 000 days).

Assuming 250 trading days per year, this translates to $T = 20$ years of "in-sample" data. A further consideration is that in the setting of continuous 24-hour trading, the opening price of day t equals the closing price of the previous day. For each trading day, the daily variance is estimated by means of the high-frequency estimators introduced in section 3.2 at the following sampling frequencies: 1, 2, 3, 4, 5, 10, 15, 20, 30, and 60 minutes. As a point of reference and because of its use in the construction of the scaled estimators for the estimation of the mean long-term daily variance (θ), the spot variance estimators are also computed at a daily (1 440 minutes) sample frequency. The daily sample frequency is not used for the evaluation of the moment estimators.

With respect to the implementation of $RV^{\Delta}_{TTS,t}$, for the construction of $RV^{\Delta}_{Subs,t}$ a subsample shift for the construction of $\delta = 10$ (i.e. 10 seconds) is used, therein following Martens and Van Dijk (2007). Thus every tenth observed price serves as a starting point for a subsample. Furthermore, the every (observed) data point is used for the construction of $RV^{\Delta}_{Max,t}$. Thus, in the case of fully observed price paths, its sample frequency is 1 second.

Furthermore, to construct the scaled realized range and scaled realized variance estimators, Martens and Van Dijk (2007) initially set q to 5 000, being the total number of simulated trading days. The Monte Carlo simulation then has a lead-in period of 5 000 days to provide enough 'lags' for the scaling factor of the first day, thus simulating a total of 10 000 days. As stated before, in this thesis the scaling factors for $RR^{\Delta}_{S,t}$ and $RV^{\Delta}_{S,t}$ are constructed with $q = 1\,000$. In Martens and Van Dijk (2007), the accuracy of these estimators improved little in their scenario with infrequent trading for $q > 1\,000$ at the higher sampling frequencies and the marginal contribution to the accuracy of higher q already declined after $q = 500$. For low frequencies the added value of more lags is less and a possible cut-off value could be found to be below the selected q. They comment that choosing the number of lags should be tuned to whether or not the spread and trading intensity is constant over time, indicating that a high q is beneficial in the affirmative case and a low q is better when these noise factors are varying. In this thesis the trading intensity and spread within each scenario will be kept constant over time. Therefore the selected q is deemed to supply enough accuracy for these scaled estimators and allows for considerably less computations.

The parameters in the table below are for the GARCH and Heston model, respectively. The set of parameters for the GARCH model matches those used by Renò (2006) and originate from Andersen and Bollerslev (1998), representing parameters calibrated to the DM-USD exchange rate over the period from October 1, 1987, through September 30, 1992. The original parameters of Andersen and Bollerslev (1998) are used 'as-is' and are originally in a daily format based on 250 trading days per year. For the Heston model, the parameters are borrowed from Lord et al. (2009), where they are used in the setting of pricing plain vanilla European options, and are converted from annualized to daily parameters. When dealing with daily variance, it is natural and standard practice to focus on the daily dynamics, hence the choice for the daily parameters. Adopting the representation in Andersen and Bollerslev (1998) and Bandi and Renò (2009), the parameters reflect daily percentage returns. Thus, daily variance will carry a factor of 10 000 accordingly.

Table 4.1: Parameter configurations (daily)

Model	V_0	κ	ω	ρ	θ	α	μ
GARCH	0.636	0.035	0.144	0	0.636	0.5	0
Heston	1.6	0.5	0.632	−0.9	1.6	1	0.05

Some comments are in order regarding the chosen parameter sets. Starting with the properties discussed in Lord et al. (2009) relating to the relative parameter values leading to 0 and ∞ being (un)attainable boundaries for the variance process and how this relates to the full truncation method. Given the CEV process employed and a set of parameters for which $\omega^2 > 2\kappa\theta$ and $\alpha = 0.5$, then 0 is an attainable boundary. This is clearly the case for the parameter set used for the Heston simulation. However, for the GARCH model it isn't (as) relevant as $\alpha > \frac{1}{2}$ and thus 0 is unattainable (Lord et al. (2009)). The parameter selection for the Heston model clearly advocates the use of the full truncation method to deal with the potential negativity of the variance process. For the GARCH model no negative effects will be encountered due to the method's neutrality towards positive values.

Furthermore, the high value for the volatility of volatility, ω, in the Heston model allows capturing the heavy skew encountered in equity and foreign exchange markets. The set of parameters for the GARCH model, with a relatively small ω, further reduces the probability of obtaining a negative value of V_t. However the probability of obtaining a negative value of the discretized variance process is still non-zero.

As a final note, the choice to use daily variance estimates to estimate daily dynamics has implications for the moment estimators, in that the ratios n/T and $\phi_{n,T}$ are equal to 1. If the choice had been to investigate annualized dynamics with daily variance estimates, n/T would equal 250 (i.e. $5\,000/20$) with $\phi_{n,T}$ still equal to 1. Thus, the daily spot variance estimates are annualized by $\phi_{n,T}$ and the time step between the spot variances is annualized by n/T. Additionally, it must be noted that for the estimation of the drift and diffusion, the bandwidth parameter $h_{T/n}$ is computed for each individual price path.

4.2 Adding Market Microstructure Noise

As the extension of interest on the previously described basic simulation, the effects of infrequent trading and bid-ask bounce on the estimation of the variance dynamics are explored. Seeing as the various estimators of the spot variance are all to different degrees susceptible or robust to the effects of market microstructure, it will be interesting to see in what way these types of noise affect, if at all, the performance of the model function estimators. The implementation of the types of noise that are included is described below and follows Martens and Van Dijk (2007).

Infrequent trading

To emulate infrequent trading, the fully observed path of X_t of the basic simulation is filtered to reflect the probability of observing a price once every 10 seconds to create the filtered series $X_{inf,t}$. Thus with, say, 100 prices per second, the probability of observing a price is $\pi = 1/(100) = 0.01$. When no price is observed at time $t+1$, then the last observation serves as the current price, i.e. $X_{inf,t+1} = X_{inf,t}$. To randomize price observations, a series of independent Bernoulli random variables is used with success probability π. The random numbers are generated using MATLAB's `binornd` function, which generates binomial random numbers as a sum of n Bernoulli random variables, so with $n = 1$ this generates a Bernoulli random variable. Thus with BI_{t+1} taking the value 1 if the price is observed and 0 when not:

$$X_{inf,t+1} = X_{t+1} \cdot BI_{t+1} + X_{inf,t}(1 - BI_{t+1}). \tag{21}$$

Bid-ask bounce

The implementation of bid-ask bounce simply entails subtracting/adding half the spread from/to the fully observed path X_t, where observing a bid or ask price is equally likely. Similar to the addition of infrequent trading, to simulate this behavior, a series of independent Bernoulli random variables are used. When the series equals one, the observed price is an ask price, and otherwise it is a bid price (although this assignment is clearly arbitrary). The spread s is set equal to 0.05% of the initial price. So where BB_{t+1} equals 1 if it is an ask price and -1 if it is a bid price:

$$X_{bounce,t+1} = X_{t+1} + 0.5s \cdot BB_{t+1} \tag{22}$$

Combining infrequent trading and bid-ask bounce

The last step is to combine both types of noise to explore the combined effect on the model function estimators. To achieve this, the path X_t is first adjusted to reflect bid and ask prices and then filtered for infrequent trading, thus creating the series $X_{combo,t}$. Thus:

$$X_{combo,t+1} = (X_{t+1} + 0.5s \cdot BB_{t+1}) \cdot BI_{t+1} + X_{combo,t} \cdot (1 - BI_{t+1}). \tag{23}$$

Clearly, the order of applying bid-ask bounce and infrequent trading should not make a difference, given that both are independent random processes.

4.3 Performance measures

As in Martens and Van Dijk (2007), the performance of the estimators of integrated variance is measured by the mean estimated daily variance and root mean squared error (RMSE) over all price paths in the simulation, excluding the lead-in period. To be specific on the calculation of the latter and using RR_t^{Δ} as an example, the RMSE is computed as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{N}(RR_t^{\Delta} - V_{\text{Day},t})^2}{N}} \tag{24}$$

with N the total number of days (i.e. 500 x 5 000) and $V_{\text{Day},t} = \frac{1}{m+1}\sum_{j=1}^{m+1} V_{t+j}$, the sum over all intraday values on day t of the true discretized variance process V_t. Its sample average $\bar{V}_{\text{Day},t}$ noted under each table of results, naturally does not include the lead-in period of length q.

Apart from the performance measures for the spot variance estimators, performance measures need to be defined for the drift and diffusion estimators. In addition to graphical evaluation, to obtain a quantified feel for the performance of the estimators, Kanaya and Kristensen (2009) is consulted. They evaluate by means of approximating the integrated squared bias, variance, and mean squared error on a closed interval that represents 95% of values reached by the true variance process. The three measures are related in that the sum of the integrated squared bias and the mean integrated variance equal the mean integrated squared error.

The performance measures used here are based on those in Kanaya and Kristensen (2009) and are defined as follows. Using the same definition for the interval on which the performance measures are approximated, let $x \in [a, b]$ represent 95% of the generated variances. Instead of the integrated squared bias, the integrated absolute bias (or root of the square) is used here to avoid explosive results due to the use of daily percentages in the returns. The integrated absolute bias for the drift estimates is defined as

$$\text{Bias} = \int_a^b |m(x) - \bar{m}(x)|\mathrm{d}x \tag{25}$$

where $m(x)$ is the theoretical value at x and $\bar{m}(x) = \frac{1}{S}\sum_{s=1}^{S} \hat{m}_s(x)$ represents the mean estimated drift at x over all $S = 500$ price paths. Accordingly, the root mean integrated

squared error (RMSE)[1] is defined as

$$\text{RMSE} = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\int_{a}^{b}(\hat{m}_s(x) - m(x))^2 \mathrm{d}x} \tag{26}$$

where the mean integrated variance is left unreported as it can easily be computed as the difference between the squared Bias and square of RMSE. Also note that the MISE is not a new notion when dealing with kernel estimation, as it has been studied and used by Tarter (1986) and Marron and Wand (1992), among others. The integrals in both measures are estimated by means of the sum over even-spaced grid points on the interval [a,b]. For the GARCH model this interval is approximated by [0.15, 1.25], while for the Heston model this comes down to [0.5, 7.68]. Clearly, the higher the number of grid points, the better the approximation of the integral.

Furthermore, relevant results for each moment estimator are presented in graphs, that include the mean estimate and 95% confidence intervals over the entire sample and are constructed as (using the first moment as an example):

$$CI = \bar{m}(x) \pm 1.96 * s_{\hat{m}(x)}, \tag{27}$$

with $s_{\hat{m}(x)}$ as the standard deviation of $\hat{m}(\cdot)$ at grid point x. Construction of the confidence intervals for $\hat{\Lambda}^2(x)$ is analogous. The graphs will show the mean estimated moments and the 95% confidence intervals set off against the true dynamics and the mean estimated moments using the generated variance series at the highest frequency.

It will become obvious that graphical inspection of the performance of the estimators is helpful in the identification of optimal sampling frequencies of the estimators when using the selected performance measures. Integrating the bias removes the information on the behavior of the estimates at different distances from θ, therefore using the performance measures in combination with graphical representation allows for the best interpretation of the results. Given the large number of potential graphs to be generated, only relevant graphs are presented in-text and in the appendices. The graphs will at least contain the true dynamics, denoted as 'theoretical', and the dynamics as estimated with the generated

[1]Strictly speaking the acronym is RMISE, but this gives aesthetically unpleasant table headers and thus is shortened to RMSE. The 'I' will be implied by context.

variance series V_t, denoted as 'generated'. Furthermore, the dynamics with using estimated series of interest are included, which is denoted by 'estimated' followed by the name of the spot variance estimator and the used sampling frequency. The relevant confidence intervals are denoted by '95% Confidence Interval'.

4.4 Practical information

As a concluding note for this section, there are some practical comments in order. First of all, the simulations were conducted in Matlab 7.5 (R2007b) on a laptop with a 2.4 GHz Intel Core2 Duo-processor with 4 Gb of DDR2 SDRAM. Given this system, the selected setup, and the programming, the computational time for a full set of 500 price paths lies in the neighborhood of 11 days.

Furthermore, special care must be taken to avoid identical simulated paths due to restarting the software package, i.e. Matlab. The nature of semi-random number generators employed in mathematical software such as Matlab, is that usually the state of the number generator is reset to the same default state at every restart of the program. Failing to change the state would result in the generation of identical sequences of random numbers, effectively reducing the actual number of repetitions one would simulate per session of the software. Assurance of a different state of the number generator can be achieved by linking its initial state to the system date and time, code for which has been included in the Appendix.

5 Results

This section presents the results of the simulation study by considering each of the four scenarios. The results for the basic scenario without market microstructure noise are discussed first. Followed by the results for the cases with either bid-ask bounce or infrequent trading present. The scenario in which both noise types are present is discussed last. The analysis of each scenario starts with a discussion of how the estimators of integrated variance perform in the setting of stochastic volatility, followed by their performance in estimating the drift and diffusion of the variance process for the GARCH and Heston models. Most tables and graphs are placed in the appendices.

5.1 No microstructure noise

Tables A.1 and A.4 in Appendix A show the results for the estimation of the long-term mean daily variance using the introduced estimators of the integrated variance for resp. the GARCH and Heston models. What immediately comes to attention is how the realized range underestimates the mean daily variance, with better performance on the mean at the lower frequencies. However, the closer estimate of the mean at the daily frequency comes at the cost of efficiency, yet is still roughly 5 times as efficient as the RV-based estimators at the same frequency. While the RMSE for realized range at the highest (presented) frequencies is clearly the worst among the estimators, the RMSE is best across the board at the lower frequencies of 30 and 60 minutes and already at the 15 and 20 minutes for the Heston model. This observation is in line with the expectations, as Martens and Van Dijk (2007) and others report similar relative efficiencies between the realized range and realized variance in the setting of constant volatility in the price process.

Where the results on the realized range differ from those found by Martens and Van Dijk (2007) is that, here, the estimated mean shows heavier underestimation and poorer efficiency. The difference is a direct result of observing 1 price per second versus the 100 prices per second that Martens and Van Dijk (2007) employ. Thus, the lower frequency of the data generating process effectively form an (unintended) intermediate version of the fully observed case and infrequent trading scenario in Martens and Van Dijk (2007), where respectively 100 prices per second and 1 price per 10 seconds are observed. The scenario with infrequent trading should confirm this, as fewer observations should exacerbate the results.

The underestimation of the realized range is in itself expected, as by construction the estimator relies on taking advantage of all the observed data within a sampling interval. Whereas, by comparison, the realized variance will maintain the same estimate regardless of how much data points are contained by the same interval, by only considering the begin and end points of that interval. To achieve their results, Martens and Van Dijk (2007) point out that the time step between observations should be a small as possible to minimize the under-/overestimation of the maximum and minimum prices in the interval.

Nevertheless, in agreement with their results, the performance of the realized range is considerably improved by implementing the scaling procedure of Martens and Van Dijk (2007), achieving a better estimate of the mean and improving on the efficiency. The scaled range performs at the highest efficiency at all reported frequencies lower than 30 minutes, otherwise only beaten by the realized range, while only slightly underestimating the mean. Furthermore, in this setting, the scaled range achieves about the same efficiency at the 10-minute frequency as the scaled realized variance at the 1-minute frequency in the results for the GARCH model. In the considerably noisier results of the Heston model, this efficiency gap is quite smaller, but nonetheless quite clear at the 15 and 20 minute sample frequencies.

The scaled realized variance has the poorest performance among the RV-based estimators, which are dominated by the efficient performance of the TTS estimator. The superior performance of the TTS estimator is attributed by Martens and Van Dijk (2007) to the use of subsamples to decrease the variance of the estimates, which is in line with the intentions of its proposers, being Zhang et al. (2005). It is in the RMSE that the TTS estimator wins over the realized variance, as the mean estimates are almost identical at the highest frequencies. The dominant results of the TTS estimator are quite clear for the estimates of the mean daily variance of both the GARCH and Heston model.

As a closing observation on the estimators for the daily variance, the parameter choices for the respective models come through quite clearly, as the higher long-term mean variance and volatility of variance of the Heston model drives up the RMSE of the estimates. The RMSE of the estimates for the Heston model are almost 11 times larger than those for the GARCH model at the same sampling frequencies. The effects of which are even more pronounced in the results for the estimates of the drift and diffusion dynamics in tables A.2 and A.3 for the GARCH model, and tables A.5 and A.6 for the Heston model,

respectively. Clearly, the RMSE of the drift estimates of the Heston model are excessively large (although this could also be accentuated by a possible scaling issue). Underlining these troublesome results is the fact that at 1.6998, the long-term mean daily variance of the generated variance series deviates from its true counterpart of 1.6, which naturally is quite unexpected.

Accordingly, the composite graphs A.9 through A.12 for the Heston model, using scaled realized range and TTS estimators as examples, show a considerably less distinct picture than those for the GARCH model (at least the drift estimates). As becomes clear from the graphs of the generated series for the Heston model, figure 5.1, the generated variance series itself also provides a fairly loose fit to the true dynamics, which is especially the case for the estimated drift. The high volatility of variance ω selected for the generation of the data is one of the culprits of this. Its influence is clearly noticeable in the RMSE of all estimates and quite visible in the large confidence bands in figure 5.1 as well as the considerably larger range of daily variance values under consideration (compared to the presented GARCH model). Furthermore, when the high volatility of variance is used in combination with the strong mean-reversion, the changes in the variance contributable to the drift from one observation to the next will likely be quite large, ultimately leading to larger variations in the estimated dynamics. Another related factor is the number of independent price paths used. Clearly, the number of price paths (500) is relatively low with respect to comparable simulation studies (often 1 000), so combined with particularly high volatility of variance, the estimated dynamics will have a poorer performance than those estimated in the presence of a lower ω. Not surprising as there are the same number of observations distributed over a larger interval.

Apart from the impact the aforementioned factors have on the drift and diffusion estimates of the Heston model, there are additional considerations. Despite the high amount of variation in the estimates, the smoothing parameter h seems to undersmooth the estimates for the drift. A large standard deviation in the 'observed' variances should produce a larger smoothing parameter, which in turn would lead to oversmoothing rather than undersmoothing. Nevertheless, the RMSE for the diffusion is not quite as unreasonable as for the drift estimates, and the estimate seems to have a good amount of smoothing to approximate its underlying functional form (as can be seen in e.g. the composite graph for the diffusion). Unfortunately, even the smoother estimates of the diffusion appear

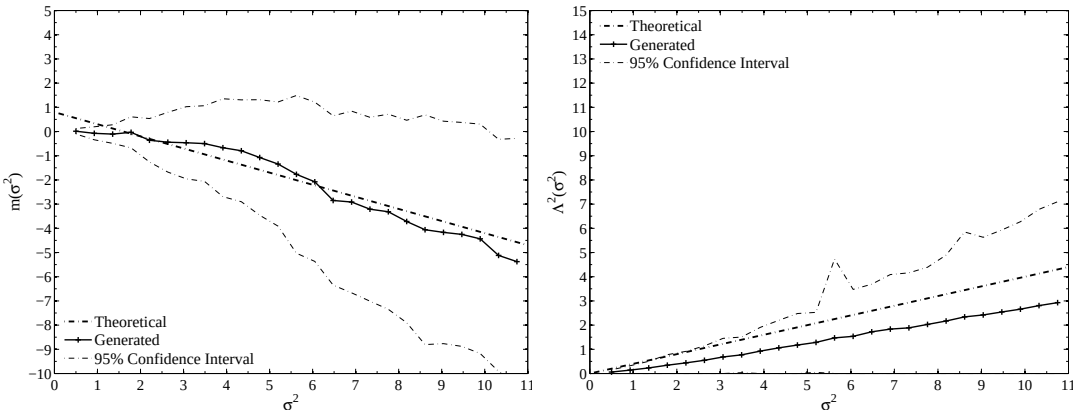

Figure 5.1: Average estimates of the drift function $m(\cdot)$ (left panel) and diffusion function $\Lambda^2(\cdot)$ (right panel) using generated daily variance at 1-second sampling. Daily variance σ^2 has been multiplied by 10 000. The variance series is generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift and diffusion functions on the domain of the graph.

to struggle to capture the underlying functional form, implying a slight curvature more resembling of the GARCH model. An interesting note is that the general functional shape of the diffusion estimate for the presented Heston model coincides with the preliminary graphical results for the Heston model of Renò (2006), despite using different parameter sets.

Having noted the excessive RMSE of the results for the Heston model, from here on the results for the GARCH model will be leading in the discussion on the relative performances, as it yields less noisy information. Nevertheless, the relative performances of the different daily variance estimators can generally still be made up from the results of the Heston model and thus should not be dismissed.

Returning to the relative performance of the estimators of integrated variance, recall that generally the smallest RMSE are found at the highest sampling frequency for the estimation of the daily variance. It appears this is generally not the case when estimating the true (theoretical) dynamics. What is seen in the tables for both the drift and diffusion estimates is that relatively lower frequencies dominate the estimation of the true dynamics. To bar thoughts of a possible inconsistency in the implementation, the results appear to agree with the graphical results for the GARCH model of Renò (2006), which are replicated in figure A.5 in Appendix A. For the replication the realized variance estimates of the daily

variance at a sampling frequency of 3 minutes (i.e. 180 seconds) are used, which lie just under the 200-second frequency employed in the cited paper.

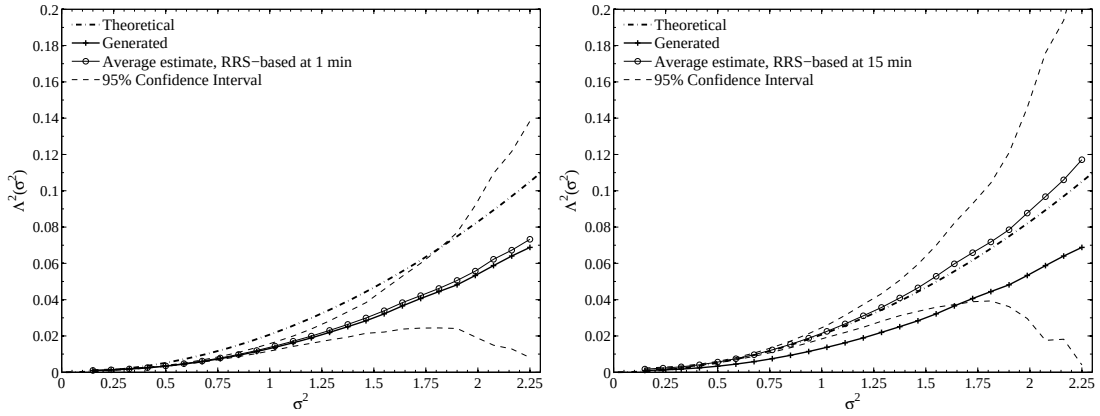

Figure 5.2: Average estimates of the diffusion function $\Lambda^2(\cdot)$ using RR_S estimates of daily variance with 1-minute sampling (left panel) and 15-minute sampling (right panel). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the diffusion estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true diffusion function on the domain of the graph.

With the consistency of the implementation confirmed, the dominance of the lower frequencies over the higher for estimating the true dynamics can be illustrated particularly well with the estimates of the diffusion as shown by figure 5.2. Showing results for the GARCH model, it gives a clear indication that the scaled realized range estimator at the 1-minute frequency gives a close estimate of the diffusion implied by the generated series, while at the 15-minute frequency a close approximation can be found for the true underlying diffusion. Thus the approximations of the mean daily variance, that are best estimated at the 1 minute frequency for most of the estimators, neatly translate to the best estimates for the dynamics as implied by the generated variance series.

To further confirm the translation of the results for the estimates of the daily variance, compare the results for the scaled realized range to the graphs in figure 5.3 for the TTS-based estimates, with respective best frequencies at 1-minute and 4-minute sampling. The graphs for the drift estimates for the scaled realized range and TTS estimators at the selected frequencies are included in appendix A as figures A.6 and A.7 and show the same results.

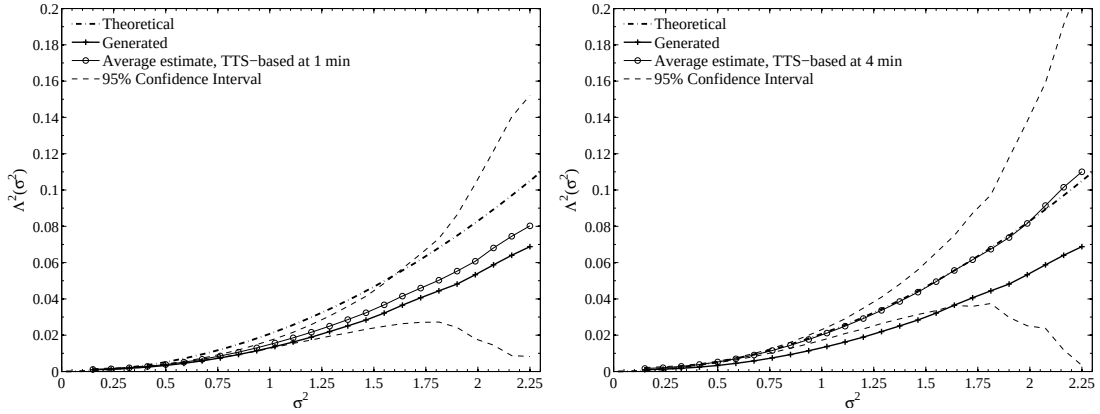

Figure 5.3: Average estimates of the diffusion function $\Lambda^2(\cdot)$ using RV_{TTS} estimates of daily variance with 1-minute sampling (left panel) and 4-minute sampling (right panel). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the diffusion estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true diffusion function on the domain of the graph.

Thus relative performance of the estimators of daily variance is also portrayed by these graphs and tables for the estimation of the drift and diffusion. The higher efficiency of the realized range estimators relative to the RV-based estimators can be recognized in the estimates for the drift and especially the diffusion. The lowest bias and RMSE for the range-based estimators are in the region of the 10 minutes frequency for the GARCH model, especially the diffusion estimates indicate a higher efficiency at the 15-minute level for both the realized range and its scaled variant. The higher overall efficiency is quite nicely illustrated by the composite graphs for the GARCH model (figures A.1 through A.4), where clearly the set of graphs of the scaled realized range are considerably more compact than the TTS-based graphs.

What causes this result is the nature of the two estimation steps of the methodology. On the one hand, the drift and diffusion estimates based on the generated variance series tend to underestimate the true dynamics (figures A.8 and 5.1). For the drift estimates this entails underestimating the level of mean-reversion, while for the diffusion estimates it means a lower estimated variance (GARCH) or volatility (Heston) of the daily variance. On the other hand, the noisy estimators of daily variance generally overestimate the dynamics of the generated variance, as can be seen for the GARCH model in the composite graphs A.1 through A.4 for the scaled realized range and TTS estimators. The best estimates found at the lower frequencies with respect to the true dynamics are such that the upward bias in

estimating the generated daily variance dynamics offset the downward bias of the dynamics of the generated daily variance. These findings are in agreement with those described by Kanaya and Kristensen (2009).

In this regard it should be interesting to pay attention to the performance of the realized range. It shows particularly interesting results, because when keeping in mind that the estimates for the mean daily variance at the higher frequencies carry relatively high RMSE values, the estimated variance series performs remarkably well for the estimation of the dynamics. The drift and diffusion estimates for the GARCH model using the realized range and scaled realized range are quite close with a slightly higher RMSE, while for the daily variance estimates they were rather different.

This basic scenario has brought a number of interesting observations. The performance of the estimates of integrated variance in the setting of stochastic volatility is in line with those in the setting of constant volatility (in the absence of noise) presented in Martens and Van Dijk (2007). Furthermore, the chosen parameter set (and possibly implementation) of the Heston model has undermined the reliability of those results for making inferences with respect to the (asymptotic) performance of the moment estimators. However, the results for the Heston model still generally preserve the relative performances of the different estimated series of the daily variance seen in the results of the GARCH model. From which it is clear that the best performances are by the scaled realized range and the TTS estimators for the estimation of the daily variance as well as for the estimation of the drift and diffusion functions. For the evaluation of the moment estimators, the better performance of lower frequencies to estimate the true dynamics underlying the data marks the reliance on offsetting biases to do so. This will make the estimation in the presence of market microstructure noise the more interesting as the biases in the estimation of the daily variance series should be affected.

5.2 Bid-ask bounce

As the first extension on the basic simulation of the previous section bid-ask bounce is added to the simulated price paths. By adding or subtracting half the spread (0.5 x 0.05% of the initial price) as described in section 4.2, estimation of the variances should become noisier. Now, apart from not observing the variance, the true price is no longer observed either, making the estimation of the daily variance the more difficult.

To see the effects, consider the results for the estimation of the long-term mean daily variance presented in tables B.1 and B.4. Clearly, the realized range and realized variance estimators are severely affected by the presence of bid-ask bounce. Both horribly overestimate the mean, with worsening performances as the sampling frequency increases, with the realized range affected even more than the realized variance. Having encountered the same, Martens and Van Dijk (2007) explain these observations by stating that in the limit the realized range will overestimate the true range by exactly the spread, as the maximum observed price on an interval will be an ask price, while the minimum price will be a bid price. The bias found in the realized variance estimates is produced by the fact that the returns are computed using the squared difference between ask-ask, bid-bid, bid-ask, or ask-bid prices. So in half the cases (the last two) the return is overestimated by the full spread, causing the estimated variance to contain the squared spread, thus causing the upward bias.

In terms of RMSE, the realized variance outperforms the realized range for all frequencies except the daily frequency in the GARCH results and the 5-minute frequency for the Heston model. For the results of the Heston model, the realized range dominates all estimators at frequencies of 10 minutes and lower when it comes to efficiency. Meanwhile, the scaled realized range has slightly poorer RMSE scores for the Heston estimates, but it clearly has a better grasp on the mean (generated) daily variance than the realized range. For the results of the GARCH model, the scaled realized range has the best RMSE at the lower frequencies and again severely improves on the performance of the realized range. The scaled realized variance also improves on its basic counterpart, scoring better RMSE than the realized variance at all frequencies for the GARCH model. The TTS estimator dominates both models at the higher frequencies for the Heston model and up to and including the 20-minute frequency for the GARCH model, yielding the estimates with the highest efficiency.

The results for the estimators of daily variance don't fully coincide with those found by Martens and Van Dijk (2007) in that the scaled realized range performs relatively poorly here, while it is the most efficient estimator in their simulation experiment. The likely cause of the diminished performance is the selected number of lags q used to construct the scaled estimators. The lower number of lags used here are selected based on their results for the setting with infrequent trading in Martens and Van Dijk (2007). While the selection of the lower number of lags is defendable for that setting, the marginal contribution of more

lags to the accuracy of the estimators in the setting of bid-ask bounce could very well lead to requiring more lags to obtain the same levels of accuracy.

Next, the estimated variance series contaminated with bid-ask bounce are used to estimate the drift and diffusion of the GARCH and Heston models. The results are presented in tables B.2 and B.3 for the GARCH model, and B.5 and B.6 for the Heston model in Appendix B. What remains apparent in the results for the GARCH model is the relative efficiency advantage of the range-based estimators at the lower sampling frequencies, where the estimators are less affected by market microstructure noise. As implied by the estimates of the mean daily variance, the scaled realized range outperforms the realized range in the GARCH estimates. This is vice versa for the diffusion estimates of the Heston model and in terms of RMSE for the drift estimates. For both models, the TTS-estimators has the least amount of bias at the higher frequencies, where for the GARCH model, it also has the lowest RMSE.

However, the presence of bid-ask bounce has an unexpected effect on the estimation of the diffusion using the range-based estimators. Consider the graphs in figure B.1 to illustrate what the tabular results imply. On the interval under consideration, the best sampling frequencies for the drift estimates are at 1 and 15 minutes for the scaled realized range, the same as in the scenario without noise. The diffusion estimates show a different story, as the 1-minute based estimates strongly underestimate the generated dynamics, which are now best estimated by using a 15-minute sampling frequency. The 15-minute frequency previously provided the closest estimate of the true dynamics, but is now best described by the 30-minute estimate. The results for the Heston model are similar albeit at a lower level of clarity, with best estimates not as easily selected for the true dynamics. The figure B.3 shows the results at 1, 10, and 20-minute sampling. The 20 minute frequency is selected as a midpoint between the optimal 15-minute frequency implied by the tabular drift-estimates (lowest bias) and the 30-minute sampling frequency for the diffusion estimates (lowest bias).

The results for the TTS-estimator are more encouraging as the best frequencies for estimating the generated and true dynamics remain at 1 and 4 minute sampling. The tabular results and the graphs in figure B.2 show almost identical results for the GARCH model. Moreover, the TTS estimator has the same or lower bias and RMSE in this setting with bid-ask bounce than in the case without noise, despite slightly poorer performance of

the daily variance estimates. The results for the Heston model also show stable estimates in the presence of bid-ask bounce, as the best frequencies are only slightly influenced. Here there is a slight discrepancy between the best drift estimate and the best diffusion estimate of the true dynamics, being at 3 and 5 minutes respectively. In figure B.4, the 4-minute frequency is shown as an indication at what can only be a slight improvement offered by the best frequencies for the Heston model.

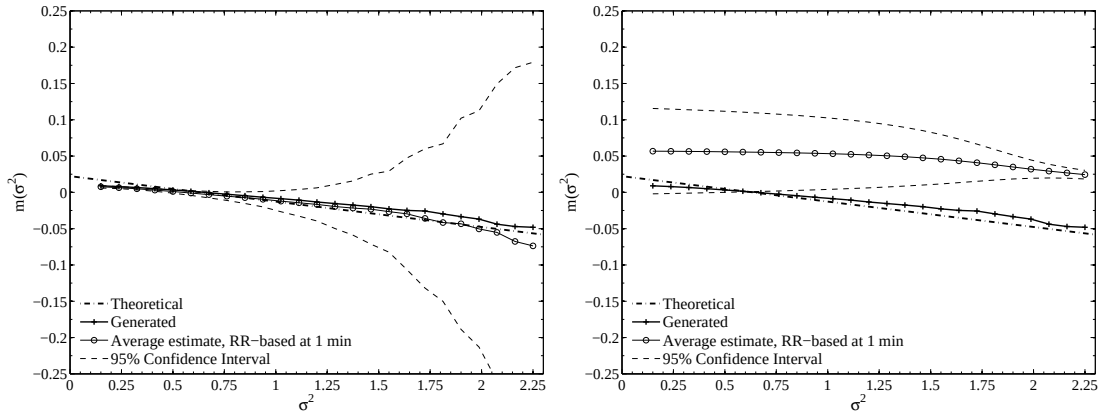

Figure 5.4: Average estimates of the drift function $m(\cdot)$ using the realized range estimates of daily variance at 1-minute sampling without noise (left panel) and with bid-ask bounce (right panel). Daily variance σ^2 has been multiplied by 10 000. The data series is generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift and diffusion functions on the domain of the graph.

After the remarkably good estimates for dynamics with the realized range despite the strong underestimation of the mean daily variance without any noise present, the results are to a degree different in the presence of bid-ask bounce. The strong upward bias in the daily variance estimates are quite noticeable in the drift estimates in figure 5.4 for the GARCH model. On the one hand, the drift is overestimated in the scenario with bid-ask bounce (in the right panel) and on the other hand its oversmoothed due to excessive variance in the daily variance estimates. The oversmoothing is apparent from the horizontal nature of the graph indicating that local characteristics are completely smoothed away. Nevertheless, as is apparent in the tables, the performance in the estimation of the diffusion is very similar to that of the scaled realized range.

As a final observation, recall that the realized variance has a lower upward bias than the realized range at the higher frequencies for the estimation of the mean daily variance. The opposite is true for the estimation of the true dynamics at those frequencies as the

realized variance has a considerably poorer performance for the estimation of both the drift and diffusion. This discrepancy illustrates the main differences between the two types of estimators, as the realized range will generally have a lower variability of the daily variance estimates, but on average it tends to miss the mark. The realized variance has the opposite characteristics, yielding a good estimate of the mean but at the cost of higher variance of the estimates.

The presence of bid-ask bounce clearly cripples the realized range and realized variance estimators. Bias-correcting using scaling partially achieves the desired effect, however the TTS estimator appears to deal with the presence of bid-ask bounce the best in the setting of estimating the mean daily variance as well as the underlying dynamics. Best performances are again by the scaled realized range and the TTS estimator. Where the scaled realized range performs best at lower frequencies and the TTS estimator has the advantage in the higher frequencies. An important observation here is the impact the bid-ask bounce has on the variance of the daily variance estimates, particularly noticeable in the results of the realized range and realized variance, and by extension on the smoothing parameter.

5.3 Infrequent trading

Having seen the devastating effects of bid-ask bounce on the realized range and realized variance estimators, the impact of infrequent trading is discussed in this section. As previously stated, the presence of infrequent trading is implemented by filtering the fully observed price path by only observing one price every ten seconds. So effectively there is less information on the real price movement on the daily interval.

Tables C.1 and C.4 in Appendix C show the results for estimating the daily variance. The tables show results that, as suggested in section 5.1, are more pronounced than the case without noise. With even less information on the price movement, the performance of the realized range takes a hit. This is quite visible in the results as its downward bias is even bigger than in section 5.1. Similar to what is seen in the results of Martens and Van Dijk (2007), the RMSE first decreases as the sampling frequency increases and then RMSE starts increasing again. As Martens and Van Dijk (2007) state, this is due to the fact that at first the lower variance of the estimates has a stronger effect than the underestimation caused by the infrequent trading. The point at which the underestimation starts dominating is clearly dependent on the trading intensity. This is illustrated by considering that in the

case without noise, the most efficient frequency is at 10-minutes while in this scenario with infrequent trading this is at 30-minutes.

Scaling the realized range removes most of the downward bias and strongly improves on the RMSE; however some bias still remains due to the fact that the daily range is also somewhat affected by infrequent trading (Martens and Van Dijk (2007)). The RMSE for the scaled range is only slightly higher than that of the realized variance and TTS estimator at the 1 minute frequency (and also at 2 and 3 minutes for the Heston model). It has the best RMSE for the rest of the sampling frequencies and displays the efficiency advantage over the realized variance estimators.

The realized variance outperforms the scaled realized variance, but both are largely unaffected by the effects of infrequent trading and only show slightly higher RMSE than in the scenario without noise. As expected, these results are in concurrence with those of Martens and Van Dijk (2007).

Estimating the variance dynamics using the estimated daily variance series provides the results in tables C.2 and C.3 for the GARCH model, and tables C.5 and C.6 for the Heston model, respectively. Similar to the previous two sections, the relative efficiency of the range-based estimators at the lower frequencies is apparent in both the estimates of daily variance and the estimates of the dynamics. With the better performance of the range-based estimators coming at the 10 minute mark for the GARCH model and generally lower frequencies for the Heston model, the realized variance-based estimators are most effective at the high frequencies.

As is clear from the tables and illustrated by figure C.1, the best frequency for estimating the true drift and diffusion of the GARCH model with the range-based estimators is at 10-minute sampling. For estimating the generated dynamics then the 1-minute frequency remains the best choice. While the true dynamics was best estimated by using 15-minute sampling in the case without noise, the difference between that frequency and the 10-minute frequency in the presence of infrequent trading is fairly small. The two frequencies are the best in terms of bias for the scaled realized range. For the results of the Heston model, the scaled range finds its best fit using the 1-minute frequency for the true drift and 15-minute frequency for the true diffusion (displayed in fig C.3). Note however that

the scaled realized range provides the lower biases, but competes directly with the realized range for the lower RMSE.

What is interesting to see in the results for the higher frequencies of the GARCH model, is that the best performance is not solely provided by the TTS estimator. The realized variance and its scaled version have relatively lower or similar biases and RMSE at first three frequencies. Nevertheless, the TTS-estimator once again finds the best performance at the 4-minute frequency (in terms of bias) yielding an almost identical behavior to the case without noise and lower bias than the best frequencies of the realized variance and scaled realized variance. Its estimates are graphed in figure C.2 and C.4 for the GARCH and Heston models respectively and again hardly differ from the previous two sections for the GARCH model. For the results of the Heston model, the graphs show the 4-minute frequency for a situation similar to the previous section with bid-ask bounce. From the tabular results, the 3 and 5 minute frequencies offer smaller biases for the drift and diffusion respectively, but as is clear from the tight fit of the graphs, the improvement is relatively small.

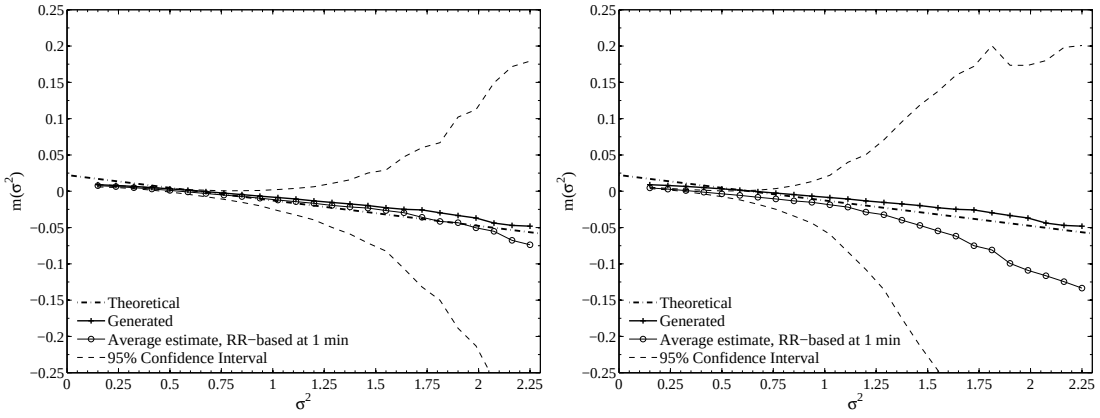

Figure 5.5: Average estimates of the drift function $m(\cdot)$ using the realized range estimates of daily variance at 1-minute sampling without noise (left panel) and with infrequent trading (right panel). Daily variance σ^2 has been multiplied by 10 000. The data series is generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift and diffusion functions on the domain of the graph.

To continue the surveillance of the realized range's performance and to illustrate the effects of infrequent trading on the unadjusted estimator, from the tables it is clear that the realized range finds its best performance at the same frequency as the scaled variant.

However, its bias is quite larger and the estimate is noisier. Nevertheless, the realized range offers a slightly higher bias at the 1-minute frequency than the scaled realized range with respect to the generated diffusion and a similar result for the drift. With respect to the results of the scenario without noise, the realized range offers an only slightly downward biased albeit much noisier estimate of the drift as illustrated by the figure 5.5, despite the considerable downward bias of the daily variance estimates.

To recap the observations, the presence of infrequent trading has brought the accentuated effects predicted in section 5.1. The decreased amount of information caused by a lower trading intensity affects the realized range-based estimators the most, exacerbating the downward bias of the estimator in such a way that it is even noticeable in the estimate of the drift. As expected, that influence on the realized variance-based estimators is minimal, as less information in an interval is not as important as the begin and endpoint thereof for creating these estimates. The only effect noticeable there is slightly noisier estimates for the realized variance and scaled realized variance. The TTS estimator no longer dominates the performances at the highest frequencies, but still offers the most stable estimates at higher sampling intensity.

5.4 Bid-ask bounce and infrequent trading

After studying the impact of bid-ask bounce and infrequent trading independently, the simulated paths are now subjected to the combined effects of the two noise-types. After adjusting the fully observed price paths to reflect bid and ask prices, the adjusted path is filtered such that only one price every 10 seconds is observed.

The combined impact of the noise types can be seen in the results in tables D.1 and D.4 in appendix D for the GARCH and Heston models respectively. As in Martens and Van Dijk (2007) it is apparent for the realized range based estimators that the downward bias of the infrequent trading and upward bias of the bid-ask bounce partially offset each other. This is especially the case for the 1-minute frequency, such that the realized variance is outperformed by the realized range. Clearly, for the rest of the frequencies, the upward bias caused by the bid-ask bounce is far more pronounced than the effects of infrequent trading.

As is to be expected after the results in the previous sections, the scaled realized range performs better than the unadjusted version. Nevertheless, as previously stated the

presence of bid-ask bounce combined with the number of lags used for scaling wrecks havoc on the performance of this estimator, despite its improvement over the realized range. Due to the smaller number of observations, the estimates of the scaled realized range are even noisier than those in case with only bid-ask bounce. The scaled realized range becomes efficient off all estimators when using 30 minute sampling (GARCH) and 10-minute sampling (Heston) and lower just as when considering bid-ask bounce alone. Furthermore, for the GARCH model the scaled realized range has an upward bias as the offsetting biases are dominated by the upward bias of the bid-ask bounce. For the Heston model, notice that the scaled range underestimated the mean in the presence of both bid-ask bounce infrequent trading, causing the results here to reflect a downward bias in the estimate.

Scaling the realized variance also proves successful as it outperforms its unadjusted counterpart at all frequencies for the GARCH model and while for the Heston model the scaled realized variance fails to remove the biases (with the exception of the 1-minute frequency). The reason for this remains unclear, but it is consistent with the performance of the scaled realized variance over all the results for the Heston model, as its efficacy is lacking there as well.

Given the performance of the TTS estimator in the last the rest of the results, it performs well again in this setting of combined noise as expected. It gives the best performance of all the estimators at the higher frequencies, yielding good approximations of the mean at the best efficiencies. The TTS estimator is the most efficient at all the frequencies up to and including 20-minute sampling in the GARCH model. For the Heston model, it slightly overestimates the mean and yields the best efficiencies up to and including the 5-minute frequency. As seen in previous results, at the 10-minute frequency the scaled realized range is more efficient, but the TTS-estimator yields the better approximation of the mean.

The performance measurements for the moment estimators are given in tables D.1 and D.4 in Appendix D. Judging from the tables, the same general assessment can be made as before that the realized range-based estimators have the best performances at the lower frequencies and the realized variance-based estimators own the higher frequencies for the estimation of both the drift and diffusion.

As usual the scaled realized range yields the lower biases as compared to its unadjusted counterpart. An exception to this is found at the 30 and 60 minutes frequencies in the drift and diffusion estimates of both models, where the realize range yields the lower RMSE and sometimes also the lower bias. With respect to the realized variance estimates, the range-based estimates yield the higher biases in estimating the true dynamics. The downward bias in the daily variance estimates of the scaled range directly translates in a strongly underestimated diffusion estimate. Within the interval of evaluation, the drift estimate seems to perform well enough, but trails off toward the higher daily variance values, as is apparent in the figure D.1. Furthermore, the same situation as for the bid-ask bounce presents itself here, where different frequencies yields best estimates for the drift and diffusion. As illustrated by the graphs and made explicit in the tables, the true drift is best estimated using the 10-minute frequency, which also yields the best estimate for the generated variance series. The true diffusion is best estimated by the 30-minute frequency, but yields a poor drift estimate.

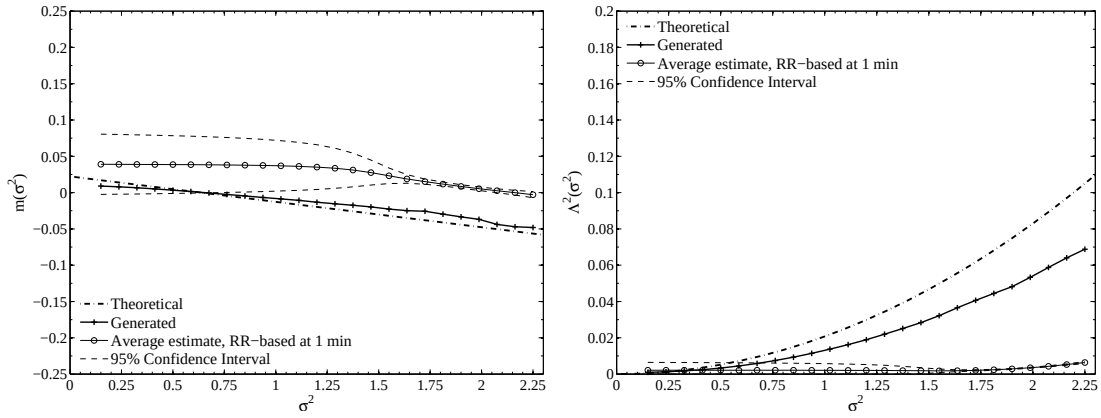

Figure 5.6: Average estimates of the drift function $m(\cdot)$ (left panel) and the diffusion function $\Lambda^2(\cdot)$ (right panel) using the realized range estimate of daily variance at 1-minute sampling in the presence of infrequent trading and bid-ask bounce. Daily variance σ^2 has been multiplied by 10 000. The data series is generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift and diffusion functions on the domain of the graph.

With the focus on the realized range estimators, consider the results of the previous sections. In the case of infrequent trading, the realized range underestimated the drift and yielded a fair diffusion estimate. In the case of bid-ask bounce, the realized range strongly overestimated the drift and was oversmoothed, yet it yielded a good diffusion estimate. In the case where both types of noise are present as figure 5.6 shows, the realized range

produces poor estimates for both the drift and diffusion. Plagued by underestimation of the diffusion and strong upward bias in the drift combined with oversmoothing, the realized range provides a clear picture of the impact of market microstructure noise on the estimation of the dynamics.

The performance of the TTS estimator has proven to be stable in the previous sections and continues to generally outperform the other realized variance estimators. Its performance with respect to the true dynamics is better than that of the realized range-based estimators at the higher frequencies. As shown by the graphical results in figure D.2 for the GARCH model, while the TTS estimator still does quite well in describing the true dynamics, now best at the 3-minute frequency with the 4-minute frequency a close second. What is quite apparent is the poor performance of the TTS estimator at the 1-minute frequency when trying to approximate the generated dynamics. The 1-minute frequency has become a contender for the estimation of the true drift and diffusion. The results for the Heston model are more encouraging in that sense, as those are reminiscent of the results of the previous sections, as is illustrated by figure D.4.

Wrapping up the discussion, the combined noise types has brought some interesting observations. The realized range is, as expected, absolutely useless when left unadjusted in the presence of market microstructure noise, which is of course not surprising. Its sensitivity to these effects is well-known, but has served the purpose to illustrate the effects of the noise types on the estimation of the variance dynamics. The scaled realized range requires careful selection of the number of lags used to construct it, as the bid-ask bounce appears to require more lags to eliminate its effects. While improving on the realized range, it too succumbs to a certain extent to the effects of infrequent trading and bid-ask bounce. The TTS estimator provides a stable estimate of the true dynamics, despite the presence of both noise types, yet loses touch with the generated dynamics it strictly speaking estimates.

6 Conclusions and further research

In this thesis the effects of market microstructure noise on the nonparametric estimation of the stochastic volatility models have been studied. The procedure is a two-step method to estimate the underlying variance dynamics of price paths simulated with a GARCH model and the Heston model. First, estimates of the latent daily variance were constructed by means of nonparametric estimators based on high frequency intra-daily prices, therein extending the work of Martens and Van Dijk (2007) into the realm of stochastic volatility. Using the series of daily variance estimates, kernel-based infinitesimal moment estimators are utilized to estimate the drift and diffusion of the variance process underlying the generated prices. Various unadjusted and bias-corrected estimators of the daily variance were implemented to discuss the impact of infrequent trading and bid-ask bounce on the estimation of the daily variance in the setting of stochastic volatility and ultimately the estimation of the variance dynamics.

To start off, some comments need to be made on the setup of the simulation experiment. The number of prices per second could ideally have been set higher to get a better feel of continuous trading, such as implemented in Martens and Van Dijk (2007), however had to be decreased to limit the computational load. A further decrease in the number of price paths simulated per model to reduce the computational load caused some results to remain unclear. Especially when combined with the parameter choices for the Heston model this proved to be somewhat unfortunate for that set of results, undermining the usefulness thereof. Due to the high variance in the results of the Heston model, convergence of the estimates remained somewhat elusive, limiting the possibility of strong conclusions and therefore forcing a downgrade of the set to a supportive role to the GARCH results. However, the results in the Heston model seem to globally agree with the GARCH models results. Additionally it should be included that, as Martens and Van Dijk (2007) indicate, there is plenty of evidence against assuming that the market microstructure noise is temporally independent and independent of the price process. Nevertheless, the simple implementation of the noise types serve its purpose for this thesis.

In general, the results of Martens and Van Dijk (2007) in the setting of constant volatility extend pretty well to the setting of stochastic volatility. Bias-correcting the realized range and realized variance estimators in the presence of infrequent trading and bid-ask bounce proves to be essential for these estimators to be useful. To that effect, the estimates of

the daily variance are quite clear as to what proves to be an effective method. Scaling the realized range is quite effective, but requires more sophisticated selection of the number of lags to be included in the scaling factor. The selected number of lags proved effective for removing most of the downward bias caused by infrequent trading, but not for removing enough of the upward bias caused by the bid-ask bounce. Ultimately, it is the two-time-scales estimator of Zhang et al. (2005) that proves most stable across the scenarios in estimation of the daily variance, noticing only slight increases in the RMSE in the presence of infrequent trading and/or bid-ask bounce. Generally, the RV-based estimators are best at the higher frequencies, but the realized range-based estimators have better performances at the lower frequencies.

The results for the estimation of the drift and diffusion of the variance process show that generally the performance of the estimators for approximating the daily variance is not a strong indicator of how the performance will be when estimating the dynamics (as illustrated by the performance of the realized range in the case without noise). However there are some signs, as the overall efficiency of the range-based estimators is generally higher than that of the realized variance, which shows in all results in such a way that the best estimates of the true dynamics are also found to be at the lower sampling frequencies for the (scaled) realized range and at the higher frequencies for the TTS estimator.

The results for the realized range shows the strong influence the noise types can have when estimating daily variance, and by extension its underlying dynamics. In regard of the drift and diffusion estimates, these results highlight the sensitivity of the overall results to the bandwidth selection. Prime examples are the drift and diffusion estimates when the variance of the estimates is high, resulting in a large bandwidth parameter under the current bandwidth selection procedure and thus to oversmoothing of the estimate. More sophisticated procedures in selecting the bandwidth could perhaps lead to results more robust to excessive variance in the estimates.

As a further remark towards the bandwidth selection, the results of the Heston model imply that perhaps selecting separate bandwidths for the drift and diffusion could lead to better estimates of the dynamics. With a larger smoothing parameter, the drift estimates might have been better, as now they seem to imply that the procedure fails to capture the true form of the dynamics. Further research is warranted to study the effects hereof. There

are some studies into alternatives to the bandwidth selection procedure such as the cross-validation procedure implemented by Kanaya and Kristensen (2009) or the automated bandwidth procedure suggested by Bandi et al. (2009). While the implication that the methodology appears to have trouble capturing the drift dynamics of the Heston model could simply mean that the bandwidth selection was dissatisfactory, the diffusion estimates might not be as easily improved. Therefore, an implementation on more complicated models than the Heston and GARCH models could shed more light on the performance depth of the infinitesimal moment estimators.

On that note, an additional observation came up when bypassing the first step of the estimation procedure by using the actual generated variance series, which would otherwise be estimated by the spot variance estimators. Implementing the moment estimators using the otherwise unobservable variance series, lead to underestimated dynamics in both models. It is not directly clear what causes this downward bias in the estimates, which preliminary evidence suggests is around 0.63 for the diffusion of both models. Regardless, consider that the daily variance estimators can at best estimate the daily variance dynamics implied by the generated data. Then it should be noted that the method appears to hinge on the fact that all best estimates of the true dynamics are found by the mere virtue that they are noisy estimates of the generated variance, thus relying on offsetting biases to make inferences on the true dynamics.

Nevertheless, what stands clear among the results is that strong performance of the TTS estimator in estimating the true dynamics of primarily of both models. Although the TTS-estimator seems to suffer from the same problems when attempting to capture the functional form of the dynamics of the Heston model, it still does so without much influence by the presence of the implemented noise types. Throughout the sections presented here, the best sampling frequency for estimating the true drift and diffusion has been with (approximately) 4-minute sampling. Therefore, this implies that controlling for market microstructure noise by means of robust bias-corrected estimators of the spot variance is sufficient to successfully implement the infinitesimal moment estimators. A logical next step in subsequent research is developing a way to determine the optimal sampling frequency for the use in the methodology presented here.

References

Alexander, C., Lazar, E., November 2005. On the continuous limit of garch. ICMA Centre Discussion Papers in Finance 13.

Andersen, T. G., Bollerslev, T., November 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39 (4), 885–905.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., Diebold, F. X., March 2005. Volatility forecasting. Prepared for the forthcoming Handbook of Economic Forecasting.

Andersen, T. G., Bollerslev, T., Diebold, F. X., July 2002. Parametric and nonparametric volatility measurement. Prepared for the forthcoming book Handbook of Financial Econometrics (L.P. Hansen and Y. Ait-Sahalia, Eds.).

Bandi, F. M., Corradi, V., Moloche, G., October 2009. Bandwidth selection for continuous-time markov processes. Working Paper.

Bandi, F. M., Phillips, P. C. B., May 2002. Nonstationary continuous-time models. Prepared for the forthcoming book Handbook of Financial Econometrics (L.P. Hansen and Y. Ait-Sahalia, Eds.).

Bandi, F. M., Renò, R., Mar. 2009. Nonparametric stochastic volatility. Global COE Hi-Stat Discussion Paper Series gd08-035, Institute of Economic Research, Hitotsubashi University.

Christensen, K., Podolskij, M., 2007. Realized range-based estimation of integrated variance. Journal of Econometrics 141, 323–349.

Drost, F. C., Werker, B. J. M., September 1996. Closing the garch gap: Continuous time garch modeling. Journal of Econometrics 74 (1), 31–57.

Ghysels, E., Harvey, A. C., Renault, E., 1996. Stochastic volatility. Prepared for the Handbook of Statistics (Vol.14: "Statistical Methods in Finance").

Hull, J. C., May 2008. Options, futures, and other derivatives (7th Edition), 7th Edition. Prentice Hall.

Kanaya, S., Kristensen, D., 2009. Estimation of stochastic volatility models by nonparametric filtering. Working Paper.

Lord, R., Koekkoek, R., Van Dijk, D. J. C., 2009. A comparison of biased simulation schemes for stochastic volatility models. Quantitative Finance, 1469–7696.

Marron, J. S., Wand, M. P., 1992. Exact mean integrated squared error. The Annals of Statistics 20 (2), 712–736.

Martens, M., Van Dijk, D. J. C., 2007. Measuring volatility with the realized range. Journal of Econometrics 138, 181–207.

McAleer, M., 2005. Automated inference and learning in modeling financial volatility. Econometric Theory 21 (13), 232261.

McAleer, M., Medeiros, M. C., 2008. Realized volatility: a review. Econometric Reviews 27 (13), 1045.

Renò, R., 2006. Nonparametric estimation of stochastic volatility models. Economics Letters 90 (3), 390–395.

Renò, R., 2008. Nonparametric estimation of the diffusion coefficient of stochastic volatility models. Econometric Theory 24 (5), 1174–1206.

Scott, D., 1992. Multivariate Density Estimation. John Wiley & Sons.

Shephard, N., Andersen, T. G., March 2008. Stochastic volatility: origins and overview. Prepared for the Handbook of Financial Time Series, Springer Verlag.

Tarter, M., March 1986. Mean integrated squared error sampling. Journal of the American Statistical Association 81 (393), 234–242.

Zhang, L., Mykland, P., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Appendix A Tables and graphs - no noise

This appendix contains results for the GARCH and Heston model in two separate parts that each hold three tables with the results for the estimation of the daily variance, the estimated drift, and the estimated diffusion for the simulation scenario without market microstructure noise. Following the three tables in each part are the graphs referenced to in the main body of this thesis.

Results for the GARCH model

Table A.1:
Estimators of daily variance using fully observed price paths without noise - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	0.5356	0.1199	0.6328	0.0232	0.6356	0.0283	0.6354	0.0489	0.6356	0.0236
2	0.5629	0.0885	0.6328	0.0266	0.6356	0.0400	0.6354	0.0563	0.6356	0.0330
3	0.5754	0.0748	0.6328	0.0297	0.6356	0.0490	0.6354	0.0631	0.6356	0.0403
4	0.5831	0.0673	0.6328	0.0324	0.6355	0.0565	0.6354	0.0691	0.6355	0.0464
5	0.5884	0.0627	0.6328	0.0348	0.6355	0.0632	0.6354	0.0746	0.6355	0.0518
10	0.6018	0.0563	0.6328	0.0452	0.6356	0.0894	0.6354	0.0978	0.6355	0.0732
15	0.6078	0.0586	0.6328	0.0535	0.6355	0.1094	0.6354	0.1165	0.6355	0.0897
20	0.6115	0.0630	0.6328	0.0607	0.6356	0.1264	0.6355	0.1326	0.6355	0.1037
30	0.6158	0.0728	0.6328	0.0730	0.6356	0.1549	0.6355	0.1601	0.6356	0.1274
60	0.6216	0.0992	0.6328	0.1011	0.6357	0.2191	0.6356	0.2228	0.6356	0.1820
1440	0.6330	0.4844	0.6330	0.4844	0.6355	1.0730	0.6355	1.0730	0.6355	1.0730

Note The table shows the results of a Monte Carlo simulation estimating daily variance using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the GARCH model. Sample average of the daily variance $\bar{V}_{\mathrm{Day},t} \approx 0.6356$ and $\theta = 0.636$. Daily variance has been multiplied by 10 000.

Table A.2:
Results for estimates of the drift for the case without noise - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	0.0159	0.0289	0.0162	0.0234	0.0117	0.0224	0.0123	0.0225	0.0134	0.0224
2	0.0144	0.0266	0.0149	0.0230	0.0063	0.0226	0.0068	0.0224	0.0096	0.0220
3	0.0133	0.0256	0.0136	0.0226	0.0037	0.0247	0.0037	0.0244	0.0062	0.0225
4	0.0121	0.0249	0.0124	0.0225	0.0075	0.0277	0.0067	0.0275	0.0039	0.0237
5	0.0110	0.0246	0.0112	0.0222	0.0126	0.0315	0.0118	0.0310	0.0048	0.0254
10	0.0068	0.0253	0.0058	0.0229	0.0396	0.0563	0.0391	0.0563	0.0221	0.0396
15	0.0070	0.0278	0.0044	0.0256	0.0655	0.0803	0.0652	0.0804	0.0403	0.0562
20	0.0108	0.0312	0.0085	0.0290	0.0911	0.1055	0.0903	0.1051	0.0582	0.0733
30	0.0218	0.0398	0.0200	0.0378	0.1383	0.1528	0.1374	0.1523	0.0924	0.1069
60	0.0539	0.0701	0.0525	0.0683	0.2638	0.2779	0.2614	0.2763	0.1863	0.2006

Note The table shows the results of a Monte Carlo simulation estimating daily drift using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The results use 500 independent samples of 5 000 estimates of daily variance generated using prices from a GARCH model. The bias when estimating the daily drift using the generated variance $V_{\text{Day},t}$ using all data is approx. 0.0176 with RMSE 0.0239. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.15, 1.25]. Daily variance has been multiplied by 10 000.

Table A.3:
Estimators of the daily diffusion for the case without noise - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	0.0212	0.0219	0.0214	0.0218	0.0154	0.0160	0.0153	0.0160	0.0178	0.0183
2	0.0195	0.0202	0.0197	0.0201	0.0075	0.0091	0.0075	0.0091	0.0124	0.0133
3	0.0177	0.0185	0.0179	0.0184	0.0020	0.0063	0.0020	0.0064	0.0071	0.0088
4	0.0161	0.0169	0.0163	0.0169	0.0089	0.0112	0.0089	0.0112	0.0024	0.0061
5	0.0144	0.0153	0.0145	0.0152	0.0168	0.0184	0.0167	0.0183	0.0043	0.0075
10	0.0060	0.0084	0.0062	0.0082	0.0555	0.0567	0.0556	0.0568	0.0306	0.0319
15	0.0030	0.0072	0.0030	0.0068	0.0940	0.0953	0.0939	0.0953	0.0570	0.0582
20	0.0109	0.0132	0.0108	0.0129	0.1305	0.1320	0.1305	0.1320	0.0831	0.0844
30	0.0274	0.0288	0.0272	0.0285	0.2014	0.2033	0.2017	0.2036	0.1345	0.1360
60	0.0757	0.0771	0.0756	0.0768	0.3926	0.3957	0.3935	0.3966	0.2799	0.2822

Note The table shows the results of a Monte Carlo simulation estimating daily diffusion using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the GARCH model. The bias when estimating the daily diffusion using the generated variance $V_{\text{Day},t}$ using all data is approx. 0.0235 with RMSE 0.0238. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.15, 1.25]. Daily variance has been multiplied by 10 000.

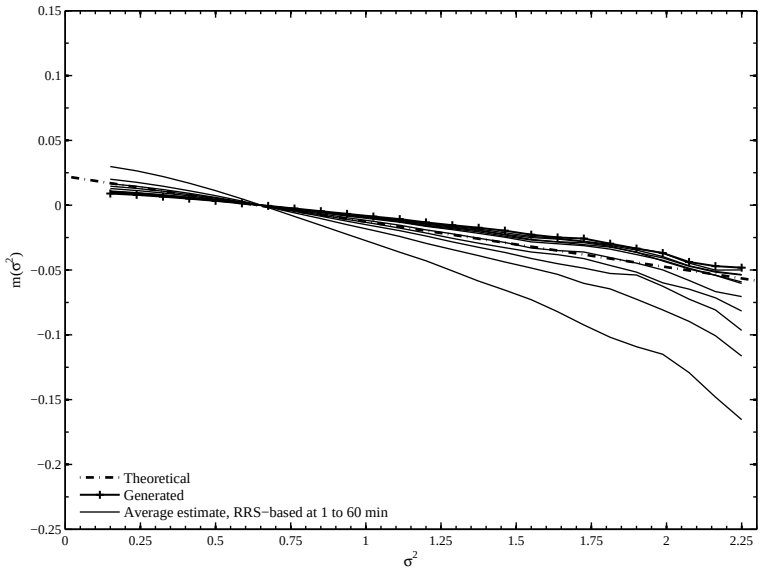

Figure A.1: Average estimates of the drift function $m(\cdot)$ using RR_S estimates of daily variance, where the 60 minute frequency gives the largest bias and the 1 minute frequency the lowest versus the generated dynamics. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

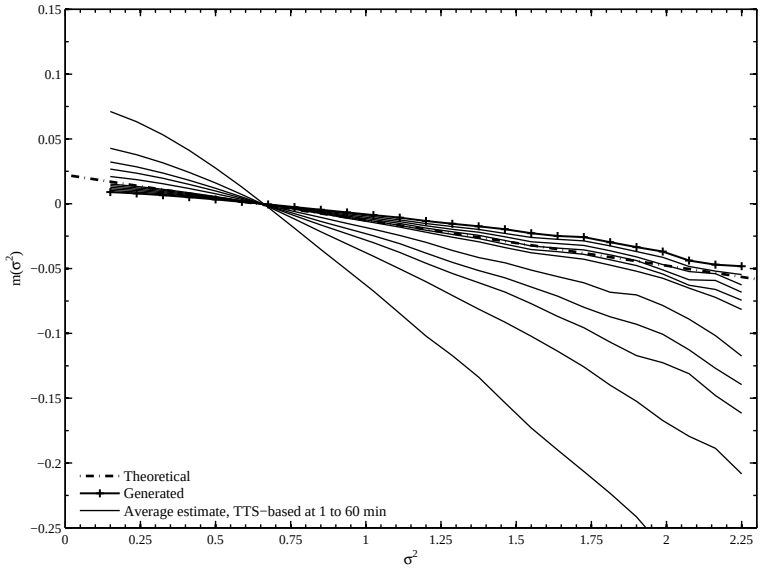

Figure A.2: Average estimates of the drift function $m(\cdot)$ using RV_{TTS} estimates of daily variance, where the 60 minute frequency gives the largest bias and the 1 minute frequency the lowest versus the generated dynamics. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

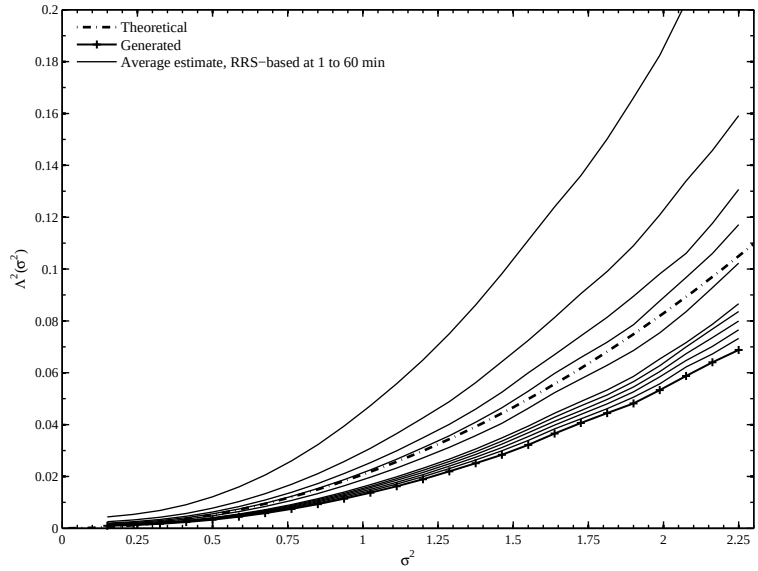

Figure A.3: Average estimates of the drift function $\Lambda^2(\cdot)$ using RR_S estimates of daily variance, where the 60 minute frequency gives the largest bias and the 1 minute frequency the lowest versus the generated dynamics. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the diffusion estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true diffusion function on the domain of the graph.

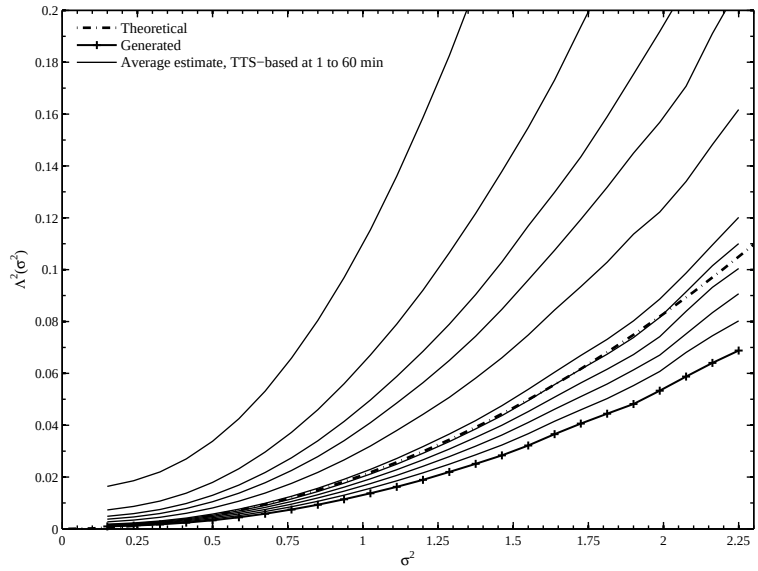

Figure A.4: Average estimates of the diffusion function $\Lambda^2(\cdot)$ using RV_{TTS} estimates of daily variance, where the 60 minute frequency gives the largest bias and the 1 minute frequency the lowest versus the generated dynamics. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the diffusion estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true diffusion function on the domain of the graph.

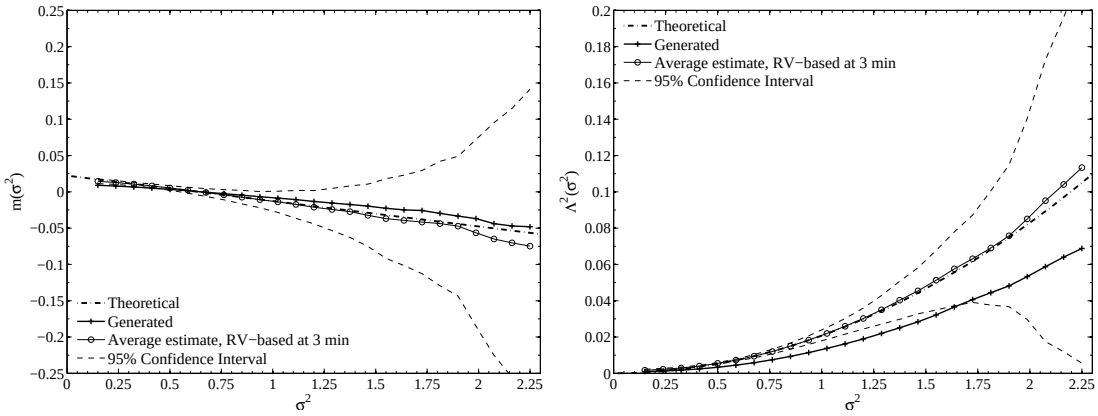

Figure A.5: Average estimates of the drift function $m(\cdot)$ (left panel) and diffusion function $\Lambda^2(\cdot)$ (right panel) using RV_{TTS} estimates of daily variance with 3-minute sampling, which is just under the 200 second-frequency of Renò (2006). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

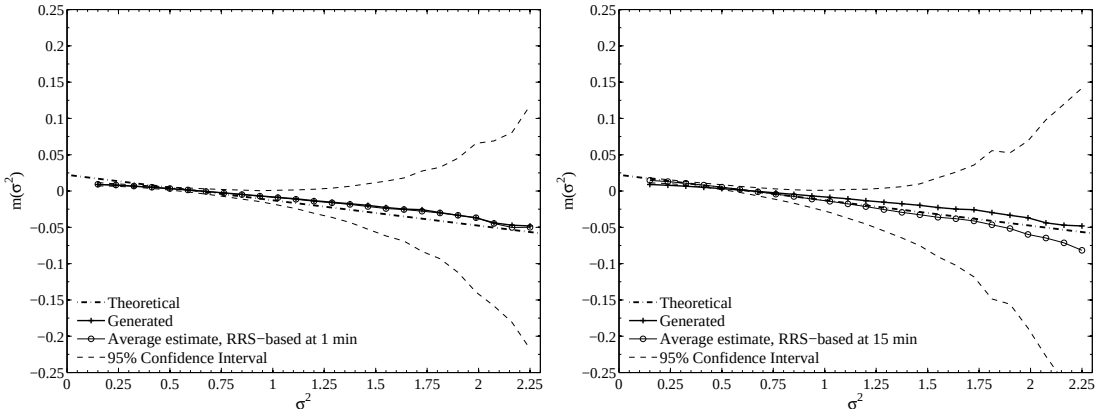

Figure A.6: Average estimates of the drift function $m(\cdot)$ using RR_S estimates of daily variance with 1-minute sampling (left panel) and 15-minute sampling (right panel). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

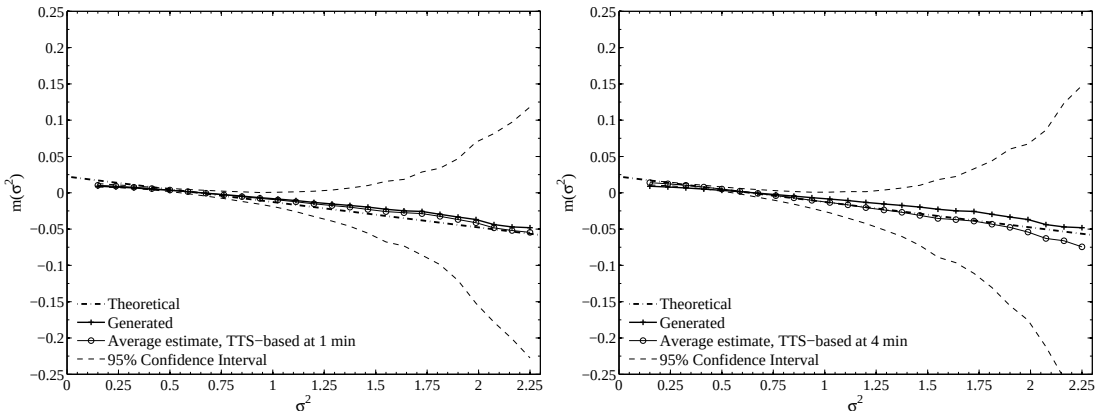

Figure A.7: Average estimates of the drift function $m(\cdot)$ using RV_{TTS} estimates of daily variance with 1-minute sampling (left panel) and 4-minute sampling (right panel). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

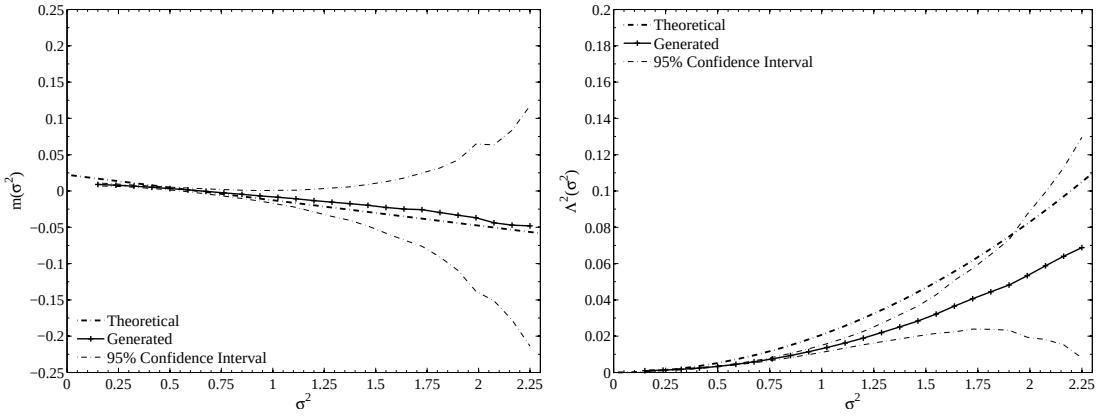

Figure A.8: Average estimates of the drift function $m(\cdot)$ (left panel) and diffusion function $\Lambda^2(\cdot)$ (right panel) using generated daily variance at 1-second sampling. Daily variance σ^2 has been multiplied by 10 000. The variance series is generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift and diffusion functions on the domain of the graph.

Results for the Heston model

Table A.4:
Estimators of daily variance using fully observed price paths without noise - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	1.4323	1.3120	1.6930	0.3698	1.6995	0.3075	1.7084	0.8018	1.6996	0.2574
2	1.5051	0.9690	1.6930	0.3965	1.6995	0.4345	1.7083	0.8580	1.6997	0.3585
3	1.5389	0.8190	1.6930	0.4215	1.6995	0.5295	1.7082	0.9139	1.6996	0.4369
4	1.5593	0.7377	1.6931	0.4429	1.6995	0.6128	1.7084	0.9628	1.6996	0.5038
5	1.5736	0.6865	1.6931	0.4661	1.7002	0.6870	1.7085	1.0172	1.6998	0.5634
10	1.6094	0.6152	1.6930	0.5606	1.7005	0.9786	1.7084	1.2337	1.7002	0.8005
15	1.6260	0.6384	1.6931	0.6435	1.7012	1.1942	1.7087	1.4129	1.7002	0.9830
20	1.6355	0.6875	1.6929	0.7142	1.6993	1.3841	1.7074	1.5763	1.7000	1.1352
30	1.6474	0.7944	1.6929	0.8406	1.7001	1.6953	1.7078	1.8607	1.6994	1.3916
60	1.6620	1.0831	1.6932	1.1344	1.6978	2.3842	1.7089	2.5212	1.6985	1.9853
1440	1.6886	5.2937	1.6886	5.2937	1.6984	11.7942	1.6984	11.7942	1.6984	11.7942

Note The table shows the results of a Monte Carlo simulation estimating daily variance using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the Heston model. Sample average of the daily variance $\bar{V}_{\mathrm{Day},t} \approx 1.6998$ and $\theta = 1.6$. Daily variance has been multiplied by 10 000.

Table A.5:
Results for estimates of the drift for the case without noise - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	1.0852	18.9038	1.3794	19.7258	1.3252	19.8646	1.2988	23.8826	1.4423	19.4277
2	1.1893	18.9939	1.4051	20.0680	1.2980	21.9283	1.4633	24.9031	1.2817	20.6959
3	1.1431	19.5697	1.4754	20.3510	1.6275	22.5436	1.8372	25.7087	1.2339	21.8012
4	1.2903	20.0154	1.4638	19.7577	1.4928	23.3663	1.6803	26.0639	1.4481	22.6272
5	1.1135	20.7101	1.4507	21.1903	1.9225	24.9192	1.8761	29.6282	1.6219	23.1564
10	1.3842	21.9367	1.8298	22.4000	2.1449	30.2867	2.3667	31.6661	1.8491	26.0680
15	1.1594	23.4311	1.4097	23.5530	2.8359	34.4870	3.1650	36.5673	2.1127	29.2705
20	1.7972	23.6337	1.8907	24.1523	3.8548	37.0197	3.8627	39.6504	2.7245	32.7952
30	1.6874	26.3081	1.7189	25.9370	5.0427	43.3658	5.8144	45.9311	3.9473	37.1242
60	2.7131	32.1143	2.7504	32.3564	11.0986	60.2013	11.3047	62.6153	7.2667	48.9427

Note The table shows the results of a Monte Carlo simulation estimating daily drift using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The results use 500 independent samples of 5 000 estimates of daily variance generated using prices from a Heston model. The bias when estimating the daily drift using the generated variance $V_{\mathrm{Day},t}$ using all data is approx. 1.3812 with RMSE 17.8848. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.5, 7.68]. Daily variance has been multiplied by 10 000.

Table A.6:
Estimators of the daily diffusion for the case without noise - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	2.5258	3.2229	1.9634	3.3954	1.7841	3.0809	1.2896	5.0397	1.9091	3.1131
2	2.2924	3.2295	1.8661	3.3087	1.3687	3.3606	1.0204	5.1559	1.6341	3.1303
3	2.1567	3.1643	1.7847	3.2809	1.0881	3.3030	0.8695	4.4537	1.3783	3.2022
4	2.0292	3.1190	1.7387	3.1064	0.8732	3.2705	0.8784	4.5358	1.1588	3.3155
5	1.8802	3.1951	1.5935	3.4283	0.9483	3.8424	1.5063	11.2781	0.9899	3.3669
10	1.4385	3.3354	1.2525	3.2901	2.8245	6.2319	3.1908	6.9998	1.3168	4.0646
15	1.0904	3.2395	0.9979	3.4236	5.1285	8.3880	5.5029	10.7080	2.7439	5.3653
20	0.9129	3.6294	0.8620	3.7606	7.3744	10.2530	8.0095	12.1931	4.3908	7.0969
30	1.2025	3.9164	1.1961	4.1424	12.4629	15.7614	13.1563	17.8156	7.5524	10.4245
60	3.8025	6.9448	3.9191	6.5967	29.5380	36.1016	30.1007	36.7486	17.9954	21.8755

Note The table shows the results of a Monte Carlo simulation estimating daily diffusion using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the Heston model. The bias when estimating the daily diffusion using the generated variance $V_{\mathrm{Day},t}$ using all data is approx. 2.2087 with RMSE 3.2190. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.5, 7.68]. Daily variance has been multiplied by 10 000.

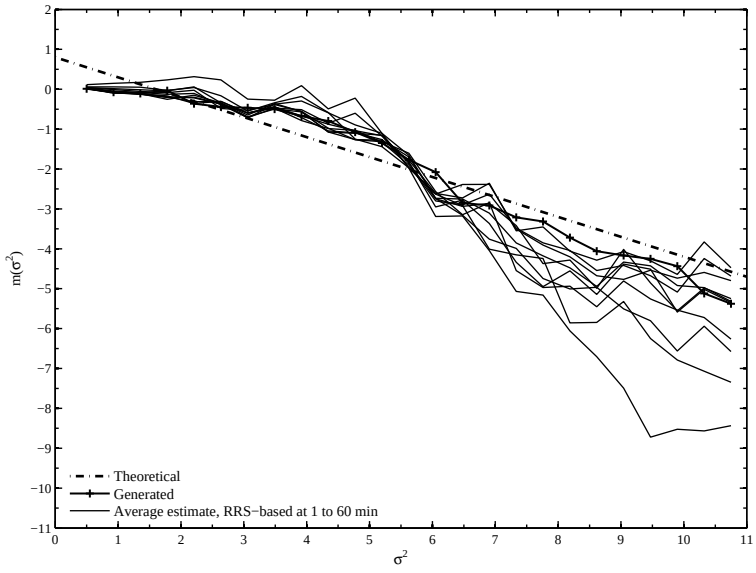

Figure A.9: Average estimates of the drift function $m(\cdot)$ using RR_S estimates of daily variance, where the 60 minute frequency generally gives the largest bias and the 1 minute frequency the lowest versus the generated dynamics. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

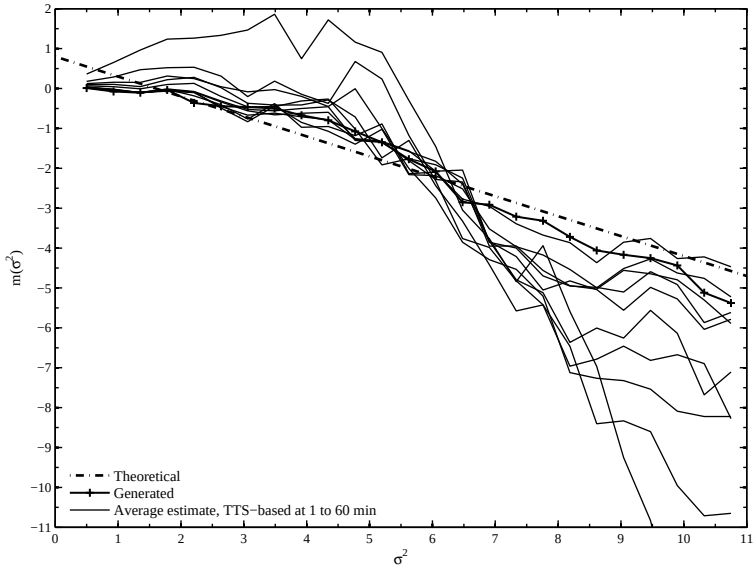

Figure A.10: Average estimates of the drift function $m(\cdot)$ using RV_{TTS} estimates of daily variance, where the 60 minute frequency generally gives the largest bias and the 1 minute frequency the lowest versus the generated dynamics. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

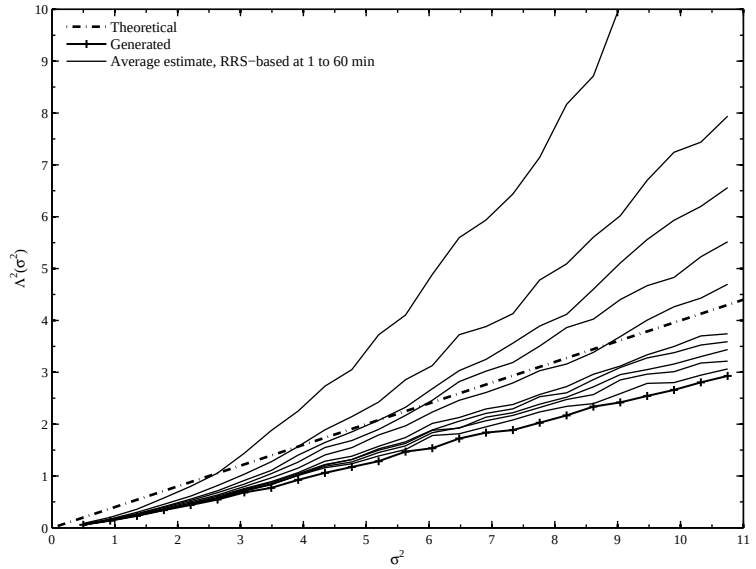

Figure A.11: Average estimates of the drift function $\Lambda^2(\cdot)$ using RR_S estimates of daily variance, where the 60 minute frequency gives the largest bias and the 1 minute frequency the lowest versus the generated dynamics. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the diffusion estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true diffusion function on the domain of the graph.

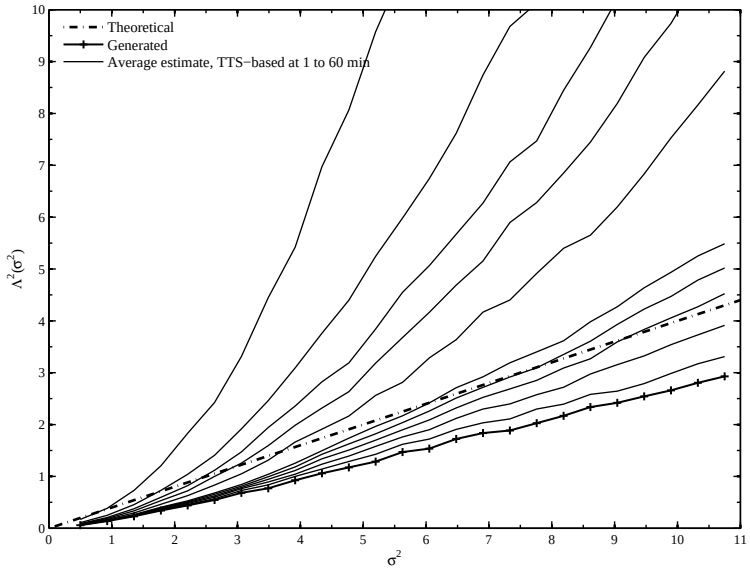

Figure A.12: Average estimates of the diffusion function $\Lambda^2(\cdot)$ using RV_{TTS} estimates of daily variance, where the 60 minute frequency gives the largest bias and the 1 minute frequency the lowest versus the generated dynamics. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the diffusion estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true diffusion function on the domain of the graph.

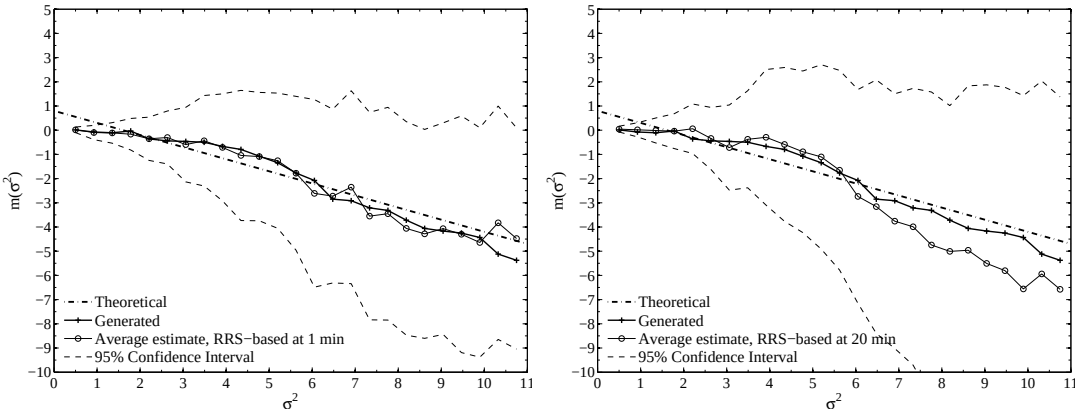

Figure A.13: Average estimates of the drift function $m(\cdot)$ using RR_S estimates of daily variance with 1-minute sampling (left panel) and 20-minute sampling (right panel). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

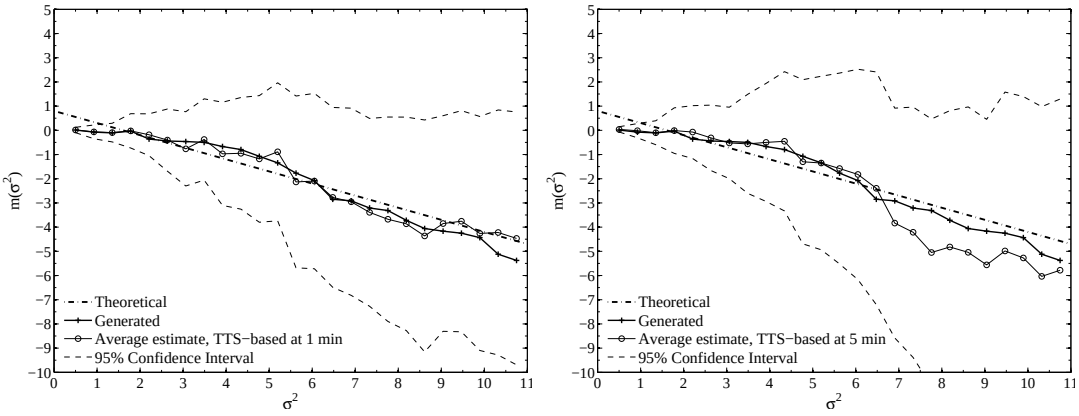

Figure A.14: Average estimates of the drift function $m(\cdot)$ using RV_{TTS} estimates of daily variance with 1-minute sampling (left panel) and 5-minute sampling (right panel). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

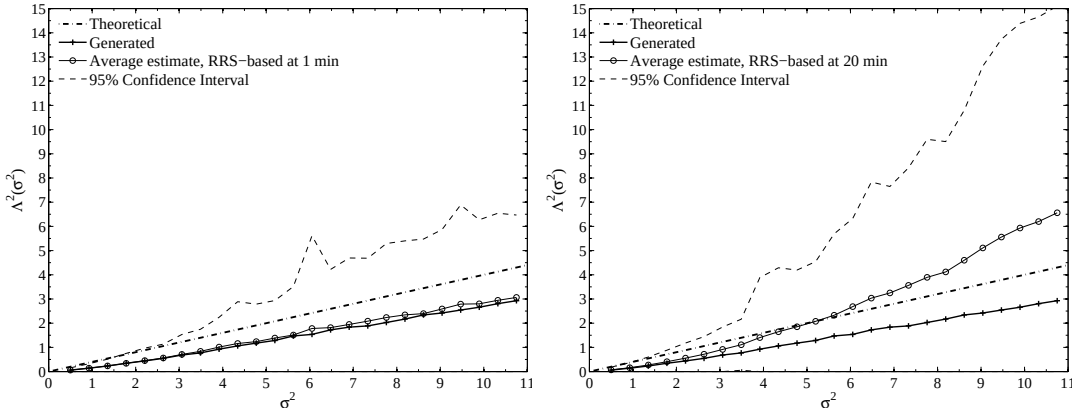

Figure A.15: Average estimates of the diffusion function $\Lambda^2(\cdot)$ using RR_S estimates of daily variance with 1-minute sampling (left panel) and 20-minute sampling (right panel). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the diffusion estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true diffusion function on the domain of the graph.

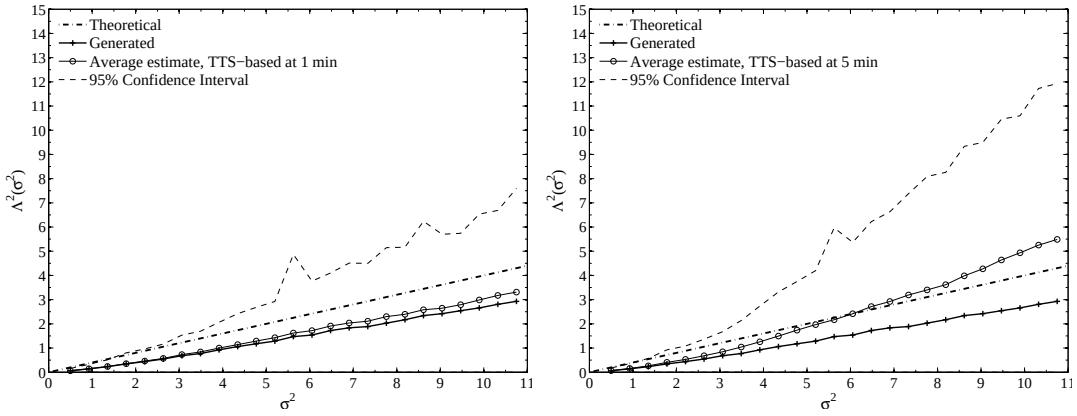

Figure A.16: Average estimates of the diffusion function $\Lambda^2(\cdot)$ using RV_{TTS} estimates of daily variance with 1-minute sampling (left panel) and 5-minute sampling (right panel). Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the diffusion estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true diffusion function on the domain of the graph.

Appendix B Tables and graphs - bid-ask bounce

This appendix contains results for the GARCH and Heston model in two separate parts that each hold three tables with the results for the estimation of the daily variance, the estimated drift, and the estimated diffusion for the simulation scenario with bid-ask bounce present. Following the three tables in each part are the graphs referenced to in the main body of this thesis.

Results for the GARCH model

Table B.1:
Estimators of daily variance using fully observed price paths with bid-ask bounce - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	2.1787	1.5459	0.6639	0.3073	2.4336	1.7998	0.6358	0.3018	0.6335	0.0567
2	1.6578	1.0241	0.6635	0.2560	1.5359	0.9036	0.6354	0.2399	0.6368	0.0459
3	1.4354	0.8016	0.6634	0.2250	1.2358	0.6055	0.6353	0.2015	0.6368	0.0478
4	1.3093	0.6758	0.6634	0.2040	1.0857	0.4578	0.6353	0.1761	0.6368	0.0516
5	1.2264	0.5932	0.6634	0.1885	0.9957	0.3706	0.6354	0.1586	0.6368	0.0558
10	1.0323	0.4007	0.6635	0.1462	0.8156	0.2093	0.6357	0.1266	0.6368	0.0751
15	0.9518	0.3223	0.6636	0.1271	0.7556	0.1724	0.6359	0.1279	0.6368	0.0910
20	0.9054	0.2783	0.6637	0.1166	0.7256	0.1654	0.6361	0.1371	0.6368	0.1047
30	0.8519	0.2306	0.6638	0.1076	0.6956	0.1756	0.6363	0.1600	0.6368	0.1282
60	0.7850	0.1838	0.6641	0.1114	0.6657	0.2282	0.6366	0.2210	0.6369	0.1825
1440	0.6649	0.4935	0.6649	0.4935	0.6367	1.0745	0.6367	1.0745	0.6367	1.0745

Note The table shows the results of a Monte Carlo simulation estimating daily variance using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the GARCH model. Sample average of the daily variance $\bar{V}_{\mathrm{Day},t} \approx 0.6356$ and $\theta = 0.636$. Daily variance has been multiplied by 10 000.

Table B.2:
Results for estimates of the drift for the case with bid-ask bounce - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	0.2182	0.2403	0.0165	0.0427	0.5974	0.6392	0.0324	0.0956	0.0060	0.0216
2	0.1538	0.1647	0.0164	0.0337	0.3389	0.3579	0.0354	0.0743	0.0043	0.0213
3	0.1216	0.1281	0.0148	0.0286	0.2213	0.2339	0.0415	0.0670	0.0030	0.0224
4	0.1038	0.1085	0.0134	0.0269	0.1676	0.1755	0.0464	0.0661	0.0027	0.0240
5	0.0923	0.0960	0.0119	0.0248	0.1373	0.1428	0.0508	0.0691	0.0051	0.0258
10	0.0680	0.0700	0.0063	0.0231	0.0970	0.1028	0.0746	0.0887	0.0226	0.0399
15	0.0588	0.0606	0.0039	0.0241	0.1051	0.1127	0.0984	0.1123	0.0405	0.0564
20	0.0542	0.0562	0.0075	0.0269	0.1222	0.1314	0.1210	0.1347	0.0582	0.0733
30	0.0509	0.0539	0.0180	0.0345	0.1632	0.1740	0.1652	0.1797	0.0924	0.1069
60	0.0613	0.0671	0.0492	0.0631	0.2790	0.2915	0.2829	0.2970	0.1861	0.2003

Note The table shows the results of a Monte Carlo simulation estimating daily drift using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The results use 500 independent samples of 5 000 estimates of daily variance generated using prices from a GARCH model. The bias when estimating the daily drift using the generated variance $V_{\text{Day},t}$ using all data is approx. 0.0176 with RMSE 0.0239. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.15, 1.25]. Daily variance has been multiplied by 10 000.

Table B.3:
Estimators of the daily diffusion for the case with bid-ask bounce - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	0.0510	0.0538	0.0532	0.0534	0.0734	0.1062	0.0441	0.0462	0.0116	0.0125
2	0.0537	0.0544	0.0488	0.0490	0.0534	0.0615	0.0316	0.0335	0.0099	0.0110
3	0.0525	0.0527	0.0458	0.0460	0.0398	0.0437	0.0205	0.0234	0.0057	0.0077
4	0.0503	0.0504	0.0432	0.0434	0.0272	0.0294	0.0108	0.0142	0.0021	0.0061
5	0.0480	0.0481	0.0409	0.0410	0.0179	0.0195	0.0052	0.0115	0.0052	0.0081
10	0.0376	0.0377	0.0314	0.0316	0.0439	0.0447	0.0454	0.0472	0.0309	0.0321
15	0.0287	0.0288	0.0232	0.0236	0.0851	0.0860	0.0853	0.0870	0.0571	0.0582
20	0.0206	0.0208	0.0156	0.0164	0.1242	0.1252	0.1236	0.1252	0.0830	0.0843
30	0.0067	0.0076	0.0030	0.0073	0.1964	0.1979	0.1962	0.1982	0.1342	0.1357
60	0.0409	0.0415	0.0437	0.0449	0.3906	0.3934	0.3883	0.3914	0.2792	0.2814

Note The table shows the results of a Monte Carlo simulation estimating daily diffusion using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the GARCH model. The bias when estimating the daily diffusion using the generated variance $V_{\text{Day},t}$ using all data is approx. 0.0235 with RMSE 0.0238. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.15, 1.25]. Daily variance has been multiplied by 10 000.

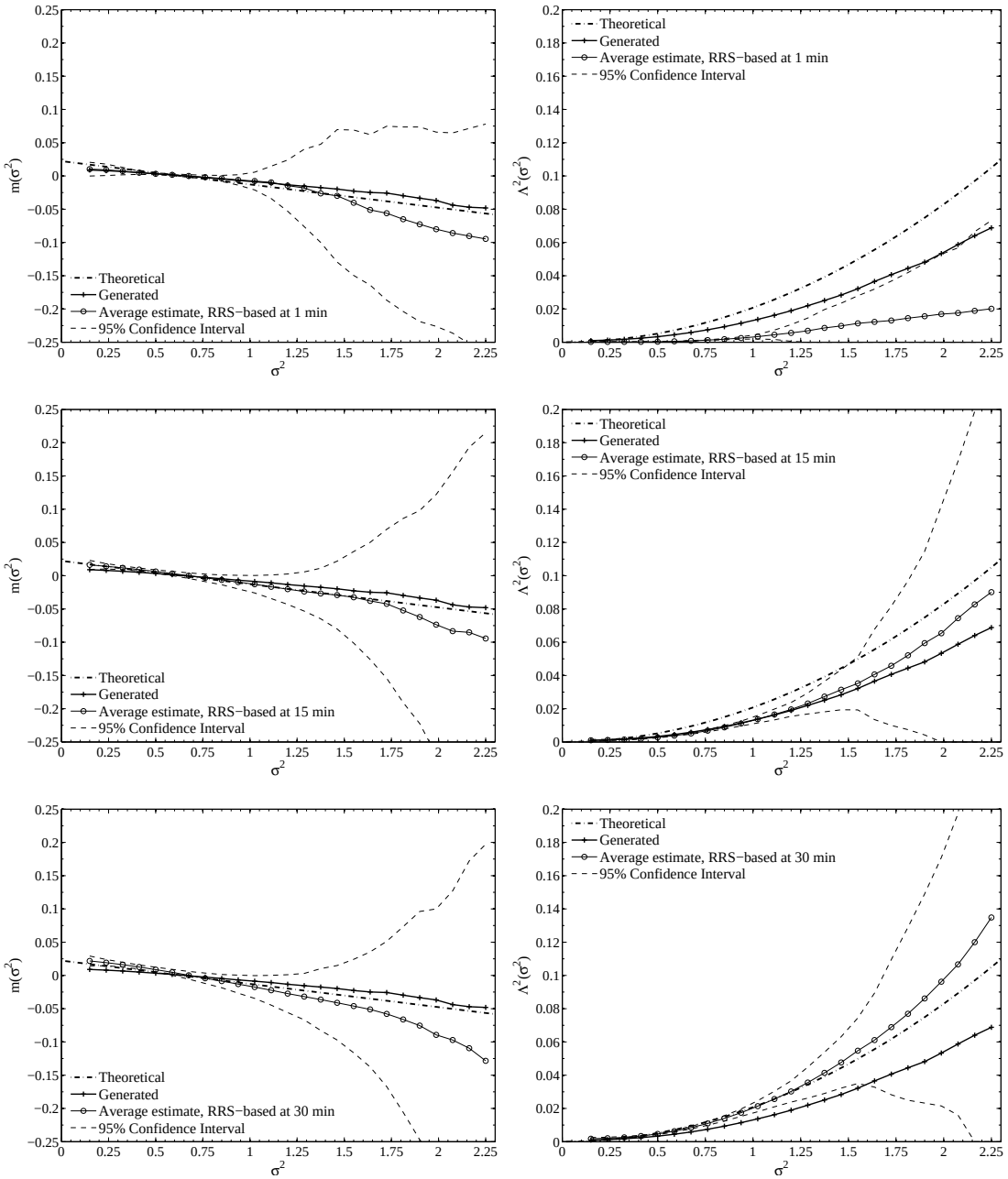

Figure B.1: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $m(\cdot)$ (right panels) using RR_S estimates of daily variance with 1, 15, and 30-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

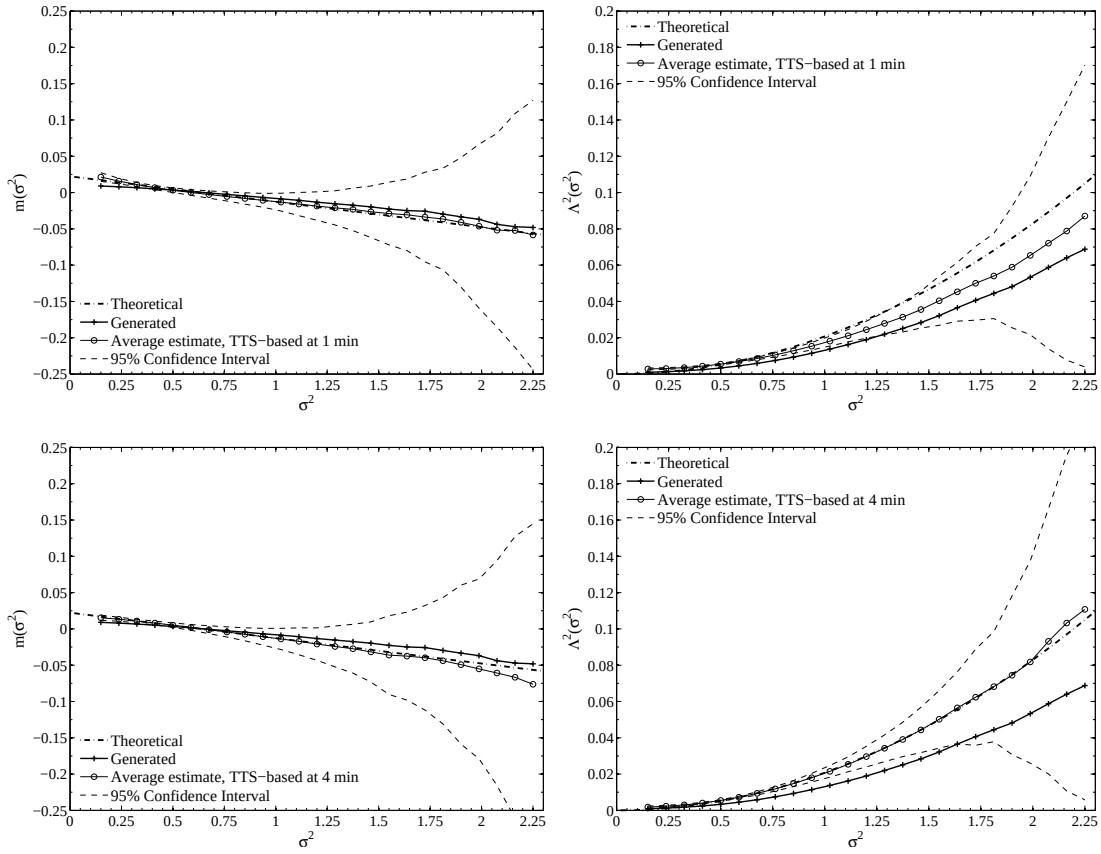

Figure B.2: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RV_{TTS} estimates of daily variance with 1-minute sampling and 4-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

Results for the Heston model

Table B.4:
Estimators of daily variance using fully observed price paths with bid-ask bounce - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	3.3910	1.9410	1.5710	1.6722	3.4995	1.8287	1.5767	1.7132	1.7006	0.2610
2	2.6517	1.2037	1.5974	1.2405	2.5995	1.0038	1.6017	1.3421	1.7008	0.3595
3	2.3889	0.9423	1.6132	1.0446	2.2995	0.8057	1.6173	1.2152	1.7008	0.4374
4	2.2509	0.8107	1.6240	0.9314	2.1495	0.7659	1.6283	1.1665	1.7008	0.5041
5	2.1651	0.7346	1.6320	0.8597	2.0602	0.7820	1.6364	1.1593	1.7010	0.5636
10	1.9802	0.6260	1.6545	0.7291	1.8806	0.9993	1.6588	1.2662	1.7014	0.8006
15	1.9116	0.6465	1.6658	0.7268	1.8212	1.2041	1.6700	1.4185	1.7015	0.9831
20	1.8740	0.6943	1.6727	0.7580	1.7894	1.3893	1.6755	1.5720	1.7012	1.1352
30	1.8336	0.8034	1.6815	0.8492	1.7601	1.6985	1.6839	1.8504	1.7006	1.3916
60	1.7860	1.0934	1.6936	1.1195	1.7278	2.3859	1.6955	2.5105	1.6998	1.9853
1440	1.7115	5.3076	1.7115	5.3076	1.6997	11.7961	1.6997	11.7961	1.6997	11.7961

Note The table shows the results of a Monte Carlo simulation estimating daily variance using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the Heston model. Sample average of the daily variance $\bar{V}_{\mathrm{Day},t} \approx 1.6998$ and $\theta = 1.6$. Daily variance has been multiplied by 10 000.

Table B.5:
Results for estimates of the drift for the case with bid-ask bounce - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	2.5641	6.2015	2.1488	9.0391	3.1211	12.3442	1.6705	11.1555	1.2717	19.3339
2	2.5484	9.1220	1.9413	10.9860	2.0565	17.4039	1.4192	15.4666	1.3500	20.8182
3	2.5073	10.9403	1.7993	12.3183	2.2341	20.3599	1.4761	17.9523	1.2695	21.7778
4	2.3408	12.4102	1.3700	13.8108	2.0394	21.5846	1.0793	20.6006	1.3713	22.7094
5	2.1922	13.7204	1.2981	15.0336	1.5080	23.4124	1.2609	22.0760	1.5999	23.2582
10	2.0884	16.0184	1.2983	18.2875	2.3179	28.2732	2.1582	29.4233	1.8428	26.1054
15	2.0089	18.2670	0.9431	19.5883	3.0350	32.7351	1.9020	34.9077	2.1137	29.1118
20	2.3755	18.7123	1.0303	21.0562	3.8758	35.8845	2.7640	39.1354	2.7428	32.8916
30	2.2114	22.3275	1.1463	23.6496	5.3797	43.5324	4.0280	46.8411	3.9003	37.2139
60	2.9554	27.3161	1.9877	29.8758	10.0091	59.3703	9.8631	64.5608	7.2508	48.9962

Note The table shows the results of a Monte Carlo simulation estimating daily drift using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The results use 500 independent samples of 5 000 estimates of daily variance generated using prices from a Heston model. The bias when estimating the daily drift using the generated variance $V_{\mathrm{Day},t}$ using all data is approx. 1.3812 with RMSE 17.8848. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.5, 7.68]. Daily variance has been multiplied by 10 000.

Table B.6:
Estimators of the daily diffusion for the case with bid-ask bounce - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	3.3603	3.4659	4.4061	3.9362	3.4669	3.7243	3.9436	3.4855	1.8612	3.0711
2	2.9333	3.1263	4.0482	3.7761	2.3029	3.0809	3.0474	3.1352	1.5970	3.1761
3	2.7177	2.9812	3.7996	3.6477	1.7056	3.3771	2.4995	3.0252	1.3666	3.2209
4	2.5388	2.9073	3.5612	3.5014	1.2804	3.0505	1.9039	2.8182	1.1514	3.3217
5	2.3947	2.8798	3.3594	3.3940	1.0702	3.6946	1.4837	3.0725	0.9852	3.4086
10	1.9279	2.6924	2.6661	3.1546	2.5082	5.1408	2.1476	5.1678	1.3293	4.0656
15	1.5156	2.6969	2.1735	2.9482	4.9041	7.4477	4.6695	7.5270	2.7387	5.3191
20	1.2406	2.7970	1.7649	2.9324	7.1751	10.1038	7.2124	10.8536	4.3915	7.1630
30	1.0051	3.3823	1.1443	3.1704	12.3673	15.8775	12.2068	15.9562	7.5696	10.4603
60	3.0677	5.4060	2.7554	5.1691	29.3413	35.4452	29.8124	37.2242	17.9990	21.8760

Note The table shows the results of a Monte Carlo simulation estimating daily diffusion using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the Heston model. The bias when estimating the daily diffusion using the generated variance $V_{\text{Day},t}$ using all data is approx. 2.2087 with RMSE 3.2190. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.5, 7.68]. Daily variance has been multiplied by 10 000.

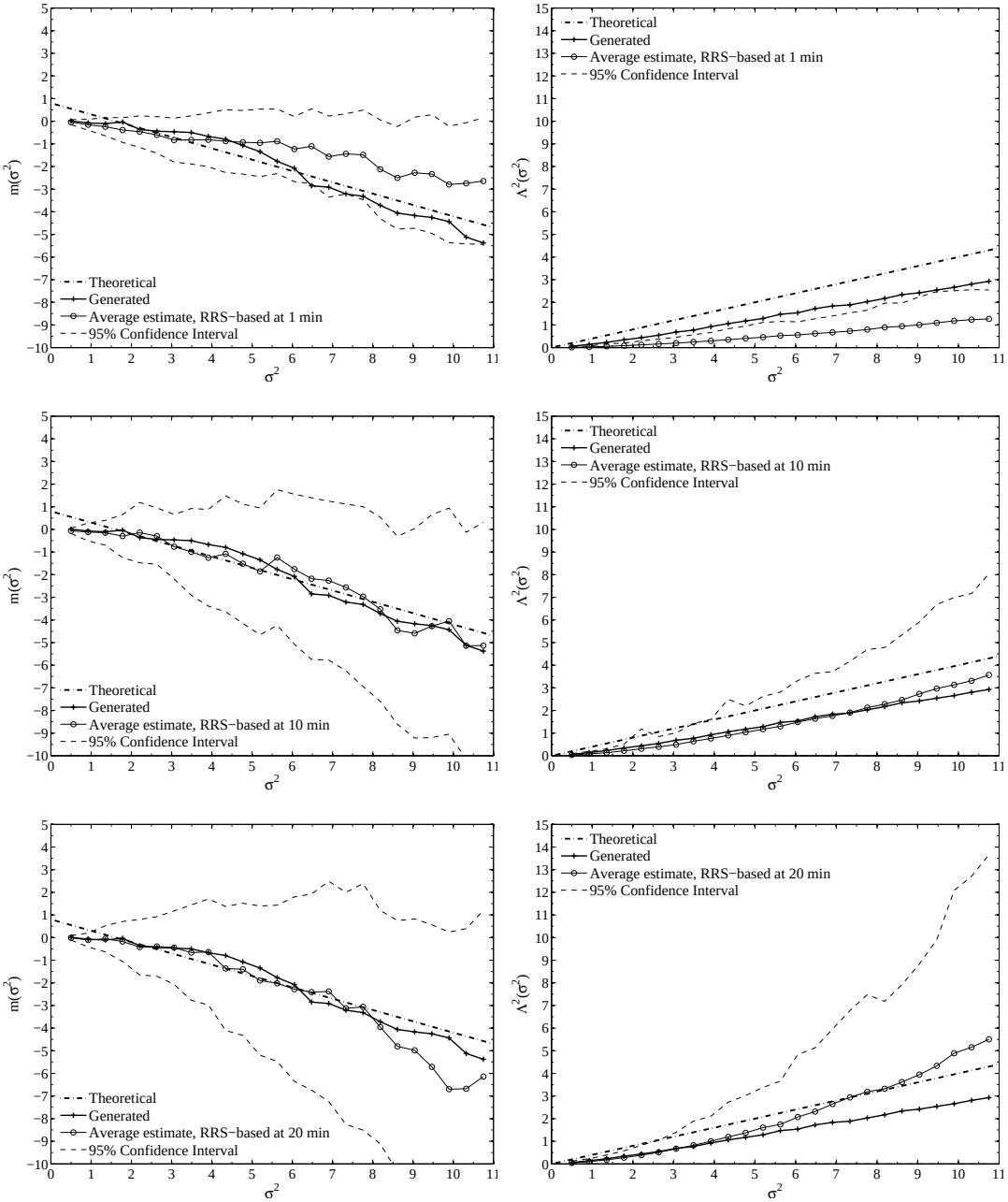

Figure B.3: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $m(\cdot)$ (right panels) using RR_S estimates of daily variance with 1, 10, and 20-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

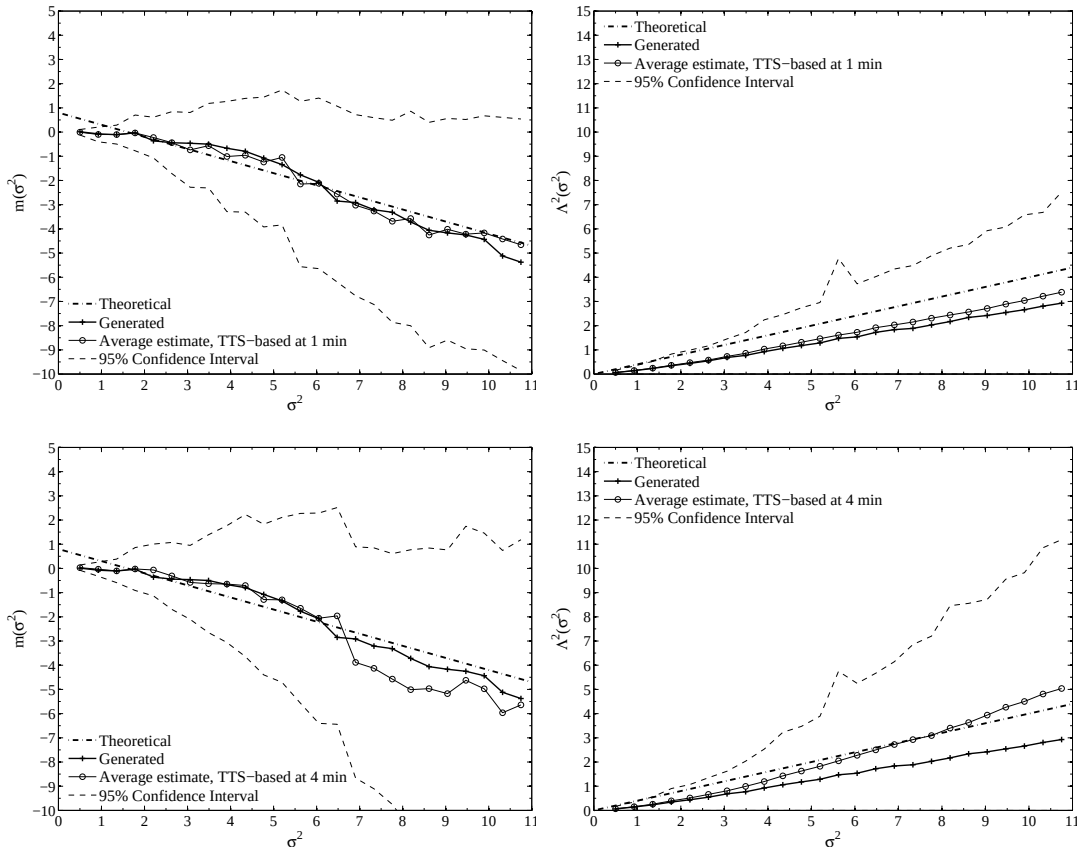

Figure B.4: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RV_{TTS} estimates of daily variance with 1-minute sampling and 4-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

Appendix C Tables and graphs - infrequent trading

This appendix contains results for the GARCH and Heston model in two separate parts that each hold three tables with the results for the estimation of the daily variance, the estimated drift, and the estimated diffusion for the simulation scenario with infrequent trading. Following the three tables in each part are the graphs referenced to in the main body of this thesis.

Results for the GARCH model

Table C.1:
Estimators of daily variance in the presence of infrequent trading - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	0.3540	0.3363	0.6249	0.0294	0.6356	0.0290	0.6354	0.0493	0.6356	0.0276
2	0.4116	0.2678	0.6249	0.0323	0.6356	0.0402	0.6354	0.0566	0.6356	0.0356
3	0.4432	0.2304	0.6249	0.0351	0.6356	0.0493	0.6354	0.0634	0.6356	0.0424
4	0.4640	0.2061	0.6249	0.0376	0.6356	0.0566	0.6354	0.0692	0.6355	0.0482
5	0.4790	0.1887	0.6249	0.0399	0.6355	0.0633	0.6354	0.0747	0.6355	0.0534
10	0.5193	0.1436	0.6249	0.0496	0.6356	0.0894	0.6354	0.0979	0.6355	0.0743
15	0.5385	0.1247	0.6249	0.0577	0.6355	0.1094	0.6354	0.1165	0.6355	0.0906
20	0.5505	0.1149	0.6249	0.0646	0.6356	0.1264	0.6355	0.1326	0.6355	0.1045
30	0.5650	0.1072	0.6249	0.0765	0.6356	0.1548	0.6355	0.1600	0.6356	0.1281
60	0.5847	0.1129	0.6249	0.1040	0.6357	0.2190	0.6356	0.2227	0.6356	0.1824
1440	0.6251	0.4817	0.6251	0.4817	0.6355	1.0729	0.6355	1.0729	0.6355	1.0729

Note The table shows the results of a Monte Carlo simulation estimating daily variance using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the GARCH model. Sample average of the daily variance $\bar{V}_{\mathrm{Day},t} \approx 0.6356$ and $\theta = 0.636$. Daily variance has been multiplied by 10 000.

Table C.2:
Results for estimates of the drift for the case with infrequent trading - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	0.0308	0.0847	0.0147	0.0232	0.0115	0.0224	0.0120	0.0223	0.0120	0.0222
2	0.0237	0.0579	0.0133	0.0229	0.0062	0.0226	0.0066	0.0225	0.0084	0.0221
3	0.0210	0.0467	0.0119	0.0227	0.0038	0.0250	0.0037	0.0248	0.0052	0.0229
4	0.0193	0.0429	0.0105	0.0226	0.0076	0.0278	0.0068	0.0273	0.0038	0.0242
5	0.0184	0.0410	0.0093	0.0225	0.0128	0.0316	0.0118	0.0311	0.0056	0.0261
10	0.0172	0.0393	0.0045	0.0243	0.0397	0.0563	0.0391	0.0564	0.0233	0.0406
15	0.0188	0.0401	0.0068	0.0281	0.0654	0.0803	0.0649	0.0803	0.0414	0.0572
20	0.0225	0.0441	0.0121	0.0320	0.0915	0.1059	0.0905	0.1053	0.0592	0.0743
30	0.0322	0.0525	0.0242	0.0417	0.1385	0.1530	0.1376	0.1525	0.0934	0.1079
60	0.0649	0.0829	0.0580	0.0741	0.2639	0.2780	0.2616	0.2763	0.1871	0.2014

Note The table shows the results of a Monte Carlo simulation estimating daily drift using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The results use 500 independent samples of 5 000 estimates of daily variance generated using prices from a GARCH model. The bias when estimating the daily drift using the generated variance $V_{\text{Day},t}$ using all data is approx. 0.0176 with RMSE 0.0239. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.15, 1.25]. Daily variance has been multiplied by 10 000.

Table C.3:
Estimators of the daily diffusion for the case with infrequent trading - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	0.0177	0.0247	0.0195	0.0200	0.0149	0.0156	0.0149	0.0156	0.0157	0.0163
2	0.0166	0.0209	0.0176	0.0182	0.0073	0.0089	0.0072	0.0089	0.0106	0.0117
3	0.0149	0.0176	0.0157	0.0163	0.0021	0.0064	0.0021	0.0065	0.0055	0.0077
4	0.0129	0.0157	0.0138	0.0146	0.0090	0.0113	0.0090	0.0114	0.0018	0.0061
5	0.0110	0.0141	0.0119	0.0129	0.0169	0.0186	0.0169	0.0185	0.0059	0.0086
10	0.0023	0.0087	0.0031	0.0066	0.0556	0.0568	0.0557	0.0568	0.0322	0.0335
15	0.0070	0.0112	0.0066	0.0093	0.0941	0.0954	0.0940	0.0954	0.0586	0.0598
20	0.0157	0.0184	0.0151	0.0169	0.1304	0.1318	0.1304	0.1318	0.0847	0.0859
30	0.0328	0.0348	0.0322	0.0336	0.2012	0.2031	0.2015	0.2035	0.1360	0.1375
60	0.0828	0.0844	0.0824	0.0837	0.3922	0.3953	0.3931	0.3963	0.2812	0.2835

Note The table shows the results of a Monte Carlo simulation estimating daily diffusion using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the GARCH model. The bias when estimating the daily diffusion using the generated variance $V_{\text{Day},t}$ using all data is approx. 0.0235 with RMSE 0.0238. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.15, 1.25]. Daily variance has been multiplied by 10 000.

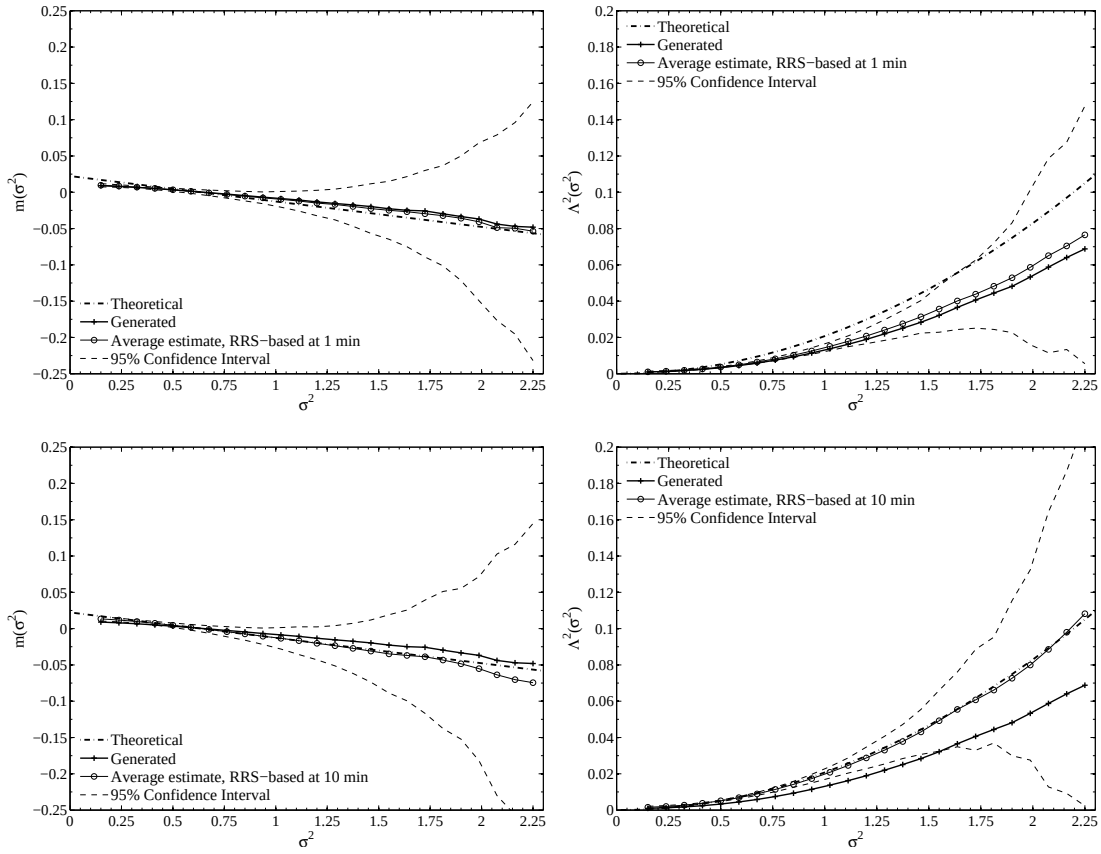

Figure C.1: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RR_S estimates of daily variance with 1-minute sampling and 10-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

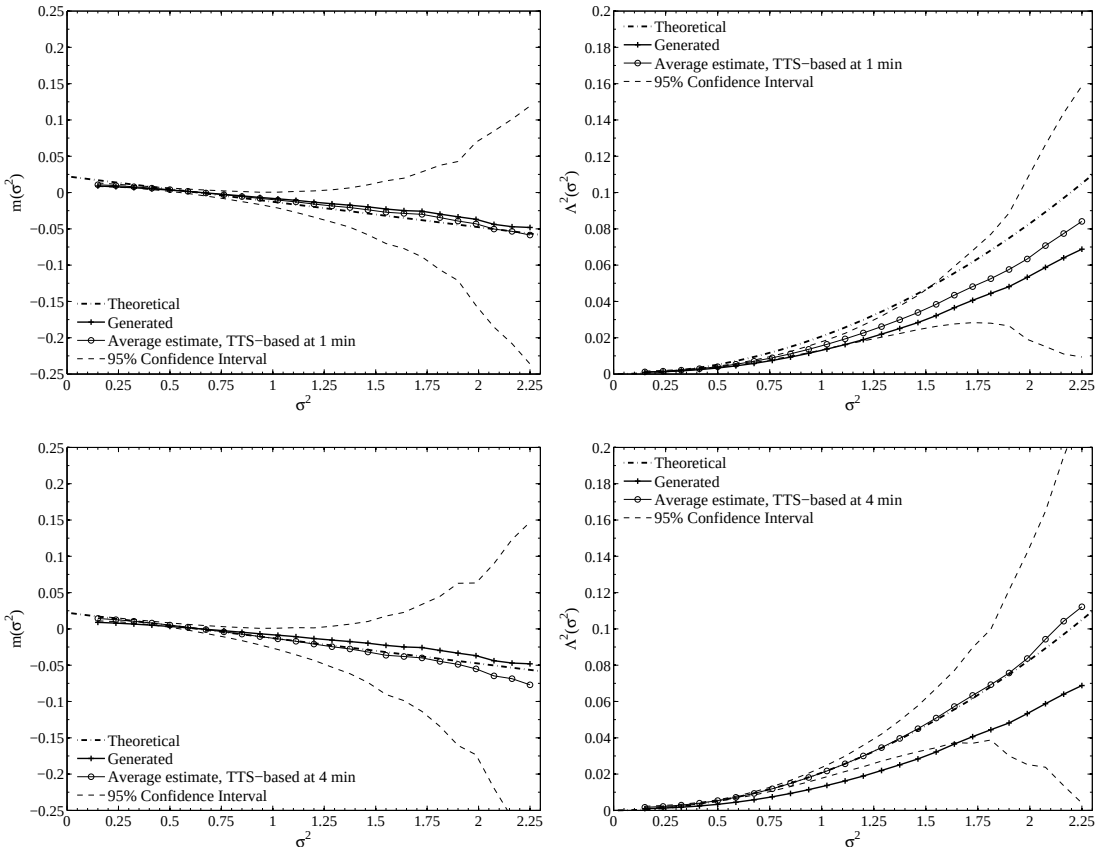

Figure C.2: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RV_{TTS} estimates of daily variance with 1-minute sampling and 4-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

Results for the Heston model

Table C.4:
Estimators of daily variance in the presence of infrequent trading - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	0.9466	3.6753	1.6720	0.4151	1.6995	0.3158	1.7080	0.8036	1.6996	0.3008
2	1.1006	2.9273	1.6720	0.4404	1.6997	0.4365	1.7080	0.8588	1.6997	0.3864
3	1.1853	2.5185	1.6720	0.4646	1.6997	0.5323	1.7079	0.9145	1.6996	0.4587
4	1.2407	2.2537	1.6720	0.4860	1.6997	0.6131	1.7082	0.9626	1.6997	0.5227
5	1.2810	2.0620	1.6720	0.5082	1.7003	0.6867	1.7081	1.0138	1.6998	0.5804
10	1.3888	1.5699	1.6720	0.6010	1.7005	0.9769	1.7080	1.2309	1.7002	0.8125
15	1.4407	1.3594	1.6721	0.6826	1.7013	1.1953	1.7085	1.4134	1.7002	0.9928
20	1.4723	1.2551	1.6719	0.7518	1.6994	1.3834	1.7072	1.5751	1.7000	1.1437
30	1.5114	1.1708	1.6719	0.8753	1.7000	1.6958	1.7074	1.8606	1.6994	1.3985
60	1.5635	1.2357	1.6721	1.1629	1.6977	2.3829	1.7085	2.5189	1.6985	1.9901

Note The table shows the results of a Monte Carlo simulation estimating daily variance using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the Heston model. Sample average of the daily variance $\bar{V}_{\text{Day},t} \approx 1.6998$ and $\theta = 1.6$. Daily variance has been multiplied by 10 000.

Table C.5:
Results for estimates of the drift for the case with infrequent trading - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	1.6446	17.7963	1.1141	19.3391	1.5618	19.2034	1.4607	23.1583	1.2985	19.7263
2	1.6007	19.0445	1.2168	19.4163	1.7096	21.5222	1.6776	27.1095	1.3252	20.8001
3	1.7656	19.9815	1.4583	20.4696	1.7193	22.3421	1.6697	25.1447	1.2775	22.0384
4	1.8919	20.9620	1.4095	20.1469	1.5037	23.3150	1.7824	26.1011	1.5369	22.9093
5	1.7412	20.7370	1.5566	21.0704	1.5544	24.9416	1.8416	29.5372	1.6818	23.3381
10	1.9476	23.4511	1.7287	23.0440	2.3657	29.6913	2.2246	33.4938	1.8327	26.0522
15	1.5176	24.5471	1.3879	24.9667	2.7072	34.3909	3.0764	35.1520	2.1661	29.4114
20	2.1237	25.0270	1.8099	24.9920	3.9021	37.6427	3.5071	40.2352	2.8125	32.9201
30	1.8720	28.2506	1.7645	27.2511	5.1546	43.5449	5.7353	45.9867	4.0428	37.1546
60	3.1445	33.0241	3.0181	33.3752	10.5654	59.8255	11.3248	62.7090	7.2796	49.0128

Note The table shows the results of a Monte Carlo simulation estimating daily drift using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The results use 500 independent samples of 5 000 estimates of daily variance generated using prices from a Heston model. The bias when estimating the daily drift using the generated variance $V_{\text{Day},t}$ using all data is approx. 1.3812 with RMSE 17.8848. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.5, 7.68]. Daily variance has been multiplied by 10 000.

Table C.6:
Estimators of the daily diffusion for the case with infrequent trading - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\text{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	3.3434	3.4984	1.9149	3.2776	1.8016	3.1017	1.3140	4.4190	1.7978	3.0907
2	2.9212	3.3228	1.8284	3.1207	1.3715	3.5932	1.0290	10.8392	1.5432	3.1750
3	2.6699	3.2543	1.7141	3.2689	1.0936	3.3021	0.8327	4.2158	1.2956	3.2321
4	2.4094	3.2259	1.6406	3.1366	0.8813	3.3624	0.8838	4.6912	1.0967	3.3565
5	2.2889	3.1543	1.5295	3.2529	0.9661	4.3920	1.4584	11.9295	0.9491	3.4074
10	1.5801	3.2144	1.1464	3.6161	2.7501	5.7096	3.3580	8.0746	1.3808	4.0968
15	1.1563	3.3267	0.9033	3.8039	5.0808	7.9514	5.4743	8.9432	2.8409	5.4752
20	0.9353	3.3930	0.9087	3.9858	7.3940	10.4605	8.0147	12.2917	4.4859	7.2104
30	1.3390	4.4552	1.4769	4.8749	12.4270	15.7310	13.1047	18.2632	7.6348	10.4965
60	4.0823	6.7374	4.3043	6.9759	29.5513	35.6970	30.1845	37.2454	18.0969	21.9927

Note The table shows the results of a Monte Carlo simulation estimating daily diffusion using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\text{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the Heston model. The bias when estimating the daily diffusion using the generated variance $V_{\text{Day},t}$ using all data is approx. 2.2087 with RMSE 3.2190. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.5, 7.68]. Daily variance has been multiplied by 10 000.

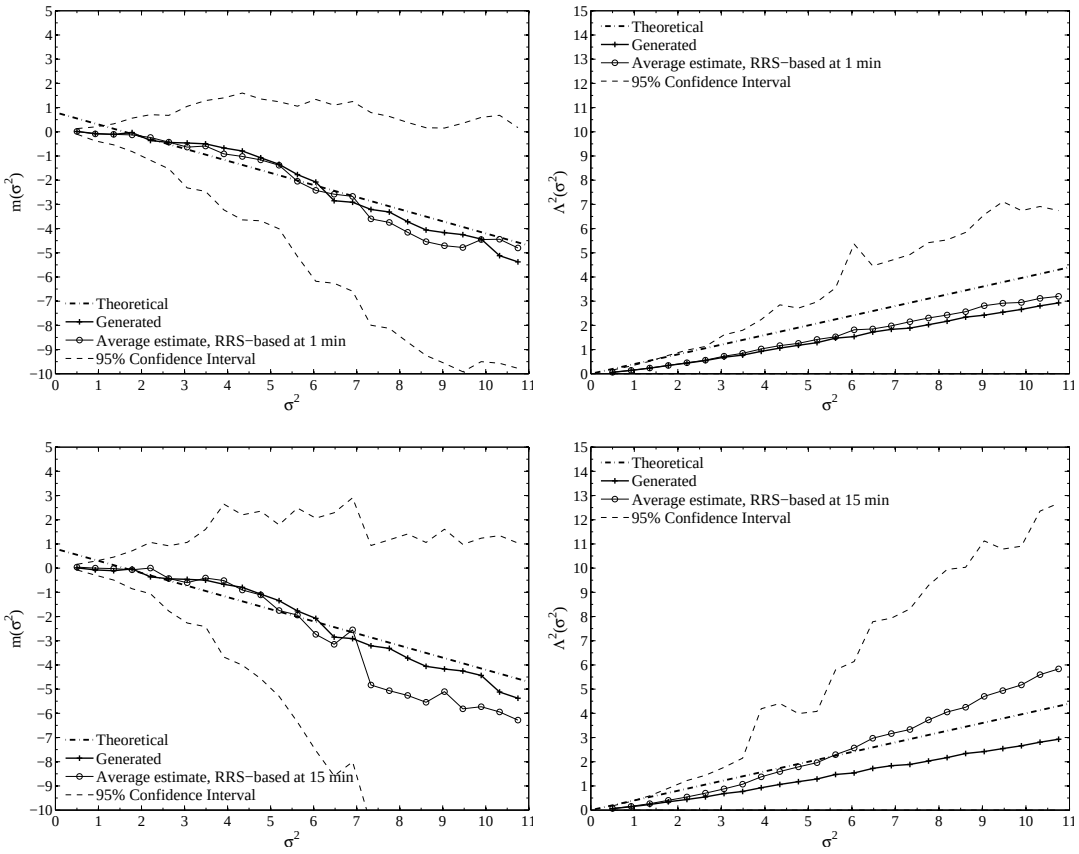

Figure C.3: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RR_S estimates of daily variance with 1-minute sampling and 15-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

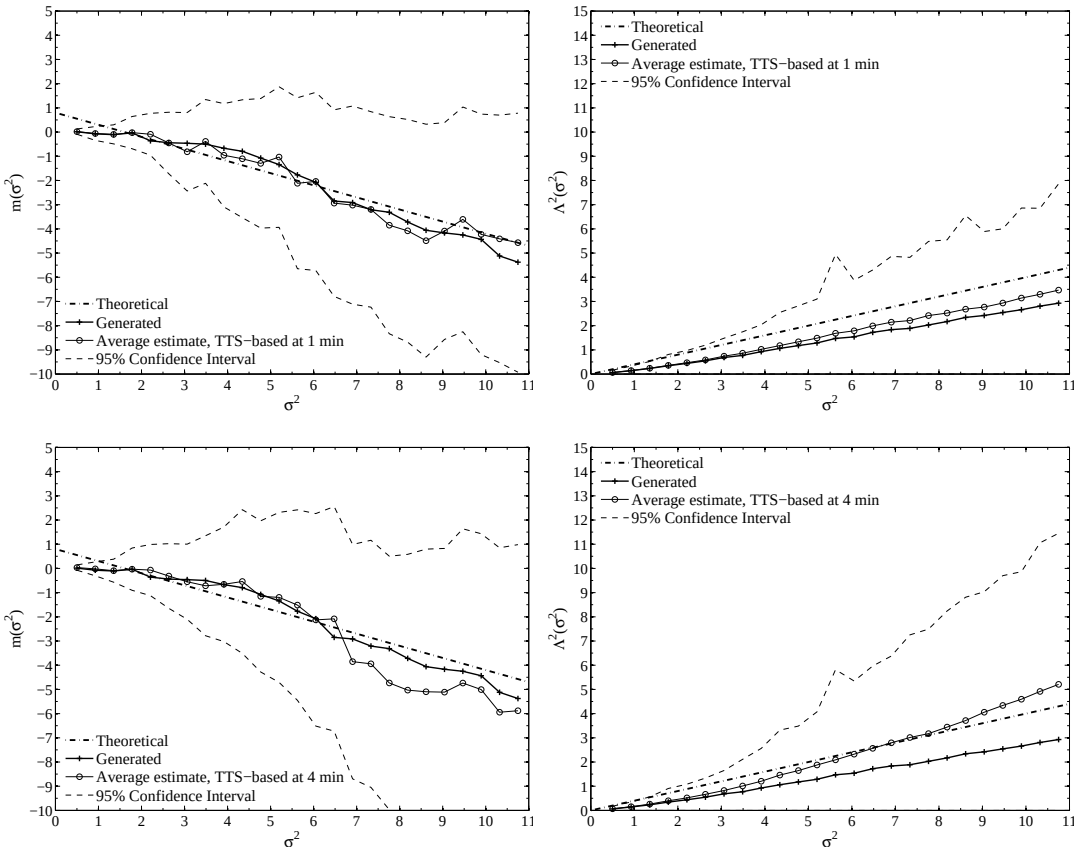

Figure C.4: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RV_{TTS} estimates of daily variance with 1-minute sampling and 4-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

Appendix D Tables and graphs - both noise types

This appendix contains results for the GARCH and Heston model in two separate parts that each hold three tables with the results for the estimation of the daily variance, the estimated drift, and the estimated diffusion for the simulation scenario with bid-ask bounce and infrequent trading present. Following the three tables in each part are the graphs referenced to in the main body of this thesis.

Results for the GARCH model

Table D.1:
Estimators of daily variance with bid-ask bounce and infrequent trading- GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	2.1787	1.5459	0.6639	0.3073	2.4336	1.7998	0.6358	0.3018	0.6335	0.0567
2	1.6578	1.0241	0.6635	0.2560	1.5359	0.9036	0.6354	0.2399	0.6368	0.0459
3	1.4354	0.8016	0.6634	0.2250	1.2358	0.6055	0.6353	0.2015	0.6368	0.0478
4	1.3093	0.6758	0.6634	0.2040	1.0857	0.4578	0.6353	0.1761	0.6368	0.0516
5	1.2264	0.5932	0.6634	0.1885	0.9957	0.3706	0.6354	0.1586	0.6368	0.0558
10	1.0323	0.4007	0.6635	0.1462	0.8156	0.2093	0.6357	0.1266	0.6368	0.0751
15	0.9518	0.3223	0.6636	0.1271	0.7556	0.1724	0.6359	0.1279	0.6368	0.0910
20	0.9054	0.2783	0.6637	0.1166	0.7256	0.1654	0.6361	0.1371	0.6368	0.1047
30	0.8519	0.2306	0.6638	0.1076	0.6956	0.1756	0.6363	0.1600	0.6368	0.1282
60	0.7850	0.1838	0.6641	0.1114	0.6657	0.2282	0.6366	0.2210	0.6369	0.1825
1440	0.6649	0.4935	0.6649	0.4935	0.6367	1.0745	0.6367	1.0745	0.6367	1.0745

Note The table shows the results of a Monte Carlo simulation estimating daily variance using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the GARCH model. Sample average of the daily variance $\bar{V}_{\mathrm{Day},t} \approx 0.6356$ and $\theta = 0.636$. Daily variance has been multiplied by 10 000.

Table D.2:
Results for estimates of the drift for the case with bid-ask bounce and infrequent trading - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	0.1567	0.1721	0.0135	0.0638	0.6155	0.6596	0.0335	0.0938	0.0276	0.0372
2	0.1084	0.1177	0.0120	0.0482	0.3407	0.3604	0.0363	0.0726	0.0047	0.0229
3	0.0878	0.0940	0.0118	0.0389	0.2218	0.2332	0.0412	0.0665	0.0025	0.0234
4	0.0752	0.0795	0.0106	0.0341	0.1672	0.1747	0.0461	0.0668	0.0046	0.0250
5	0.0671	0.0704	0.0091	0.0283	0.1359	0.1420	0.0511	0.0694	0.0075	0.0270
10	0.0494	0.0518	0.0031	0.0258	0.0961	0.1018	0.0742	0.0892	0.0244	0.0413
15	0.0426	0.0453	0.0059	0.0273	0.1037	0.1115	0.0972	0.1112	0.0420	0.0576
20	0.0397	0.0430	0.0116	0.0311	0.1215	0.1304	0.1211	0.1353	0.0596	0.0746
30	0.0395	0.0445	0.0232	0.0401	0.1626	0.1735	0.1659	0.1803	0.0936	0.1080
60	0.0585	0.0665	0.0563	0.0707	0.2781	0.2908	0.2818	0.2962	0.1870	0.2012

Note The table shows the results of a Monte Carlo simulation estimating daily drift using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The results use 500 independent samples of 5 000 estimates of daily variance generated using prices from a GARCH model. The bias when estimating the daily drift using the generated variance $V_{\mathrm{Day},t}$ using all data is approx. 0.0176 with RMSE 0.0239. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.15, 1.25]. Daily variance has been multiplied by 10 000.

Table D.3:
Estimators of the daily diffusion for the case with bid-ask bounce and infrequent trading - GARCH

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	0.0582	0.0587	0.0544	0.0548	0.0794	0.1168	0.0442	0.0459	0.0104	0.0118
2	0.0581	0.0583	0.0499	0.0503	0.0543	0.0632	0.0317	0.0332	0.0057	0.0078
3	0.0548	0.0549	0.0463	0.0468	0.0397	0.0432	0.0204	0.0226	0.0032	0.0065
4	0.0514	0.0515	0.0435	0.0439	0.0270	0.0290	0.0106	0.0141	0.0038	0.0071
5	0.0483	0.0484	0.0410	0.0412	0.0177	0.0195	0.0052	0.0117	0.0080	0.0103
10	0.0359	0.0360	0.0300	0.0304	0.0440	0.0448	0.0455	0.0473	0.0330	0.0342
15	0.0259	0.0261	0.0209	0.0216	0.0851	0.0860	0.0859	0.0875	0.0590	0.0601
20	0.0171	0.0174	0.0126	0.0139	0.1238	0.1249	0.1236	0.1252	0.0848	0.0860
30	0.0044	0.0061	0.0052	0.0092	0.1966	0.1981	0.1956	0.1975	0.1358	0.1373
60	0.0478	0.0485	0.0503	0.0516	0.3895	0.3923	0.3884	0.3916	0.2805	0.2828

Note The table shows the results of a Monte Carlo simulation estimating daily diffusion using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the GARCH model. The bias when estimating the daily diffusion using the generated variance $V_{\mathrm{Day},t}$ using all data is approx. 0.0235 with RMSE 0.0238. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.15, 1.25]. Daily variance has been multiplied by 10 000.

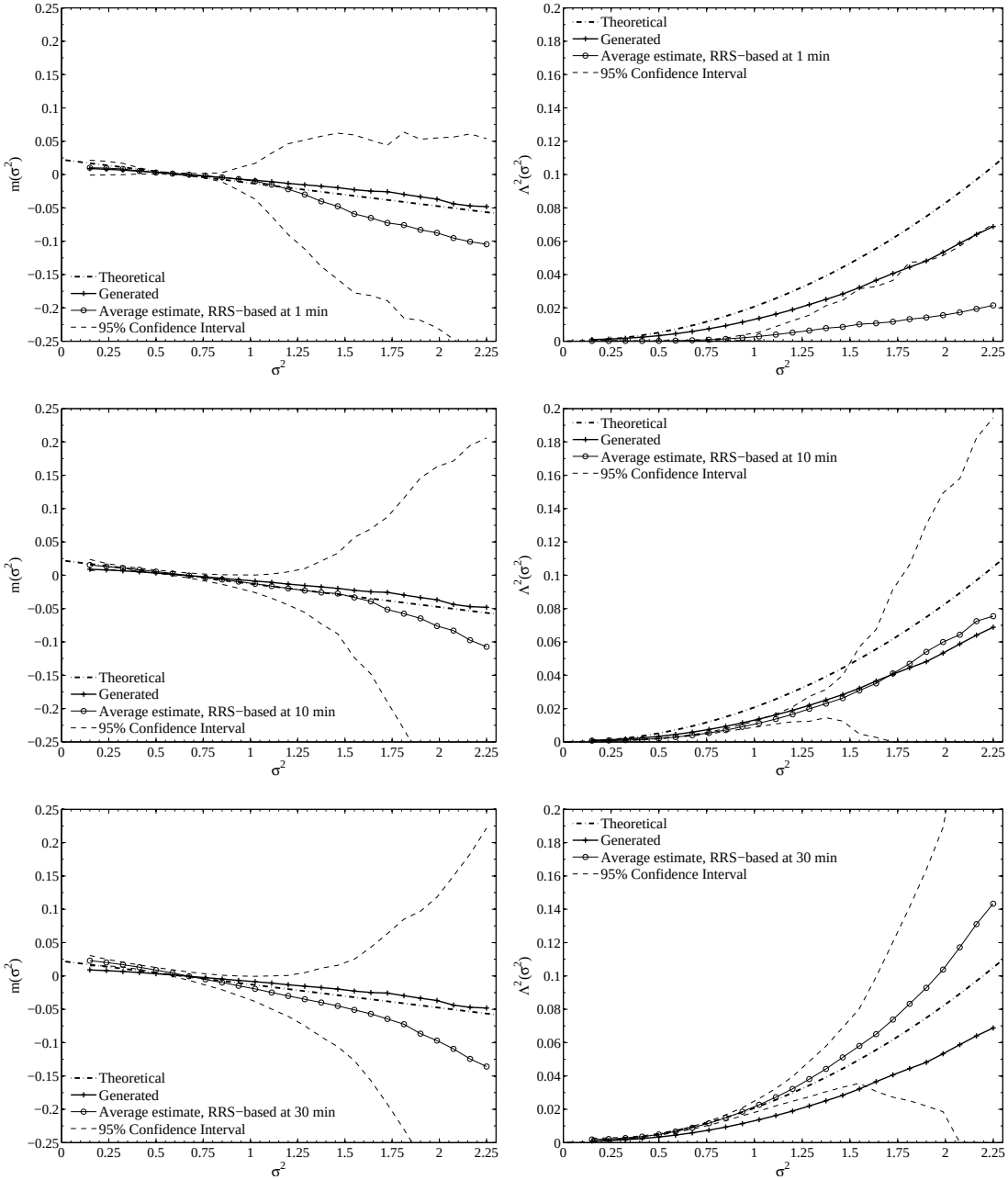

Figure D.1: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RR_S estimates of daily variance with 1, 10, and 30-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

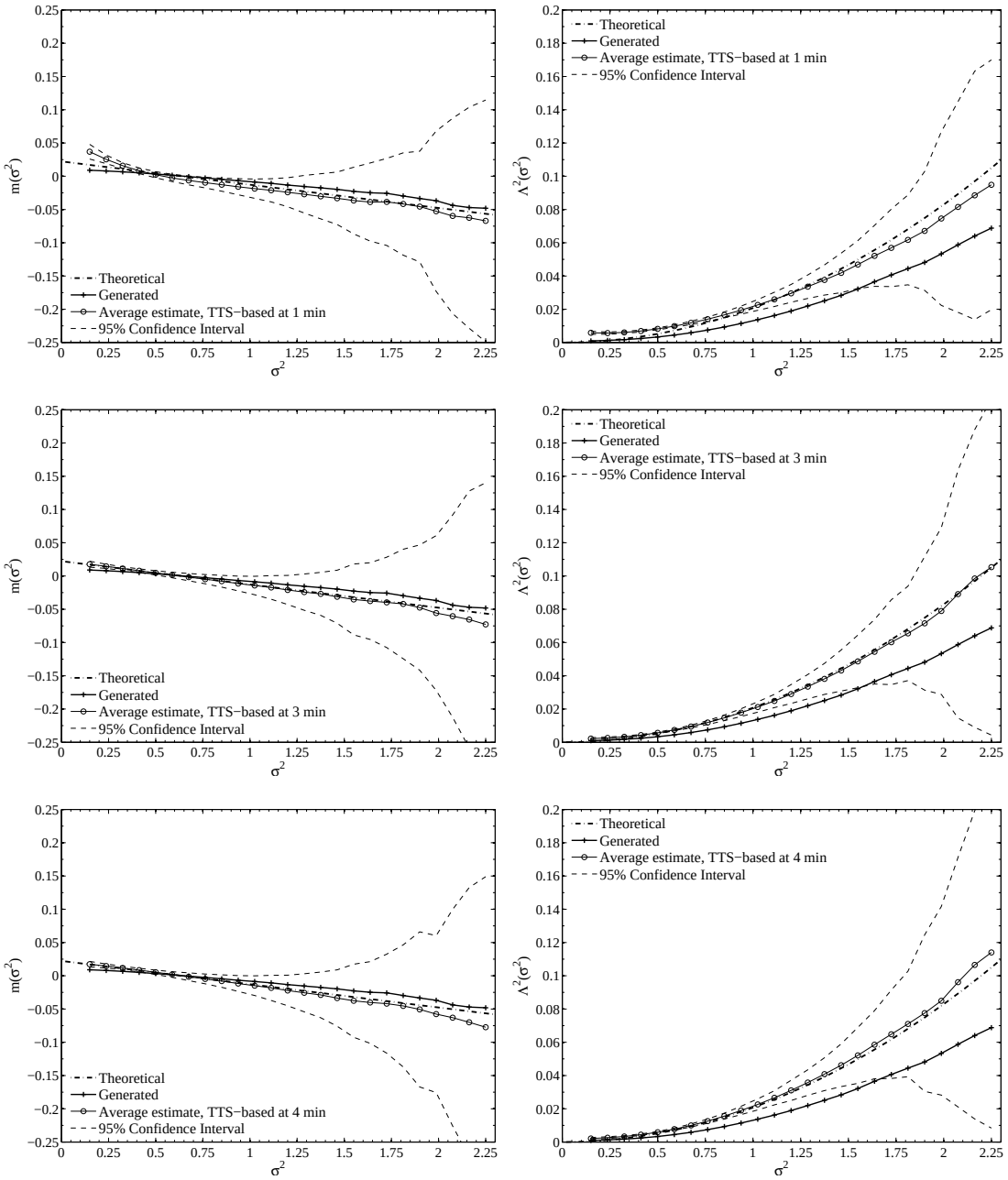

Figure D.2: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RV_{TTS} estimates of daily variance with 1, 3, and 4-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the GARCH model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

Results for the Heston model

Table D.4:
Estimators of daily variance with bid-ask bounce and infrequent trading - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	2.3888	3.4074	1.5426	1.9383	3.4967	1.8274	1.5763	1.7141	1.6976	0.3082
2	1.9628	2.6004	1.5660	1.4242	2.5996	1.0049	1.6013	1.3434	1.7009	0.3887
3	1.8220	2.2023	1.5823	1.1873	2.2996	0.8074	1.6169	1.2164	1.7009	0.4600
4	1.7573	1.9561	1.5939	1.0525	2.1496	0.7663	1.6280	1.1665	1.7010	0.5236
5	1.7219	1.7820	1.6025	0.9672	2.0603	0.7815	1.6360	1.1570	1.7011	0.5811
10	1.6639	1.3538	1.6266	0.8097	1.8805	0.9973	1.6584	1.2636	1.7015	0.8128
15	1.6522	1.1835	1.6386	0.7969	1.8212	1.2050	1.6697	1.4192	1.7015	0.9930
20	1.6488	1.1086	1.6458	0.8210	1.7893	1.3885	1.6753	1.5709	1.7012	1.1438
30	1.6492	1.0650	1.6549	0.9031	1.7599	1.6991	1.6836	1.8510	1.7006	1.3986
60	1.6551	1.1876	1.6672	1.1592	1.7275	2.3846	1.6950	2.5086	1.6998	1.9901
1440	1.6846	5.2733	1.6846	5.2733	1.6996	11.7935	1.6996	11.7935	1.6996	11.7935

Note The table shows the results of a Monte Carlo simulation estimating daily variance using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the Heston model. Sample average of the daily variance $\bar{V}_{\mathrm{Day},t} \approx 1.6998$ and $\theta = 1.6$. Daily variance has been multiplied by 10 000.

Table D.5:
Results for estimates of the drift for the case with bid-ask bounce and infrequent trading - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	1.9356	10.0620	1.9169	8.9213	3.1869	11.8788	1.8475	11.3990	1.1491	20.2654
2	1.6660	12.2776	2.0510	10.7078	2.1804	16.3120	1.8203	15.3332	1.1712	21.6553
3	1.6231	13.9871	1.7083	12.1815	1.8470	19.0856	1.2389	18.7377	1.1243	22.4597
4	1.5522	14.9623	1.3329	13.8066	1.8890	20.4551	1.1561	21.6316	1.4168	23.0155
5	1.5410	15.6610	1.1820	14.8570	1.9461	22.9794	1.3240	22.9218	1.5820	23.3158
10	1.5279	18.8681	1.2342	18.5379	2.2708	28.8541	1.8857	30.1692	1.7772	26.0837
15	1.3444	20.5494	1.0714	20.5284	2.7659	33.4955	1.9087	33.9165	2.1887	29.4392
20	1.7685	20.9391	1.0113	21.5866	3.8549	36.7994	2.8416	39.5168	2.9009	33.0415
30	1.8483	23.5986	1.2181	24.5299	5.0749	42.7305	4.0780	45.7911	4.1573	36.8162
60	2.8821	29.1266	2.3360	31.5677	10.6422	59.9666	9.6060	64.1018	7.2576	49.0062

Note The table shows the results of a Monte Carlo simulation estimating daily drift using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The results use 500 independent samples of 5 000 estimates of daily variance generated using prices from a Heston model. The bias when estimating the daily drift using the generated variance $V_{\mathrm{Day},t}$ using all data is approx. 1.3812 with RMSE 17.8848. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.5, 7.68]. Daily variance has been multiplied by 10 000.

Table D.6:
Estimators of the daily diffusion for the case with bid-ask bounce and infrequent trading - Heston

Frequency	RR_t^{Δ}		$RR_{S,t}^{\Delta}$		RV_t^{Δ}		$RV_{S,t}^{\Delta}$		$RV_{\mathrm{TTS},t}^{\Delta}$	
(minutes)	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	4.2512	4.3128	4.5356	3.9109	3.4810	3.6686	3.9176	3.4834	1.6317	3.0500
2	3.6818	3.8434	4.1039	3.7386	2.3444	2.9663	3.0597	3.1613	1.4576	3.4215
3	3.3306	3.6029	3.8321	3.6111	1.7123	2.8817	2.4232	3.0863	1.2538	3.4395
4	3.0784	3.4373	3.5859	3.4843	1.3042	2.8152	1.8481	2.8769	1.0751	3.3879
5	2.8731	3.3599	3.3687	3.3616	1.1139	4.0965	1.4406	3.0842	0.9434	3.3706
10	2.1293	3.1438	2.5708	3.0820	2.6428	5.6604	2.2449	5.7485	1.3985	4.0509
15	1.6312	2.9881	2.0377	2.8614	4.9439	7.8270	4.4957	7.1358	2.8411	5.4667
20	1.2814	3.0445	1.6348	2.9293	7.2638	10.2735	7.2241	10.6504	4.4959	7.3005
30	1.0476	3.4511	1.0990	3.2979	12.2979	15.6333	12.1026	15.4728	7.5963	10.5018
60	3.3674	5.6623	3.2554	5.6837	29.5437	35.8730	29.9242	37.1227	18.0776	21.9620

Note The table shows the results of a Monte Carlo simulation estimating daily diffusion using the realized range (RR_t^{Δ}), realized variance (RV_t^{Δ}), the two time scales realized variance ($RV_{\mathrm{TTS},t}^{\Delta}$) estimator, as well as the scaled versions of the realized range and realized variance ($RR_{S,t}^{\Delta}$ and $RV_{S,t}^{\Delta}$ resp.) with $q = 1\,000$. The experiment simulates 500 independent price paths of 6 000 days with 86 400 log prices per day (1 price per second for a 24-hour trading day) using the Heston model. The bias when estimating the daily diffusion using the generated variance $V_{\mathrm{Day},t}$ using all data is approx. 2.2087 with RMSE 3.2190. The bias and RMSE are computed with respect to the theoretical drift on the interval [0.5, 7.68]. Daily variance has been multiplied by 10 000.

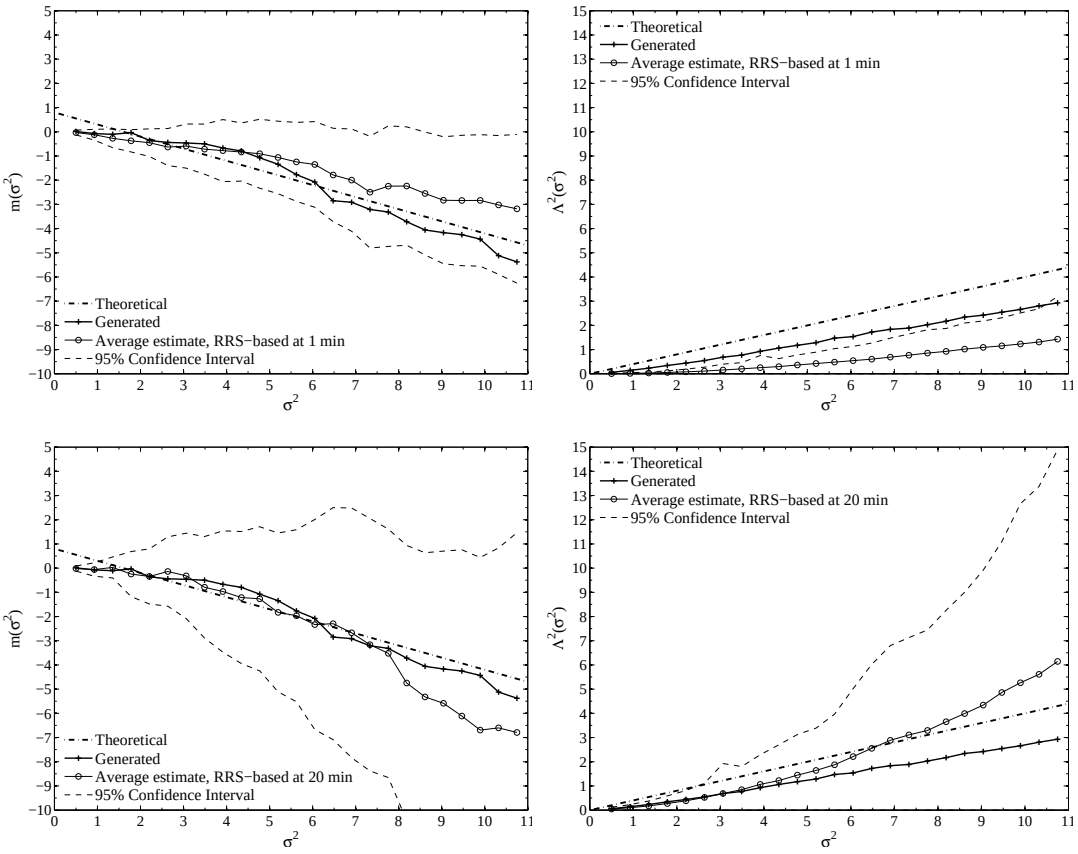

Figure D.3: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RR_S estimates of daily variance with 1-minute sampling and 20-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.

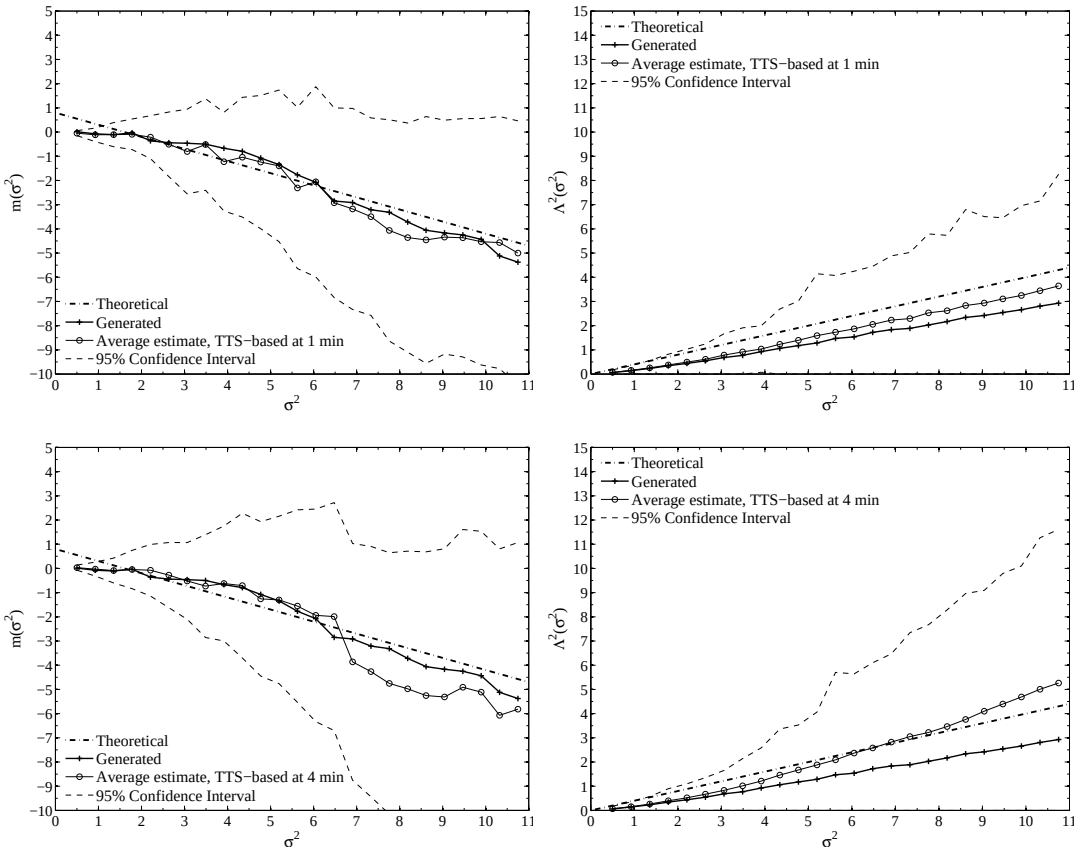

Figure D.4: Average estimates of the drift function $m(\cdot)$ (left panels) and diffusion function $\Lambda^2(\cdot)$ (right panels) using RV_{TTS} estimates of daily variance with 1-minute sampling and 4-minute sampling. Daily variance σ^2 has been multiplied by 10 000. The 'generated' graph refers to the drift estimate using the variance series generated with the Heston model. The 'theoretical' graph is obtained by evaluating the true drift function on the domain of the graph.