

Summary

When looking at prices in the stock market, it becomes clear that these prices are rather volatile. This means that prices vary over the day. Estimating the degree of variation through the day is of great importance to practitioners in the financial markets, and has therefore become a popular topic in financial econometric literature. The more prices are observed throughout the day the more accurately one can estimate the daily volatility. However, as is the problem with any estimation, we cannot say with 100% certainty that the estimated value is equal to the true value. What we can do is construct a confidence interval around our estimate. A confidence interval states that the true value lies within a specified range with a certain probability. In order to do this we need two things: we need to know the variance of our estimator and we need to know its distribution. Although the first issue can be dealt with, the second issue proves itself to be difficult.

The first approach is to derive an asymptotic distribution. In deriving this distribution, one makes the assumption that the number of returns observed within a day is infinite. By making this assumption, it can be shown that the difference between the true value and the estimated value converges to the normal distribution. This is a nice result as it is easy to interpret and work with this distribution. A property of this distribution is that it is symmetric around its mean. In reality we do not observe an infinite amount of returns within a day. When we observe only a finite number of returns, we make an error in assuming the difference between the true value and the estimator is normally distributed. In case of a finite number of observations the true distribution is asymmetric as opposed to a symmetric one like the normal distribution.

To deal with this issue the idea of applying the bootstrap technique was introduced. When performing a bootstrap, one takes the original sample on which the original estimator is based, and create a new sample (the bootstrap sample). Using the bootstrapped sample we can re-estimate the statistic of interest (bootstrapped estimate). There are many ways to con-

struct the bootstrap sample. The i.i.d. bootstrap and the wild bootstrap are considered in my bachelor thesis. With the i.i.d. bootstrap the new sample is created by randomly picking observations from the original sample, allowing for the same observation to be picked multiple times. The wild bootstrap constructs a new sample by multiplying the observations of the original sample by a random variables. Different bootstrapping methods have different properties and it is therefore interesting to consider more than just one.

Analogously to the difference between the true value and the estimator, it can be shown that the difference between the original estimate and the bootstrapped one also converges to a normal distribution when the number of observations goes to infinity. Again, this only holds in case the number of observations is infinite. The nice thing about the bootstrap, however, is that it is created by the investigator and can be done as many times as desired. This means that even in the finite case, although analytically unknown, the distribution of the difference between the bootstrapped estimate and the original estimate can be generated by the investigator.

This is not possible for the difference between the original estimator and the true value, as there is only one sample is observed in reality. As mentioned before when we assume a normal distribution when constructing a confidence interval for the difference between the estimator and the true value we make an error. If rather than assuming a normal distribution, we assume it has the same distribution as the difference between the bootstrapped estimate and the original one, under the right conditions the error made becomes smaller. If this is the case we can improve our confidence interval by making use of the bootstrap method.

The whole story above is very much dependent on the assumptions made on the model that the stock prices follow. The goal of my bachelor thesis was to investigate the effect of assuming an intra-day periodicity factor in the volatility of the stock price on the ability of the bootstrap methods to provide better confidence intervals.

4.1 Introduction

The availability of intra-day return data has given rise to statistics such as Realized Volatility (RV). Instead of measuring the daily volatility by the difference in opening and closing prices, the true volatility path could be investigated in more detail. Under certain conditions RV has been shown to be a consistent estimator of the Integrated Volatility (IV), a quantity of great importance to modelling daily returns using GARCH-like models. In order to draw inference on this quantity, a normal asymptotic theory for RV was derived by Barndorff-Nielsen and Shephard (2002). In small samples the true distribution of the standardized RV is quite skewed, which leads to an under-coverage of the true IV by confidence intervals based on normal asymptotic theory. Goncalves and Meddahi (2009) therefore proposed the bootstrapping method. Using Edgeworth expansions and Monte Carlo simulations they showed that the bootstrap error in estimating the true distribution of the standardized RV is smaller than the error of the normal asymptotic approximation. In this paper, we want to look at the effect of including an intra-day periodicity factor in the volatility term. We find the bootstrapping methods to provide better confidence intervals than the normal asymptotic theory, also when a periodicity component is included in the underlying price model. Estimating the periodicity component first, and filtering it out of the returns, leads to an improvement in our results. In section 2 the relevant literature is reviewed. In section 3 the simulation process and the construction of confidence intervals are described. In section 4 we will discuss the results of the Monte Carlo simulation, and in the last section a conclusion is given.

4.2 Literature Review

4.2.1 Underlying Model and Realized Volatility

In most literature related to volatility estimation the following continuous time model is assumed for the log price:

$$d \log S_t = \mu_t dt + v_t dW_t \quad (4.1)$$

Here $\log S_t$ denotes the log of the price level, μ_t is a deterministic trend component, and v_t is a time-varying stochastic volatility process. W_t denotes a standard Wiener process. At high data frequencies the drift term becomes negligible, as its order of dt is much smaller than the order of the volatility term. This term will be excluded such that the focus goes to the

effect of the extension to this model we want to investigate later on. With the exclusion of the drift term equation (1) is simplified to:

$$d \log S_t = v_t dW_t \quad (4.2)$$

Lets assume returns are observed over equal spaced time intervals of size Δ . The i^{th} return observed is then equal to:

$$r_i \equiv \log S_i - \log S_{i-\Delta} = \int_{i-\Delta}^i v_u dW_u$$

In case of a time-invariant volatility factor, $v_t = v, \forall t \geq 0$, this return can be rewritten as:

$$r_i = \int_{i-\Delta}^i v dW_u = v(W_i - W_{i-\Delta}) \sim N(0, \Delta v^2)$$

When the volatility process is time varying, conditional on the volatility path, the returns over Δ are heteroskedastic and can be expressed as $r_i \sim i.i.d. N(0, \sigma_i^2)$, where $\sigma_i^2 = \int_{i-\Delta}^i v_u^2 du$ for $i = 1, \dots, 1/\Delta$.

The quantity of interest in this paper is the daily volatility. As the volatility component in the underlying model can be time-variant we need to know the path of this volatility component throughout the day if we want to estimate the true volatility. This quantity is referred to as the Integrated Volatility:

$$IV = \int_0^1 v_u^2 du \quad (4.3)$$

As v_t is not observed and therefore IV can not directly be computed, we need to find a consistent estimator. One could simply take the difference between the first and the last observed price level. However, this would be a naive estimation as it would completely disregard the volatility dynamics during the day. Andersen, Bollerslev, Diebold and Labys and Barndorff-Nielsen and Shephard (2002, 2001) have shown that Realized Volatility is a consistent and efficient estimator of IV:

$$RV = \sum_{i=1}^{1/\Delta} r_i^2 \quad (4.4)$$

$$E(RV) = \sum_{i=1}^{1/\Delta} E(r_i^2) = \sum_{i=1}^{1/\Delta} E(\sigma_i^2 u_i^2) = \sum_{i=1}^{1/\Delta} \int_{i-\Delta}^i v_u^2 du = IV$$

4.2.2 Intra-day Periodicity

An extension to the basic model we would like to investigate is the inclusion of an intra-day periodicity factor. Andersen and Bollerslev (1997) investigated intra-day return behaviour of both exchange rate and price levels. They used 5-minute-interval data over 1 year for the Deutschemark-U.S.Dollar exchange rate and 5-minute-interval returns for Standard and Poor's composite index futures contract over a period of three years. They looked at the means of the returns during each of the intervals, as well as their respective variances. Although there appeared to be no systematic appreciation or depreciation pattern in the returns, the return volatility did seem to exhibit a pattern. Taking the absolute value of the returns the pattern became more clear. The ACF of the absolute returns also confirmed the suspicion of a periodicity in the returns. These patterns were different for the exchange rate and the futures contract data.

Although in the study of Andersen and Bollerslev, the periodicity did not seem to differ much on the different days of the week. For some data slight differences have been found in the pattern. For the future contracts the data was taken from one specific organized market which trades during specific opening and closing times. This is not the same for the exchange rates, they are traded during the whole 24 hour period of the day. The future contract absolute returns exhibited a U- shaped to a J- shaped function. Implying that volatility reaches its highest after opening and before closing, and its lowest during the lunch period of the day. For the exchange rates, volatility has been found to be highly correlated with trading volumes. This results into a sinusoid-like function with its peaks depending on the opening and closing times of the major financial markets across the world. Andersen and Bollerslev proposed to implement the periodicity into the basic model by redefining the volatility term:

$$\sigma_i = s_i f_i$$

For identification purposes the following normalization is assumed on the periodicity factor:

$$\Delta \sum_{i=1}^{1/\Delta} f_i^2 = 1$$

Two different approaches will be looked at when constructing confidence intervals for IV. In the first approach the normal returns are taken to compute the relevant statistics. In the second approach the periodicity factor is estimated and filtered out of the returns. The filtered returns are then

used to compute the relevant statistics. Therefore in the first approach the periodicity is dealt with in an indirect way by relying on the bootstrapping method to capture its effect on the distribution of the standardized RV. The second approach is a more direct approach as the periodicity has been mostly filtered out before the bootstrapping method has been applied. The way f_i is simulated for the underlying process is discussed in section 3. The estimation techniques will be described here.

The intra-day periodicity component can be estimated both in a parametric and a non-parametric way. The method applied in our Monte Carlo simulations is a simple non-parametric estimation. The reasoning behind the method is that during a specific local window, most of the volatility in the returns can be accounted for by the periodicity. If we define $\bar{r}_i = r_i/\hat{s}_i$, where \hat{s}_i is an estimate of s_i , then $\bar{r}_i \sim N(0, f_i)$. Andersen and Bollerslev assume that in the local window $s_t = s \forall t \geq 0$ and estimate $\hat{s} = \sqrt{\Delta RV}$. If one has several days of data with a common periodicity, one can use the standard deviation of \bar{r}_i for each i to compute f_i :

$$\hat{f}_i = \frac{SD_i}{\sqrt{\Delta \sum_{i=1}^{1/\Delta} SD_i^2}}$$

where $SD_i = \sqrt{\frac{1}{n} \sum_{j=1}^n r_{i,j}^2}$ and n denotes the number of days observed.

This estimator is very similar to the one used by Taylor and Xu (1997) If the data allows it, \hat{f}_i could be computed for each day of the week separately. In the presence of jumps, using the standard deviation this estimator becomes inconsistent. Boudt, Croix, and Laurent (2008) propose the use of the Median Absolute Deviation (MAD) instead of the standard deviation. As the normal asymptotic theory presented later does not hold in presence of jumps, we do not consider this scenario. Although s_t is assumed to be constant in this estimation, it provides a decent estimation even in case of time varying volatility.

An alternative approach is the parametric one. Here we once again rely on the assumption made by Andersen and Bollerslev (1998b) that $r_i/\hat{s} \approx f_i u_i$ with $u_i \sim N(0,1)$. Under this assumption $\log(\frac{|r_i|}{\hat{s}}) \approx \log(f_i) + \log(|u_i|)$, which allows f_i to be isolated as follows:

$$\log(\frac{|r_i|}{\hat{s}}) - c = \log(f_i) + \epsilon_i$$

where c equals the expected value of the log of the absolute value of a standard normal variable and ϵ_i is i.i.d. distributed with mean 0 and has

the density function of the centred absolute value of the log of a standard normal variable. $\log f_i$ can then be linearly estimated by $x_i' \theta$ where x_i is a vector of variables such as sinusoids and polynomial transformations of the time of the day, and θ the respective estimated coefficients. The OLS and ML estimates are not robust to the presence of jumps, which led to the proposal of the Truncated Maximum Likelihood estimator by Boudt, Croix, and Laurent (2008), which gives zero weight to observations suspicious of being affected by jumps. Boudt, Croix, and Laurent have shown that the parametric estimators outperform the non-parametric ones. For investigating the difference between the approaches of filtering and not filtering out the returns when constructing confidence intervals, the simple non-parametric estimator will suffice.

4.2.3 Normal Asymptotic Theory and the Bootstrapping Methods

Our goal is to check the ability of the bootstrapping methods to refine confidence intervals based on normal asymptotic theory in the presence of intra-day periodicity. The CLT is based on results of Jacod (1994), Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (BN-S)(2002)):

$$S_{\Delta} = \frac{RV - IV}{\sqrt{\Delta V}} \rightarrow^d N(0, 1) \quad (4.5)$$

where

$$V = 2 \int_0^1 v_u^4 du$$

Here S_{Δ} is the standardized value of RV, which, as $\Delta \rightarrow 0$, converges in distribution to a standard normal variable. S_{Δ} can not be computed in practice as V is unknown. BN-S show that if V is replaced by the consistent estimator \hat{V} the CLT still holds:

$$T_{\Delta} = \frac{RV - IV}{\sqrt{\Delta \hat{V}}} \rightarrow^d N(0, 1) \quad (4.6)$$

where

$$\hat{V} = \frac{2}{3} \sum_{i=1}^{1/\Delta} r_i^4$$

As the name gives away, the normal asymptotic theory works nicely for constructing confidence intervals in an asymptotic scenario, but in small samples the distribution of T_Δ is not standard normal. Its left-side tail is much fatter than the right-side, and its mean is slightly negative rather than 0. Using the CLT means assuming a mean of 0, and symmetry; two conditions that are both violated. Goncalves and Meddahi (2009) therefore propose the use of the bootstrapping method. They show using Edgeworth Expansions and Monte Carlo simulations that the bootstrap error in estimating the distribution of T_Δ is smaller than the error in assuming the distribution of T_Δ to be standard normal.

Goncalves and Meddahi considered two types of bootstrapping methods: the i.i.d. bootstrap and the wild bootstrap. In the i.i.d. bootstrap returns are re-sampled with replacement. This method will work best when returns are i.i.d. distributed, which is the case if the volatility factor is constant over time. Even when volatility is time-varying and returns are not i.i.d. distributed any more, Goncalves and Meddahi show that this bootstrap still provides a refinement over the standard normal approximation. The bootstrapped returns denoted by $r_i^* = r_{I_i}$ with $I_i \sim$ i.i.d. uniform on $\{1, 2, \dots, 1/\Delta\}$. The bootstrap analogue of RV is equal to:

$$RV^* = \sum_{i=1}^{1/\Delta} r_i^{*2} \quad (4.7)$$

Goncalves and Meddahi construct a standardization RV^* using the RV of the original data:

$$T_\Delta^* = \frac{RV^* - RV}{\sqrt{\Delta \hat{V}^*}} \quad (4.8)$$

Although the true variance is known in this case - $V^* = R_4 - RV^2$ - an estimator based on the bootstrapped data is used: $\hat{V}^* = R_4^* - RV^{*2}$. This is because we are trying to approximate the distribution of T_Δ , rather than to perform the best inference possible on RV. Under certain conditions Goncalves and Meddahi show that as $\Delta \rightarrow 0$:

$$\sup_{x \in \mathbf{R}} |P^*(T_\Delta^* \leq x) - \Phi(x)| \xrightarrow{P} 0 \quad (4.9)$$

Combining equations (6) and (9) we get that $P^*(T_\Delta^* \leq x) - P(T_\Delta \leq x) = o_P(1)$, which is a first order validation for the use of the bootstrapping method.

The second bootstrapping method considered is the wild bootstrap. The returns of the original sample are multiplied by an i.i.d random variable: $r_i^* = r_i \eta_i$ and η_i is chosen externally. Let $\mu_q^* = E^*(|\eta_i|^q)$ and the bootstrapped RV be defined as $RV^* = \sum_{i=1}^{1/\Delta} r_i^*$. Then the statistics for the wild bootstrap are:

$$T_{\Delta}^* = \frac{RV^* - \mu_2^* RV}{\sqrt{\Delta \hat{V}^*}} \tag{4.10}$$

where

$$\hat{V}^* = \frac{\mu_4 - \mu_2^2}{\mu_4} R_4^*$$

here \hat{V}^* again is a consistent estimator of the true conditional variance. Equation (9) also holds for the wild bootstrap, and therefore combined with equation (6) gives a first order asymptotic validity for the use of this bootstrap. This result does not depend on the choice of η_i as long as R^* is carefully standardized. The distribution of η_i does, however, affect the ability to provide second order refinements.

4.2.4 Edgeworth Expansions

Goncalves and Meddahi take two different approaches in investigating the refinement ability of the bootstrapping method. One method is using Monte Carlo simulations. This is the approach we take here as well, and is described in sections 3 and 4. However, the theoretical justification for the use of the bootstrap method is provided by Goncalves and Meddahi using Edgeworth Expansions. They consider the formal one-term expansions which are defined as follows for the normal asymptotic approximation:

$$P(T_{\Delta} \leq x) = \Phi(x) + \sqrt{\Delta} q_1(x) \phi(x) + O(\Delta) \tag{4.11}$$

uniformly in $x \in \mathbf{R}$, and where $\Phi(x)$ denotes the cdf of a standard normal variable and $\phi(x)$ its density function. q_1 is a function of x whose coefficients depends on the cumulants of T_{Δ} :

$$q_1(x) = -(\kappa_1 + \frac{1}{6} \kappa_3 (x^2 - 1)) \tag{4.12}$$

where κ_1 and κ_3 are the leading terms of the first and third order cumulants of T_{Δ} . The formal first term Edgeworth expansion for the bootstrap distribution, T_h , is of the same form:

$$P^*(T_\Delta^* \leq x) = \Phi(x) + \sqrt{\Delta}q_1^*(x)\phi(x) + O(\Delta) \quad (4.13)$$

where

$$q_1^*(x) = -(\kappa_1^* + \frac{1}{6}\kappa_3^*(x^2 - 1)) \quad (4.14)$$

Goncalves and Meddahi show that, conditional on v , as $\Delta \rightarrow 0$:

- $q_1(x) = \frac{4(2x^2+1)}{6\sqrt{2}} \frac{\bar{\sigma}^6}{\bar{\sigma}^4{}^{3/2}}$
- $q_1^{i.i.d.} = \frac{1}{6}(2x^2 + 1) \frac{R_6 - 3R_4RV + 2RV^3}{(R_4 - RV^2)^{3/2}}$
- $q_1^{wild} = -(-\frac{A_1^*}{2} + \frac{1}{2}(B_1^* - 3A_1^*)(x^2 - 1)) \frac{R_6}{R_4^{3/2}}$,
 where $A_1^* = \frac{\mu_6^* - \mu_2^*\mu_4^*}{\mu_4^*(\mu_4^* - \mu_2^{*2})^{1/2}}$ and $B_1^* = \frac{\mu_6^* - 3\mu_2^*\mu_4^* + 2\mu_2^{*3}}{(\mu_4^* - \mu_2^{*2})^{3/2}}$

Here $\bar{\sigma}^q = \int_0^1 v_u^q du$.

Using equation (11) we can define the error made by the normal asymptotic approach as:

$$P(T_\Delta \leq x) - \Phi(x) = \sqrt{\Delta}q_1(x)\phi(x) + O(\Delta) \quad (4.15)$$

uniformly in $x \in \mathbf{R}$. In a similar way the error made in approximating the distribution of T_Δ using the bootstrap can be defined as:

$$P^*(T_\Delta^* \leq x) - P(T_\Delta \leq x) = \sqrt{\Delta}(q_1^* - q_1(x))\phi(x) + O_P(\Delta) \quad (4.16)$$

Note that the error made in relying on asymptotic normality is a function of $q_1(x)$, whereas the bootstrap error is a function of the difference between $q_1^*(x)$ and $q_1(x)$. Taking the plim of $q_1^*(x)$, the right hand side of the equality sign of equation (16) converges to:

$$\sqrt{\Delta}(\text{plim}_{\Delta \rightarrow 0} q_1^*(x) - q_1(x))\phi(x) + o_P(\sqrt{h})$$

This implies that if $\text{plim}_{\Delta \rightarrow 0} q_1^*(x) - q_1(x) = 0$, i.e. the bootstrapping method can perfectly match the coefficients of the first three cumulants of T_Δ^* to those of T_Δ to order $O(\sqrt{\Delta})$, the first term in error becomes zero, and thus the bootstrapping error is of order $o_P(\sqrt{\Delta})$ which is smaller than

the normal asymptotic error of order $O(\Delta)$.

For the i.i.d. bootstrap they found the following properties as $\Delta \rightarrow 0$, conditional on the path of v_t :

- $plim_{\Delta \rightarrow 0} q_1^*(x) - q_1(x) = \frac{1}{6}(2x^2 + 1) \left[\frac{15\bar{\sigma}^6 - 9\bar{\sigma}^4\bar{\sigma}^2 + 2(\bar{\sigma}^2)^3}{(3\bar{\sigma}^4 - (\bar{\sigma}^2)^2)^{3/2}} - \frac{4}{\sqrt{2}} \frac{\bar{\sigma}^6}{(\bar{\sigma}^4)^{3/2}} \right]$
- when $v_t = v \forall t$ then: $plim_{\Delta \rightarrow 0} q_1^*(x) - q_1(x) = 0$
- In the general case:

$$|plim_{\Delta \rightarrow 0} q_1^*(x) - q_1(x)| \leq |q_1(x)|$$

The second point shows that when volatility is time-invariant the difference goes to 0 and hence the bootstrap method provides an asymptotic refinement through order $O(\sqrt{\Delta})$ over the normal asymptotic approximation. This result relies on the fact that when volatility is constant $\bar{\sigma}^q = (\bar{\sigma})^q$. However, this does not hold in case of time-varying volatility. The rate of convergence of the bootstrap error is then of order $O_P(\sqrt{\Delta})$ which is the same as the order of the normal asymptotic error. This contradicts the results found in their Monte Carlo simulations. They argue that to order $O(\sqrt{\Delta})$ the magnitude of normal approximation error is of $q_1(x)$ and the magnitude of the bootstrap error equal to $plim_{\Delta \rightarrow 0} q_1^*(x) - q_1(x)$. As a consequence of the last point made above:

$$\frac{|plim_{\Delta \rightarrow 0} q_1^*(x) - q_1(x)|}{|q_1(x)|} \equiv r_1(x) \leq 1$$

The inequality above states that the bootstrap error is always smaller than the normal approximation error. In a similar way it can be shown that the difference between the quantiles of the distributions T_Δ and T_Δ^* is smaller than the difference between the quantiles of the distribution of T_Δ and the standard normal distribution.

For the wild bootstrap the error is defined as:

- $plim_{\Delta \rightarrow 0} q_1^*(x) - q_1(x) = - \left[(plim_{\Delta \rightarrow 0} \kappa_1^*(x) - \kappa_1(x)) + \frac{1}{6} (plim_{\Delta \rightarrow 0} \kappa_3^*(x) - \kappa_3(x))(x^2 - 1) \right]$
- $plim_{\Delta \rightarrow 0} \kappa_1^*(x) - \kappa_1(x) = -\frac{1}{2} \frac{\bar{\sigma}^6}{(\bar{\sigma}^4)^{3/2}} \left(\frac{5}{\sqrt{3}} A_1^* - A_1 \right)$
- $plim_{\Delta \rightarrow 0} \kappa_3^*(x) - \kappa_3(x) = \frac{\bar{\sigma}^6}{(\bar{\sigma}^4)^{3/2}} \left[\left(\frac{5}{\sqrt{3}} B_1^* - B_1 \right) - 3 \left(\frac{5}{\sqrt{3}} A_1^* - A_1 \right) \right]$

Here $A_1 = B_1 = \frac{4}{\sqrt{2}}$. Again the ability to provide an asymptotic refinement relies on the ability of the bootstrap to match the cumulants of T_Δ . Matching the cumulants is done by choosing the distribution of η_i . In our simulations we chose $\eta_i \sim N(0,1)$. Goncalves and Meddahi actually showed that by choosing the standard normal distribution $\text{plim}_{\Delta \rightarrow 0} q_1^*(x) - q_1(x) = (\frac{5}{\sqrt{3}-1})q_1(x) \approx 1.89q_1(x)$. This implies that the wild bootstrap error is actually bigger than the normal asymptotic approximation. This is confirmed by the Monte Carlo simulations for both the one-sided and symmetric two sided confidence interval. The $N(0,1)$ -wild bootstrap significantly overestimates these types of confidence intervals. In case of equal-tailed confidence intervals the $N(0,1)$ bootstrap does seem to provide an improvement over the normal asymptotic approximation. This is also the type of confidence interval we investigate here, and choosing η_i to be standard normally distributed we do find better coverage rates compared to the CLT. Instead of the standard normal distribution they suggest to chose a Bernoulli distribution for η_i as this allows for the cumulants to be completely matched to order $O(\Delta)$. However, in the case of equal tailed confidence intervals, for low number of observations the Monte Carlo simulations show the $N(0,1)$ -bootstrap to provide better results than the matched-Bernoulli-bootstrap.

4.3 Simulation

As the underlying model and the bootstrapping methods have been reviewed, we move on to the simulation process. The idea is to test the ability of the bootstrap to provide higher order refinements when the underlying model for the price is extended to incorporate intra-day periodicity. In this section the simulation of the various underlying processes of interest are discussed as well the construction of confidence intervals.

Three different models are investigated. The first is a model with constant volatility and no intra-day periodicity. The results of this model will be used as a benchmark for the other models. In the second model the volatility term consists of a constant volatility component and a deterministic periodicity factor. The last model investigated includes both an intra-day periodicity and a time-varying volatility component. The price simulation process uses a standard Euler discretization scheme. For all models the number of underlying returns is 3456. This is a nice multiple of the number of returns which are actually observed. The number of actual observations

can equal: 12, 48, 96, 288, 576, and 1152. For each model 3456 underlying returns are simulated and these returns are used for all frequencies combined with all IC construction methods (CLT, i.i.d , and wild). For example, in case 12 returns are observed, each of these returns is based on $3456/12 = 288$ underlying returns. This makes it easier to compare different methods and different data frequencies. For the last two models, which both include intra-day periodicity, the filtered as well as the non-filtered returns are used. As the underlying process does not include a jump component, we use the non-parametric filtering method based on the standard deviation of \bar{r}_i .

In order to incorporate a periodicity factor in line with empirical findings, we follow Hecq, Laurent, and Palm (2012) in their Monte Carlo simulations and assume f_i to be the sum of various sinusoids over the period of 1 day:

$$\log f_i^* = \sum_{l=1}^4 \gamma_{2l-1} \cos\left(\frac{i2\pi l}{1/\Delta}\right) + \sum_{l=1}^4 \gamma_{2l} \sin\left(\frac{i2\pi l}{1/\Delta}\right) \quad (4.17)$$

where $\gamma = (-0.24422, -0.49756, -0.054171, 0.073907, -0.26098, 0.32408, -0.11591, -0.21442)$. This leads to a function of the time of the day with three peaks corresponding to the activity on the three main regions across the world.

When the volatility component is time-varying, it follows the GARCH(1,1) diffusion model studied in Andersen and Bollerslev (1998a):

$$dv_t^2 = 0.035(0.636 - v_t^2)dt + 0.144v_t^2 dW_{1t}$$

Note that here W_{1t} is assumed to be independent of W_t in equations (1) and (2). Goncalves and Meddahi did consider the possibility of a dependence between the Wiener processes, i.e. leverage effect, in their Monte Carlo simulations. The results found for the inclusion of this extension were not significantly different from the results of the baseline model.

Relying on the normal asymptotic theory the 95% two sided confidence interval for IV is given by:

$$IC_{0.95}^{(2)} = \left(RV - z_{0.025} \sqrt{\Delta \hat{V}}, RV + z_{0.975} \sqrt{\Delta \hat{V}} \right)$$

where the z-values denote the quantiles of the standard normal distribution. As Goncalves and Meddahi argued, the distribution of T_Δ is not standard normal when the number of observations is low, and therefore the IC given above will not be able to cover 95% of the IV. It is not possible to derive the true distribution of T_Δ along with its quantiles, As the true IV is unknown, because the true IV is not known. However, the bootstrap T_Δ^* can be simulated, since both the bootstrapped statistics as well as RV itself are known. Goncalves and Meddahi showed using Edgeworth expansions that using the quantiles of T_Δ^* should give better coverage rates. The bootstrap confidence interval is given by:

$$IC_{0.95}^{(2)*} = \left(RV - q_{0.025}^* \sqrt{\Delta \hat{V}}, RV + q_{0.975}^* \sqrt{\Delta \hat{V}} \right)$$

where q_α^* is the α -quantile of the distribution of T_Δ^* .

We are only looking at equal tailed bootstrap intervals here. Goncalves and Meddahi also considered symmetric bootstrap intervals, where instead of using the 2.5%- and 97.5%- quantiles of the distribution of T_Δ^* , the 95%-quantile of the distribution of $|T_\Delta^*|$ is used. Due to the asymmetry of the distribution of T_Δ the symmetric intervals underperform compared to the equal-tailed ones. Only equal-tailed IC are therefore considered here. Goncalves and Meddahi also considered the log transform of the of the confidence interval. For the IC based on the CLT and for the symmetric bootstrap IC, this led to a significant improvement in the coverage rates. However, this was not the case for equal tailed IC.

4.4 Results

In this section the results of the Monte Carlo simulations are discussed. The results for all models are given in table 1. Note that table 1 reports the simulations using raw returns, instead of filtered. Before we move on to the models with intra-day periodicity, we first take a look at the benchmark model. The results found for the constant-volatility model are in line with theory and the results found by Goncalves and Meddahi (2009). Relying on normal asymptotic theory leads to an under-coverage at lower frequencies, but as the number of intra-day observations increases the coverage rate converges to 95%. Both bootstrapping methods provide a refinement over the normal asymptotic theory. In fact, even at low observation frequencies, the coverage rate already lies around 95%. The results for the i.i.d. bootstrap are in line with the theory that when $v_t = v \forall t$, $\text{plim}_{\Delta \rightarrow 0} q_1^*(x) - q_1(x) = 0$

and thus the i.i.d. bootstrap provides an asymptotic refinement over the normal asymptotic approximation. According to the theory the wild bootstrap should perform worse in comparison to the i.i.d. bootstrap, but this is not the case.

Next to the results of the benchmark model, the coverage rates for the model with a constant volatility component and intra-day periodicity are given. As expected, due to the time-varying periodicity factor the returns become heteroskedastic, which causes the rate of convergence to a standard normal distribution of T_Δ to become slower. However, the convergence still takes place. Due to the inclusion of a periodicity factor, although still deterministic, v_t becomes time-varying. The i.i.d. bootstrap does not provide an asymptotic refinement through order $O(\sqrt{\Delta})$ over the normal asymptotic approximation any more. However, as Goncalves and Meddahi have shown, the magnitude of the first term of bootstrap error is always smaller than the magnitude of the first term of the normal approximation error. As a consequence the results of the i.i.d. bootstrap are slightly worse compared to the benchmark model, but still indicate a refinement over normal asymptotic theory. When v_t is time-varying the wild bootstrap would seem like a natural choice. However, the theory suggests when the external variable is chosen to be standard normally distributed, the i.i.d. bootstrap outperforms the wild bootstrap. This is confirmed by our results. We find that even for the choice of the standard normal distribution, the wild bootstrap does provide a refinement over the normal asymptotic theory.

The last model in table 1 is the one where also the volatility component s_t becomes time-varying. Although both components of the volatility factor are now time-varying, in general the results do not deteriorate much. This could be explained by the fact that the variance of the periodicity factor is much higher than the conditional variance of the stochastic volatility term. Thus the bootstrapping methods are still able to improve upon the asymptotic theory and the i.i.d. outperforms the wild bootstrap.

In table 2, the results for the filtered returns are presented. Comparing the results of the constant volatility without periodicity based on raw returns with the constant volatility model with intra-day-periodicity based on the filtered returns, the results are strikingly similar. This is expected, as according to the assumption made by Andersen and Bollerslev, the filtered returns are now approximately i.i.d. normally distributed with variance s^2 . As the returns have been filtered using an estimate of the true periodicity factor f_i , some degree of heteroskedasticity remains, but very little. As the filtered

returns are almost i.i.d. distributed, the rate of convergence to a standard normal distributed variable of T_{Δ} should be similar between both models. This is confirmed by the results presented in tables 1 and 2. For the filtered returns, the results for the models with and without a time-varying volatility component are very similar. This is in accordance with the fact that the conditional variance of the stochastic volatility component is relatively small compared to the variance of the periodicity factor in our simulation.

4.5 Conclusion

In this paper literature related to bootstrapping realized volatility and intra-day periodicity has been reviewed. A Monte Carlo simulation has been performed to investigate the impact of intra-day periodicity on the normal asymptotic theory and the ability of the bootstrapping methods to provide a refinement. When using raw returns to compute the relevant statistic, we found the periodicity factor to have a significant impact on the coverage rates. Although the volatility component might not be stochastic and constant over time, when there is periodicity present in the volatility term, the returns are not i.i.d. distributed any more. This has an impact on both the normal asymptotic approximation as well as the bootstrapping methods. In addition to using raw returns, filtered returns have been used as well. This led to an improvement in coverage rates for both models with and without a time-varying volatility component. Filtering the returns leads to a reduction in the degree of heteroskedasticity. In general we found the i.i.d. bootstrap to provide better results than the wild bootstrap. This is in line with the theory derived by Goncalves and Meddahi. The wild bootstrap on the other hand performs better than suggested by their study.

Note that although the Monte Carlo simulations suggest the bootstraps to perform well, even in presence of intra-day periodicity, problems might occur when implementing them in practice. None of the results presented above are robust to the inclusion of a jump component in the underlying process. In the presence of jumps, Realized Volatility is not a consistent estimator of IV any more and therefore one can not rely on this estimator. Although asymptotic theories have been derived for consistent estimators, the asymptotic theories themselves are derived under the null of no jumps. The distribution of the difference between the estimator and the IV will depend on the jump size and jump frequency and does not have to be asymptotically normally distributed. As a lot of empirical research has shown the presence of jumps in return data, this is a problem of serious nature for confidence interval estimation.

4.6 Tables

Table 1: Equal Tailed CI using raw returns

	CV			CV + Per.			GARCH(1,1) + Per.		
	CLT	i.i.d.	wild	CLT	i.i.d.	wild	CLT	i.i.d.	wild
12	85.82	95.46	95.28	79.42	93.00	94.70	79.60	92.14	94.04
48	92.38	95.30	94.54	89.10	94.34	93.38	88.08	94.04	93.68
96	93.22	95.18	94.94	91.58	95.00	94.18	90.84	94.92	94.10
288	94.50	95.10	95.48	93.86	95.00	94.32	92.82	94.08	93.52
576	94.60	94.84	95.04	94.46	94.98	94.68	93.18	94.04	93.74
1152	94.92	95.30	95.18	94.56	95.02	95.04	92.26	94.31	93.08

Table 2: Equal Tailed CI using filtered returns

	CV + Per.			GARCH(1,1) + Per.		
	CLT	i.i.d.	wild	CLT	i.i.d.	wild
12	85.90	94.44	94.54	86.36	95.46	93.90
48	92.64	95.60	93.96	92.18	95.36	94.66
96	93.36	95.56	94.06	93.70	95.36	94.88
288	94.90	95.56	95.20	94.78	95.32	94.78
576	94.72	95.02	94.80	94.50	94.76	94.22
1152	94.52	94.56	94.42	95.28	95.62	95.60

Bibliography

[Andersen, T. G., and T. Bollerslev (1997)] Andersen, T. G., and T. Bollerslev, Intraday Periodicity and Volatility Persistence in Financial Markets, *Journal of Empirical Finance*, 4, 115–158, 1997.

[Andersen, T. G., and T. Bollerslev (1998a)] Andersen, T. G., and T. Bollerslev, Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, *International Economic Review*, 39, 885–905, 1998.

[Andersen, T. G., and T. Bollerslev (1998b)] Andersen, T. G., and T. Bollerslev, DM-Dollar volatility: intraday activity patterns, macroeconomic announcements and longer run dependencies, *The Journal of Finance*, 53, 219–265, 1998.

[Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001)] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys, The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Association*, 96, 42–55, 2001.

[Barndorff-Nielsen, O.E. and N. Shephard (2001)] Barndorff-Nielsen, O.E. and N. Shephard, Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics, (with discussion), *Journal of the Royal Statistical Society, Series B*, 63, 167–241, 2001.

[Barndorff-Nielsen, O. and N. Shephard, (2002)] Barndorff-Nielsen, O. and N. Shephard, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society, Series B*, 64, 253–280, 2002.

[Boudt, K., C. Croux, and S. Laurent (2008)] Boudt, K., C. Croux, and S. Laurent, Robust estimation of intraweek periodicity in volatility and jump detection, *Mimeo*, 2008.

[Goncalves, S. and Meddahi, N. (2009)] Goncalves, S. and Meddahi, N., Bootstrapping Realized Volatility, *Econometrica*, Volume 77, Issue 1, pages 283–306, 2009.

[Hecq, A., Laurent, S. and Palm, F. (2012)] Hecq, A., Laurent, S. and Palm, F., "Common Intraday Periodicity," *Journal of Financial Econometrics*, Oxford University Press, vol. 10(2), pages 325–353, 2012.

[Jacod, J. (1994)] Jacod, J., Limit of random measures associated with the increments of a Brownian semimartingale, Preprint number 120, *Laboratoire de Probabilités, Université Pierre et Marie Curie, Paris*, 1994.

[Jacod, J. and P. Protter (1998)] Jacod, J. and P. Protter, Asymptotic error distributions for the Euler method for stochastic differential equations, *Annals of Probability* 26, 267–307, 1998.

[Taylor, S. J., and X. Xu (1997)] Taylor, S. J., and X. Xu, The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance*, 4, 317–340, 1997.