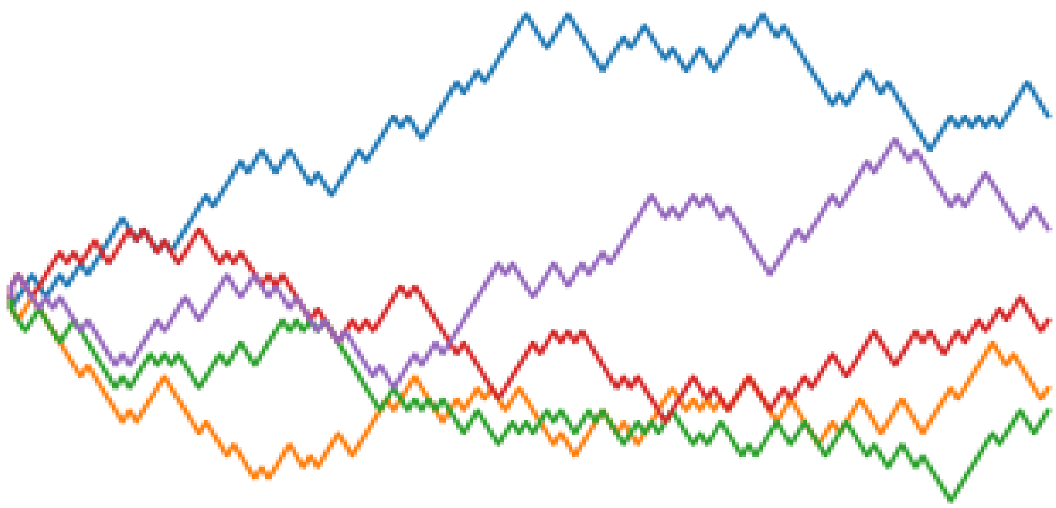
Volatility Modeling:
HAR & GARCH Models

A comparative study in forecasting accuracy



Aurouet Lucas

A thesis presented for the Master's degree of
Applied Econometrics, under the direction of Pr. B.Sevi

Master EKAP
IAE Nantes
31/05/2020

Abstract

In this thesis, the goal is to compare different volatility models; mainly the GARCH model along with the HAR model proposed by F.Corsi. HAR-family models are expected to outperform GARCH models in terms of out-of-sample forecasting accuracy. We test this hypothesis and find results in line with the existing literature; HAR does outperform GARCH both in and out-sample. Because they are also easily computed with OLS, HAR models are already becoming the new benchmark model for volatility modeling in financial series. This is why we dedicated a part to a deeper analysis of the HAR model's performance. We investigate on how different estimation methods such as WLS and LAD can yield different results than standard OLS. Although we find that some estimators induce more reliable forecasts, the results are not different enough for us to discriminate any of the methods with statistical significance.

Contents

Chapter 1

Introduction

Since the 2008 financial crisis, the measure of risk has become a major field in econometrics. Former risk metrics have been criticized along with traditional assumptions on the behavior of stock prices. A large amount of empirical work on assets return's density distribution have since proven that stock returns do not follow a Gaussian Normal distribution. Assumption which implies underestimation of the occurrence of extreme values, hence underestimation of risk itself. Nowadays, being able to accurately measure and forecast risk has become an essential issue for financial institutions. New tools have been developed to obviate the use of such old-fashioned hypothesis and to try to give a broader and more precise measure of risk on financial markets hopefully fitting the stylized facts of asset prices time series. Volatility, which is the standard measure of risk is the core principle of these new econometrics models and the best way to approximate and forecast it is the on-going and relatively recent debate among econometricians and statisticians. In 2009, F. Corsi developed the HAR model [1] (Heterogenous Auto Regressive) which quickly became a benchmark model for volatility forecasting. Its simplicity and predictive power make it a very interesting model to implement, even thought older models such as the GARCH-family models have been considered state of the art for a long time in the econometric literature. The main goal here is to verify the hypothesis whereby HAR outperforms GARCH models in volatility forecasting. Furthermore, we will also focus on different HAR models and how they perform compared to the benchmark HAR estimated by OLS. This idea is strongly suggested by A.Clements and D.Preve, who found statistically different forecasting accuracies inside the HAR family models,

[1] A Simple Approximate Long-Memory Model of Realized Volatility - F.Corsi [8]

depending on the estimation method used. As we will see, HAR can be estimated by standard Least Squares estimation, as opposed to GARCH which requires Maximum Likelihood Estimation. The availability of traditional regression techniques opens a door to a variety of different experiments on HAR modeling. We will try and compare different estimation methods such as OLS, WLS and LAD. This study will rely on three data sets from Tickdata, in particular the S&P 500, the WTI and the USD/CAD exchange rate. The first chapter will set the background required fort he following chapters. We will introduce the notion of Realized Variance, as a proxy for volatility. For this purpose, we need an overview of Brownian Motions and stochastic integrals. This will help us understand the fundamental differences between HAR and GARCH mdoeling. We will also focus on the theory behind estimation methods, both for the HAR model (OLS, WLS and LAD) as well as for the GARCH (MLE and QMLE), as this also is a nuance between these two kinds of volatility models. With the theoretical ground covered we will estimate and compare different HAR and GARCH models. We end up with a total of seven models (three GARCH(p, q) and four HAR), allowing us to do an extensive review of each model's predictive accuracy.

Chapter 2

Methodology

2.1 Estimating Volatility With Realized Variance

2.1.1 Brownian Motions and Price Diffusion Processes

Volatility has to be estimated and the choice of its estimator is essential and also a research-intensive debate in quantitative finance.

It is usually assumed (and, under some hypothesis, mathematically verified [1]) that prices follow a stochastic process called Brownian Motion (BM). Formally, we define a Brownian Motion as a stochastic process that verifies:

- Independent increments: for all $t, s, t > s$ the $Bt - Bs$ increment is independent from the previous states of the process $B_u 0 < u < s$. Where B_u is the value of the Brownian Motion at time u.
- Gaussian stationary increments: $B_t - B_s$ is a normal random variable with mean 0 and variance $t - s$.
- B(t) is continuous [2]

Also, if $B_0 = 0$, the process is called a Standard Brownian Motion. Less formally, Brownian Motion is a generalization of a Random Walk in continuous time. Which

[1] For any n-dimensional arbitrage-free vector price process with finite mean, the associated logarithmic vector price process, p, may be written uniquely as the sum of a finite variationand predictable component, A, and a local martingale, M = (M1 , ... , Mn). [19]

[2] We often see "almost surely" continuous. We will ignore this detail here.

is why Random Walk and Brownian Motion have slightly different definitions, however both processes are identical in the broad sense. Random Walk is constituted of discrete (independent and normally distributed) increments meanwhile Brownian Motion's increments are defined in continuous time. Now that Brownian Motion is defined, let's imagine a risk-free asset (with a null volatility, by definition). And take p_t the price at date t.We can write:

$$p_t = p_0 e^{\mu t} \tag{2.1}$$

The price grows exponentially with respect to time scaled by a factor μ as interests are compounded [3]. Now, we can add a volatility term, adding risk in the equation. According to the Efficient Market Hypothesis[4], the stochastic variation in an asset price follows a Random Walk process or, in continuous time a Brownian Motion. The asset's price grows exponentially but we are adding randomness in the path, with the help of a Brownian Motion. We write this as:

$$p_t = p_0 e^{\mu t + \sigma B_t} \tag{2.2}$$

Because B_t, the Brownian Motion is a chain of i.i.d normal increments, σB_t is a chain of random increments scaled by a factor σ σ can be time dependent inducing periods of high and low volatility. When σ increases σB_t increases. σ then becomes a measure of the magnitude of the random fluctuations around the risk-free path taken by an asset with no volatility. In volatility Modeling, we are interested in σ, will then tells us if prices will move by large amounts or stay of the same order. A precision is to be made here: both μ and σ can be either constants or functions of time and asset price. This equation simplifies as:

$$\frac{p_t}{p_0} = e^{\mu t + \sigma B_t}$$

$$\ln(\frac{p_t}{p_0}) = \mu t + \sigma B_t$$

We know that:

$$\ln(\frac{p_t}{p_0}) = \ln(p_t) - \ln(p_0) = \frac{dp_t}{t}$$

[3] in a risk-free asset, Treasury bills for example μ would be the risk-free rate of return

[4] See Efficient Capital Markets: A Review of Theory and Empirical Work - E.Fama [6]

So far we have:

$$\frac{dp_t}{t} = \mu t + \sigma B_t$$
$$dp_t = \mu t dt + \sigma t dB_t$$

dp_t, the change in the asset price is dictated by a certain value μ and a random (stochastic) process of i.i.d normal increments, scaled by a factor σ. The last equation is known as a Stochastic Differential Equation (SDE). Precisely, a process that satisfies an SDE of this form is called a Geometric Brownian Motion. What we can drive from this result is that the change in price process dp_t of an asset follows a stochastic process with finite and predictable drift term $\mu t dt$ (which in finance would be equivalent to the expected rate of return) along with a locale martingale $\sigma t dB_t$[5]. The asset's price process itself can be identified by integrating the former equation:

$$dp_t = \mu t dt + \sigma t dB_t$$

$$p_t = \int_0^t \mu t dt + \int_0^t \sigma t dB_t \text{[6]} \tag{2.3}$$

2.1.2 Quadratic Variation

We are interested in $\int_0^t \sigma t dB_t$, which represents the volatility component of the asset path. Solving for $\int_0^t \sigma t dB_t$ requires stochastic calculus, this problem cannot be solved using standard calculus as B_t is a stochastic process. This type of integrals are called Ito's integrals[7].And if a random variable can be expressed as Ito's integrals it follows a particular stochastic process named Ito's process. We will not address the resolution of such integrals, as it requires advanced calculus. However, we need to use one of their properties to be able to forecast volatility. As mentioned earlier, σ represents the magnitude of the stochastic process, which is volatility. It can be shown that for Ito's processes, the quadratic variation (quadratic is another terminology for 'squared') of the process is equal to:

[5] As a reminder a martingale is a stochastic process that satisfies $E(X_{n+1} \| X_1, X_2, ..., X_n) = X_n$ expected value for next period is equal to the most recent observation

[6] Here we assume $p_0 = 0$

[7] In honor of the Japanese mathematician Kiyoshi Ito, for more information see "Ito's Lemma"

$$< p >_t = \int_0^t \sigma^2 t dt \tag{2.4}$$

Also called Integrated Variance (IV), the Quadratic variation (QV)"reveals exactly the actual variance of a stochastic volatility" [19]. Although, σ^2 is the true parameter, which is unknown[8]. σ^2 then has to be estimated. One efficient and consistent exists, especially because we supposed that the price process was an ito's process. For such processes, $\int_0^t \sigma^2 t dt$ represents the quadratic variation. But for any process, the definition of QV is given by:

$$< X >_t = \lim_{\pi \to \infty} \sum_{k=1}^{n} (X_{tk} - X_{tk-1})^2 \tag{2.5}$$

With π, the number of partitions of the [t-1;t] interval. As these partitions become smaller (as we divide time into smaller and smaller sub-intervals) $\sum_{k=1}^{n} (X_{tk} - X_{tk-1})^2$ approaches the true QV. Because we made the assumptions that prices follow an ito's process, we have:

$$\lim_{\pi \to \infty} \sum_{k=1}^{n} (X_{tk} - X_{tk-1})^2 =< QV >= \int_0^t \sigma^2 t dt \tag{2.6}$$

We do not know the true value of $\int_0^t \sigma^2 t dt$, but we know $\lim_{\pi \to \infty} \sum_{k=1}^{n} (X_{tk} - X_{tk-1})^2$ tends towards that value as we divide time in sub-intervals. In fact QV represents the squared difference between two time stamps as the time elapsed between these two time stamps (π) tends to 0. RV is a daily variable hence π could be 24 if we had hourly data, 1400 if we have one value per minutes etc...More importantly, it means that with intraday data (π ¿ 1), one can compute the difference between two time stamps, and as the number of intraday data points increases, we asymptotically converge to QV, thus to $< p >_t = \int_0^t \sigma^2 t dt$. The sum of squared difference between two periods (or sum of squared returns), which we will now refer to as Realized Variance, is an asymptotically consistent estimator of the variance of a price process with stochastic volatility.

$$RV_t = \sum_{t=1}^{n} r_{t,t+n}^2 \tag{2.7}$$

[8] σ^2 is unobserved, otherwise we could perfectly capture the price process

With r the intraday returns equal to $X_{tk} - X_{tk-1}$ as long as we have intraday data. RV approximates the stochastic part of the price diffusion process (if & only if the process if a semi-martingale). Although RV has been proven to be sensible to micro-structure noise[9] and jumps or leverage effect, we will not go into so much details here. However robust QV estimates (such as the bipower variation) and forecasting methods take these factors into account. See: The Role Of Jumps and Leverage in Forecasting Volatility In Internationals Equity Markets - D.Buncic K.Gisler. [5]

2.2 HAR & GARCH modeling

2.2.1 HAR Models

With the ease of accessing high frequency (intra-day) data, volatility forecasting models started relying on on RV to produce forecasts. The Heterogeneous Auto Regressive model has gained popularity due to its simplest form ans its forecasting accuracy outperforming a large class of pre-existing models. We are going to focus on the standard HAR which is the most accessible version. Other HAR-family models do include additional terms to take account for volatility jumps (HAR-CJ) or leverage effect (HAR-CJL). These models wont be covered here although D.Buncic and K.I.M.Gisler found that they were statistically improving out-of-sample accuracy[10]. The original model, as proposed by Corsi in 2009 is expressed as follows:

$$RV_t = \beta_0 + \beta_1 RV^d_{t-1} + \beta_2 RV^w_{t-1} + \beta_3 RV^m_{t-1} + U_t \tag{2.8}$$

Where RV is a linear combination of past; daily (RV^d_{t-1}), weekly (RV^w_{t-1}) and monthly (RV^m_{t-1}), RV's. The HAR hence is auto-regressive as it's values are computed based on past values of itself. U_t is a supposedly Normally distributed white noise process (Gaussian white noise). We will see later on that depending on which estimation method we use to find β's, that the Gauss-Markov conditions[11] do not necessarily have to be respected.

The daily, weekly and monthly Realized Variances are respectively defined as:

[9] Bid-Ask jumps, latency, discreteness for example

[10] [5]

[11] As a reminder the Gauss-Markov conditions are: $u_t\ N(0, \sigma^2), Cov(u_t, u_{t+1}) = 0$

$$RV_t^d = \sum_{i=1}^{n} r_{i,i-1}^2 \tag{2.9}$$

$$RV_t^w = \frac{1}{5}\sum_{j=1}^{5} RV_{t+1-j}^d \tag{2.10}$$

$$RV_t^m = \frac{1}{22}\sum_{j=1}^{22} RV_{t+1-j}^d \tag{2.11}$$

2.2.2 GARCH(p, q) Models

Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) is a volatility modeling tool close from an ARMA model on squared returns. We will begin with an ARCH(1) model which easily extents to ARCH(p) and GARCH(p,q) models. If we have y_t the log returns of an asset, the ARCH model is specified in the following manner:

$$y_t = \sqrt{\beta_0 + \beta_1 y_{t-1}^2}\, u_t \tag{2.12}$$

Where u_t is a strong white noise with mean 0 and variance 1. If we square the log returns, we can write:

$$y_t^2 = \beta_0 + \beta_1 y_{t-1}^2 + \eta_t$$

With η_t, a martingale difference sequence.

$$\sigma_t^2 = \beta_0 + \beta_1 RV_{t-1} + \eta_t$$

Where σ_t^2 is proxied by y_t^2, the daily squared log-returns. While σ_t^2 was proxied by the sum of intraday squared log returns in RV. RV necessarily is a better estimator of volatility as RV converges to QV as π increases, if $\pi = 1$ RV and squared log returns are strictly identical.

For an ARCH(p) model we would have:

$$y_t^2 = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i}^2 + \eta_t$$

The ARCH model is an auto regressive model on the squared daily log-returns of an asset. The Generalized-ARCH follows the same reasoning adding a Moving Average component. The GARCH(p,q) model is then expressed as:

$$y_t = \sqrt{h_t} u_t$$

Where h_t:

$$h_t = \beta_0 + \sum_{i=1}^{q} \beta_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

We can rewrite the GARCH(p,q) in an alternative form that gives an ARMA(max(p,q),q) representation model of the squared log returns:

$$y_t^2 - \eta_t = \beta_0 + \sum_{i=1}^{q} \beta_i y_{t-i}^2 + \sum_{j=1}^{p} \alpha_j (y_{t-j}^2 - \eta_{t-j})$$

$$y_t^2 = \beta_0 + \sum_{i=1}^{max(p,q)} \beta_i + \alpha_j y_{t-i}^2 + \eta_t - \sum_{j=1}^{p} \alpha_j \eta_{t-j}) \tag{2.13}$$

In the following sections we will focus on these two models and we will evaluate their forecasting accuracy. The already available literature suggests the HAR should outperform the GARCH model in terms of out-of-sample predictions. We will use our results to assess if that intuition is correct.

2.3 Estimation

2.3.1 HAR models: Minimizing The Sum of Residuals

The HAR model is broadly adopted for its computational efficiency, a regular OLS is sufficient to estimate the 4 parameters. The GARCH model is usually estimated via Maximum Likelihood Estimation (MLE) or, Quasi-Maximum Likelihood (QMLE). However, every statistical model is sensible to the estimation method. D.Preve and A.Clements found that various estimations yielded different out-of-sample accuracies for the same HAR model[12].

[12] A Practical Guide to Harnessing the HAR Volatility Model – D.Preve, A.Clements [2]

Among the many different ways of estimating coefficients for the HAR model, OLS, LAD and WLS have been tested and compared. We will go over these 3 different methods and their drawbacks before addressing MLE and QMLE. Ordinary Least Squares is the most basic estimation method. For an HAR model, it consists in solving the following minimization problem:

$$min \sum_{t=23}^{n} (RV_t - \hat{\beta}_0 - \hat{\beta}_1 RV_{t-1}^d - \hat{\beta}_2 RV_{t-1}^w - \hat{\beta}_3 RV_{t-1}^m)^2 \tag{2.14}$$

Although OLS allows one to compute the estimates for β's, Least Squares are restricted by the Gauss-Markov conditions. For this minimization problem to yield the BLUEs, the error term u_t is supposed to be normally distributed and homoskedastic. If not, the estimators are biased and might induce large prediction errors. The other issue with OLS is its sensitivity to outliers. Therefore, OLS is not optimal when fitting an HAR model. Empirically, it has been shown that RV data usually contains outliers and possibly influential points inducing estimation bias. Outliers have by definition, high residuals which will be squared when using OLS. When minimizing the sum of squared residuals (eq. 2.14), the fitted values will be drawn in, in order to reduce the residual. The closer the fitted value is to the actual value, the smaller the residual is. As a counterpart, fitted value will also mova away from the rest of the data that contains the reliable and relevant information. Not squaring the residuals might address this issue.

This is the main reason why we could consider Least Absolute Deviation (LAD). The motive behind LAD is to try and avoid squaring residuals to reduce the outliers' impact on the regression. To do this, instead of minimizing the sum of squared residuals, we minimize the sum of the absolute values of residuals. By taking absolute values we ensure to have positive residuals only[13], we then have to solve for:

$$min \sum_{t=23}^{n} \left| RV_t - \hat{\beta}_0 - \hat{\beta}_1 RV_{t-1}^d - \hat{\beta}_2 RV_{t-1}^w - \hat{\beta}_3 RV_{t-1}^m \right| \tag{2.15}$$

The LAD equation's form will automatically give less weight to outliers than regular OLS. That is why this LAD estimator has been considered for HAR models.

Althought LAD is a reasonable method to deal with some of the stylized facts of Realized Variance, it still gives the same weight to all the observations. But one

[13]As negative and positive residuals might cancel out each other

might assume that some points are more relevant than others and should have higher impact on the regression. Therefore, another method was considered; Weighted Least Squares (WLS). The idea is to minimize the weighted sum of residuals, expressed as:

$$min \sum_{t=23}^{n} w_t(RV_t - \hat{\beta}_0 - \hat{\beta}_1 RV_{t-1}^{d} - \hat{\beta}_2 RV_{t-1}^{w} - \hat{\beta}_3 RV_{t-1}^{m})^2 \tag{2.16}$$

With w_t the weight of the $t^t h$ observation.

These weights can be computed in many different ways depending on what we are aiming for. Because the idea is to give less weights to observations with higher residuals, weights have to be inversely proportional to the $t^t h$ observation's value. One way of distributing these weight value would then be:

$$w_t = \frac{1}{\hat{RV_t}}$$

We then obtain a $WLS_{\hat{RV}} - HAR$ model, where $\hat{RV}$ are the fitted values obtained from the standard OLS regression. Another possible technique would be using an $WLS_{GA\hat{R}CH} - HAR$ model. If the residuals from an HAR estimated by OLS indeed display conditional heteroskedasticity, we can then fit a GARCH model on the residuals and compute the w_t as:

$$w_t = \frac{1}{\hat{y^2}}$$

Where $\hat{y^2}$ is the conditional variance estimated by the GARCH model on the residuals.

2.3.2 GARCH models: Maximum Likelihood Estimation

The GARCH-family models often are estimated by QMLE, however the R package we are using uses MLE. We will briefly go over the MLE and understand the difference between standard MLE and QMLE.

If we have a random sample of variables X_i following an assumed probability density function with unknown parameter θ such as this mass function is $f(x_i; \theta)$, then, the joint probability density function of X_i's would be:

$$\mathcal{L}(\theta) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = f(x_1;\theta) * f(x_2;\theta) * ... * f(x_n;\theta)$$

Which is, if we take the product notation:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(x_n;\theta) \tag{2.17}$$

If we solve the first order condition:

$$\frac{\partial \mathcal{L}(\theta)}{\theta} = 0$$

We find the θ parameter that maximizes the likelihood of observing x_i's under the assumption that x_i's follow a distribution with a pdf such as $f(xi;\theta)$. For computational purposes we usually use the logarithm of the likelihood function and we then solve the first order condition of the so-called Log-Likelihood function:

$$\ln \mathcal{L}(\theta) = \sum_{i=1}^{n} f(x_n;\theta) \tag{2.18}$$

In theory solvable with traditional calculus. However, in practice θ is often achieved via iterative algorithms. The Quasi-Maximum Likelihood Estimation is based on the same principle but allows for less restrictive hypothesis. One only has to specify a variance function[14] when using QMLE meanwhile the complete specification of a probability density function is required when using MLE. One can then avoid making the assumption that data follows a particular pdf when using QMLE. In most circumstances, the true distribution is unknown, therefore assumed. The Gauss-Markov conditions, for example, assume that residuals are normally distributed, but in the case where this is not true, the residuals' distribution is mispecified. The advantage of QMLE over MLE is that QMLE allows for mispecification while still being asymptotically efficient[15]. Explaining why QMLE is often preferred over MLE.

[14] a function which links the variance and mean of a variable with the variance being a function of the mean of that same variable. For example, the variance function of a Gaussian distribution is constant

[15] Maximum Likelihood Estimation of Misspecified Models - Halbert White [10]

Chapter 3

Empirical Results

3.1 Data

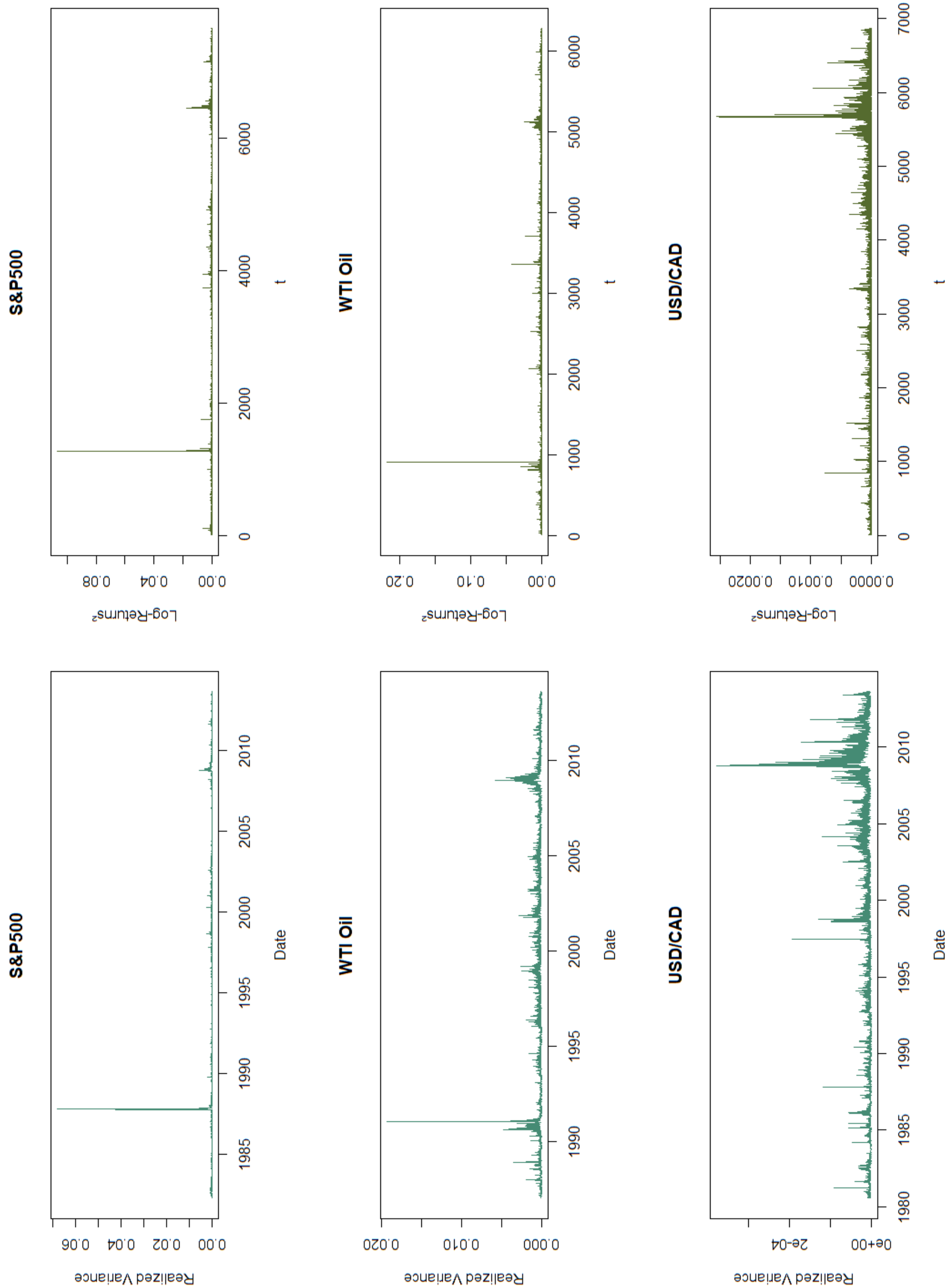
3.1.1 Outlook

The data we will be using has been provided by Tickdata. The full data set is composed of 3 series, one for each asset we will be studying. We will apply the models we discussed on the S&P500, the WTI and the USD/CAD exchange rate. Although we possess data on the jump components for these assets we will restrict our analysis to the standard HAR model(s)[1]. The S&P500 data set consists in 7661 observations from 22/04/1982 to 16/08/2013. The WTI data set ranges from 02/01/1987 to 16/08/2013 totaling 6279 observations. Finally the USD/CAD ranges from 21/07/1980 to 16/08/2013 with 6869 observations.

We will divide the data into two different sub-samples. One will be reserved for estimation and in-sample goodness-of-fit tests. The other one will compose the testing set on which we will fit the previously estimated models to compare out-of-sample forecasting accuracies. The testing sample will be composed of 7% of the observations. This was partly motivated by our results the results found by H.S. Ng and K.P. Lam[9] and also because using percentages allows for a clearer R script.

We present the series in the following figures. The first three ones represent the Realized Variance series, used for the HAR regression models, the three last ones represent the squared log returns, used for the GARCH models.

[1] As mentioned earlier, jumps can be added to HAR models as an external regressor


S&P500
WTI Oil
USD/CAD
Realized Variance
Date
Log-Returns²
t

3.1.2 Addressing Outliers

We can notice extreme values in the three series, considering both RV and squared log returns. The SP500 volatility takes very large values in 1987, because of the Black Monday, in october 1987, when the New York Stock Exchange plunged over a one day course. This incredibly fast decrease is responsible for the record-breaking volatility in the series. The WTI also displays high volatility. We notice a peak in 1991, probably explained by the Gulf War and the oil crisis that followed. Meanwhile the USDCAD volatility peaks in 2008, during the subprime crisis. Before going into further analysis, we have to treat these points. Outliers detection and treatment is a major issue when trying to fit statistical models on a series. As discussed in the first part, outliers have high influence on the regression line. Because of their value, when minimizing the residuals the line will tend to get as close as possible to these points neglecting the rest of the data. Hence, having outliers in our data-set might affect the HAR model's performance. Furthermore, outliers can impact estimation, by changing the coefficients, but also inference and forecasting because the model is unstable. A model fitted on one-off data points will be worse at predicting average values. On the other hand we showed that LAD and WLS can be used to reduce the impact of influential points by giving them a lesser weight in the objective function (when minimizing the residuals). In the next part, we will try to assess whether or not these points have to be removed from our data-set in order to keep the results adapted to the rest of the data points. We will run the four regressions with and without these outliers (possibly influential) and see if the exclusion of these extreme values yields statistically different results. We will also use Cook's Distance, which is a concrete measure of the change in standard deviation and residual values between two models with and without outliers. It is one way of measuring the statistical differences between regression fitted with and without outliers. Cook's Distance is defined as:

$$D_i = \frac{\sum_{j=1}^{n}(\hat{y}_j - \hat{y}_{j(i)})^2}{ps^2} \tag{3.1}$$

Where s^2 is the Mean Squared Error (MSE) of the model, n is the number of observations and p the number of predictors. Different authors use different cut-off values to determine if a point can be considered an influential point. 0.5 and 1 are common values so we will consider removing the i^{th} point if it's Cook's Distance is greater than 1. We will only use the residuals from the standard OLS to measure the Cook's Distance as WLS and LAD regressions are, by construction, not impacted by outliers the same way as OLS. The following table and figures summarize the results.

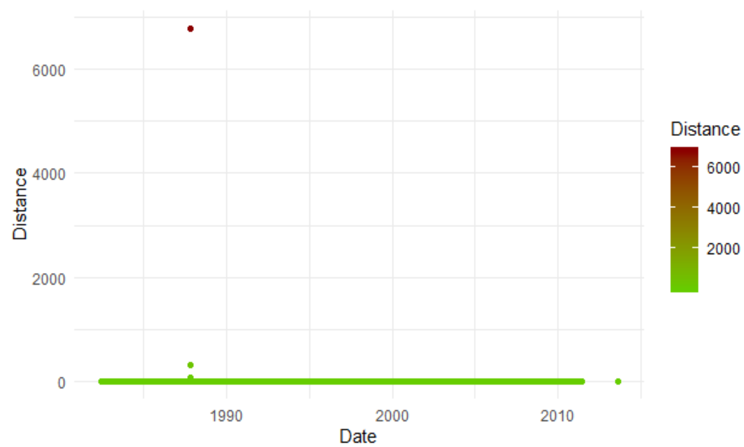


Figure 3.1: Cook's Distances from the OLS residuals for the S&P 500

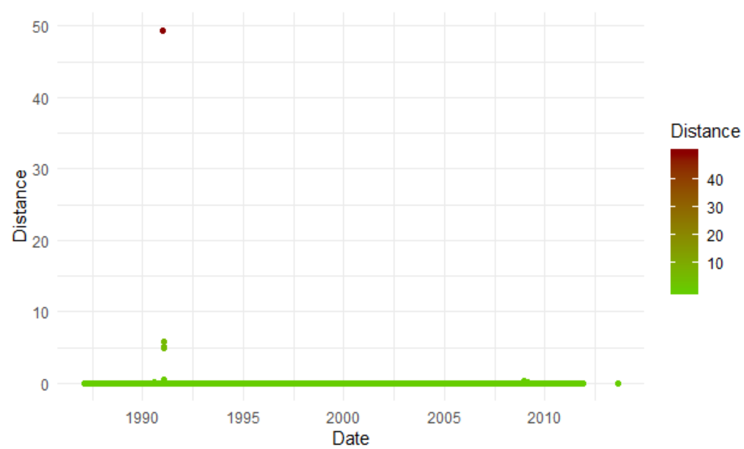


Figure 3.2: Cook's Distances from the OLS residuals for the WTI

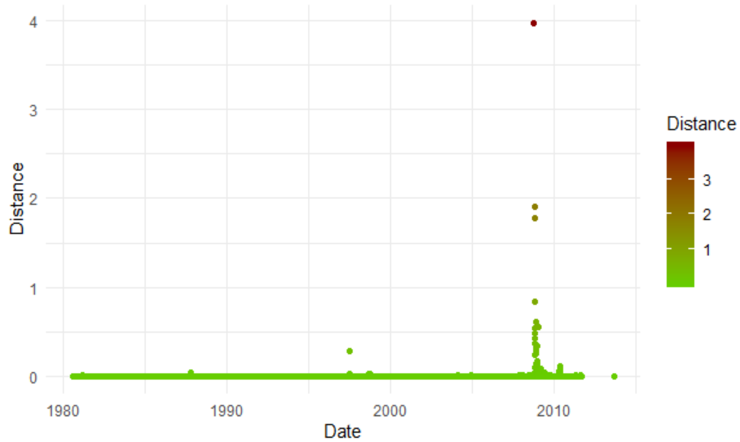


Figure 3.3: Cook's Distances from the OLS residuals for the USD/CAD

The previous plots help us identify data points falling outside of the average range of the data. In addition, Cook's Distance also measures the impact of these extreme points on the regression line. As we can notice, the 3 series contains points that are heavily impacting the model. To help us decide whether or not to remove these points, we will also use the fits presented bellow. Nevertheless, the outliers seem to confirm the possibilities we evoked previously. The outliers in the S&P 500 series are situated around 1987 on the figure, and after looking at the data itself we find that the point with the highest Cook's Distance is the 1256^{th} point, which corresponds to 1987-10-21, right during the financial crash we mentioned earlier. For the WTI series, RV spikes in January 1991, on the 1991-01-09, resulting in the outlier one can see on the figure above. This point is the 883^{th} point in the series, and the date leaves us thinking that the Gulf War at the beginning of 1991 might be responsible. And last, the USD/CAD exchange rate also seems to contain outliers, even thought the Cook's Distance is smaller it still exceeds the threshold of 1. In comparison, the other two series that display incredibly large Cook's Distances compared to the USD/CAD. The 5646^{th} point has a Cook's Distance of 3.97, hence, may have to be removed from the data set. The 5646^{th} point corresponds to 2008-10-10, at the end of the subprime crisis.

One important distinction is to be made between the three figures above, one can notice that both WTI and S&P 500 contain either one or a handful of potential outliers, the USD/CAD contains a lot more around 2008. This provides additional information; RV spiked during a very short amount of time in the first two series meanwhile for the USD/CAD Cook's Distance suggests a period of very high volatility, going on for several weeks, or months. The outliers in both the WTI and S&P 500 resulted from a quick crash over one day whereas the USD/CAD has been impacted throughout the entirety of the sub-prime crisis. This should not be surprising now that we can observe the full consequences of the 2008 financial crisis; among others, its spread and duration.

Now that we identified points that are likely to be outliers (that have an influence on the fitted model), we are going to look at the same models fitted on different data sets. One with the alleged influential points and one without. These regression should yield different results, which would be a reason for why these points have a large Cook's Distance [2]. The Cook's Distance could be sufficient to discriminate these points as it already is a measure of change between two models. However, we find it interesting to analyze the results in details.

[2]As discussed earlier the Cook's Distance is a measure of change in residuals between the same model fitted on different data

With Outliers				Without Outliers (1:1260)			
Model	Coefficients	p-values	R^2	Model	Coefficients	p-values	R^2
OLS	$-7.920e^{-02}$	$5.35e^{-10}$ ***	0.285	OLS	$4.245e-02$	0.00205 **	0.6273
	1.002	$< 2e^{-16}$ ***			$9.186e-01$	$< 2e^{-16}$ ***	
	$-6.646e^{-02}$	0.0682 .			$-1.864e-02$	0.00455 **	
$WLS_{\hat{RV}}$	NA	NA	NA	$WLS_{\hat{RV}}$	$-5.312e^{-02}$	0.000257 ***	0.6085
	NA	NA			1.034	$< 2e^{-16}$ ***	
	NA	NA			$3.753e^{-03}$	0.346682	
WLS_{GARCH}	$-4.979e^{-01}$	$< 2e{-16}$ ***	0.5928	WLS_{GARCH}	$-1.542e^{-01}$	$< 2e^{-16}$ ***	0.5582
	2.590	$< 2e{-16}$ ***			1.338	$< 2e^{-16}$ ***	
	$-8.555e^{-01}$	$< 2e{-16}$ ***			$-1.199e^{-01}$	$< 2e^{-16}$ ***	

Table 3.1: S&P 500

With Outliers				Without Outliers (1:889)			
Model	Coefficients	p-values	R^2	Model	Coefficients	p-values	R^2
OLS	$-2.052e^{-01}$	$< 2e^{-16}$ ***	0.5057	OLS	$-1.817e^{-01}$	$< 2e^{-16}$ ***	0.6407
	1.206	$< 2e^{-16}$ ***			1.191	$< 2e^{-16}$ ***	
	$-2.070e^{-02}$	0.459			$-2.373e^{-02}$	0.378	
$WLS_{\hat{RV}}$	$-1.815e^{-01}$	$< 2e^{-16}$ * **	0.5842	$WLS_{\hat{RV}}$	$-1.818e-01$	$< 2e^{-16}$ ***	0.6023
	1.204	$< 2e^{-16}$ ***			1.218	$< 2e^{-16}$ ***	
	$-3.617e^{-02}$	0.0558 .			$-4.706e^{-02}$	0.0481 *	
WLS_{GARCH}	$-3.221e-01$	$< 2e^{-16}$ ***	0.5683	WLS_{GARCH}	$-2.365e^{-01}$	$< 2e^{-16}$ ***	0.586
	1.679	$< 2e^{-16}$ ***			1.416	$< 2e^{-16}$ ***	
	$-3.074e^{-01}$	$< 2e^{-16}$ ***			$-1.631e^{-01}$	$2.78e^{-08}$ ***	

Table 3.2: WTI

With Outliers				Without Outliers (5400:6847)			
Model	Coefficients	p-values	R^2	Model	Coefficients	p-values	R^2
OLS	$-1.840e^{-01}$	$< 2e^{-16}$ ***	0.6752	OLS	$-1.824e^{-01}$	$< 2e^{-16}$ ***	0.5095
	1.158	$< 2e^{-16}$ ***			1.184	$< 2e^{-16}$ ***	
	$1.251e^{-02}$	0.615			$-2.803e^{-02}$	0.350	
$WLS_{\hat{RV}}$	$-1.885e^{-01}$	$< 2e^{-16}$ ***	0.6611	$WLS_{\hat{RV}}$	$-1.869e-01$	$< 2e^{-16}$ ***	0.5751
	1.215	$< 2e^{-16}$ ***			1.218	$< 2e^{-16}$ ***	
	$-3.379e^{-02}$	0.106			$-4.889e^{-02}$	0.0320 *	
WLS_{GARCH}	$-3.158e^{-01}$	$< 2e^{-16}$ ***	0.6277	WLS_{GARCH}	$-3.849e^{-01}$	$< 2e^{-16}$ ***	0.5577
	1.649	$< 2e^{-16}$ ***			1.835	$< 2e^{-16}$ ***	
	$-3.076e^{-01}$	$< 2e^{-16}$ ***			$-4.259e^{-01}$	$< 2e^{-16}$ ***	

Table 3.3: USD/CAD

We start with the S&P 500. The first striking result is that we are unable to compute the $WLS_{\hat{RV}}$ due to negative weights. We cannot use negative weights in WLS, however, because of the way we compute these weights, some of them end up being negative. As a reminder, for the $WLS_{\hat{RV}}$, weights are calculated using: $\frac{1}{\hat{RV}}$. $\hat{RV}$ being the fitted values from the OLS model. First, because this gives us negative weights we cannot fit the $WLS_{\hat{RV}}$ which is a problem in itself. But more importantly, this means the OLS model has some of its fitted values that are negative. Because of the auto-regressive characteristic of the HAR model, when the large RV value on the 1987-10-21 is included in the model as lagged RV (RV^d_{t-1}), the resulting fitted value is negative [3]. Which also is a problem on its own as volatility cannot be negative. Finding negative RV values is irrational because the price spread from one day to another has to be positive. Hence RV, which is a proxy for this spread, should always be positive. This first element strongly suggests to remove the outlier [4]. Otherwise, we notice a substantial change in statistical significance of the coefficients. the lagged monthly RV is significant at a 10% confidence level with the outlier and 5% when removing them. The daily lagged RV looses significance but still is at a 1% level. And the R^2 goes from 0.285 with outliers to 0.6273 without, which indicates that the model fits the data a lot better when we exclude the outliers. The coefficients themselves change, the daily lagged RV becomes positive when removing the outliers and the monthly RV remains negative but its coefficient is much smaller after the outliers have been removed. The $WLS_{GA\hat{R}CH}$ has a slightly worse goodness-of-fit with the outliers than without, with an R^2 going from 0.593 to 0.558. We think this is a consequence of the functional form of the model itself. Because of the weighting of extreme observations, it is possible that when removed, the model captures less variance, as these weights are specifically designed to overcome outliers effects [5].R^2 being a measure of variance, it will be impacted by outliers which by definition have high variance, when removed, or when adjusted by weighting, the overall variance of the data should decrease, the presence of outliers and the goodness-of-fit of models are closely related, especially when using WLS. In the broad sense, we expect the WLS

[3] the coefficient for lagged daily RV is negative, when multiplied by a large positive value, it should give a large negative value

[4] As we will discuss in the later part, we will remove the entire period and not just this single point

[5] and/or heteroskedasticity, which is not a problem here as it is exactly what we are trying to find and model

models to perform better than OLS in presence of outliers but to perform as good as traditional OLS once outliers have been removed, as weights should come closer to each other (if weights are all equal then WLS is strictly equivalent to standard OLS). Because of the striking differences between the models, and the fact that we obtain negative fitted values, we decide to remove the outliers in the S&P 500 series.

The WTI fits shows the same patterns, with increasing R^2 for each and every model when removing outliers. As well as stronger statistical significance of the coefficients. Lagged monthly RV seems to only be partially significant for the WTI series, but we are interested in the change in between the models and not the models themselves. We also notice WLS outperforming OLS with outliers and being outperformed when removing outliers although the goodness-of-fit of the WLS models increases. This might be due to the functional form of the models as discussed above. However, everything is suggesting to remove the outliers in the WTI series; higher significance of the coefficients [6], higher R^2 implying better goodness-of-fit. The coefficients don't vary as much as they did for the S&P 500 series, they all conserve the same sign and for the most part we notice that their variation is of the same order in every model. Daily lagged RV coefficient seems to decrease when removing the outliers, and this is true for the three models, meanwhile weekly lagged RV remains the same and monthly lagged RV coefficient increases. This last observation doesn't hold for the WLS_{GARCH} model, however the similar patterns in changes of the coefficients values leads us to believe we should also remove the outliers from the WTI series.

The coefficients for the USD/CAD also seem to increase when removing outliers although the overall goodness-of-fit decreases for every model. Monthly lagged RV is barely significant, as we first observed in the WTI series, and overall this coefficient has the worst statistical significance in every series, but it always is in the WLS_{GARCH} model, maybe indicating some strong monthly GARCH effects. Opposite to the two other series, the coefficients values do not vary much when removing outliers. Alongside the decrease in R^2, it suggests that the outliers might not be this problematic in this series. Our hypothesis is that even though influential points observed in the Cook's Distance figure have some kind of effect on the model, they are pretty well explained. They still have high residuals, because of their large value, so removing them will automatically change the Cook's Distance. But the higher R^2 values of the first models tell us that the values for these points is well captured. This might be due

[6]except for the monthly lagged RV in the WLS_{GARCH}, which still remains significant at 0.1%

the sub-prime crisis during for a longer period of time, the auto-regressive successively captured the relatively high volatility for this period. Volatility model are designed to find volatility clustering, alternation of high and low volatility periods. Meanwhile for the two other series potential outliers where only spotted in 1 or 2 points, in the USD/CAD series, 2008 seems to be a period of high volatility overall. And because we have daily data, hence numerous data points in a year, the model might be able to interpret these extreme values, as high as they are, as a high volatility cluster. Which could not be done by the same models on data with very few extreme and intermittent values, over such short periods that the auto-regressive part could not adjust to it. For this reason, we decide to use the full data set for the USD/CAD series.

Because we are using time series, we cannot remove only the outliers, but have to remove the entire period prior to or after the point. This is because if we only remove one point, which is the value for one day, the explanatory variables are not representing the same effects anymore. The daily lagged RV will be the data from two days ago, instead of RV_{t-1}^d, we would be using RV_{t-2}^d. Which would change the change the functional form of the model, as RV_{t-1}^w and RV_{t-1}^m would also change as they are moving averages of the previous days. In conclusion, we will be using 6379 observations for the S&P 500, from 1987-11-18 to 2013-08-16. 5368observations for the WTI, from 1991-01-23 to 2013-08-16. And 6869 observations from 1980-07-21 to 2013-08-16 for the USD/CAD (the original data set). We will now proceed to the estimation with these final data sets.

3.2 In-Sample Estimation

With some of the models fitted on clean data (with outliers removed), some of the coefficients still weren't statistically significant. Monthly RV especially, was often founded to be significant at a 10% level only, in some cases, not at all. We could take this information into account and decide to remove the monthly variable, as we would do when carrying any kind of linear regression. Here, we want to verify if the HAR model as proposed by Corsi in 2009 can outperform the oldest GARCH model. The specification in Corsi's paper includes monthly RV as a regressor when modeling RV, hence we will leave the variable in the equation. Furthemore, the coefficient's significance seems to increase when carrying Weighted Least Squares

Regression instead of Ordinary Least Squares, which should yield interesting results when comparing the HAR models between each other. We estimate the three HAR models on the entire data sets mentioned earlier [7]. Along with different GARCH(p, q) models $p, q = \{1, 2, 3\}$. We will settle on three final GARCH models, having a total of five models, which will be used for forecasting later on. We decide to keep two GARCH model and not just the best one to have a chance of them yielding different forecasts, then illustrating different effects that might play a role in volatility clustering.

We already estimated every HAR model above [8], we will now move toward GARCH modeling. The first model we will use is the GARCH(1, 1) model, often considered as sufficient. The second and third will be chosen upon the Akaike's Information Criterion (AIC), as well as the Baye's Information Criterion (BIC). Let us first briefly go over both criteria.

AIC is defined as :

$$AIC = 2k - 2\ln(\mathcal{L}) \tag{3.2}$$

BIC is defined as :

$$BIC = \ln(n)k - 2\ln(\mathcal{L}) \tag{3.3}$$

Where n is the number of observations, k the number of estimated parameters (here, ranging from 2; (GARCH(1, 1)) to 6; (GARCH(3, 3))). $\mathcal{L}$ is the maximum of the Log-Likelihood function. BIC carries out a larger penalization for the number of parameters, compared to AIC; $2k$ in he AIC against $\ln(n)k$ in the BIC[9]. $\mathcal{L}$ hold equal, BIC will be larger than AIC. Note as $\mathcal{L}$ tends to ∞, both AIC and BIC tend to $-\infty$. This means than the better the model model fits he data (as $\mathcal{L}$ increases) the smaller the criteria values will be. Hence the best model is the one with the smallest criterion. Because BIC will always be larger than AIC, BIC tends to choose more parsimonious models, with fewer parameters to estimate, sacrificing some goodness-of-fit in favor

[7] We divided the data into two different sub-samples, with the last 7% of the data as discussed before

[8] The integrity of the results is presented in the Appendice

[9] $\ln(n) > 2, \forall n > exp(2) \approx 7.4$

of stability of the model over different data sets.

As we discussed in the first part of this paper, GARCH models are estimated by MLE, which implies a specification of an assumed density function for the series. This is imposed by the rugarch package we are using for the computation, GARCH can be estimated by QMLE, we then would not need to assume a distribution, but the rugarch package uses standard MLE. hence we have to verify which distribution is the more likely to fit the data. The standard assumption is that log-returns are normally distributed, although, stylized facts of financial returns often display fat tails and smaller kurtosis than Normal distribution. We suspect a Student's distribution might fit the log-returns better than a Normal distribution. on the figure below can can see the kernel density estimation[10] of the log-returns along with the two density functions.

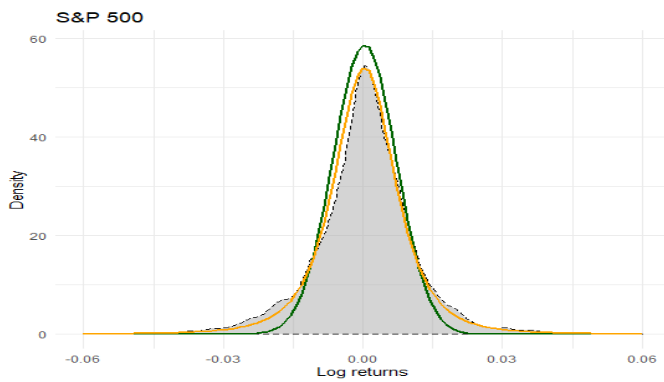


Figure 3.4: S&P 500 Data (black) against Normal distribution (green), and Student's distribution (yellow)

[10]Kernel density is is some way, equivalent to a smoothed histogram

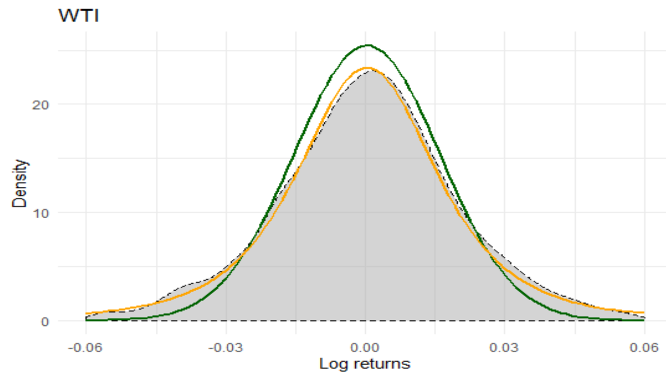


Figure 3.5: WTI Data (black) against Normal distribution (green), and Student's distribution (yellow)

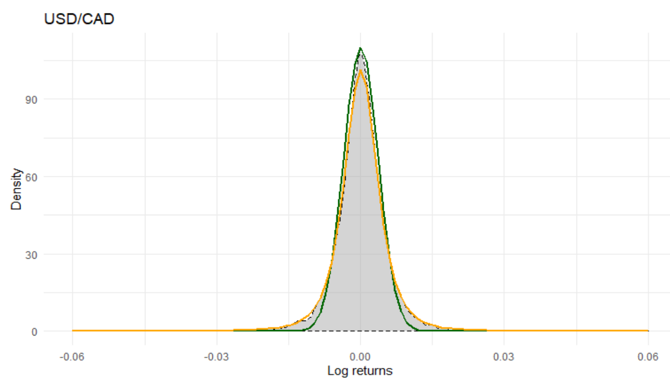


Figure 3.6: USD/CAD Data (black) against Normal distribution (green), and Student's distribution (yellow)

As expected, it seems like the Student's distribution fits the data better than the Normal distribution. Student's density function naturally is heavy-tailed, which is why it is a better candidate to represent financial series. Therefore, when we will be specifying a distribution for the MLE, we will use the Student's distribution.

We fit the GARCH(p, q) models on the data and we extract the AIC and BIC from the results. Their values appear in the Table bellow:

Model	S&P 500	-	WTI	-	USD/CAD	-
(p, q)	AIC	BIC	AIC	BIC	AIC	BIC
(1, 1)	-41022.7	-40986.7	-26354.3	-26319.2	-55517.3	-55480.9
(1, 2)	-41020.2	-40977.1	-26352.3	-26310.3	-55514.8	-55471.2
(1, 3)	-41016.3	-40966	-26350.3	-26301.3	-55512.9	-55462.1
(2, 1)	-41031.2	-40988.1	-26354.1	-26312.1	-55514.6	-55470.9
(2, 2)	NA	NA	-26352.2	-26303.1	-55512.9	-55462.1
(2, 3)	-41030.3	-40972.8	-26350.8	-26294.8	-55511.3	-55453.2
(3, 1)	-41028.4	-40978.1	-26352.2	-26303.1	-55511.7	-55460.8
(3, 2)	-41028.4	-40970.9	-26350.2	-26294.2	-55510.2	-55452.0
(3, 3)	-41030.7	-40965.9	-26351.2	-26288.2	-55509.1	-55443.7

Table 3.4: Information criteria for every series, every GARCH(p, q) model

The smallest values have been highlighted. For the WTI as well as for the USD/CAD series, both AIC and BIC suggest the same model; GARCH(2, 1) for the WTI and GARCH(1, 2) for the USD/CAD. For the S&P 500 series, AIC suggests a GARCH(3, 3) and BIC suggests a GARCH(2, 1). Because the main goal will be forecasting, we think choosing a more parsimonious model will help us gaining in forecasting accuracy, as models with fewer parameters are more likely to be usable on another data set.

We know have every model we need and we can start forecasting. We will be using the models fitted on the training data sets to forecast the out-of-sample RV and squared log-returns. AS a reminder, the HAR models are used to forecast Realized Variance, meanwhile GARCH models are used to forecast squared log-returns, which also are a proxy for volatility.

3.3 Forecasting

3.3.1 HAR vs GARCH

Usually when dealing with time series, we can use different types of forecasting schemes. Rolling forecast, n-ahead, or recursive forecasts for example. These schemes will be applied to the GARCH model, however the HAR model is not really a time series model. Whereas in time series modeling we are looking for auto-correlation between successive periods, in the HAR, we are looking at the correlation of one variable with another, which appends to be the lagged value of that same variable. Hence both models are, technically auto-regressive, but the HAR models functions as a standard Simple Linear Regression model. This is why it can be estimated by OLS allowing for less complicated computation. In terms of forecasts, HAR models will be treated as simple linear regression models, we will feed all testing data points (RV values for the testing set) in the models, which will output a value for the response variable; today's RV. GARCH model will perform 1-step-ahead rolling forecasts, meaning we will produce one forecasted value, then introduce the true value in the data set, perform another 1-step-ahead forecast, add the real value in the data set etc...Both methods are equivalent, none of the GARCH nor the HAR model receive any additional information as both are forecasting one value based on a data set containing every known value so far. The only difference is that HAR model produce all these forecast at the same time, whereas GARCH works value by value.

To assess forecasting accuracy we will use the Diebold&Mariano test, as well as the Mean Squared Error (MSE) and Mean Absolute Error (MAE) and the Mean Percentage Absolute Error (MAPE). We will briefly go over these concepts, in order to highlight one major issue we ran into during this study: scale-variance.

The Diebold&Mariano test is a statistical test for forecasting accuracy difference between two models. it's statistic is formally defined as:

$$DM = \frac{d - \mu}{\sqrt{\frac{\hat{\omega}}{T}}} \tag{3.4}$$

Where d is equal to

$$d = \sum_{t=1}^{T} g(e_{it}) - g(e_{jt}) \tag{3.5}$$

And is the overall difference between the values of the loss function of models i and j (MSE, and MAE are both loss functions[11]). $\hat{\omega}$ is an estimator of the variance of $\sqrt{T}d$, and T is the sample size. Under H_0, the DM statistic is asymptotically (as T, tends ton ∞) distributed as N(0, 1). The null being that boss models are equivalent in terms of forecasting accuracy, that their the value outputted by their loss functions are equal, hence d = 0. If two models' MSE are the same then, they have the same

[11]For the MSE, d would equal $\sum_{t=1}^{T}(\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{y_{it}})^2 - \frac{1}{T}\sum_{t=1}^{T}(y_{jt} - \hat{y_{jt}})^2)$

forecasting accuracies.

Although, the results might be, and will almost surely be, biased as RV and log-returns do not fit the same scale. All other things equal, GARCH and HAR models will have different MSE simply because log-returns values are defined between -0.005 and 0.005, whereas RV values are between 0.0000004 and 0.0003 [12]. The GARCH MSE (resp. MAE) will almost surely be larger than the HAR MSE (resp. MAE), even thought they might have similar forecasting accuracy. one solution would be to use Mean Absolute Percentage Error (MAPE, another loss function), as percentages would overcome the difference of scale. However to calculate the MAPE we need positive values only [13]. We can see that log-returns can be negative, hence we cannot use MAPE as it is. RV is already strictly positive, and poses no issue whatsoever, but log-returns, estimated and forecasted by the GARCH models can be negative (by default the rugarch packages estimates log-return, even though GARCH models are used for volatility estimation). We will only look at the variance forecasts and not the forecasts for the log-returns themselves. Squared log-returns are, by definition, strictly positive and are defined between 0 and 0.002. We hope this would allow us to compute the MAPE, which is a scale invariant loss function. But MSE and MAE might still be biased due to the scale difference. However scale-invariant loss functions are not easily accessible[14], making the HAR-GARCH comparison a problematic subject. Log-returns and RV are not strictly proportional, hence we cannot use a transformed version of the data to obtain identical scaling (otherwise we could transform log-returns by scaling it and both MSE and MAE would become unbiased). We therefore decide to use MAPE to compare HAR and GARCH models, this is partially motivated by the previous explanations, but also by the results we found, which show a clear difference in terms of performance. Both models yields very different results and we think that one loss function is enough to discriminate one of the models.

[12] This interval is built on the USD/CAD data as an example

[13] $MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{A_t - F_t}{A_t}\right|$, with A_t the actual value and F_t the forecasted value

[14] Such functions are abundant in Machine Learning and computer science

Model	S&P 500	WTI	USD/CAD
$OLS - HAR$	0.5008	0.5445	0.5157
$WLS_{\hat{RV}}$	0.4801	0.5112	0.5119
$WLS_{GA\hat{R}CH}$	0.5045	0.5425	0.5414
$LAD - HAR$	0.4272	0.4592	0.4517
$GARCH(1,1)$	∞	∞	∞
$GARCH(1,2)$	∞	∞	∞
$GARCH(2,1)$	∞	∞	∞

MAPE has many drawbacks, one of them is particularly problematic here; MAPE can produce infinite values if the series contains 0, or close to 0 values. Log-returns sometimes takes the value of 0, which happens to yield a MAPE with infinite value. We could remove observations where log-returns are equal to 0. But first; we would loose information and removing outliers should be motivated and supported by other factors than this. Second, even by removing null values from the data frame we will still be unable to compute MAPE. We tried implementing this strategy, and the MLmetrics::MAPE() function we are using in R still yields infinite MAPEs. our hypothesis is that this is due to GARCH not performing good enough. Sometimes, large residuals, caused by poor goodness-of-fit can also cause issues when computing MAPE. Some loss functions have been proposed to address this issue; symmetric MAE (sMAE), Mean Absolute Scaled Error (MASE), MAE/mean or Root Relative Squared Error (RRSE) for example. Because we were unable to use any of the standard loss functions to compare forecasting accuracies, we decide to try and test if RRSE can be achieved. RRSE is a scale-invariant loss function with the form:

$$RRSE = \sqrt{\frac{\sum_{j=1}^{T} \left(F_{ij} - A_j\right)^2}{\sum_{j=1}^{T} \left(A_j - \bar{A}_j\right)^2}} \tag{3.6}$$

We divide the accuracy of model i by the sum of the values minus the arithmetic mean of the series. Along with the aforementioned metrics, RRSE is commonly used in Machine Learning, where we can find theory behind scale-adjusted loss functions. Here, we will simply use the RRSE as it is to try and determine which of the GARCH or HAR model is performing best on their respective data sets.

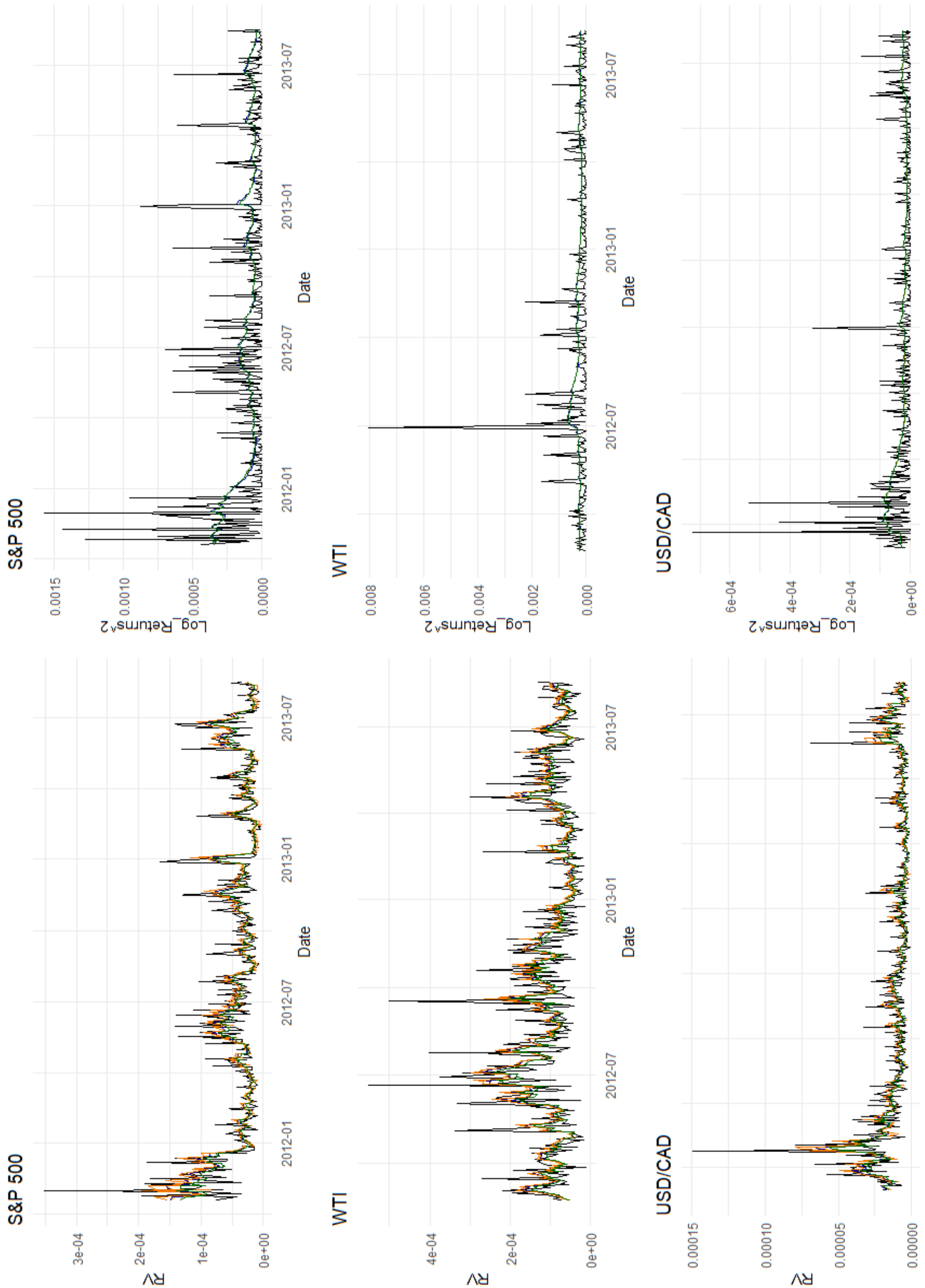
Model	S&P 500	WTI	USD/CAD
$OLS-HAR$	0.6821	0.7965	0.6827
$WLS_{\hat{RV}}$	0.6788	0.7939	0.6822
$WLS_{GA\hat{R}CH}$	0.7052	0.8047	0.6986
$LAD-HAR$	0.6918	0.8014	0.6952
$GARCH(1,1)$	1.417	1.4684	1.435
$GARCH(1,2)$	1.415	1.4682	1.436
$GARCH(2,1)$	1.418	1.4680	1.439

If the model fits the data perfectly, then the numerator in RRSE is 0, therefore RRSE ranges between 0 and ∞, the smallest values indicates the best model. We can clearly see that HAR, independent of the estimation method, outperforms every GARCH models. this is true for the S&P 500, the WTI and the USD/CAD. There is no absolute good or bad RRSE value as they are all scaled therefore the absolute value of RRSE is of little interest. As with many other loss functions, its usefulness arises when comparing different model between each other. Independent of the actual RRSE value, the model with the smallest one has the best fit. But here we can see GARCH RRSEs are almost twice as large than HAR RRSEs, indicating that HAR does outperform GARCH model in volatility forecasting. For the HAR, the RRSE ranges from 0.678 ($WLS_{\hat{RV}}$) to 0.705 ($WLS_{GA\hat{R}CH}$) In the S&P 500 series. Against the GARCH RRSE ranging from 1.415 (GARCH(1, 2)) to 1.418 (GARCH(2, 1)), fitted on the same data. For the two other series, the values are of the same order. We can clearly see that the RRSE is smaller for the HAR than for the GARCH, indicating that HAR has a better out-of-sample fit. This results confirms what has already been observed by other authors as we mentioned before. HAR is a stronger model for volatility forecasting despite its simplicity, justifying its status of benchmark model in the quantitative finance literature.

Even if we know MSE and MAE are biased due to the scale difference, they still display considerable differences between HAR and GARCH models. We know part of this difference is due to Log Returns being naturally larger than RV. However, it still strongly suggests that HAR outperforms GARCH, even taking into account the scale bias. We present these values in the Appendix.

We plotted the series along with the forecasts for each model to give a visual representation of the out-of-sample accuracy. Int he following figures, we present the

RV along with the four HAR models along with the squared log returns and the corresponding foretasted values from the three GARCH models.


S&P 500
WTI
USD/CAD
RV
Log_Returns^2
Date
2012-01
2012-07
2013-01
2013-07

The HAR models are represented in; yellow (OLS-HAR), blue ($WLS_{\hat{RV}}$), orange ($WLS_{GA\hat{R}CH}$), green (LAD-HAR). The forecasted value overlap for the most part, due to every HAR models being similar, as we will develop in the next part. The $WLS_{GA\hat{R}CH}$ forecasts seems to be higher than the other three models during periods of high volatility and smaller during calm periods. We can impute this to the weights based on potential GARCH effects in the OLS-HAR residuals. This is generally true for every series, but more noticeable in the S&P 500 series, which seems to be more subject to volatility clustering than the WTI or the USD/CAD. The GARCH models (GARCH(1, 1) in orange, GARCH(1, 2) in green, GARCH(2, 1) in blue) all overlap in every series, and we can notice the same pattern as before with the GARCH(2, 1) forecasting higher values during high volatility clusters and lower values during low volatility periods. Due to the additional auto-regressive term, the (2, 1) model has a longer and more intense persistence.

In general we can notice that, visually, HAR forecasts do fit the true values a lot better than GARCH models. This is especially true for the WTI and the USD/CAD where the GARCH forecasted values are almost constant, meanwhile the HAR forecasts fit RV pretty well. The S&P 500 series contains more lengthy clusters, reasonably well captured by the GARCH models, even though HAR performs better, as we can see here.

3.3.2 HAR With Different Estimators

We showed how HAR models can outperform GARCH models in terms of forecasting accuracy. Now we will focus on the difference in results for every estimation methods used for HAR modeling (OLS, LAD, $WLS_{\hat{RV}}$, $WLS_{GA\hat{R}CH}$). The goal is to determine if the standard OLS is sufficient or, on the opposite, if more advanced estimators yield more precise results. In the first chapter we saw that OLS has many drawbacks; it's sensitiveness to outliers, Gauss-Markov assumptions are the main issues when using OLS in any regression analysis. We saw that Weighted Least Squares or Least Absolute Deviation could overcome such problems. A.Clements and D.Preve found that results where statistically different depending on the estimator used. Hence we expect to see one of the models outperforming the benchmark OLS. Because all HAR models we fitted on realized Variance, we will not encounter the same problems as we previously did comparing GARCH and HAR models. We can use either scalevariant or scale-invariant loss functions and, as a consequence, the Diebold&Mariano test will be unbiased. We will use the MSE, the MAE and the MAPE to find which model seems to be performing better, and we will then use the DM test to find if the differences are statistically significant or not.

Model	S&P 500	WTI	USD/CAD
$OLS - HAR$	$6.2357e^{-10}$	$3.2594e^{-09}$	$6.8915e^{-11}$
$WLS_{\hat{RV}}$	$6.1752e^{-10}$	$3.2379e^{-09}$	$6.8814e^{-11}$
$WLS_{GA\hat{R}CH}$	$6.6658e^{-10}$	$3.3268e^{-09}$	$7.2163e^{-11}$
$LAD - HAR$	$6.4141e^{-10}$	$3.2991e^{-09}$	$7.1470e^{-11}$

Table 3.5: MSE

Model	S&P 500	WTI	USD/CAD
$OLS - HAR$	$1.6084e^{-05}$	$4.0579e^{-05}$	$4.7132e^{-06}$
$WLS_{\hat{RV}}$	$1.5912e^{-05}$	$3.9850e^{-05}$	$4.7168e^{-06}$
$WLS_{GA\hat{R}CH}$	$1.6876e^{-05}$	$4.0725e^{-05}$	$4.9889e^{-06}$
$LAD - HAR$	$1.5474e^{-05}$	$3.9036e^{-05}$	$4.5400e^{-06}$

Table 3.6: MAE

Model	S&P 500	WTI	USD/CAD
$OLS - HAR$	0.6821	0.7965	0.6827
$WLS_{\hat{RV}}$	0.6788	0.7939	0.6822
$WLS_{GA\hat{R}CH}$	0.7052	0.8047	0.6986
$LAD - HAR$	0.6918	0.8014	0.6952

Table 3.7: MAPE

For the S&P500, $WLS_{\hat{RV}}$ has the smallest MSE as well as the smallest MAPE. It also has the second smallest MAE behind the LAD-HAR. $WLS_{GA\hat{R}CH}$ seems to under-perform according to all three metrics. The benchmark OLS performs pretty well compared to other estimators, the $WLS_{\hat{RV}}$ might be more accurate, but we will have to use the DM test to be certain the difference is statistically significant. We find similar results for the WTI. $WLS_{\hat{RV}}$ has the smallest MSE and MAPE and the second smallest MAE, behind LAD. The GARCH is strictly under-performing. Same goes for the USD/CAD. So far the $WLS_{\hat{RV}}$ seems to have to best out-of-sample fit, with LAD, they have the smallest values in two out of three data sets (S&P500 and WTI). $WLS_{GA\hat{R}CH}$ is the worst model with the highest MSE, MAE and MAPE values on all three data sets. We will now use the DM test to see if these difference are statistically significant.

Model 2 \ Model 1	$OLS-HAR$	$WLS_{\hat{RV}}$	$WLS_{GA\hat{R}CH}$	$LAD-HAR$
OLS-HAR	-	0.1684	0.8473	0.7269
$WLS_{\hat{RV}}$	0.8316	-	0.9086	0.7704
$WLS_{GA\hat{R}CH}$	0.1527	0.0913	-	0.7704
LAD-HAR	0.2731	0.2296	0.6464	-

Table 3.8: S&P 500

Model 2 \ Model 1	$OLS-HAR$	$WLS_{\hat{RV}}$	$WLS_{GA\hat{R}CH}$	$LAD-HAR$
$OLS-HAR$	-	0.0827	0.8643	0.6237
$WLS_{\hat{RV}}$	0.9173	-	0.9134	0.6977
$WLS_{GA\hat{R}CH}$	0.1357	0.0866	-	0.4396
$LAD-HAR$	0.3763	0.3023	0.5604	-

Table 3.9: WTI

Model 2 \ Model 1	$OLS-HAR$	$WLS_{\hat{RV}}$	$WLS_{GA\hat{R}CH}$	$LAD-HAR$
$OLS-HAR$	-	0.4355	0.7058	0.7524
$WLS_{\hat{RV}}$	0.5645	-	0.7323	0.7299
$WLS_{GA\hat{R}CH}$	0.2942	0.2677	-	0.4713
$LAD-HAR$	0.2476	0.2701	0.5287	-

Table 3.10: USD/CAD

The table above shows the p-value for the Diebold&Mariano test. As we discussed before, the null hypothesis for this test is equal predictive accuracy. Therefore, a p-value smaller than 0.05 indicates a significant difference in out-of-sample performance between two models at a 5% confidence level. We chose the uni-sided test instead of the two-sided test. The two-sided test indicates if models hold different accuracy, the uni-sided test allows us to determine if one is better or worse than the other. Here, we test for Model 2 is worse than Model 1. We can see that we do not find any model strictly outperforming the OLS estimator. $WLS_{\hat{RV}}$ and LAD have the largest p-values for both the S&P 500 and the WTI indicating that their accuracy is not inferior to the OLS's one. $WLS_{GA\hat{R}CH}$ has relatively small p-values, although it is not significantly under-performing compared to the OLS. however the $WLS_{\hat{RV}}$ is better than the $WLS_{GA\hat{R}CH}$ and than the OLS at a 10% confidence level for the WTI series. We noticed before that $WLS_{\hat{RV}}$ often had the smallest MSE, MAE and MAPE, which is consistent with the latest results from the DM-test. It seems to confirm that $WLS_{\hat{RV}}$ is doing slightly better than the OLS. But we cannot say it is significantly outperforming the benchmark OLS model. For the USD/CAD, the LAD and the $WLS_{GA\hat{R}CH}$ models seem to perform slightly better than other models. But once again, no model is statistically better than the others for this data set, and none of them is capable of outperforming the OLS. We find no statistical evidence that some estimators are more appropriate when using HAR models. The standard OLS is sufficient and yields forecasts at least as accurate as WLS or LAD estimation. Although we now $WLS_{\hat{RV}}$ performs better, the marginal gain in accuracy is not statistically significant. In practice one still might consider using $WLS_{\hat{RV}}$ instead of OLS, because they both have the same degree of freedom. Meaning the slight increase in accuracy comes with no cost in statistical power, even though we found no significant difference, one may prefer a more accurate model.

We also computed an additional variable S_{resid} capturing the difference in residuals between each model and the benchmark OLS. This variable is calculated as follows:

$$S_{resid,i,t} = e_{OLS,t} - e_{i,t} \tag{3.7}$$

$$S_{i,t} = \sum_{t=1}^{t} S_{resid,i,t} \tag{3.8}$$

$S_{i,t}$ is a series representing the cumulative sum of $e_{OLS,t}$ minus $e_{i,t}$ (the residuals from the OLS model and the residuals from the i^{th} model). Here $i = \{1, 2, 3\}$. For any point t, if $e_{i,t}$ is smaller than $e_{OLS,t}$, $S_{resid,i,t}$ will be positive. in other words, if model i is more accurate than OLS (hence, has a smaller residual) $S_{resid,i,t}$ will be positive. Using a cumulative sum (instead of simply using the difference) allows us to see if one model is consistently under or over performing the OLS. We computed S_i for every model and every series, giving us nine new S_{resid} series. The S_{resid} series are plotted and can be found in the Appendix as their was nine of them, for readability we preferred having these nine additional plots in the Appendix. They highlight the differences in forecasting accuracies between our models. And as we noticed before, S_{resid}, is only positive for the $WLS_{\hat{RV}}$ model, indicating it is the only one outperforming the benchmark OLS (once again, not with enough statistical significance). This is true for every series. As an example we use the $S_{WLS_{\hat{RV}},t}$ from the S&P 500 series.

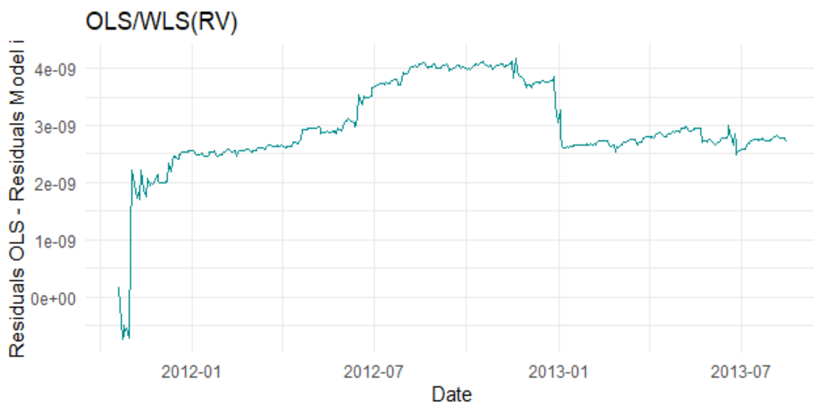


Figure 3.7: Cumulative Sum of $resid(OLS) - resid(WLS_{\hat{RV}})$ - S&P 500

The series is positive, indicating that the residuals from the OLS are usually larger than those from $WLS_{\hat{RV}}$. However the series is not steadily increasing, indicating the $WLS_{\hat{RV}}$ is not consistently outperforming the OLS (hence the use of cumulative sum instead of simple difference in residuals). These plots give us a visual representation of which model is performing better (has the smallest residuals) but we have to remember they do not give any information on the statistical significance of this difference. As we saw earlier, $WLS_{\hat{RV}}$ is slightly more accurate than OLS, but not sufficiently better to say is it significantly outperforming the OLS estimator. The previous plot simply helps us visually identify this phenomenon of $WLS_{\hat{RV}}$having smaller residuals than OLS.

Chapter 4

Conclusions

In this paper we compared different volatility forecasting models. The goal was to find out whether or not the HAR model proposed by Corsi could outperform the GARCH model that have been a benchmark model for volatility forecasting since the work of Robert F. Engle [1] in 1982. Inspired by the paper of A.Clements and D.Preve [2] we estimated the HAR model with diverse methods (WLS, OLS and LAD). We expected HAR to outperform GARCH as this phenomenon has been observed and relayed in different existing papers. We also expected the WLS and LAD to outperform OLS for HAR modeling. Because we were using different data for each model, and because of the shape of the data itself, comparing HAR and GARCH models could not be done by using standard metrics such as MSE or MAE, preventing us from running the Diebold&Mariano test between both models. We had to rely on Root Relative Squared Error (RRSE) which is a scale-invariant loss function. After cleaning the data from outliers, we find that HAR does indeed easily outperforms GARCH models, independent of the estimator used. However we could not find any statistically significant difference between the different HAR models, even though $WLS_{\hat{RV}}$ seems to have a better forecasting accuracy. HAR models are a recent breakthrough in quantitative finance and its characteristics are extremely helpful as it requires only OLS to be estimated and yet yields impressively accurate forecasts. Apart from their interest in quantitative finance, HAR models can be implemented in practice for VaR computation or risk management. They can also be upgraded by adding jump or leverage effects as external regressors, further increasing their

[1] [15]
[2] [2]

accuracy. One interesting development would be to compare HAR models to Machine Learning volatility modeling which is, as one would naturally expect, becoming a noteworthy feature in quantitative finance.[2]

Chapter 5

Appendix

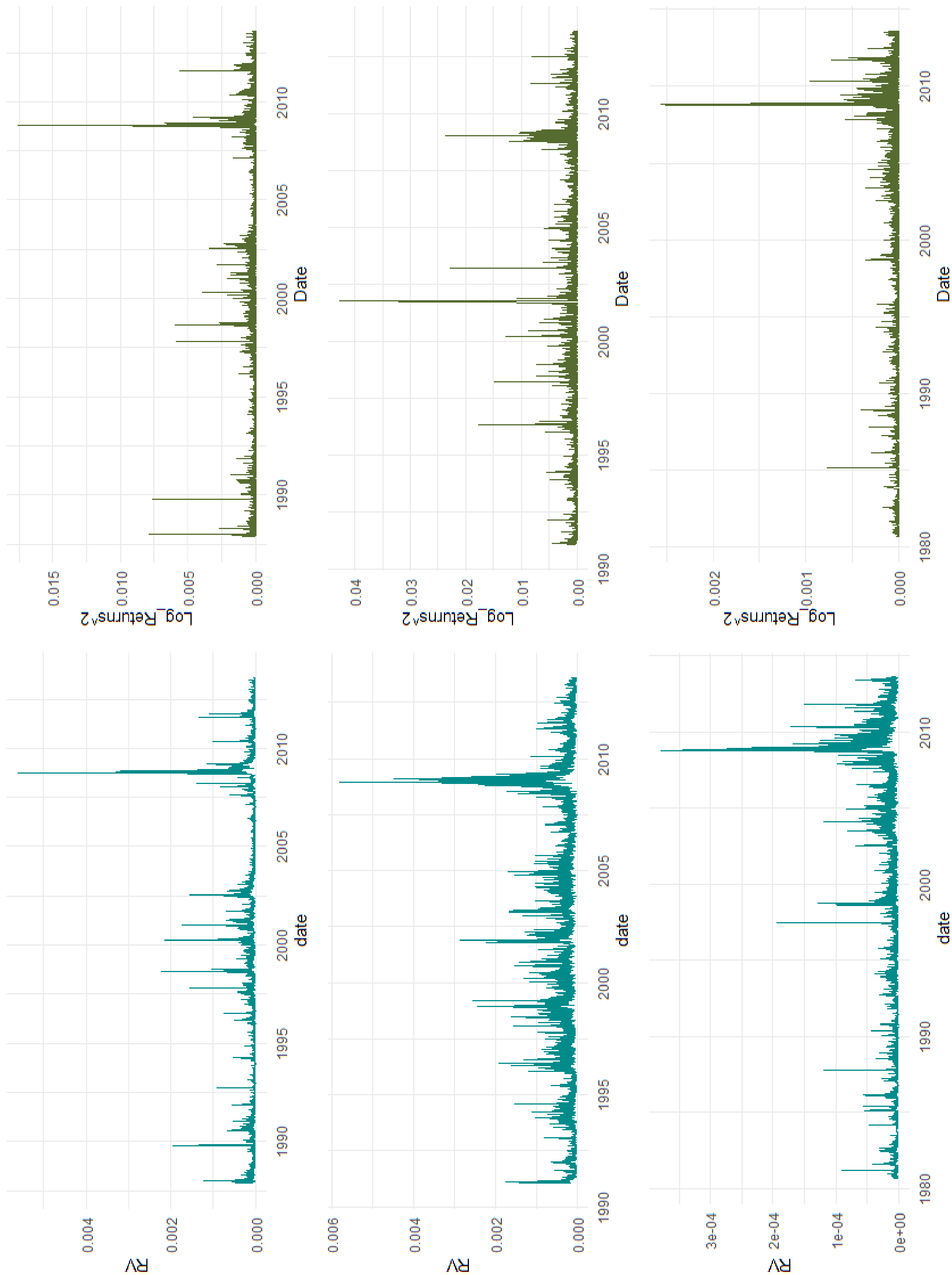
In this last section we give some additional information; the out-of-sample MSE and MAE, the figures for every series cleaned off outliers, as well as the in-sample HAR results and the S_{resid} series (cumulative sum of the difference in residuals between model i and OLS).

Model	S&P 500	WTI	USD/CAD
$OLS-HAR$	$6.235e^{-10}$	$3.259e^{-09}$	$6.891e^{-11}$
$WLS_{\hat{RV}}$	$6.175e^{-10}$	$3.237e^{-09}$	$6.881e^{-11}$
$WLS_{GA\hat{R}CH}$	$6.665e^{-10}$	$3.326e^{-09}$	$7.216e^{-11}$
$LAD-HAR$	$6.414e^{-10}$	$3.299e^{-09}$	$7.147e^{-11}$
$GARCH(1,1)$	$1.776e^{-04}$	$4.787e^{-04}$	$5.283e^{-05}$
$GARCH(1,2)$	$1.771e^{-04}$	$4.786e^{-04}$	$5.293e^{-05}$
$GARCH(2,1)$	$1.777e^{-04}$	$4.785e^{-04}$	$5.312e^{-05}$

Table 5.1: MSE

Model	S&P 500	WTI	USD/CAD
$OLS-HAR$	$1.608e^{-05}$	$4.057e^{-05}$	$4.713e^{-06}$
$WLS_{\hat{RV}}$	$1.591e^{-0}5$	$3.985e^{-05}$	$4.716e^{-06}$
$WLS_{GA\hat{R}CH}$	$1.687e^{-05}$	$4.072e^{-05}$	$4.989e^{-06}$
$LAD-HAR$	$1.547e^{-05}$	$3.903e^{-05}$	$4.540e^{-06}$
$GARCH(1,1)$	$1.047e^{-02}$	$1.766e^{-02}$	$5.711e^{-03}$
$GARCH(1,2)$	$1.043e^{-02}$	$1.767e^{-02}$	$5.718e^{-03}$
$GARCH(2,1)$	$1.047e^{-02}$	$1.766e^{-02}$	$5.732e^{-03}$

Table 5.2: MAE



RV
date
Log_Returns^2
Date
0.000
0.002
0.004
0.006
0e+00
1e-04
2e-04
3e-04
0.005
0.010
0.015
0.00
0.01
0.02
0.03
0.04
0.001
1980
1990
1995
2000
2005
2010

Table 5.3: OLS-HAR for the WTI

	Dependent variable:
	RV
Daily_Lag	-0.159^{***}
	(0.014)
Weekly	1.152^{***}
	(0.026)
Monthly	-0.031
	(0.023)
Constant	0.00001^{**}
	(0.00000)
Observations	4,992
R^2	0.631
Adjusted R^2	0.631
Residual Std. Error	0.0002 (df = 4988)
F Statistic	$2{,}840.045^{***}$ (df = 3; 4988)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.4: LAD-HAR for the WTI

	Dependent variable:
	RV
Daily_Lag	-0.115^{***}
	(0.017)
Weekly	0.955^{***}
	(0.038)
Monthly	0.038^{**}
	(0.019)
Constant	0.00000
	(0.00000)
Observations	4,992
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.5: $WLS_{\hat{RV}}$ for the WTI

	Dependent variable:
	RV
Daily_Lag	-0.171^{***}
	(0.013)
Weekly	1.193^{***}
	(0.025)
Monthly	-0.040^{**}
	(0.020)
Constant	0.00000
	(0.00000)
Observations	4,992
R^2	0.600
Adjusted R^2	0.599
Residual Std. Error	0.009 (df = 4988)
F Statistic	$2{,}489.221^{***}$ (df = 3; 4988)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.6: $WLS_{GA\hat{R}CH}$ for the WTI

	Dependent variable:
	RV
Daily_Lag	-0.216^{***}
	(0.016)
Weekly	1.363^{***}
	(0.029)
Monthly	-0.147^{***}
	(0.026)
Constant	0.00001^{**}
	(0.00000)
Observations	4,992
R^2	0.580
Adjusted R^2	0.580
Residual Std. Error	0.016 (df = 4988)
F Statistic	$2{,}297.079^{***}$ (df = 3; 4988)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.7: OLS-HAR for the S&P 500

	Dependent variable:
	RV
Daily_Lag	-0.024^{*}
	(0.014)
Weekly	1.004^{***}
	(0.018)
Monthly	-0.003
	(0.006)
Constant	0.00000
	(0.00000)
Observations	5,932
R^2	0.648
Adjusted R^2	0.648
Residual Std. Error	0.0001 (df = 5928)
F Statistic	$3{,}639.737^{***}$ (df = 3; 5928)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.8: LAD-HAR for the S&P 500

	Dependent variable:
	RV
Daily_Lag	-0.009
	(0.009)
Weekly	0.833^{***}
	(0.078)
Monthly	0.022
	(0.072)
Constant	0.00000^{**}
	(0.00000)
Observations	5,932
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.9: $WLS_{\hat{RV}}$ for th S&P 500

	Dependent variable:
	RV
Daily_Lag	−0.076***
	(0.014)
Weekly	1.070***
	(0.018)
Monthly	0.001
	(0.005)
Constant	0.00000
	(0.00000)
Observations	5,932
R^2	0.619
Adjusted R^2	0.618
Residual Std. Error	0.006 (df = 5928)
F Statistic	3,204.296*** (df = 3; 5928)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.10: $WLS_{\hat{GARCH}}$ for th S&P 500

	Dependent variable:
	RV
Daily_Lag	−0.285***
	(0.019)
Weekly	1.461***
	(0.023)
Monthly	−0.027***
	(0.007)
Constant	−0.00000***
	(0.00000)
Observations	5,932
R^2	0.593
Adjusted R^2	0.593
Residual Std. Error	0.011 (df = 5928)
F Statistic	2,880.929*** (df = 3; 5928)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.11: OLS-HAR for the USD/CAD

	Dependent variable:
	RV
Daily_Lag	-0.184^{***}
	(0.013)
Weekly	1.158^{***}
	(0.027)
Monthly	0.013
	(0.025)
Constant	0.00000
	(0.00000)
Observations	6,367
R^2	0.675
Adjusted R^2	0.675
Residual Std. Error	0.00001 (df = 6363)
F Statistic	$4{,}409.950^{***}$ (df = 3; 6363)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.12: LAD-HAR for the USD/CAD

	Dependent variable:
	RV
Daily_Lag	-0.138^{***}
	(0.012)
Weekly	0.958^{***}
	(0.035)
Monthly	0.078^{***}
	(0.021)
Constant	0.00000
	(0.00000)
Observations	6,367
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.13: $WLS_{\hat{RV}}$ for the USD/CAD

	Dependent variable:
	RV
Daily_Lag	-0.189^{***}
	(0.011)
Weekly	1.215^{***}
	(0.025)
Monthly	-0.034
	(0.021)
Constant	0.00000
	(0.00000)
Observations	6,367
R^2	0.661
Adjusted R^2	0.661
Residual Std. Error	0.002 (df = 6363)
F Statistic	$4,137.148^{***}$ (df = 3; 6363)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.14: $WLS_{GA\hat{R}CH}$ for the USD/CAD

	Dependent variable:
	RV
Daily_Lag	-0.316^{***}
	(0.015)
Weekly	1.649^{***}
	(0.030)
Monthly	-0.308^{***}
	(0.028)
Constant	0.00000
	(0.00000)
Observations	6,367
R^2	0.628
Adjusted R^2	0.628
Residual Std. Error	0.004 (df = 6363)
F Statistic	$3,576.384^{***}$ (df = 3; 6363)
Note:	*p<0.1; **p<0.05; ***p<0.01

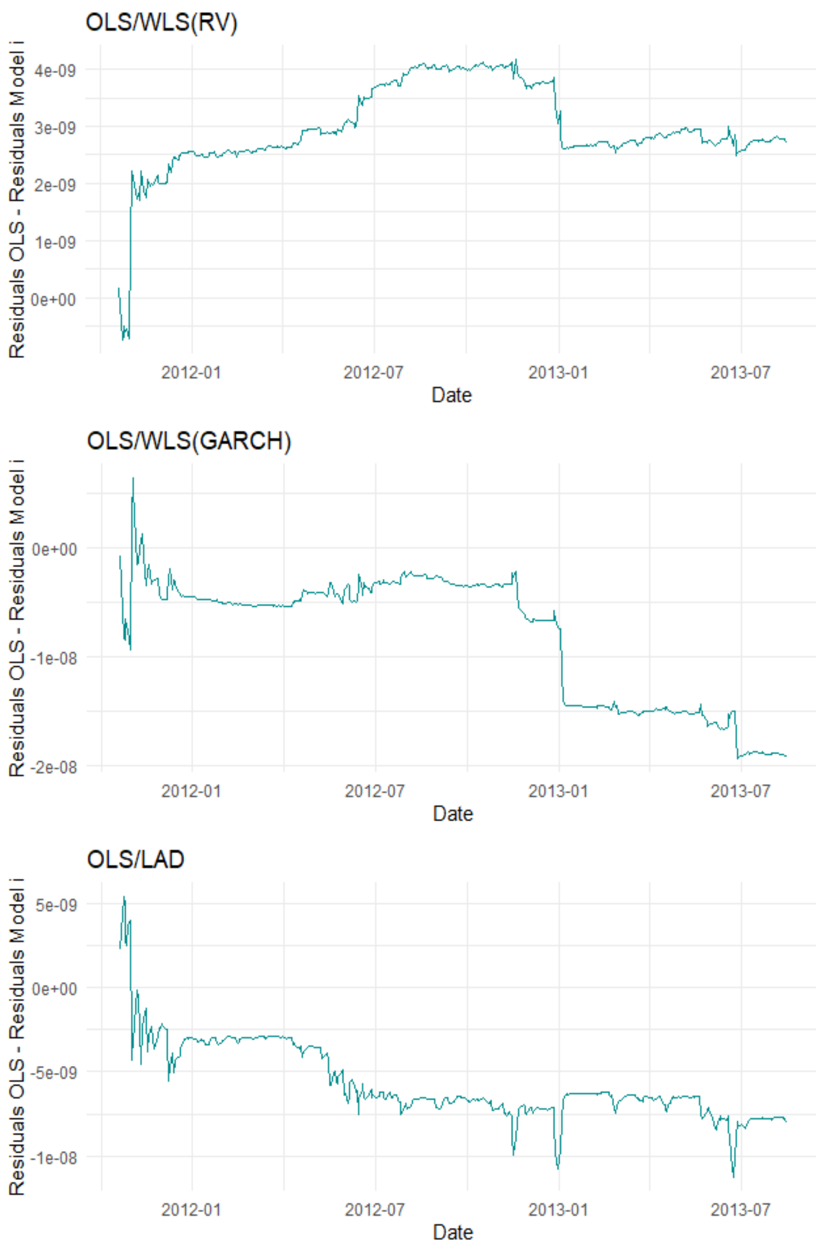


Figure 5.1: Cumulative Sum of $resid(OLS) - resid(model_i)$ - S&P 500

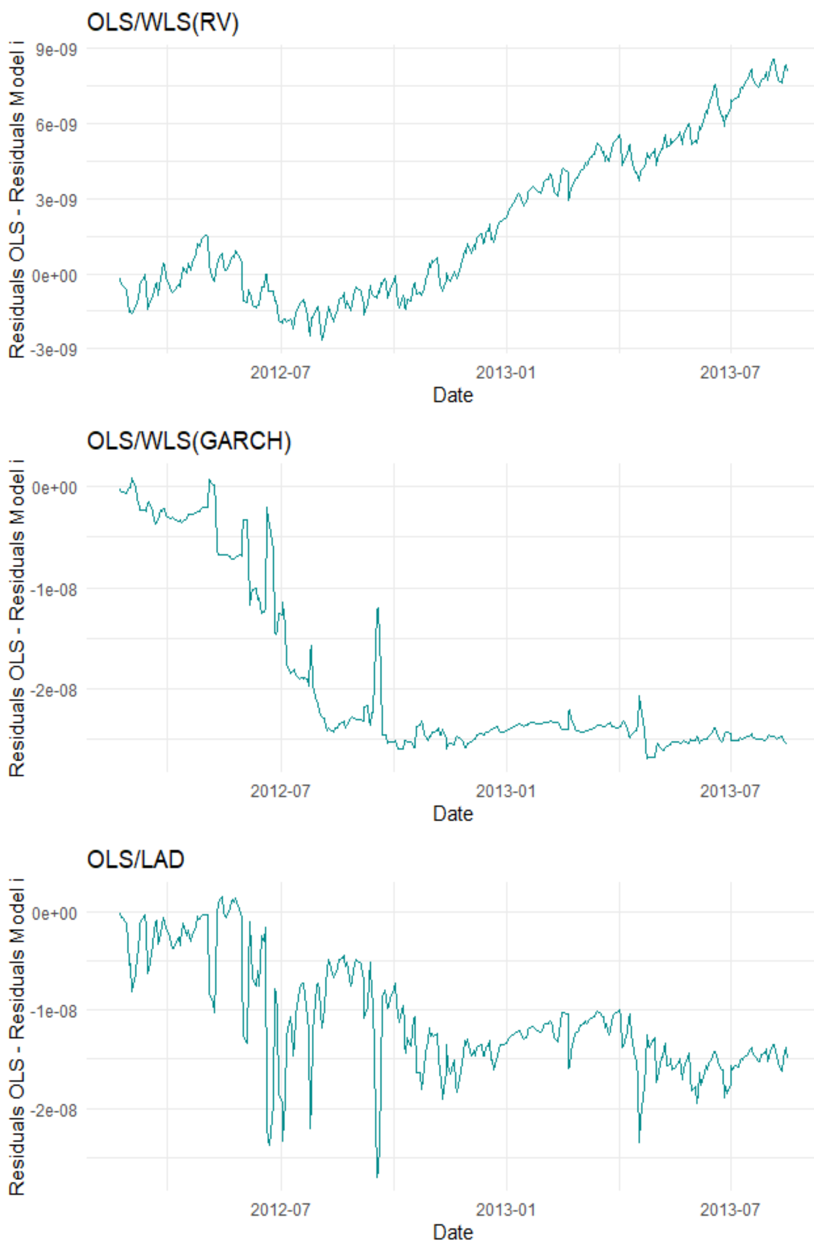


Figure 5.2: Cumulative Sum of $resid(OLS) - resid(model_i)$ - WTI

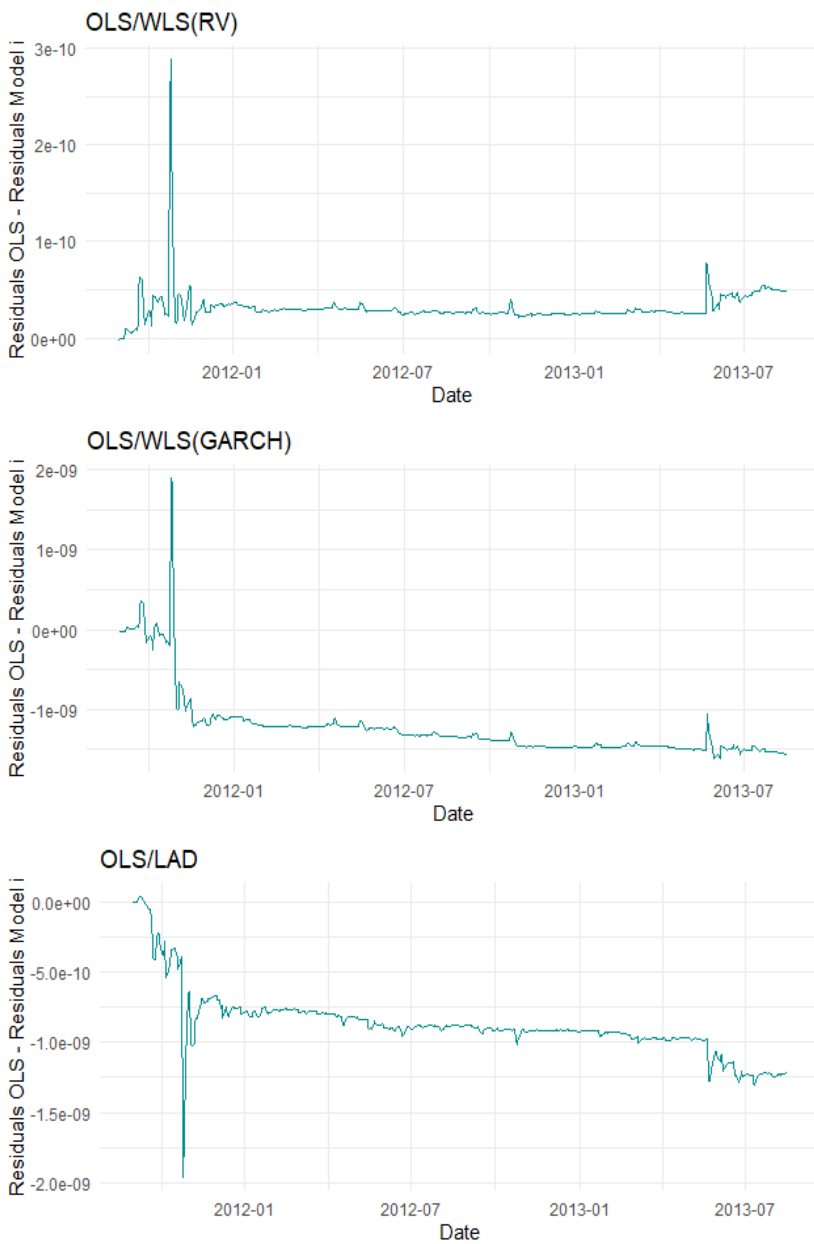


Figure 5.3: Cumulative Sum of $resid(OLS) - resid(model_i)$ - USD/CAD

Bibliography

[1] Botchkarev A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology, 2019.

[2] Clements A and Preve D. A practical guide to harnessing har volatility model. 2019.

[3] Dmouj A. *Stock price modelling:Theory and Practice*. PhD thesis, Vrije University, Netherlands, 2006.

[4] du Toit C and Conradie WJ. A realised volatility measurement using quadratic variation and dealing with microstructure effects. *ORiON*, 2006.

[5] Buncic D and Gisler K. The role of jumps and leverage in forecasting volatility in international equity markets. *Journal of International Money and Finance*, pages No. 79, 1–19, 2017.

[6] Fama E. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance No.25*, pages 383–417, 1969.

[7] Asiri F. *The Price of Stocks, Geometric Brownian Motion,and Black Scholes Formula*. PhD thesis, Windsor University, Ontario, Canada, 2018.

[8] Corsi F. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics No. 2*, pages 174–196, 2009.

[9] Ng H, S and Lam K, P. How does the sample size affect garch model? 2006.

[10] White H. Maximum likelihood estimation of misspecified models. *Econometrica No.50*, pages 1–25, 1982.

[11] Guidolin M. An introduction to the use of realized variance and covariance in risk management, 2019.

[12] Barndorff-Nielsen O, E and Shepard N. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics No. 17*, pages 457–477, 2002.

[13] Hansen P, R and Lunde A. An unbiased measure of realized variance. 2004.

[14] Tahar P and Turinici G. *Mouvement Brownien et Evaluation d'actifs contingents*. PhD thesis, Universite Paris Dauphine, 2013.

[15] Engle R, F. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, pages No. 50, 987–1008, 1982.

[16] Kim S and Kim H. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, pages No. 32, 669–679, 2016.

[17] Volgushev S and al. Significance testing in quantile regression. *Electronic Journal of Statistics*, pages No. 7, 105–145, 2013.

[18] Andersen T, G and Bollerslev T. Answering the skeptics:yes, standard volatility models do provide accurate forecasts. *International Economic Review*, pages No. 39, 885–905, 1998.

[19] Andersen T, G, Bollerslev T, and al. Modeling and forecasting realized volatility. *National Bureau Of Economic Research*, page Working paper No. 8160, 2001.

[20] Bollerslev T. Generalzied autoregressive conditional heteroskedasticity. *Journal of Econometrics*, pages No. 31, 307–327, 1986.