ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

MASTER THESIS [QUANTITATIVE FINANCE]

Option prices leading commodity futures prices

Author:

Edwin Ma Bsc (581003)

Supervisor:

Terri van der Zwan

Company Supervisors:

Daniël Schotanus

Ralph van de Klundert

Second assessor:

dr. Rutger-Jan Lange

June 23, 2023

Abstract

This research aims to investigate the notion of option prices leading the underlying commodity futures prices using several research questions. The most important result of this research is that implied volatility contains information on the future realised volatility of the underlying commodity futures contract beyond what is known from historical volatility. This is mostly in line with current understanding of implied volatility as a forward looking measure. This result was obtained by comparing the out-of-sample forecasts of a standard heterogeneous auto-regressive (HAR) model, which only includes historical volatility as inputs, to an extended HAR model, which additionally includes implied volatility as input. The forecasts of these two models were compared using the Diebold-Mariano test with mean-squared error as loss function. In fourteen out of the fifteen tested cases, the forecasts of the extended HAR model had a lower mean-squared error, than the standard HAR model. In nine of those cases, the extended HAR model also performed significantly better. This shows that in general the extended HAR model, which includes implied volatility as input, performs better than the standard HAR model, which means that implied volatility contains information on the future realised volatility beyond what is known from historical volatility.

Contents

1 Introduction

Information from option prices if often thought to be leading the future return distribution of the underlying asset. The idea behind this is that option buyers and sellers adjust their willingness to buy or sell an option based on their expectation of the future realised volatility of the underlying asset. Option buyers and sellers can adjust these expectations on the basis of likely future events, whereas a measure such as historical volatility does not encapsulate any information from likely future events. This would make option-implied information such as implied volatility a welcome variable to forecast the future realised volatility with. Empirical research has indeed shown that option prices in many ways contain information on the future return distribution of the underlying asset (Christoffersen et al. (2013) for a full overview). Most research agrees that option-implied volatility contains information on the future realised volatility of the underlying asset beyond what is known from historical realised volatility. Moreover, research suggests option markets contain information about the direction of the return distribution of the underlying, due to informed traders who would prefer to trade in option market rather than directly in the market for the underlying, see Cremers and Weinbaum (2010) and Xing et al. (2010). However, not a lot of research has been done on these phenomena when it comes to options on commodity futures. In this thesis I explore if the notion of implied volatility being a forward looking measure holds up when it comes to commodity futures as underlying. I do this on the basis of four research questions to be introduced in this section.

Market participants in the commodity market typically trade commodities through futures contracts, rather than the commodity itself. There also exist option contracts with commodity futures contracts as underlying asset. The ability to forecast commodity futures markets can be useful to market participants in a variety of ways, from portfolio optimization to risk management. Aside from market participants, government agencies such as the United Nations are also interested in the ability to foresee times of high or extremely high volatility. They recognize that high volatility can have a large and usually negative effect on producers and consumers, especially in the developing world, see Greb and Prakash (2017). The ability of these option contracts to forecast the future return distribution of the underlying commodity futures contract is the subject of this research.

This research considers five commodities with great impact on the economy, sufficient liquidity and that are all driven by their own distinct factors. These commodities are Corn (C), Crude Light (CL), Gold (GC), Natural Gas (NG) and Soybeans (S). All of these commodity futures and option contracts are traded on the Chicago Mercantile Exchange (CME). The data used in this research originates from the CME and is provided for this research by Transtrend. Throughout this thesis implied volatility is referred to as IV. Realised volatility, that is measured subsequently after IV is measured, is referred to as future realised volatility (FRV). Realised volatility, that is measured right before IV is measured, is referred to as historical volatility (HV). There are different methods to calculate IV. This thesis uses the CVOL method, which is a model-free method developed by the CME. The CVOL method uses option prices across all strike prices to come up with one estimate for implied volatility. Unlike the Black-Scholes model, the CVOL method does not make any assumptions on the return distribution of the underlying asset.

The first hypothesis of this thesis is: IV contains information on the future realised volatility of the underlying commodity futures contract beyond what is known from HV. This is in line with the current understanding of options and research, with a few notable exceptions. However, not a lot of research has been done on this phenomenon when it comes to commodity futures as underlying. I test this hypothesis by comparing the out-of-sample forecasts of the normal HAR model to the forecasts of an extended HAR model. The normal HAR model, as proposed by Corsi (2009), uses the HV of the last trading day, week, and month as exogenous variables to forecast the future realised volatility. The extended HAR model is similar to the normal HAR model, but includes IV as an exogenous variable. In fourteen out of the fifteen tested cases, the extended HAR model had a better out-of-sample performance in terms of mean squared error than the normal HAR model. In eleven out of these fourteen cases, the extended HAR model also performed significantly better according to a Diebold-Mariano test (again in terms of mean-squared error). This largely confirms the hypothesis that IV contains information on the future realised volatility of the underlying beyond what is known from HV.

These results suggest that market participants in option markets adjust their willingness to buy or sell an option based on their expectation of the future realised volatility of the underlying and that these expectation are good or at least add information to that is not captured in a model based that only takes HV into account. This is in accordance with the idea behind the Black-Scholes model and other option-pricing models.

The second hypothesis is that HV contains information on future IV beyond what is known from current IV. This hypothesis is informed by the fact that IV and HV seemingly move simultaneously and by Triantafyllou et al. (2015) who shows that HV contains information on the future IV, although crucially it does not show that HV contains information content on future IV beyond current IV. This hypothesis should therefore be seen as rather exploratory, whereas the first hypothesis is more in line with past research. To test the second hypothesis, I apply a Granger-Causality test to a VEC model to see if there is Granger causality from HV to IV. If this is the case, then I check if that VEC model can beat a random walk out-of-sample. If this is not the case, then it would not be fair to conclude that there is Granger causality from HV to IV, because if this was really the case, then surely the model would have beat a random walk. The second hypothesis is thereby rejected. In addition to Granger causality, I've also applied an instant causality test to test if FRV is informative of future IV. The test yielded a positive result in all cases. These result suggests that market participants in option markets have priced in all information of HV into IV and that new information of realised volatility is priced into IV on he same day.

The third hypothesis is that certain derivatives measures of IV contain information on the future direction of the underlying commodity futures contract. This hypothesis is informed by research including Cremers and Weinbaum (2010), who show that volatility spread predict returns of the underlying, Xing et al. (2010) who show that volatility skew predicts returns of the underlying and Bohmann (2020) who shows that volatility spread predicts return of underlying commodity futures. The CVOL allows for several different derivative measures. I've chosen the ones that are most similar to volatility skew as typically defined in the literature, although they are not exactly the same. The trading days of the sample are sorted five quantiles based on the tested measure. Then, a t-test is applied to see if the returns of the highest quantile are significantly different than the returns of the lowest quantile. The results were weak enough that they could be explained by pure coincidence. This means that this research found no evidence of this phenomenon, which could either be because it doesn't exist or because there is something unaccounted for in the tested measures or method.

The fourth hypothesis is that IV accounts for seasonal differences in FRV. Agricultural commodities have a clear seasonal pattern because of their harvest and energy commodities often have a seasonal pattern because there is more demand for energy in winter months, which could also affect volatility. Typically IV is thought of as a forward looking measure, meaning that it takes into account things that are known to be happening such as change of seasons. This would imply that IV would take seasonal differences in volatility into account. If the hypothesis is true, this means that it should already take seasonality into account and that models for forecasting

FRV don't have to incorporate seasonality, but can simply use IV as exogenous variable. This would make IV an even more desirable variable to forecast FRV. For this particular hypothesis I have only performed a heuristic check by plotting a a scatter plot of data points with IV as x-axis and FRV as y-axis sorted by season. In general it does appear that IV takes seasonality into account. This again supports the notion that IV is a forward-looking measure for FRV.

The remainder of this paper is structured as follows: Chapter 2 covers the existing literature on this topic; Chapter 3 describes the data used in this research, described the method by which implied volatility and realised volatility are calculated and reports summary statistics of these implied and realised volatility; Chapter 4 describes the methodology of this research; Chapter 5 discusses the results; Chapter 6 summarizes the main findings and Chapter 7 discusses the limitation of this research together with ideas for future research.

2 Literature review

2.1 Option pricing models and implied volatility

Throughout the economic literature, several models have been proposed to price options. Black and Scholes (1973) designed an option pricing model called the Black-Scholes model BS, that is still widely used today. Black (1976) created an option model specifically for commodities that is only slightly different compared to the original BS model. Black and Scholes argue that under certain assumptions, the payoff of an option contract can be replicating through the construction of a dynamic portfolio that takes positions in the underlying. By the law-of-one-price the value of an option should be equal to the value of the replicating portfolio. The value of a European option on a non-dividend-paying asset can then simply be expressed as a function of five variables namely: the price of the underlying asset, the time to maturity, the strike price, the interest rate and the volatility of the underlying asset. The only unknown variable in the equation is the volatility of the underlying. Since the price of the option is known, we can inverse the (BS) formula and get what is known as the implied volatility (IV).

One of the BS assumptions is that the return of the underlying asset is log-normally distributed. This is not in accordance with empirical findings on stocks and commodity futures, which show that their return distributions exhibit non-zero skewness and excess kurtosis. Based on the assumptions of the Black-Scholes model, IV's should be equal across strike prices. In practice, however, IV's are different across strike prices. Options with a strike that is close to the price of the underlying usually have a lower IV than strikes that are further away from the price of underlying, see Christoffersen et al. (2013) The pattern that usually emerges is refered to as the volatility smirk or skew, because of ATM options have lower BSIV than deep OTM or deep ITM options. The fact that this pattern remains so frequent shows that the BS model is misspecified. Despite this known misspecification, the BS model is still widely used in the literature and by practitioners, probably because of its simplicity, see Christoffersen et al. (2013).

It is possible to infer the entire risk-neutral density of the return distribution of the underlying asset from a continuum of quoted option prices, as proven by Breeden and Litzenberger (1978). All of this can be done in a model-free way, meaning no assumptions about the return distribution of the underlying are made. In practice only finitely many option prices are quoted. This means one needs to interpolate and extrapolate option prices between quoted strikes, in order to infer the risk-neutral density. This is either done by interpolating/extrapolating option prices directly or by interpolating/extraplating the BSIV between strikes, see Grith et al. (2012).[1]

In the 1990's and early 2000's model-free methods were developed to estimate implied volatility. These methods use the idea of variance swaps. Variance swaps are over-the-counter derivatives, where one party agrees to pay or receive a monetary sum equal to realized variance over a given period minus the agreed strike price. It can be shown that these variance swaps can be replicated by holding a dynamic portfolio of out-of-the-money (OTM) puts and calls. This replicating portfolio can thus be used to express implied volatility. This is the idea behind volatility indexes such as the VIX, a volatility index for the S&P 500, and the CVOL, a volatility index for, among others, commodities futures markets. The CVOL has certain derivative measures that indicative the volatility over one part of the volatility surface such as upward volatility, downward volatility and at-the-money volatility. These can be combined to create measures such as skewness and convexity.

A question often asked is whether implied volatility can act as a good predictor for future realised volatility. Research is mostly affirmative on this question. Poon and Granger (2003) considers the research up to that point and concludes option-implied volatility forecast seem strong compared to models based on historical volatility. This is probably because implied volatility is a forward looking measure, meaning it takes into account future events, whereas historical volatility is only backward looking. There are some papers, especially earlier ones,

[1] It should be noted that Grith et al. (2012) include a third approach as well, although to the best of my knowledge this approach is not suitable for commodities because it involves making some assumption on the pricing kernel.

that are very skeptical of the informative role of opion-implied information. The most notable of which is probably Canina and Figlewski (1993), who claim that IV has virtually no correlation with future realised volatility.

Jiang and Tian (2005) find that model-free implied volatility (MFIV) contains more information than BSIV when it comes to forecasting variance. They argue that this is due to the fact that MFIV contains information about BSIV because it contains information of option prices with all strikes, as opposed to BSIV which only contains information of one strike. This is a motivating factor to use model-free methods in this research.

It should be noted that implied volatility is only a risk-neutral indicator of future volatility. In simple terms this means that implied volatility only tells us what investors are willing to pay for things like variance swaps, but it does not directly tell us what the market expects future variance to be. The difference between implied variance and future realised variance is called the variance risk premium and is on average negative for stocks, see Carr and Wu (2009). This means that investors on average pay to be long in stock variance. It has been shown that energy markets exhibit a negative risk premium, see Bakshi and Kapadia (2003) and Doran and Ronn (2008).

There are methods to transform the risk-neutral measure into a physical measure, so as to get unbiased volatility forecasts. One of these is the Heston model, see Heston (1993). However, the Heston model does not seem to outperform implied volatility based on Black-Scholes or model-free methods (Christoffersen et al. (2013)).

Another method to transform the risk-neutral measure to the physical measure includes making some assumptions about the pricing kernel. A stock index, such as the S&P500 is expected to give positive returns over time, meaning that it can be assumed that times of low S&P500 return are correlated with higher marginal utility and vice versa, in order for the law-of-one-price to hold. Historical return data is then used to fit a pricing kernel.

To the best of my knowledge this approach is not suitable for commodity futures. The most important reason is that there is no natural long position in many commodities, in stark contrast to stock markets. Commodity futures are bought or sold by suppliers and users to hedge risks they hold, or by traders to make a profits. These commodity traders can be either long or short in the commodity future. Some commodity contracts are in normal backwardation, meaning that the price of the future is expected to rise as futures contract comes closer to expiration and some commodity contracts are in contango, meaning that futures prices are expected to decline as the future gets closer to expiration.[2] Sometimes the futures market for a commodity can switch from being in backwardation to contango or vice versa, see Galán-Gutiérrez and Martín-García (2022) and Fama and French (2016). All of this makes it extremely difficult to make any assumption on the pricing kernel of a commmodity future.

2.2 Informed trading in option markets

Evidence from the academic literature suggests that public information contained in the options market can predict the direction of the returns of the underlying equities. The most important and quoted papers in this area are Xing et al. (2010), who find that stocks with high volatility skew underperform stocks with low volatility skew, Cremers and Weinbaum (2010), who find that stocks with a high volatility spread outperform stocks with a low volatility spread, and more recently Bohmann (2020), who finds that both these measures are good predictors for the future return of a commodity futures markets during scheduled events. Additionally, Krams (2014), a master thesis, shows that agricultural commodities can be traded by using information from volatility spread. It is suspected that this market inefficiency is due to informed traders, who prefer to trade in options rather than directly in the underlying, due to a variety of reasons such as the added leverage of options.

The volatility skew is typically defined as the the BSIV of OTM put options minus the BSIV of OTM call options. Xing et al. (2010) sort stocks into quintiles based on the value of this volatility skew and show that, on average, the highest quintile underperforms the lowest quintile in the following week. Consequently the portfolio, consisting of long position in the lowest quintile and a long position in the highest quintile, has a positive returns on average. By using a Fama-Macbeth regression, they show that the returns from this portfolio cannot be explained by factor return.

The option volatility spread, a measure of apparent put-call parity violations, is also often thought of as a sign of informed trading and a predictor for future returns. The volatility spread is often defined as the difference between the implied volatility of at-the-money call options and put options matched on strike price and duration, the implied volatility is calculated using BS in the existing literature. Papers who use volatility spread as a predictor are Ofek et al. (2004) and Cremers and Weinbaum (2010). Cremers and Weinbaum (2010) sort stocks into quantiles based

[2]The definitions of normal backwardation and contango are taken from investopedia. See https://www.investopedia.com/articles/07/contango_backwardation.asp

on the level of their volatiliy spread. They show that the highest quantile outperforms the lowest quantile and create a portfolio that goes long in the highest quantile and short in the lowest. They show that the returns of this portfolio cannot be explained by factor returns. Additionally they find that the adjusted returns in the latter half of their data sample from January 2001-December 2005 are lower than in the first half of their sample from January 1996– December 2000. They suspect this is due to the market becoming more efficient.

Bohmann (2020) looks at the volatility skew and spread in commodity options before announced events. He calculates volatility skew and spread in the same way as Xing et al. (2010) and Cremers and Weinbaum (2010). He takes the abnormal volatility skew (or spread) as the difference between the current vol skew minus the average volatility skew (or spread) of futures of the same commodity before that date. Besides that he also corrects the volatility spread for (a lack of) liquidity, the idea being that violations of the put-call parity are higher on average when liquidity is low, because arbitrageurs have a harder time exploiting arbitrage opportunities due to a lack of liquidity. This means that volatility spread is less indicative of activity from insider trading, when liquidity is low. He then uses a panel regression to regress the abnormal vol skew or vol spread of the days before a scheduled event against the returns of the underlying in the days after the event. He finds that vol skew and spread are significant predictors. My research will differ from Bohmann (2020) in that I will use model free methods to compute vol skew and spread, and that I will look at how the general relationship evolves through time instead of just days before an event.

Krams (2014) specifically looks at vol spread as a predictor in agricultural commodities. He uses the same definition of volatility spread as Cremers and Weinbaum (2010). Every day the volatility spread is calculated for each commodity and compared against the last 40 trading days. If the volatility spread of a certain commodity future is in the top 10% of the last trading days, that commodity will get a long signal, and if the volatility spread is in the bottom 10% of the last 40 trading days, the commodity future will get a short signal. However if there is a long signal, but the value of the vol spread is still negative, the long signal will be canceled and vice versa for a short signal. A portfolio is then crafted which goes long in commodity futures with a long signal and short in commodity futures with a short signal. This portfolio yields a high positive alpha.

3 Data

3.1 Data source

The data used in this research comes from the Chicago Mercentile Exchange(CME) on five commodities namely: Corn(C), Crude Light(CL), Gold(GC), Natural Gas(NG) and Soybeans(S). The data consists of daily futures and option prices of the Open, High Low and Close. On any day there are multiple active contracts of one commodity with different delivery dates. Taking all these contracts into account could lead to overlapping problems, which could lead to statistical problems. To counter this I will work with a continuous contract, which can be seen as a portfolio that shifts from contract to contract. The portfolio of the continuous contract holds the contract with the shortest maturity on the condition that the days to expiration of the option contract is more or equal to 22 days.

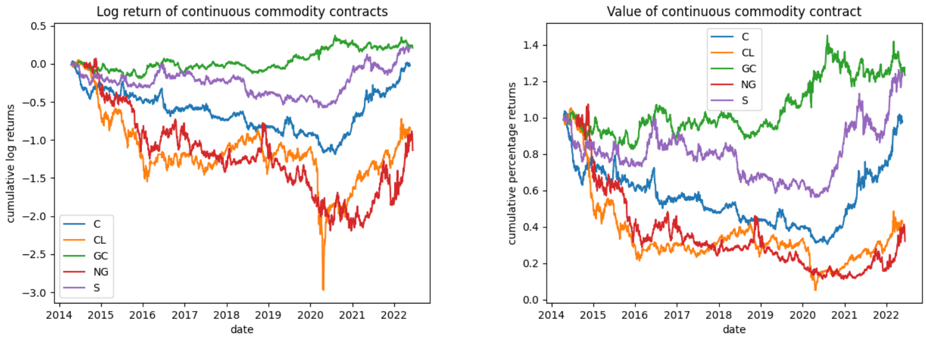
Table 1 shows the traded futures expiration months of the commodities. All of these will be used in this research. Gold futures contracts with expiration months January, March, May, July, September, and November also exist and are traded. There is however so little volume and open interest in these contracts that analysing them is difficult and perhaps not representative of the broader market. They will therefore not be used in this research.

Figure 2a shows the cumulative log returns of the continuous contract and figure 2b shows the cumulative percentage returns of the contract. As we can see all commodities, except for gold, were going down until the middle of 2020. After that they mostly went up. There are a couple of date points where the price of crude light futures was under zero. These date points will be excluded from the data, meaning they will be seen as non-market days (such as weekends) , because realised volatility can't be calculated by conventional methods during those days.

Table 1: **Products considered in this research**

Commodity	**Exchange**	**Traded future expiration months**	**Symbol**
Corn	CME	MAR,MAY,JUL,SEP,DEC	C
Crude Light	CME	JAN,FEB,MAR,APR,MAY,JUN,JUL,AUG,SEP,OCT,NOV,DEC	CL
Gold	CME	FEB,APR,JUN,AUG,OCT,DEC	GC
Natural Gas	CME	JAN,FEB,MAR,APR,MAY,JUN,JUL,AUG,SEP,OCT,NOV,DEC	NG
Soybeans	CME	JAN,MAR,MAY,JUL,AUG,SEP,NOV	S

Figure 2: **Returns of continuous contract**



(a) Log returns of continuous commodity contract

(b) Percentage returns of continuous commodity contract

3.2 Choice of IV and RV estimator

As discussed in the literature section, multiple methods have been proposed to calculate IV. In this research, the CVOL method is used as a measure of implied volatility. The CVOL method is a model-free method, based on the idea of a variance swap, that calculates IV using all available prices of OTM options. For a brief introduction to the CVOL method, see Appendix A, for the exact procedure to calculate see the CVOL whitepaper.[3] The only alteration made in this thesis is to calculate the CVOL of one specific futures contract rather than a combination of them. This ensures that there is a fair comparison between the IV and the realised volatility of the underlying. The CVOL method can also be used to describe certain characteristics of the implied volatility curve such as skew and convexity, which will be used to investigate a possible relationship between the IV surface and the future return of the underlying.

Realized volatility is not directly visible, but has to be estimated using an appropriate estimator. Multiple estimators hasve been proposed throughout the literature. For this research the Yang-Zhang Volatility estimator (YZ-estimator) is used, which is an unbiased and drift-independent estimator, that uses the prices at the open, high, low, and close of a trading day, see Yang and Zhang (2000).[4] Another possibility would be to use the close-to-close estimator. However the close-to-close estimator is drift-dependent, whereas the YZ-estimator is drift-

[3] https://www.cmegroup.com/market-data/cme-group-benchmark-administration/files/cvol-methodology.pdf

[4] Sometimes the Yang-Zhang estimator is called the Yang-Zhang-Garman-Klass volatility estimator, because it is an extension of the Garman-Klass estimator. For brevity, I will refer to it only as the Yang-Zhang estimator.

independent. Furthermore the YZ-estimator takes into account intraday volatility by using the high-low prices, whereas the close-to-close estimator does not.

$$\hat{\sigma}_{t,t+n} = \sqrt{\frac{252}{n} * \sum_{i=t}^{n+t} \left[\log^2\left(\frac{O_i}{C_{i-1}}\right) + \frac{1}{2}\log^2\left(\frac{H_i}{L_i}\right) - (2\log(2)-1)\log^2\left(\frac{C_i}{O_i}\right) \right]}, \qquad (1)$$

where $\sigma_{t,t+n}$ is the realised volatility from day t to day $t+n$, O_t, H_t, L_t, C_t denote the price of the open, high, low and close of trading day t.

Realised volatility, that happens right before IV is measured, will be referred to as historical volatility (HV) and realised volatility, that happens right after volatility is measured, will be referred to as future realised volatility (FRV). This means that FRV is the same as time-shifted HV of the same contract. HV is known at the time of measurement, while FRV is unknown at the time. This means HV and FRV are defined as follows:

$$HV_{d,t} = \sigma_t \qquad (2)$$

$$HV_{w,t} = \sigma_{t-4,t} \qquad (3)$$

$$HV_{m,t} = \sigma_{t-21,t} \qquad (4)$$

$$FRV_{d,t} = \sigma_{t+1} = HV_{d,t+1} \qquad (5)$$

$$FRV_{w,t} = \sigma_{t+1,t+5} = HV_{w,t+5} \qquad (6)$$

$$FRV_{m,t} = \sigma_{t+1,t+22} = HV_{m.t+22} \qquad (7)$$

3.3 Summary statistics

Table 2 shows the means of the variables. The mean of IV is higher than the mean of FRV_m for Corn, Gold, Nat Gas and Soybeans. This would indicate that the variance risk premium is negative for these commodity futures, which is in line with the literature. On the other hand, the mean IV of CL is lower than the mean of FRV_m, which would indicate that the variance risk premium is positive for Crude Light futures, which is not in line with the literature. Table 2 shows the standard deviation of the variables. Figure 6 in Appendix B show the realised volatility over the last month (HV_m) together with the IV.

Tables 10, 11, 12, 13, and 14 show the correlation between variables. There are a couple of things worth noting. The first thing to notice is that FRV is positively correlated with HV. This indicates volatility clustering, which is a widely recognized phenomenon in the financial sector and academia. The second thing to notice is that FRV is more correlated with IV than with HV, which would indicate that IV holds predictive information on FRV that is not reflected in HV. A third thing to notice is that IV is more correlated with HV than it is with FRV. This could indicate that IV is reacting to changes in price swings of the underlying or that events that influence volatility and implied volatility happen simultaneously.

Table 2: Table of means

	C	CL	GC	NG	S
IV	25.060	38.712	15.726	48.011	20.158
HV_d	19.817	34.097	13.263	37.800	17.082
HV_w	20.634	35.436	13.858	38.815	17.752
HV_m	21.060	36.025	14.115	37.807	18.034
FRV_d	19.845	34.172	13.268	37.994	17.102
FRV_w	20.715	36.137	13.919	39.865	17.797
FRV_m	21.468	39.460	14.395	42.768	18.306

This table reports the means of IV and RV's.

3.4 Unit root test

An important prerequisite for using time series data in an econometric model, such as OLS model, is that the time series do not contain any unit roots. If there does exist a unit root, then the model will generally be misspecified (Phillips (1987)). Therefore, the realised volatility data and implied volatility data is tested for the existence of a unit root using the Augmented Dickey-Fuller (ADF) test. The null hypothesis of an ADF test is that the time series does contains a unit root. If the p-value of the ADF test is significant at the 5% level, the null hypothesis can be rejected. If, on the other hand, the p-value of the ADF test is not significant, then the null hypothesis cannot be rejected, but this does not mean there must exist a unit root. It could just be that there is simply not enough evidence to rule out the existence of a no unit root, even though it is not present.

The results of the ADF test can be found in Table 3. The null hypothesis of a unit root can be rejected at the 5% significance level for all commodities and time spans, except for weekly realised volatility of Nat Gas, which has a p-value of 0.0545, which is just above the significance level. This is probably due to the low power of the test, rather than the existence of an actual unit root. Thus, we continue with the analysis as usual. Unit roots will not be further mentioned in the methodology or results section.

Table 3: **p-values of Augmented Dickey-Fuller (ADF) test**

	ADF p-values			
variable	HV_d	HV_m	HV_w	IV
Commodity				
C	0.0000**	0.0068**	0.0000**	0.0019**
CL	0.0023**	0.0019**	0.0043*	0.0019**
GC	0.0000**	0.0053**	0.0000**	0.0089**
NG	0.0236*	0.0147*	0.0545	0.0001**
S	0.0000**	0.0032**	0.0000**	0.0000**

($^{*}p < 0.05,^{**}p < 0.01$)

4 Methodology

This research is about the relationship between implied volatility and the future return distribution of the underlying future, particularly the volatility of the underlying future. There are several aspects of this relationship that are investigated and different methods are used depending on which relationship is analysed. The hypothesis of this research are as follows:

H1: IV contains information on the future realised volatility of the underlying commodity futures contract beyond what is known from HV.

H2: HV contains information on future IV beyond what is known from current IV.

H3: Certain derivative measures of IV contain information on the future direction of the underlying commodity future.

H4: IV accounts for seasonal differences in FRV. (This hypothesis is only tested for heuristically.)

The first hypothesis, **H1** is tested for by comparing the out-of-sample forecasts of a standard heterogeneous auto-regressive (HAR) model, which only includes HV as inputs, to an extended HAR model, which additionally includes IV as input. The forecasts of these two models are compared using the Diebold-Mariano test with mean-squared error as loss function. This is further explained in section 4.1. The second hypothesis, **H2** is tested for by applying a Granger-causality test on a VEC model, which uses realised volatility and implied volatility as inputs and outputs. This is further explained in section 4.2. The third hypothesis, **H3** is tested for by sorting the trading days of a commodity futures in five quantiles based on a derivative measure of IV. A Welch t-test is performed to check if the highest quantile outperforms or underperforms the lowest quantile. This is further explained in section 4.3. The fourth hypothesis, **H4**, is heuristically explored by plotting the IV of a commodity futures contract against their FRV in a scatterplot. The data points on the scatterplot are given a colour based on the season in which the futures contract ends. This is further explainend in 4.4.

4.1 HAR models

In order to test **H1**, the heterogeneous auto-regressive (HAR) model is extended to include IV as input. The forecasts of the original HAR model are compared against the forecasts of the extended HAR model using a Diebold-Mariano test with mean-squared error as loss function. If the forecasts of the extended HAR model performs significantly better according to the Diebold-Mariano test, it is concluded that IV contains information on the future return distribution of the underlying future beyond what is known from historical volatility and **H1** is accepted. This procedure is done for all five commodities and for three forecasting time spans namely: one day (1 trading day), one week (5 trading days), and one month (22 trading days). This means that there is a total of 15 cases. The model parameters of the HAR models is estimated by the OLS algorithm. However, the OLS algorithm can be sensitive to outliers, therefore the analysis is repeated with a robust regression. If the robust regression performs better out-of-sample in terms of mean-squared error than with OLS under the same commodity, time span and model, the conclusions are drawn from the forecasts that arose from the robust regression.

Additionally the original HAR model and the extended HAR model are compared in-sample together with the IV-only model, which as the name suggests only takes IV as input. In-sample testing is done to shed light on the out-of-sample results. It is also used to see if the inlcusion of HV as input improves the model.

The model specifications of the HAR models and the IV-only are defined in subsection 4.1.1, the procedure for in-sample testing is explained in subsection 4.1.2, the details of the estimation window, the loss function and the math behind the Diebold-Mariano test are explained in 4.1.3 and the robust regression is explained in 4.1.4.

4.1.1 HAR models specification

Volatility forecasting is a wide subject in which many models have been proposed. However, two models seem to dominate the field at present. The first of which is the generalised autoregressive conditional heteroscedasticity (GARCH) model. The second model, which is seemingly gaining in popularity, is the heterogeneous autoregressive (HAR) model. The HAR model was first proposed by Corsi (2009) and has gained in popularity ever since. It has been lauded for its consistently good performance and the simplicity of the model. The HAR model is the most suitable for this research, because of its good performance and its easy applicability to the term structure of commodity futures.

The standard HAR model, which will be referred to as HAR-RV throughout this thesis, is defined as:

$$FRV_t = \beta_0 + \beta_1 HV_{d,t} + \beta_2 HV_{w,t} + \beta_3 HV_{m,t} + \varepsilon_t. \tag{8}$$

The term on the left hand side FRV_t can stand for $FRV_{d,t}$, $FRV_{w,t}$ or $FRV_{m,t}$. The HAR-RV can be extended by adding a term for implied volatility. This model will be referred to as the HAR-IV model throughout this thesis, in accordance with the terminology of Liang et al. (2020). The HAR-IV model is defined as:

$$FRV_t = \beta_0 + \beta_1 HV_{d,t} + \beta_2 HV_{w,t} + \beta_3 HV_{m,t} + \beta_{IV} IV_t + \varepsilon_t. \tag{9}$$

The HAR-RV and HAR-IV models will be compared by their in-sample and out-of-sample forecasts. Additionally, the IV-only model is defined by:

$$FRV_t = \beta_0 + \beta_{IV} IV_t \tag{10}$$

The IV-only model will be compared in-sample against HAR-IV. Notice that the HAR-RV model and the IV-only model are nested in the HAR-IV model.

4.1.2 In-sample testing

Before testing the forecasts of the HAR models for out-of-sample, the results will first be tested in-sample. While we are more interested in out-of-sample results, in-sample testing can aid us in shedding light on the out-of-sample results. It also has the advantage of taking into account the whole sample. When comparing model fits in-sample, it is important that the independent variables are independent of each other. The month-ahead realised volatility, $FRV_{m,t}$, is used as the regressant. The sample is sorted into a monthly sampling frequency to avoid overlap in the dependent variables of the regression.

A regression is performed on the HAR-IV, HAR-RV, and IV-only models. A partial F-test is performed comparing the the HAR-IV model and HAR-RV model, to evaluate if the inclusion of IV as a regressor significantly improves the R^2 of the model. Another partial F-test is performed comparing the HAR-IV model and IV-only model, to evaluate if the combined inclusion of HV_d, HV_w and HV_m as regressors significantly improves the R^2 of the model.

4.1.3 Out-of-sample testing

This section explains the out-of-sample testing of the HAR-IV and HAR-IV models used in this research. The models will forecast over a period of one day, one week, and one month. The regression parameters of the HAR models will be estimated over a rolling window. The size of the rolling window will consist of 20% of the sample for a forecast of the volatility of the next day. 30% of the sample for a forecast of the volatility of the next week and 40% of the sample for the volatility of the next month.

The parameters of the HAR-IV and HAR-RV model will be estimated with the Oridnary Least Squares (OLS) method. The OLS algorithm can be sensitive to outliers, which can have a disproportionate effect on the estimated model parameters. To test if this disproportionate effect is present, the analysis is performed again in the same way, except that the regression estimated will be estimated with a robust regression as proposed by Clements and Preve (2021). The details of this robust regression are explained in subsection 4.1.4. The forecasts of the models will be compared using a Diebold-Mariano test with the mean-squared error (MSE) as loss function, which is further explained in this subsection.

Loss functions are used to compare forecasts. In this research the mean squared error (MSE) loss function is used to evaluate the out-of-sample performance. The MSE loss function is defined by:

$$L_t^{MSE}(y_t, \hat{y}_t) = (y_t - \hat{y}_t)^2 \tag{11}$$

This loss function is chosen because the OLS method minimizes the in-sample MSE, which means that the same loss function is used to judge the out-of-sample forecasts as to estimate the in-sample regression parameters. This is the fairest way to compare models. Besides this fact, it is also a very commonly used loss function. Remember that $R^2 = 1 - (MSE/TSS)$, where TSS is the total sum of squares. This means that R^2 and MSE are directly related. Because of this, and the fact that R^2 is far easier to interpret and compare across commodities, the R^2 will be reported as a measure for forecast performance in the tables.

The Diebold-Mariano test (DM-test) will be used to test if the HAR-IV forecast are significantly better than the HAR-RV forecasts. Let the loss differential, d_t, be defined by:

$$d_t = L(f_1(t)) - L(f_2(t)). \tag{12}$$

In essence, the Diebold-Mariano test is a Z-test on the loss differential. As in the standard Z-test the test statistics is obtained by standardizing the mean loss differential

$$\frac{\sqrt{T}(\bar{d} - \mu)}{\sigma_d} \to N(0,1), \tag{13}$$

where σ_d is a heteroskedasticity and autocorrelation consistent estimator (HAC-estimator).

Under the null hypothesis, the average difference between the loss function of the HAR-IV and HAR-RV model is equal to zero. This means that the added IV factor does not improve forecast whatsoever. This implies that the true slope coefficient β_{IV} should be equal to zero. This means that as the window over which the slope coefficients tends to infinity, the slope coefficients of the extra variables will tend to zero. This in turn would mean that, as the estimation window tends to infinity, both models would tend towards the same predictions. This would mean that the loss differential is constant and equal to zero and that therefore variance of the loss differential is infinite. Which would means that equation 13 does not hold anymore, which would be a problem. However if the window to estimate the slope coefficients does not tend to infinity, this problem is not present because there is always some inherent inaccuracy in estimating coefficients. In practice this means taking a rolling window to estimate the parameters rather than an expanding window will prohibit this problem from occurring. This is also what is done in this research.

When comparing out-of-sample forecasting between nested models, a one-sided test is standard practice. The argument is that we are interested in the information content of a certain variable. In this case we are interested in the information content of IV on FRV. A variable either has information content or not. It cannot have negative information content. IV can either help us better predict FRV or not, but it is non-sensical to state that implied volatility will help us predict FRV worse. If the nested model does perform better than the overarching model, then this is due to over-fitting of the overarching model. It is not due to IV having negative information content. This means we can use a one-sided DM-test to evaluate out-of-sample forecasts of nested models. For further discussion on the justifications for the use of a one-sided test see Clark and McCracken (2013). This research also uses a one-sided test to compare the HAR-IV and HAR-RV forecasts.

The DM-test gives a conservative p-value when small sample sizes are used. To overcome this problem, the method of Harvey et al. (1997) is used in this thesis. This method slightly alters the DM-statistic to a test statistic that follows a t-distribution.[5]

4.1.4 Robust regression

As stated before the OLS estimator can be sensitive to outliers, therefore the analysis is performed again using a robust regression. The robust regression is done using the Tukey-Biweight estimator. The Tukey-Biweight estimator is a so called M-estimator, which means that it estimates the model parameters by minimizing a certain in-sample loss function ρ. In the case of HAR-IV, this means::

$$\min_{\beta_0,\beta_1,\beta_2,\beta_3} \left[\sum_{t=1}^{n} \rho \left(FRV_t - \beta_d - \beta_1 HV_{d,t} - \beta_2 HV_{w,t} - \beta_3 HV_{m,t} \right) \right] \tag{14}$$

where ρ is a pre-specified symmetric function with a unique minimum at zero. The OLS estimator is also an M-estimator, where $\rho(e) = x^2$. The mean absolute deviation estimator is defined by setting $\rho(e) = |e|$. The loss function ρ of the Tukey-Biweight estimator function is defined as:

$$\rho(e) = \begin{cases} \frac{k^2}{6}\left(1 - \left[1 - \left(\frac{e}{k}\right)^2\right]^3\right), & |e| \leq k \\ \frac{k^2}{6}, & |e| > k \end{cases}, \tag{15}$$

where k is the tuning constant. For this research, the Statsmodels Robust Linear Models module in Python is used to calculate the Tukey-Biweight estimator with its defualt setting of $k = 4.685$. This choice of value for the tuning constant is perhaps arbitrary, but any choice of tuning constant is ultimately arbritrary. Besides, other research also sets the tuning constant based on the default

[5] A small summary of this method can be found on https://www.lem.sssup.it/phd/documents/Lesson19.pdf

setting of the program that they happen to be using. The Tukey-Biweight loss function is plotted in Figure 3 to get a general idea of what the values of this function are. The Tukey-Biweight estimator is considered a robust estimator, because it gives less value to outliers and is therefore less affected by them. Calculating the Tukey-Biweight estimator is done by iteratively applying weighted least squares. More information on how this last bit is done can be found in Clements and Preve (2021).

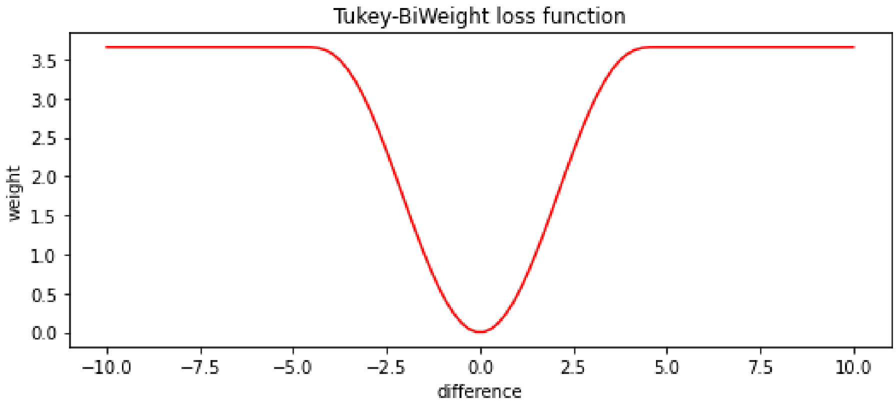


Figure 3: The Tukey-Biweight loss function (k= 4.685)

To be clear the Tukey-Biweight loss function is <u>not</u> used to compare the out-of-sample forecasts. The regression parameters are estimated using the robust regression by minimizing Tukey-Biweight loss-function over the estimation window. The out-of-sample forecasts that arise from this robust regression are then compared by the mean-squared error loss function through a Diebold-Mariano test, identical to how the forecasts of the OLS regression are compared. This means the OLS regression compares forecasts based on the same loss function that is used to estimate the regression parameters, whereas the forecasts of the robust regression are compared using a different loss function to the one that is used to estimate the regression parameters. If the forecasts of the robust regression are better (meaning lower mean-squared error) than the forecasts from the OLS regression, it is concluded that the OLS regression was too sensitive to outliers. In this case, the result from the Diebold-Mariano test is taken from the comparison of the robust regressions and the result from the Diebold-Mariano test from the OLS regressions is dismissed. If this is not the case, the result from the Diebold-Mariano test from the OLS regression is taken as true result.

4.2 VEC model

In order to test **H2**, a vector error correction (VEC) model is utilized with realized volatility and implied volatility as inputs and outputs. A Granger-causality test is applied to check if historical volatility contains information on the future implied volatility beyond what is known from current implied volatility. If this is the case, the VEC model is compared against a random walk out-of-sample. If the forecasts of the VEC model do not significantly outperform a random walk, the VEC model is considered to be misspecified and the result of the Granger-causality test is deemed inconclusive. If the forecasts of the VEC model do outperform a random walk, the result form the Granger-causality test is considered conclusive and **H2** is accepted. The vector error correction (VEC) model allows us to apply a vector autoregressive (VAR) model on two co-integrated variables. Judging from Figure 6 in Appendix B, there does seem to be co-integration between realised volatility and implied volatility. This naturally leads us to wonder if a VEC model can be utilized effectively.

Before applying the model, an Engle-Granger co-integration test is performed on IV and HV. If this test shows that there is co-integration between the two variables, then a VEC model as is applied as described below. I will use the fewest amount of lags in this model, namely one, because the added effect of HV and IV will likely be more clear to witness with fewer lags. The VEC model as is used here is defined by:

$$\begin{bmatrix} IV_{t+1} \\ HV_{t+1} \end{bmatrix} = \begin{bmatrix} IV_t \\ HV_t \end{bmatrix} + \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} (IV_t - \beta_0 - \beta_1 HV_t). \tag{16}$$

Aside from Granger causality, we can also test for instantaneous causality. A variable x instant causes another variable y if the future value of y, y_{t+1} can be better predicted using the future value of x, $_{t+1}$ in addition to the current and past values of y and x. Instantaneous causality is symmetric meaning that x instant causes y if and only if y instant causes x. Instanteneous causality is tested by using an instantaneous causality test, that tests if the residuals of x and y are correlated. Instant causality cannot aid us in forecasting, at least when we do not know the future value of x or y. It can only help us in potentially providing an explanation for the strong relationship between variables.

If the Granger causality test shows that there is Granger-causality from RV to IV, out-of-sample forecasts for IV from the VEC model will be taken and compared against the forecasts of a random walk. The comparison will be done with a Diebold-Mariano test with mean-squared error as loss function, similar to how the forecasts from the HAR models were compared. The estimation window for the variables will consist of 50% of the data sample. The DM test will be double sided since there is no reason for a one-sided test as with the nested models. If the forecasts from the VEC model could not significantly outperform the random walk, the VEC model is considered the previous result from the Granger-causality test is deemed in conclusive

4.3 Directional analysis

In order to test **H3**, the performance of a commodity futures on trading days when the derivative measure is high are compared to trading days when the derivative measure is low. I was not able to recreate the exact derivative measure that is use in Xing et al. (2010), but I was able to recreate some derivative measures that are related to volatility skew as defined in the literature by using the CVOL methodology as described in the whitepaper. In addition, other derivative measures that are not related to skew are also tested. An overview of the CVOL methodology can be found in appendix A. In short, the CVOL method prices from out-of-the-money puts and calls to calculate IV. Additionally one can calculate IV based on out-of-the-money calls alone, which is called $UpVar$ or from out-of-the-money puts alone, which is called $DnVar$. This naturally leads to two derivative measures for skew, which are defined by:

$$\text{skew1} = \text{UpVar} - \text{DnVar} \tag{17}$$

$$\text{skew2} = \text{UpVar}/\text{DnVar} \tag{18}$$

The derivative measures that I will test are skew1, skew2, Δskew1, which is the difference between skew1 of today and yesterday, abnormal skew, which is the difference between the average skew of the last three days minus the average skew of the last 40 trading days, convexity, which is a measure of the curvature of the volaitlity curve, and IV itself, as calculated by the CVOL methodology.

I sort the trading days of a commodity into 5 quantiles based on the tested derivative measure. I perform a Welch t-test on the one-day ahead return of the highest and lowest quantile to see if there is a significant difference between the two. The same procedure will also be done for five-day-ahead returns.

4.4 Seasonal graphs

In order to heuristically explore **H4**, the commodity futures contracts are sorted based on their expiration month. Loosely speaking, agricultural commodities futures such as on Corn and Soybeans can be sorted into old crop and new crop contracts, where old crop refers to crop that has already been harvested and new crop refers to crop that is yet to be planted or harvested. Futures contracts on soybeans with expiration dates in January, March, May and June are considered old crop contracts, whereas contracts for soybeans with expiration dates in August September and November are considered new crop contracts. Futures contracts for Corn with expiration dates in March, May and June are considered old crop contracts, whereas futures contracts for corn with expiration dates in September and December are considered new crop contracts. Energy contracts can be sorted in winter and summer contracts. The realised volatility of the last 22 days till expiration are plotted together with their IV right before this period in a scatterplot. The data points on the scatterplot from old crop contract have a different colour than the new crop contracts, likewise for summer and winter contracts. These scatterplots are by no means decisive but they could give an idea as to whether IV information encapsulates seasonal information or seasonal information should be its own factor in future research for forecasting realised volatility.

5 Results

This chapter discusses results from this research in the same order as they are discussed in the Methodology section of this thesis. First the results from the HAR models will be discussed, then the results from the VEC model, then the results from the directional analysis and lastly the results from the seasonal graphs.

5.1 HAR models

5.1.1 In-sample results

This section lays out the results of the in-sample comparison between the HAR-IV model, HAR-RV model and IV-only model as described in subsection 4.1.2 of the Methodology. The models are compared to see if the inclusion of HV and IV significantly contributes to the model fit. Table 4 shows the results of this comparison.

The null hypothesis, that IV does not improve the the HAR-RV model, is rejected for Corn, Gold, Nat Gas and Soybeans. This means that the inclusion of IV as a regressor significantly improves the R^2 of the model. This is very much in line with the expectations of **H1**. However, this null hypothesis cannot be rejected for Crude Light. In fact, the inclusion of IV only improves the R^2 of the model only very marginally for Crude light from 0.686 to 0.688. The finding that IV does not significantly contribute as a regressor for Crude Light futures is in contrast with Szakmary et al. (2003). I believe this is likely due to the fact that the sample used in this research includes the year 2020, which was a very volatile year for Crude Light futures, which could have a disproportionate effect on the estimated parameters.

The other null hypothesis, that the inclusion of RV as regressors does not improve forecasts is rejected for Crude Light, Gold and Nat Gas. This means that, at least for those commodities, IV does not contain the full information content of future volatility. This raised the questions why IV does not seem to encapsulate all information from HV, after all option traders can adjust their willingness to buy or sell options based on past volatility. There does not seem to be a commonly agreed upon answer to this question. Interestingly, this null hypothesis cannot be rejected for corn and soy. In the HAR-IV model of Corn and Soybeans, the regression parameter of IV, β_{IV}, is far bigger than the regression parameters for RV, β_d, β_w and β_m, which indicates that the inclusion of the inclusion of RV as a regressor simply isn't that important for Corn and Soybeans. It could be that the IV-only model is the best model the use for Corn and Soybeans, and this is a possible future research direction, although it is not the direction that of this thesis.

A last thing to notice is that the R^2 differs quite a lot between commodities. Nat Gas has the highest R^2 under HAR-IV of 0.706, while Gold has the lowest R^2 of 0.408. This goes to show that the volatility of certain commodity futures is easier to predict than others.

Table 4: **Linear Regression results (in-sample)**

commodity	model	N	β_0	β_d	β_w	β_m	β_{IV}	R^2	F-test
C	HAR-IV	91	0.548	0.123	-0.099	-0.188	0.977	0.658	-
	HAR-RV	91	9.012	0.097	0.044	0.457	-	0.401	<0.001**
	IV-only	91	0.342	-	-	-	0.846	0.640	0.228
CL	HAR-IV	92	-1.432	1.330	-0.190	-0.215	0.267	0.688	-
	HAR-RV	92	1.515	1.309	-0.069	-0.108	-	0.686	0.529
	IV-only	92	-21.453	-	-	-	1.592	0.488	<0.001**
GC	HAR-IV	90	3.324	-0.384	0.650	-0.136	0.576	0.480	-
	HAR-RV	90	5.878	-0.273	0.684	0.192	-	0.415	0.002**
	IV-only	90	3.641	-	-	-	0.684	0.340	<0.001**
NG	HAR-IV	91	6.809	-0.170	0.372	0.335	0.321	0.706	-
	HAR-RV	91	8.560	-0.211	0.521	0.582	-	0.663	0.001**
	IV-only	91	10.589	-	-	-	0.675	0.628	<0.001**
S	HAR-IV	91	2.879	0.052	0.116	-0.210	0.799	0.576	-
	HAR-RV	91	8.113	0.016	0.210	0.341	-	0.385	<0.001**
	IV-only	91	2.427	-	-	-	0.784	0.551	0.176

The first columns reports the commodity. The second column reports the model. The HAR-IV model consists $FRV_m = \beta_0 + \beta_d HV_d + \beta_w HV_w + \beta_m HV_m + \beta_{IV} IV$. The HAR-RV model is the same as the HAR-IV model but does not include IV as regressor. The IV-only model only includes IV as regressor. The third column reports the amount of data points used after scrapping data points due to overlap. The fourth to eight column report the estimated parameters of the regression. The ninth column reports the R^2 of the regression. The last column reports the p-value of the partial F-test, with the null hypothesis that all β coefficients of that are removed from the HAR-IV model are equal to zero. ($^* p < 0.05$,$^{**} p < 0.01$)

5.1.2 Out-of-sample results

This section lays out the results from the out-of-sample comparison of the HAR-IV and HAR-RV model as described in subsection 4.1.3 and subsection 4.1.4 of the Methodology. Table 5 shows the out-of-sample performance results of the HAR-models including the R^2 of the out-of-sample forecast of HAR-RV and HAR-IV together with the Diebold Mariano test statistic and p-value. Graphs of the out-of-sample forecast together with the realised volatility that is being forecasted can be found in Appendix C.

Table 5: **Forecast evaluation of HAR models estimated with OLS Results (out-of-sample)**

		(R^2, HAR-RV)	(R^2, HAR-IV)	DM-stat	p-value
C	daily	0.372	0.424	2.72	0.003**
	week	0.472	0.561	2.95	0.002**
	monthly	0.500	0.700	3.19	<0.001**
CL	daily	0.423	0.479	1.46	0.072
	week	0.451	0.463	0.69	0.245
	monthly	0.432	0.427	-0.38	0.646
GC	daily	0.332	0.362	1.90	0.029*
	week	0.403	0.474	0.933	0.176
	monthly	0.273	0.354	2.258	0.012*
NG	daily	0.478	0.495	1.27	0.102
	week	0.590	0.616	2.37	0.009**
	monthly	0.528	0.574	1.05	0.147
S	daily	0.272	0.308	3.00	0.001**
	week	0.340	0.403	2.29	0.011*
	monthly	0.381	0.495	2.13	0.017*

This table reports the out-of-sample results for the HAR-RV and HAR-IV forecasts, where the model parameters are estimated using the OLS estimator. The first column reports the commodity. The second column reports the time span. The third column with header "(R^2, HAR-RV)" reports the R^2 of the out-of-sample forecasts from the HAR-RV model. The fourth column with header "(R^2, HAR-IV)" reports the R^2 of the out-of-sample forecasts from the HAR-IV model. The fifth column reports the Diebold-Mariano test statistic. The last column reports the p-value of the DM test. ($^{*}p < 0.05, ^{**}p < 0.01$)

The out-of-sample R^2 of the HAR-IV forecasts is higher than the out-of-sample R^2 of the HAR-RV forecasts in fourteen out of the fifteen cases. The only exception is Crude Light with a monthly time span, which is possibly due to the extreme volatility spike in Crude Light futures. The HAR-IV model outperforms the HAR-RV model for both corn and soybeans in all three time spans. This is very solid evidence that IV has predictive information beyond what is known in HV. Results for other commodities are a little less clear. For Gold, HAR-IV significantly outperforms HAR-RV on the monthly level, but not significantly on the daily and weekly level. For Nat Gas, HAR-IV significantly outperforms HAR-RV on the weekly level, but not on the daily and monthly level. This is a bit weird and might be a signal that the results from this research are on the conservative side. Possible reasons for this, such as a small sample size, are discussed in the Discussion in Chapter 7.

Table 6: **Forecast evaluation of HAR models estimated with Tukey-Biweight estimator (out-of-sample)**

		(R^2, HAR-RV)	(R^2, HAR-IV)	DM-stat	p-value
C	daily	0.346	0.399	3.61	<0.001**
	weekly	0.454	0.550	3.26	<0.001**
	monthly	0.468	0.652	3.05	0.001**
CL	daily	0.442†	0.498†	1.20	0.023*
	weekly	0.458†	0.494†	1.86	0.032*
	monthly	0.412	0.410	-0.07	0.529
GC	daily	0.320	0.358	1.95	0.026*
	weekly	0.316	0.421	1.61	0.054
	monthly	0.222	0.318	2.32	0.010*
NG	daily	0.455	0.475	1.73	0.042*
	weekly	0.571	0.596	1.60	0.055
	monthly	0.515	0.566	1.25	0.105
S	daily	0.247	0.281	2.68	0.004**
	weekly	0.331	0.390	2.65	0.004**
	monthly	0.350	0.460	2.06	0.020*

This table reports the out-of-sample results for the HAR-RV and HAR-IV forecasts, where the model parameters are estimated using the Tukey-Biweight estimator. The first column reports the commodity. The second column reports the time span. The third column with header "(R^2, HAR-RV)" reports the R^2 of the out-of-sample forecasts from the HAR-RV model. The fourth column with header "(R^2, HAR-IV)" reports the R^2 of the out-of-sample forecasts from the HAR-IV model. The fifth column reports the Diebold-Mariano (DM) test statistic from the Diebold-Mariano test. The last column reports the p-value of the DM test. A "†" denotes that the out-of-sample R^2 value is higher when using the Tukey-Biweight estimator, than using OLS estimator for this commodity, model and time span. ($^* p < 0.05, ^{**} p < 0.01$)

Table 6 shows the out-of-sample results of the HAR-models estimated with a robust regression. In the majority of cases, parameter estimation using robust linear regression yields a less good out-of-sample results in term of R^2. This is unsurprising, since OLS minimizes the R^2

in-sample, whereas Tukey-Biweight minimizes another loss functions in-sample. On the other hand Clements and Preve (2021) did find that using the Tukey-Biweight estimator yielded better out-of-sample results. The only thing that I could think of that could explain this difference in result is that Clements and Preve (2021) uses the HAR model to forecast a stock index volatility, whereas I forecast commodity futures volatility. This would suggest that the choice of how to estimate model parameters of a volatility model should also depend on the kind of asset that is being forecasted.

Clements and Preve (2021) uses the HAR model to forecast a stock index, whereas I do this on commodities, which might explain the difference in results? The only times when using the Tukey-Biweight estimator improves forecasts is with Crude Light on the daily and weekly level. Crude Light has had a extreme values of IV and FRV. These data points would have had a larger impact on the OLS estimator than on the Tukey-Biweight estimator. This could explain why the robust regression yielded better out-of-sample results. Since in these two cases the robust regression performs better out-of-sample in terms of mean-squared error than with OLS under the same commodity, time span and model, the conclusions are drawn from the forecasts that arose from the robust regression. This means that IV is considered to contain information on the future realised volatility of the underlying beyond what is known from HV under Crude Light as underlying and forecasting time spans daily and weekly.

In total, this mean that out of a total of fifteen combinations of commodity and time spans, the HAR-IV model performed better in fourteen of them. Out of these fourteen cases, eleven of them were significant at the 5% significance level. In general, the HAR-IV model performs better than the HAR-IV model. Thus, **H1** is accepted in general terms, meaning that in general IV contains information on the future realised volatility of the underlying commodity futures contract beyond what is known from HV.

5.2 VEC model

This section lays out the result of the tests on the VEC model as described in section 4.2. An Engle-Granger co-integration test is performed on IV and HV of the last trading day, week or month. A Granger Causality test is performed on HV and IV in both directions. Additionally an instanteneous causality test is performed. Table 7 shows the results of these tests.

The null hypothesis that IV and FRV are not co-integrated can be rejected at the 1% level for all commodities and time-spans, except for weirdly enough (Nat Gas, monthly). In the 14 out of the 15 cases I continue to apply a VEC model as specified in the Methodology Section.

Table 7: **Results of tests on co-integration, Granger causality and instantaneous causality**

		IV~FRV	IV~HV	IV GC RV	HV GC IV	IC
commodity	timespan					
C	day	<0.001**	<0.001**	<0.001**	0.181	<0.001**
	week	<0.001**	<0.001**	<0.001**	0.286	<0.001**
	month	<0.001**	<0.001**	<0.001**	0.006**	<0.001**
CL	day	<0.001**	<0.001**	<0.001**	0.000**	<0.001**
	week	<0.001**	0.013**	<0.001**	0.000**	<0.001**
	month	<0.001**	<0.001**	<0.001**	0.126	<0.001**
GC	day	<0.001**	<0.001**	<0.001**	0.061	<0.001**
	week	<0.001**	<0.001**	<0.001**	0.115	<0.001**
	month	<0.001**	<0.001**	<0.001**	0.701	<0.001**
NG	day	<0.001**	<0.001**	<0.001**	0.529	<0.001**
	week	<0.001**	<0.001**	<0.001**	0.228	<0.001**
	month	0.600	0.679	-	-	-
S	day	<0.001**	<0.001**	<0.001**	0.412	<0.001**
	week	<0.001**	<0.001**	<0.001**	0.508	<0.001**
	month	<0.001**	<0.001**	<0.001**	0.915	<0.001**

This table report the p-values of the statistical test that were performed on the VEC model. The first column represents the p-values of the co-integration test between IV and RV, the second column represents the p-values of the co-integration test between IV and HV, the third column represents the Granger causality test of IV Granger causing RV, the fourth column represents the HV Granger causing IV. The fifth column represent the p-value of the instantaneous causality test. (${}^{*}p < 0.05, {}^{**}p < 0.01$)

The Granger causality test for IV Granger causing RV is significant for all cases. This means we can continue using a VEC model. The Granger causality test from IV to RV yields that IV Granger causes RV in all cases. The Granger causality test from RV to IV yields that does not Granger cause in 12 of the 15 cases. The instantaneous causality test shows there is instantaneous causality between RV and IV. This is intuitive, since if we knew the realised volatility we would surely be able to predict IV better. The instantaneous causality explains the high correlation between IV and RV, but it does not help in prediction, since anybody who wants to predict IV will have the current value of IV at its disposal. There are three exceptions where the Granger causality test does indicate that HV Granger causes IV. We analyse these cases further by testing if the out-of-sample forecast of the VEC model outperform the forecast of a random walk. The forecast of a random walk is simply the previous value of the variable that is being forecasted. Results van be found in Table 8 below.

Table 8: **out-of-sample results VECM**

Commodity	time span	predicting	(R^2,VECM)	(R^2, RW)	DM-stat	p-value
C	monthly	IV	0.551	0.536	-0.64	0.524
		FRV	0.579	0.257	-3.02	0.004**
CL	daily	IV	0.938	0.939	0.28	0.781
		FRV	0.523	0.323	-3.64	< 0.001**
	weekly	IV	0.790	0.796	0.47	0.641
		FRV	0.452	0.361	-2.34	0.020*

This table reports the evaluations of the out-of-sample forecasts of a VEC model and a random walk estimator. Only those combination of commodity and time span, of which the p-value of the granger causality test from HV to IV is significant, are considered here. The first column reports the commodity. The second column report the time span. The third column reports which value is predicted in that row. The VECM model predicts IV and RV simultaneously, so two rows of IV and RV should be considered as the same VEC model prediction. The fourth column reports the R^2 of the out-of-sample forecasts of the VEC model, with rolling window of 50% of the sample. The fifth column reports the R^2 of the random walk. The sixth column reports the Diebold Mariano (DM) test statistic. The last column reports the double-sided p-value of the Diebold-Mariano test. ($^*p < 0.05,^{**} p < 0.01$) signals significance at the 1

Table 8 shows that in the three cases where HV was supposed to Granger cause IV, the VECM forecast for IV actually underperforms a random walk. The first combination (Corn, monthly) does perform better than a random walk for IV, but not significantly so. The VECM forecast for RV do outperform the random walk significantly. Thereby this research finds no evidence for **H2**. It could be that this is due to ill-fitted methodology or models. Even there was a model that could supply evidence for this hypothesis, it would still be hard to find this model in practice, since the seemingly most straightforward method yields no evidence for it.

5.3 Directional Analysis

This section lays out the results of the directional analysis as described in section 4.3. The trading days of a commodity are sorted into five quantiles based on certain IV characteristics. The definition of these tested characteristics can be found in section 4.3. Table 15 in Appendix D reports the one-day-ahead results and Table 16 reports the five-day-ahead results. This means that during the sample period the continuous contract had a higher subsequent day return when convexity was high than days when convexity was low. As we can see for the one-day-ahead return only one combination (Crude Light, convexity) of the 30 combinations testing yields a statistically significant result at the 5% level on the Welch t-test. On trading days where convexity of a crude light futures contract was high, the subsequent day return was higher than the subsequent day return of trading days where the convexity was low. The five-day-ahead result have 5 combinations that with a significant p-value of the Welch t-test. These combinations are (Crude Light, convexity), (Gold, cvol), (Gold, skew1), (Gold, skew2), (Gold, Δskew1). On trading days when the skew of a gold futures contract is high, we expect higher returns to follow in the next five days than when skew is low. However on trading days when the delta skew is high we expect the lower returns in the next five days than when the delta skew is low. This findings seems paradoxical and not in line with the hypothesis of skew being a reflection of informed trading. These findings should be taken with a bit of scrutiny. Only 6 of the total 60 combinations yield a significant Welch t-test at the 5% level. This result would not be unusual and could very well be the result of pure chance. Under the assumption that these 60 trials were independent the chance of having 6 or more significant results at the 5% level is 7.82%, which is already more than the usual 5% level that is usually required. This 7.82% is also a (very) conservative estimate, since the trials are not independent, meaning that the distribution has fatter tails. The results from this subsection are thereby considered inconclusive.

Table 9: **Directional Analysis**

commodity	measure	n	μ highest	μ lowest	t-stat	P(t)
CL	conv	one	0.003	-0.003	-2.232	0.026*
CL	conv	five	0.014	-0.011	-4.886	0.000**
GC	cvol	five	0.005	0.001	-2.004	0.046*
GC	skew1	five	0.008	0.000	-3.970	0.000**
GC	skew2	five	0.005	0.000	-2.710	0.007**
GC	Δ_skew1	five	0.000	0.004	2.028	0.043*

This table reports part of results from Table 15 and Table 16. The first column denotes the commodity. The second column denotes the tested derivative measure. Trading days of the commodity are sorted in five quantiles based on the level of the derivative. The third column denotes the amount of days after which the returns are compared. The fourth column reports the mean of the returns of the highest quantiles after n trading day(s) (in %). The fifth column reports the mean of the returns of the lowest quantile after n trading day(s) (in %). The sixth column report the t-statistic of the Welch t-test that is performed on the returns of the highest and lowest quantile. The last column report the p-value of the t-test. Only cases where $P(t) < 0.05$ are reported in this table. ($^{*}p < 0.05,^{**}p < 0.01$)

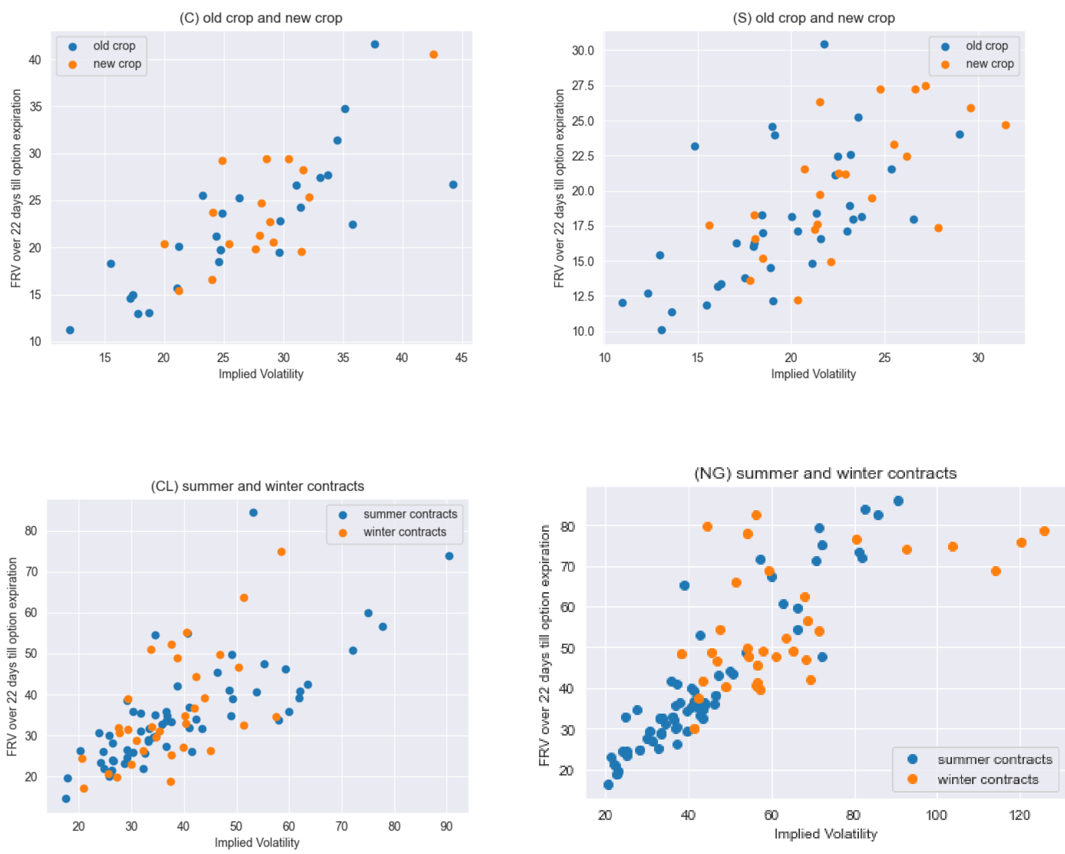
5.4 Seasonal graphs

This section discusses the seasonal graphs that were made using the procedure laid out in section 5.4. The results can be found in 4. For completeness, these graphs are also sorted based on expiration month alone in Appendix D.

Of the four commodities, Nat Gas has the clearest seasonal pattern. Winter contracts, which end in November, December, January, and February, have a higher average IV and a higher FRV than summer contracts. This also shows the practicality of IV as a predictor for future realised volatility, since by its forwards-looking nature, it appears to already takes into account this seasonal structure. Additionally it looks like the variance in volatility is higher in winter contracts for Nat Gas. This could be subject of further research. Soybeans also appears to have a similar seasonal pattern, where new crop contracts have an average higher FRV and IV than old crop contracts. Corn, and Crude Light do not seem to have any seasonal pattern in terms of realised volatility or IV. Judging by these plots, there seems to be no good reason to take seasonality as a factor in forecasting realised volatility since, there does not seem to be any seasonal patterns for some commodities such as Corn and Crude Light, and if there is a seasonal pattern, this can be largely accounted for by taking IV as a factor in forecasting.

Another thing to notice is that in all four of these graphs the relationship between IV and FRV looks relatively linear. This is an indication that non-linear models don't need to be used here.

Figure 4: IV and FRV sorted by season



These graphs show the future realized volatility over 22 days (FRV_m) till expiration compared against the implied volatility (IV). The dots are sorted on new crop and old crop for the agricultural commodities corn and soy and summer or winter for the energy commodities Crude Light and Nat Gas

6 Conclusion

This research set out to investigate whether information from option prices can be said to be leading commodity futures prices. This was done on the basis of four hypotheses, namely:

H1: IV contains information on the future realised volatility of the underlying commodity futures contract beyond what is known from HV.

H2: HV contains information on future IV beyond what is known from current IV.

H3: Certain derivative measures of IV contain information on the future direction of the underlying commodity future.

H4: IV accounts for seasonal differences in FRV. (This hypothesis is only tested for heuristically.)

The most important result of this research found strong evidence in favour of **H1**. The HAR-IV model, which takes into account IV as additional facter, performed better out-of-sample in terms of mean-squared error, than the HAR-RV model in fourteen out of fifteen cases. In eleven out of those, the HAR-IV model also performed significantly better than the HAR-RV model. This is clear evidence that the inclusion of IV as input in a model improves the out-of-sample performance and thereby that IV contains information on the future realiseed volatility of the underlying commodity futures contract beyond what is known from HV.

No evidence was found for **H2**. Even though in 3 out of 15 cases there was Granger causality according to the Granger-causality test, the VEC model in these cases could not outperform a random walk for IV. It was concluded that the model was ill-fitted and the result of the Granger-causality test was not representative. This leads me to suggest that a potential forecaster of IV has no or little use in using historical volatility of the underlying as a variable in their model. There was no evidence found for **H3** either. Since the few significant cases could be explained as a result of pure chance. The last hypothesis, **H4**, was only tested for heuristically. Judging by the graphs it produced, there seemed to be no seasonal difference in FRV or IV for Corn and Crude Light. There did appear to be seasonal differences in the FRV of Soybeans and Nat Gas. These seasonal differences were largely accounted for by IV. This leads me to believe a potential forecaster has more use for IV than seasonality as additional variable in a forecasting model.

It should be noted that a lack of evidence for **H2** and **H3** is necessarily evidence against **H2** and **H3**. It could be that the choice of model or choice of tested derivative measure was inadequate. This does mean it is difficult to find these phenomena in practice.

7 Discussion

The results from the HAR models is largely affirmative on the idea that IV holds information on the FRV beyond the information of HV. In this sense it extends the findings of Liang et al. (2020)

to commodity futures markets. However, the results are not as overwhelmingly affirmative on this question as with Liang et al. (2020). This difference could be explained by the possibility that commodity futures are harder to predict with IV than stock indexes. The results of this research were stronger for Corn and Soybeans than they were for Crude Light, Gold and Nat Gas, so it could well be that for some products the IV is simply less informative than for other products. It could also be that the sample in Liang et al. (2020) was larger, which would lead to more statistically significant results. Nevertheless, the results still feel rather conservative given the in-sample results of section 5.1.1 and the fact that the HAR-IV model outperformed the HAR-RV model in 14 out of 15 cases. In the ideal scenario, a joint test is performed on all forecasting time windows and commodities combined. The relatively new method developed in Quaedvlieg (2021) could be a way to perform a joint test on time windows. Perhaps this method combined with a method for a joint test across products could be a way to, for once and for all, settle the debate over the informativeness of IV on FRV, although this would also depend on the results from this test.

One could wonder if the models used in this research were appropriate or perhaps too simple. There are of course a great many volatility models. As discussed in the methodology, the GARCH model was not used because it was harder to adapt the model to the term stucture of commodity futures. There are of course also many variations and extensions of the HAR model, such as the use of time-varying windows in which the historical variance of the regressors is calculated, such as used in Tian et al. (2017). Degiannakis et al. (2022), however, finds that the standard HAR-RV performs best for agricultural commodities in the US compared to extended HAR-RV models. They conclude that the future research on the volatility of agricultural should not focus on extending the HAR-RV model to include things like jump component and other additions to the HAR model, but rather focus on in things like including exogenous predictors. By looking at the advantage of including IV in the standard HAR-RV model, this research is very much in line with this recommendation.

References

Bakshi, G. and Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *The Review of Financial Studies*, 16(2):527–566.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1-2):167–179.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Bohmann, M. J. M. (2020). *Price discovery and information asymmetry in equity and commodity futures options markets*. PhD thesis.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, pages 621–651.

Canina, L. and Figlewski, S. (1993). The informational content of implied volatility. *The Review of Financial Studies*, 6(3):659–681.

Carr, P. and Wu, L. (2009). Variance risk premiums. *The Review of Financial Studies*, 22(3):1311–1341.

Christoffersen, P., Jacobs, K., and Chang, B. Y. (2013). Forecasting with option-implied information. *Handbook of Economic Forecasting*, 2:581–656.

Clark, T. and McCracken, M. (2013). Advances in forecast evaluation. *Handbook of Economic Forecasting*, 2:1107–1201.

Clements, A. and Preve, D. P. (2021). A practical guide to harnessing the har volatility model. *Journal of Banking & Finance*, 133:106285.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Cremers, M. and Weinbaum, D. (2010). Deviations from put-call parity and stock return predictability. *Journal of Financial and Quantitative Analysis*, 45(2):335–367.

Degiannakis, S., Filis, G., Klein, T., and Walther, T. (2022). Forecasting realized volatility of agricultural commodities. *International Journal of Forecasting*, 38(1):74–96.

Doran, J. S. and Ronn, E. I. (2008). Computing the market price of volatility risk in the energy commodity markets. *Journal of Banking & Finance*, 32(12):2541–2552.

Fama, E. F. and French, K. R. (2016). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. In *The World Scientific Handbook of Futures Markets*, pages 79–102. World Scientific.

Galán-Gutiérrez, J. A. and Martín-García, R. (2022). Fundamentals vs. financialization during extreme events: From backwardation to contango, a copper market analysis during the covid-19 pandemic. *Mathematics*, 10(4):559.

Greb, F. and Prakash, A. (2017). Assessing volatility patterns in food crops. *FAO Commodity and Trade Policy Research Working Papers*, (53).

Grith, M., Härdle, W. K., and Schienle, M. (2012). Nonparametric estimation of risk-neutral densities. In *Handbook of Computational Finance*, pages 277–305. Springer.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial studies*, 6(2):327–343.

Jiang, G. J. and Tian, Y. S. (2005). The model-free implied volatility and its information content. *The Review of Financial Studies*, 18(4):1305–1342.

Krams, M. (2014). Giving it away: Option trades revealing information in commodity futures markets.

Liang, C., Wei, Y., and Zhang, Y. (2020). Is implied volatility more informative for forecasting realized volatility: An international perspective. *Journal of Forecasting*, 39(8):1253–1276.

Ofek, E., Richardson, M., and Whitelaw, R. F. (2004). Limited arbitrage and short sales restrictions: Evidence from the options markets. *Journal of Financial Economics*, 74(2):305–342.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301.

Poon, S.-H. and Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539.

Quaedvlieg, R. (2021). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, 39(1):40–53.

Szakmary, A., Ors, E., Kim, J. K., and Davidson III, W. N. (2003). The predictive power of implied volatility: Evidence from 35 futures markets. *Journal of Banking & Finance*, 27(11):2151–2175.

Tian, F., Yang, K., and Chen, L. (2017). Realized volatility forecasting of agricultural commodity futures using the har model with time-varying sparsity. *International Journal of Forecasting*, 33(1):132–152.

Triantafyllou, A., Dotsis, G., and Sarris, A. H. (2015). Volatility forecasting and time-varying variance risk premiums in grains commodity markets. *Journal of Agricultural Economics*, 66(2):329–357.

Xing, Y., Zhang, X., and Zhao, R. (2010). What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial and Quantitative Analysis*, 45(3):641–662.

Yang, D. and Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, 73(3):477–492.

A The CVOL method

The CVOL method is a method to calculate the model-free options-implied (risk-neutral) volatility. The idea behind the CVOL method comes from model-free methods to create a replicating portfolio for a variance swap. Below you will find a short summary of the CVOL method. For the exact procedure to calculate the CVOL and its derivative measures, see the CVOL whitepaper.[6] To calculate the CVOL we take the prices of call and put options as in figure 5. We then calculate the area under the surface, using the available option prices. The CVOL method does not further interpolate or extrapolate option prices to estimate the volatility under the surface. This could lead to inaccuracies, see Christoffersen et al. (2013).

Point F is defined as the point on the x-axis, where the lines of the call prices and put prices cross, see figure 5. Because of the put-call parity, the spot price S of the underlying future is the same as the the strike price at which the call and put option have the same price. In other words $F = S$. This means that the area under the surface is determined solely by the OTM options. Let A_{left} be the area under the surface to the left of F and let A_{right} be the area under the surface to the left of F. Let $DnVar$ be equal to $2 * A_{left}$ and let $UpVar$ be equal to $2 * A_{right}$. Notice that $DnVar$ is determined by OTM put options and gives a measure of downward volatility and $UpVar$ is determined by OTM call options and gives a measure of upward volatility. The volatility skewness can be represented by two measures namely: $Skew1 = UpVar - DnVar$ or $Skew2 = UpVar/DnVar$.

The CVOL measures as calculated by CME combines different contracts to come up with a estimate for 30- or 60-day ahead forecast. The only change made in this thesis is that the CVOL measures are reported that are specific to one contract. The rest of the CVOL methodology is similar to the way that the CME calculates it.

[6] CVOL whitepaper: https://www.cmegroup.com/market-data/cme-group-benchmark-administration/files/cvol-methodology.pdf

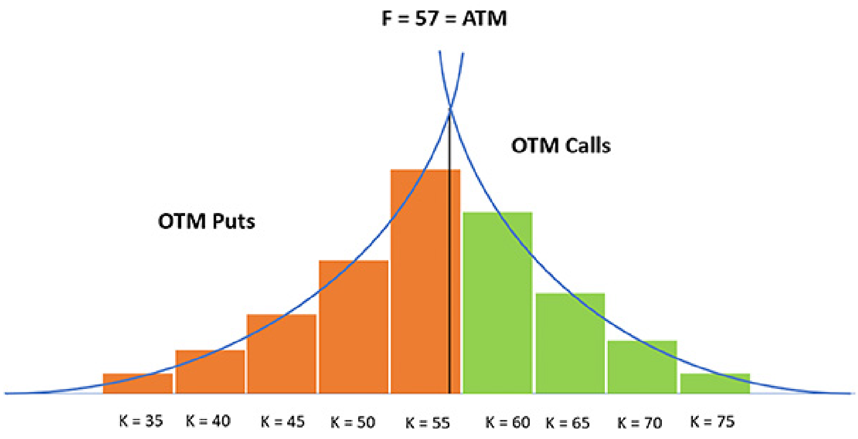


Figure 5: The two lines represent the option put and call prices. The area under the surface is calculated as the CVOL estimate.

B Data

Table 10: **Corn (C) correlation matrix**

	IV	HV_d	HV_w	HV_m	FRV_d	FRV_w	FRV_m
IV	1.000	0.617	0.755	0.820	0.615	0.726	0.801
HV_d	0.617	1.000	0.856	0.645	0.503	0.490	0.511
HV_w	0.755	0.856	1.000	0.795	0.563	0.593	0.614
HV_m	0.820	0.645	0.795	1.000	0.550	0.624	0.620
FRV_d	0.615	0.503	0.563	0.550	1.000	0.524	0.523
FRV_w	0.726	0.490	0.593	0.624	0.524	1.000	0.754
FRV_m	0.801	0.511	0.614	0.620	0.523	0.754	1.000

This table reports the correlation coefficients of IV and RV of Corn.

Table 11: **Crude Light (CL) correlation matrix**

	IV	HV_d	HV_w	HV_m	FRV_d	FRV_w	FRV_m
IV	1.000	0.777	0.872	0.933	0.716	0.631	0.666
HV_d	0.777	1.000	0.889	0.736	0.641	0.508	0.531
HV_w	0.872	0.889	1.000	0.845	0.662	0.580	0.618
HV_m	0.933	0.736	0.845	1.000	0.645	0.615	0.664
FRV_d	0.716	0.641	0.662	0.645	1.000	0.525	0.526
FRV_w	0.631	0.508	0.580	0.615	0.525	1.000	0.640
FRV_m	0.666	0.531	0.618	0.664	0.526	0.640	1.000

This table reports the correlation coefficients of IV and RV of Crude Light.

Table 12: **Gold (GC) correlation matrix**

	IV	HV_d	HV_w	HV_m	FRV_d	FRV_w	FRV_m
IV	1.000	0.602	0.748	0.854	0.558	0.651	0.608
HV_d	0.602	1.000	0.838	0.626	0.445	0.457	0.427
HV_w	0.748	0.838	1.000	0.785	0.517	0.561	0.520
HV_m	0.854	0.626	0.785	1.000	0.510	0.563	0.491
FRV_d	0.558	0.445	0.517	0.510	1.000	0.487	0.449
FRV_w	0.651	0.457	0.561	0.563	0.487	1.000	0.714
FRV_m	0.608	0.427	0.520	0.491	0.449	0.714	1.000

This table reports the correlation coefficients of IV and RV of Gold.

Table 13: **Natural Gas (NG) correlation matrix**

	IV	HV_d	HV_w	HV_m	FRV_d	FRV_w	FRV_m
IV	1.000	0.668	0.791	0.831	0.656	0.741	0.792
HV_d	0.668	1.000	0.886	0.740	0.578	0.624	0.626
HV_w	0.791	0.886	1.000	0.874	0.663	0.714	0.741
HV_m	0.831	0.740	0.874	1.000	0.668	0.760	0.790
FRV_d	0.656	0.578	0.663	0.668	1.000	0.643	0.635
FRV_w	0.741	0.624	0.714	0.760	0.643	1.000	0.825
FRV_m	0.792	0.626	0.741	0.790	0.635	0.825	1.000

This table reports the correlation coefficients of IV and RV of Nat Gas

Table 14: **Soybean (S) correlation matrix**

	IV	HV_d	HV_w	HV_m	FRV_d	FRV_w	FRV_m
IV	1.000	0.570	0.735	0.816	0.544	0.667	0.713
HV_d	0.570	1.000	0.828	0.607	0.391	0.433	0.453
HV_w	0.735	0.828	1.000	0.777	0.476	0.551	0.571
HV_m	0.816	0.607	0.777	1.000	0.488	0.578	0.568
FRV_d	0.544	0.391	0.476	0.488	1.000	0.454	0.467
FRV_w	0.667	0.433	0.551	0.578	0.454	1.000	0.732
FRV_m	0.713	0.453	0.571	0.568	0.467	0.732	1.000

This table reports the correlation coefficients of IV and RV of Soybeans

Figure 6: These graphs show the implied volatility together with the historical realised volatility of the last trading month (HV_m).

C HAR out-of-sample results

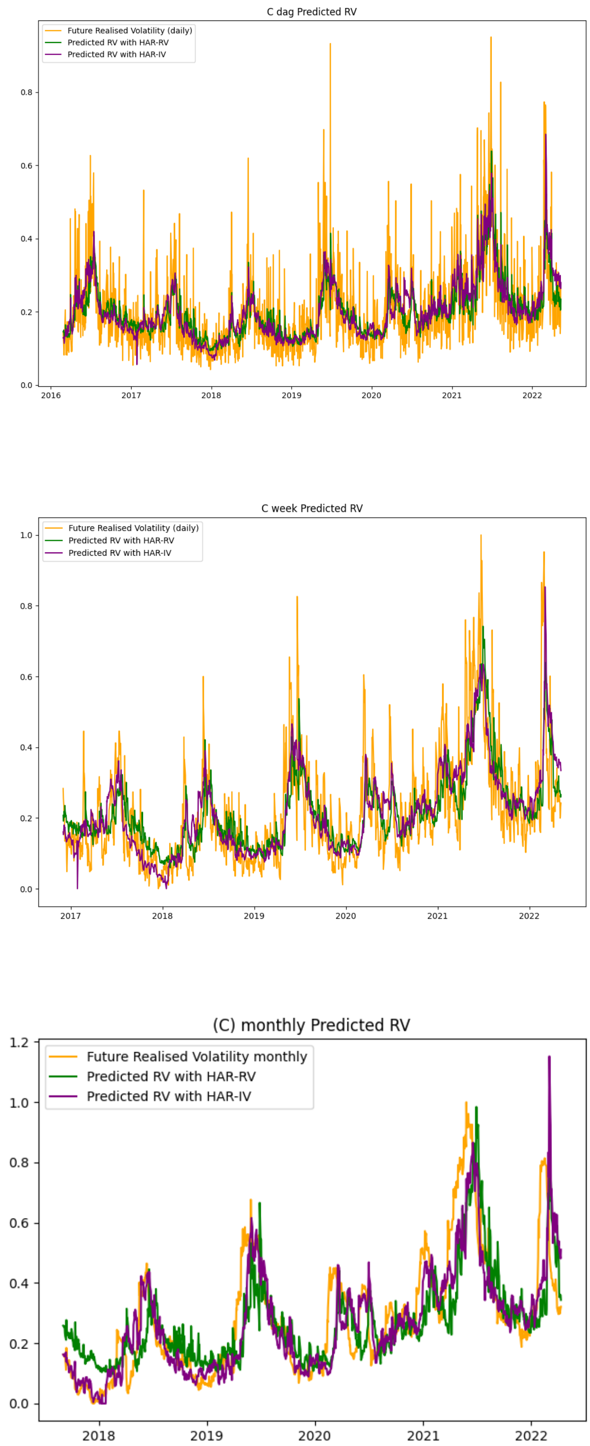
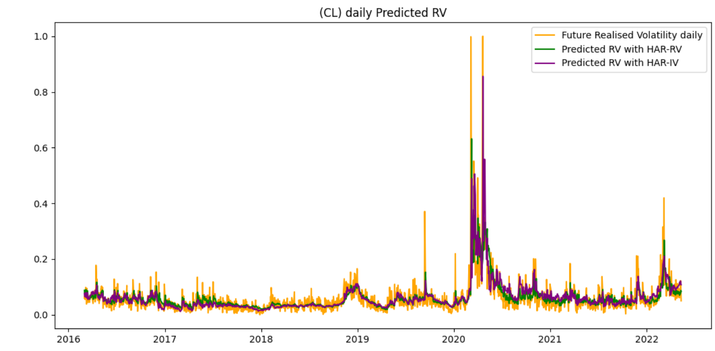
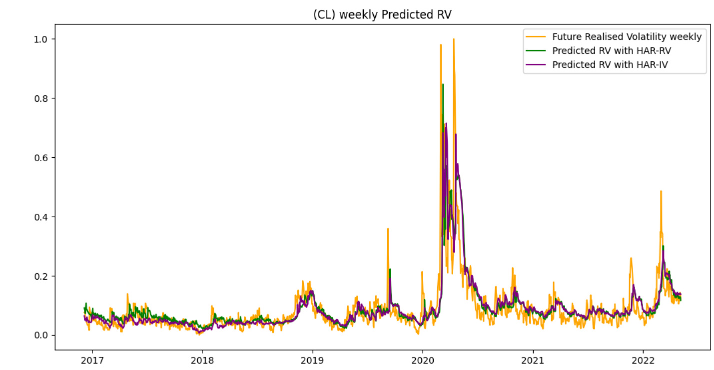


Figure 7: Out-of-sample forecast of HAR-RV and HAR-IV Corn





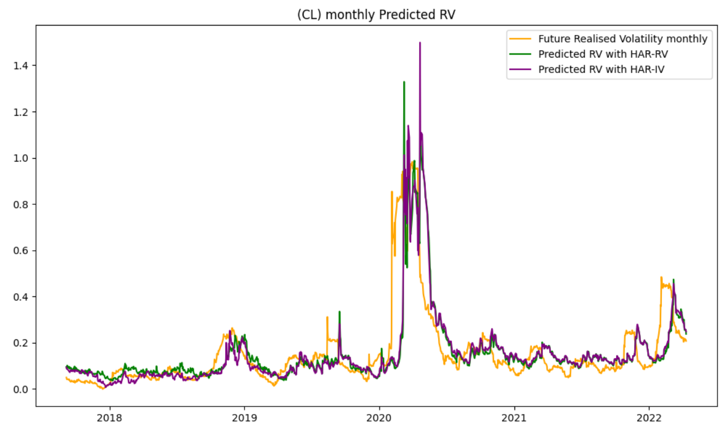
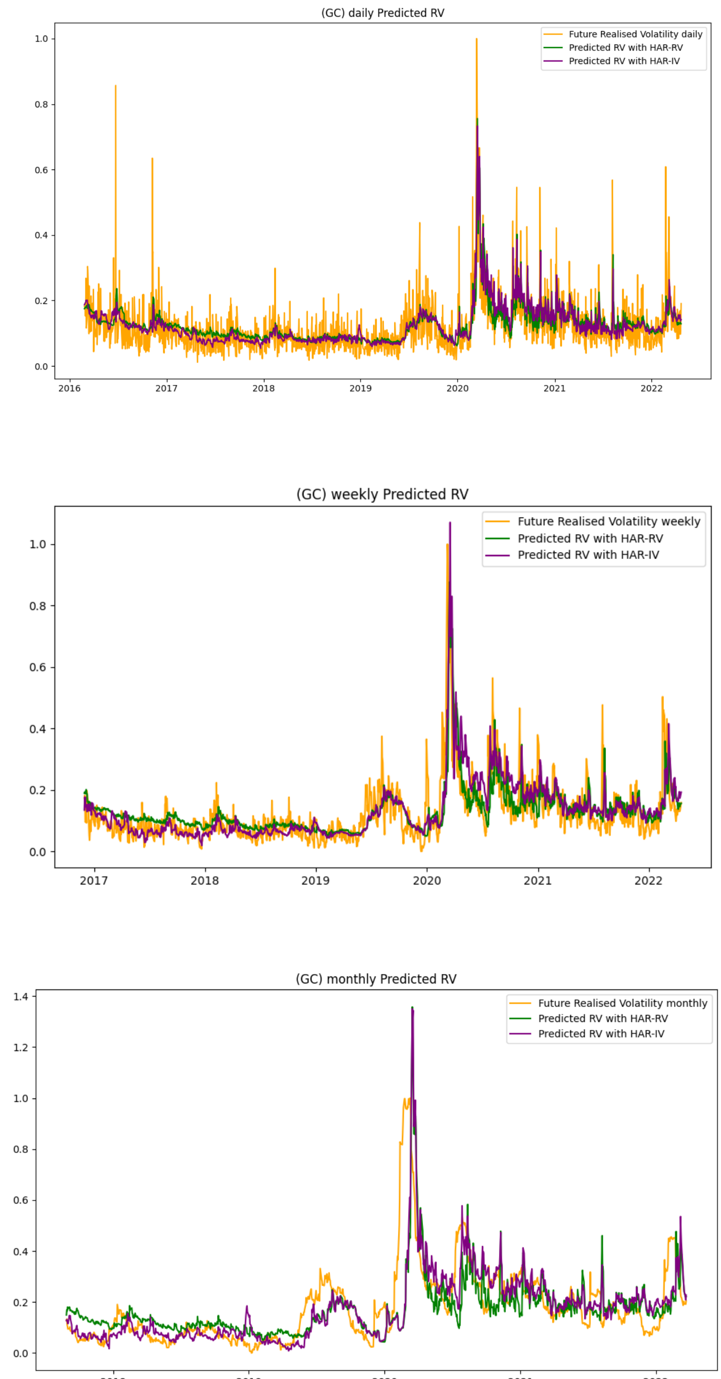


Figure 8: Out-of-sample forecast of HAR-RV and HAR-IV Crude Light



(a) As we can see both HAR-RV and HAR-IV did not see the volatility spike in 2020 coming.

Figure 9: Out-of-sample forecast of HAR-RV and HAR-IV Gold

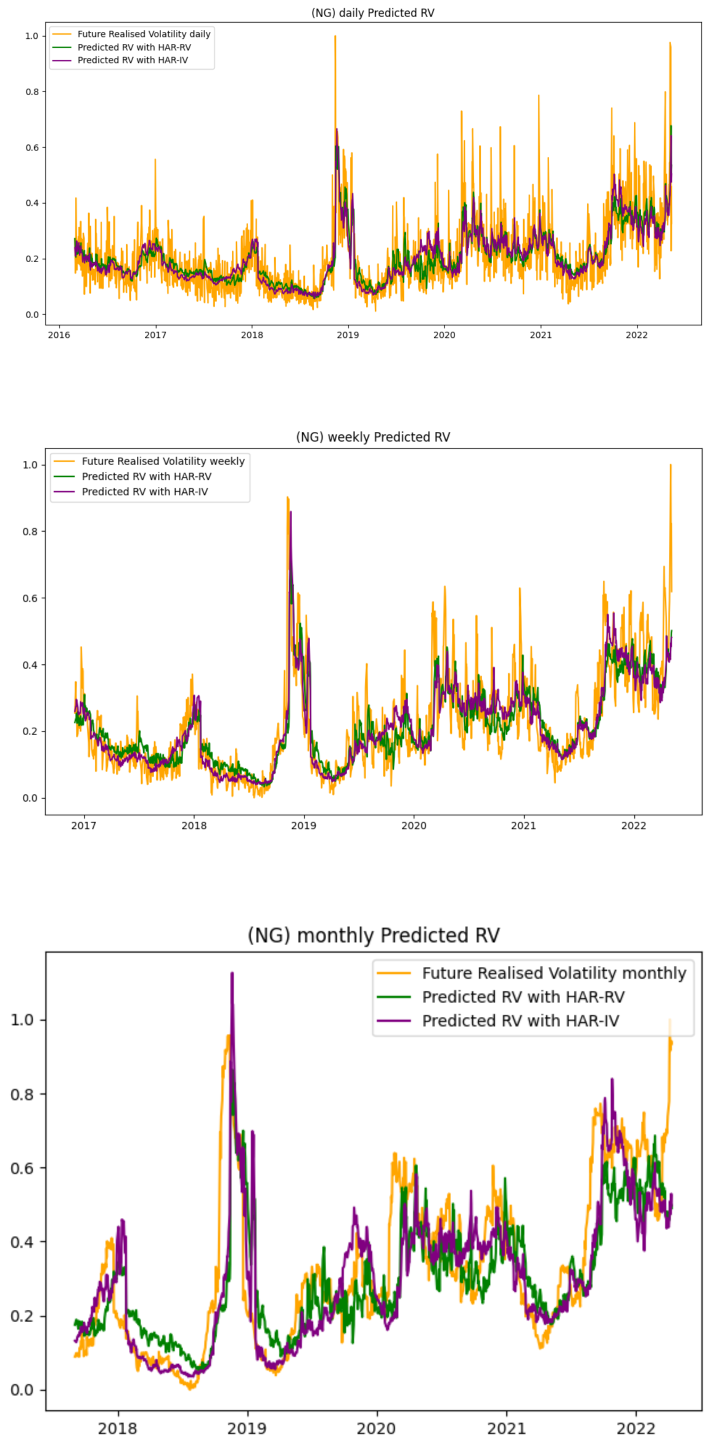


Figure 10: Out-of-sample forecast of HAR-RV and HAR-IV Nat Gas

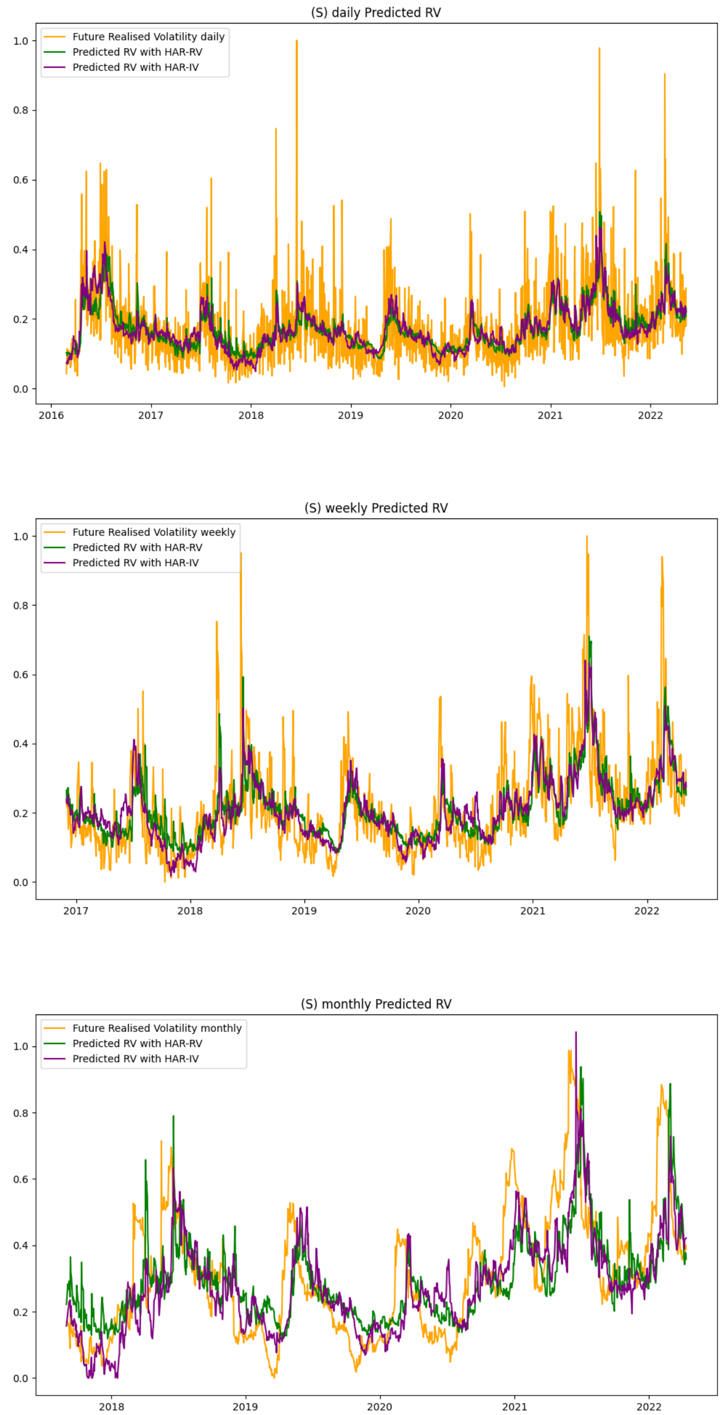


Figure 11: Out-of-sample forecast of HAR-RV and HAR-IV Soybeans

D Directional Analysis

Table 15: **Directional analysis of one-day-ahead returns**

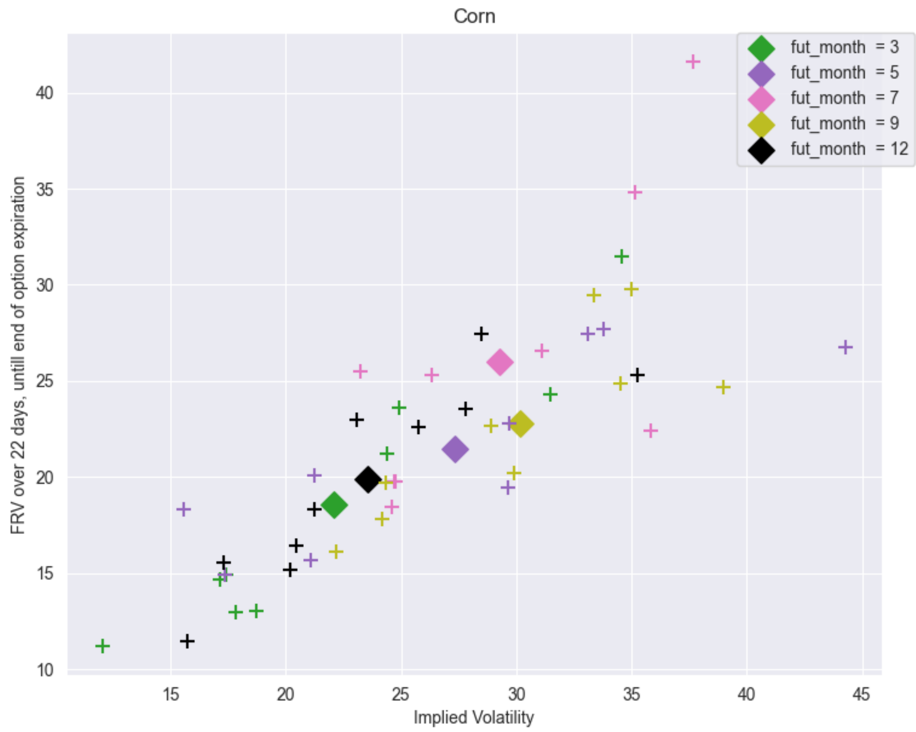
commodity	measure	μ highest	μ lowest	t-stat	P(t)
C	IV	0.001	-0.000	-0.824	0.410
C	skew1	-0.000	-0.000	0.002	0.998
C	skew2	-0.001	-0.000	0.344	0.731
C	Δskew1	-0.001	-0.001	0.052	0.959
C	abn_skew1	-0.000	-0.001	-0.746	0.456
C	conv	0.000	0.000	0.002	0.998
CL	cvol	0.001	0.000	-0.301	0.764
CL	skew1	0.002	-0.000	-0.751	0.453
CL	skew2	0.001	0.000	-0.199	0.842
CL	Δskew1	0.001	-0.001	-0.509	0.611
CL	abn_skew1	-0.002	-0.000	0.639	0.523
CL	conv	0.003	-0.003	-2.232	0.026*
GC	cvol	0.001	0.000	-0.984	0.326
GC	skew1	0.001	0.000	-1.357	0.175
GC	skew2	0.001	0.000	-0.679	0.497
GC	Δskew1	-0.000	0.000	0.444	0.657
GC	abn_skew1	-0.001	0.001	0.994	0.321
GC	conv	0.000	-0.000	-0.558	0.577
NG	cvol	-0.002	-0.001	0.503	0.615
NG	skew1	0.001	-0.000	-0.605	0.545
NG	skew2	0.001	-0.001	-0.710	0.478
NG	Δskew1	0.001	-0.001	-0.563	0.574
NG	abn_skew1	-0.003	-0.002	0.503	0.615
NG	conv	0.001	-0.002	-1.539	0.124
S	cvol	0.000	-0.000	-0.506	0.613
S	skew1	0.001	-0.001	-1.775	0.076
S	skew2	0.001	-0.000	-1.037	0.300
S	Δskew1	-0.000	-0.000	0.366	0.714
S	abn_skew1	0.001	0.001	0.246	0.806
S	conv	0.001	0.001	0.071	0.944

The first column denotes the commodity. The second column denotes the tested derivative measure. Trading days of the commodity are sorted in five quantiles based on the level of the derivative. The third column reports the mean of the returns of the highest quantiles after one trading day (in %). The fourth column reports the mean of the returns of the lowest quantile after one trading day (in %). The fifth column report the t-statistic of the Welch t-test that is performed on the returns of the highest and lowest quantile. The last column report the p-value of the t-test.

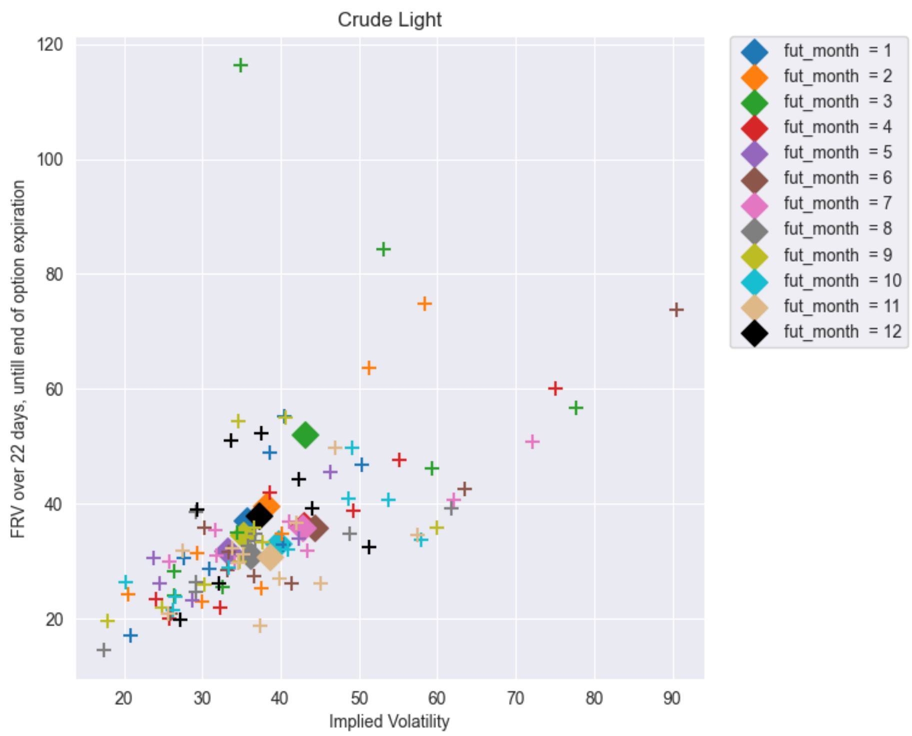
Table 16: **Directional Analysis of five-day-ahead returns**

commodity	measure	μ highest	μ lowest	t-stat	P(t)
C	IV	0.000	-0.001	-0.528	0.598
C	skew1	-0.001	-0.000	0.300	0.764
C	skew2	-0.004	-0.000	1.452	0.147
C	Δskew1	-0.004	-0.001	1.396	0.163
C	abn_skew1	-0.003	-0.000	1.167	0.244
C	conv	0.001	-0.001	-1.082	0.280
CL	cvol	0.008	0.001	-1.255	0.210
CL	skew1	0.002	0.007	0.812	0.417
CL	skew2	0.000	0.007	1.459	0.145
CL	Δskew1	-0.002	0.006	1.437	0.151
CL	abn_skew1	-0.002	-0.008	-1.051	0.294
CL	conv	0.014	-0.011	-4.886	0.000**
GC	cvol	0.005	0.001	-2.004	0.046*
GC	skew1	0.008	0.000	-3.970	0.000**
GC	skew2	0.005	0.000	-2.710	0.007**
GC	Δskew1	0.000	0.004	2.028	0.043*
GC	abn_skew1	-0.001	0.001	0.482	0.630
GC	conv	0.001	-0.000	-0.456	0.649
NG	cvol	-0.009	-0.003	1.282	0.200
NG	skew1	-0.001	-0.003	-0.447	0.655
NG	skew2	-0.002	-0.003	-0.149	0.881
NG	Δskew1	-0.002	-0.002	0.018	0.985
NG	abn_skew1	-0.006	-0.006	-0.086	0.931
NG	conv	-0.000	-0.007	-1.338	0.181
S	cvol	0.002	-0.001	-1.578	0.115
S	skew1	0.004	-0.001	-2.486	0.013*
S	skew2	0.002	0.000	-1.077	0.282
S	Δskew1	-0.001	0.001	1.267	0.206
S	abn_skew1	0.002	0.003	0.731	0.465
S	conv	0.003	0.003	-0.072	0.943

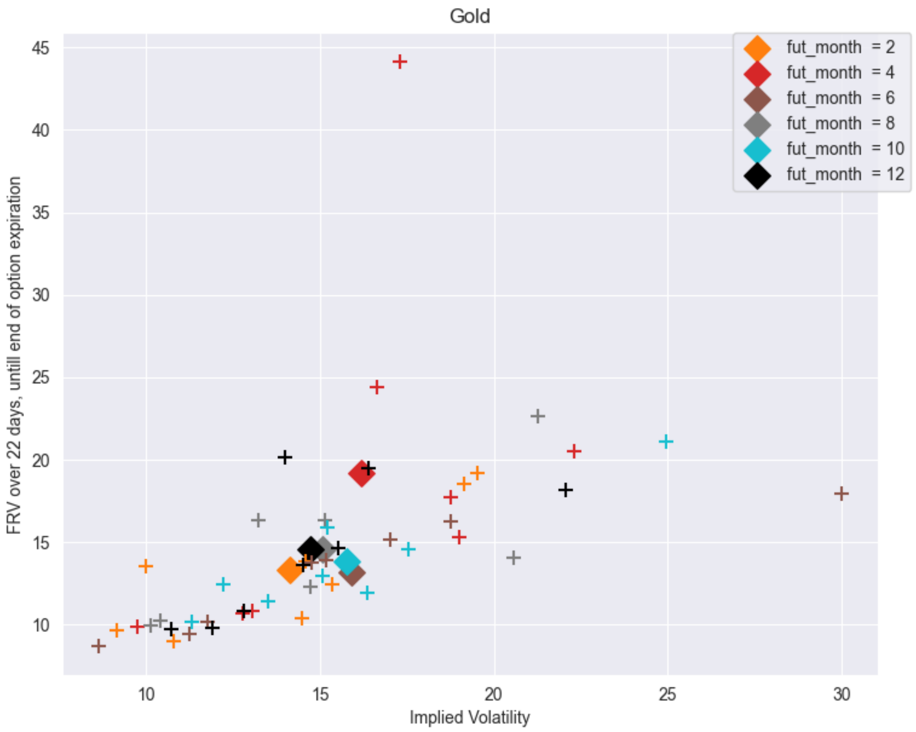
The first column denotes the commodity. The second column denotes the tested derivative measure. Trading days of the commodity are sorted in five quantiles based on the level of the derivative. The third column reports the mean of the returns of the highest quantiles after five trading days (in %). The fourth column reports the mean of the returns of the lowest quantile after five trading days (in %). The fifth column report the t-statistic of the Welch t-test that is performed on the returns of the highest and lowest quantile. The last column report the p-value of the t-test.



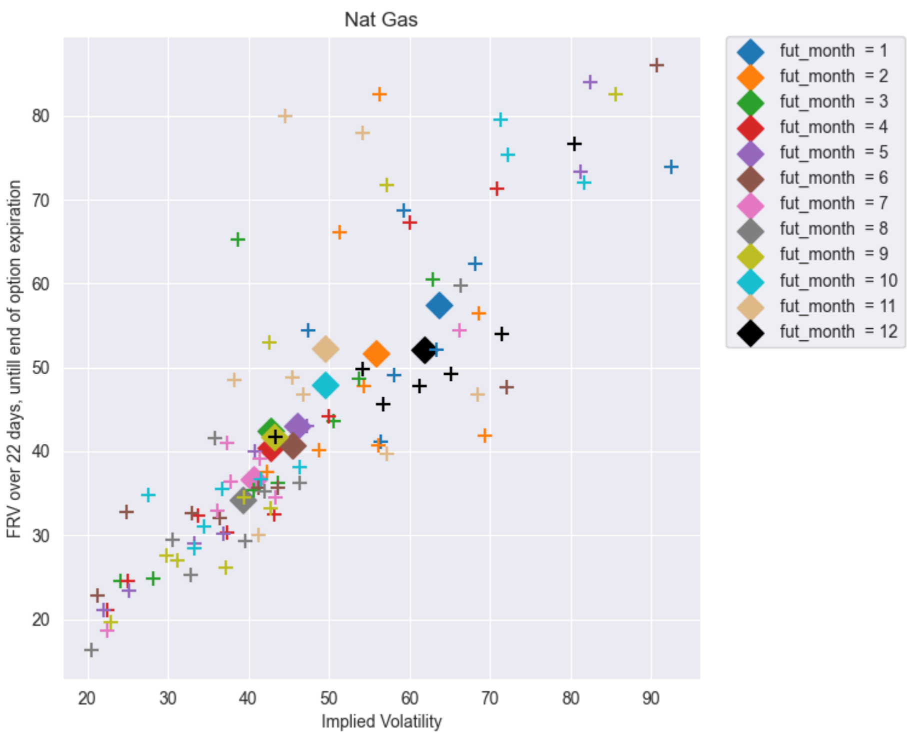
(a) Scatterplot of Implied Volatility and future realised volatility of Corn futures. Different Colors denote different future expiration months.



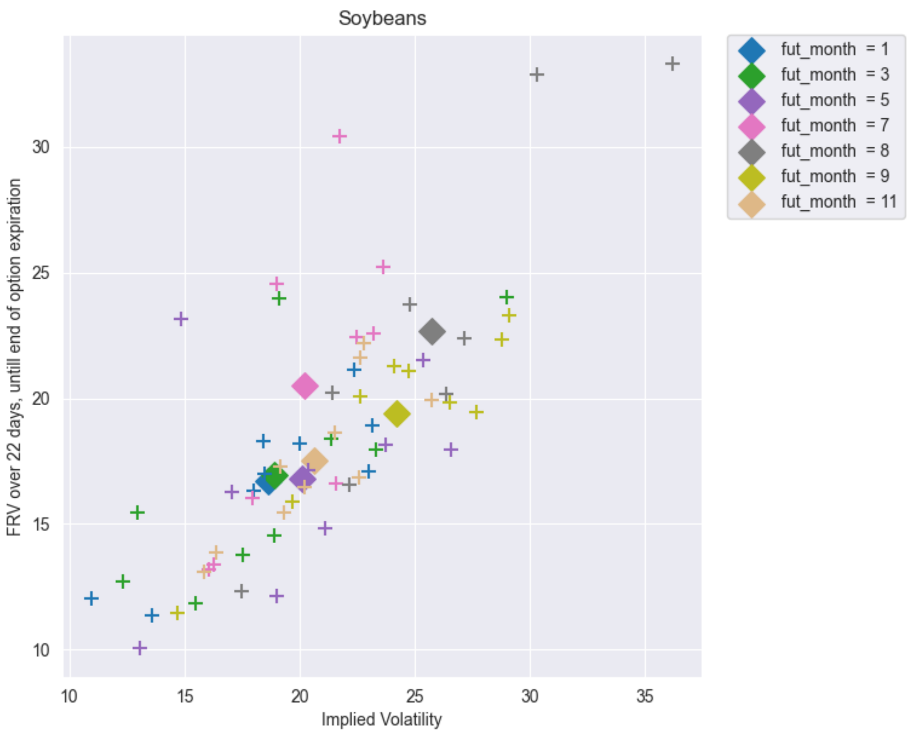
(b) Scatterplot of Implied Volatility and future realised volatility of Crude Light futures. Different Colors denote different future expiration months.



(a) Scatterplot of Implied Volatility and future realised volatility of Gold futures. Different Colors denote different future expiration months.



(b) Scatterplot of Implied Volatility and future realised volatility of Nat Gas futures. Different Colors denote different future expiration months.



(a) Scatterplot of Implied Volatility and future realised volatility of Soybeans futures. Different Colors denote different future expiration months.