Volatility Forecasts Embedded in the Prices of Crude-Oil Options

Dudley Gilder[*] and Leonidas Tsiaras[**]

[*]Aston Business School, e-mail: d.gilder@aston.ac.uk
[**]Aston Business School, e-mail: l.tsiaras@aston.ac.uk

June 2018

Abstract

This paper evaluates and compares the ability of alternative option-implied volatility measures to forecast the monthly realized volatility of crude-oil returns. We find that a corridor implied volatility measure that aggregates information from a narrow range of option contracts consistently outperforms forecasts obtained by the popular Black-Scholes and model-free volatility expectations, as well as those generated by a high-frequency realized volatility model. In particular, this measure ranks favorably in all regression-based tests, delivers the lowest forecast errors under either symmetric or asymmetric loss functions, and generates economically significant gains in volatility timing exercises. Our results also show that the CBOE's "oil-VIX" (OVX) index performs poorly, as it routinely produces the least accurate forecasts.

Keywords: Volatility forecasting, option-implied volatility, realized variance
JEL classifications: C53, C58, G17

1 Introduction

In economic terms, crude-oil is the most important traded commodity. Unsurprisingly, a wide range of economic agents, from individual investors to policy makers, closely monitor its price and routinely attempt to make predictions about the future. Unlike standard financial assets, however, one salient feature of crude-oil prices is that they can experience dramatic shifts for reasons that are largely unrelated to global macroeconomic conditions, such as OPEC policy changes or geopolitical instability in oil-producing regions. It is therefore tempting to expand the information set of standard time-series models, which rely exclusively on the record of historical prices, with measures that have "forward-looking" features by construction.

Traded options appear as a natural candidate for this task. In theory, market efficiency dictates that their prices should instantaneously adjust to any piece of relevant information pertaining to their expiration horizon.[1] In practice, option-implied measures have indeed shown remarkable forecast performance across a variety of markets and asset classes. As evidenced by the voluminous literature, this is particularly true for volatility forecasting.[2,3]

However, given a panel of option prices, different measures of expected, risk-neutral, return variation can be explored, so empirical evidence contrasting the various alternatives are needed in order to proceed. In this paper, we attempt to fill this gap for the case of crude-oil volatility. We do so by comparing the forecast performance, using both statistical and economic criteria, of various popular alternatives, such as the Black-Scholes at-the-money implied volatility (ATMIV), the so-called model-free implied volatility (MFIV) and CBOE's Crude Oil Volatility Index (OVX). More importantly, we investigate, for the first time in the commodities literature, corridor implied volatility measures introduced by Carr and Madan (1998) and Andersen and Bondarenko (2007) and obtain very promising results.

The main strand of the crude-oil volatility forecasting literature focuses on the performance of models belonging either to the GARCH family, introduced by Bollerslev (1986), or those based on realized variance measures, notably the HAR model of Corsi (2009) and its extensions. Sadorsky (2006) finds that simple GARCH models perform better than more complicated alternatives such as bivariate GARCH, vector autoregression and state-space models. Kang, Kang, and Yoon (2009) compare various GARCH models with a particular focus on those that exhibit long memory. More recently, Sévi (2014) and Prokopczuk, Symeonidis, and Wese (2016) utilize high frequency returns and compare the original HAR model of Corsi (2009) with its variants that separately model the continuous and discontinuous variance components. Interestingly, both studies find that jump decompositions do not outperform the simple HAR model in terms of out-of-sample forecast accuracy.

Commencing with Latané and Rendleman (1976), there is a large collection of papers examining the information content of option-implied volatilities (OIV) for the purpose of forecasting, with a handful of papers focusing on crude-oil. Naturally, early studies focused on producing forecasts using information in the ATMIV. Day and Lewis (1993) were the first to examine this measure and found that ATMIV-based forecasts encompass those generated by GARCH and EGARCH models. Szakmary, Ors, Kim, and Davidson (2003) also compared ATMIV with GARCH-based forecasts for 35 different assets, including crude-oil, and found that ATMIV was superior in their forecasting comparisons. Similarly, Martens and Zein (2004) find that option-implied forecasts rank favourably against GARCH alternatives, although the best models are those that combine both option-implied and high-frequency return information. Agnolucci (2009) also favours models that combine GARCH and option-implied forecasts, however the former appear to be more accurate than the latter in individual forecast comparisons.

While the early literature has examined the information content of Black-Scholes implied

---

[1]Melick and Thomas (1997) provide an insightful study of how the option-implied probability distributions of oil prices reflected market commentary during the first Gulf War.

[2]See Christoffersen, Jacobs, and Chang (2013) for a recent review of literature concerning the forecasting performance of various option-implied measures, including volatility.

[3]Although, strictly speaking, this paper focuses on forecasting variance, we use the terms "volatility" and "variance" interchangeably henceforth.

volatilities calculated from different strikes and maturities (Trippi (1977); Chiras and Man-aster (1978), Beckers (1981); Gemmill (1986); Fung, Lie, and Moreno (1990)) the consensus is that the simple ATMIV of a contract expiring as close to the forecast horizon appears to provide the most reliable results. More recently, ATMIV forecasts have been compared to the so-called *model-free implied volatility* (MFIV) that has a number of appealing theoretical properties.[4] The empirical evidence, however, has produced inconclusive results. Jiang and Tian (2005) study the S&P500 index and find MFIV to be more informative than ATMIV, while the opposite conclusion is reached by Andersen and Bondarenko (2007) for the same underlying asset. Taylor, Yadav, and Zhang (2010) examine individual U.S. stocks and report that ATMIV provides more accurate volatility forecasts than its model-free counter-part. Finally, in their study of three energy markets, Prokopczuk and Simen (2014) find that MFIV is more informative than ATMIV in predicting either crude-oil, heating oil or natural gas volatility. They also find that a simple adjustment for volatility risk-premia enhances the forecast performance of all option-implied measures.

When the task at hand is to predict future return variation MFIV is not without short-comings. This is mainly for two reasons. First, some options included in the calculation of this measure (such as deep out-of-the money puts for the case of equities) tend to be very sensitive to volatility risk-premia fluctuations. This can introduce substantial variation in the option-implied measures that is largely unrelated to the forecast target, i.e. integrated variance. Second, calculating MFIV requires that market prices of options with extreme strikes are observed. In practice, this means that either some extrapolation scheme must be implemented, or that options beyond a certain strike range should be excluded from the calculation.

The most popular estimates of MFIV measures are the volatility indices produced and published by CBOE, such as the VIX index for the case of the S&P500 and the OVX for the case of crude-oil. CBOE's implementation algorithm, which is common for both the VIX and OVX indices, adopts a liquidity-based cut-off point that determines the range of options to be included in the MFIV measure calculation. The choice of this algorithm by CBOE has recently attracted some criticism. Andersen and Bondarenko (2007) were the first to note that the VIX is in fact an ex ante measure of corridor integrated variance, rather than integrated variance. In two comprehensive empirical studies, Andersen, Bondarenko, and Gonzalez-Perez (2015) and Andersen, Fusari, and Todorov (2017) use high-frequency option data and report that the VIX calculation method introduces systematic biases to the extracted measure, including artificial jumps, which become particularly pronounced during periods of market stress. From a different viewpoint, Griffin and Shams (2018) put forth evidence pointing towards market manipulation of the VIX futures market. In essence, this is facilitated by CBOE's adopted cut-off algorithm, as speculators can temporarily boost the liquidity of deep out-of-the money S&P 500 options, increasing the level of the VIX just before the settlement price for VIX futures is determined. Given that the same method-ology is used to calculate both the VIX and OVX indices, all the above raise reasonable concerns regarding the informational efficiency of OVX-based forecasts. In addition, since

---

[4]As shown in Carr and Madan (1998), Demeterfi, Derman, Kamal, and Zou (1999) and Britten-Jones and Neuberger (2000), MFIV is a, risk-neutral, *ex ante* expectation of future return variation and, unlike the Black-Scholes implied volatilities, can be directly obtained from observed option price data without assuming a particular option pricing model.

the popularity of volatility indices has recently become widespread in the finance industry, a comparison between the OVX and other option-implied alternatives appears to be long overdue.

Our work builds on the study of Andersen and Bondarenko (2007) who propose an alternative measure of ex ante risk-neutral expectation of volatility, the so-called *corridor implied volatility* (CIV). Similar to the MFIV, and unlike the Black-Scholes model, this measure aggregates volatility information from several options and does not depend on a particular option pricing model. However, the extracted measure is not a risk-neutral expectation of integrated variance but corridor integrated variance, i.e., return variation accumulated only when the asset price lies within a corridor of two pre-specified price levels. The advantage of this approach is that one can select a corridor width that, while containing a wide-range of option prices, excludes those with extreme strikes, avoiding both price extrapolations and liquidity-driven cut-off points that may influence the reliability of the extracted measure.

The contribution of this paper is twofold. First, we examine the forecast performance of CIV measures vis-à-vis a collection of competing alternatives, including HAR, MFIV, OVX and ATMIV forecasts, for the case of crude-oil. Our paper builds, but significantly expands, on the work done by Prokopczuk and Simen (2014) who compare the performance of MFIV and ATMIV forecasts. Besides considering additional option-implied measures, our study also includes models that utilize high-frequency return information, while forecasts are ranked using both statistical and economic criteria. Second, we provide the first empirical evaluation of the OVX index, used in the forecasting study of Haugom, Langeland, Molnár, and Westgaard (2014), against other option-implied alternatives.

Our empirical results provide insights on a number of issues. We find that a particular CIV measure, that uses a relatively narrow range of option prices, consistently ranks favorably against all other competing measures using a variety of statistical and economic criteria. In particular, model forecasts that utilize this measure achieve the highest R^2 in Minzer-Zarnowitz regressions, remain significant in encompassing regression tests, and deliver the most accurate forecasts for under both the symmetric and asymmetric loss functions we consider. Moreover, volatility timing exercises show that utilizing this measure results in significant economic gains. With respect to the performance of the CBOE's OVX index, we find clear evidence that this measure is problematic, as it is routinely outperformed by all other option-based alternatives. Finally, in contrast to Prokopczuk and Simen (2014), we find that ATMIV is more informative about future crude-oil volatility compared to the MFIV measure[5].

The structure of the paper is as follows. Section 2 discusses various measures of volatility that we use in this study. Section 3 describes the dataset and the details of our methodology. The empirical results are presented in Section 4. Robustness checks can be found in Section 5. Section 6 concludes.

---

[5]The same result is also reported in Andersen and Bondarenko (2007) for the case of the S&P 500, and Taylor et al. (2010), for the case of individual stocks.

2 Volatility Measures

In this section we describe the alternative volatility measures we use to construct forecasts. However, before doing so, we state the assumptions we make about asset price dynamics and outline the relevant theory on which all our volatility measures are based.

2.1 The Dynamics of Futures Prices

Assume that over the period $t \in [0, \mathbb{T}]$ investors can continuously trade in a frictionless and arbitrage-free market. In the filtered probability space $(\Omega, \mathbb{F}, \mathbb{P}; \mathbb{F}_{t \in [0, \mathbb{T}]})$, the futures price of a contract expiring at time $0 < T < \mathbb{T}$, denoted as F_t, evolves according to the following general diffusion,

$$\frac{dF_t}{F_t} = \mu_t dt + \sigma_t dW_t \quad , \tag{1}$$

where W_t is a Wiener process. The drift μ_t and volatility σ_t can change across time according to the filtration $\mathbb{F}_t$.

The constraint imposed on the futures price dynamics is that the stochastic process is a semimartingale without jumps in prices[6]. It is worth noting that the only restriction imposed on the volatility dynamics is that σ_t is a strictly positive (càdlàg) stochastic process, so volatility can exhibit jumps.

According to these price dynamics, the total variation of logarithmic futures price changes from $t = 0$ to T is given by the *integrated variance (IVAR)*, defined as,

$$IVAR_{0,T} = \int_0^T \sigma_t^2 dt \tag{2}$$

Although total return variation is the forecast target of this paper, we also utilize the concept of *corridor integrated variance (CIVAR)*, i.e., variance accumulated only when the underlying asset (F_t in our case) lies between two "barrier" price levels B_1 and B_2. Defining the indicator function I_t that takes the value of 1 if $B_1 \leq F_t \leq B_2$ and 0 otherwise, CIVAR is given by the following expression,

$$CIVAR_{0,T} = \int_0^T \sigma_t^2 I_t(B_1, B_2) dt \tag{3}$$

Obviously, when the corridor defined by B_1 and B_2 is sufficiently wide to contain all levels that the futures price can reach with positive probability under $\mathbb{P}$, CIVAR will be equal to IVAR. In other words, IVAR can be seen as a special case of CIVAR, since for $B_1 = 0$ and $B_2 = \infty$ the definitions of the two measures coincide.

---

[6]Price jumps are excluded from this representation because the OIV expectations, discussed later in the paper, will be biased when prices are subject to discontinuous movements. Jiang and Tian (2005) and Carr and Wu (2009) note that the bias will not be sizeable for small or moderate jumps, although large jumps could have a significant impact, as argued by Carr, Lee, and Wu (2012).

2.2 Volatility Expectations From Option Prices

Option markets may be informative about future volatility, since observed prices can be utilized to extract forward-looking expectations of the aforementioned volatility measures. In particular, suppose European plain vanilla options, written on an underlying futures contract F_t and expiring at time $t = T$, trade for a continuum of strike prices K. As shown in Carr and Madan (1998), Demeterfi et al. (1999) and Britten-Jones and Neuberger (2000), *ex ante risk-neutral* expectations of the future integrated variance can be obtained by calculating the value of a static position in a portfolio of European options. Specifically, the expected integrated variance from time $t = 0$ to time T, under the risk-neutral measure $\mathbb{Q}$, can be calculated from:

$$MFIV_{0,T} = E^{\mathbb{Q}}\Big[IVAR_{0,T}\Big] = E^{\mathbb{Q}}\bigg[\int_0^T \sigma_t^2 dt\bigg] = 2e^{rT}\int_0^\infty \frac{M_{0,T}(K)}{K^2}dK \quad , \qquad (4)$$

where $M_{0,T}(K)$ is the price of a European out-of-the money option (i.e., either put or call), with strike price K and maturity T. Since this expectation does not depend on a particular option pricing model (such as the Black-Scholes model for example), it is referred to as *Model Free Implied Variance* (MFIV).

Similarly, as shown in Carr and Madan (1998) and Andersen and Bondarenko (2007), *Corridor Implied Variance* (CIV), i.e. the risk-neutral expectation of future integrated corridor variance, can be obtained by calculating the value of a static position in European options with strikes ranging from B_1 to B_2,

$$CIV_{0,T} = E^{\mathbb{Q}}\Big[CIVAR_{0,T}\Big] = E^{\mathbb{Q}}\bigg[\int_0^T \sigma_t^2 I_t(B_1, B_2)dt\bigg] = 2e^{rT}\int_{B_1}^{B_2} \frac{M_{0,T}(K)}{K^2}dK. \quad (5)$$

2.3 CBOE Crude Oil Volatility Index (OVX)

The last option-based measure we consider is the Crude Oil ETF Volatility Index (OVX), also known as the "Oil VIX". The OVX, which is produced and disseminated by the CBOE, intends to measure the market's (risk-neutral) expectation of crude-oil price volatility over the next month. It is defined as the square root of MFIV, given in Equation 4. The data underlying the OVX computation are options written on the United States Oil Fund (USO), an ETF that is designed to track the price of West Texas Intermediate light sweet crude-oil.

For the construction of the OVX the CBOE adopts exactly the same methodology as the one employed for the popular S&P 500 VIX index. Notably, CBOE applies a liquidity criterion to determine the range of option contracts included in the calculation of the index. In particular, moving from high (low) strike, out-of-the-money, put (call) options towards those with lower (higher) strikes, once two contracts with consecutive strike prices have zero bid prices a cut-off point is applied and no further contracts are considered. Therefore, both the VIX and the OVX are, in reality, CIV measures, with a corridor width determined by market liquidity.

2.4 Realized Variance

While our forecast target, i.e. integrated variance, is inherently latent, accurate ex post IVAR estimates can be obtained using high-frequency price observations. In particular, Barndorff-Nielsen and Shephard (2002), Meddahi (2002) and Andersen, Bollerslev, Diebold, and Labys (2003) show that summing squared intraday returns leads to an estimator which converges in probability to IVAR and is referred to as *realized variance* (RV). To calculate RV, suppose on day t there are $M+1$ equally spaced intraday price observations at times t_i, $i = 0, \ldots, M$. We will also assume the interval between the observations is $1/M$, i.e., the length of a day is standardised to unity. If the log price at time t_i is p_{t_i}, then the intraday return between times t_{i-1} and t_i is $r_{t_i} = p_{t_i} - p_{t_{i-1}}$. It is then straightforward to calculate RV on day t as,

$$RV_t = \sum_{i=1}^{M} r_{t_i}^2.$$

Theoretically, RV becomes more accurate as M increases, i.e., as more intraday prices are observed over shorter and shorter intervals. However, if prices are observed over very short intervals, RV will be contaminated by microstructure noise, which causes an upward bias (Andersen, Bollerslev, Diebold, and Labys, 2000).[7] A common remedy is to use prices observed over a relatively coarse set of intraday times so that the effects of microstructure noise are mitigated. Typically, prices recorded over 5-min intervals are used, despite transactions occurring at a much higher frequency. Although using a coarse intraday sampling interval solves the microstructure noise problem, it results in information being discarded. In order to recover some of this information, ensuring our RVs are estimated with as much accuracy as possible, and to continue avoiding microstructure noise by using a coarse sampling interval, we use sub-sampled RVs (Zhang, Mykland, and At-Sahalia, 2005) which are calculated as follows,[8]

$$\widetilde{RV}_t = \frac{1}{\Delta} \left(\frac{M/\Delta}{(M/\Delta) - 1} \sum_{i=2}^{M-\Delta-1} \bar{r}_{\Delta,t_i}^2 + \sum_{i=1}^{M/\Delta} \bar{r}_{t_{\Delta(i-1)+1}}^2 \right), \tag{6}$$

where $\bar{r}_{\Delta,t_i} = \sum_{j=1}^{\Delta} r_{t_{i+(j-1)}} = p_{t_{i-1+\Delta}} - p_{t_{i-1}}$ is the Δ-period intraday return between times t_{i-1} and $t_{i-1+\Delta}$, and M and Δ are chosen such that M/Δ is an integer.

---

[7]In the context of measuring IVAR, microstructure noise refers to the difference between observed transaction prices and efficient, true prices. Examples of factors that contribute to microstructure noise include: transactions tending to "bounce" between the bid and ask prices, i.e., transactions must occur at either the bid or ask price, neither of which may be the efficient, true price; and price discreteness, i.e., the minimum price change is one cent but the efficient, true price may change by fractions of a cent.

[8]Alternative microstructure-robust estimators of IVAR, which use prices observed at higher frequencies, include the kernel realized variance introduced by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and the pre-averaged realized variance developed by Podolskij and Vetter (2009) and Jacod, Li, Mykland, Podolskij, and Vetter (2009). However, in a volatility forecasting context, Liu, Patton, and Sheppard (2015) show that these estimators lead to marginal improvements over realized variance estimated using a 5-min frequency.

3 Methodology

3.1 Data and Sample Construction

Our main dataset consists of options and high-frequency prices for the WTI Light Sweet Crude Oil futures, which currently trade at CME, the world's most liquid commodity derivatives market. Our options and futures datasets start in January 1996 and end in April 2016.

3.1.1 Option Data

Option contracts are written on futures contracts for physical delivery of light sweet crude-oil. In particular, the underlying asset is the futures contract whose delivery date is three business days after the expiration of the option. These contracts have American-style exercise and are settled in cash. Our option data consist of daily settlement prices, which are recorded at 14:30 ET each day.

Our empirical study focuses on monthly variance forecasts. Along these lines, we study options that have approximately 22 trading days to expiration.[9] Throughout the sample, the day that we collect option prices is never before the maturity date of the previous option chain we studied, i.e. all our option-based forecasts are non-overlapping.

We make several adjustments to the raw data before we proceed with the estimation of the OIV measures. As the latter require price data on European options rather than American ones for their calculation, we attempt to alleviate this problem in two ways. First, we eliminate all in-the-money options and only keep out-of-the money options, for which the early exercise premium is significantly lower and, for the case of deep-out-of the money options, almost quantitatively negligible. Second, we estimate the early exercise premium of each option using the Barone-Adesi and Whaley (1987) American option pricing formula and subsequently calculate the price of their European-style counterparts.

Finally, in order to guard against recording errors and other market microstructure effects, we eliminate options with a price less than $0.01 and filter all call/put prices that violate standard arbitrage bounds. An overview of our final option data sample is provided in Table 1. It is noteworthy that the number of traded option contracts has increased substantially over the last few years.

3.1.2 High-Frequency Futures Data

Our high-frequency data comprise of transaction prices recorded at 30-sec intervals. Until June 2006, futures were traded between 09:00 and 14:30 ET using an open outcry system in a trading pit, resulting in 661 price observations per day. Subsequently, they have been traded between 18:00 and 17:00 ET the following day on the electronic GLOBEX platform, resulting in 2,761 price observations per day.

To construct RV we use 5-min intraday returns. This frequency is commonly used in the empirical literature, e.g., see Andersen, Bollerslev, Diebold, and Labys (2001), Andersen et al.

---

[9]If for a given month options with an expiration horizon of 22 trading days are not available, we shorten the target horizon by one or (if needed) two days and scale the resulting OIV such that its expiration horizon corresponds to 22 days. For example, if the expiration horizon of the OIV was N trading days, we scale the OIV by multiplying it by $22/N$.

(2003) and Liu et al. (2015) among others, as it is deemed to provide an appropriate trade-off between the objective of incorporating as much information as possible from intraday prices and the necessity to avoid contamination from microstructure noise. Hence, in applying our sub-sampled RV estimator in Equation (6) we set $M = 660$ or $M = 2760$, depending on which dates the data are from, and $\Delta = 10$.

3.1.3 Construction of Implied Volatility Measures

The computation of the MFIV and CIV measures requires the existence of options trading for a continuum of strike prices, an assumption which is of course not satisfied in practice. In order to address this, we first estimate a risk-neutral distribution using the prices of observed options for each relevant date, which enables us to subsequently generate option prices for arbitrary strike prices. Our preferred risk-neutral distribution is the flexible Generalized Beta Distribution of the second kind (GB2) which, as discussed in Taylor (2005), has a number of appealing properties.[10]

The calculation of the CIV measures also requires a selection of the relevant corridor width, i.e., the barrier price levels B_1 and B_2. We consider four CIV measures in total, with the barriers determined by evaluating the quantile function of the risk-neutral distribution F_Q. Specifically, defining $B_1 = F_Q^{-1}(p)$ and $B_2 = F_Q^{-1}(1-p)$, the CIV1 to CIV4 measures are obtained by first setting $p = 0.45, 0.35, 0.25, 0.10$, respectively, and subsequently evaluating Equation (5).

Figure 1 plots the four CIV measures together with the MFIV estimates during our full sample (1996-2016) period. As expected, all measures exhibit a strong degree of covariation, although narrow CIV measures appear more stable than their wide corridor counterparts. As shown in Table 2, all CIV measures are highly correlated with both MFIV and ATMIV. The autocorrelation patterns for 1, 6, and 12 lags also reveal that CIV measures become less persistent as the width of the corridor widens. As expected, the unconditional moments of the CIV measures approach those of the full MFIV as more options are included in the calculations, while for the case of CIV4 the two measures have almost identical properties.

3.2 Volatility Forecast Construction

All our variance forecasts are non-overlapping and correspond to a monthly horizon (22 trading days). We consider two samples of different length, one that includes the OVX and one that does not. If we exclude the OVX from the set of OIVs evaluated, then our sample starts in January 1996 and ends in April 2016. We refer to this as our *full sample*. Conversely, if we include the OVX, then our sample starts in May 2007, when the OVX is first reported, and ends in April 2016. We refer to this as our *OVX sample*. Our full and OVX samples consist of 244 and 108 non-overlapping monthly observations, respectively.[11]

Our OIVs provide *ex ante* expectations of monthly variance, hence they form stand-alone forecasts of variance. In contrast, RV is an *ex post* estimator of variance. Although we could

---

[10]The GB2 distribution, firstly proposed by Bookstaber and McDonald (1987), allows for general levels of skewness and kurtosis, while European option prices can be computed in closed-form.

[11]Note, the number of non-overlapping monthly observations stated here are for the combined sample of OIVs and RVs and may differ to the number reported in Table 1, which referred to the sample of OIVs only.

use RV in a random-walk model, i.e., use the lagged RV as a stand-alone forecast, there is a considerable literature demonstrating the superiority of forecasts generated from time-series models of RV (e.g., Corsi (2009), Patton and Sheppard (2015), Bollerslev, Patton, and Quaedvlieg (2016) and Bollerslev, Hood, Huss, and Pedersen (2018)). We use the heterogeneous autoregressive model (HAR) of Corsi (2009) to generate RV-based forecasts,

$$RV_{t+1,t+22} = \beta_0 + \beta_1 RV_{t,t} + \beta_2 RV_{t-4,t} + \beta_3 RV_{t-21,t} + \epsilon_{t+22},$$

where, $RV_{u,v} = \sum_{s=u}^{v} \widetilde{RV}_s$. While various extensions of the HAR model could be considered, our study relies exclusively on the baseline version. This is because, as shown in the comprehensive studies of Sévi (2014) and Prokopczuk et al. (2016), sophisticated HAR extensions do not outperform the simple HAR benchmark[12]. We estimate the HAR model parameters using a rolling window of 60 monthly observations. Forecast evaluation is conducted out-of-sample. Thus, in the following sections, reference to our full sample will be to the 184 out-of-sample observations. Similarly, reference to our OVX sample will be to the 48 out-of-sample observations.

4 Empirical Results

We examine the forecast accuracy of competing forecasts using several techniques. Firstly, we evaluate the information content of each of our forecasts using Mincer-Zarnowitz regressions. Secondly, we make comparisons between the information content of our forecasts using encompassing regressions. Thirdly, we analyse the prediction errors of our forecasts using statistical loss functions. Lastly, we assess the economic value of our forecasts by implementing a volatility timing exercise.

4.1 Mincer-Zarnowitz Regressions

Our first evaluation procedure assesses the information content of our HAR-based forecasts and the OIVs. This is done by running Mincer-Zarnowitz regressions (Mincer and Zarnowitz, 1969), whereby we regress our variance target, the RV calculated over each out-of-sample monthly forecast horizon, against the competing forecasts. More precisely, for each forecast i, we run the following regression,

$$RV_{t+1,t+22} = \beta_0 + \beta_1 f_{t,t+22}^i + \epsilon_{t+22}, \tag{7}$$

where $f_{t,t+22}^i$ is the forecast from model i using information available up until day t for the variance between days $t+1$ and $t+22$. The information content of each forecast is measured by the R^2 of this regression.

In Table 3 we present results for the Mincer-Zarnowitz regressions. In Panel A we report results for the full sample, whilst in Panel B we summarize results for the OVX sample. The values of R^2 suggest the following. Firstly, OIVs have markedly higher information content. Forecasts based solely on RV, i.e., the HAR forecasts, result in the lowest R^2 values, whilst

---

[12]Experimentation with alternative HAR models corroborated this finding for our sample. Since our results replicate those of Sévi (2014) and Prokopczuk et al. (2016) we do not report them to conserve space.

the R^2 for the OIV forecasts are larger by 5-7 percentage points for the full sample and 35-37 percentage points for the OVX sample. Secondly, of the OIVs, CIV1 appears to have the highest information content. The CIV1 forecasts result in the largest R^2s amongst all of the OIVs in both the full and OVX samples.

It should also be noted that the parameter values differ substantially between the forecasts. This is as expected and a consequence of the differing levels of bias in the forecasts. If a forecast is unbiased, then we would expect $\beta_0 = 0$ and $\beta_1 = 1$. Whilst all of the β_0 parameters are close to zero, there are large differences between the values of β_1. Overall, the MFIV, ATMIV and OVX are upwardly biased, consistent with the presence of a variance risk-premium, whilst CIV1, CIV2 and CIV3 are downwardly biased, reflecting the fact that they provide risk-neutral expectations for CIVAR. The observed biases are consistent with the mean values for the OIVs reported in Table 2.

4.2   Encompassing Regressions

Our second evaluation procedure assesses the relative performance of the alternative forecasts by running encompassing regressions. We make comparisons between two forecasts by running the following bivariate regressions,[13]

$$RV_{t+1,t+22} = \beta_0 + \beta_1 f^i_{t,t+22} + \beta_2 f^j_{t,t+22} + \epsilon_{t+22}. \tag{8}$$

We can determine whether the forecast from one model encompasses the other by examining the significance of the individual regression parameters. If the information contained in the forecast from model i is subsumed by the information in the forecast from model j, then we expect β_1 to be insignificant and β_2 to be significant. If this occurs, then we say that the forecast from model i encompasses the forecast from model j and vice versa.

We make the following comparisons using our encompassing regressions: (i) we compare the information content of the HAR model forecasts, which are based on RV alone, against our OIVs to examine whether the forward-looking information in our OIVs is useful vis-à-vis the backward-looking information contained in RV; and (ii) we compare the information content of our alternative OIVs to examine which of our OIVs contains the most useful forecasting information.

The results of the encompassing regressions are summarized in Table 4. Results for comparisons between the OIVs and the HAR forecasts are summarized in Panels A and B for the full and OVX samples, respectively. Panel A shows that CIV1, CIV2 and CIV3 encompass the HAR forecasts, at the 1% level. In the remaining encompassing regressions, although the parameters on MFIV, ATMIV and CIV4 are significant at the 1% level, the parameters on HAR are also significant; albeit only at the 10% level for the encompassing regressions involving ATMIV and CIV4. Thus, although they do not encompass the HAR forecasts, there appears to be incremental information in the MFIV, ATMIV and CIV4 OIVs which is not captured by RV. In Panel B, all OIVs encompass the HAR forecasts, at the 1% level.

Results for comparisons between the OIVs are summarized in Panels C and D for the full and OVX samples, respectively. The results in Panels C and D provide more refined insights

---

[13]This is the most common way to implement encompassing regressions.

regarding the relative information content of the OIVs. Focusing on the full sample results in Panel C, it can be seen that the MFIV variances are encompassed by all the other OIVs, at the 5% level. This result is consistent with the arguments of Andersen and Bondarenko (2007) that illiquid options with strike prices in the tails of the risk-neutral distribution (RND) and the requirement to extrapolate the tails of the RND reduce the accuracy of the MFIV forecast. Hence, despite its inferior theoretical foundations, the ATMIV is able to provide superior forecasts to the MFIV variance.[14] Comparisons of the ATMIV and CIVs show that ATMIV is encompassed by CIV1, CIV2 and CIV3, at the 5% level. Thus, there is a trade-off between excluding imprecisely measured information from the tails of the RND and limiting the information used to construct OIVs to that contained in the at-the-money (ATM) option. It appears that CIV1, CIV2 and CIV3 make a trade-off which enables them to harness the information contained in the liquid strikes that straddle the ATM contract without being contaminated by measurement errors in the tails of the RND. The encompassing regressions involving CIV1, CIV2, CIV3 and CIV4 show that CIV4 is encompassed by CIV1, CIV2 and CIV3, at the 5% level. This is indicative of the corridor being too wide for CIV4 such that it too is contaminated by noise in the tails of the RND.

The results in Panel D are much weaker than those in Panel C and likely reflect the smaller size of the OVX sample. Nonetheless, although often significant at the 10% level only, the results show that CIV1 encompasses all other OIVs.

Thus, the results from the encompassing regressions support those from the Mincer-Zarnowitz regressions. The information content of the OIVs appears to be superior relative to those based on RV alone, i.e., the HAR forecasts. Further, of the OIVs, the CIV1 appears to be informationally superior.

4.3 Bias-corrected and Augmented Models

Unsurprising, our previous results showed that our OIV forecasts are biased. By comparing the adjusted R^2s of the encompassing regressions in Table 4 to the R^2s of the Mincer-Zarnowitz regressions in Table 3, we can also see that a linear combination of RVs and OIVs leads to a higher information content than can be attained with any individual forecast. Therefore, we introduce two additional sets of forecasting models. Firstly, instead of using our OIVs as stand-alone forecasts, we use predicted values from univariate regressions of $RV_{t+1,t+22}$ on each of our OIVs. Intuitively, the objective of this set of models is to remove the bias from the OIVs. Hence, we refer to the predicted values from these models as bias-corrected option-implied forecasts. We use BC-MFIV to denote the bias-corrected MFIV forecast, BC-ATMIV to denote the bias-corrected ATMIV forecast and so on.

Secondly, we augment the HAR model by adding each of the OIVs. Thus, the specification of our augmented HAR models is,

$$RV_{t+1,t+22} = \beta_0 + \beta_1 RV_{t,t} + \beta_2 RV_{t-4,t} + \beta_3 RV_{t-21,t} + \beta_4 OIV_{t,t+22} + \epsilon_{t+22},$$

where $OIV_{t,t+22}$ is an OIV calculated on day t for return variation between days $t + 1$ and $t + 22$. We use HAR-MFIV to denote the forecasts from an HAR model augmented with MFIV, HAR-ATMIV the forecasts from an HAR model augmented with ATMIV and so on.

---

[14]This is consistent with the evidence in Taylor et al. (2010) for individual stock variances.

Therefore, in total, we examine thirteen competing forecasts. Our benchmark is the HAR model, since this uses information from RV exclusively. We then have six bias-corrected option-implied forecasts and six augmented HAR forecasts.

4.4 Forecast Evaluation using Statistical Criteria

Although Mincer-Zarnowitz and encompassing regressions provide insights into the information content of our forecasts, they do not provide much information about their precision. From the perspective of economic agents, quantifying forecast accuracy is of paramount importance. Thus, we now turn to assessing prediction errors by means of statistical loss functions.

4.4.1 Statistical Loss Functions

To evaluate the accuracy of our forecasts, we use a symmetric, the mean squared error (MSE), and an asymmetric, the quasi-likelihood (QLIKE), loss function. A lower MSE and/or QLIKE corresponds to smaller prediction errors. These loss functions were chosen because they are commonly employed in the volatility forecasting literature and, as shown in Patton (2011), they are robust to measurement error in the IVAR proxy. More precisely, using the MSE and QLIKE loss functions ensures that the ranking of two forecasts in terms of expected loss is preserved even when the true integrated variance is replaced by a conditionally unbiased, but imperfect, proxy.[15]

For forecast i, the MSE loss function we use is defined as,

$$L_{MSE,i} = \frac{1}{\lfloor T/22 \rfloor - 1} \sum_{s=1}^{\lfloor T/22 \rfloor - 1} \left(f^i_{22(s-1),(22(s-1))+22} - RV_{22(s-1)+1,22(s-1)+22}\right)^2,$$

where $\lfloor x \rfloor$ rounds x to an integer that is less than or equal to x. Similarly, the QLIKE loss function we use is defined as,

$$L_{QLIKE,i} = \frac{1}{\lfloor T/22 \rfloor - 1} \sum_{s=1}^{\lfloor T/22 \rfloor - 1} \ln(f^i_{22(s-1),22(s-1)+22}) - \frac{f^i_{22(s-1),22(s-1)+22}}{RV_{22(s-1)+1,22(s-1)+22}}.$$

In order to test for significant differences between the MSE and QLIKE of competing forecasts we use Diebold-Mariano tests (Diebold and Mariano, 1995) with Newey-West (Newey and West, 1987) standard errors.

4.4.2 Forecast Evaluation Results

The out-of-sample MSEs and QLIKEs of our competing forecasts are presented in Table 5. Panel A reports results for the full sample whilst Panel B summarizes results for the OVX sample. In each Panel we also report the results of Diebold-Mariano tests in which

---

[15]It should also be noted that while measurement errors in the IVAR proxy can affect statistical inference, proxies constructed using high-frequency return data (such as our RV measure) offer large gains in power and improvements in finite-sample size, as shown in Patton and Sheppard (2009).

the HAR forecasts are used as a benchmark. Therefore, any significance reported indicates the MSE or QLIKE is significantly different to that of the HAR forecasts. In Panel B, to assess the accuracy of forecasts which incorporate information from the OVX, we additionally report results from Diebold-Mariano tests where the HAR-OVX forecasts are treated as the benchmark.[16]

From Panel A it can be seen that, overall, the bias-corrected forecasts outperform the corresponding HAR forecasts according to the MSE. The BC-CIV1 forecasts result in the lowest MSE and, of the HAR-based forecasts, the lowest MSE is associated with HAR-CIV1. However, Diebold-Mariano tests show that no forecast results in an MSE which is significantly lower than that generated by the HAR forecasts.

The QLIKE results in Panel A differ slightly to those for the MSE. Under this loss function the HAR-based forecasts outperform the bias-corrected forecasts, with the most accurate forecasts being HAR-CIV1. However, in contrast to the MSE results, we find that the QLIKEs for HAR-CIV1, HAR-CIV2, HAR-CIV3 and HAR-ATMIV are all significantly lower than for HAR at the 5% level. The QLIKEs for HAR-MFIV and HAR-CIV4 are also significantly lower at the 10% level. The fact that we find significant differences when using the QLIKE loss function is most likely associated with the ability of this loss function to more accurately discriminate between competing variance forecasts (Patton and Sheppard, 2009). Overall, the results in Panel A suggest that using information in CIV1 leads to the most accurate variance forecasts.

The results in Panel B are analogous to those in Panel A and support our overall conclusion that incorporating information from CIV1 leads to the most accurate forecasts. Importantly, it can also be seen that forecasts based on OVX perform poorly. When the HAR-OVX forecasts are used as a benchmark, it can be seen that nearly all non-OVX option-implied forecasts produce significantly lower QLIKEs and MSEs at the 5% level.[17] Although the BC-OVX forecasts result in an MSE that is significantly lower than that for the HAR-OVX forecasts, of the bias-corrected forecasts, BC-OVX result in the largest MSE.

In summary, the results show that prediction errors can be minimized when CIV1 is employed. The results are consistent with those from the Mincer-Zarnowitz and encompassing regressions, which showed that the OIVs were informationally superior to forecasts based on RV and that CIV1 had a higher information content relative to other OIVs. Our results also suggest that forecasts based on the OVX perform significantly worse than those based on our alternative OIVs. Thus, the results suggest there is value in constructing our OIVs directly from option prices rather than relying on the CBOE's methodology.

4.5 Forecast Evaluation using Economic Criteria

Although evaluating variance forecasts with MSE and QLIKE is common, they are statistical loss functions. It is not clear how minimising the MSE and/or QLIKE translates into eco-

---

[16]Between the OVX-based forecasts, the BC-OVX forecasts result in a lower MSE, whilst the HAR-OVX forecasts generate a lower QLIKE. However, to avoid unnecessary complexity, we only use the HAR and HAR-OVX forecasts as benchmarks in our Diebold-Mariano tests reported in Table 5. This choice has no impact on our conclusions.

[17]The BC-CIV4 (HAR-CIV1) forecasts generate a QLIKE (MSE) that is significantly lower than the QLIKE (MSE) for the HAR-OVX forecasts at the 10% level only.

nomic gains for the forecaster. Therefore, in order to ascertain whether the improved forecast accuracy we observe leads to economic gains, we perform a volatility timing exercise.

We consider an agent whose investment opportunity set consists of the WTI crude-oil futures and a risk-free asset and assume the agent's objective is to maximise the utility of a portfolio consisting of these two assets. We follow the set-up of Bollerslev et al. (2018) and assume Sharpe ratios are constant and that a quadratic utility function provides an accurate approximation to investors' true utility functions. With these assumptions, the following function describes investors' utility per unit wealth (UoW),

$$UoW_t = \frac{U(w_t)}{W_t} = \left(w_t SR\sqrt{E_t(RV_{t+1,t+22})} - \frac{\gamma}{2}w_t^2 E_t(RV_{t+1,t+22}) \right), \tag{9}$$

where W_t is the investors wealth, w_t is the proportion of the investors wealth held in crude-oil futures, SR is the Sharpe ratio and γ is the investors level of risk aversion.[18] It can be shown that constructing a portfolio to maximise expected utility is equivalent to forming a portfolio with a specific volatility target, where the optimal proportion of wealth to invest in crude-oil futures is,

$$w_t^* = \frac{SR/\gamma}{\sqrt{E_t(RV_{t+1,t+22})}}. \tag{10}$$

The ratio in the numerator corresponds to the volatility target and the denominator is the expected volatility. To operationalise the strategy, $E_t(RV_{t+1,t+22})$ in denominator is replaced with a variance forecast. We also follow Bollerslev et al. (2018) and set $SR = 0.4$ and $\gamma = 2$, which they argue are sensible parameters when forecasting variance over a monthly horizon, and results in a volatility target of 20%. Using Equation 10 to substitute for w_t in Equation 9, replacing $E_t(RV_{t+1,t+22})$ with our variance forecasts, and plugging-in our assumed values of SR and γ leads to the following expression for the utility per unit wealth based on forecast $f_{t+1,t+22}$,

$$UoW_t = 8\%\frac{\sqrt{RV_{t+1,t+22}}}{\sqrt{f_{t+1,t+22}}} - 4\%\frac{RV_{t+1,t+22}}{\sqrt{f_{t+1,t+22}}}. \tag{11}$$

Comparisons between models are then made using realized utility,

$$RUoW = \frac{1}{T}\sum_{t=1}^{T}UoW_t. \tag{12}$$

Note, realized utility is expressed as a percentage return and, given our assumptions, can take a maximum value of 4%.

4.5.1 Realized utility results

Panels A and B of Table 6 summarize the realized utility results for our full and OVX samples, respectively. The first column of Panel A reports the realized utility. It can be seen

---

[18]See Bollerslev et al. (2018) for more detail.

that the HAR-based forecasts outperform the bias-corrected forecasts. In addition, of the HAR-based forecasts, HAR-CIV1 generates the highest realized utility.

In the first column of Panel A we also report the results from Diebold-Mariano tests of whether the realized utility associated with a forecast is significantly different to the realized utility generated by the HAR forecasts. It can be seen that HAR-CIV1, HAR-CIV2 and HAR-CIV3 produce realized utilities that are significantly higher, at the 5% level for HAR-CIV1 and HAR-CIV2 and the 10% level for HAR-CIV3, than the realized utility attained with the HAR forecasts. Therefore, the pattern in forecasting performance observed with the statistical loss functions is retained when we use an economic loss function. The results also confirm that the improvements in forecasting accuracy we observed with the statistical loss functions translate into economic benefits.

The difference between the value of the realized utility for the HAR-CIV1 forecasts relative to the HAR forecasts in Panel A is 2 bp. Although this may appear to be a relatively modest difference, there are two reasons why this represents a material economic improvement. Firstly, as highlighted by Bollerslev et al. (2018), there has been a drive by investment management companies, in particular mutual funds and ETFs, towards lowering fees. The fees now charged by low-cost funds are of the order of tens of basis points. As argued by Bollerslev et al. (2018), this means that a single digit basis point increase in fees is relatively substantial. The difference in realized utility of 2 bp means that a fund using the HAR-CIV1 forecasts instead of the HAR forecasts will be able to increase its fees by 2 bp and remain equally attractive to investors.

Secondly, the realized utilities reported are unconditional. Therefore, in any given month the economic benefit of using the HAR-CIV1 over the HAR forecasts could be much larger than 2 bp. In order to examine this further, we report in columns 2-7 of Panel A in Table 6 the 2.5, 5, 10, 25, 50 and 75% quantiles of the UoW for each set of forecasts. Comparing the HAR-CIV1 to the HAR forecasts, it is clear that there are substantial differences, of 5-10 bp, in the UoW at the 2.5, 5 and 10% quantiles. The difference between the lower quantiles of the UoW distributions suggests the HAR-CIV1 forecasts tend to outperform the HAR forecasts precisely when it is most difficult to forecast volatility and when an accurate forecast is in greatest demand, e.g., when there is a volatility shock which causes UoW to be low.

The results for the OVX sample in Panel B of Table 6 are analogous to those for the full sample in Panel A. Again, the HAR-CIV1 results in the highest realized utility, with all the non-OVX augmented HAR forecasts generating a realized utility higher than that attained by the HAR forecasts. The differences between the realized utilities of each non-OVX augmented HAR forecast and the HAR forecasts are also larger and approximately 2-5 bp.

To examine the economic benefit of using our OIVs over the OVX, in Panel B we test for a significant difference between each forecast and the HAR-OVX forecasts.[19] It can be seen that the realized utility of all of the non-OVX augmented HAR forecasts are significantly higher, typically at the 1% level, than the realized utility associated with the HAR-OVX forecasts. Therefore, these results further support our conclusion that there is

---

[19]We use the HAR-OVX forecasts rather than the BC-OVX forecasts as a benchmark because, of the two, the former generate a higher realized utility.

value in constructing our OIVs directly from option prices rather than relying on the CBOE's methodology.

Similar to our findings in Panel A, there are potentially large differences between the conditional values of UoW for the OVX sample. For the 25-50% UoW quantiles the difference between the realized utilities associated with the HAR and augmented HAR forecasts is approximately 4-13 bp, whilst the difference between the HAR-OVX and non-OVX augmented HAR forecasts is approximately 9 bp.

Of course, the magnitude of the economic benefits derived from each forecast depends on the assumptions employed. As Bollerslev et al. (2018) highlight, the value of the realized utility is a linear function of the volatility target. Thus, if the volatility target doubles, e.g., through a doubling of the Sharpe ratio or a halving of the coefficient of risk aversion, the size of the economic benefits also double. Nevertheless, the framework above provides a sensible approximation and therefore the magnitude of the economic benefits presented should be reasonable.

In conclusion, the results demonstrate that the improvements in forecasting accuracy observed when using OIVs translate into economic benefits under reasonable assumptions. They also further corroborate our preference for the CIV1 amongst our OIVs.

4.5.2 Realized utility results with transaction costs

In order to make our analysis more realistic, we also take into consideration transaction costs. Specifically, transaction costs are calculated as being a proportion of turnover,

$$TC = \sum_{t=1}^{T} c \cdot turnover_t, \tag{13}$$

where,

$$turnover_t = |w_t - w_{t-1}|.$$

The precise level of transaction costs is controlled by c. We follow Wang, Liu, Ma, and Wu (2016) and Caldeira, Moura, Nogales, and Santos (2017) and set c to be either 0.033% or 0.15%. It should be noted that transaction costs are low in futures markets (Locke and Venkatesh, 1997) and have decreased markedly over the past few decades. Realised utility net of transaction costs is then given by,

$$NetRUoW = RUoW - TC. \tag{14}$$

Panels A and B of Table 7 summarize the net realized utility and transaction costs for the full and OVX samples, respectively. In Panel A (Panel B) we also use Diebold-Mariano tests to formally evaluate whether the net realized utility and transaction costs associated with each forecast are significantly different to the net realized utility and transaction costs generated by the HAR (HAR-OVX) forecasts.

In both Panels A and B it can be seen that transaction costs, whether $c = 0.015\%$ or $c = 0.0033\%$, do not differ substantially between the competing forecasts. In Panel A, it can be seen that there are no significant differences in transaction costs, whilst in Panel

B, BC-CIV1 and BC-CIV2 lead to significantly lower transaction costs at the 10% level. Consequently, because none of the forecasts lead to unusually high transaction costs, the ranking of the forecasts in Panels A and B of Table 7 are identical to those in Panels A and B of Table 6. In particular, the HAR-CIV1 forecasts produce the highest net realized utility in both the full and OVX samples. Within the full sample, the HAR-CIV1 and HAR-CIV2 forecasts result in net realized utilities that are significantly higher than the net realized utility of the HAR forecasts at the 5% level. Whilst for the OVX sample, all the non-OVX augmented HAR forecasts result in significantly higher net realized utilities than the HAR-OVX forecasts, typically at the 1% level. Thus, differences in trading volumes are small and do not have a material impact on the relative economic benefits of the forecasts. Therefore, the results also support our conclusions in Section 4.5.1

5 Robustness Checks

In this section we analyse the robustness of our results to: (i) the size of the rolling-window used to estimate the models; (ii) the choice of out-of-sample period; and (iii) the method used to bias correct the OIVs.

5.1 Estimation window

Thus far, the parameters for our bias-correction and HAR models have been estimated using rolling windows of 60 observations, or approximately five years worth of data. In this section we analyse the robustness of our results to the choice of estimation window. Specifically, we vary the estimation window between 66 and 96 observations, or between approximately five and eight years, and then re-evaluate our forecasts using the MSE, QLIKE and realized utility loss functions. We do not use an estimation window below 60 observations since this results in poorly estimated parameters and recalcitrant forecasts. Due to the limited number of observations, we are also unable to conduct this robustness test for the OVX sample. Table 8 summarizes the results of this analysis. Reassuringly, our main conclusions continue to hold. For all estimation windows, the BC-CIV1 model minimizes the MSE, whilst the HAR-CIV1 minimizes the QLIKE and maximises realized utility. Furthermore, our results using an estimation window of 60 observations are superior to those found using estimation windows between 66 and 96 observations.

5.2 Sub-sample analysis

In our analysis in Section 4, our out-of-sample period started in Jan 2001 and ended in April 2016. In this section we check the robustness of our results to variations in the out-of-sample period. Specifically, we vary the out-of-sample start date to be either Jan 2002, Jan 2003, ..., or Jan 2012. This ensures we have sub-samples that both include and exclude the 2007/8 financial crisis. We do not consider a start date beyond 2012 because this would result in an insufficient number of out-of-sample observations.

In Table 9 we summarize the results from using the MSE, QLIKE and realized utility loss functions. Each column refers to the year in which the out-of-sample evaluation begins, with

each analysis commencing in January of the associated year, and also reports the number of observations in each sub-sample. The pattern observed over the full out-of-sample period is retained in each of the sub-samples; the BC-CIV1 provides the most accurate forecasts according to the MSE whilst the HAR-CIV1 forecasts are the most accurate according to the QLIKE and realized utility. Therefore, we are convinced our results are robust to the choice of out-of-sample period.

5.3 Alternative Bias Correction Procedure

Our final robustness check evaluates an alternative bias correction technique. Thus far, the bias in OIVs has been corrected for by using a regression model or by including the OIVs in the HAR model, where the parameter multiplying the OIV is able to modulate the bias. In this section we use a technique inspired by Prokopczuk and Simen (2014), who show that the forecasting performance of the MFIV can be improved significantly by making a non-parametric adjustment for the variance risk premium. We apply their technique to all our OIVs.[20] Since this method relies on averages of ratios, we will refer to it as a relative bias-correction.

   More precisely, to implement the relative bias-correction, the average ratio of OIV to RV must be computed to get an estimate of the relative bias, which, in our case, is estimated using a window of τ monthly non-overlapping observations,

$$RB_t^{IV} = \frac{1}{\tau - 1} \sum_{s=1}^{\tau-1} \frac{OIV_{t-(22s),t-(22s)+22}}{RV_{t-(22s)+1,t-(22s)+22}}. \tag{15}$$

The relative bias-corrected OIV can then be estimated as follows,

$$RBCIV_t = \frac{OIV_{t,t+22}}{RB_t^{IV}}. \tag{16}$$

In Table 10 we summarize the MSE, QLIKE and realized utility for the relative bias-correction forecasts. As benchmarks, we include our, thus far, best-performing (non-relative) bias-corrected and HAR-based forecasts: the BC-CIV1 and HAR-CIV1 forecasts. To evaluate the effect of the window size used in the relative bias-correction, we report results for $\tau = \{6, 12, \ldots, 60\}$.[21] Several conclusions can be made. Firstly, the relative bias-corrected forecasts are most accurate when $\tau = 12$. Secondly, the relative bias-corrected forecasts, for all values of τ, are more accurate than the BC-CIV1 and HAR-CIV1 forecasts under the MSE loss function. Except for when $\tau = 24$, the MSE is minimized by the RBC-CIV2 forecasts. Lastly, under the QLIKE and realized utility loss functions, the HAR-CIV1 forecasts provide the most accurate forecasts for all values of τ. Overall, although there is some evidence that the relative bias-correction may produce more accurate forecasts under the MSE loss function, our conclusions concerning the value of the information content of OIVs and particularly CIV1 remain unaltered.

---

[20]Although the approach can be thought of as a correction for the variance risk premium when using the MFIV, the same interpretation does not hold for all of our OIVs, particularly the CIVs. Instead, we take a pragmatic view and treat it as an alternative bias-correction method.

[21]These effectively correspond to window sizes of $0.5, 1, 1.5, \ldots, 5$ years.

6   Conclusion

In this paper we evaluated, using both economic and statistical criteria, the information content of monthly crude-oil volatility forecasts extracted from the prices of traded options. We examined a variety of alternative option-implied measures including Black-Scholes at-the-money implieds (ATMIV), model-free volatility expectations (MFIV), CBOE's oil-VIX (OVX) and, notably, corridor implied volatilities (CIV). Besides stand-alone comparisons, option-implied forecasts were also contrasted, and combined, with those obtained by a realized volatility model (HAR) that utilizes high-frequency return information.

Our key finding is that a particular CIV measure (CIV1), that utilizes a narrow range of option contracts, consistently generates the most accurate forecasts compared to all other alternatives. In Mincer-Zarnowitz regressions, CIV1 achieves the highest R^2, whilst cncompassing regression tests show that CIV1 subsumes the information contained in ATMIV, MFIV, OVX and HAR forecasts. Furthermore, under either a symmetric (MSE) or an asymmetric (QLIKE) loss function, CIV1-based forecasts deliver the lowest forecast errors. In terms of economic significance, incorporating the CIV1 into the HAR model leads to forecasts that generate a significantly higher realized utility, even when transaction costs are taken into account. All these findings remain intact for both our full sample (1996-2016) and a shorter sub-sample that starts when the OVX index was first disseminated (2007-2016).

Our results also provide valuable insights regarding the information content of the OVX index. In particular, we find that the OVX-based forecasts perform rather poorly. Regression-based tests show that the OVX is encompassed by CIV1, whilst OVX-based forecasts are typically the least accurate according to either statistical or economic loss functions. To the best of our knowledge, this is the first time the reliability of the OVX measure has been scrutinized, so the concerns we raise have direct implications for practitioners who often rely on the CBOE's volatility indices.

Overall, this paper contributes to the academic literature that assesses the forward-looking information embedded in the prices of crude-oil options. Given that measuring crude-oil risk is of paramount importance for a variety of economic agents, our empirical study is of value for policy-makers and investors alike.

References

Agnolucci, Paolo, 2009, Volatility in crude-oil futures: A comparison of the predictive ability of GARCH and implied volatility models, *Energy Economics* 31, 316 – 321.

Andersen, Torben G., T. Bollerslev, Francis X. Diebold, and Paul Labys, 2000, Great realizations, *RISK* 13, 105–108.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 96, 42–55.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* 71, 579–625.

Andersen, Torben G., and Oleg Bondarenko, 2007, Construction and interpretation of model-free implied volatility, in Israel Nelken, ed., *Volatility As An Asset Class*, 141–181 (Risk Publications).

Andersen, Torben G., Oleg Bondarenko, and Maria T. Gonzalez-Perez, 2015, Exploring return dynamics via corridor implied volatility, *The Review of Financial Studies* 28, 2902–2945.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov, 2017, Short-term market risks implied by weekly options, *The Journal of Finance* 72, 1335–1386.

Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard, 2008, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, *Econometrica* 76, 1481–1536.

Barndorff-Nielsen, Ole E., and Neil Shephard, 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 253–280.

Barone-Adesi, Giovanni, and Robert E. Whaley, 1987, Efficient analytic approximation of American option values, *The Journal of Finance* 42, 301–320.

Beckers, Stan, 1981, Standard deviations implied in option prices as predictors of future stock price variability, *Journal of Banking and Finance* 5, 363 – 381.

Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307 – 327.

Bollerslev, Tim, Benjamin Hood, John Huss, and Lasse Heje Pedersen, 2018, Risk everywhere: Modeling and managing volatility, *Review of Financial Studies* forthcoming.

Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg, 2016, Exploiting the errors: A simple approach for improved volatility forecasting, *Journal of Econometrics* 192, 1 – 18.

Bookstaber, Richard M., and James B. McDonald, 1987, A general distribution for describing security price returns, *Journal of Business* 60, 401–424.

Britten-Jones, Mark, and Anthony Neuberger, 2000, Option prices, implied price processes, and stochastic volatility, *The Journal of Finance* 55, 839–866.

Caldeira, Joo F., Guilherme V. Moura, Francisco J. Nogales, and Andr A. P. Santos, 2017, Combining multivariate volatility forecasts: An economic-based approach, *Journal of Financial Econometrics* 15, 247–285.

Carr, Peter, Roger Lee, and Liuren Wu, 2012, Variance swaps on time-changed Lévy processes, *Finance and Stochastics* 16, 335–355.

Carr, Peter, and Dilip Madan, 1998, Towards a theory of volatility trading, *Volatility: New estimation techniques for pricing derivatives* 417–427.

Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *The Review of Financial Studies* 22, 1311–1341.

Chiras, Donald P., and Steven Manaster, 1978, The information content of option prices and a test of market efficiency, *Journal of Financial Economics* 6, 213 – 234.

Christoffersen, Peter, Kris Jacobs, and Bo Young Chang, 2013, Chapter 10 - forecasting with option-implied information, in Graham Elliott, and Allan Timmermann, eds., *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, 581 – 656 (Elsevier).

Corsi, Fulvio, 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.

Day, Theodore E., and Craig M. Lewis, 1993, Forecasting futures market volatility, *Journal of Derivatives* 1, 33–50.

Demeterfi, Kresimir, Emanuel Derman, Michael Kamal, and Joseph Zou, 1999, A guide to volatility and variance swaps, *The Journal of Derivatives* 6, 9–32.

Diebold, Francis X., and Robert S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

Fung, HungGay, ChinJen Lie, and Abel Moreno, 1990, The forecasting performance of the implied standard deviation in currency options, *Managerial Finance* 16, 24–29.

Gemmill, Gordon, 1986, The forecasting performance of stock options on the London traded options market, *Journal of Business Finance and Accounting* 13, 535–546.

Griffin, John M., and Amin Shams, 2018, Manipulation in the VIX?, *The Review of Financial Studies* 31, 1377–1417.

Haugom, Erik, Henrik Langeland, Peter Molnár, and Sjur Westgaard, 2014, Forecasting volatility of the u.s. oil market, *Journal of Banking and Finance* 47, 1 – 14.

Jacod, Jean, Yingying Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter, 2009, Microstructure noise in the continuous case: The pre-averaging approach, *Stochastic Processes and their Applications* 119, 2249 – 2276.

Jiang, George J, and Yisong S Tian, 2005, The model-free implied volatility and its information content, *Review of Financial Studies* 18, 1305–1342.

Kang, Sang Hoon, Sang-Mok Kang, and Seong-Min Yoon, 2009, Forecasting volatility of crude oil markets, *Energy Economics* 31, 119 – 125.

Latané, Henry A., and Richard J. Rendleman, 1976, Standard deviations of stock price ratios implied in option prices, *The Journal of Finance* 31, 369–381.

Liu, Lily Y., Andrew J. Patton, and Kevin Sheppard, 2015, Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes, *Journal of Econometrics* 187, 293 – 311.

Locke, Peter R., and P. C. Venkatesh, 1997, Futures market transaction costs, *Journal of Futures Markets* 17, 229–245.

Martens, Martin, and Jason Zein, 2004, Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility, *Journal of Futures Markets* 24, 1005–1028.

Meddahi, Nour, 2002, A theoretical comparison between integrated and realized volatility, *Journal of Applied Econometrics* 17, 479–508.

Melick, William R., and Charles P. Thomas, 1997, Recovering an asset's implied pdf from option prices: An application to crude oil during the gulf crisis, *Journal of Financial and Quantitative Analysis* 32, 91115.

Mincer, J., and V. Zarnowitz, 1969, The evaluation of economic forecasts, in J. Mincer, ed., *Economic forecasts and expectations*, 3–46 (Cambridge: National Bureau of Economic Research).

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Patton, Andrew J., 2011, Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics* 160, 246 – 256.

Patton, Andrew J., and Kevin Sheppard, 2009, Evaluating volatility and correlation forecasts, in Thomas Mikosch, Jens-Peter Kreiß, Richard A. Davis, and Torben Gustav Andersen, eds., *Handbook of financial time series*, 801–838 (Springer).

Patton, Andrew J., and Kevin Sheppard, 2015, Good volatility, bad volatility: Signed jumps and the persistence of volatility, *Review of Economics and Statistics* 97, 683–697.

Podolskij, Mark, and Mathias Vetter, 2009, Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps, *Bernoulli* 15, 634–658.

Prokopczuk, Marcel, and Chardin Wese Simen, 2014, The importance of the volatility risk premium for volatility forecasting, *Journal of Banking and Finance* 40, 303–320.

Prokopczuk, Marcel, Lazaros Symeonidis, and Simen Chardin Wese, 2016, Do jumps matter for volatility forecasting? Evidence from energy markets, *Journal of Futures Markets* 36, 758–792.

Sadorsky, Perry, 2006, Modeling and forecasting petroleum futures volatility, *Energy Economics* 28, 467 – 488.

Sévi, Benoît, 2014, Forecasting the volatility of crude oil futures using intraday data, *European Journal of Operational Research* 235, 643 – 659.

Szakmary, Andrew, Evren Ors, Jin Kyoung Kim, and Wallace N Davidson, 2003, The predictive power of implied volatility: Evidence from 35 futures markets, *Journal of Banking and Finance* 27, 2151 – 2175.

Taylor, Stephen J., 2005, *Asset price dynamics, volatility, and prediction* (Princeton university press).

Taylor, Stephen J, Pradeep K Yadav, and Yuanyuan Zhang, 2010, The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks, *Journal of Banking and Finance* 34, 871–881.

Trippi, R., 1977, A test of option market efficiency using a random-walk valuation model, *Journal of Economics and Business* 29, 93–98.

Wang, Yudong, Li Liu, Feng Ma, and Chongfeng Wu, 2016, What the investors need to know about forecasting oil futures return volatility, *Energy Economics* 57, 128 – 139.

Zhang, Lan, Per A. Mykland, and Yacine At-Sahalia, 2005, A tale of two time scales, *Journal of the American Statistical Association* 100, 1394–1411.

FIGURE 1. Corridor Volatility Expectations
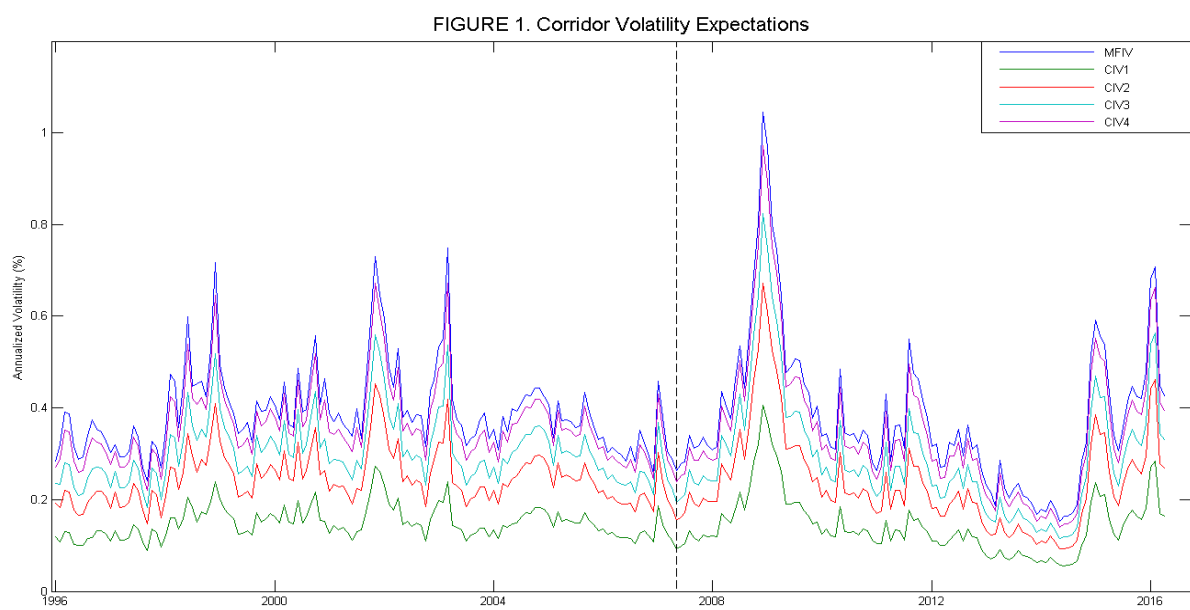
Figure 1: Time-series plots of different corridor implied volatility (CIV) measures for the 1996-2016 period. All expectations represent annualized standard deviations corresponding to a monthly horizon. The dashed vertical line indicates the beginning of the OVX sample.

Table 1: Summary statistics of our two option datasets. We report the number of unique option cross-sections (Obs) used to construct our non-overlapping, option-implied, monthly volatility forecasts, as well as the mean, minimum and maximum number of option contracts used in our estimations.

| | | | All Options | | | Calls (Puts) | | |
	Period	Obs	Mean	Min	Max	Mean	Min	Max
Full Sample	1996-2016	280	61.1	7	213	33.8 (27.3)	3 (3)	184 (94)
OVX Sample	2007-2016	108	112.5	54	213	63.2 (49.3)	29 (16)	184 (94)

Table 2: Descriptive statistics of option-implied volatility estimates. The first four columns show the unconditional sample moments for each measure. Sample autocorrelations for 1, 6 and 12 lags are denoted as ρ_1, ρ_6 and ρ_{12}, respectively. The last two columns show the correlation of each measure with the model-free (ρ_{MF}) and Black-Scholes (ρ_{BS}) implied volatilities.

	Mean	StDev	Skew	Kurt	ρ_1	ρ_6	ρ_{12}	ρ_{MF}	ρ_{BS}
CIV1	0.146	0.051	1.531	7.514	0.856	0.33	0.125	0.981	0.987
CIV2	0.241	0.084	1.564	7.65	0.85	0.319	0.116	0.988	0.993
CIV3	0.298	0.102	1.572	7.68	0.845	0.314	0.114	0.993	0.995
CIV4	0.356	0.12	1.583	7.699	0.837	0.306	0.112	0.999	0.995
MFIV	0.386	0.129	1.591	7.711	0.829	0.301	0.112	1	0.993
ATMIV	0.365	0.124	1.549	7.509	0.836	0.306	0.115	0.993	1

Table 3: Mincer-Zarnowitz regression results for the HAR, MFIV, ATMIV, CIV1, CIV2, CIV3 and CIV4 forecasts. This table reports parameter estimates and values of R^2 for the Mincer-Zarnowitz regressions in Equation (7). Panel A presents results for the full sample, which excludes the OVX forecasts. Panel B presents results for the OVX sample, which includes the OVX forecasts. Newey-West standard errors were estimated. t-statistics are reported in parentheses. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.

Forecast	Panel A: Full Sample			Panel B: OVX Sample		
	β_0	β_1	R^2	β_0	β_1	R^2
HAR	7.499E-05***	0.785***	0.700	4.159E-02**	0.633***	0.368
	(3.749)	(14.197)		(2.055)	(4.549)	
MFIV	-9.079E-06	0.776***	0.747	3.099E-03	0.877***	0.720
	(<0.001)	(11.458)		(0.308)	(9.898)	
ATMIV	-4.743E-06	0.857***	0.757	5.837E-03	0.964***	0.728
	(<0.001)	(12.049)		(0.607)	(9.801)	
CIV1	2.746E-06	5.141***	0.766	1.197E-02	5.576***	0.737
	(0.127)	(12.278)		(1.272)	(9.695)	
CIV2	1.351E-06	1.907***	0.766	9.836E-03	2.109***	0.734
	(0.064)	(12.545)		(1.031)	(9.722)	
CIV3	-1.839E-06	1.268***	0.764	8.059E-03	1.411***	0.731
	(<0.001)	(12.410)		(0.836)	(9.783)	
CIV4	-6.678E-06	0.903***	0.756	5.187E-03	1.013***	0.725
	(<0.001)	(11.942)		(0.527)	(9.871)	
OVX				-8.312E-04	0.886***	0.725
				(<0.001)	(9.525)	

Table 4: Encompassing regression results. The table reports parameter estimates and values of the adjusted R^2 for the encompassing regressions in Equation (8). Estimates reported in the columns labeled β_1 are for the forecast listed in the column labeled *Forecast 1*. Estimates reported in the columns labeled β_2 are for the forecast listed in the column labeled *Forecast 2*. Panel A presents results for encompassing regressions involving comparisons between the HAR forecasts and each of the option-implied variances for the full sample, which excludes the OVX forecasts. Panel B presents results for encompassing regressions involving comparisons between the HAR forecasts and each of the option-implied variances for the OVX sample, which includes the OVX forecasts. Panel C presents results for encompassing regressions involving comparisons between the option-implied variances for the full sample. Panel D presents results for encompassing regressions involving comparisons between the option-implied variances for the OVX sample. Newey-West standard errors were estimated. *t*-statistics are reported in parentheses. *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level.

		Parameters				Parameters			
Forecast 1	Forecast 2	β_0	β_1	β_2	adj R^2	β_0	β_1	β_2	adj R^2
		Panel A: HAR vs. Option-Implied Forecasts (Full Sample)				Panel B: HAR vs. Option-Implied Forecasts (OVX Sample)			
HAR	MFIV	5.360E-06 (0.279)	0.284** (2.141)	0.530*** (4.121)	0.763	4.867E-03 (0.492)	-0.390 (<0.001)	1.210*** (4.595)	0.756
HAR	ATMIV	6.529E-06 (0.342)	0.247* (1.813)	0.621*** (4.205)	0.768	8.965E-03 (0.981)	-0.413 (<0.001)	1.350*** (4.589)	0.769
HAR	CIV1	1.045E-05 (0.541)	0.218 (1.554)	3.898*** (4.138)	0.774	1.807E-02** (2.047)	-0.434 (<0.001)	7.909*** (4.815)	0.783
HAR	CIV2	9.438E-06 (0.490)	0.214 (1.522)	1.453*** (4.180)	0.774	1.487E-02* (1.677)	-0.428 (<0.001)	2.982*** (4.774)	0.778
HAR	CIV3	7.475E-06 (0.387)	0.223 (1.611)	0.953*** (4.210)	0.772	1.224E-02 (1.365)	-0.423 (<0.001)	1.989*** (4.755)	0.774
HAR	CIV4	5.372E-06 (0.278)	0.253* (1.875)	0.648*** (4.199)	0.767	7.958E-03 (0.849)	-0.407 (<0.001)	1.413*** (4.677)	0.765
HAR	OVX					-1.157E-04 (<0.001)	-0.364 (<0.001)	1.195*** (4.598)	0.757
		Panel C: Option-Implied Forecasts (Full Sample)				Panel D: Option-Implied Forecasts (OVX Sample)			
MFIV	ATMIV	6.383E-07 (0.031)	-0.549 (<0.001)	1.457** (2.099)	0.758	1.267E-02 (0.963)	-1.750 (<0.001)	2.875 (1.426)	0.728
MFIV	CIV1	2.848E-06 (0.145)	-0.004 (<0.001)	5.166** (2.546)	0.765	2.317E-02 (1.593)	-0.945 (<0.001)	11.491** (2.032)	0.738
MFIV	CIV2	7.549E-06 (0.393)	-0.279 (<0.001)	2.579** (2.464)	0.766	2.094E-02 (1.506)	-1.229 (<0.001)	5.028* (1.875)	0.735
MFIV	CIV3	8.505E-06 (0.441)	-0.647 (<0.001)	2.309** (2.397)	0.766	2.055E-02 (1.495)	-1.856 (<0.001)	4.368* (1.848)	0.735
MFIV	CIV4	8.481E-06 (0.435)	-2.501 (<0.001)	3.794** (2.265)	0.767	1.967E-02 (1.466)	-5.038 (<0.001)	6.806* (1.798)	0.735
ATMIV	CIV1	2.314E-06 (0.115)	0.031 (0.087)	4.956** (2.019)	0.765	2.623E-02* (1.928)	-2.000 (<0.001)	17.055** (2.040)	0.740
ATMIV	CIV2	5.580E-06 (0.289)	-0.394 (<0.001)	2.775** (1.995)	0.766	2.222E-02* (1.969)	-2.681 (<0.001)	7.948* (1.773)	0.735
ATMIV	CIV3	2.377E-06 (0.120)	-0.734 (<0.001)	2.348** (1.980)	0.764	1.696E-02** (2.002)	-3.449 (<0.001)	6.447 (1.535)	0.731
ATMIV	CIV4	-5.149E-06 (<0.001)	0.726 (0.874)	0.138 (0.159)	0.756	7.344E-03 (0.640)	2.648 (0.708)	-1.773 (<0.001)	0.724
CIV1	CIV2	2.049E-06 (0.097)	3.130 (0.630)	0.747 (0.430)	0.765	2.647E-02* (1.680)	39.351* (1.704)	-12.809 (<0.001)	0.742
CIV1	CIV3	1.901E-06 (0.093)	4.436 (1.348)	0.175 (0.239)	0.765	2.481E-02 (1.632)	21.911* (1.709)	-4.153 (<0.001)	0.739
CIV1	CIV4	2.399E-06 (0.120)	5.018** (2.123)	0.022 (0.061)	0.765	2.386E-02 (1.604)	14.032* (1.865)	-1.552 (<0.001)	0.738
CIV2	CIV3	6.274E-06 (0.325)	4.282 (1.332)	-1.582 (<0.001)	0.766	2.109E-02 (1.493)	14.244 (1.362)	-8.132 (<0.001)	0.734
CIV2	CIV4	7.217E-06 (0.375)	2.868** (2.045)	-0.460 (<0.001)	0.766	2.130E-02 (1.509)	6.628* (1.661)	-2.185 (<0.001)	0.735
CIV3	CIV4	8.374E-06 (0.434)	3.090** (1.988)	-1.305 (<0.001)	0.766	2.110E-02 (1.506)	6.925 (1.641)	-3.976 (<0.001)	0.736
OVX	MFIV					-9.208E-04 (<0.001)	0.911 (0.768)	-0.025 (<0.001)	0.719
OVX	ATMIV					3.244E-03 (0.286)	0.308 (0.459)	0.631 (0.866)	0.723
OVX	CIV1					1.051E-02 (0.961)	0.089 (0.200)	5.025* (1.676)	0.732
OVX	CIV2					7.877E-03 (0.735)	0.144 (0.286)	1.772 (1.412)	0.728
OVX	CIV3					6.003E-03 (0.565)	0.182 (0.322)	1.125 (1.212)	0.726
OVX	CIV4					1.999E-03 (0.182)	0.433 (0.555)	0.521 (0.587)	0.721

Table 5: MSE and QLIKE values for bias-corrected and HAR-based forecasts. Panel A reports the values of MSE and QLIKE for the full sample, which excludes the BC-OVX and HAR-OVX forecasts. Panel B reports values of MSE and QLIKE for the OVX sample, which includes the BC-OVX and HAR-OVX forecasts. Boldface is used to highlight minimum values. The results of Diebold-Mariano tests, based on Newey-West standard errors, are also summarised. In Panel A the HAR forecasts are the benchmark against which all other forecasts are compared. In panel B the HAR and HAR-OVX forecasts are used as benchmarks against which all other forecasts are compared. Where the HAR forecasts are a benchmark, *** indicates significance at the 1% level; ** indicates significance at the 5% level; * indicates significance at the 10% level. Where the HAR-OVX forecasts are a benchmark, † † † indicates significance at the 1% level; †† indicates significance at the 5% level; † indicates significance at the 10% level.

| | Panel A: Full Sample | | Panel B: OVX Sample | |
Forecast	MSE	QLIKE	MSE	QLIKE
HAR	2.721	-1.634	9.760	-1.504
BC-MFIV	2.562	-1.637	4.916††	-1.504
BC-ATMIV	2.427	-1.639	4.626*,††	$-1.519^{††}$
BC-CIV1	**2.229**	-1.636	**4.279***,††	$-1.528^{††}$
BC-CIV2	2.268	-1.638	4.413*,††	$-1.527^{††}$
BC-CIV3	2.314	-1.639	4.517*,††	$-1.523^{††}$
BC-CIV4	2.430	-1.638	4.729*,††	$-1.514^{†}$
BC-OVX			5.010††	-1.445
HAR-MFIV	2.721	-1.642^{*}	6.447†††	$-1.519^{*,††}$
HAR-ATMIV	2.600	-1.644^{**}	5.583††	$-1.526^{**,†††}$
HAR-CIV1	2.393	**-1.646****	4.440†	**-1.533*****,†††
HAR-CIV2	2.460	-1.645^{**}	4.900††	$-1.531^{***,†††}$
HAR-CIV3	2.518	-1.644^{**}	5.276††	$-1.528^{**,†††}$
HAR-CIV4	2.629	-1.643^{*}	5.943††	$-1.523^{**,†††}$
HAR-OVX			7.196	-1.494

Table 6: Realized utility and quantiles of the utility per unit wealth (UoW) for the bias-corrected and HAR-based forecasts. The realized utility is given in Equation (12) and the per period UoW is given in Equation (11). All values in the table are reported as percentages. In the columns reporting realized utility, boldface is used to highlight maximum values. Panel A reports the realized utility and quantiles of the UoW for forecasts calculated using the full sample. Panel B reports the realized utility and quantiles of the UoW for forecasts calculated using the OVX sample. We also report the results of Diebold-Mariano tests for differences between the realized utility of the HAR forecasts, which are the benchmark forecasts, and each of the remaining competing forecasts. Newey-West standard errors were used. ***, ** and * indicate a significant difference to the benchmark at the 1, 5 and 10% levels, respectively.

| | Panel A: Full Sample | | | | | | | Panel B: OVX Sample | | | | | | |
| | Realized | Quantiles | | | | | | Realized | Quantiles | | | | | |
Forecast	Utility	2.5%	5%	10%	25%	50%	75%	Utility	2.5%	5%	10%	25%	50%	75%
HAR	3.807	2.610	3.106	3.518	3.826	3.937	3.977	3.713	2.624	2.756	3.122	3.578	3.890	3.980
BC-MFIV	3.805	2.567	3.060	3.561	3.824	3.941	3.985	3.691	1.670	2.574	2.936	3.666	3.934	3.982
BC-ATMIV	3.811	2.661	3.169	3.527	3.815	3.944	3.988	3.725	1.805	2.613	3.266	3.680	3.940	3.996
BC-CIV1	3.804	2.294	3.174	3.568	3.800	3.937	3.989	3.748	2.037	2.826	3.287	3.705	3.931	3.989
BC-CIV2	3.808	2.528	3.171	3.580	3.818	3.942	3.990	3.744	1.920	2.739	3.258	3.706	3.939	3.996
BC-CIV3	3.809	2.605	3.163	3.563	3.818	3.942	3.989	3.735	1.842	2.686	3.268	3.689	3.939	3.997
BC-CIV4	3.809	2.576	3.121	3.565	3.817	3.943	3.989	3.713	1.695	2.604	3.192	3.663	3.938	3.991
BC-OVX								3.547	0.566	1.433	2.728	3.634	3.886	3.970
HAR-MFIV	3.822	2.684	3.145	3.509	3.815	3.951	3.988	3.735**	2.476	2.920	3.057	3.701	3.935	3.992
HAR-ATMIV	3.825*	2.706	3.124	3.536	3.828	3.953	3.989	3.747***	2.505	2.755	3.185	3.711	3.926	3.994
HAR-CIV1	**3.829****	2.702	3.222	3.575	3.828	3.952	3.987	**3.765****	2.503	2.789	3.248	3.755	3.927	3.994
HAR-CIV2	3.826**	2.714	3.163	3.570	3.820	3.952	3.988	3.757***	2.413	2.776	3.231	3.701	3.929	3.994
HAR-CIV3	3.825*	2.713	3.128	3.552	3.811	3.953	3.988	3.752***	2.442	2.772	3.201	3.699	3.931	3.994
HAR-CIV4	3.823	2.703	3.083	3.545	3.815	3.954	3.988	3.742***	2.481	2.823	3.123	3.705	3.932	3.994
HAR-OVX								3.696	2.502	2.759	3.076	3.622	3.835	3.984

32

Table 7: Net realized utility and transaction costs (TC) for the bias-corrected and HAR-based forecasts. The net realized utility is given in Equation (14) and TC is given in Equation (13). Net realized utility and TC are presented for $c = 0.15\%$ and $c = 0.033\%$. All values in the table are reported as percentages. In the columns reporting net realized utility, boldface is used to highlight maximum values. Panel A reports the net realized utility and TC for forecasts calculated using the full sample. Panel B reports the net realized utility and TC for forecasts calculated using the OVX sample. We also report the results of Diebold-Mariano tests for differences between the net realized utility of the HAR forecasts, which are the benchmark forecasts, and each of the remaining competing forecasts. Analogous Diebold-Mariano tests are conducted for TC, where the TC of the HAR forecasts are the benchmark. Newey-West standard errors were used in all Diebold-Mariano tests. ***, ** and * indicate a significant difference to the benchmark at the 1, 5 and 10% levels, respectively.

| | Panel A: Full Sample | | | | Panel B: OVX Sample | | | |
| | c=0.15% | | c=0.033% | | c=0.15% | | c=0.033% | |
Forecast	Net Realized Utility	Transaction Costs	Net Realized Utility	Transaction Costs	Net Realized Utility	Transaction Costs	Net Realized Utility	Transaction Costs
HAR	3.793	0.026	3.804	0.0056	3.697	0.0079	3.710	0.0017
BC-MFIV	3.791	0.026	3.802	0.0054	3.674	0.0081	3.687	0.0018
BC-ATMIV	3.797	0.025	3.808	0.0055	3.710	0.0073	3.721	0.0016
BC-CIV1	3.791	0.025	3.801	0.0054	3.734	0.0070*	3.745	0.0015*
BC-CIV2	3.795	0.024	3.805	0.0054	3.729	0.0069*	3.740	0.0015*
BC-CIV3	3.796	0.025	3.807	0.0055	3.720	0.0071	3.732	0.0016
BC-CIV4	3.795	0.025	3.806	0.0056	3.698	0.0076	3.710	0.0017
BC-OVX					3.524	0.0109	3.542	0.0024
HAR-MFIV	3.808	0.026	3.819	0.0057	3.718**	0.0080	3.731**	0.0018
HAR-ATMIV	3.811*	0.026	3.822*	0.0057	3.731***	0.0079	3.744***	0.0017
HAR-CIV1	**3.815****	0.026	**3.826****	0.0057	**3.749****	0.0076	**3.761****	0.0017
HAR-CIV2	3.812**	0.026	3.823**	0.0057	3.741***	0.0076	3.754***	0.0017
HAR-CIV3	3.811*	0.026	3.822*	0.0057	3.735***	0.0078	3.748***	0.0017
HAR-CIV4	3.809	0.026	3.820	0.0057	3.725***	0.0079	3.738***	0.0017
HAR-OVX					3.679	0.0080	3.692	0.0018

Table 8: MSE, QLIKE and realized utility for the bias-corrected and HAR-based forecasts generated using alternative in-sample rolling-windows for parameter estimation. To generate forecasts from the bias-correction model and the augmented and non-augmented HAR models, parameters were estimated using in-sample rolling-windows of either 66, 72, 78, 84, 90 or 96 observations (each observation corresponds to a 22-day non-overlapping interval). Panel A reports the MSE for the forecasts generated using the alternative in-sample rolling-windows, whilst Panel B and C report the QLIKE and realized utility (reported as a percentage), respectively. Boldface is used to highlight minimum values of MSE and QLIKE and maximum values of realized utility. All forecast evaluations were for the full sample. We also report the results of Diebold-Mariano tests for differences between the MSE, QLIKE or realized utility of the HAR forecasts, which are the benchmark forecasts, and the MSE, QLIKE and realized utility, respectively, of each of the remaining competing forecasts. Newey-West standard errors were used in all Diebold-Mariano tests. ***, ** and * indicate a significant difference to the benchmark at the 1, 5 and 10% levels, respectively.

Models	In-Sample Estimation Window					
	66	72	78	84	90	96
	Panel A: MSE					
HAR	3.0302	2.5269	2.5764	2.4522	2.4578	2.5943
BC-MFIV	2.6303	2.3433	2.3651	2.3652	2.3069	2.4333
BC-ATMIV	2.4810*	2.2502	2.2779	2.2733	2.1977	2.3227
BC-CIV1	**2.2687****	**2.1639**	**2.2097**	**2.1940**	**2.0902**	**2.2161**
BC-CIV2	2.3092**	2.1734	2.2143	2.2012	2.1096	2.2343
BC-CIV3	2.3584**	2.1956	2.2317	2.2235	2.1415	2.2655
BC-CIV4	2.4870*	2.2640	2.2917	2.2897	2.2217	2.3462
HAR-MFIV	2.9416	2.5111	2.5841	2.3901	2.4332	2.5467
HAR-ATMIV	2.8332	2.4364	2.5242	2.2903	2.3060	2.4182
HAR-CIV1	2.6174*	2.2821	2.3790	2.2247	2.1983	2.3246
HAR-CIV2	2.6796	2.3325	2.4202	2.2460	2.2262	2.3511
HAR-CIV3	2.7357	2.3743	2.4574	2.2752	2.2676	2.3904
HAR-CIV4	2.8464	2.4513	2.5278	2.3373	2.3553	2.4732
	Panel B: QLIKE					
HAR	-1.6197	-1.6425	-1.6583	-1.6604	-1.6763	-1.6631
BC-MFIV	-1.6251	-1.6319	-1.6475	-1.6480*	-1.6588**	-1.6431**
BC-ATMIV	-1.6284	-1.6367	-1.6517	-1.6510	-1.6622*	-1.6461*
BC-CIV1	-1.6264	-1.6345	-1.6464	-1.6453	-1.6571*	-1.6389**
BC-CIV2	-1.6280	-1.6364	-1.6493	-1.6486	-1.6606*	-1.6436*
BC-CIV3	-1.6283	-1.6363	-1.6499	-1.6494	-1.6612*	-1.6445*
BC-CIV4	-1.6273	-1.6345	-1.6492	-1.6492	-1.6604**	-1.6442**
HAR-MFIV	-1.6258	-1.6440	-1.6600	-1.6621	-1.6764	-1.6628
HAR-ATMIV	-1.6284*	-1.6466	-1.6615	-1.6635	-1.6781	-1.6641
HAR-CIV1	**-1.6316****	**-1.6493****	**-1.6637***	**-1.6644***	**-1.6795**	**-1.6645**
HAR-CIV2	-1.6301**	-1.6480*	-1.6628	-1.6639*	-1.6790	-1.6643
HAR-CIV3	-1.6290*	-1.6471	-1.6622	-1.6636	-1.6785	-1.6640
HAR-CIV4	-1.6273	-1.6455	-1.6610	-1.6629	-1.6774	-1.6633
	Panel C: Realized Utility					
HAR	3.7967	3.8241	3.8246	3.8289	3.8316	**3.8237**
BC-MFIV	3.8017	3.7899*	3.7892*	3.7891**	3.7786**	3.7645**
BC-ATMIV	3.8087	3.8005*	3.7981*	3.7951*	3.7855**	3.7704**
BC-CIV1	3.8029	3.7942*	3.7841*	3.7797*	3.7714**	3.7515**
BC-CIV2	3.8070	3.7991*	3.7915*	3.7881*	3.7804**	3.7631**
BC-CIV3	3.8079	3.7992*	3.7932*	3.7906*	3.7822**	3.7658**
BC-CIV4	3.8060	3.7954*	3.7923*	3.7910**	3.7814**	3.7662**
HAR-MFIV	3.8087	3.8253	3.8263	3.8309	3.8291	3.8199
HAR-ATMIV	3.8133	3.8296	3.8279	3.8320	3.8310	3.8210
HAR-CIV1	**3.8201****	**3.8353***	**3.8326**	**3.8333**	**3.8335**	**3.8210**
HAR-CIV2	3.8165*	3.8323	3.8305	3.8324	3.8325	3.8206
HAR-CIV3	3.8144	3.8305	3.8293	3.8321	3.8317	3.8204
HAR-CIV4	3.8113	3.8279	3.8278	3.8315	3.8303	3.8201

Table 9: MSE, QLIKE and realized utility for the bias-corrected and HAR-based forecasts evaluated over alternative out-of-sample periods. Forecasts were evaluated for out-of-sample periods starting in January of 2002, 2003, ..., 2012. The number of observations (each observation corresponds to a 22-day non-overlapping interval) in each of the alternative out-of-sample periods is reported in parentheses. Panel A reports the MSE for the forecasts evaluated over the alternative out-of-sample periods, whilst Panel B and C report the QLIKE and realized utility (reported as a percentage), respectively. Boldface is used to highlight minimum values of MSE and QLIKE and maximum values of realized utility. We also report the results of Diebold-Mariano tests for differences between the MSE, QLIKE or realized utility of the HAR forecasts, which are the benchmark forecasts, and the MSE, QLIKE and realized utility, respectively, of each of the remaining competing forecasts. Newey-West standard errors were used in all Diebold-Mariano tests. ***, ** and * indicate a significant difference to the benchmark at the 1, 5 and 10% levels, respectively.

Out-of-Sample Start Date
(No. of Out-of-Sample Observations)

Models	2002 (232)	2003 (220)	2004 (208)	2005 (196)	2006 (184)	2007 (172)	2008 (160)	2009 (148)	2010 (136)	2011 (124)	2012 (112)
Panel A: MSE											
HAR	2.7477	2.9174	3.0982	3.3510	3.6569	4.0213	4.4742	3.4550*	3.3163	3.8358*	4.0015*
BC-MFIV	2.6070	2.7477	2.9369	3.1738	3.4651	3.8119	4.2385	2.2690	2.2261	2.5815	2.5668
BC-ATMIV	2.4594	2.5867	2.7505	2.9714	3.2419	3.5635	3.9566	2.0663	2.0798*	2.4054*	2.4119
BC-CIV1	2.2504	2.3722	2.4895	2.6892	2.9234	3.2086	3.5559	1.9903	1.9379*	2.2330*	2.2288
BC-CIV2	2.2941	2.4166	2.5526	2.7569	3.0012	3.2954	3.6541	2.0027	1.9850*	2.2907*	2.2986
BC-CIV3	2.3435	2.4682	2.6185	2.8285	3.0822	3.3857	3.7569	2.0469	2.0315*	2.3474*	2.3539
BC-CIV4	2.4675	2.5992	2.7712	2.9945	3.2673	3.5920	3.9906	2.1590	2.1329*	2.4698	2.4670
HAR-MFIV	2.7553	2.9074	3.1097	3.3642	3.6756	4.0493	4.5071	3.0290	2.8181	3.2334	3.2459
HAR-ATMIV	2.6245	2.7690	2.9604	3.2027	3.4980	3.8517	4.2858	2.8181	2.5667	2.9428	2.9185
HAR-CIV1	2.4061	2.5393	2.7016	2.9240	3.1884	3.5065	3.8999	2.5735	2.2875	2.6105	2.5217
HAR-CIV2	2.4776	2.6127	2.7862	3.0152	3.2895	3.6190	4.0252	2.6697	2.3993	2.7445	2.6864
HAR-CIV3	2.5401	2.6777	2.8592	3.0939	3.3768	3.7164	4.1340	2.7573	2.5003	2.8632	2.8214
HAR-CIV4	2.6579	2.8026	2.9965	3.2421	3.5409	3.8993	4.3389	2.9142	2.6828	3.0763	3.0645
Panel B: QLIKE											
HAR	-1.6408	-1.6530	-1.6550	-1.6518	-1.6297	-1.5865	-1.5538	-1.6595	-1.7601	-1.7491	-1.8430
BC-MFIV	-1.6382	-1.6519	-1.6522	-1.6488	-1.6260	-1.5826	-1.5501	-1.6600	-1.7610	-1.7429	-1.8368
BC-ATMIV	-1.6412	-1.6558	-1.6568	-1.6538	-1.6316	-1.5889	-1.5582	-1.6690	-1.7705	-1.7550	-1.8509
BC-CIV1	-1.6380	-1.6516	-1.6538	-1.6505	-1.6291	-1.5865	-1.5564	-1.6693	-1.7707	-1.7566	-1.8603
BC-CIV2	-1.6399	-1.6538	-1.6557	-1.6526	-1.6310	-1.5885	-1.5583	-1.6710	-1.7726	-1.7585	-1.8589
BC-CIV3	-1.6404	-1.6544	-1.6558	-1.6528	-1.6309	-1.5883	-1.5577	-1.6699	-1.7715	-1.7567	-1.8554
BC-CIV4	-1.6399	-1.6538	-1.6545	-1.6514	-1.6290	-1.5860	-1.5544	-1.6653	-1.7667	-1.7502	-1.8463
HAR-MFIV	-1.6453	-1.6593	-1.6599	-1.6570	-1.6348	-1.5911	-1.5591	-1.6649	-1.7664	-1.7578	-1.8538
HAR-ATMIV	-1.6477*	-1.6620**	-1.6630*	-1.6601*	-1.6383	-1.5952	-1.5638	-1.6699	-1.7722	-1.7638	-1.8609
HAR-CIV1	-1.6490**	-1.6631**	-1.6647**	-1.6618**	-1.6407**	-1.5983**	-1.5672**	-1.6724**	-1.7774**	-1.7699**	-1.8686**
HAR-CIV2	-1.6479*	-1.6621**	-1.6636*	-1.6607*	-1.6394*	-1.5967*	-1.5654*	-1.6724*	-1.7751*	-1.7671*	-1.8649**
HAR-CIV3	-1.6471	-1.6614*	-1.6627*	-1.6598*	-1.6383	-1.5953	-1.5639	-1.6706	-1.7731	-1.7648*	-1.8622*
HAR-CIV4	-1.6461	-1.6603	-1.6612	-1.6583	-1.6364	-1.5930	-1.5613	-1.6675	-1.7694	-1.7609	-1.8574
Panel C: Realized Utility											
HAR	3.8242	3.8162	3.8100	3.7984	3.7849	3.7713	3.7520	3.7489	3.7210	3.7034	3.7470
BC-MFIV	3.8088	3.8039	3.7928	3.7800	3.7640	3.7496	3.7298	3.7342	3.7046	3.6676	3.7068
BC-ATMIV	3.8153	3.8122	3.8029	3.7907	3.7758	3.7633	3.7472	3.7535	3.7254	3.6940	3.7395
BC-CIV1	3.8073	3.8021	3.7953	3.7822	3.7685	3.7556	3.7407	3.7519	3.7234	3.6948	3.7624
BC-CIV2	3.8116	3.8070	3.7996	3.7871	3.7732	3.7606	3.7456	3.7563	3.7283	3.6997	3.7580
BC-CIV3	3.8130	3.8085	3.8001	3.7878	3.7734	3.7607	3.7448	3.7543	3.7263	3.6963	3.7496
BC-CIV4	3.8123	3.8077	3.7977	3.7852	3.7700	3.7565	3.7384	3.7453	3.7166	3.6831	3.7287
HAR-MFIV	3.8289	3.8248	3.8150	3.8035	3.7894	3.7752	3.7570	3.7529	3.7258	3.7117	3.7569
HAR-ATMIV	3.8334	3.8299*	3.8208	3.8095	3.7961	3.7828	3.7660	3.7622	3.7367	3.7226	3.7701
HAR-CIV1	3.8365*	3.8326**	3.8251*	3.8137*	3.8017*	3.7897*	3.7735**	3.7724*	3.7485**	3.7366**	3.7879**
HAR-CIV2	3.8337	3.8301*	3.8221	3.8107	3.7981	3.7856	3.7691	3.7672	3.7425	3.7293	3.7784
HAR-CIV3	3.8321	3.8286	3.8202	3.8088	3.7958	3.7829	3.7660	3.7636	3.7383	3.7246	3.7725
HAR-CIV4	3.8302	3.8265	3.8173	3.8058	3.7922	3.7785	3.7610	3.7577	3.7314	3.7172	3.7633

Table 10: MSE, QLIKE and realized utility for relative bias-corrected forecasts. The table reports the MSE, QLIKE and realized utility for option-implied variance forecasts generated using the alternative relative bias-correction procedure described in Equations (15) and (16). Results are reported for alternative window sizes, τ, used in applying the relative bias-correction. Results for the BC-CIV1 and HAR-CIV1 forecasts, which do not employ the relative bias-correction procedure, are also included as benchmarks. All forecast evaluations were conducted using the full sample. Panel A reports the MSE, Panel B the QLIKE and Panel C the realized utility (reported as a percentage) for the forecasts. Boldface is used to highlight minimum values of MSE and QLIKE and maximum values of realized utility. We also report the results of Diebold-Mariano tests for differences between the MSE, QLIKE and realized utility of the HAR-CIV1 forecasts, the benchmark forecasts, and the MSE, QLIKE and realized utility, respectively, of each of the competing relative bias-corrected forecasts. Newey-West standard errors were used in all Diebold-Mariano tests. ***, ** and * indicate a significant difference to the benchmark at the 1, 5 and 10% levels, respectively.

Models	$\tau = 6$	$\tau = 12$	$\tau = 18$	$\tau = 24$	$\tau = 30$	$\tau = 36$	$\tau = 42$	$\tau = 48$	$\tau = 54$	$\tau = 60$
					Panel A: MSE Loss Function					
BC-CIV1	2.2287	2.2287	2.2287	2.2287	2.2287	2.2287	2.2287	2.2287	2.2287	2.2287
HAR-CIV1	2.3934	2.3934	2.3934	2.3934	2.3934	2.3934	2.3934	2.3934	2.3934	2.3934
RBC-MFIV	2.1888	1.7208*	1.7172*	1.6662*	1.7041*	1.8882	2.0070	2.0328	2.1289	2.1252
RBC-ATMIV	2.0849	1.6638*	1.6583*	**1.6133***	1.6538*	1.8198	1.9272	1.9542	2.0477	2.0408
RBC-CIV1	2.1002	1.6604	1.6422*	1.6594	1.6451	1.8333	1.9133	1.9345	2.0092	1.9833
RBC-CIV2	**2.0703**	**1.6395***	**1.6210***	1.6230*	**1.6281***	**1.8048**	**1.8900**	**1.9143**	**2.0005**	**1.9776**
RBC-CIV3	2.0747	1.6419*	1.6265*	1.6136*	1.6295*	1.8081	1.9031	1.9295	2.0204	2.0017
RBC-CIV4	2.1198	1.6741*	1.6654*	1.6296*	1.6595*	1.8424	1.9528	1.9807	2.0757	2.0651
					Panel B: QLIKE Loss Function					
BC-CIV1	-1.6364*	-1.6364*	-1.6364*	-1.6364*	-1.6364*	-1.6364*	-1.6364*	-1.6364*	-1.6364*	-1.6364*
HAR-CIV1	**-1.6457**	**-1.6457**	**-1.6457**	**-1.6457**	**-1.6457**	**-1.6457**	**-1.6457**	**-1.6457**	**-1.6457**	**-1.6457**
RBC-MFIV	-1.6397	-1.6393	-1.6371	-1.6335	-1.6322*	-1.6269**	-1.6256**	-1.6236**	-1.6229**	-1.6233**
RBC-ATMIV	-1.6403	-1.6378	-1.6355	-1.6310*	-1.6286**	-1.6236**	-1.6227**	-1.6212***	-1.6209***	-1.6212***
RBC-CIV1	-1.6361	-1.6321	-1.6300	-1.6233**	-1.6203**	-1.6154***	-1.6152***	-1.6146***	-1.6146***	-1.6153***
RBC-CIV2	-1.6384	-1.6357	-1.6336	-1.6271*	-1.6243**	-1.6193***	-1.6189***	-1.6181***	-1.6180***	-1.6186***
RBC-CIV3	-1.6396	-1.6378	-1.6358	-1.6299*	-1.6274**	-1.6224**	-1.6218***	-1.6207***	-1.6205***	-1.6210**
RBC-CIV4	-1.6403	-1.6395	-1.6374	-1.6328*	-1.6310*	-1.6258**	-1.6248**	-1.6233**	-1.6228**	-1.6232**
					Panel C: Realized Utility Loss Function					
BC-CIV1	3.8041**	3.8041**	3.8041**	3.8041**	3.8041**	3.8041**	3.8041**	3.8041**	3.8041**	3.8041**
HAR-CIV1	**3.8286**	**3.8286**	**3.8286**	**3.8286**	**3.8286**	**3.8286**	**3.8286**	**3.8286**	**3.8286**	**3.8286**
RBC-MFIV	3.8088	3.8057	3.8000	3.7911*	3.7883**	3.7760**	3.7731***	3.7682***	3.7667***	3.7678***
RBC-ATMIV	3.8102	3.8022	3.7965*	3.7855**	3.7801**	3.7686***	3.7664***	3.7628***	3.7622***	3.7630***
RBC-CIV1	3.8015*	3.7887*	3.7833*	3.7674**	3.7603**	3.7490***	3.7485***	3.7466***	3.7466***	3.7479***
RBC-CIV2	3.8063	3.7971	3.7917*	3.7762**	3.7697**	3.7582***	3.7571***	3.7549***	3.7546***	3.7559***
RBC-CIV3	3.8090	3.8021	3.7967	3.7827**	3.7771**	3.7655***	3.7640***	3.7612***	3.7607***	3.7618***
RBC-CIV4	3.8103	3.8061	3.8007	3.7895**	3.7855**	3.7735***	3.7712***	3.7673***	3.7662***	3.7673***