NCER Working Paper Series

The role in index jumps and cojumps in forecasting stock index volatility: Evidence from the Dow Jones index

A E Clements
Y Liao

Working Paper #101
June 2014

The role of index jumps and cojumps in forecasting stock index volatility: Evidence from the Dow Jones index

A.E. Clements and Y. Liao
School of Economics and Finance, Queensland University of Technology, NCER.

Abstract

Modeling and forecasting realized volatility is of paramount importance. Previous studies have examined the role of both the continuous and jump components of volatility in forecasting. This paper considers how to use index level jumps and cojumps across index constituents for forecasting index level volatility. In combination with the magnitude of past index jumps, the intensity of both index jumps and cojumps are examined. Estimated jump intensity from a point process model is used within a forecasting regression framework. Even in the presence of the diffusive part of total volatility, and past jump size, intensity of both index and cojumps are found to significantly improve forecast accuracy. An important contribution is that information relating to the behaviour of underlying constituent stocks is useful for forecasting index level behaviour. Improvements in forecast performance are particularly apparent on the days when jumps or cojumps occur, or when markets are turbulent.

Keywords
Realized volatility; diffusion; jumps; point process; Hawkes process; forecasting

JEL Classification Numbers
C22, G00.

Corresponding author
Adam Clements
School of Economics and Finance
Queensland University of Technology
Brisbane, 4001
Qld, Australia

email `a.clements@qut.edu.au`

Acknowledgments The authors thank seminar participants at the University of Melbourne, and conference participants at the Frontiers in Financial Econometrics meetings at Princeton University for their thoughts and feedback on earlier version of this work. Any remaining errors or omissions are the responsibility of the authors.

1 Introduction

Understanding how the volatility of asset returns evolves is of paramount importance. Many financial applications such as risk management, portfolio allocation and derivative pricing utilize such information and as a consequence there has been a vast literature relating to the estimation and forecasting of volatility. Much of this literature has stemmed from the development of the univariate GARCH class of models, Engle (1982) and Bollerslev (1986). In recent years, this literature has benefited from the availability of high-frequency intraday data which has led to the development of Realized Volatility (RV) by Andersen, Bollerslev, Diebold, and Labys (2001, 2003) and improved measures of volatility at a daily frequency. As opposed to the GARCH approach of treating volatility as latent, RV provides an observable proxy upon which time-series models can be directly used to generate forecasts

A number of times series approaches have been applied to forecasting total RV. Among these are the Mixed Interval Data Sampling (MIDAS) of Ghysels, Santa-Clara, and Valkanov (2006) and long-memory ARMA models Oomen (2001), Andersen et al. (2003) and Koopman, Jungbacker, and Hol (2005). Estimates of RV converge to the total quadratic variation in the asset return process, which may be attributable to both continuous diffusion and discrete jump processes. Barndorff-Nielsen and Shephard (2006) and Andersen, Bollerslev, and Diebold (2007) develop methods for measuring the contribution to total quadratic variation from both sources. Andersen et al. (2007) propose a formal test for identifying when significant contributions from jump activity occur by employing the Heterogeneous AutoRegressive (HAR) model of Muller, Dacorogna, Dave, Olsen, Pictet, and Weizsacker (1997) and Corsi (2009) to forecast volatility using separate jump and diffusion components. They find that most of the predictability in volatility stems from the persistent diffusive component.

This paper extends the work of Andersen et al. (2007) for forecasting Dow Jones index volatility along two dimensions. First, while Andersen et al. (2007) only consider jumps at the index level, here the role of cojumps across the index constituents for forecasting index level volatility are also examined. There is now a growing literature that attempts to document and study simultaneous discrete jumps across many assets (cojumps). Progress on this front includes the development of tests for cojumps in a pair of asset returns Barndorff-Nielsen and Shephard (2003), Gobbi and Mancini (2007) and Jacod and Todorov (2009), as well as a cojump test developed by Bollerslev, Law, and Tauchen (2008) that is applicable to a large panel of high-frequency returns. Lahaye and Neely (2011) relate the cojumps extracted from a panel of U.S. stocks to U.S. macroeconomic releases and formally modeled how news surprises explain cojumps. Dungey and Hvozdyk (2011) studied the cojumps in spot and futures prices in high frequency U.S Treasury data, and find that an anticipated macroeconomic news announcement is sufficient to change the probability of observing cojumps. While these tests pave the way to identify the factors which drive the occurrence of cojumps in different markets or assets, there is little literature studying the impact of cojumps. This paper considers the role of cojumps across the Dow Jones constituents in the context of forecasting index level volatility. Second, estimated intensities of both index jumps and cojump activity are also used as exogenous variables for forecasting. In the HAR forecasting regressions of Andersen et al. (2007) only the lagged magnitude of index jumps were used. Here, estimates of the intensity of the arrival of both types of jumps are extracted from a point process model and are included in HAR forecasting regressions.

Overall, the results indicate that the intensities of both index jumps and cojumps significantly contribute to the predicability of volatility. Models utilising jump intensities were found to produce volatility forecasts that were significantly superior to current benchmark models. This result extends the finding of Andersen et al. (2007) where the history of the index jump size exhibited no significant forecasting power. The improvements in forecast performance are robust to periods of market turmoil such as the recent financial crisis.

The paper proceeds as follows. Section 2 describes the methods employed to estimate RV and identify index jumps and cojumps. Section 3 describes the Dow Jones stock index data used in the empirical analysis. Section 4 presents the econometric models employed while Section 5 outlines the empirical results. Section 6 provides concluding comments.

2 Theoretical Framework

In this section, the method for forming realized volatility from high frequency index returns, along with how the diffusion and jump components are extracted is discussed. The testing procedure for detecting cojumps in the panel of index constituents is also outlined.

2.1 Realized variance and jumps

Begin by assuming that the logarithm of the asset price within the active part of the trading day evolves in continuous time as a standard jump-diffusion process given by

$$dp(t) = u(t)dt + \sigma(t)dw(t) + \kappa(t)dq(t), \tag{1}$$

where $u(t)$ denotes the drift term that has continuous and locally bounded variation, $\sigma(t)$ is a strictly positive spot volatility process and $w(t)$ is a standard Brownian motion. The $\kappa(t)dq(t)$ term refers to a pure jump component, where $k(t)$ is the size of jump and $dq(t) = 1$ if there is a jump at time t (and 0 otherwise). The $j-th$ discrete-time return within day t is given by

$$r_{t_j} = p_{(t-1)+\frac{j}{M}} - p_{(t-1)+\frac{j-1}{M}}, \quad j = 1, 2, ...M, \tag{2}$$

where M refers to the number of intraday equally spaced returns over the trading day t,. As such, the daily return for the active part of the trading day equals $r_t = \sum_{j=1}^{M} r_{t_j}$.

As noted in Andersen and Bollerslev (1998), Andersen et al. (2003), and Barndorff-Nielsen and Shephard (2002), the quadratic variation of the process in equation (1) can be estimated by realized volatility (RV), which is defined as the sum of the intraday squared returns, i.e.

$$RV_t \equiv \sum_{j=1}^{M} r_{t_j}^2, \tag{3}$$

whereas the integrated variance of the process in equation (1) is typically estimated using realized bi-power variation (BV) defined by

$$BV_t \equiv \frac{\pi}{2}(\frac{M}{M-1}) \sum_{j=2}^{M} |r_{t_j}||r_{t_{j-1}}|. \tag{4}$$

Therefore, given appropriate regularity conditions, we have

$$\underset{M\to\infty}{plim}\, RV_t = \int_{t-1}^{t} \sigma^2(s)ds + \sum_{t<s<t+1} \kappa^2(s), \tag{5}$$

$$\underset{M\to\infty}{plim}\, BV_t(\Delta) = \int_{t-1}^{t} \sigma^2(s)ds, \tag{6}$$

Naturally, the difference between the two estimators can be used to estimate the contribution of jump component to the volatility.

However, the theoretical justification for equations (3) and (4) is based on the notion of increasingly finer sampled returns, or $M \to \infty$. Of course, any practical implementation with a finite fixed sampling frequency, or $M < \infty$, is invariably subject to measurement error, and hence it is desirable to treat small jumps as measurement error and only identify significantly large jumps. Barndorff-Nielsen and Shephard (2006) developed such a test, which is modified to account for microstructure noise and improve the finite sample performance following Huang and Tauchen (2005) as

$$Z_t \equiv \Delta^{-1/2} \times \frac{[RV_t - BV_t]RV_t^{-1}}{[(\mu_1^{-4} + 2\mu_1^{-2} - 5)max\{1, TQ_t BV_t^{-2}\}]^{1/2}}, \tag{7}$$

$\Delta = 1/M$ and TQ_{t+1} is the realized tripower quarticity measure, whose general expression is

$$TQ_t = \Delta^{-1}\mu_{4/3}^{-3} \sum_{j=3}^{M} |r_{t_j}|^{4/3}|r_{t_{j-1}}|^{4/3} r_{t_{j-2}}|^{4/3}, \tag{8}$$

and $\mu_{4/3} \equiv E(|Z|^{4/3}) = 2^{2/3} \cdot \Gamma(7/6) \cdot \Gamma(1/2)^{-1}$. The BV and TQ measures are generated based on staggered returns to remove the microstructure noise.

Significant jumps can be identified by the realizations of Z_t in excess of some critical value Φ_α,

$$J_{t,\alpha} \equiv I(Z_t > \Phi_\alpha) \cdot [RV_t - BV_t], \tag{9}$$

where $I(\cdot)$ denotes an indicator function, and α is the significance level. The continuous component is then defined as

$$C_{t,\alpha} \equiv I(Z_t \leq \Phi_\alpha) \cdot RV_t + I(Z_t \leq \Phi_\alpha) \cdot BV_t \tag{10}$$

to ensure that the continuous and jump components add to the realized volatility.

2.2 Identifying Cojumps in Stock Prices

The above section has discussed a test to detect the existence of jumps in asset prices. However, many financial problems involve multiple assets and thus there is a need for tests that can detect simultaneous jumps in many assets (cojump tests). The cojump test developed by Bollerslev et al. (2008) (henceforth denoted as the BLT test) that is applicable to a large panel of high-frequency returns fills this void. The intuition behind this test is that idiosyncratic noise in individual returns can hide the presence of a synchronous component. Therefore a test based on the cross products of returns in a panel avoids this problem while still being sensitive to systematic movements across all stocks. In this study, we use the BLT test to detect the occurrence of cojumps across a panel of stock prices.

Begin by assuming that each of the individual stocks, denoted by the index i, follow the stochastic process in equation (1). M intraday returns are observed for each stock with the $j-th$ within-day return on day t of an equi-weighted portfolio of n stocks given by

$$r_{EQW,t_j} = \frac{1}{n}\sum_{i=1}^{n} r_{i,t_j}.$$

The daily realized variance for this equi-weighted portfolio is given by

$$RV_{EQW,t} = \sum_{j=1}^{M}(\frac{1}{n}\sum_{i=1}^{n} r_{i,t_j})^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{M} r_{i,t_j}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\sum_{j=1}^{M} r_{i,t_j} r_{l,t_j}, \tag{7}$$

and when this is decomposed into its continuous and jump components, Bollerslev et al. (2008) show that most of the jump contribution to $RV_{EQW,t}$ originates from the covariation term (i.e from within $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1,l\neq i}^{n}\sum_{j=1}^{M} r_{i,t_j} r_{l,t_j}$) in equation (7) when n is large, while the effects of idiosyncratic jumps (i.e. that originate from the $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{M} r_{i,t_j}^2$ term) are essentially diversified away.

The notion that cojumps can cause the price of a portfolio to jump when n is large forms the basis for the cojump test. Their derivation was based on an equi-weighted portfolio, but their conclusion that most of the information about cojumps is contained in the covariation between stock returns is valid given any well-diversified portfolio.

The *zmcp* test statistic proposed by Bollerslev et al. (2008) is given by

$$z_{mcp,t_j} = \frac{mcp_{t_j} - \overline{mcp_t}}{s_{mcp,t}}, \; j = 1, 2, ...M, \qquad \text{where} \tag{8}$$

$$mcp_{t_j} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{l=i+1}^{n} r_{i,t_j} r_{l,t_j}, \qquad j = 1, 2, \ldots, M, \tag{8a}$$

$$\overline{mcp_t} = \frac{1}{M} \sum_{j=1}^{M} mcp_{t_j} = \frac{1}{M}[\frac{n}{n-1} RV_{ew,t} - \frac{1}{n(n-1)} \sum_{i=1}^{n} RV_{i,t}], \tag{8b}$$

$$s_{mcp,t} = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} (mcp_{t_j} - \overline{mcp_t})^2}, \tag{8c}$$

and it can be used as a test for common jumps because the jump (but not the continuous) component in the second term in $mcp_t = \frac{n}{n-1} RV_{ew,t} - \frac{1}{n(n-1)} \sum_{i=1}^{n} RV_{i,t} \approx RV_{ew,t} - \frac{1}{n(n-1)} \sum_{i=1}^{n} RV_{i,t}$ is diversified away as n grows large.

It is easy to rearrange the expression for mcp_{t_j} to obtain

$$mcp_{t_j} = \frac{n}{n-1} \left[(\frac{1}{n} \sum_{i=1}^{n} r_{i,t_j})^2 - \sum_{i=1}^{n} (\frac{1}{n})^2 r_{i,t_j}^2 \right] = \frac{n}{n-1} r_{ew,t_j}^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} r_{i,t_j}^2, \tag{9}$$

and from this we can see that it is possible to calculate the test statistic directly from the squared returns of the equally weighted portfolio and the individual stocks. We work with this alternative expression for mcp_{t_j} in the subsequent empirical analysis.

The z_{mcp,t_j} statistic is not well approximated by any of the standard distribution, but it is relatively straightforward to bootstrap its empirical distribution under the null hypothesis of no jumps, to find critical values that are relevant for a given application[1]. Meanwhile, the *zmcp* cojump test relies on three assumptions.

First, the studentization of the $mcp_{t,j}$ test statistic each day relies on an assumption that the location and scale of this statistic remains approximately constant over the day. This assumption may be at odds with the well-known U-shaped pattern associated with intra-day stock volatility. Bollerslev et al. (2008) suggest a way to deal with this issue by scaling the return for each stock over a certain time-interval by the reciprocal of the square root of the corresponding unconditional bi-power variation for that particular time-interval. This deflates

[1] Simulations conducted by Bollerslev et al. (2008) show that the distribution of the z_{mcp,t_j} statistic is centered to the left of zero and has a very strong right skew.

returns near the beginning and the end of the day while inflating returns in the middle of the day. This approach will be taken in the subsequent empirical analysis when detecting cojumps. Second, the $mcp_{t,j}$ realizations are assumed to be serially uncorrelated, making it appropriate to simply standardized each of the within-day mcp statistics by using the corresponding daily sample standard deviation $s_{mcp,t}$. Third, it is important to note that the sample mean used in the $zmcp$ test statistic incorporates the cojump contribution relating to each day and although the contribution of a few jumps on a day might be negligible, the contribution of several cojumps is unlikely to be negligible and then relatively large intra-day $mcp_{t,j}$ realizations might be masked by the correspondingly large sample mean $\overline{mcp_t}$. Therefore, the test implies an assumption that cojumps occur rarely, particularly no more than one cojump over a day. This assumption finds empirical support in Bollerslev et al. (2008) along with the current sample of 30 Dow Jones constituents.

3 Data and Summary Statistics

Data relating to the Dow Jones Industrial Average for the period 3 January 2000 to 25 June 2012 (3113 trading days) is used in this study. Data at both an index and constituent stock level is employed. Index level data is represented by prices for Dow Jones index futures contracts. Futures prices sampled at a 5-minute frequency were gathered to construct the volatility estimates described in Section 2.1, including RV, BPV and jumps in the index returns. Split and dividend adjusted prices for the 30 constituents of the Dow Jones Index were also sampled at a 5-minute interval to identify cojumps across the constituent stocks using the method described in Section 2.2. While the individual constituents are not reported here, periodic changes in the membership of the index are taken into account. Both index and constituent level data were collected from Thomson Reuters Tick History.

Figure 1 shows plots of RV, jumps in index returns and the occurrence of cojumps across constituent stocks. [2] It is clear that RV and jumps in the index returns follow a broadly similar pattern and are dominated by the period of 2000-2003 and the Global Financial Crisis (GFC) period of 2008-2010. RV exhibits the familiar pattern of higher volatility during 2000-2003, and reached historically high levels during the GFC. While a number of smaller jumps in index returns occur during the earlier part of the sample, jumps occur more frequently and exhibit a great deal of clustering during and after the GFC. Overall, it appears as though there is a positive relationship between the level of volatility and jump activity. The lower panel of Figure 1 shows that while the occurrence of cojumps are more evenly spread through the sample, there is clear evidence of clustering. Given the contrast between the index jumps and cojumps, they

[2]A plot of continuous component of index volatility is not shown as there is little difference in comparison RV.

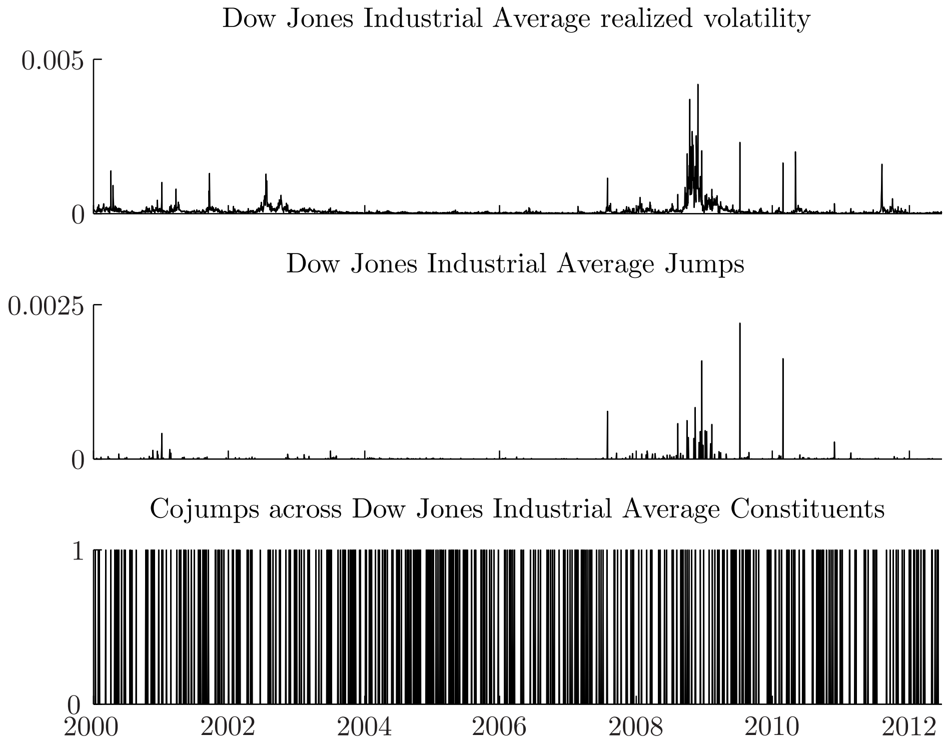

Figure 1: Dow Jones Industrial Average realized volatility (top panel), index jumps (middle panel) and cojumps (lower panel).

may contain distinctly different information for forecasting volatility. Table 1 reports various summary statistics for each of the data series. One clear conclusion is that all series exhibit varying degrees of persistence as indicated by large Ljung-Box test statistics (at a lag of 10) reported in the final row. Unconditionally, index jumps occur with an intensity of 0.075 and account for 6.5% of the mean of RV (5.8×10^{-6} relative to 9.01×10^{-5}). Consistent with previous studies, the logarithm of RV is much closer to being normally distributed than raw RV. Relative to index jumps cojumps occur slightly more frequently at an unconditional intensity of 0.131.

4 Econometric models

4.1 A Hawkes model for jump intensity

To model the intensity of jumps and cojumps, their occurrence is viewed as a point process. Such an approach is common when dealing with events in financial markets such as the arrival of trades or quotes, for an overview of the literature see Bauwens and Hautsch (2009). To begin, a number of definitions are required. Let $\{t_i\}_{i\in 1,\cdots,n}$ be a random sequence of increasing event times $0 \geq t_1 > \cdots > t_n$ which describe a simple point process. Given $N(t) := \sum_{i\geq 1} \mathbf{1}_{t_i \geq t}$ is a counting function, the conditional intensity, $\lambda(t)$ can be viewed as the expected change in $N(t)$

	RV_t	$RV_t^{1/2}$	$log(RV_t)$	J_t	$J_t^{1/2}$	$log(J_t+1)$	J_t^*	CJ_t^*
Mean	0.901*	0.008	-10.087	0.058*	4.556*	0.058*	0.075	0.131
St. dev.	2.127*	0.006	1.184	0.668*	0.002	6.680*	0.263	0.337
Skewness	8.867	3.450	-0.312	22.458	9.801	22.448	3.231	2.191
Kurtosis	113.420	21.970	6.586	605.746	134.357	605.227	11.441	5.799
Min	0.0003*	1.803*	-17.242	0	0	0	0	0
Max	0.004	0.065	-5.477	0.002	0.047	0.002	1	1
LB_{10}	8,042	13,769	12,642	71.005	189.941	71.117	57.294	5.721

Table 1: Summary statistics for daily realized volatility and jumps in Dow Jones industry average index, and cojumps in its constituents. The first six rows report the sample mean, standard deviation, skewness, and kurtosis, along with the sample minimum and maximum. The rows labeled LB10 give the Ljung-Box test statistic for up to tenth-order serial correlation. The daily realized volatilities and jumps for the index, and the cojumps across the constituent stocks are constructed from five-minute returns spanning the period from 3 January 2000 to 25 June 2012, for a total of 3,113 daily observations. While J_t denotes the series of jump size, J_t^* and CJ_t^* respectively denote the binary series of jump and cojump occurrences. The numbers marked by an asterisk is scaled by 10^4.

(as a reflection of the probability of an event occurring) over a small time horizon,

$$\lambda(t) = \lim_{s \downarrow t} \frac{1}{s-t} E[N(s) - N(t)] \tag{11}$$

Bauwens and Hautsch (2009) provide a discussion of various specifications for $\lambda(t)$.

A common specification for $\lambda(t)$ is the self-exciting Hawkes process, attributable to Hawkes (1971)

$$\lambda(t) = \mu + \int_0^t w(t-u)dN(u) = \mu + \sum_{t_i<t} w(t-t_i) \tag{12}$$

where μ is a constant and $w()$ is a non-negative weight function. This process is self-exciting in the sense that $\text{Cov}[N(a,b),\, N(b,c)] > 0$ where $0 > a \geq b < c$. The weight function $w()$ is a decreasing function of $t-u$ meaning that subsequent to a spike the intensity decays. Bowsher (2007), Large (2007) and Ait-Sahalia, Cacho-Diaz, and Laeven (2011) model events in financial markets such as trading activity, liquidity supply and contagion across markets using Hawkes processes.

To implement the Hawkes model in equation 12, the common approach is to replace the integral by a discrete sum over past events such that the intensity is given by

$$\lambda(t) = \mu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)} \tag{13}$$

where α captures the immediate impact on intensity after an event occurs, and β controls the rate of the decay in the exponential weighting function as $t - t_i$ grows.

4.2 Forecasting volatility: Heterogeneous autoregressive (HAR) model

Once estimates of jump and cojump intensity are obtained, they are incorporated into a volatility forecasting model. With the widespread availability of high-frequency financial data, recent literature has focussed on RV to build forecasting models for time-varying financial volatility. Among these forecasting models of RV, the Heterogeneous Autoregressive (HAR) model proposed by Corsi (2009) has gained popularity due to its simplicity. The HAR formulation is based on a straightforward extension of the so-called Heterogeneous ARCH, or HARCH, class of models analyzed by Muller et al. (1997). In such models the conditional variance of the discretely sampled returns is parameterized as a linear function of the lagged squared returns over the identical return horizon together with the squared returns over longer and/or shorter return horizons. Corsi (2009) applied this type of model to RV itself to develop a HAR-RV model and Andersen et al. (2007) separated the RV regressors into their continuous and jump components to build up a HAR-RV-CJ model.

The original HAR-RV model specifies RV as a function of a daily, weekly and monthly realized volatility component, and is expressed as

$$RV_t^d = \alpha_0 + \alpha_1 RV_{t-1}^d + \alpha_2 RV_{t-1}^w + \alpha_3 RV_{t-1}^m + \varepsilon_t, \tag{14}$$

where RV_t^d denotes the daily realized volatility, $RV_{t-1}^w = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i}$ is the weekly realized volatility and $RV_{t-1}^m = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i}$ represents the monthly realized volatility. Andersen et al. (2007) extended this model by explicitly decomposing the realized volatilities into the continuous sample path variability and the jump variation utilizing the separate nonparametric measurements based on a statistical jump test to the HAR-RV-CJ model, which is expressed as

$$\begin{aligned} RV_t^d &= \alpha_0 + \alpha_1 C_{t-1}^d + \alpha_2 C_{t-1}^w + \alpha_3 C_{t-1}^m \\ &+ \alpha_4 J_{t-1}^d + \alpha_5 J_{t-1}^w + \alpha_6 J_{t-1}^m + \varepsilon_t, \end{aligned} \tag{15}$$

where C_t and J_t are respectively the continuous sample path variation and the jump variation at time t, and $C_{t-1}^w = \frac{1}{5}\sum_{i=1}^{5} C_{t-i}$, $C_{t-1}^m = \frac{1}{22}\sum_{i=1}^{22} C_{t-i}$, $J_{t-1}^w = \frac{1}{5}\sum_{i=1}^{5} J_{t-i}$, and $J_{t-1}^m = \frac{1}{22}\sum_{i=1}^{22} J_{t-i}$.

The current analysis extends the HAR-RV-CJ model in three directions. First, we further explore the information reflected in the realized jump component. Andersen et al. (2007) only considers the magnitude of jumps in the context of volatility forecasting, however, there is a second dimension to jump activity that being the intensity of a jump occurring that has thus far been ignored. Therefore, the estimated intensity of jumps λ_{J_t} from a Hawkes model is incorporated into a HAR framework in two ways. One is to replace the magnitude of jumps with the intensity of jumps to develop a HAR-RV-CI model, and the other is to include both dimensions of the jump component to produce a HAR-RV-CJI model.

The HAR-RV-CI model simply replaces past jump sizes with the estimated intensity

$$RV_t^d = \alpha_0 + \alpha_1 C_{t-1}^d + \alpha_2 C_{t-1}^w + \alpha_3 C_{t-1}^m + \alpha_7 \lambda_{J_t} + \varepsilon_t, \tag{16}$$

whereas the HAR-RV-CJI model

$$RV_t^d = \alpha_0 + \alpha_1 C_{t-1}^d + \alpha_2 C_{t-1}^w + \alpha_3 C_{t-1}^m + \alpha_4 J_{t-1}^d + \alpha_5 J_{t-1}^w + \alpha_6 J_{t-1}^m + \alpha_7 \lambda_{J_t} + \varepsilon_t \tag{17}$$

incorporates information relating to both past jumps size and occurrence. From equation 13, it can be seen that λ_{J_t} is estimated based on the past jump intensities, which actually incorporates all the historical information of jump occurrences. Hence, this additional information is useful to examine whether future volatility tends to increase or decrease when jumps have been occurring in the near past. Second, the information content in cojumps for future volatility is examined. The empirical data shows that the cojumps detected across constituent stocks appear to follow a different pattern from the index jump component, so it is plausible that the two kinds of jumps play distinct roles in predicting future volatility of the index returns. We add the estimated intensity of the cojumps λ_{CJ_t} from a Hawkes model into a HAR model to examine the impact of the probability of cojumps on future volatility, and denote the resulting model HAR-RV-CCJ. The HAR-RV-CCJ model is expressed as

$$RV_t^d = \alpha_0 + \alpha_1 C_{t-1}^d + \alpha_2 C_{t-1}^w + \alpha_3 C_{t-1}^m + \alpha_4 \lambda_{CJ_t} + \varepsilon_t. \tag{18}$$

Similarly, λ_{CJ_t} actually contains all the historical information of cojump occurrences, and hence this additional information may help forecast volatility when cojumps have been occurring in the near past. Third, the joint impact of the jump component of index returns and cojumps in constituent stocks on the future volatility of the index are examined. The joint effect of the intensity of both types of jumps, λ_{J_t} and λ_{CJ_t} are included in a HAR model leading to the HAR-RV-CICJ model,

$$RV_t^d = \alpha_0 + \alpha_1 C_{t-1}^d + \alpha_2 C_{t-1}^w + \alpha_3 C_{t-1}^m + \alpha_4 \lambda_{J_t} + \alpha_5 \lambda_{CJ_t} + \varepsilon_t. \tag{19}$$

Beyond the intensities, all possible jump related information including the magnitude of index jumps are included in the HAR-RV-CJCICJ model,

$$RV_t^d = \alpha_0 + \alpha_1 C_{t-1}^d + \alpha_2 C_{t-1}^w + \alpha_3 C_{t-1}^m + \alpha_4 J_{t-1}^d + \alpha_5 J_{t-1}^w + \alpha_6 J_{t-1}^m + \alpha_7 \lambda_{J_t} + \alpha_9 \lambda_{CJ_t} + \varepsilon_t. \tag{20}$$

The magnitude of cojumps cannot be included in any forecasting models as the Bollerslev et al. (2008) cojump test only detects the existence of a cojump across multiple assets. Hence no reference can be made to its size as it is an event that occurs across many assets. The nonlinear specifications (logarithm and square root) of the above models are also estimated in our empirical analysis as a robustness check.

4.3 Forecast evaluation

Here there are two obvious benchmark forecasts, HAR-RV and HAR-RV-CJ, which are models considered by Andersen et al. (2007). To highlight the role of jumps and cojump for forecasting index volatility, the forecast performance of the extended HAR models will compared to both the HAR-RV and HAR-RV-CJ models. To achieve this, the pairwise test for equal predictive accuracy (EPA) of Diebold and Mariano (1995) and West (1996) (DMW) is employed. Let $\mathcal{L}(f_t^a)$ and $\mathcal{L}(f_t^b)$ represent a generic loss function defined on two competing volatility forecasts f_t^a and f_t^b, then the relevant null and alternative hypotheses are

$$\begin{aligned} H_0 &: \mathbb{E}[\mathcal{L}(H_t^a)] = \mathbb{E}[\mathcal{L}(H_t^b)] \\ H_A &: \mathbb{E}[\mathcal{L}(H_t^a)] \neq \mathbb{E}[\mathcal{L}(H_t^b)]. \end{aligned} \tag{21}$$

The test is based on the computation of

$$DMW_T = \frac{\overline{d}_T}{\sqrt{\widehat{\mathrm{var}}[\overline{d}_T]}}\,, \qquad \overline{d}_T = \frac{1}{T}\sum_{t=1}^{T} d_t\,, \qquad d_t = \mathcal{L}(f_t^a) - \mathcal{L}(f_t^b)\,, \tag{22}$$

where $\widehat{\mathrm{var}}[\overline{d}_T]$ is an estimate of the asymptotic variance of the average loss differential, $\overline{d}_T$.

To begin, forecast performance will be compared using the simple root mean squared forecast error of the $i-th$ forecast, defined as

$$RMSE^i = \sqrt{\frac{1}{T}(RV_t - f_t^i)^2}, \tag{23}$$

where T is the total number of forecast periods (2113 in this case), f_t^i is the forecast from the $i-th$ model and RV_t is the target. To implement the DMW test, given the $i-th$ forecast, the MSE loss function is chosen to represent $\mathcal{L}()$,

$$MSE_t^i = (RV_t - f_t^i)^2. \tag{24}$$

5 Empirical Results

The first 2113 days of the sample from January 3, 2000 to June 25, 2008 are used as the initial estimation window with the last 1000 days used for out-of-sample forecast analysis. Subsequent to the initial estimation period, a recursive window is used for parameter estimation. These estimated parameters are used to construct one-step-ahead out-of-sample forecasts that incorporates new information as it becomes available. Analysis will be based on linear models for the level of RV and nonlinear models for the natural logarithm and the square root of RV. The following two subsections outline in-sample estimation results and analyze forecast performance.

Hawkes Model: $\lambda(t) = \mu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}$				
	μ	α	β	loglike
λ_J	0.0487	0.0101	0.0356	520.3883
	(0.0127)	(0.0045)	(0.0175)	
λ_{CJ}	0.0951	0.0019	0.0060	849.3978
	(0.0281)	(0.0009)	(0.0031)	

Table 2: Estimation results of Hawkes model for index jump and cojump intensity. This table reports the MLE estimates for Hawkes model of index jump and cojump intensity from January 3, 2000 to June 25, 2008, for a total of 2113 trading days. The standard errors are reported in parentheses. The last column reports the negative value of the sum of log likelihood function across time evaluated at the estimates.

5.1 In-sample estimation

The parameters of the Hawkes model in equation (13) estimated for index jumps and cojumps to obtain their intensities are reported in Table 2. The first row represents the model results for index jumps and the second row shows the results for cojumps. The estimates for β and α in the two models are significantly positive, confirming that both jumps and cojumps self-excite and exhibit a persistent intensity process. The estimated jump and cojump intensity from the Hawkes models are plotted in Figure 2. The plot shows that the index jump intensity and cojump intensity respectively fluctuates around 0.06 and 0.13 during period 2000-2008, which are very close to the proportions of trading days with jumps and cojumps reported in Section 3. After a slight increase between 2000-2002, the jump intensity reaches a high between 2003-2005. The cojump intensity continues rising, and reaches a very high level in the middle of 2005.

Table 3 reports estimation results of all the linear HAR models, including HAR-RV and HAR-RV-CJ models as the two benchmark models, and the five extended HAR models. Table 4 and Table 5 present estimation results of all the nonlinear (the logarithm and square root transformation) HAR models. These results reveal several noteworthy points. First, consistent with Andersen et al. (2007), the estimates of α_1, α_2 and α_3 in all models confirm the strong predictive power of the continuous part of volatility. The estimates of α_4, α_5 and α_6 in the related models show that the magnitude of jump component (jump size) contains no predictive power, a result broadly consistent with Andersen et al. (2007). In contrast to Andersen et al. (2007), the improvement in in-sample fit from separately modeling the continuous and jump components implied by the change of R^2 from the HAR-RV model to the HAR-RV-CJ model is relatively modest. Second, the intensity of the jump component exhibits strong predictive power for volatility and the effect appears stronger under the nonlinear specifications. The estimated coefficient of the jump intensity (α_7) is systematically positive, and overwhelmingly significant

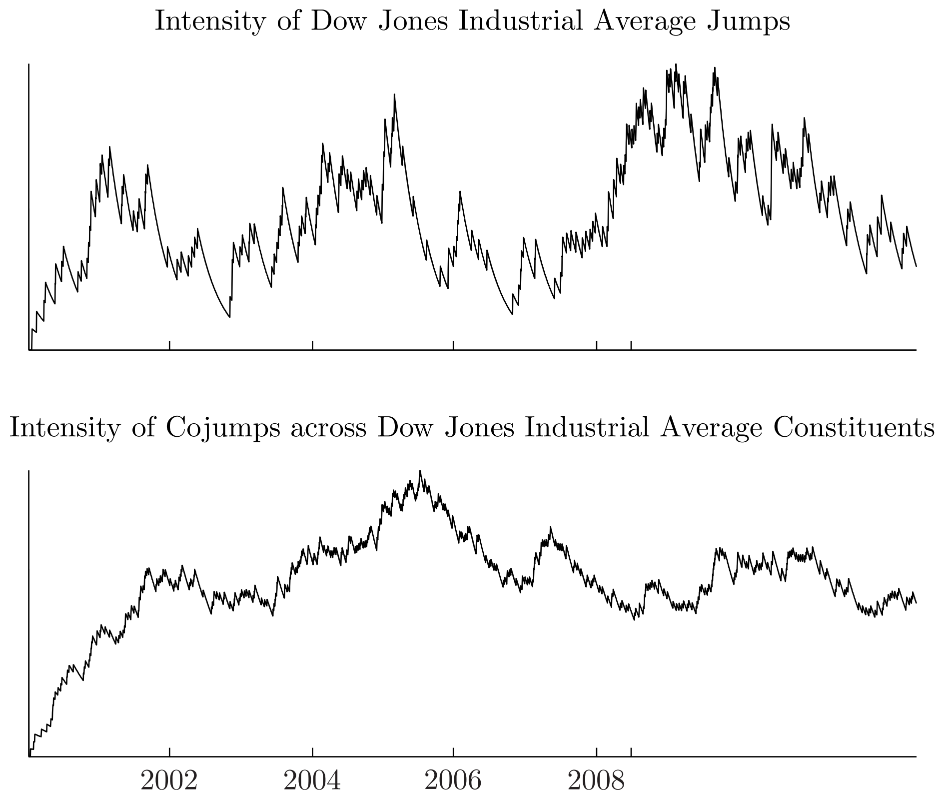

Figure 2: The estimated index jumps and cojumps intensity from the Hawkes model as defined in equation (13)

in most of the related models. This reveals that when jumps are more likely to occur volatility will be higher, and implies beyond the size of jumps, the probability with which they occur is important. While future volatility is significantly related to jump intensity, improvements in in-sample model fit are still modest. R^2 for any of the linear or nonlinear versions only increase marginally from the HAR-RV-CJ to either the HAR-RV-CI or HAR-RV-CJI models. Thirdly, the cojumps contain strong predictive information for future volatility. It is clear that the estimates of α_8 (the coefficient on cojump intensity) is uniformly negative and statistically significant regardless of the model form. This suggests that when cojumps are more likely to occur, volatility is subsequently lower. Again, in terms of in-sample fit, the change of R^2 from HAR-RV-CJ model to HAR-RV-CCJ is relatively small. The HAR-RV-CJICJ model, in which all the jump and cojump related information are included has the highest R^2, implying that both dimensions of jump activity are important for volatility prediction. Even though improvements in in-sample fit are small, the crucial issue of out-of-sample forecast performance of these models will be considered later.

Meanwhile, the magnitude of the estimated coefficients in Hawkes model quantifies the impact of the occurrences of past jumps and cojumps on future volatility. For example, if a jump occurs at time $t-1$, the probability of a jump occurring at time t will increase by $\alpha e^{-\beta} = 0.0097$ according to the estimation results for the Hawkes model in equation (13), and thereby the volatility at time t will increase by $\alpha_7 \times 0.0097 = 1.8226 \times 10^{-6}$ according to the estimation results of linear HAR-RV-CJICJ. Similarly, if a cojump occurs at time $t-1$, the probability of a cojump occurring at time t will increase by $\alpha e^{-\beta} = 0.0019$, and the volatility at time t will decrease by $\alpha_8 \times 0.0019 = 9.6835 \times 10^{-7}$. It is worth noting that the occurrence of past jumps and cojumps on the future volatility is not considered in the standard HAR-RV-CJ model.

To consider these impacts in further detail, we assume that both a jump of size 1.2641×10^{-5} and a cojump occurred at time $t-1$.[3] Figure 3 plots the impact this jump and cojump on volatility over the following five days implied by the linear HAR-RV-CJICJ model and the two benchmark linear models, HAR-RV and HAR-RV-CJ models. Under the HAR-RV the impact of the index jump occurs via an increase in lagged RV where as under the HAR-RV-CJ model index jumps only enter directly as exogenous variables. On the other hand, jumps and cojumps enter the HAR-RV-CJICJ model by their influence on both intensities along with lagged jump size and subsequently future volatility. The curve from the linear HAR-RV model which only considers the impact of index jumps implies that volatility increases by 4.96182×10^{-6} in the next period after a jump, and the impact falls to 9.11646×10^{-7} in the following four periods. The linear HAR-RV-CJ model indicates that when a jump and cojump occur, the volatility in the next period will decrease by around -3.3615×10^{-7}, this impact is reduced to -1.3136×10^{-7} in the following four periods. After adding in the influence of the jump and cojump intensity, as shown by the curve from the linear HAR-RV-CJICJ model, the volatility in the next period increases by 3.8452×10^{-7} after a jump and a cojump occur, and the impact on the volatility keeps increasing until the fifth day. It is clear that the impact of a jump and cojump on future volatilities with respect to its magnitude and direction is distinctly different after considering jump and cojump intensity, showing that the impact of incorporating jump and cojump intensity within the HAR framework has important implications for volatility forecast, an issue now considered.

[3] This magnitude of the jump is randomly selected jump and is quite close to the mean value of jump size in the data.

$$
\begin{aligned}
\text{HAR-RV: } RV_t^d &= \alpha_0 + \alpha_1 RV_{t-1}^d + \alpha_2 RV_{t-1}^w + \alpha_3 RV_{t-1}^m + \varepsilon_t, \\
\text{HAR-RV-CJICJ: } RV_t^d &= \alpha_0 + \alpha_1 C_{t-1}^d + \alpha_2 C_{t-1}^w + \alpha_3 C_{t-1}^m + \alpha_4 J_{t-1}^d \\
&+ \alpha_5 J_{t-1}^w + \alpha_6 J_{t-1}^m + \alpha_7 \lambda_{J_t} + \alpha_8 \lambda_{CJ_t} + \varepsilon_t
\end{aligned}
$$

Model	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	adjusted R^2
HAR-RV	9.6938×10^{-6}	0.3204	0.3095	0.2248						0.4090
	(3.1083)	(3.9624)	(2.6063)	(2.9777)						
HAR-RV-CJ	1.1895×10^{-5}	0.3499	0.2963	0.2321	-0.0162	-0.1038	0.2281			0.4137
	(4.0787)	(3.8806)	(2.2364)	(2.9684)	(-0.0975)	(-0.5275)	(-0.5811)			
HAR-RV-CI	1.6511×10^{-6}	0.3489	0.2949	0.2401				1.5156×10^{-4}		0.4140
	(0.1850)	(3.9210)	(2.2474)	(3.0729)				(1.0816)		
HAR-RV-CJI	6.4350×10^{-7}	0.3497	0.2964	0.2386	-0.0171	-0.1138	0.0884	1.6744×10^{-4}		0.4141
	(0.0704)	(3.8679)	(2.2306)	(3.0750)	(-0.1039)	(-0.5679)	(0.2255)	(1.2494)		
HAR-RV-CCJ	1.2893×10^{-5}	0.3507	0.2928	0.2351						0.4145
	(4.0726)	(3.9666)	(2.2294)	(3.0220)						
HAR-RV-CICJ	7.4473×10^{-5}	0.3466	0.2938	0.1991				1.4121×10^{-4}		0.4170
	(2.1516)	(3.8647)	(2.2871)	(2.9488)				(1.0184)	(-2.1171)	
HAR-RV-CJICJ	7.1571×10^{-5}	0.3484	0.2936	0.2002	-0.0141	-0.1089	-0.0539	1.8792×10^{-4}	-5.0966×10^{-4}	0.4171
	(2.0491)	(3.8364)	(2.2534)	(2.9275)	(-0.0861)	(-0.5350)	(-0.1426)	(1.3966)	(-2.1270)	

Note: This table reports the OLS estimates for linear HAR-RV model, HAR-RV-CJ model and our seven extended HAR models, in which the continuous variation, jump component and cojumps are constructed from the original five-minute returns spanning from January 3, 2000 to June 25, 2008, for a total of 2113 trading days. The Newey-West t-statistics are reported in parentheses. The last column labeled adjusted R2 is adjusted r-square of the regression

Table 3: Estimation results of linear HAR type models on realized volatility

$$
\begin{aligned}
\text{HAR-RV: } log(RV_t^d) &= \alpha_0 + \alpha_1 log(RV_{t-1}^d) + \alpha_2 log(RV_{t-1}^w) + \alpha_3 log(RV_{t-1}^m) + \varepsilon_t, \\
\text{HAR-RV-CJICJ: } log(RV_t^d) &= \alpha_0 + \alpha_1 log(C_{t-1}^d) + \alpha_2 log(C_{t-1}^w) + \alpha_3 log(C_{t-1}^m) + \alpha_4 log(J_{t-1}^d + 1) \\
&+ \alpha_5 log(J_{t-1}^w + 1) + \alpha_6 log(J_{t-1}^m + 1) + \alpha_7 log(\lambda_{J_t}) + \alpha_8 log(\lambda_{CJ_t}) + \varepsilon_t
\end{aligned}
$$

Model	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	adjusted R^2
HAR-RV	-0.8696	0.3113	0.3786	0.2307						0.6423
	(-5.2244)	(8.8898)	(7.8689)	(6.6538)						
HAR-RV-CJ	-0.8342	0.2938	0.3951	0.2273	-362.9104	-105.1590	-1088.4			0.6458
	(-4.8960)	(8.8390)	(7.9801)	(6.4421)	(-0.6035)	(-0.0600)	(-0.3714)			
HAR-RV-CI	-1.0049	0.2927	0.3928	0.2301				0.1631		0.6468
	(-5.7931)	(8.8553)	(8.0976)	(6.6727)				(2.1603)		
HAR-RV-CJI	-0.9670	0.2930	0.3949	0.2371	-380.5989	-310.2357	-4079.8	0.2295		0.6469
	(-5.5319)	(8.8641)	(7.9844)	(6.7000)	(-0.6585)	(0.1718)	(-1.3288)	(2.8126)		
HAR-RV-CCJ	-0.7376	0.2918	0.3941	0.2033					-0.3443	0.6467
	(-4.0867)	(8.8157)	(8.0862)	(5.6925)					(-1.9892)	
HAR-RV-CICJ	-0.8767	0.2917	0.3926	0.2067				0.1659	-0.3508	0.6474
	(-4.7616)	(8.8251)	(8.0671)	(5.8156)				(2.2144)	(-2.0334)	
HAR-RV-CJICJ	-0.8437	0.2928	0.3935	0.2140	-364.9849	-292.3250	-4236.0	0.2344	-0.3409	0.6475
	(-4.5234)	(8.8149)	(7.8925)	(5.9293)	(-0.6274)	(-0.1593)	(-1.3893)	(2.9076)	(-1.9918)	

Note: This table reports the OLS estimates for nonlinear (the natural logarithm transformation) HAR-RV model, HAR-RV-CJ model and our seven extended HAR models, in which the continuous variation, jump component and cojumps are constructed from the original five-minute returns spanning from January 3, 2000 to June 25, 2008, for a total of 2113 trading days. The Newey-West t-statistics are reported in parentheses. The last column labeled adjusted R2 is adjusted r-square of the regression.

Table 4: Estimation results of nonlinear (the natural logarithm transformation) HAR type models on realized volatility

$$
\begin{aligned}
\text{HAR-RV: } (RV_t^d)^{1/2} &= \alpha_0 + \alpha_1 (RV_{t-1}^d)^{1/2} + \alpha_2 (RV_{t-1}^w)^{1/2} + \alpha_3 (RV_{t-1}^m)^{1/2} + \varepsilon_t, \\
\text{HAR-RV-CJICJ: } (RV_t^d)^{1/2} &= \alpha_0 + \alpha_1 (C_{t-1}^d)^{1/2} + \alpha_2 (C_{t-1}^w)^{1/2} + \alpha_3 (C_{t-1}^m)^{1/2} + \alpha_4 (J_{t-1}^d)^{1/2} \\
&+ \alpha_5 (J_{t-1}^w)^{1/2} + \alpha_6 (J_{t-1}^m)^{1/2} + \alpha_7 \lambda_{J_t}^{1/2} + \alpha_8 \lambda_{CJ_t}^{1/2} + \varepsilon_t
\end{aligned}
$$

Model	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	adjusted R^2
HAR-RV	6.8528×10^{-4}	0.3663	0.3139	0.2081						0.5984
	(3.9340)	(7.7985)	(4.6580)	(4.3707)						
HAR-RV-CJ	7.8555×10^{-4}	0.3806	0.3144	0.2081	0.0088	-0.0205	0.0575			0.6049
	(4.7250)	(8.8400)	(4.3028)	(4.2611)	(0.1319)	(-0.3675)	(0.8125)			
HAR-RV-CI	-3.2928×10^{-4}	0.3811	0.3137	0.2149				0.0043		0.6056
	(-0.5083)	(8.7119)	(4.2600)	(4.3489)				(1.7287)		
HAR-RV-CJI	-5.4780×10^{-4}	0.3804	0.3148	0.2159	0.0070	-0.0331	5.9336×10^{-4}	0.0052		0.6057
	(-0.6754)	(8.8139)	(4.2767)	(4.4813)	(0.1056)	(-0.5824)	(0.0073)	(1.6968)		
HAR-RV-CCJ	0.0055	0.3796	0.3135	0.1793					-0.0122	0.6068
	(2.7188)	(8.6643)	(4.2736)	(3.8973)					(-2.3703)	
HAR-RV-CICJ	0.0044	0.3793	0.3125	0.1840				0.0042	-0.0121	0.6072
	(2.1911)	(8.6258)	(4.2607)	(4.0043)				(1.6922)	(-2.3653)	
HAR-RV-CJICJ	0.0039	0.3802	0.3119	0.1872	0.0106	-0.0370	-0.0374	-2.0073×10^{-4}	-0.0123	0.6073
	(1.9162)	(8.7567)	(4.2180)	(4.0849)	(0.1585)	(-0.6497)	(-0.4686)	(2.1292)	(-2.4076)	

Note: This table reports the OLS estimates for nonlinear (the square root transformation) HAR-RV model, HAR-RV-CJ model and our seven extended HAR models, in which the continuous variation, jump component and cojumps are constructed from the original five-minute returns spanning from January 3, 2000 to June 25, 2008, for a total of 2113 trading days. The Newey-West t-statistics are reported in parentheses. The last column labeled adjusted R2 is adjusted r-square of the regression.

Table 5: Estimation results of nonlinear (the square root transformation) HAR type models on realized volatility

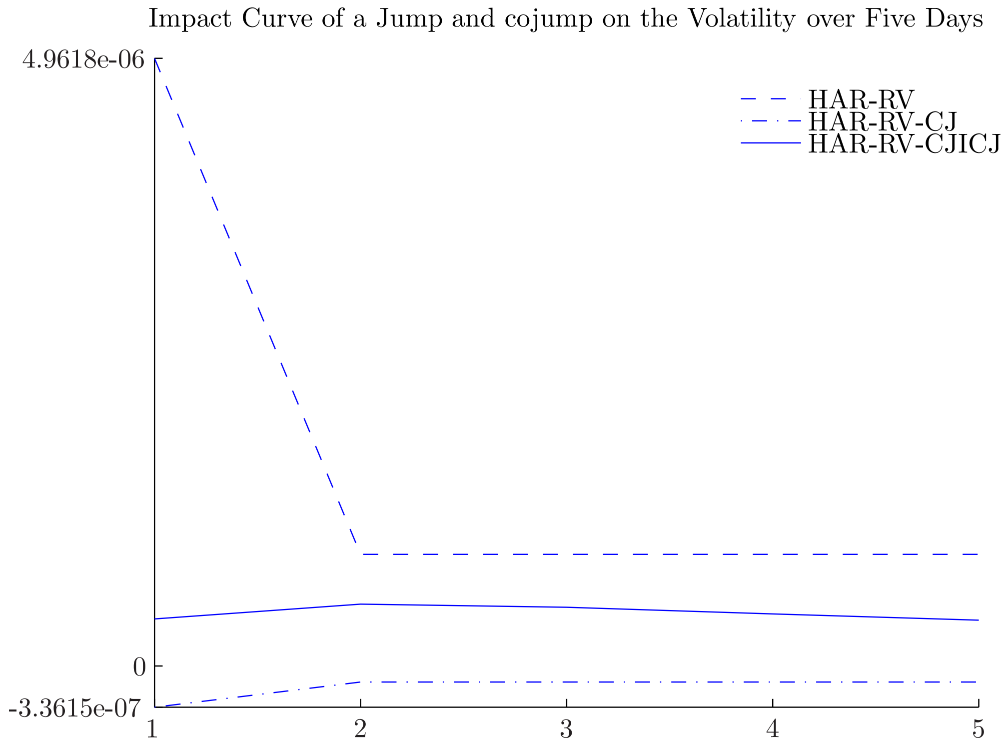

Figure 3: Impact curve of a jump and cojump on the volatility over five days

5.2 Out-of-sample forecast performance

The volatility forecast performance of all HAR models will first be compared given a simple root mean squared one-step ahead forecast errors (RMSE). While doing so indicates relative forecast accuracy, this gives no indication of whether any differences in performance are significant. The test of equal predictive accuracy of Diebold and Mariano (1995) and West (1996) (DMW) will be used to achieve this. All models will be individually compared to the two benchmark models, HAR-RV and HAR-RV-CJ. The forecast performance of these models will be considered in four settings, including the full sample of 1000 days, days on which jumps occur, days on which cojumps occur and the period of high volatility from June 2008 to December 2009 associated with the Global Financial Crisis. Forecast performance in these subsamples is of great interest, as these are times when volatility is rising rapidly and decision making must adapt quickly to changing conditions.

Table 6 contains the RMSE for all models, with Panels A, B and C reporting results based on linear and nonlinear (logarithm and square root) specifications respectively. Beginning with the whole sample results, for both linear and nonlinear models, RMSE reduces moving from HAR-RV and HAR-RV-CJ to other HAR models. The middle columns of Table 6 report equivalent results for days on which jumps or cojumps occur. In the RV and $log(RV)$ cases, all other HAR models outperform the HAR-RV and HAR-RV-CJ models with the improvement being more pronounced than the full sample period. This result also applies in the $RV^{1/2}$ case with one exception. The relative forecast performance during 2008-2009 also follows the same pattern, with the differences in performance once again being more marked than the whole sample period. Overall, the HAR-RV-CI model achieves the smallest RMSE in all cases with only one exception. While there are improvements in forecast accuracy, results from the DMW test reveal whether these are significant improvements. Here, the forecast performance of each model is individually compared to both HAR-RV and HAR-RV-CJ by applying the DMW test. Results for the DMW test report a single * when a model significantly outperforms HAR-RV, and a double ** if a model exhibits significant improvement over HAR-RV-CJ. For the whole sample period, most of the models are significantly superior to both the HAR-RV and HAR-RV-CJ benchmarks. On days with jumps or cojumps the HAR-RV-CJ offers no improvement over the simpler HAR-RV model under RV or $RV^{1/2}$. In all cases with one exception there are significant improvements over HAR-RV and in most cases improvements over the HAR-RV-CJ benchmark. In the turbulent period, these result are very similar to the full sample with all forecasts significantly outperforming HAR-RV and in most cases HAR-RV-CJ.

Hence, using the probability of jump and cojump occurrences in HAR type models leads to the out-of-sample volatility forecasts that are statistically superior to those obtained from the HAR-RV and HAR-RV-CJ models. These results show that moving beyond simply modeling the continuous and jump components separately as in the HAR-RV-CJ model, to utilising the information in the intensity of jumps is important for forecasting index volatility. A novel contribution of these results is that cojumps across the constituent stocks are found to contain useful information for forecasting volatility at an index level. These results is robust to the crucial times when volatility is rising due to jumps or general market uncertainty.

Panel A: $y = RV$

	The Whole Sample	The Days with Jumps	The Days with Cojumps	The Turbulent Volatility
HAR-RV	2.5558×10^{-4}	3.9225×10^{-4}	2.1113×10^{-4}	3.8213×10^{-4}
HAR-RV-CJ	$2.4948\times10^{-4*}$	3.9212×10^{-4}	2.0153×10^{-4}	$3.7355\times10^{-4*}$
HAR-RV-CI	$2.4583\times10^{-4**}$	$3.8358\times10^{-4**}$	$1.9227\times10^{-4**}$	$3.6720\times10^{-4**}$
HAR-RV-CJI	$2.4783\times10^{-4**}$	$3.8752\times10^{-4**}$	$2.0139\times10^{-4*}$	$3.7074\times10^{-4**}$
HAR-RV-CCJ	$2.4657\times10^{-4**}$	$3.8748\times10^{-4**}$	$1.9431\times10^{-4**}$	$3.6846\times10^{-4**}$
HAR-RV-CICJ	$2.4622\times10^{-4**}$	$3.8406\times10^{-4**}$	$1.9286\times10^{-4**}$	$3.6787\times10^{-4**}$
HAR-RV-CJICJ	$2.4824\times10^{-4**}$	$3.8483\times10^{-4**}$	$2.0109\times10^{-4*}$	$3.7144\times10^{-4**}$

Panel B: $y = log(RV)$

	The Whole Sample	The Days with Jumps	The Days with Cojumps	The Turbulent Volatility
HAR-RV	2.4430×10^{-4}	3.8916×10^{-4}	2.1214×10^{-4}	3.6415×10^{-4}
HAR-RV-CJ	$2.4005\times10^{-4*}$	$3.8899\times10^{-4*}$	$1.9695\times10^{-4*}$	$3.5689\times10^{-4*}$
HAR-RV-CI	$2.3630\times10^{-4**}$	$3.8238\times10^{-4**}$	$1.9401\times10^{-4**}$	$3.5039\times10^{-4**}$
HAR-RV-CJI	$2.3788\times10^{-4**}$	$3.8252\times10^{-4**}$	$1.8778\times10^{-4**}$	$3.5321\times10^{-4**}$
HAR-RV-CCJ	$2.3857\times10^{-4*}$	$3.8748\times10^{-4*}$	$1.9191\times10^{-4**}$	$3.5564\times10^{-4*}$
HAR-RV-CICJ	$2.3752\times10^{-4**}$	$3.8406\times10^{-4**}$	$1.8930\times10^{-4**}$	$3.5267\times10^{-4**}$
HAR-RV-CJICJ	$2.3904\times10^{-4*}$	$3.8483\times10^{-4**}$	$1.9517\times10^{-4*}$	$3.5538\times10^{-4*}$

Panel C: $y = RV^{1/2}$

	The Whole Sample	The Days with Jumps	The Days with Cojumps	The Turbulent Volatility
HAR-RV	2.4901×10^{-4}	3.9025×10^{-4}	1.9330×10^{-4}	3.7193×10^{-4}
HAR-RV-CJ	$2.4181\times10^{-4*}$	3.8884×10^{-4}	1.9279×10^{-4}	$3.5902\times10^{-4*}$
HAR-RV-CI	$2.3976\times10^{-4**}$	$3.8701\times10^{-4**}$	$1.9131\times10^{-4**}$	$3.5718\times10^{-4**}$
HAR-RV-CJI	$2.4064\times10^{-4**}$	$3.8751\times10^{-4**}$	$1.9157\times10^{-4**}$	$3.5875\times10^{-4*}$
HAR-RV-CCJ	$2.4136\times10^{-4*}$	$3.8964\times10^{-4*}$	$1.9221\times10^{-4*}$	$3.5899\times10^{-4*}$
HAR-RV-CICJ	$2.4044\times10^{-4**}$	$3.8800\times10^{-4*}$	$1.9231\times10^{-4*}$	$3.5844\times10^{-4**}$
HAR-RV-CJICJ	$2.4128\times10^{-4*}$	$3.8844\times10^{-4*}$	1.9268×10^{-4}	$3.5893\times10^{-4*}$

Note: This table reports the root mean squared one-step-ahead volatility forecast errors of HAR models. Panel A, Panel B and Panel C respectively reports the corresponding results for linear and nonlinear (logarithm and square root) versions.

Table 6: Root Mean Squared Forecast Errors of linear and nonlinear HAR models

6 Conclusion

The modelling and forecasting of financial volatility is a critical issue and hence attracts a great deal of research attention. In recent years, this literature has benefited from the availability of high frequency price data. Such data has led to improvements in the estimation of realized volatility and identification of its diffusion and jump components. While a great deal of work has focused on forecasting realized volatility, less has considered how to use both components of volatility effectively. This paper contributes to this literature and examines how best to use both index level jumps and cojumps across index constituents for forecasting index level volatility. Specifically, the role of index jump size and the probability of index jumps and cojumps occurring were considered. A point process model was developed to capture the intensity, or probability of index jump and cojump occurrence. These estimated intensities were then used in a volatility forecasting framework. It was found that by extending existing forecasting models with jump and cojump intensity, superior forecasts were obtained. Therefore information about the behaviour of underlying constituent stocks and the probability with which jumps occur is useful for forecasting index level behaviour. This improvement in performance was robust to critical periods when jumps or cojumps occurred, or when volatility high during periods of financial crisis.

References

Y. Ait-Sahalia, J. Cacho-Diaz, and R.J.A Laeven. Modeling financial contagion using mutually exciting jump processes. 2011.

T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001.

T.G. Andersen and T. Bollerslev. Dm-dollar volatility: intraday activity patterns, macroeconomic announcements and longer run dependencies. *Journal of Finance*, 53:219–265, 1998.

T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71:579–625, 2003.

T.G. Andersen, T. Bollerslev, and F.X. Diebold. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89:701–720, 2007.

O.E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of Royal Statisitcal Society, Series B*, 64: 253–280, 2002.

O.E. Barndorff-Nielsen and N. Shephard. Measuring the impact of jumps in multivariate price processes using bi-power covariation. Working Paper, Oxford University, 2003.

O.E. Barndorff-Nielsen and N. Shephard. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30, 2006.

L. Bauwens and N. Hautsch. Modelling financial high frequency data using point processes. In T.G. Andersen, R.A. Davis, J-P. Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*. Springer, 2009.

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

T. Bollerslev, T. H. Law, and G. Tauchen. Risk, jumps, and diversification. *Journal of Econometrics*, 144:234–256, 2008.

C.G. Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141:876–912, 2007.

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

F.X. Diebold and R.S. Mariano. Comparing predictive accuracy. *Journal of Business and Economics Statistics*, 13:253–263, 1995.

M. Dungey and L. Hvozdyk. Cojumping: Evidence from the us treasury bond and futures markets. Working Paper, 2011.

R.F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1008, 1982.

E. Ghysels, P. Santa-Clara, and R. Valkanov. Predicting volatility: how to get most out of returns data sampled at different frequencies. *Journal of Econometrics*, 131:59–95, 2006.

F. Gobbi and C. Mancini. Identifying the covariation between the diffusion parts and the co-jumps given discrete observations. Working Paper, Dipartimento di Matematica per le Decisioni, Universita degli Studi di Firenze, 2007.

A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.

X. Huang and G. Tauchen. The realtive contribution of jumps to total price variation. *Journal of Financial Econometrics*, 3:456–499, 2005.

J. Jacod and V. Todorov. Testing for common arrivals of jumps for discretely observed multidimensional processes. *The Annals of Statistics*, 37(4):1792–1838, 2009.

S.J. Koopman, B. Jungbacker, and E. Hol. Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12:445–475, 2005.

S. Lahaye, L. Laurent and C. Neely. Jumps, cojumps and macro announcement. *Journal of Applied Econometrics*, 26(6):893–921, 2011.

J. Large. Measuring the resiliency of an electronic order book. *Journal of Financial Markets*, 10:1–25, 2007.

U. Muller, M. Dacorogna, R. Dave, R. Olsen, O. Pictet, and J. Weizsacker. Volatilities of different time resolutions – analysing the dynamics of market components. *Journal of Empirical Finance*, 4:213–239, 1997.

R. Oomen. Using high frequency stock market index data to calculate, model and forecast realized return variance. 2001.

K.D. West. Asymptotic inference about predictive ability. *Econometrica*, 64:1067–1084, 1996.