

The Economic Value of Volatility Timing with Realized Jumps *

Ingmar Nolte [†] Qi Xu [‡]

This Version: January 29, 2015

Abstract

This paper comprehensively investigates the role of realized jumps detected from high frequency data in predicting future volatility from both statistical and economic perspectives. Using seven major jump tests, we show that separating jumps from diffusion improves volatility forecasting both in-sample and out-of-sample. Moreover, we show that these statistical improvements can be translated into economic value. We find a risk-averse investor can significantly improve her portfolio performance by incorporating realized jumps into a volatility timing based portfolio strategy. Our results hold true across the majority of jump tests, and are robust to controlling for microstructure effects and transaction costs.

JEL classification: C58, C53, G11

Keywords: high frequency data, jumps, nonparametric tests, asset allocation, volatility forecasting, realized volatility.

*We thank Marcelo Fernandes, Sandra Nolte, Alessandro Palandri, and conference and seminar participants at EEA-ESEM 2013 (Gothenburg), IFABS 2013 (Nottingham), and Warwick Business School for helpful comments.

[†]Accounting and Finance Department, Lancaster University Management School, Lancaster University, Lancaster, LA1 4YX, United Kingdom. Email: I.Nolte@lancaster.ac.uk.

[‡]Corresponding author: Finance Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, United Kingdom. Email: Qi.Xu.11@mail.wbs.ac.uk.

1 Introduction

The importance of jumps in asset pricing, option pricing, and risk management is widely recognized (Ait-Sahalia 2004). Although, resorting on jumps as a modeling device is not new, realized jumps were generally overlooked until recently. In this paper, we comprehensively investigate the role of realized jumps detected from high frequency data for the prediction of future volatility. Different from previous studies with similar focus, we not only conduct an extensive statistical evaluation of volatility forecasting using all major jump tests, but also provide new economic insights in the form of whether a risk-averse investor can significantly benefit from considering realized jumps in volatility timing based portfolio allocation strategies.

The literature on this topic can be broadly categorized into two streams: The parametric literature starting with Merton (1976) includes jump-diffusion and stochastic-volatility with jumps (SVJ) models in continuous time (Eraker, Johannes, and Polson 2003, Eraker 2004, Chernov, Gallant, Ghysels, and Tauchen 2003) and GARCH-J models in discrete time (Maheu and McCurdy 2004, Duan, Ritchken, and Sun 2006, Christoffersen, Jacobs, and Ornthanalai 2012). These parametric models are widely used in portfolio choice, option pricing, and risk management applications and the jumps introduced in models are *ex ante* in nature. As Backus, Chernov, and Martin (2011) admit: “jumps are usually treated as a modeling device to generate non-normal returns, which is economically meaningless *per se*”.

The second stream of the literature considers nonparametric approaches. Recently, many nonparametric jump tests (Barndorff-Nielsen and Shephard 2006, Andersen, Bollerslev, and Dobrev 2007, Ait-Sahalia and Jacod 2009) use high frequency data to estimate *ex post* realized jumps. This stream of the literature primarily focuses on issues such as why asset prices jump (e.g. macroeconomic new announcements) or how often asset prices jump (e.g. less than one per day). However, only a very few studies consider economic applications of realized jumps. We therefore aim to fill this gap between two related but different streams of literature by considering economic applications of realized jumps.

We focus on two research questions: Firstly, we are interested in whether realized jumps can forecast future volatility. We apply seven main stream nonparametric jump tests to identify realized jumps, decompose realized variance into jump and diffusion

components, and then adapt them into a forecasting framework. Our findings suggest that realized jumps do contain predictive information for future volatility for the majority of jump tests both in-sample and out-of-sample. We find that jump models in general generate higher adjusted R squares and lower Mean Squared Prediction Errors (MSPE) compared to the benchmark model, which does not separate jumps from diffusion. Results hold true across the majority of jump specifications, and different forecasting horizons. Existing studies investigate similar issues. However, they mainly rely on one particular jump test and their results are mixed. For example, Andersen, Bollerslev, and Diebold (2007) find a negative (but insignificant) relationship between jumps and one period ahead volatility. Corsi, Pirino, and Reno (2010) on the contrary show statistically significant evidence to support a positive relationship if a modified jump test is applied. By using all major jump tests, our results contribute to the debate whether in general realized jumps help to forecast volatility.

Incorporating realized jumps into volatility forecasting require accessing intraday high frequency data and applying sophisticated nonparametric jump tests. Therefore, a natural question arises whether it is worth to estimate and use realized jumps. Even though separating jumps from diffusion improves volatility forecasting, it is interesting to know whether the improvement is large, and more importantly whether the improvement is economically valuable. Therefore, our second research question explicitly asks whether the potential statistical forecasting improvement obtained by separating jumps from diffusion can be translated into tangible economic benefits for a risk-averse investor. We construct a mean-variance portfolio strategy based on the predicted volatility obtained from the previous step. Our findings suggest that the statistical improvements are also economically significant. Under different risk aversion levels and jump specifications, jump strategies can in general generate positive and statistically significant performance fees relative to the benchmark strategy. A few existing papers also consider the role of jumps in asset allocations. For example, Liu, Longstaff, and Pan (2003) provide an analytical solution to the optimal portfolio choice problem when event risk or jumps are considered. They find that jumps play an important role in determining the optimal portfolio choice. Two recent studies by Chen, Hyde, and Poon (2010) and Maheu, McCurdy, and Zhao (2012) are also close to us in considering jumps in asset allocation. However, we differ from those studies on a few aspects. Firstly, our nonparametric framework enables us to exploit the information embedded in jump variations in a model free fashion while previous papers rely on a parametric specification. Secondly, we use high frequency data to

separate the jumps and the diffusion component precisely, while they mainly rely on daily data to obtain relatively noisy proxies for jumps (i.e. large extraordinary movement or middle size jumps etc). The high frequency data we use also allows us to access intraday information, which is overlooked by previous studies.

We then conduct comprehensive robustness checks, and find that further controlling for market microstructure effects and transaction costs does not change our main results. We also investigate the predictive ability of realized jumps on alternative realized moments. We find that realized jumps can predict realized volatility and its signed components, but can hardly predict realized higher moments. We further show that a mean-variance portfolio strategy based on predicting positive and negative conditional volatilities separately can outperform the benchmark strategy based on predicting total volatility, and incorporating realized jumps can additionally improve economic benefits.

A few other studies are also related to ours. Firstly, our paper can be viewed as a natural extension of the stream of literature considering the economic value of volatility timing. Previous studies (Fleming, Kirby, and Ostdiek 2003, Bandi and Russell 2006, Bandi, Russell, and Zhu 2008, Liu 2009) already document that volatility timing performance can be improved by using high frequency data, optimal sampling, and optimal rebalancing frequencies. We extend the above studies by considering realized jumps. Secondly, our paper is also related to other uses of realized jumps or applications of jump tests. For example, Dumitru and Urga (2012) and Theodosiou and Zikes (2011) conduct comprehensive simulation studies to compare size and power of jump tests. Tauchen and Zhou (2011) and Jiang and Yao (2012) use detected realized jumps to predict bond risk premia and the cross-section of stock returns respectively. We distinguish ours from previous studies by focusing on the role of realized jumps in volatility timing.

The rest of the paper is structured as follows: Section 2 discusses the theoretical setup and the jump tests. Section 3 describes the data and methodology. Section 4 discusses empirical findings from both statistical and economic perspectives. Section 5 conducts comprehensive robustness checks. Section 6 concludes.

2 Jumps in Asset Prices

2.1 Theoretical Setup

Let p_t denote the logarithmic price which follows a jump diffusion process given by

$$dp_t = \mu_t dt + \sigma_t dW_t + dJ_t \quad (1)$$

where μ_t , σ_t , and dW_t are drift, diffusion parameter, and standard brownian motion respectively. J_t is a jump process, and $J_t = \sum_{j=1}^{N_t} c_{t_j}$, where c_{t_j} is the jump size and N_t a counting process. For simplicity, we only consider finite activity jumps and we assume that the jump and the diffusion components are independent. Let

$$r_{j,t} = p_{(t-1)h + \frac{hj}{M}} - p_{(t-1)h + \frac{h(j-1)}{M}}, \quad j = 1, \dots, M,$$

where h is the length of the intraday sampling interval and M is the number of intraday returns during the day. Then the realized variance can be written as the sum of the squared intraday returns

$$RV_t = RV_{t,M} = \sum_{j=1}^M r_{j,t}^2$$

Given the price dynamics of the jump diffusion process, the realized variance as an approximation to the price's quadratic variance can be further written as follows

$$\lim_{M \rightarrow \infty} RV_{t,M} = \int_{t-1}^t \sigma_s^2 ds + \sum_{j=1}^{N_t} c_j^2 \quad (2)$$

Here $\int_{t-1}^t \sigma_s^2 ds$ is the integrated variance (IV_t), and $\sum_{j=1}^{N_t} c_j^2$ is the quadratic variation of the jump part (JV_t) over the period from $t-1$ to t (often a day). Jump tests are therefore designed to estimate JV_t using high frequency data.

2.2 Jump Tests

We consider the use of seven major jump tests developed in the literature, including Barndorff-Nielsen and Shephard (2006) (BNS), Ait-Sahalia and Jacod (2009) (AJ), Jiang and Oomen (2008) (JO), Andersen, Dobrev, and Schaumburg (2012) (Med, Min), Corsi, Pirino, and Reno (2010) (CPR), and Podolskij and Ziggel (2010) (PZ). In this part, we only describe the BNS jump test in detail. Specifications of all other six tests are described in Appendix 1. Andersen, Bollerslev, and Dobrev (2007) and Lee and

Mykland (2008) are two other important tests that are however not studied in this paper.¹

Barndorff-Nielsen and Shephard (2006) developed a bipower based jump estimator. The idea is using the calculated realized bipower variation to proxy the integrated variance. Since jumps are rare and are unlikely to occur for two consecutive intraday returns, when intervals are small enough, the realized bipower variation will converge in probability to the integrated variance. The difference between realized variance and realized bipower variation is then an estimator of the jump variation. The realized bipower statistic is defined as

$$BV_{t,M} = \frac{\mu_1^{-2}M}{M-1} \sum_{j=2}^M |r_{t_{j-1}}||r_{t_j}|,$$

$$\lim_{M \rightarrow \infty} BV_{t,M} \rightarrow IV = \int_0^t \sigma_s^2 ds.$$

Following Huang and Tauchen (2005), the standardized $BV_{t,M}/RV_{t,M}$ ratio converges to a standard normal distribution and the test statistic is given by

$$J_{t,M} = \frac{1 - \frac{BV_{t,M}}{RV_{t,M}}}{\sqrt{[(\frac{2}{\pi})^2 + \pi - 5] \frac{1}{M} \max(1, \frac{TQ_{t,M}}{BV_{t,M}^2})}} \rightarrow N(0, 1), \quad (3)$$

where TQ_t refers to the tripower quarticity given by

$$TQ_{t,M} = M\mu_{4/3}^{-3} \frac{M}{M-2} \sum_{j=3}^n |r_{t_{j-2}}|^{4/3} |r_{t_{j-1}}|^{4/3} |r_{t_j}|^{4/3},$$

where $\mu_p = E(|U|^p) = \pi^{-1/2} 2^{p/2} \Gamma(\frac{p+1}{2})$.

The jump variation can then be obtained as

$$JV_t = (RV_t - BV_t) I_{[J_{t,M} \geq \phi_\alpha^{-1}]}, \quad (4)$$

where the ϕ_α^{-1} is the α quantile of the normal distribution.

¹The reason is that both of them are intraday jump tests that average daily bipower variation to local volatility. Therefore, although they can identify which return contains a jump, they are not significantly different from BNS in volatility forecasting when only daily jump variation is considered.

3 Data and Methodology

3.1 Realized Jumps and Volatility Forecasting

We first investigate the role of realized jumps in volatility forecasting. There are many approaches to forecast volatility using high frequency data. We use the Heterogeneous Autoregressive (HAR) model of Corsi (2009) because it can be implemented easily and it can capture the long memory property of volatility processes in a straightforward way. Our benchmark model considers the use of daily, weekly, and monthly lagged realized variances to forecast one step ahead realized variance. The HAR-RV specification is as follows:

$$RV_{t,t+h-1} = \beta_0 + \beta_{RVD}RV_{t-1} + \beta_{RVW}RV_{t-5,t-1} + \beta_{RVM}RV_{t-22,t-1} + \epsilon_{t,t+h-1}. \quad (5)$$

To assess the role of jumps, we consider the following HAR-RV-CJ specification:

$$RV_{t,t+h-1} = \beta_0 + \beta_{IVD}IV_{t-1} + \beta_{IVW}IV_{t-5,t-1} + \beta_{IVM}IV_{t-22,t-1} + \beta_{JVD}JV_{t-1} + \epsilon_{t,t+h-1}, \quad (6)$$

where $RV_{t,t+h} = h^{-1}[RV_t + RV_{t+1} \dots + RV_{t+h-1}]$ is the averaged h -periods realized variance. IV_{t-1} , $IV_{t-5,t-1}$, $IV_{t-22,t-1}$ are jump robust integrated variation estimators over a lagged daily, weekly, and monthly horizon, JV_{t-1} is the daily lagged jump variation detected using the jump tests introduced before.

3.2 Realized Jumps and Volatility Timing Based Portfolio Strategy

We then conduct our economic evaluations by constructing volatility timing based portfolio allocation strategies. We consider a risk-averse investor with mean-variance preferences, who allocates her wealth into one risky asset (a market index ETF) and one risk-free asset. Our one risky asset specification is similar to Marquering and Verbeek (2004). The economic intuition for this strategy is simple. Given the expected return, when the volatility level is high, the investor allocates more wealth into the risk-free asset, and when the volatility level is low, the investor allocates more wealth into the risky asset. If separating jumps from diffusion components lead to more accurate prediction of future volatility, then we should expect the investor to improve her portfolio performance by actively rebalancing the portfolio based on the signal of the predicted volatility.

Although more sophisticated utility functions can be used, we stick to mean-variance preferences because we are primarily interested in whether statistical improvements in volatility forecasting by separating jumps from diffusion can be translated into economic values. To concentrate on the impact of jumps, we consider only one market index as the risky asset. We choose this setting to avoid dealing with jump tests in multivariate settings and controlling for non-synchronicity of different assets, while we still maintain the generality of our empirical results. We also assume that the investor is myopic. Namely, the investor dynamically rebalances her portfolio period by period, and she does not consider the intertemporal hedging demand in the portfolio selection. We make this assumption both for simplicity to directly translate our volatility forecasting results into portfolio performance and for consistency with the existing volatility timing literature.

Hence, the investor solves the following optimization problem:

$$\text{Max}_{w_t} U[E_t(r_{p,t+1}), \text{Var}_t(r_{p,t+1})],$$

where $E_t(r_{p,t+1})$ is the conditional expected portfolio return and $\text{Var}_t(r_{p,t+1})$ is the conditional variance of the portfolio return. The portfolio return is $E_t(r_{p,t+1}) = r_{f,t+1} + w_t(E_t(r_{m,t+1}) - r_{f,t+1})$, where w_t is the portfolio weight of the risky asset, $E_t(r_{m,t+1})$ is the conditional expected return of the risky asset and $r_{f,t+1}$ is the return for the risk free asset, which we know ex ante.

The mean-variance utility function is given by

$$U[E_t(r_{p,t+1}), \text{Var}_t(r_{p,t+1})] = E_t(r_{p,t+1}) - \frac{\gamma}{2}\text{Var}_t(r_{p,t+1}),$$

with γ the risk-aversion parameter. Hence, the optimal portfolio weight is given by

$$w_t = \frac{E_t(r_{m,t+1}) - r_{f,t+1}}{\gamma \text{Var}_t(r_{m,t+1})}. \quad (7)$$

We set the expected return for the risky asset equal to the in-sample mean ², as over a short horizon, expected return changes are negligible and we are only interested in the volatility timing effect of realized jumps.

Fleming, Kirby, and Ostdiek (2003) show that using a dynamic volatility timing strategy relying solely on realized volatilities can outperform both static strategies and dynamic volatility timing strategies using lagged daily volatilities only. Therefore

²We also tried using other specifications such as rolling windows, results are similar.

the question whether realized jumps are economically valuable can be translated into whether our jump augmented volatility timing strategy (HAR-RV-CJ) can outperform the benchmark strategy (HAR-RV) which does not separate jumps from the diffusion component.

To implement the above strategy, we conduct two further adjustments. Firstly, we impose a short selling constraint. Following Marquering and Verbeek (2004), we restrict the negative portfolio weights to zero and the greater than one portfolio weights to one. Secondly, we also match the high frequency trading period (6.5 hour) to daily frequency (24 hour). Therefore, rather than directly plugging in the predicted realized variance, we adjust the predicted realized variance with a bias-correction factor. We follow existing studies (Fleming, Kirby, and Ostdiek 2003, Bandi and Russell 2006) and construct the bias-correction factor as follows:

$$BCF = \frac{1/n \sum_{t=1}^n r_t^2}{1/n \sum_{t=1}^n RV_t}, \quad (8)$$

where RV_t is the daily realized variance for 6.5 trading hours and r_t is the daily return for 24 hours. We construct the bias-correction factor using all data from the in-sample period. The conditional variance is estimated by predicted realized variance scaled by the bias correction factor, namely $Var_t(r_{m,t+1}) = BCF \cdot \widehat{RV}_{t+1}$, where \widehat{RV}_{t+1} is the predicted realized variance obtained from the volatility forecasting part. We can then plug in the conditional variance into the optimal portfolio weight function to estimate the optimal portfolio weight.

3.3 Performance Evaluations

Following the volatility timing literature, we focus on the utility based performance evaluation measure to assess whether investors can benefit from including realized jumps into their information set. We follow Fleming, Kirby, and Ostdiek (2003) and Marquering and Verbeek (2004) and rely on averaged realized utility to compare jump strategies to the benchmark strategy.

The sample averaged realized utility for portfolio strategy p is given by

$$\bar{U}(R_p) = \frac{1}{T} \sum_{t=0}^{T-1} [r_{p,t+1} - \frac{\gamma}{2} Var_t(r_{p,t+1})]. \quad (9)$$

Given the optimal portfolio weights, we can compute daily time series of ex post portfolio returns $r_{p,t+1} = r_{f,t+1} + w_t(r_{m,t+1} - r_{f,t+1})$ and variances $Var_t(r_{p,t+1}) = (r_{p,t+1} - \bar{r}_p)^2$, and then plug that in to obtain the averaged realized utility.³

To quantify the economic benefit relative to the benchmark strategy, we use the performance fee Δ (in basis points), which is the fee an investor is willing to pay to switch from the benchmark strategy (with portfolio return $r_{bm,t}$) to our strategy (e.g. strategy p with portfolio return $r_{p,t}$). In our analysis, we consider different risk aversion levels ($\gamma = 2, 6, 10$). The performance fee is computed as follows:

$$\frac{1}{T} \sum_{t=0}^{T-1} [(r_{p,t+1} - \Delta) - \frac{\gamma}{2} Var_t(r_{p,t+1})] = \frac{1}{T} \sum_{t=0}^{T-1} [r_{bm,t+1} - \frac{\gamma}{2} Var_t(r_{bm,t+1})]. \quad (10)$$

3.4 Statistical Significance of Economic Values

One of the major concerns of existing studies on the economic value of return/volatility prediction is the statistical significance of the economic value obtained. The economic value computed is just a figure and usually not large (e.g in the unit of basis points), hence we don't know whether the value is significantly different from zero across different strategies. Therefore different methods have been used to investigate the statistical significance of economic values. Following Engle and Colacito (2006) and Bandi, Russell, and Zhu (2008), we address this concern by viewing the economic gains as loss differential in which we compare one portfolio to the benchmark portfolio. The approach is in the spirit of Diebold and Mariano (1995). The Diebold-Mariano (DM) test was designed to examine whether the loss differential of two forecasts is statistically significantly different from zero. The test can be used when the loss differential series is covariance stationary. Engle and Colacito (2006) and Bandi, Russell, and Zhu (2008) applied it to examine whether the ex post portfolio-volatility-difference between a candidate strategy and a benchmark strategy is statistically significantly different from zero. In our study, we investigate whether the performance fees (viewed as loss differential) are significantly different from zero. We first compute the time series of daily "spot" realized utilities and then the time series of daily performance fees for each strategy in comparison to the benchmark. Afterwards, we construct the DM statistics and test whether the alternative strategies do not outperform the benchmark

³Alternatively, we can estimate the portfolio variance from the variance of the risky asset (we can use ex post realized variance scaled with the bias correction factor) scaled by the squared weight, $Var_t(r_{p,t+1}) = w_t^2 Var_t(r_{m,t+1}) = w_t^2 BCF \cdot RV_{t+1}$, results are similar.

(null hypothesis) using a one-sided t-test with a robust variance covariance estimator.⁴

3.5 Data Description

In our empirical analysis we use the S&P500 ETF or SPDR contract (SPY) as the risky asset, obtained from NYSE TAQ database. The contract tracks the S&P500 index and is very liquidly traded. Trading spans from 9:30 EST to 16:00 EST. Our sample spans from Jan 2nd 2001 to Dec 31st 2010. To start with we compute the realized volatility estimator based on equidistant observations sampled at the conventional five minute frequency in order to control for market microstructure noise. In the robustness checks section, we also report the results of a more advanced estimator to control for market microstructure noise – the average RV estimator of Andersen, Bollerslev, Christoffersen, and Diebold (2011). This estimator is a sub-sample estimator that can also be constructed easily. Starting from one minute regular spaced log-returns, we compute the average RV as an equally-weighted average of five overlapping five minute RV estimators. Andersen, Bollerslev, and Meddahi (2011) found that the average RV estimator can perform as well as more complex estimators (realized kernel, multiple time scale, pre-averaging etc) in volatility forecasting. We implement two further adjustments: Firstly, we remove the overnight periods. Secondly, we use linear interpolation to correct for different trading hours, especially in December 2008 and afterwards. We also collect S&P500 ETF or SPDR contract (SPY) daily data from CRSP in order to match intraday trading period to daily frequency. As the risk-free asset we use the daily average of the one month US Treasury bill series.

4 Empirical Findings

4.1 Statistical Findings

This part presents volatility forecasting results using realized jumps. Table 1 documents the descriptive statistics for the realized variance and realized jump variations using different jump tests. Although the statistics from different jump tests look different, they all share the same features, including high skewness and high kurtosis,

⁴An alternative way to assess the statistical significance of economic gains is to use bootstrapping methods. A recent study by McCracken and Valente (2013) provides a formal test of economic values using the bootstrapping method.

supporting the asymmetric and rare event nature of jumps. We first conduct in-sample statistical evaluations by estimating our models with the whole sample data from 2001 and 2010.

Table 2 shows in-sample volatility forecasting results for the benchmark model and models using different jump tests. We follow Andersen, Bollerslev, and Diebold (2007) and use the Newey-West variance covariance matrix estimator with 5, 10 and 44 lags for daily, weekly, and monthly ahead forecasts. For the benchmark HAR-RV model, the one day ahead forecast shows that only the coefficient of weekly lagged realized variance is significant at the 5% level, while for one week and one month ahead forecasts all three lagged realized variance coefficients are significant. The $adjR^2$ takes values of 0.562, 0.682 and 0.644 for the different forecasting horizons. We then look at the HAR-RV-CJ models using different jump tests. At least four points are worth mentioning. Firstly, although at weekly and monthly horizons, coefficients for the integrated variances are all significant as in the benchmark model, the HAR-RV-CJ results differ from the HAR-RV model at daily horizon. While weekly and monthly lagged integrated variances remain insignificant, the daily lagged integrated variances now become significant, indicating that jump robust integrated variation is more important than total realized variance in daily volatility forecasting. Secondly, the jump signs are almost all negative.⁵ Our result is consistent with Andersen, Bollerslev, and Diebold (2007), but is different from Corsi, Pirino, and Reno (2010). Therefore, Corsi, Pirino, and Reno (2010)'s explanation that larger jumps lead to higher future volatility due to an increased level of disagreement may not hold in our setup. Instead, our findings are more consistent with Andersen, Bollerslev, and Diebold (2007)'s explanation that jumps are quickly mean reverting and hence can lead to a lower volatility rather than a higher one. Thirdly, jump coefficients differ in terms of the significance. Jump coefficients are significant across all horizons for BNS, JO and PZ, significant only at daily horizon for Med, Min, CPR, and insignificant for AJ across all horizons at the 5% level. Finally, we look at the goodness-of-fit of the models. We find that almost all models with different jump specifications can outperform the benchmark HAR-RV model at all forecasting horizons. At daily level, the highest $adjR^2$ is PZ of 0.604 and the lowest is AJ of 0.562. BNS has an $adjR^2$ of 0.592. Compared to BNS, CPR, JO, and Med have higher $adjR^2$ s while Min has a lower $adjR^2$. For weekly and monthly ahead forecasts, $adjR^2$ s are all close to 0.70 and 0.65 respectively. Although

⁵The exception of AJ and its later relative weak statistical performance can mainly be justified by its finite sample properties in a simulation analysis reported in the Appendix.

we observe a clear inverse U shape pattern of $adjR^2$'s levels across forecasting horizons for all models, the improvements in $adjR^2$'s compared to the benchmark model are diminishing from about 3% on daily horizon to about 1% on average on weekly and monthly horizons.

Although in-sample findings document a clear improvement in volatility forecasting by separating jumps from diffusion, we are also interested in whether results hold true out-of-sample. We first estimate model parameters using the first 1000 days of the whole sample as the in-sample period, and then use the rest of the sample from 2006 to 2010 as the out-of-sample period. Table 3 reports our out-of-sample volatility forecasting results. We report the Mean Squared Error (MSE) for the predicted value compared to the realized value. Similar to the in-sample analysis, we find that all HAR-RV-CJ models using different jump tests can outperform the benchmark HAR-RV in terms of lower MSEs. This finding holds true for all daily, weekly, and monthly horizons. Similar to the in-sample findings, we observe i) the largest statistical improvements at daily horizons and ii) improvements diminish when forecasting horizons increase. When we compare out-of-sample findings across different jump tests, we find that AJ has the lowest out-of-sample performance. Models using PZ, CPR, Med, or JO outperform BNS while the model using Min underperforms BNS. Results are consistent with in-sample findings and hold true across forecasting horizons.

4.2 Out-of-Sample Economic Findings

Given the significant statistical improvement by separating jump and diffusion components, we are now interested in whether such statistical accuracy can be translated into economic value for a risk-averse investor. We construct volatility timing strategies as discussed above for our out-of-sample period (2006 to 2010). The largest statistical forecasting improvement was observed for a daily horizon and given that the jump effect is quickly mean reverting we concentrate on volatility timing with daily re-balancing. To calculate the optimal portfolio weights we use the model-predicted volatility as a predictor for conditional volatility, and then adjust it with the bias-correction factor as illustrated in equation (8).

Table 4 reports the out-of-sample economic findings. Our main performance measure is performance fee, interpreted as the fee that an investor is willing to pay to switch from a benchmark strategy to a jump augmented strategy. We consider three

risk aversion levels $\gamma = 2, 6, 10$. We show that all jump strategies generate positive performance fees in comparison to the benchmark strategy, and the economic values generated depend on different jump strategies and risk aversion levels. For the moderate risk aversion level of 6, we show that highest performance fees are 20 basis points for Med and Min, followed by 19 basis points for BNS and CPR, and 18 and 17 basis points for JO and PZ. AJ generates positive but very small performance fee; a result that is consistent with its negligible forecasting improvement in the statistical part. The economic magnitude is also affected by the change of the risk aversion level, ranging from 59 basis points ($\gamma = 2$) to 11 basis points ($\gamma = 10$), indicating that the strategy seems to work better for less risk averse investors. Around 0.6% annualized performance fee looks small in magnitude, and we therefore also assess the statistical significance of the economic value generated. We find that except for AJ, all jump strategies generate positive and statistically significant performance fees with DM t-statistics above 2. To summarize, we find that the separation of jumps from diffusion components improves volatility timing strategies for almost all jump tests. The out-of-sample economic findings are generally consistent with in-sample and out-of-sample statistical findings, although it does not necessarily match with the ranking of the in and out-of-sample volatility forecasting analyses. One possible explanation could be that jumps not only affect the volatility process, but also the return process, which is not captured by our volatility timing strategies.

5 Robustness Checks

In this section, we conduct comprehensive robustness checks. We focus on three issues: Firstly, our main results are based on the RV estimator sampled as the conventional five minutes sampling frequency. It is interesting to see whether our results still hold true under a more stringent control of market microstructure noise. Secondly, although we show that incorporating realized jumps in volatility timing generates economic value, we are also interested in whether it is feasible for an investor to exploit this in the presence of transaction costs. Thirdly, we also discuss whether realized jumps can help to predict realized higher moments and semi-variances, and whether performances can be improved using these in the portfolio allocation. Further extensions and robustness checks including simulation analysis, good and bad jumps,

and sub-sample analysis can be found in the Appendix⁶.

5.1 Market Microstructure Noises

We follow Andersen, Bollerslev, Christoffersen, and Diebold (2011) and Andersen, Bollerslev, and Meddahi (2011) and construct average realized variances and bipower variations. Table 5 reports the in-sample volatility forecasting results after further controlling for market microstructure noises. Although the jump coefficient is still negative and significant as shown in Section 3.2, the $adjR^2$ is different. For the one day ahead forecast, the $adjR^2$ for the benchmark model raises from 0.562 to 0.588 when using the average RV estimator. Similarly, the adjusted $adjR^2$ for the HAR-RV-CJ with BNS raises from 0.592 to 0.628. A similar statistical improvement is also found in the out-of-sample evaluation as shown in panel 1 of Table 6. Such statistical improvements by using subsample estimators also indicate potential economic improvements. The out-of-sample portfolio allocation results are shown in panel 2 of Table 6. We find that performance fees remain positive and statistically significant. Moreover, we show that economic magnitudes are larger using the microstructure noise robust estimators compared to the conventional five minutes estimator. A risk-averse investor is willing to pay performance fees ranging from 62 basis points ($\gamma = 2$) to 12 basis points ($\gamma = 10$) to use a jump strategy. Our findings suggest that the statistical and economic improvements by separating jumps from diffusion are not likely to be driven by market microstructure noises. Instead, we show that controlling for market microstructure noises strengthens our findings. Our results are also consistent with previous studies (Bandi and Russell 2006, Bandi, Russell, and Zhu 2008, Liu 2009) that controlling for microstructure noises improve portfolio performances.

5.2 Transaction Costs

We then analyze the impact of transaction costs on our results. Different from existing studies comparing dynamic and static strategies (Fleming, Kirby, and Ostdiek 2001) or comparing two dynamic strategies which are based on high frequency and daily information respectively (Fleming, Kirby, and Ostdiek 2003), our analysis compares two dynamic strategies both using high frequency information. Therefore, we expect that the effect of transaction costs will not be as strong as documented in the existing

⁶In our earlier version of the paper, we also show the results of an alternative parametric portfolio allocation strategies using realized jumps, the economic values are positive but small in magnitude and in general statistically insignificant.

literature. Following Bandi, Russell, and Zhu (2008), we define the transaction cost adjusted portfolio return in the following way:

$$\bar{r}_{p,t+1} = r_{p,t+1} - \rho(1 + r_{p,t+1})|\Delta w_{t+1}|, \quad (11)$$

where $\bar{r}_{p,t+1}$ is the transaction cost adjusted portfolio return, $r_{p,t+1}$ is pre-adjusted return, ρ is transaction cost parameter, where we choose a high value of 0.0025, corresponding to a 2.5 cent half spread on a 10 dollar stock, Δw_{t+1} is the change of the weight from t to $t + 1$, a proxy of trading turnover.

Table 7 presents the out-of-sample volatility timing results for all jump tests when transaction costs are taken into consideration. All specifications yield positive performance fees, implying that they can outperform the benchmark even when transaction costs are considered. Moreover, those performance fees are also statistically significant except for AJ. Jump strategies require incorporating recent information more quickly while the benchmark strategy is smoother, therefore we are not surprised by the higher turnover of the jump strategies compared to the benchmark strategy. Although the performance fees are slightly lower when controlling for transaction costs, we find that our results are generally consistent with our main findings in Section 4.4. Moreover, the relative performance of alternative jump specifications is also consistent with that in the main analysis, indicating that transaction costs have similar and only marginal effects for most of jump based volatility timing strategies.

5.3 Realized Jumps and Alternative Realized Moments

In practice, portfolio allocations are also subject to the impact of higher moments. Considering more general utility functions usually requires the prediction of higher moments. In addition to the sophistication of incorporating realized jumps into the portfolio allocation problem beyond mean-variance preferences, we present some statistical evidence of the predictive ability of realized jumps for alternative realized moments. We consider the use of realized jumps for predicting realized upside and downside volatilities, skewness, and kurtosis. We follow Barndorff-Nielsen, Kinnebrock, and Shephard (2008) and Amaya, Christoffersen, Jacobs, and Vasquez (2013), and

construct alternative realized moments in the following way,

$$\begin{aligned}
RV_{t,M}^+ &= \sum_{j=1}^M r_{j,t}^2 1_{r_{j,t}>0}, & RV_{t,M}^- &= \sum_{j=1}^M r_{j,t}^2 1_{r_{j,t}<0} \\
RSK_{t,M} &= \frac{\sqrt{M} \sum_{j=1}^M r_{j,t}^3}{(\sum_{j=1}^M r_{j,t}^2)^{3/2}}, & RKU_{t,M} &= \frac{M \sum_{j=1}^M r_{j,t}^4}{(\sum_{j=1}^M r_{j,t}^2)^2}
\end{aligned}$$

We first investigate the contemporaneous relationship between realized jumps and alternative realized moments using the following regression equation:

$$RM_t = \beta_0 + \beta_{RJ} RJ_t + \epsilon_t, \quad (12)$$

where RM_t is the realized moment including RV , RV^+ , RV^- , RSK , and RKU . Table 8 presents the contemporaneous regression results. We find that realized jump variation is a significant determinant of contemporaneous realized variance, positive and negative variances, and kurtosis, explaining 18% to 80% variation of realized moments respectively. A large jump variation is associated with a large variance, upside and downside variance, and kurtosis. Jump variation is also negatively related to realized skewness, however the relation is only marginally significant, and jumps can only explain about 1% variation in skewness.

We are more interested in the predictive relationship of realized jump variation and realized moments. Therefore we forecast realized moments using daily, weekly, and monthly lagged realized moments in the fashion of the HAR model, just as we did for realized variance in Section 4.1. We consider daily, weekly, and monthly ahead forecasting horizons. The models are specified as follows:

$$RM_{t,t+h-1} = \beta_0 + \beta_{RMD} RM_{t-1} + \beta_{RMW} RM_{t-5,t-1} + \beta_{RMM} RM_{t-22,t-1} + \epsilon_{t,t+h-1}. \quad (13)$$

To investigate the impact of jump, we then augment the model with realized jumps.

$$RM_{t,t+h-1} = \beta_0 + \beta_{RMD} RM_{t-1} + \beta_{RMW} RM_{t-5,t-1} + \beta_{RMM} RM_{t-22,t-1} + \beta_{JVD} JV_{t-1} + \epsilon_{t,t+h-1}. \quad (14)$$

Table 9 reports in-sample forecasting results for realized moments. The results vary across different realized moments and forecasting horizons. Firstly, we find that realized jump helps to forecast realized variance, even though we do not use jump robust integrated variance as we did in the main part of the analysis. For all the forecasting

horizons, we observe negative and statistically significant jump coefficients. Moreover, the adjusted R squares of the models including jumps are higher than those of the models without jumps. This finding suggests that jumps do contain incremental information for predicting future volatility and the statistical and economic improvements we documented in the main analysis do not purely come from a better measurement of jump robust integrated variance. We now discuss the role of jumps in predicting alternative realized moments: We find that realized jumps have negative and statistically significant impacts on future downside volatility for all forecasting horizons and improvements in adjusted R squares range from 1% to 3%. We also find that realized jumps help to predict upside volatility at the daily horizon and generate improvements in adjusted R squares of about 0.2%. However, we show that realized jumps do not predict future realized skewness. Although realized jumps predict realized kurtosis at daily and weekly horizons, improvements in adjusted R squares are almost negligible. These results, suggest that realized jumps may not contain predictive information beyond second moments at least in our empirical setups.

We then consider a simple portfolio allocation within the mean variance framework to quantify the predictive ability of realized jumps on alternative realized moments. We focus on upside and downside volatilities. Since the optimal portfolio weight is a function of the conditional variance of the risky asset, we can decompose it, as:

$$Var_t(r_{m,t+1}) = Var_t(r_{m,t+1})^+ + Var_t(r_{m,t+1})^- = BCF(\hat{RV}_{t,t+1}^+ + \hat{RV}_{t,t+1}^-). \quad (15)$$

Since jumps have different predictive abilities to forecast upside and downside volatilities, an investor may improve portfolio performances by forecasting $RV_{t,t+1}^+$ and $RV_{t,t+1}^-$ separately and combining and scaling them by the Bias Correction Factor to obtain the total conditional variance $Var_t(r_{m,t+1})$, which can then be plugged into the portfolio weights function as shown in equation (7). We construct two portfolio strategies: The first strategy is based on predicting upside and downside volatilities with their lagged values in the HAR fashion. To improve forecasting performance, we include both upside and downside components at daily, weekly, and monthly lagged levels to predict each component one day ahead.

$$RV_{t,t+h-1}^+ = \beta_0 + \beta_{RVPD}RV_{t-1}^+ + \beta_{RVMD}RV_{t-1}^- + \beta_{RVPW}RV_{t-5,t-1}^+ + \beta_{RVMW}RV_{t-5,t-1}^- + \beta_{RVPM}RV_{t-22,t-1}^+ + \beta_{RVMM}RV_{t-22,t-1}^- + \epsilon_{t,t+h-1}, \quad (16)$$

$$RV_{t,t+h-1}^- = \beta_0 + \beta_{RVPD}RV_{t-1}^+ + \beta_{RVMD}RV_{t-1}^- + \beta_{RVPW}RV_{t-5,t-1}^+ + \beta_{RVMW}RV_{t-5,t-1}^- + \beta_{RVPM}RV_{t-22,t-1}^+ + \beta_{RVMM}RV_{t-22,t-1}^- + \epsilon_{t,t+h-1}. \quad (17)$$

The second strategy augments the first strategy with daily lagged realized jump variation as additional regressor.

$$RV_{t,t+h-1}^+ = \beta_0 + \beta_{RVPD}RV_{t-1}^+ + \beta_{RVMD}RV_{t-1}^- + \beta_{RVPW}RV_{t-5,t-1}^+ + \beta_{RVMW}RV_{t-5,t-1}^- + \beta_{RVPM}RV_{t-22,t-1}^+ + \beta_{RVMM}RV_{t-22,t-1}^- + \beta_{JVD}JV_{t-1} + \epsilon_{t,t+h-1}, \quad (18)$$

$$RV_{t,t+h-1}^- = \beta_0 + \beta_{RVPD}RV_{t-1}^+ + \beta_{RVMD}RV_{t-1}^- + \beta_{RVPW}RV_{t-5,t-1}^+ + \beta_{RVMW}RV_{t-5,t-1}^- + \beta_{RVPM}RV_{t-22,t-1}^+ + \beta_{RVMM}RV_{t-22,t-1}^- + \beta_{JVD}JV_{t-1} + \epsilon_{t,t+h-1}. \quad (19)$$

We focus on out-of-sample performances and consider three comparisons. The first comparison is between the first strategy based on equations (16) and (17) and the benchmark strategy in the main analysis, which predicts the total variance using HAR-RV in equation (5). The purpose is to assess whether predicting each volatility component separately can be economically valuable in comparison to predicting the total volatility. The second comparison is between the second strategy using jumps in equations (18) and (19) and the benchmark strategy in equation (5). The third comparison is between the first and second strategies, showing whether jumps convey incremental economic improvements. Table 10 reports out-of-sample portfolio performance fees for those three cases. We find that strategies based on predicting each volatility component separately outperform the benchmark strategy based on predicting total volatility, and can generate positive and statistically significant economic values. To be specific, in the first comparison, if we only use lagged upside and downside volatilities, we can generate annualized performance fees ranging from 13 basis points ($\gamma = 2$) to 2 basis points ($\gamma = 10$). If we include realized jumps, then the performance fees increase to range from 45 basis points ($\gamma = 2$) to 8 basis points ($\gamma = 10$). The third comparison suggests that including jump is important and can generate incremental economic improvements from 31 basis points ($\gamma = 2$) to 6 basis points ($\gamma = 10$).

To summarize, we show that jumps do contain incremental predictive information for future volatility and its signed components, however realized jumps can hardly predict future realized higher moments. Therefore, the results suggest that realized jumps do not contribute much to moment timing based portfolio strategies beyond mean-variance approaches. If we remain in the mean-variance framework, predicting positive and negative volatility components separately can generate tangible economic improvements compared to predicting total volatility, and incorporating jumps can further improve the magnitude of the economic value.

6 Conclusion

Although a number of different nonparametric jump tests were developed in the literature, only very few studies analyze the potential use of realized jumps. Using high frequency data and seven major nonparametric jump tests, this paper investigates the predictive information content of realized jumps on volatility timing from both statistical and economic perspectives.

Covering all major jump tests, we confirm that separating jumps from the diffusion component does improve volatility forecasting in general. The result holds true both in-sample and out-of-sample. Moreover, we show that using a simple volatility timing strategy, a risk-averse investor can generate a significant economic value by separating jumps from the diffusion component. We conduct comprehensive robustness checks. We show that after controlling for microstructure noise and transaction costs, our main results still hold. We also find that realized jumps can predict realized volatility and its signed up and down components, and portfolio performance can be improved by separately predicting each component.

Our paper contributes to the field on a few aspects: Firstly, we contribute to the existing literature on the role of jumps in volatility forecasting. By using seven different jump tests, we show that jumps in general help to forecast volatility. Secondly, we show that the statistical improvement can also be exploited in a mean-variance portfolio allocation strategy. Hence, we also contribute to the literature on economic value of volatility timing. Thirdly, we contribute to the literature on the use of high frequency data and nonparametric jump tests. Our study can be viewed as evaluations of alternative jump tests using real world data while most previous studies focus on simulations. Further extensions include dealing with multivariate jumps (co-jumps), using more sophisticated utility functions, and considering alternative economic applications. They are beyond the purpose of our paper, and we leave them to future studies.

References

AIT-SAHALIA, Y. (2004): “Disentangling diffusion from jumps,” *Journal of Financial Economics*, 74(3), 487–528.

- AIT-SAHALIA, Y., AND J. JACOD (2009): “Testing for Jumps in a Discretely Observed Process,” *Annal of Statistics*, 37(1), 184–222.
- AMAYA, D., P. CHRISTOFFERSEN, K. JACOBS, AND A. VASQUEZ (2013): “Does Realized Skewness Predict the Cross-Section of Equity Returns?,” Creates research papers, School of Economics and Management, University of Aarhus.
- ANDERSEN, T. G., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2011): “Financial Risk Measurement for Financial Risk Management,” Creates research papers, School of Economics and Management, University of Aarhus.
- ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2007): “Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility,” *The Review of Economics and Statistics*, 89(4), 701–720.
- ANDERSEN, T. G., T. BOLLERSLEV, AND D. DOBREV (2007): “No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications,” *Journal of Econometrics*, 138(1), 125–180.
- ANDERSEN, T. G., T. BOLLERSLEV, AND N. MEDDAHI (2011): “Realized volatility forecasting and market microstructure noise,” *Journal of Econometrics*, 160(1), 220–234.
- ANDERSEN, T. G., D. DOBREV, AND E. SCHAUMBURG (2012): “Jump-Robust Volatility Estimation using Nearest Neighbor Truncation,” *Journal of Econometrics*, 169(1), 75–93.
- BACKUS, D., M. CHERNOV, AND I. MARTIN (2011): “Disasters Implied by Equity Index Options,” *Journal of Finance*, 66(6), 1969–2012.
- BANDI, F., J. RUSSELL, AND Y. ZHU (2008): “Using High-Frequency Data in Dynamic Portfolio Choice,” *Econometric Reviews*, 27(1-3), 163–198.
- BANDI, F. M., AND J. R. RUSSELL (2006): “Separating microstructure noise from volatility,” *Journal of Financial Economics*, 79(3), 655–692.
- BARNDORFF-NIELSEN, O. E., S. KINNEBROCK, AND N. SHEPHARD (2008): “Measuring downside risk — realised semivariance,” Creates research papers, School of Economics and Management, University of Aarhus.

- BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2004): “Power and Bipower Variation with Stochastic Volatility and Jumps,” *Journal of Financial Econometrics*, 2(1), 1–37.
- (2006): “Impact of jumps on returns and realised variances: econometric analysis of time-deformed Levy processes,” *Journal of Econometrics*, 131(1-2), 217–252.
- CHEN, K., S. HYDE, AND S.-H. POON (2010): “Economic Implications of Extraordinary Movements in Stock Markets,” Working paper, University of Manchester.
- CHERNOV, M., R. GALLANT, E. GHYSELS, AND G. TAUCHEN (2003): “Alternative models for stock price dynamics,” *Journal of Econometrics*, 116(1-2), 225–257.
- CHRISTOFFERSEN, P., K. JACOBS, AND C. ORNTHANALAI (2012): “Dynamic jump-intensities and risk premiums: Evidence from S&P500 returns and options,” *Journal of Financial Economics*, 106(3), 447–472.
- CORSI, F. (2009): “A Simple Approximate Long-Memory Model of Realized Volatility,” *Journal of Financial Econometrics*, 7(2), 174–196.
- CORSI, F., D. PIRINO, AND R. RENO (2010): “Threshold bipower variation and the impact of jumps on volatility forecasting,” *Journal of Econometrics*, 159(2), 276–288.
- DIEBOLD, F. X., AND R. S. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13(3), 253–63.
- DUAN, J.-C., P. RITCHKEN, AND Z. SUN (2006): “Approximating Garch-Jump Models, Jump-Diffusion Processes, And Option Pricing,” *Mathematical Finance*, 16(1), 21–52.
- DUMITRU, A., AND G. URGA (2012): “Identifying Jumps in Financial Assets: a Comparison between Nonparametric Jump Tests,” *Journal of Business and Economic Statistics*, 30(02), 242–255.
- ENGLE, R. F., AND R. COLACITO (2006): “Testing and Valuing Dynamic Correlations for Asset Allocation,” *Journal of Business and Economic Statistics*, 24, 238–253.
- ERAKER, B. (2004): “Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices,” *Journal of Finance*, 59(3), 1367–1404.

- ERAKER, B., M. JOHANNES, AND N. POLSON (2003): “The Impact of Jumps in Volatility and Returns,” *Journal of Finance*, 58(3), 1269–1300.
- FLEMING, J., C. KIRBY, AND B. OSTDIEK (2001): “The Economic Value of Volatility Timing,” *Journal of Finance*, 56(1), 329–352.
- (2003): “The Economic Value of Volatility Timing Using “Realized Volatility,”” *Journal of Financial Economics*, 67(3), 473–509.
- HUANG, X., AND G. TAUCHEN (2005): “The Relative Contribution of Jumps to Total Price Variance,” *Journal of Financial Econometrics*, 3(4), 456–499.
- JIANG, G., AND T. YAO (2012): “Stock Price Jumps and Cross-Sectional Return Predictability,” *Journal of Financial and Quantitative Analysis*, p. forthcoming.
- JIANG, G. J., AND R. C. OOMEN (2008): “Testing for jumps when asset prices are observed with noise—a “swap variance” approach,” *Journal of Econometrics*, 144(2), 352–370.
- LEE, S. S., AND P. A. MYKLAND (2008): “Jumps in Financial Markets: A New Non-parametric Test and Jump Dynamics,” *Review of Financial Studies*, 21(6), 2535–2563.
- LIU, J., F. A. LONGSTAFF, AND J. PAN (2003): “Dynamic Asset Allocation with Event Risk,” *Journal of Finance*, 58(1), 231–259.
- LIU, Q. (2009): “On portfolio optimization: How and when do we benefit from high-frequency data?,” *Journal of Applied Econometrics*, 24(4), 560–582.
- MAHEU, J. M., AND T. H. MCCURDY (2004): “News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns,” *Journal of Finance*, 59(2), 755–793.
- MAHEU, J. M., T. H. MCCURDY, AND X. ZHAO (2012): “Do Jumps Contribute to the Dynamics of the Equity Premium?,” Working paper, University of Toronto.
- MARQUERING, W., AND M. VERBEEK (2004): “The Economic Value of Predicting Stock Index Returns and Volatility,” *Journal of Financial and Quantitative Analysis*, 39(02), 407–429.
- MCCRACKEN, M., AND G. VALENTE (2013): “Testing Economic Value of Asset Return Predictability,” working paper, Federal Reserve Bank of St. Louis.

- MERTON, R. C. (1976): “Option pricing when underlying stock returns are discontinuous,” *Journal of Financial Economics*, 3(1-2), 125–144.
- PATTON, A., AND K. SHEPPARD (2011): “Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility,” working paper, Duke University.
- PODOLSKIJ, M., AND D. ZIGGEL (2010): “New tests for jumps in semimartingale models,” *Statistical Inference for Stochastic Processes*, 13(1), 15–41.
- TAUCHEN, G., AND H. ZHOU (2011): “Realized jumps on financial markets and predicting credit spreads,” *Journal of Econometrics*, 160(1), 102–118.
- THEODOSIOU, M., AND F. ZIKES (2011): “A Comprehensive Comparison of Nonparametric Tests for Jumps in Asset Prices,” Working paper, Imperial College London.

Tables

Table 1: Descriptive Statistics

	<i>RV</i>	<i>JV_{BNS}</i>	<i>JV_{AJ}</i>	<i>JV_{JO}</i>	<i>JV_{Med}</i>	<i>JV_{Min}</i>	<i>JV_{CPR}</i>	<i>JV_{PZ}</i>
<i>Mean</i>	1.324E-4	4.693E-6	1.313E-6	4.919E-6	6.377E-6	5.906E-6	8.223E-6	6.855E-6
<i>Std</i>	3.109E-4	7.692E-5	1.748E-5	7.689E-5	9.836E-5	9.805E-5	1.042E-4	7.836E-5
<i>Skew</i>	10.321	43.829	40.756	43.881	42.242	43.188	40.479	41.555
<i>Kurt</i>	158.187	2.056E+3	1.867E+3	2.060E+3	1.945E+3	2.040E+3	1.827E+3	1.906E+3
<i>Min</i>	3.468E-6	0	0	0	0	0	0	0
<i>Max</i>	0.0065	0.0037	8.131E-4	0.0037	0.0046	0.0046	0.0048	0.0037

Notes: The table summarizes the main descriptive statistics. We report realized variance (RV), and realized jump variations for seven different jump tests (BNS, AJ, JO, Med, Min, CPR, PZ). The sample period spans from Jan 1st 2001 to Dec 31st 2010. Both realized variance and realized jump variations are computed as shown in Section 2 using the 5 minutes high frequency Spyder contract (SPY) tracking the S&P500 index.

Table 2: In-Sample Volatility Forecasting Results

$$RV_{t,t+h-1} = \beta_0 + \beta_{RVD}RV_{t-1} + \beta_{RVW}RV_{t-5,t-1} + \beta_{RVM}RV_{t-22,t-1} + \epsilon_{t,t+h-1}$$

$$RV_{t,t+h-1} = \beta_0 + \beta_{IVD}IV_{t-1} + \beta_{IDW}IV_{t-5,t-1} + \beta_{IVM}IV_{t-22,t-1} + \beta_{JVD}JV_{t-1} + \epsilon_{t,t+h-1}$$

	β_0	β_{RVD}	β_{RVW}	β_{RVM}	$adjR^2$	
<i>BM</i>						
$h = 1$	1.436E-5 (1.953)	0.278 (1.887)	0.425 (4.115)	0.187 (1.767)	0.562	
$h = 5$	2.204E-5 (2.779)	0.180 (2.029)	0.370 (4.643)	0.282 (2.568)	0.682	
$h = 22$	4.103E-5 (4.249)	0.092 (2.211)	0.290 (2.951)	0.303 (3.401)	0.644	
	β_0	β_{IVD}	β_{IVW}	β_{IVM}	β_{JVD}	$adjR^2$
<i>BNS</i>						
$h = 1$	1.601E-5 (2.548)	0.467 (3.897)	0.325 (3.745)	0.132 (1.372)	-0.359 (-4.121)	0.592
$h = 5$	2.354E-5 (3.130)	0.279 (3.197)	0.335 (3.840)	0.242 (2.143)	-0.158 (-2.451)	0.699
$h = 22$	4.223E-5 (4.256)	0.145 (4.390)	0.284 (3.840)	0.275 (2.855)	-0.097 (-2.001)	0.654
<i>AJ</i>						
$h = 1$	1.362E-5 (1.845)	0.280 (1.883)	0.419 (4.061)	0.188 (1.791)	2.931 (1.641)	0.562
$h = 5$	1.991E-5 (2.436)	0.177 (2.038)	0.369 (4.544)	0.276 (2.471)	6.662 (1.608)	0.693
$h = 22$	3.996E-5 (4.261)	0.091 (2.213)	0.290 (2.944)	0.299 (3.306)	3.599 (1.647)	0.648
<i>JO</i>						
$h = 1$	1.532E-5 (2.503)	0.466 (3.848)	0.338 (3.958)	0.128 (1.349)	-0.366 (-4.093)	0.593
$h = 5$	2.306E-5 (3.173)	0.287 (3.161)	0.337 (4.039)	0.239 (2.175)	-0.195 (-2.936)	0.702
$h = 22$	4.187E-5 (4.219)	0.147 (4.418)	0.281 (2.806)	0.279 (2.920)	-0.099 (-2.258)	0.654
<i>Med</i>						
$h = 1$	1.594E-5 (2.576)	0.489 (4.040)	0.321 (1.283)	0.124 (1.283)	-0.214 (-5.021)	0.597

	β_0	β_{IVD}	β_{IVW}	β_{IVM}	β_{JVD}	$adjR^2$
$h = 5$	2.368E-5	0.297	0.313	0.254	-0.062	0.697
	(3.181)	(3.022)	(2.358)	(2.358)	(-1.712)	
$h = 22$	4.224E-5	0.146	0.280	0.285	-0.018	0.652
	(4.253)	(4.365)	(2.962)	(2.962)	(-0.362)	
<i>Min</i>						
$h = 1$	1.670E-5	0.464	0.337	0.121	-0.162	0.592
	(2.664)	(3.887)	(3.894)	(1.283)	(-3.046)	
$h = 5$	2.431E-5	0.275	0.352	0.228	-0.049	0.699
	(3.262)	(3.174)	(3.971)	(2.010)	(-1.355)	
$h = 22$	4.307E-5	0.143	0.294	0.265	-0.024	0.653
	(4.261)	(4.569)	(2.817)	(2.748)	(-0.617)	
<i>CPR</i>						
$h = 1$	1.725E-5	0.509	0.319	0.108	-0.188	0.598
	(2.838)	(4.019)	(3.402)	(1.133)	(-3.836)	
$h = 5$	2.490E-5	0.311	0.324	0.232	-0.070	0.702
	(3.334)	(3.327)	(3.580)	(2.057)	(-1.849)	
$h = 22$	4.333E-5	0.154	0.285	0.272	-0.021	0.653
	(4.276)	(4.932)	(2.807)	(2.787)	(-0.505)	
<i>PZ</i>						
$h = 1$	1.665E-5	0.537	0.316	0.095	-0.498	0.604
	(2.884)	(4.311)	(3.500)	(1.052)	(-4.035)	
$h = 5$	2.406E-5	0.320	0.336	0.219	-0.248	0.705
	(3.319)	(3.306)	(3.717)	(2.001)	(-2.481)	
$h = 22$	4.268E-5	0.165	0.288	0.263	-0.133	0.655
	(4.261)	(4.781)	(2.776)	(2.621)	(-3.403)	

Notes: The table reports in-sample volatility forecasting results for the SPY contract from 2001 to 2010 using different jump test. The HAR-RV model is used as the benchmark model. The HAR-RV-CJ models use realized jump variation detected using different jump tests: BNS, AJ, JO, Med, Min, CPR, and PZ. We forecast one day, one week, and one month ahead realized variances. The figures in parentheses are t-statistics computed using Newey-West corrected standard error for autocorrelation orders 5, 10, and 44 respectively. The $adjR^2$ is the adjusted R square.

Table 3: Out-of-Sample Volatility Forecasting Results

MSE	h=1	h=5	h=22
<i>BM</i>	5.657E-8	3.532E-8	3.504E-8
<i>BNS</i>	4.830E-8	3.195E-8	3.268E-8
<i>AJ</i>	5.656E-8	3.517E-8	3.483E-8
<i>JO</i>	4.794E-8	3.150E-8	3.273E-8
<i>Med</i>	4.733E-8	3.219E-8	3.258E-8
<i>Min</i>	4.881E-8	3.179E-8	3.238E-8
<i>CPR</i>	4.653E-8	3.107E-8	3.222E-8
<i>PZ</i>	4.523E-8	3.017E-8	3.214E-8

Notes: The table reports out-of-sample volatility forecasting results for the SPY contract for 2006 to 2010 using alternative jump tests. The HAR-RV model is used as the benchmark model. The HAR-RV-CJ model use realized jump variation detected using different jump tests: BNS, AJ, JO, Med, Min, CPR, and PZ. The out-of-sample period ranges from 2006 to 2010. We report the Mean Squared Error (MSE) for predicted volatility over one day, one week, and one month forecasting horizons.

Table 4: Out-of-Sample Portfolio Performance: Daily Rebalancing

Strategies	Volatility Timing: Performance Fees		
	$\gamma = 2$	$\gamma = 6$	$\gamma = 10$
<i>BNS</i>	0.0056 (2.5684)	0.0019 (2.5332)	0.0011 (2.4980)
<i>AJ</i>	3.2391E-5 (0.2951)	1.0797E-5 (0.2895)	6.4781E-6 (0.2839)
<i>JO</i>	0.0055 (2.5355)	0.0018 (2.5003)	0.0011 (2.4655)
<i>Med</i>	0.0059 (2.6870)	0.0020 (2.6521)	0.0012 (2.6171)
<i>Min</i>	0.0059 (2.5301)	0.0020 (2.4944)	0.0012 (2.4586)
<i>CPR</i>	0.0056 (2.6071)	0.0019 (2.5713)	0.0011 (2.5355)
<i>PZ</i>	0.0046 (2.2351)	0.0017 (2.2001)	9.5992E-4 (2.1652)

Notes: The table reports out-of-sample portfolio allocation results for the SPY contract for 2006 to 2010 using alternative jump tests. The benchmark strategy uses the HAR-RV model to predict one day ahead volatility, other strategies using HAR-RV-CJ models with realized jump variation detected from the respective jump tests. We report performance fees relative to the benchmark strategy under risk aversion levels of 2, 6, and 10. Figures in parentheses are t-statistics for the DM test, where under the null hypothesis it is assumed that the (mean) performance fee equals zero.

Table 5: In-Sample Volatility Forecasting: Average RV

$$RV_{t,t+h-1} = \beta_0 + \beta_{RVD}RV_{t-1} + \beta_{RVW}RV_{t-5,t-1} + \beta_{RVM}RV_{t-22,t-1} + \epsilon_{t,t+h-1}$$

$$RV_{t,t+h-1} = \beta_0 + \beta_{IVD}IV_{t-1} + \beta_{IDW}IV_{t-5,t-1} + \beta_{IVM}IV_{t-22,t-1} + \beta_{JVD}JV_{t-1} + \epsilon_{t,t+h-1}$$

	β_0	β_{RVD}	β_{RVW}	β_{RVM}		$adjR^2$
<i>BM</i>						
$h = 1$	1.348E-5 (1.897)	0.330 (1.887)	0.369 (2.984)	0.197 (1.936)		0.588
$h = 5$	2.150E-5 (2.834)	0.207 (1.971)	0.331 (3.869)	0.294 (2.647)		0.685
$h = 22$	3.960E-5 (4.202)	0.108 (1.997)	0.268 (2.738)	0.313 (3.311)		0.651
	β_0	β_{IVD}	β_{IVW}	β_{IVM}	β_{JVD}	$adjR^2$
<i>BNS</i>						
$h = 1$	1.066E-5 (1.834)	0.669 (5.556)	0.195 (1.891)	0.097 (1.098)	-0.176 (-6.362)	0.628
$h = 5$	1.966E-5 (2.645)	0.372 (4.442)	0.292 (2.916)	0.225 (1.911)	-0.054 (-2.624)	0.714
$h = 22$	3.915E-5 (4.031)	0.195 (5.958)	0.274 (2.420)	0.262 (2.368)	-0.051 (-3.238)	0.664

Notes: The table reports in-sample volatility forecasting results for the SPY contract from 2001 to 2010 controlling for market microstructure noises. The HAR-RV model and the HAR-RV-CJ model using a BNS jump test are applied to forecast one day, one week, and one month ahead realized variances. The average realized variance is used in order to control for market microstructure noise. The figures in parentheses are t-statistics using Newey-West standard errors for autocorrelation orders of 5, 10, and 44. $adjR^2$ is the adjusted R square.

Table 6: Out-of-Sample Statistical and Economic Performances: Average RV

Panel 1: Volatility Forecasting			
MSE	h=1	h=5	h=22
<i>BM</i>	4.501E-8	3.246E-8	3.231E-8
<i>BNS</i>	3.262E-8	2.764E-8	2.846E-8
Panel 2: Volatility Timing: Daily			
Strategies	$\gamma = 2$	$\gamma = 6$	$\gamma = 10$
<i>BNS</i>	0.0062	0.0021	0.0012
	(2.7019)	(2.6660)	(2.6301)

Notes: The table reports out-of-sample statistical and economic performances from 2006 to 2010 after controlling for market microstructure noises. Average realized variance, integrated variance, and jump variation are used to control for market microstructure noise. Panel 1 reports out-of-sample volatility forecasting results. The HAR-RV model is used as the benchmark model. The HAR-RV-CJ model using a BNS jumps is also used. We report the Mean Squared Error for predicted volatility over one day, one week, and one month forecasting horizons. Panel 2 reports out-of-sample volatility timing results using the average RV estimator. Parameters are all estimated in-sample (2001-2005) and out-of-sample volatility forecasting and volatility timing are conducted out-of-sample (2006-2010). We report performance fees relative to the benchmark strategy under risk aversion levels of 2, 6, and 10. Figures in parentheses are t-statistics for the DM test, where under the null hypothesis it is assumed that the (mean) performance fee equals zero.

Table 7: Out-of-Sample Portfolio Allocation with Transaction Costs

Strategies	$\gamma = 2$		$\gamma = 6$		$\gamma = 10$	
	PF	TO	PF	TO	PF	TO
<i>BM</i>		0.0040		0.0013		7.9190E-4
<i>BNS</i>	0.0054 (2.5088)	0.0267	0.0018 (2.4736)	0.0089	0.0011 (2.4385)	0.0053
<i>AJ</i>	3.1067E-5 (0.2827)	0.0044	1.0356E-5 (0.2772)	0.0015	6.2134E-6 (0.2716)	8.8352E-4
<i>JO</i>	0.0053 (2.4767)	0.0251	0.0018 (2.4415)	0.0084	0.0011 (2.4064)	0.0050
<i>Med</i>	0.0058 (2.6438)	0.0263	0.0019 (2.6109)	0.0088	0.0012 (2.5760)	0.0053
<i>Min</i>	0.0054 (2.5083)	0.0284	0.0018 (2.4732)	0.0095	0.0011 (2.4381)	0.0057
<i>CPR</i>	0.0055 (2.5537)	0.0254	0.0018 (2.5180)	0.0085	0.0011 (2.4823)	0.0051
<i>PZ</i>	0.0045 (2.1772)	0.0267	0.0015 (2.1427)	0.0088	9.1207E-4 (2.1081)	0.0053

Notes: The table reports out-of-sample portfolio allocation results for the SPY contract for 2006 to 2010 controlling for transaction costs. The parameters are estimated in-sample from 2001 to 2005 and the ex post portfolio returns are obtained out-of-sample (2006 to 2010). We report performance fees (PF) and turnovers (TO) for risk aversion level of 2, 6, and 10. Figures in parentheses are t-statistics for the DM test. The test has the null hypothesis that the (mean) performance fee equals to zero. TO is the mean value of the absolute change of portfolio weight.

Table 8: Contemporaneous Regressions of Realized Jumps

$$RM_t = \beta_0 + \beta_{JVD}JV_t + \epsilon_t$$

RM	β_0	β_{JVD}	$adjR^2$
<i>RV</i>	1.2392E-4 (10.5855)	1.7351 (124.7009)	0.3135
<i>RV</i> ⁺	6.4057E-5 (10.8169)	0.3005 (2.6581)	0.1831
<i>RV</i> ⁻	5.9861E-5 (10.2246)	1.4346 (11.9167)	0.4437
<i>RSK</i>	0.0310 (2.1479)	-1.1907E+3 (-1.6645)	0.0145
<i>RKU</i>	4.1987 (88.9212)	1.7231E+4 (3.7758)	0.8055

Notes: This table reports the coefficients of the contemporaneous regressions of realized moments including *RV*, *RV*⁺, *RV*⁻, *RSK*, *RKU* on realized jumps for the SPY contract for 2001 to 2010 using the BNS jump test. $adjR^2$ is the adjusted R square

Table 9: Realized Moments Forecasting

$$RM_{t,t+h-1} = \beta_0 + \beta_{RMD}RM_{t-1} + \beta_{RMW}RM_{t-5,t-1} + \beta_{RMM}RM_{t-22,t-1} + \epsilon_{t,t+h-1}$$

$$RM_{t,t+h-1} = \beta_0 + \beta_{RMD}RM_{t-1} + \beta_{RMW}RM_{t-5,t-1} + \beta_{RMM}RM_{t-22,t-1} + \beta_{JVD}JV_{t-1} + \epsilon_{t,t+h-1}$$

	β_0	β_{RMD}	β_{RMW}	β_{RMM}	β_{JVD}	$adjR^2$
<i>RV</i>						
$h = 1$	1.436E-5 (1.923)	0.278 (1.887)	0.425 (4.115)	0.187 (1.767)		0.562
	1.518E-5 (2.4670)	0.485 (4.012)	0.286 (3.549)	0.147 (1.659)	-0.914 (-4.687)	0.591
$h = 5$	2.204E-5 (2.568)	0.180 (1.872)	0.370 (4.137)	0.282 (2.413)		0.682
	2.252E-5 (2.760)	0.301 (2.966)	0.289 (3.247)	0.258 (2.340)	-0.536 (-3.292)	0.697
$h = 22$	4.103E-5 (6.329)	0.092 (1.675)	0.290 (2.903)	0.303 (2.898)		0.644
	4.133E-5 (6.318)	0.167 (2.784)	0.239 (2.383)	0.288 (2.765)	-0.333 (-3.086)	0.651
<i>RV⁺</i>						
$h = 1$	6.326E-6 (1.923)	0.289 (2.628)	0.454 (3.714)	0.159 (1.420)		0.604
	6.634E-6 (2.047)	0.308 (2.716)	0.442 (3.589)	0.156 (1.399)	-0.108 (-2.545)	0.606
$h = 5$	9.825E-6 (2.481)	0.186 (2.336)	0.411 (4.089)	0.252 (2.087)		0.717
	9.914E-6 (2.511)	0.191 (2.329)	0.408 (4.028)	0.251 (2.078)	-0.031 (-0.913)	0.717
$h = 22$	1.873E-5 (4.129)	0.098 (4.149)	0.301 (2.896)	0.310 (3.059)		0.668
	1.879E-5 (4.139)	0.102 (4.228)	0.299 (2.903)	0.309 (3.039)	-0.023 (-1.209)	0.668

	β_0	β_{RMD}	β_{RMW}	β_{RMM}	β_{JVD}	$adjR^2$
<i>RV-</i>						
$h = 1$	1.030E-6	0.134	0.411	0.299		0.387
	(2.072)	(0.984)	(3.493)	(2.133)		
	1.035E-5	0.401	0.268	0.221	-0.647	0.419
	(2.616)	(3.299)	(2.814)	(2.352)	(-3.501)	
$h = 5$	1.402E-5	0.103	0.321	0.365		0.583
	(2.959)	(1.253)	(3.007)	(2.932)		
	1.405E-5	0.286	0.222	0.311	-0.445	0.609
	(3.378)	(3.062)	(2.050)	(3.233)	(-3.110)	
$h = 22$	2.362E-5	0.053	0.252	0.335		0.594
	(4.113)	(1.287)	(2.307)	(3.845)		
	2.364E-5	0.172	0.188	0.301	-0.289	0.609
	(4.212)	(2.638)	(1.602)	(4.224)	(-2.725)	
<i>RSK</i>						
$h = 1$	0.027	-0.029	-0.087	0.050		0.003
	(1.697)	(-1.424)	(-1.601)	(0.422)		
	0.028	-0.030	-0.086	0.049	-102.129	0.003
	(1.739)	(-1.442)	(-1.59)	(0.413)	(-1.188)	
$h = 5$	0.027	-0.004	-0.100	0.079		0.013
	(1.908)	(-0.457)	(-2.283)	(0.746)		
	0.027	-0.004	-0.099	0.078	-42.374	0.013
	(1.927)	(-0.519)	(-2.287)	(0.742)	(-0.950)	
$h = 22$	0.026	-0.002	-0.010	0.023		0.027
	(1.924)	(-0.843)	(-0.496)	(0.225)		
	0.026	-0.002	-0.010	0.023	3.862	0.027
	(1.918)	(-0.825)	(-0.497)	(0.225)	(0.429)	
<i>RKU</i>						
$h = 1$	3.333	-0.002	0.071	0.153		0.735
	(7.858)	(-0.120)	(1.289)	(1.533)		
	3.290	0.012	0.071	0.150	-925.924	0.735
	(7.608)	(0.550)	(1.300)	(1.513)	(-2.190)	
$h = 5$	3.245	0.007	0.021	0.214		0.926
	(9.464)	(0.602)	(0.335)	(2.382)		
	3.213	0.018	0.021	0.217	-686.931	0.926
	(9.351)	(1.448)	(0.337)	(2.353)	(-2.516)	
$h = 22$	3.378	-0.002	0.013	0.201		0.980
	(12.221)	(-0.345)	(0.519)	(2.842)		

3.374	-1.081E-4	0.013	0.201	-88.993	0.980
(12.073)	(-0.020)	(0.519)	(2.849)	(-0.558)	

Notes: The table reports in-sample RM forecasting results for the SPY contract from 2001 to 2010. RV , RV^+ , RV^- , RSK , RKU are RMs used. For each RM forecasting, the HAR-RM model using its own daily, weekly, and monthly lagged RM, and the HAR-RM-J model using also the jump variation from the BNS jump test are applied to obtain one day, one week, and one month ahead forecasts. The figures in parentheses are t-statistics using Newey-West standard errors for autocorrelation orders of 5, 10, and 44 respectively. The $adjR^2$ is adjusted R square.

Table 10: Out-of-Sample Portfolio Performance: Alternative Realized Moments

Strategies	Moment Timing: Performance Fees		
	$\gamma = 2$	$\gamma = 6$	$\gamma = 10$
<i>RM</i>	0.0013 (2.0816)	4.4101E-4 (2.0566)	2.6401E-4 (2.0316)
<i>RM + JV</i>	0.0045 (2.1240)	0.0015 (2.0899)	8.9427E-4 (2.0588)
<i>JV</i>	0.0031 (1.7982)	0.0010 (1.7688)	6.2967E-4 (1.7355)

Notes: The table reports out-of-sample portfolio allocation results for the SPY contract for 2006 to 2010 by predicting upside and downside volatilities. The first case compares portfolio performances between predicting upside and downside volatilities and predicting total realized volatility. The second case compares portfolio performances between predicting upside and downside volatilities with their past values and jumps and predicting total realized volatility. The third case compares predicting upside and downside volatilities with their past values and jumps. We report performance fees relative to the benchmark strategy under risk aversion levels of 2, 6, and 10. Figures in parentheses are t-statistics for DM test. The test has the null hypothesis that the (mean) performance fee equals to zero.

Online Appendix

A Appendix 1: Jump Tests

In this part, we discuss specifications of six alternative jump tests, including Ait-Sahalia and Jacod (2009)(AJ), Jiang and Oomen (2008)(JO), Andersen, Dobrev, and Schaumburg (2012) (Med, Min) Corsi, Pirino, and Reno (2010)(CPR), and Podolskij and Ziggel (2010) (PZ).

A.1 AJ

Ait-Sahalia and Jacod (2009) find that realized power variation is invariant to different sampling scales when jumps are present. Therefore the AJ test detects the presence of jumps using the ratio of realized power variation sampled from two scales. For the realized power variation for the sampling scale h and kh with scalar $k > 0$ we have

$$PV_{t,M}(p, h) = \sum_{j=1}^{M/h} |r_j|^p, \text{ and}$$
$$\hat{S}_t(p, k, h) = \frac{\hat{P}V_{t,M}(p, kh)}{\hat{P}V_{t,M}(p, h)}.$$

Then the test statistic is given by

$$AJ_{t,M} = \frac{\hat{S}_t(p, k, h) - k^{p/2-1}}{\sqrt{\hat{V}_{t,M}}} \rightarrow N(0, 1), \quad (20)$$

where $\hat{V}_{t,M}$ is the asymptotic variance of $S_t(p, k, h)$.

$$\hat{V}_{t,M} = \frac{N(p, k)\hat{A}_{t,M(2p)}}{M\hat{A}_{t,M(p)}},$$
$$N(p, k) = (1/\mu_p^2)(k^{p-2}(1+k))\mu_{2p} + k^{p-2}(k-1)\mu_p^2 - 2k^{p/2-1}\mu_{k,p},$$
$$\mu_{k,p} = E(|U|^p|U + \sqrt{k-1}V|^p),$$
$$\hat{A}_{t,M(2p)} = \frac{(1/M)^{1-p/2}}{\mu_p} \sum_{j=1}^M |r_{t_j}|^p I_{|r_{t_j}| < \alpha(1/M)^\omega}.$$

U, V are random variables that $U \sim N(0, 1)$ and $V \sim N(0, 1)$ and p, k, α, ω are parameters.

A.2 JO

The swap variance test developed by Jiang and Oomen (2008) is in the spirit of Neuberger (1994)'s replicating strategy on the payoff of the variance swap contract. Different from

the other tests, which generate jump robust estimators, the JO test constructs a jump sensitive estimator. The difference between simple returns R_{t_j} and logarithmic returns r_{t_j} can capture one half of the integrated variance if there is no jump in the underlying sample path. Therefore the difference between twice the return difference and the realized variance can capture the jump component if a jump occurs.

$$SwV_{t,M} = 2 \sum_{j=1}^M (R_{t_j} - r_{t_j}).$$

Hence the test statistic using the ratio of $SwV_{t,M}$ and $RV_{t,M}$ is

$$JO_{t,M} = \frac{MBV_{t,M}}{\sqrt{\Omega_{SwV}}} \left(1 - \frac{RV_{t,M}}{SwV_{t,M}}\right) \longrightarrow N(0, 1), \quad (21)$$

where Ω_{SwV} is the asymptotic variance using estimated integrated sixicity.

$$\Omega_{SwV} = \frac{\mu_6}{9} \frac{M^3 \mu_{6/p}^{-p}}{(M-p-1)} \sum_{j=0}^{M-p} \prod k = 1^p |r_{t_j}|^{6/p}.$$

$p = 4$ or 6 are suggested parameters.

A.3 Med and Min

Andersen, Dobrev, and Schaumburg (2012) introduced two simple to implement but powerful estimators for integrated variance, which use nearest neighbour truncation to control for the impact of market microstructure noise and zero returns.

$$MinRV_t = \frac{\pi}{\pi - 2} \frac{m}{m - 1} \sum_{j=2}^m \min(|r_j|, |r_{j-1}|)^2, \quad (22)$$

$$MedRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} \frac{m}{m - 2} \sum_{j=3}^m \text{med}(|r_j|, |r_{j-1}|, |r_{j-2}|)^2 \quad (23)$$

The statistics are as follows,

$$\frac{1 - \frac{MinRV_t}{RV_t}}{1.81\delta \max\left(1, \frac{MinRQ_t}{MinRV_t^2}\right)} \longrightarrow N(0, 1), \quad (24)$$

$$\frac{1 - \frac{MedRV_t}{RV_t}}{0.96\delta \max\left(1, \frac{MedRQ_t}{MedRV_t^2}\right)} \longrightarrow N(0, 1), \quad (25)$$

where $MinRQ_t, MedRQ_t$ are minimum and median realized quarticity to estimate the integrated quarticity. The specifications are as follow

$$MinRQ_t = \frac{\pi m}{3\pi - 8} \frac{m}{m - 1} \sum_{j=2}^m \min(|r_j|, |r_{j-1}|)^4, \quad (26)$$

$$MedRQ_t = \frac{3\pi m}{9\pi + 72 - 52\sqrt{3}} \frac{m}{m - 2} \sum_{j=3}^m (|r_j| |r_{j-1}| |r_{j-2}|)^4. \quad (27)$$

A.4 CPR

A recent study by Corsi, Pirino, and Reno (2010) extended the multipower variation in the spirit of BNS by incorporating a threshold. The idea is that large returns can result in an underrejection of jumps using multipower variation based tests. Therefore by introducing a local variance based threshold to filter out large returns, the test expects to reduce the bias. The corrected realized threshold bipower variation and corrected threshold tripower quarticity are as follows,

$$ctBV_{t,M} = \frac{\pi}{2} \sum_{j=2}^M Z1(r_{t_j}, \theta_{t_j}) Z1(r_{t_{j-1}}, \theta_{t_{j-1}}),$$

$$ctTQ_{t,M} = \mu_{4/3}^{-3} \sum_{j=3}^M Z1(r_{t_j}, \theta_{t_j}) Z1(r_{t_{j-1}}, \theta_{t_{j-1}}) Z1(r_{t_{j-2}}, \theta_{t_{j-2}}),$$

where $\mu_{4/3} = E(|U|)^{4/3}$, $U \sim N(0, 1)$, and $\theta_{t_j} = c_\theta^2 \hat{V}_{t_j}$, c_θ is constant, and \hat{V}_{t_j} is local volatility. $Z1(r_{t_j}, \theta_{t_j})$ is the threshold function given by,

$$Z1(r_{t_j}, \theta_{t_j}) = \begin{cases} |r_{t_j}| & r_{t_j}^2 < \theta_{t_j}, \\ 1.094\theta_{t_j}^{1/2} & r_{t_j}^2 > \theta_{t_j} \end{cases}$$

The statistic is as follows

$$\frac{1 - ctBV_{t,M}/RV_t}{\sqrt{(\pi^2/4 + \pi - 5)hmax(1, ctBV_{t,M}/ctTQ_{t,M})}} \longrightarrow N(0, 1). \quad (28)$$

A.5 PZ

Similar to CPR, Podolskij and Ziggel (2010)'s test is also inspired by the Mancini (2009)'s threshold estimator. However, different from the CPR, PZ constructs the statistic based on random perturbed intraday returns. The difference of realized power variation and the truncated estimator captures the jump component. The estimator is as follows,

$$tMV_{t,M}(p) = \frac{1}{M^{1-p/2}} \sum_{j=1}^M |r_{t_j}|^p (1 - \eta_j I_{|r_{t_j}| \leq ch^\omega}),$$

and the statistic is given by,

$$PZ_{t,M}(p) = \frac{tMV_{t,M}(p)}{\sqrt{Var(\eta_j)MV_{t,M}(2p)}} \longrightarrow N(0, 1), \quad (29)$$

where the perturbing variable η_j is drawn from the distribution $P^\eta = 1/2(\delta_{1-\tau} + \delta_{1+\tau})$, δ is the Dirac measure and $\tau = 0.1$ or 0.05 .

B Appendix 2: Additional Robustness Checks

In this part, we document additional robustness check results, which have not been included in the main analysis. We first conduct a Monte Carlo simulation experiment to understand finite sample properties of different jump tests. Then, we consider decomposing jumps into positive and negative components and investigate their statistical and economic performances. Finally we conduct sub-sample analyses.

B.1 Simulations

In this section, we conduct a comprehensive Monte Carlo simulation analysis to understand the finite sample properties of different jump estimators. The purpose is to justify different statistical and economic performance of different jump estimators documented in the main part of the paper. Following Barndorff-Nielsen and Shephard (2004) and Huang and Tauchen (2005), we generate sample paths through a one factor stochastic volatility model under the null hypothesis of no jumps. The one factor stochastic volatility model is given by

$$\begin{aligned} dp_t &= \mu dt + \exp[\beta_0 + \beta_1 v_t] dW_{pt}, \\ dv_t &= \alpha_v v_t dt + dW_{vt}, \quad \text{corr}(dW_{pr}, dW_{vt}) = \rho. \end{aligned} \quad (30)$$

The one factor model shows that the price dynamic is driven by the price diffusion term dW_{pt} and the volatility diffusion term dW_{vt} , and these two terms are correlated in order to allow for a leverage effect. Under the alternative hypothesis a jump component is added to the price process, which is a compounded Poisson process with intensity λ and sizes drawn from $N(0, \sigma_{jump}^2)$. To implement the simulation, we use the Euler scheme with increments of 1 second. We generate 10000 trading days each with 6.5 hours. Then we sample with a frequency of 1, 2, 5, 15 minutes.

Table 11: Simulation Results: Size

Median Mean reversion $av = -0.100$				
h	1	2	5	15
<i>BNS</i>	0.0525	0.0485	0.0330	0.0405
<i>AJ</i>	0.0780	0.0700	0.0775	0.0890
<i>JO</i>	0.0705	0.0745	0.0840	0.1255
<i>Med</i>	0.0470	0.0495	0.0475	0.0480
<i>Min</i>	0.0500	0.0450	0.0380	0.0375
<i>CPR</i>	0.0465	0.0470	0.0485	0.0535
<i>PZ</i>	0.0650	0.0810	0.0820	0.0910

Notes: The table reports the simulation size for all seven jump tests we used, including BNS, AJ, JO, Med, Min, CPR, and PZ. We report the type I errors for 1, 2, 5, and 15 minutes frequency compared to a significant level of 5%. The values are obtained through simulations with 10,000 trading days under the null hypothesis that no jumps occur.

Table 11 shows the empirical sizes (α) for a 5% level of the different jump tests. We follow Huang and Tauchen (2005) and Dumitru and Urga (2012) to choose parameters and set

$av = -0.1$ to represent a moderate level of mean reversion. The simulated data does not include a noise component, therefore the sizes for the different frequencies should only reflect the sizes of the tests and should not be affected by market microstructure noise as in real world data. For the 1 minute frequency, Med RV, Min RV, and CPR have a better size (closer to 5%) in comparison to the other jump tests, while AJ and JO seem to be a bit oversized. The relative merits of Med RV, Min RV, and CPR generally hold true throughout the other frequencies. The bipower variation test in general performs well across different frequencies, but is slightly oversized at a 1 minute frequency.

Table 12: Simulation Results: Size Corrected Power with Varying Intensity

Median Mean reversion $av = -0.100$, Median Jump size $\sigma = 1.5$.

h	1	2	5	15
$\lambda = 2$				
<i>BNS</i>	0.8554	0.8213	0.7627	0.6107
<i>AJ</i>	0.2467	0.1532	0.0575	0.0346
<i>JO</i>	0.8897	0.8671	0.8270	0.7776
<i>Med</i>	0.8437	0.8206	0.7606	0.6155
<i>Min</i>	0.8400	0.7963	0.7256	0.5397
<i>CPR</i>	0.8584	0.8484	0.8056	0.6857
<i>PZ</i>	0.8898	0.9064	0.8769	0.7943
$\lambda = 1.5$				
<i>BNS</i>	0.7673	0.7257	0.6567	0.5367
<i>AJ</i>	0.2674	0.1688	0.0656	0.0456
<i>JO</i>	0.7999	0.7736	0.7320	0.6975
<i>Med</i>	0.7576	0.7322	0.6688	0.5541
<i>Min</i>	0.7532	0.7105	0.6393	0.4779
<i>CPR</i>	0.7740	0.7450	0.6942	0.5980
<i>PZ</i>	0.8080	0.8107	0.7941	0.7178
$\lambda = 1.0$				
<i>BNS</i>	0.5995	0.5812	0.5253	0.4252
<i>AJ</i>	0.2614	0.1667	0.0699	0.0587
<i>JO</i>	0.6299	0.6256	0.6075	0.5975
<i>Med</i>	0.5971	0.5850	0.5391	0.4627
<i>Min</i>	0.5832	0.5649	0.5005	0.3844
<i>CPR</i>	0.6051	0.6028	0.5665	0.4849
<i>PZ</i>	0.6508	0.6551	0.6220	0.5545
$\lambda = 0.5$				
<i>BNS</i>	0.3794	0.3305	0.3268	0.2465
<i>AJ</i>	0.2034	0.1403	0.0759	0.0840
<i>JO</i>	0.4024	0.3663	0.3936	0.4105
<i>Med</i>	0.3820	0.3372	0.3381	0.2757
<i>Min</i>	0.3668	0.3215	0.3160	0.2286
<i>CPR</i>	0.3907	0.3452	0.3505	0.2932
<i>PZ</i>	0.4214	0.4353	0.4101	0.3768

Notes: The table reports the simulation power for all seven jump tests we used, including BNS, AJ, JO, Med, Min, CPR, and PZ. We report one minus type II errors for 1, 2, 5, and 15 minutes frequency compared to a significant level of 5%. The values are obtained through simulations with 10,000 trading days under the alternative hypothesis with jumps. We report how the power of test varies when the jump intensity λ changes.

Tables 12 and 13 report the empirical power $(1 - \beta)$ for the different jump tests when the jump intensity and/or the jump size is varying. Similar to the size part, we fix the mean reversion parameter to a median level of $av = -0.1$. Firstly, we allow for a changing jump intensity while keeping the jump size fixed equal to its median level $\sigma = 1.5$ as shown in Table 12. When the jump intensity decreases, the power for all tests also decreases as there are less jumps in the data. Moreover, when the sampling frequency is reduced, the power of the tests monotonically decreases in general. One exception is the PZ test. Here the highest

power is found for the 2 min sampling frequency. The highest powers are found for PZ, JO, and CPR, while Med RV and Min RV have powers close to the baseline BNS test. AJ is found to have the lowest power for all frequencies and levels of intensities.

Table 13: Simulation Results: Size Corrected Power with Varying Size

Median Mean reversion $av = -0.100$, Median Jump Intensity $\lambda = 1$.

h	1	2	5	15
$\sigma = 2$				
<i>BNS</i>	0.6327	0.5949	0.5724	0.4815
<i>AJ</i>	0.2614	0.1769	0.0656	0.0582
<i>JO</i>	0.6573	0.6451	0.6496	0.6209
<i>Med</i>	0.6270	0.5986	0.5869	0.5016
<i>Min</i>	0.6232	0.5775	0.5593	0.4426
<i>CPR</i>	0.6424	0.6107	0.6143	0.5357
<i>PZ</i>	0.6535	0.6708	0.6552	0.6568
$\sigma = 1.5$				
<i>BNS</i>	0.6137	0.5770	0.5109	0.4070
<i>AJ</i>	0.2473	0.1726	0.0721	0.0615
<i>JO</i>	0.6514	0.6294	0.5944	0.5603
<i>Med</i>	0.6133	0.5813	0.5249	0.4307
<i>Min</i>	0.6053	0.5602	0.4922	0.3714
<i>CPR</i>	0.6214	0.5965	0.5518	0.4638
<i>PZ</i>	0.6567	0.6649	0.6285	0.5825
$\sigma = 1.0$				
<i>BNS</i>	0.5520	0.5255	0.4390	0.3007
<i>AJ</i>	0.2218	0.1683	0.0900	0.0703
<i>JO</i>	0.5933	0.5808	0.5393	0.4923
<i>Med</i>	0.5514	0.5397	0.4572	0.3451
<i>Min</i>	0.5326	0.5079	0.4210	0.2540
<i>CPR</i>	0.5721	0.5509	0.4924	0.3693
<i>PZ</i>	0.6134	0.6126	0.5861	0.4895
$\sigma = 0.5$				
<i>BNS</i>	0.4084	0.3594	0.2404	0.1042
<i>AJ</i>	0.1773	0.1065	0.0726	0.0873
<i>JO</i>	0.4825	0.4392	0.3652	0.2613
<i>Med</i>	0.4171	0.3730	0.2724	0.1308
<i>Min</i>	0.3800	0.3115	0.2095	0.0774
<i>CPR</i>	0.4457	0.3977	0.3079	0.1490
<i>PZ</i>	0.5947	0.5212	0.4112	0.2503

Notes: The table reports the simulation power for all seven jump tests we used, including BNS, AJ, JO, Med, Min, CPR, and PZ. We report one minus type II errors for 1, 2, 5, and 15 minutes frequency compared to a significant level of 5%. The values are obtained through simulations with 10,000 trading days under the alternative hypothesis with jumps. We report how the power of test varying when jump intensity σ changes.

Secondly, we fix the jump intensity and allow for the value of the jump size to vary. Table 13 reports that the power of the tests decreases when jump size decreases. This result holds true for all jump tests, indicating that all existing jump tests have difficulties to detect small jumps or to distinguish volatility bursts from jumps. As before, the power decreases with lower sampling frequency. Similar to the size part, we also find a peak of power for the PZ test for a 2 min frequency, but only when the jump size is large $\sigma = 2$ or $\sigma = 1.5$. The highest power is again found for the tests: PZ, JO, and CPR. Moreover, both PZ and JO seem to have high power even when the sampling frequency is relative low (15 min).

The relative ranking is similar to the case when intensity is varying. Med RV and Min RV have similar power in comparison to the baseline bipower test. To summarize, we show that the finite sample properties of different jump tests are varying across sampling frequen-

cies and jump characteristics. Med RV, Min RV, and CPR have the best size properties while PZ, JO, and CPR have the best power properties. In general, the performance of the different tests in the simulation study reflects very well the trade-off between size and power.

B.2 Good and Bad Jumps

As reported in the descriptive statistics in Table 1, the jump densities are significantly skewed. The conventional view is that negative jumps should matter more than positive jumps. Therefore, we analyze in this section whether further decomposing jump variations into positive and negative jumps can improve statistical and economic performances. We follow Barndorff-Nielsen, Kinnebrock, and Shephard (2008) and Patton and Sheppard (2011) and use realized semivariances for this purpose.

Following Patton and Sheppard (2011), we define realized semivariances as follows:

$$RV^- = \sum_{i=1}^n r_i^2 I_{[r_i < 0]}, \quad (31)$$

$$RV^+ = \sum_{i=1}^n r_i^2 I_{[r_i > 0]}. \quad (32)$$

Realized semivariances converge to the sum of half the integrated variance and jump variation for positive and negative returns respectively:

$$RV^- \rightarrow \frac{1}{2} \int_{t-1}^t \sigma_s^2 ds + \sum_{j=1}^{N_t} c_j^2 I_{[c_j < 0]}, \quad (33)$$

$$RV^+ \rightarrow \frac{1}{2} \int_{t-1}^t \sigma_s^2 ds + \sum_{j=1}^{N_t} c_j^2 I_{[c_j > 0]}. \quad (34)$$

Therefore, the difference of the two realized semivariances ensures that the half integrated variation terms vanishes, so that the signed jump variation is given by

$$\Delta J^2 = RV^+ - RV^- \rightarrow \sum_{j=1}^{N_t} c_j^2 I_{[c_j > 0]} - \sum_{j=1}^{N_t} c_j^2 I_{[c_j < 0]}. \quad (35)$$

The negative and positive jump components are given by

$$\Delta J^{2+} = (RV^+ - RV^-) I_{[(RV^+ - RV^-) > 0]}, \quad (36)$$

$$\Delta J^{2-} = (RV^+ - RV^-) I_{[(RV^+ - RV^-) < 0]}. \quad (37)$$

Using these measures we discuss four alternative HAR-RV-CJ specifications to evaluate volatility timing statistically and economically.

$$\begin{aligned}
RV_{t,t+h-1} &= \beta_0 + \beta_{IVD}IV_{t-1} + \beta_{IDW}IV_{t-5,t-1} + \beta_{IVM}IV_{t-22,t-1} + \beta_{JSD}JSD_{t-1} + \epsilon_{t,t+h-1}, \\
RV_{t,t+h-1} &= \beta_0 + \beta_{IVD}IV_{t-1} + \beta_{IDW}IV_{t-5,t-1} + \beta_{IVM}IV_{t-22,t-1} + \beta_{JND}JND_{t-1} + \epsilon_{t,t+h-1}, \\
RV_{t,t+h-1} &= \beta_0 + \beta_{IVD}IV_{t-1} + \beta_{IDW}IV_{t-5,t-1} + \beta_{IVM}IV_{t-22,t-1} + \beta_{JPD}JPD_{t-1} + \epsilon_{t,t+h-1}, \\
RV_{t,t+h-1} &= \beta_0 + \beta_{IVD}IV_{t-1} + \beta_{IDW}IV_{t-5,t-1} + \beta_{IVM}IV_{t-22,t-1} + \beta_{JND}JND_{t-1} + \beta_{JPD}JPD_{t-1} + \epsilon_{t,t+h-1}.
\end{aligned} \tag{38}$$

where JSD is the daily lagged signed jump variation, JND and JPD are the negative and positive jump components, respectively.

Table 14 presents in-sample volatility forecasting results. The sign of the jump component is positive for negative jumps and negative for positive jumps, just as suggested in Patton and Sheppard (2011). Therefore, a large price drop is likely to increase the volatility in the future while a price increase tends to reduce volatility. Moreover, the coefficients for signed jumps are positive, implying that negative jumps play a dominant role. However, we also find that for a daily horizon, the coefficients for jump specifications are insignificant, which is different from our main findings using jump variations. Then the coefficients turn significant for weekly and monthly horizons. Similar to before, we observe that all four models can lead to an improvement in adjusted R^2 s compared to the benchmark HAR-RV model. For the one day ahead horizon, the adjusted R^2 is on average about 2% higher than that of the HAR-RV model. The adjusted R^2 s are close to each other for all four specifications, however, the adjusted R^2 for the negative jump model is slightly higher than that for the positive jump model, which is consistent with Patton and Sheppard (2011).

The out-of-sample volatility forecasting results are reported in Table 15 panel 1. For the daily horizon, all four specifications lead to better forecasts than the HAR-RV model. The negative jump model generates lower MSEs than the positive jump model, which is in line with the in-sample findings. However, we show that using signed jumps without separating them into positive and negative could generate lower MSEs. Moreover, all MSEs are close to the MSEs of the baseline HAR-RV-CJ models using the BNS test. Nevertheless, although coefficients become significant for weekly and monthly horizons, the out-of-sample performance deteriorates when the forecasting horizon is extended.

Table 15 panel 2 presents volatility timing performances for these models. We find that all those four specifications can outperform the benchmark HAR-RV model, and can

Table 14: In-Sample Volatility Forecasting Results

	β_0	β_{IVD}	β_{IVW}	β_{IVM}	β_{JSD}	β_{JND}	β_{JPD}	$adjR^2$
<i>JS</i>								
$h = 1$	1.527E-5 (2.358)	0.435 (3.850)	0.350 (4.023)	0.133 (1.351)	0.089 (0.545)			0.587
$h = 5$	2.316E-5 (3.048)	0.260 (4.643)	0.349 (4.019)	0.245 (2.148)	0.009 (0.095)			0.697
$h = 22$	4.198E-5 (4.257)	0.131 (4.418)	0.294 (2.858)	0.277 (2.862)	-0.010 (-0.120)			0.653
<i>JN</i>								
$h = 1$	1.544E-5 (2.430)	0.461 (3.778)	0.341 (3.852)	0.133 (1.381)		0.181 (1.275)		0.589
$h = 5$	2.333E-5 (3.100)	0.282 (3.334)	0.340 (3.832)	0.242 (2.141)		0.105 (2.052)		0.699
$h = 22$	4.206E-5 (4.254)	0.142 (4.450)	0.289 (2.776)	0.275 (2.858)		0.042 (0.662)		0.654
<i>JP</i>								
$h = 1$	1.515E-5 (2.352)	0.433 (3.440)	0.359 (3.871)	0.145 (1.476)			-0.194 (-0.698)	0.587
$h = 5$	2.318E-5 (3.102)	0.282 (3.357)	0.351 (4.055)	0.258 (2.295)			-0.364 (-2.720)	0.701
$h = 22$	4.201E-5 (4.260)	0.147 (4.421)	0.294 (2.831)	0.284 (2.980)			-0.218 (-2.189)	0.655
<i>JN, JP</i>								
$h = 1$	1.551E-5 (2.469)	0.487 (3.629)	0.339 (3.798)	0.142 (1.453)		0.212 (1.618)	-0.285 (-1.061)	0.590
$h = 5$	2.344E-5 (3.173)	0.321 (3.586)	0.336 (3.828)	0.256 (2.275)		0.153 (2.934)	-0.429 (-2.907)	0.704
$h = 22$	4.213E-5 (4.255)	0.164 (4.637)	0.287 (2.754)	0.283 (2.961)		0.069 (1.321)	-0.247 (-2.426)	0.656

Notes: The table reports in-sample volatility forecasting results for the SPY contract from 2001 to 2010 using good and bad jumps. The HAR-RV-CJ model is applied to forecast one day, one week, and one month ahead realized variance. We consider four specifications: realized semivariance, negative jumps, positive jumps, and both negative and positive jumps. The figures in parentheses are t-statistics using Newey-West standard errors for autocorrelation order 5, 10, and 44 respectively. $adjR^2$ is the adjusted R square.

generate positive and statistically significant performance fees at the 5% significance level. For the signed jump model, we show that annualized performance fees range from 40 ($\gamma = 2$) to 8 ($\gamma = 10$) basis points. We also find that the negative jump model (from 47 basis points for $\gamma = 2$ to 9 basis points for $\gamma = 10$) outperforms its positive counterpart (37 basis points for $\gamma = 2$ to 7 basis points for $\gamma = 10$) slightly, which is consistent with both our statistical findings and the existing literature. Including both positive and negative jumps can further improve portfolio performances, and generate economic value from 52 basis points ($\gamma = 2$) to 10 basis points ($\gamma = 10$). Our findings suggest that separating jumps from diffusion improve volatility timing performances even without the use of nonparametric jump tests. And decomposing jumps into positive and negative components can further improve economic performances.

B.3 Sub-Sample Analysis

We then conduct a sub-sample analysis in order to understand whether the results hold true for different time periods. We split the whole data sample into period 1 (2001 to 2005) and period 2 (2006 to 2010). In each period, we use the first two years as the in-sample period and the remaining years as the out-of-sample period.

Table 15: Out-of-Sample Statistical and Economic Performances: Good and Bad Jumps

Panel 1: Volatility Forecasting			
MSE	h=1	h=5	h=22
<i>BM</i>	5.657E-8	3.532E-8	3.504E-8
<i>JS</i>	4.719E-8	3.100E-8	3.194E-8
<i>JN</i>	4.778E-8	4.700E-8	5.516E-8
<i>JP</i>	4.791E-8	4.881E-8	5.834E-8
<i>JN, JP</i>	4.811E-8	4.613E-8	5.348E-8
Panel 2: Volatility Timing: Daily			
Strategies	$\gamma = 2$	$\gamma = 6$	$\gamma = 10$
<i>JS</i>	(0.0040)	0.0013	8.2063E-4
	(2.6780)	(2.6450)	(2.6100)
<i>JN</i>	(0.0047)	0.0016	9.4140E-4
	(2.6512)	(2.6171)	(2.6512)
<i>JP</i>	(0.0037)	0.0012	7.4159E-4
	(3.3700)	(3.3357)	(3.3015)
<i>JN, JP</i>	(0.0052)	0.0017	0.0010
	(2.7244)	(2.6901)	(2.6558)

Notes: The table reports out-of-sample statistical and economic performances for 2001 to 2010 using good and bad jumps. Panel 1 reports out-of-sample volatility forecasting results. The HAR-RV model is used as the benchmark model. The HAR-RV-CJ model with semivariance (*JS*), negative jumps (*JN*), positive jumps (*JP*), and both negative and positive jumps (*JN, JP*). Panel 2 reports out-of-sample volatility timing results. Parameters are estimated in-sample (from 2001 to 2005). The volatility forecasting and volatility timing results are obtained out-of-sample (from 2006 to 2010). We report performance fees relative to benchmark for risk aversion levels of 2, 6, and 10. Figures in parentheses are t-statistics for DM test. The test has the null hypothesis that the (mean) performance fee equals to zero.

Table 16 summarizes the out-of-sample volatility forecasting results for different jump tests. Initially, we find that the MSE in period 1 is significantly lower than in period 2. This finding can be attributed to the financial crisis period covered in period 2, which yields time varying parameters (especially for the in-sample period for 2006 to 2008). In period 1, the estimated in-sample parameters are more stable across the out-of-sample period and hence produce better forecast. Secondly, almost all HAR-RV-CJ models can outperform the benchmark HAR-RV in period 1 for all forecasting horizons. In period 2, only CPR and Med can outperform the benchmark at daily and weekly horizons, but not at a monthly horizon. During the financial crisis, local volatility is higher, hence jumps are more difficult to detect. When the forecasting horizon increases the reduction in forecasting accuracy is expected during the financial crisis period as the parameters are not immediately updated with the changing high local volatility information.

Table 17 illustrates out-of-sample volatility timing results for period 1 and 2. In period 1, the realized utility criteria show that all tests but AJ can outperform the benchmark. This result is consistent with the out-of-sample volatility forecasting results shown in Table 16, and with the whole in and out-of-sample volatility forecasting results. The performance fees are also statistically significant for almost all models. We find that all but the AJ jump strategies can generate positive and statistically signifi-

Table 16: Out-of-Sample Volatility Forecasting HAR-RV-CJ: Sub Samples

$$RV_{t,t+h-1} = \beta_0 + \beta_{IVD}IV_{t-1} + \beta_{IDW}IV_{t-5,t-1} + \beta_{IVM}IV_{t-22,t-1} + \beta_{JVD}JV_{t-1} + \epsilon_{t,t+h-1}$$

MSE	Period 1			Period 2		
	h=1	h=5	h=22	h=1	h=5	h=22
<i>BM</i>	1.528E-9	2.071E-9	4.511E-9	3.385E-8	1.356E-8	7.585E-9
<i>BNS</i>	6.698E-10	1.118E-9	3.280E-9	3.393E-8	1.374E-8	8.370E-9
<i>AJ</i>	1.522E-9	2.080E-9	4.496E-9	3.502E-8	1.892E-8	3.483E-8
<i>JO</i>	6.684E-10	1.120E-9	3.275E-9	3.388E-8	1.464E-8	8.576E-9
<i>Med</i>	6.686E-10	1.090E-9	3.194E-9	3.833E-8	1.397E-8	7.590E-9
<i>Min</i>	6.381E-10	1.063E-9	3.199E-9	3.396E-8	1.367E-8	8.508E-9
<i>CPR</i>	6.616E-10	1.085E-9	3.197E-9	3.124E-8	1.345E-8	1.578E-8
<i>PZ</i>	7.108E-10	1.193E-9	3.350E-9	3.386E-8	1.122E-8	8.601E-9

Notes: The table reports out-of-sample volatility forecasting results for the SPY contract for period 1 (2001 to 2005) and period 2 (2006 to 2010). HAR-RV model is used as the benchmark model. HAR-RV-CJ model with different jumps are used. The parameters are estimated in-sample from 2001 to 2003 for period 1 and 2006 to 2008 for period 2. The forecasting studies are implemented in the out-of-sample period 2004 to 2005 for period 1 and 2009 to 2010 for period 2. We report the Mean Squared Error (MSE) for predicted volatility over one day, one week, and one month forecasting horizons.

cant economic values. Moreover, the economic magnitudes are higher than those using the whole sample in the main part of the paper. In period 2, only AJ can outperform the benchmark. The t-statistics are all lower than the single tail 5% critical value. The results show that economic values in period 2 are either negative or insignificant.

Table 17: Out-of-Sample Volatility Timing: Sub-Samples

Strategies	Period 1			Period 2		
	$\gamma = 2$	$\gamma = 6$	$\gamma = 10$	$\gamma = 2$	$\gamma = 62$	$\gamma = 10$
<i>BNS</i>	0.0299 (2.5814)	0.0273 (2.3650)	0.0247 (2.1464)	-1.7027E-5 (-0.161)	-7.6982E-5 (-0.0724)	-1.3654E-4 (-0.1279)
<i>AJ</i>	-1.0276E-4 (-0.3415)	-1.1435E-4 (-0.3812)	-1.2586E-4 (-0.4212)	0.0015 (0.2168)	0.0015 (0.2131)	0.00015 (0.2093)
<i>JO</i>	0.0279 (2.4002)	0.0252 (2.1824)	0.0226 (1.9622)	-3.2155E-4 (-0.1845)	-4.0598E-4 (-0.2317)	-4.8986E-4 (-0.2776)
<i>Med</i>	0.0307 (2.6058)	0.0280 (2.3884)	0.0254 (2.1687)	-0.0105 (-1.1615)	-0.0099 (-1.0998)	-0.0093 (-1.0316)
<i>Min</i>	0.0322 (2.5149)	0.0293 (2.2936)	0.0263 (2.0699)	-2.3896E-4 (-0.0865)	-4.0598E-4 (-0.1418)	-5.6343E-4 (-0.1940)
<i>CPR</i>	0.0289 (2.4808)	0.0262 (2.2623)	0.0236 (2.0416)	-0.0021 (-0.3326)	-0.0028 (-0.4455)	-0.0035 (-0.5537)
<i>PZ</i>	0.0256 (2.3422)	0.0233 (2.1933)	0.0209 (1.9149)	-0.0706 (-1.5079)	-0.0723 (-1.1793)	-0.0741 (-2.4810)

Notes: The table reports out-of-sample volatility timing results for the SPY contract for period 1 (2001 to 2005) and period 2 (2006 to 2008). The benchmark strategy uses the HAR-RV to predict one day ahead volatility, and other strategies using the HAR-RV-CJ to predict one day ahead volatility with realized jump variation detected from respective jump test. The parameters are estimated in-sample (2001 to 2003 for period 1 and 2006 to 2008 for period 2), the predicted volatility and the ex post portfolio returns are obtained out-of-sample (2004 to 2005 for period 1 and 2009 to 2010 for period 2). We report performance fee relative to benchmark strategy under risk aversion of 2, 6, and 10. Figures in parentheses are t-statistics for DM test. The test has the null hypothesis that the (mean) performance fee equals to zero.

To summarize, due to the short sample period of 10 years, our sub-sample periods covered the recent financial crisis, which leads to unstable parameter estimation and high local volatility. We suggest that separating the integrated variance from the jump components precisely is more difficult during a high local volatility regime and models have more difficulties to update time varying parameters. Hence, we find that our main result is stronger in the first sub-sample period, where the economic environ-

ment is relatively stable. Nevertheless, our main findings still holds true in general for both sub sample periods: The separation of jumps from diffusion component rather than jump itself is responsible for delivering statistically significant economic values.