A Factor Approach to Realized Volatility Forecasting in the Presence of Finite Jumps and Cross-Sectional Correlation in Pricing Errors

Alev Atak*
Essex Business School

George Kapetanios†
Queen Mary, University of London

April 3, 2013

Abstract

There is a growing literature on the realized volatility (RV) forecasting of the asset returns using the high-frequency data. We explore the possibility of forecasting the RV with the factor analysis; once considering the significant jumps. A real high-frequency financial data application suggests that the factor based approach is of significant potential interest and novel.

Keywords: realized volatility; bipower variation; jump tests; factor models; volatility forecasting; model selection

JEL classification: C32; C50; C52; C53; C58

1 Introduction

Recently, there has been increasing interest in forecasting methods that utilize large high frequency data sets. Andersen and Bollerslev (1998), Andersen et al. (2003), Barndorff-Nielsen and Shephard (2002) (termed BNS henceforth), among others, advocated the use of nonparametric realized volatility (RV). The consistency of the RV as an estimator is violated by the presence of the market microstructure noise (henceforth 'noise') which emerges due to market frictions. Another backdrop is that the nonparametric RV literature has concentrated less on distinguishing jump from non jump movements. Corsi et al. (2010) reveal that dividing volatility into jumps and continuous variation yields a substantial improvement in volatility forecasting.

*Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ. Email: aatak@essex.ac.uk

†School of Economics and Finance, Queen Mary, University of London, Mile End Road, London E1 4NS. Email: g.kapetanios@qmul.ac.uk

There is an alternative way of looking at these problems. The limitations of the traditional procedures motivate our diverse approach for measuring and forecasting the realized equity return volatility. We apply the methodology of approximate factor modelling on the nonparametric RV and also on the realized bipower variation (BV) when it is required after separately measuring the continuous sample path variation and the discontinuous jump part of the quadratic variation (QV) process. Factor methods are very appealing and extensively used for forecasting; providing a theoretical device for summarizing large data sets without running into degrees of freedom problem, while taking into account the marginal benefits that increasing information brings to forecasting. As argued by Ludvingson and Ng (2009), the fluctuations and comovements of a large number of economic and financial variables are produced by a handful of observable or unobservable factors, which in this case represent the omitted unobservable factors in the noise. Our new factor-based realized volatility model ($FB-RV-J$) fits well for large dimensional panels.

2 Theory

The dynamics of the logarithmic price process, p_t, is usually assumed to be a jump-diffusion process of the form:

$$dp_t = \mu_t dt + \sigma_t dW_t + dJ_t \tag{1}$$

where μ_t denotes the drift term with a continuous and locally bounded variation, σ_t is the diffusion parameter and W_t is a standard Brownian motion. J_t is the jump process at time t, defined as $J_t = \sum_{j=1}^{N_t} k_{t_j}$ where k_{t_j} represents the size of the jump at time t_j and N_t is a counting process, representing the number of jumps up to time t. The QV of the price process up to a certain point in time t is:

$$QV_t = \int_0^1 \sigma_s^2 ds + \sum_{j=1}^{N_t} k_{t_j}^2 \tag{2}$$

where $\int_0^1 \sigma_s^2 ds = IV_t$ is the integrated variance or volatility. Thus, QV has two parts; the diffusion component and the jump component. The two components have a different nature and should be separately analyzed and modelled. The IV is characterized by persistence, whereas jumps have an unpredictable nature.

Let the interval $[0, t]$ split into n equal subintervals of length m. The jth intra-day return r_j on day t is defined as $r_j = p_{t-1+jm} - p_{t-1+(j-1)m}$. QV_t can be estimated by the realized volatility, or variation, (RV_t), defined as (Andersen and Bollerslev 1998):

$$RV_t = \sum_{j=1}^{n} r_j^2 \xrightarrow{p} QV_t, \text{for } m \longrightarrow 0 \tag{3}$$

where $\xrightarrow{p}$ stands for convergence in probability. Hence, in the absence of discontinuities and noise the RV_t is consistent for the IV_t. Most of the jump detection procedures are based on the comparison between RV_t and a robust to jump estimator. To highlight, none of these procedures can test for the absence or presence of jumps in the model or the data generating process. Hence, it is difficult to judge whether the realization of the process is continuous or not, within a certain time interval or at a certain moment without a jump test. We turn now to the jump detection methods.

2.1 Jump Tests

We use two tests, the adjusted ratio statistic of BNS (2006) and the Lee and Mykland test (Lee and Mykland, 2008; termed LM henceforth), in order to check whether the two tests give consistent results. BNS test tells whether a jump occurred during a particular day and how much the jump-squared contributes to the total realized variance, i.e. $\int_{t-1}^{t} J_s^2 dq_s / RV_t$. The significant jump component of RV_t is:

$$\widehat{J}_t \equiv sign(r_t) \times \sqrt{(RV_t - BV_t) \times I_{t,(ZJ_{(bv)} \geq \Phi_\alpha^{-1})}} \tag{4}$$

where $BV_t = 1.57 \sum_{j=2}^{n} |r_j| \, |r_{j-1}|$. BNS test can only identify days that contain jumps. Hence, we use the "intra-day" LM test which has the additional capability of identifying specific returns that can be classified as jumps. We compute the LM test statistics for every moment t_j within a trading day and then pick up the maximum statistic as the final test for that day to determine whether both tests are consistently detecting the presence of jumps. We effectively observe the consistency in both methods.

2.2 Model

We now put the idea of separately measuring the jump component and continuous variation. The contribution to the QV_t process due to the discontinuities in the underlying price process can be estimated by:

$$RV_t - BV_t \rightarrow \sum_{j=1}^{N_t} k_{t_j}^2, \text{ for } m \longrightarrow 0 \tag{5}$$

Under this central insight and based on the above mentioned test statistics and threshold requirements, we use BV_t in our analysis if we detect jumps in the data, otherwise RV_t. So, $C_t = I_{t,(ZJ_{(bv)} < \Phi_\alpha^{-1})} RV_t + I_{t,(ZJ_{(bv)} \geq \Phi_\alpha^{-1})} BV_t$. This recognition motivates our model. We propose that our nonparametric jump-free 'realized' measure can be decomposed into the common and idiosyncratic components. We relate the common component to unobservable financial characteristics, in particular, to cross sectional correlation in pricing errors. For simplicity, we abbreviate our model $FB - RV - J$:

$$\begin{aligned} h_{it} &= \alpha_i^{'} f_t + u_{it}, \qquad (6) \\ t &= 1, \ldots, T \text{ and } i = 1, \ldots, N \end{aligned}$$

where h_{it}, is the realized measure, which is the element in the t^{th} row and i^{th} column of the data matrix, $T \times N$. f_t is a $r-$dimensional vector of common factors with

$t = 1, \ldots, T$ and α_i refers to the i^{th} row of the corresponding matrix of factor loadings. $\alpha_i^{'} f_t = W_{it}$ is the set of common components. In addition, u_{it} is the idiosyncratic component of h_{it}. We assume that in general the idiosyncratic terms are also weakly dependent processes with mild cross-sectional dependence. α_i and f_t are clearly not jointly identified since the factors can be pre-multiplied by an invertible $r \times r$ matrix without having to make changes in the model. The most crucial point here is that $r << N$, so that substantial dimension reduction can be achieved.

Factor identification and estimation of (6) is based on the set of assumptions that are used in Bai and Ng (2002, 2006). Estimation is divided into steps; we start with determining the number of factors, which is followed by estimating them along with the loadings. We estimate common factors in large panels by the method of asymptotic principal components. This approach fits well for the large panel of realized volatilities because it does not suffer from the curse of dimensionality problem.

2.2.1 The Number of Factors

We now focus on checking robustness with respect to the number of factors and consider two approaches; Bai and Ng (2002) information criteria forming a nonparametric method to determine the statistically important factors and the Onatski (2010) estimator described by an algorithm named edge distribution (ED). Kapetanios (2010) suggests a method of the determination of the number of factors using a bootstrap method, which is robust to considerable cross-sectional and temporal dependence, but we prefer to follow a simpler approach by Onatski (2010). As it is shown in the empirical application, the two methods indicate that there exist three common factors.

3 Empirical Application

The data used in this paper are extracted and compiled from the Trade and Quote (TAQ). We use 50 largest capitalization stocks included in the S&P500 index. The data consists of full record transaction prices from January 2007 to December 2010. As in Müller et al. (1993) linear interpolation of logarithmic five-minute returns are used in all measures. We use a significance level of $\alpha = 0.1\%$ to detect jumps and construct the series for J_{it}, RV_{it}, BV_{it}. We find non-negligible number of significant jumps in our series.

As common factors are unobserved, we can apply the asymptotic principal component method to extract the r largest eigenvectors from $\widehat{hh}^{'}, \widehat{h} = [h_{1,} \ldots, h_T]$. We use the Bai and Ng(2002) panel decision criteria and find $r^{\max} = 3$, shown in Table 1.

The ED estimator grants the same number of factors, a decent result for the robustness check. In addition, the regression results from equation (6) give a good estimate with an average R^2 value of 0.8315 for the 50 stocks in consideration. Respecting the fact that underestimation of the number of factors may be more challenging than its overestimation, we try different numbers of factors. We observe that

Table 1: Selection Criteria for Common Factors

r	**PC**$_1(r)$	**PC**$_2(r)$	**IC**$_1(r)$	**IC**$_2(r)$
0	0.3186	0.3176	−2.1918	−2.1848
1	0.1871	0.1804	−2.8397	−2.8257
2	0.1061	0.1075	−3.5567	−3.5288
3	**0.0922**	**0.0938**	**−3.7890**	**−3.7541**
4	0.8666	0.8666	−0.8506	−0.8506

adjusted R^2, both in individual stocks and on average, decreases gradually with an increase in the number of factors.

We proceed with a thorough forecasting competition, comparing the $FB-RV-J$ model with the HAR model of Corsi (2009) and HAR-RV-J model of Andersen et al. (2007) for three forecasting horizons, two loss functions and R^2 of the Mincer-Zarnowitz (MZ) regressions. In addition, for comparison purposes, the standard GARCH (1,1), AR(1) and AR(3) models are added. The HAR regresses RV on three terms: the past one day, five days and 22 days average RVs. We also think about a variation of Corsi's (2009)'s model by incorporating factors into the regression. We abbreviate the resulting model F-HAR-RV, and after several combinations, the best results are obtained with only daily and weekly factors:

$$\begin{aligned} RV^d_{t,t+q} &= \beta^d RV^d_{t-1,t} + \beta^w RV^w_{t-5,t} + \beta^{bw} RV^{bw}_{t-10,t} + \beta^{tw} RV^{tw}_{t-15,t} + \beta^m RV^m_{t-20,t} \\ &\quad + \widehat{\alpha}^{d\prime} \widetilde{\widehat{f}}^{(d)}_t + \widehat{\alpha}^{w\prime} \widetilde{\widehat{f}}^{(w)}_t + \epsilon^d_{t,t+q} \end{aligned} \tag{7}$$

The HAR-RV-J model of Andersen et al. (2007) is:

$$RV_{t,t+h} = c + \beta^{(d)} RV_t + \beta^{(w)} RV_{t-5,t} + \beta^{(m)} RV_{t-22,t} + \beta^{(j)} J_t + \epsilon_{t+1} \tag{8}$$

where $J_{t+1}(m) = \max\left[RV_{t+1}(m) - BV_{t+1}(m), 0\right]$, the authors do not refer to any jump tests. Obviously, the difference $RV_t - BV_t$ may be non-zero in finite samples due to sampling variation even if no jump occurred during period t, which explains the rationale behind our decision to use the tests to identify the significant jump component mentioned in Section 2.1.

In order to evaluate the volatility forecasts, a benchmark IV has to be assumed, as in empirical applications the true one is latent. Hence, the standard practice is to use the best available estimate as the true IV.[1] Part of the literature on assessing the forecasting performance of daily models (see Hansen and Lunde, 2006) recommends using RV to evaluate forecast accuracy and we follow in their footsteps. The out-of-sample forecasts are obtained by estimating rolling models, with 85 days as a rolling window size.[2] Then, we directly compare the forecast models by testing the null hypothesis of equal predictive accuracy with the Diebold and Mariano (1995) (DM) statistic.

[1] Andersen et al. (2003) and Corsi (2009) use the RV estimates as true IV.

[2] Nowadays there is no consensus about which method to use in order to find an optimal rolling window size (Pesaran and Timmermann, 2007).

3.0.2 Comparing Predictive Accuracy

In this subsection, each competing model is fitted to examine the out-of-sample forecast accuracy using the 50 stocks from S&P500, while considering multiple prediction horizons, $h_{i,t+q}$ for $q = 1, 5, 10$ days. The main tool for forecast evaluations are expected losses. Patton (2011) proposes a family of robust loss functions and he suggests RMSE when comparing two imperfect forecasts the ranking can change depending on the choice of loss function.

Table 2: Forecasting Results

Comparison of the out-of-sample performances of the 1,5 and 10 day ahead forecasts of FB-RV-J, F-HAR-RV, HAR-RV, HAR-RV-J, GARCH(1,1), AR(1) and AR(3). Performance measures are the minimum, maximum, and average values of the root mean square error (RMSE), the mean absolute error (MAE), and the averageR^2 of the Mincer Zarnowitz regressions. The best model is in bold.

		$q=1$					
		RMSE			MAE		R^2
Competing Models	**Min**	**Max**	**Average**	**Min**	**Max**	**Average**	**Average**
FB-RV-J	**0.9168**	**4.4540**	**1.6580**	**0.3595**	**1.7243**	**0.8325**	**0.6946**
HAR-RV	1.9171	5.3187	2.7999	1.5134	3.8875	1.9141	0.3931
F-HAR-RV	1.3115	4.4811	2.0467	0.7662	1.9593	1.6891	0.5071
HAR-RV-J	1.1368	4.9662	2.4534	0.4167	2.7637	1.2965	0.4639
GARCH(1,1)	2.8029	7.1083	6.4997	2.5920	5.6712	3.8002	0.1441
AR(1)	4.4660	7.5482	8.4430	4.0894	7.1776	4.4962	0.1213
AR(3)	4.7251	10.1483	7.5060	3.6947	6.1502	4.0924	0.1129
		$q=5$					
		RMSE			MAE		R^2
	Min	**Max**	**Average**	**Min**	**Max**	**Average**	**Average**
FB-RV-J	**1.0102**	**4.5865**	**1.6550**	**0.3625**	**2.2274**	**0.7963**	**0.6541**
HAR-RV	2.8812	5.3187	3.4627	2.0678	3.8875	2.5769	0.3140
F-HAR-RV	2.5980	5.1048	2.7698	1.4561	3.3151	1.7665	0.4707
HAR-RV-J	1.9442	4.7432	2.9807	1.3025	2.5860	1.8267	0.4159
GARCH(1,1)	4.3810	8.9727	8.3229	4.1561	7.4401	5.4456	0.1108
AR(1)	6.1546	9.4417	10.3978	5.7539	9.0480	6.1878	0.0933
AR(3)	6.4298	12.2166	9.3975	5.3331	7.9513	5.7559	0.0868
		$q=10$					
		RMSE			MAE		R^2
	Min	**Max**	**Average**	**Min**	**Max**	**Average**	**Average**
FB-RV-J	**1.0313**	**4.6980**	**1.4049**	**0.3956**	**2.4233**	**0.6427**	**0.5335**
HAR-RV	3.1313	5.3337	3.8213	1.9141	3.8484	2.6493	0.2632
F-HAR-RV	2.5589	5.2253	2.9746	1.4561	3.0289	2.0527	0.4307
HAR-RV-J	2.2064	4.8878	3.4386	1.1759	3.0831	2.1672	0.3859
GARCH(1,1)	5.0820	10.3154	9.5743	4.8260	8.5690	6.2952	0.0775
AR(1)	7.1037	10.8497	11.9404	6.6467	10.4018	7.1409	0.0652
AR(3)	7.4170	14.0123	10.8005	6.1667	9.1515	6.6488	0.0607

We grid search over the evaluation of the forecast accuracy and consider MAE, RMSE, and R^2 of the MZ regressions (Table 2). Three main conclusions are extracted from the results. The difference in forecasting performance between the standard models and the ones using factors that capture the persistence of the empirical data is evident. The $FB - RV - J$ has the smallest RMSE and MAE but the largest R^2 of the MZ regression among all models, followed by F-HAR-RV, HAR-RV-J, HAR-RV, GARCH (1,1), AR (3) and AR (1). So, in terms of R^2, $FB - RV - J$ forecasts are more accurate than the others. It turns out that, the $FB-RV-J$ model steadily outperforms the others at all three time horizons considered.

The inference on the statistical significance of the RMSE of all models compared to the $FB-RV-J$ benchmark is performed using the DM test, with a Newey-West covariance estimator. In Table 3, considering the one step ahead column, the HAR-RV is favored compared to the benchmark only in 3 out of 50. On the other hand, the number of rejections in favor of our model is 43 out of 50. Hence, the percentage of rejections of the null hypothesis of equal prediction errors indicates that $FB-RV-J$ is the leading model, particularly in the further step ahead forecasts. We can observe the improvement in the results especially when separating continuous and significant jump components, showing the relevance of including this criterion in the analysis. Overall, we find that a large proportion of the RMSEs of the factor based models are statistically significant, confirming that the factors tend to improve the estimation and forecast performance of the realized estimators.

Table 3: Diebold Mariano Test Results

Table reports the DM test results at the 10% significance level. FB-RV-J is the factor based RV model, HAR-RV is the HAR model of Corsi (2009), F-HAR-RV is the HAR model with factors, HAR-RV-J is the model of Andersen et al. (2007). The forecasting exercise is performed using a rolling window of 85 days. DM1 and DM2 refer to the LHS and RHS of the test, respectively.

	$q=1$		$q=5$		$q=10$	
	DM1	**DM2**	**DM1**	**DM2**	**DM1**	**DM2**
HAR-RV	3	43	1	46	1	46
F-HAR-RV	4	39	3	41	2	42
HAR-RV-J	4	40	3	39	1	40
GARCH(1,1)	1	44	0	45	0	48
AR(1)	0	45	0	47	0	49
AR(3)	1	43	0	44	0	47

In Table 4, we also provide DM test results for one stock only (IBM) to save the space.[3] In addition to the benchmark analysis, we present the pair wide test results of the competing models. The F-HAR-RV is preferred when it is compared with the HAR-RV-J and HAR-RV, pointing out how factors can enhance the results. The DM test results are generally in line with the MAE and RMSE; the $FB-RV-J$ is performing significantly better than the others.

[3] All other results for the rest of the stocks are included in the working paper version.

Table 4: Diebold Mariano Test Results (IBM)

Table reports the DM test results, the * indicates which model is favored. FB-RV-J is the factor based RV model, HAR-RV is the HAR model of Corsi (2009), F-HAR-RV-F is the HAR model with factors, HAR-RV-J is the model of Andersen et al. (2007). The forecasting exercise is performed using a rolling window of 85 days. So, if two forecasts A and B, respectively exist as $\widehat{y}^{A}_{t+h|t}$ and $\widehat{y}^{B}_{t+h|t}$, then using the forecasts, there would be two loss functions which are defined as $l^{A}_{t} = (y_{t+h} - \widehat{y}^{A}_{t+h|t})^2$ and $l^{B}_{t} = (y_{t+h} - \widehat{y}^{B}_{t+h|t})^2$, where losses do not need to be MSE. Accordingly, Diebold Mariano (DM) is implemented as a $t-test$ for $E[\delta_t] = 0$, where $\delta_t = l^{A}_{t} - l^{B}_{t}$, and $H_0 : E[\delta_t] = 0$, and $H^{A}_{1} : E[\delta_t] < 0$, $H^{B}_{1} : E[\delta_t] > 0$. The sign indicates which model is favoured. Therefore, reject if $|t| > C_\alpha$, where C_α is the critical value for a 2-sided test using a normal distribution with size α. If significant, reject in favour of model A if test statistic is negative, or reject in favour of model B if test statistic is positive.

Competing Models	$q = 1$	$q = 5$	$q = 10$
FB-RV-J* vs HAR-RV	-5.5927	-6.0222	-7.7834
FB-RV-J* vs HAR-RV-J	-4.1660	-5.4430	-5.0518
FB-RV-J* vs F-HAR-RV	-3.9344	-5.3484	-4.6222
HAR-RV-J vs F-HAR-RV*	2.7269	5.8283	-2.5538
F-HAR-RV* vs HAR-RV	-4.5503	-4.8250	-6.6222
HAR-RV vs HAR-RV-J*	3.8081	5.0859	6.0755

4 Concluding Remarks

This paper examines the role of approximate factors in forecasting future RV. For an enhanced forecasting performance, we begin with identifying the discontinuous components using the jump tests before applying factors. We then relate the common component to the unobservable financial characteristics. Both on the methodological and substantive side, our $FB - RV - J$ model outperforms the currently available approaches with regards to its forecast accuracy and efficiency at various prediction horizons. Overall, we believe that our results are appealing and complement the burgeoning RV literature.

Acknowledgments: We have greatly benefitted from comments by Oliver Linton. We are also grateful for advice from Fulvio Corsi, as well participants in the CEF, 2010 meeting in London. The usual disclaimer applies.

5 References

Andersen, TG., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility model do provide accurate forecast. International Economics Review 39, 4.

Andersen, TG., Bollerslev, T., Diebold, FX., 2007. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. Review of Economics and statistics 89, 701-720.

Andersen, TG., Bollerslev, T., Diebold, FX., Labys P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579-625.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-222.

Bai, J., Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics 131, 507-537.

Barndorff-Nielsen, OE., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of Royal Statistical Society B 64, 253-280.

Barndorff-Nielsen, OE., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1-37.

Barndorff-Nielsen, OE., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. Journal of Financial Econometrics 4, 1-30.

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

Corsi. F., Pirino, D., Rena, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics 159(2), 276-288.

Diebold, FX., Mariano, RS., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13(3), 253-63.

Hansen, PR., Lunde, A., 2006. Consistent ranking of volatility models. Journal of Econometrics 131, 97-121.

Kapetanios G. 2010. A testing procedure for determining the number of factors in approximate factor models with large datasets. Journal of Business and Economic Statistics 28(3), 397-409.

Lee, SS., Mykland, PA., 2008. Jumps in financial markets: a new nonparametric test and jump dynamics. Review of Financial Studies 21, 2535-2563.

Ludvigson, S., Ng, S., 2009. Macro factors in bond risk premia. The Review of Financial Studies 22(12), 5027-5067.

Müller, U., Dacorogna, M., Dav, R., Pictet, O., Olsen, R., Ward, J., Fractals and intrinsic time - a challenge to econometricians, WP 1993-08-16, Olsen and Associates, 1993.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economics and Statistics 92: 1004-1016.

Patton, AJ., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246-256.

Pesaran, M.H, Timmermann, A., 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics 137, 134-161.