Volatility Determinants: Heterogeneity, Leverage, and Jumps

Fulvio Corsi†  Roberto Renò‡

February 10, 2009

Abstract

We identify three main endogenous determinants in the dynamics of asset price volatility, namely heterogeneity, leverage, and jumps. We find that each of the three components plays a significant role in volatility forecasting and neglecting one of them is detrimental to the forecasting performance. Importantly, we find remarkable forecasting power for the negative past returns at all the considered frequencies, which unveils a novel heterogeneous structure in the leverage effect. We also show, using simulation studies, that the presence of jumps is important for two distinct reasons. Firstly, explicitly modeling jumps has trimming effect on the dynamics of the persistent volatility component. Secondly, they have a positive and significant impact on future volatility, although of a short-lived nature.

JEL classification: C13; C22; C51; C53

Keywords: Volatility Forecasting; High Frequency Data; HAR; Leverage Effect; Jumps.

*Comments are welcome. We would like to acknowledge Davide Pirino for research assistance and Giampiero Gallo and Roel Oomen for useful suggestions.
†Università di Siena, University of Lugano, and Swiss Finance Institute, E-mail: fulvio.corsi@lu.unisi.ch
‡Università di Siena, Dipartimento di Economia Politica, E-mail: reno@unisi.it
1 Introduction

Volatility forecasting is a key ingredient in many financial problems. However, volatility dynamics displays well known stylized facts which pose serious challenges to standard econometric models. Volatility is a clustered and highly persistent process with a memory decay of several months. Equity and stock-index volatilities show significant asymmetric response to past returns. More precisely, volatility tends to increase more after a negative shock than after a positive shock of the same magnitude (Bollerslev et al., 2006). This asymmetric return-volatility dependence, first noted by Black (1976), is usually called the “leverage effect”. Moreover, price process shows presence of sudden large price changes, the so called jumps, which arguably have an important impact on volatility dynamics.

In Corsi (2009) a simple Heterogeneous Auto-Regressive of Realized Volatility (HAR-RV) model has been proposed to capture the empirical memory persistence of volatility in a simple and parsimonious way. In this paper, we propose an extended version of the HAR-RV model which considers asymmetric responses of the realized volatility not only to previous daily returns but also to past weekly and monthly returns. Our main contribution is then to show that the heterogeneous structure applies to the leverage effect as well, thus reinforcing the Heterogenous Market Hypotesis of Muller et al. (1997).

In addition we study the impact on future volatility of jumps measured over the same three different horizons. Given the inadequacy of bipower variation in measuring volatility in presence of jumps, we use the tests and measures introduced by Corsi et al. (2008) which provide a better identification and more precise measurement of jumps, and uncover the significant impact of jumps on future volatility. We confirm this finding out-of-sample. We also show, by means of a simulation study, that the presence of jumps is important for two distinct reasons. First, it has a direct impact on volatility dynamics which may be explained by the presence of contemporaneous jumps in price and volatility. Second, it has a trimming effect on the volatility series, which allows a better fit of the realized volatility, as suggested by Andersen et al. (2007).

Moreover, we conduct robustness tests to check whether other volatility measures proposed in the literature (such as absolute variation, range, or semivariance) contain additional useful information which are not captured in our model specification. The results of these comparisons show that the other volatility measures either drop out or only marginally contribute to the performance of the model. Hence, the proposed model seems to capture the main determinants of volatility dynamics.

Summarizing, in this paper we are proposing a relatively simple model that incorporates three main determinants of volatility dynamics, namely: heterogeneity, leverage and jumps. When estimating this model on the S&P500 time series, we find that each component has a distinct forecasting power, both in-sample and out-of sample. While, other data-driven measures of volatility only marginally add to our model.

The paper is organized as follows. Section 2 reviews the HAR model and presents its possible extensions.
with heterogeneous leverage and jumps. Section 3 describes the empirical in-sample and out-of-sample analysis on a long serie of high frequency S&P500 futures data. Section 4 discuss the results in the light of two Monte Carlo simulation models with independent and contemporaneous jumps and Section 5 contains some concluding remarks.

2 Modelling volatility

Assume that the state variable $X$, which may be thought as an economic variable (an interest rate or a stock log-price), is driven by the stochastic process:

$$
\frac{dX_t}{\sigma_t} = \mu_t dt + \sigma_t dW_t + c_t dN_t
$$

where $\mu_t$ is predictable, $\sigma_t$ is càdlàg and $N_t$ is a doubly stochastic Poisson process whose intensity is an adapted stochastic process $\lambda_t$, the times of the corresponding jumps are $(\tau_j)_{j=1,...,N_T}$ and $c_j$ are i.i.d. adapted random variables measuring the size of the jump at time $\tau_j$. In practice, e.g. for risk management purposes, we are interested in forecasting the quadratic variation defined as:

$$
\tilde{\sigma}_t = \int_t^{t+1} \sigma_s^2 ds + \sum_{t \leq \tau_j \leq t+1} c_{\tau_j}^2,
$$

where the time unit is one day. We estimate quadratic variation using $n$ observations of the state variable in the interval $[0,T]$. The most popular estimator is realized volatility, which, after defining $\Delta_{t,j} X = X_{t+j/n} - X_{t+(j+1)/n}$, is given by:

$$
RV_t = \sum_{j=0}^{n-1} (\Delta_{t,j} X)^2
$$

which is a consistent estimator, as $n \to \infty$, of $\tilde{\sigma}_t$, see Andersen et al. (2003) for a review. Other estimators have been devised, such as the range (see e.g. Alizadeh et al. 2002 and Brandt and Jones 2006) or refinement of realized volatility to account for the presence of microstructure noise, as those in Zhang et al. (2005), Barndorff-Nielsen et al. (2008), or Jacod et al. (2007). We indicate by $\tilde{V}_t$ a generic unbiased estimator of quadratic variation such that (working with logarithms to avoid negativity issues):

$$
\log \tilde{\sigma}_t = \log \tilde{V}_t + \omega_t
$$

where $\omega_t$ is zero mean and finite variance measurement error.$^1$

We are interested in modelling the dynamics of $\tilde{\sigma}_t$, that is the dynamics of quadratic variation, a topic which has received growing attention in the last decade.

$^1$In our empirical analysis, we use the two-scales estimator of Zhang et al. (2005).
2.1 Heterogeneity

The need for heterogeneity of volatility components, advocated by Muller et al. (1997), has been reconsidered in the work of Corsi (2009) by making use of the concept of volatility cascades. In what follows, we review this latter approach working with logarithmic transformations to avoid negativity issues and get approximately Normal distribution for the volatility measure. Consider the aggregated values of \( \hat{\sigma}_t \), defined as:

\[
\log \hat{\sigma}_t^{(n)} = \frac{1}{n} \left( \log \hat{\sigma}_t + \ldots + \log \hat{\sigma}_{t-n+1} \right)
\]

and assume having two different time scales, of length \( n_1 \) and \( n_2 \), with \( n_1 > n_2 \). For the largest time scale, assume that \( \tilde{\sigma}_t \), once aggregated as in (2.4) is determined by:

\[
\log \tilde{\sigma}_{t+n_1}^{(n_1)} = \eta(n_1) + \beta(n_1) \log \hat{\sigma}_t^{(n_1)} + \epsilon_t^{(n_1)}
\]

where \( \epsilon_t^{(n_1)} \) is IID zero mean and finite variance noise independent on \( \omega_t \).

It has been recently suggested that volatility over longer time intervals has stronger influence on those over shorter time intervals than conversely, suggesting a volatility cascade from low to high frequencies.\(^2\) This can be economically explained by noticing that for short-term traders the level of long term volatility matters because it determines the expected future size of trends and risk. On the other hand, the level of short-term volatility does not affect the trading strategies of long-term traders. The shorter time scale \( n_2 \) is assumed to be influenced by the expected future value of the largest time scale \( n_1 \), so that:

\[
\log \tilde{\sigma}_{t+n_2}^{(n_2)} = \eta(n_2) + \beta(n_2) \log \hat{\sigma}_t^{(n_2)} + \delta(n_2) \mathbb{E}_t \left[ \log \tilde{\sigma}_t^{(n_1)} \right] + \epsilon_t^{(n_2)}
\]

with \( \epsilon_t^{(n_2)} \) IID zero mean and finite variance noise independent on \( \epsilon_t^{(n_1)} \) and \( \omega_t \).

By substitution, and using equation (2.3), this gives:

\[
\log \hat{\sigma}_t^{(n_2)} = \eta(n_2) + \beta(n_2) \log \hat{\sigma}_t^{(n_2)} + \beta(n_1) \log \hat{\sigma}_t^{(n_1)} + \epsilon_t
\]

where \( \epsilon_t \) is IID noise depending on \( \epsilon_t^{(n_1)}, \epsilon_t^{(n_2)}, \omega_t \). The model (2.7) can be easily extended to \( d \) horizons of length \( n_1 > n_2 > \ldots > n_d \). Typically, three components are used with length \( n_1 = 22 \) (monthly), \( n_2 = 5 \) (weekly), \( n_3 = 1 \) (daily). Since volatility at shorter time horizons is influenced by volatility at longer horizons, the auto-correlation function of the model and hence its memory persistence increases. Thus, even if the HAR model does not formally belong to the class of long memory processes, it fits the persistence properties of financial data as well as true long memory models, such as the fractionally integrated one, which, however, are much more complicated to estimate and to deal with (see the review of Banerjee and Urga 2005).

The HAR model has been employed in several applications in the literature. Corsi et al. (2008) use it to study the volatility of realized volatility; Ghysels et al. (2006) and Forsberg and Ghysels (2007) compare

\(^2\)See Müller et al. (1997), Arneodo et al. (1998), Lynch and Zumbach (2003). However, the HAR model would hold even if we allow the short-term volatility to affect the long-term volatility, although this would be at odds with the empirical findings (see Section 3.1).
this model with the MIDAS model; Andersen et al. (2007) use an extension of this model to forecast the volatility of stock prices, foreign exchange rates and bond prices; Clements et al. (2008) implement it for risk management with VaR measures; Bollerslev et al. (2008) use it to analyze the risk-return tradeoff.

2.2 Leverage effects

It is well known that equities and stock indexes often exhibit the so-called “leverage effect”, i.e. volatility tends to increase more after a negative shock than after a positive shock of the same size. By extending the Heterogeneous Market Hypothesis approach to leverage effect, we consider asymmetric responses of realized volatility not only to previous daily returns but also to past weekly and monthly returns.

We model such heterogeneous leverage effects by introducing asymmetric return-volatility dependence at each level of the cascade considered in the above section. Define daily returns \( r_t = X_t - X_{t-1} \) and past aggregated negative and positive returns as:

\[
\begin{align*}
    r_t^{(n)} &= \frac{1}{n} (r_2 + \ldots + r_{t-n}) I_{\{r_2 + \ldots + r_{t-n} \geq 0\}} \\
    r_t^{(-n)} &= \frac{1}{n} (r_2 + \ldots + r_{t-n}) I_{\{r_2 + \ldots + r_{t-n} < 0\}}
\end{align*}
\]

where \( I_{\{\cdot\}} \) denotes the indicator function. We assume that integrated volatility is determined by the cascade:

\[
\log \hat{\sigma}_{t+n_1} = c(n_1) + \gamma(n_1) - r_t^{(n_1)} + \beta(n_1) \log \hat{V}_t + \gamma(n_1) + r_t^{(n_1)} + \epsilon_{t+n_1}
\]

\[
\log \hat{\sigma}_{t+n_2} = c(n_2) + \beta(n_2) \log \hat{V}_t + \gamma(n_2) - r_t^{(n_2)} + \gamma(n_2) + r_t^{(n_2)} + \delta(n_2) E_t [\log \hat{\sigma}_{t+n_1}] + \epsilon_{t+n_2}
\]

which now gives:

\[
\log \hat{V}_{t+n_2} = c + \beta(n_2) \log \hat{V}_t + \beta(n_1) \log \hat{V}_t + \gamma(n_2) - r_t^{(n_2)} + \gamma(n_2) + r_t^{(n_2)} + \gamma(n_1) - r_t^{(n_1)} + \gamma(n_1) + r_t^{(n_1)} + \epsilon_t
\]

We then postulate that leverage effects influence each market component separately, and that they appear aggregated at different horizons in the volatility dynamics. Note that the inclusion of both negative and positive returns is equivalent, by linearity, to the inclusion of negative (or positive) and total returns.

2.3 Jumps

The importance of jumps in financial econometrics is rapidly growing. Recent research focusing on jumps detection and volatility measuring in presence of jumps includes Barndorff-Nielsen and Shephard (2004); Mancini (2007), Lee and Mykland (2007), Jiang and Oomen (2008), Aït-Sahalia and Mancini (2008), Aït-Sahalia and Jacod (2008), Christensen et al. (2008), and Boudt et al. (2008). Andersen et al. (2007) suggested that the continuous volatility and jump component have different dynamics and should thus
be modelled separately. In this section, we follow closely Corsi et al. (2008) using the C-Tz statistics to detect the occurrence of the jump in a single day, and threshold bipower variation to measure the continuous part of integrated volatility, defined as:

\[ \text{TBPV}_t = \frac{2}{\pi} \sum_{j=0}^{n-2} |\Delta t,j X| \cdot |\Delta t,j+1 X| I_{\{\Delta t,j X^2 \leq \theta_j \}} I_{\{\Delta t,j+1 X^2 \leq \theta_j \}} \]  

(2.11)

where \( \theta_j \) is a threshold function which we estimate as in Corsi et al. (2008). This continuous volatility estimator has better finite sample properties than standard bipower variation and provides more accurate jump tests, which allows for a corrected separation of continuous and jump components. To this purpose, we fix a confidence level \( \alpha \) and estimate the jump component as:

\[ J_t = I_{\{C-Tz > \Phi_\alpha \}} \cdot \left( \tilde{V}_t - \text{TBPV}_t \right)^+ \]  

(2.12)

where \( \Phi_\alpha \) is the value of the standard Normal distribution corresponding to the confidence level \( \alpha \), and \( x^+ = \max(x, 0) \). The corresponding continuous component is defined as:

\[ C_t = \tilde{V}_t - J_t, \]  

(2.13)

which is equal to \( \tilde{V}_t \) if there are no jumps in the trajectory and to \( \text{TBPV}_t \) if a jump is detected with the C-Tz statistics.

Figure 1 reports the percentage contribution of jumps estimated by (2.12) to total quadratic variation computed on a 3-month and 1-year moving window for the full S&P 500 futures sample. In line with the results in Andersen et al. (2007) and Huang and Tauchen (2005) we find a jumps contribution varying between 2% and 20% of total variation (with an overall sample mean of about 6%), showing a higher level at the beginning of the sample and an increasing trend toward the end of the sample period.

In the volatility cascade we assume that \( C_t \) and \( J_t \) enter separately at each level of the cascade, that is:

\[
\log \tilde{\sigma}_{t+n1}^{(n1)} = \alpha^{(n1)} \log (J_t^{(n1)} + 1) + \beta^{(n1)} \log C_t^{(n1)} + \gamma^{(n1)} - r_t^{(n1)} + \epsilon_{t+n1}^{(n1)} \\
\log \tilde{\sigma}_{t+n2}^{(n2)} = \alpha^{(n2)} \log (J_t^{(n2)} + 1) + \beta^{(n2)} \log C_t^{(n2)} + \gamma^{(n2)} - r_t^{(n2)} + \epsilon_{t+n2}^{(n2)} \\
\quad + \delta^{(n2)} E_t \left[ \tilde{\sigma}_{t+1}^{(n1)} \right] + \epsilon_t^{(n2)}
\]

originating the model:

\[
\log \tilde{V}_{t+n2}^{(n2)} = c + \alpha^{(n1)} \log (J_t^{(n1)} + 1) + \alpha^{(n2)} \log (J_t^{(n2)} + 1) + \beta^{(n2)} \log C_t^{(n2)} + \beta^{(n1)} \log C_t^{(n1)} + \gamma^{(n2)} - r_t^{(n2)} + \gamma^{(n1)} - r_t^{(n1)} + \epsilon_t^{(n2)}
\]

(2.14)

2.4 The LHAR-CJ model

Combining heterogeneity in realized volatility, leverage, and jumps we construct the Leverage Heterogeneous Auto-Regressive with Continuous volatility and Jumps (LHAR-CJ) model. As it is common in
Jump Contribution to Total Variation

Figure 1: Percentage contribution of daily jump estimated by (2.12) to total quadratic variation measured over a moving window of 3-month (dotted line) and 1-year (solid line) for the S&P500 futures from January 1990 to December 2007 (4344 days). The C-Tz statistics in (2.12) is computed with a confidence interval $\alpha = 99.9\%$.

practice, we use three components: daily, weekly and monthly. Hence, the proposed model reads:

$$\log \hat{\gamma}_{t+h}^{(h)} = c + \alpha^{(d)} \log(1 + J_t) + \alpha^{(w)} \log(1 + J_t^{(5)}) + \alpha^{(m)} \log(1 + J_t^{(22)})$$

$$+ \beta^{(d)} \log C_t + \beta^{(w)} \log C_t^{(5)} + \beta^{(m)} \log C_t^{(22)}$$

$$+ \gamma^{(d)} \frac{1}{r_t^{(1)}} + \gamma^{(w)} \frac{1}{r_t^{(5)}} + \gamma^{(m)} \frac{1}{r_t^{(22)}} + \xi_t^{(h)},$$

We estimate model (2.15) by OLS with Newey-West covariance correction for serial correlation. In order to make multiperiod predictions we will estimate the model considering the aggregated dependent variable $\log \hat{\gamma}_{t+h}^{(h)}$, with $h$ ranging from 1 to 22 i.e. from one day to one month. While, strictly speaking, models with $h > 1$ would require a cascade specification with longer frequencies multiple of $h$, for simplicity and comparison purposes, we will always retain the standard cascade specification with the three natural frequencies of one day, one week and one month. This can be viewed as a simplifying approximation justified by its empirically good performances.

3 Empirical evidence

The purpose of this section is to empirically analyze the main determinants of future asset volatility. Our data set covers a long time span of almost 18 years of high frequency data for the S&P 500 futures from the
January 1990 to June 2007 (4,344 days). In order to mitigate the impact of microstructure effects on our estimates, the daily volatilities $\hat{V}_t$ are computed with the two-scales estimator proposed by Zhang et al. (2005). Aït-Sahalia and Mancini (2008) show that using the two-scales estimator instead of standard realized volatility measures yields significant gains in volatility forecasting. The TBPV measure (2.11) for jump detection is computed at the sampling frequency of 5 minutes (corresponding to 84 returns per day).

3.1 In-sample analysis

The LHAR-CJ model is estimated using, as a dependent variable, realized volatility aggregated at different horizons. The results of the estimation of the LHAR-CJ when forecasting the S&P500 realized volatility at 1 day, 1 week, 2 weeks and 1 month are reported in Table 1, together with their statistical significance evaluated with the Newey-West robust t-statistic. As usual, all the coefficients of the three continuous volatility components are positive and, in general, highly significant. It is, however, interesting to remark that, while the coefficient which measures the impact of monthly volatility on daily volatility is highly significant at all horizons, the opposite does not hold, confirming the hierarchical asymmetric propagation of the volatility cascade presented in Section 2.

The daily and weekly jump components remain highly significant and positive for the shorter horizon realized volatility and become insignificant for the longer ones. In particular, while the daily jumps stay strongly significant up to the weekly horizon, the weekly jumps remain significant up to the 2-week horizon. Figure 2 plots the $t$-statistics of the impact of the daily jump on aggregated volatility at different time horizons, confirming, with its rapid decline, that daily jumps affects future volatilities only over a short period of about one week. The jumps aggregated at monthly level turn out to be always insignificant and might be removed by our specification.

The most interesting result is, however, the strong significance (with an economically sound negative sign) of the negative returns at all the daily, weekly and monthly frequencies which unveils an heterogeneous structure in the leverage effect as well. Not only daily negative returns affect the next day volatility (the well know leverage effect), but, in addition, also the negative returns of the past week and past month have an impact on forthcoming volatility, which is even stronger than that of the previous day. This finding leads to the conclusion that the market aggregates daily, weekly and monthly memory, observing and reacting to price declines happened in the past week and month. To our knowledge, this is a novel empirical finding that further confirms the views of the Heterogeneous Market Hypothesis.

On the contrary, positive returns are almost insignificant in predicting future volatility and may be removed from a baseline model specification. However, it is interesting to remark that the $t$-statistics

3The two-scales estimator combines two realized volatilities computed at two different frequencies, where the slower one is computed by subsampling and averaging while the faster one (being a proxy for the variance of microstructure noise) is used for bias correction. In our implementation of the two-scales estimator we use a slower frequency of ten ticks returns.
S&P500 LHAR in-sample regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>One day</th>
<th>One week</th>
<th>Two weeks</th>
<th>One month</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.571*</td>
<td>0.933*</td>
<td>1.326*</td>
<td>2.031*</td>
</tr>
<tr>
<td></td>
<td>(6.993)</td>
<td>(3.263)</td>
<td>(2.751)</td>
<td>(2.625)</td>
</tr>
<tr>
<td>$C$</td>
<td>0.236*</td>
<td>0.145*</td>
<td>0.101*</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(10.205)</td>
<td>(7.354)</td>
<td>(4.299)</td>
<td>(1.531)</td>
</tr>
<tr>
<td>$C(5)$</td>
<td>0.372*</td>
<td>0.343*</td>
<td>0.289*</td>
<td>0.177*</td>
</tr>
<tr>
<td></td>
<td>(11.450)</td>
<td>(7.961)</td>
<td>(5.652)</td>
<td>(2.760)</td>
</tr>
<tr>
<td>$C(22)$</td>
<td>0.264*</td>
<td>0.310*</td>
<td>0.326*</td>
<td>0.340*</td>
</tr>
<tr>
<td></td>
<td>(9.652)</td>
<td>(6.490)</td>
<td>(4.863)</td>
<td>(3.507)</td>
</tr>
<tr>
<td>$J$</td>
<td>0.038*</td>
<td>0.016*</td>
<td>0.008</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(5.004)</td>
<td>(3.213)</td>
<td>(1.444)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>$J(5)$</td>
<td>0.012*</td>
<td>0.013*</td>
<td>0.014*</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(3.031)</td>
<td>(2.390)</td>
<td>(2.499)</td>
<td>(1.274)</td>
</tr>
<tr>
<td>$J(22)$</td>
<td>-0.001</td>
<td>-0.003</td>
<td>-0.008</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(-0.323)</td>
<td>(-0.711)</td>
<td>(-1.251)</td>
<td>(-1.113)</td>
</tr>
<tr>
<td>$r^-$</td>
<td>-0.008*</td>
<td>-0.007*</td>
<td>-0.006*</td>
<td>-0.006*</td>
</tr>
<tr>
<td></td>
<td>(-8.074)</td>
<td>(-7.874)</td>
<td>(-5.481)</td>
<td>(-3.401)</td>
</tr>
<tr>
<td>$r(5)^-$</td>
<td>-0.012*</td>
<td>-0.006*</td>
<td>-0.008*</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(-4.629)</td>
<td>(-2.013)</td>
<td>(-2.108)</td>
<td>(-1.920)</td>
</tr>
<tr>
<td>$r(22)^-$</td>
<td>-0.011*</td>
<td>-0.026*</td>
<td>-0.035*</td>
<td>-0.051*</td>
</tr>
<tr>
<td></td>
<td>(-2.209)</td>
<td>(-2.485)</td>
<td>(-2.205)</td>
<td>(-2.256)</td>
</tr>
<tr>
<td>$r^+$</td>
<td>0.000</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.920)</td>
<td>(1.523)</td>
<td>(1.552)</td>
</tr>
<tr>
<td>$r(5)^+$</td>
<td>-0.002</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(-1.157)</td>
<td>(-0.232)</td>
<td>(0.440)</td>
<td>(0.824)</td>
</tr>
<tr>
<td>$r(22)^+$</td>
<td>0.006</td>
<td>0.014*</td>
<td>0.021*</td>
<td>0.028*</td>
</tr>
<tr>
<td></td>
<td>(1.496)</td>
<td>(2.319)</td>
<td>(2.697)</td>
<td>(2.859)</td>
</tr>
</tbody>
</table>

$R^2$        | 0.7333  | 0.7607  | 0.7178    | 0.6074    |
RMSE         | 3.3026  | 2.7087  | 2.7325    | 3.0128    |

Table 1: OLS estimate for LHAR-CJ regressions, model (2.15), for S&P500 futures from January 1990 to December 2007 (4,344 observations). The LHAR-CJ model is estimated on 1-day, 1-week, 2-week and 1-month realized volatility. The significant jump are computed using a critical value of $\alpha = 99.9\%$. Reported in parenthesis are $t$-statistics based on Newey-West correction.
of monthly positive returns increase with the forecasting horizon, becoming mildly significant at weekly and longer horizons. This suggests that, for longer horizons, price trends affect future volatility, mildly corroborating the results in Zumbach (2005).

In order to evaluate the relative contribution of the different volatility determinants we compare the in-sample prediction of the LHAR-CJ model over different horizons with those of the standard HAR model and the HAR model with jumps but with no leverage effects (the HAR-CJ model of Corsi, Pirino, and Renò 2008). For each horizon, ranging from one day to one month, the forecasts are obtained by first estimating the parameters of the models on the full sample and then performing a series of static one-step-ahead forecasts. The forecasts of the different models are evaluated on the basis of the RMSE in predicting the square root of $\hat{V}$. The results, reported in Figures 3, show unambiguously that the inclusion of both the heterogeneous jumps and the heterogeneous leverage effects considerably improve the forecasting performance of the S&P 500 volatility at any forecasting horizon. In particular, the inclusion of heterogeneous leverage effect provides a relevant overall benefit in the in-sample performance.

3.2 Robustness to other volatility measures

In the literature many volatility measures have been proposed to better capture the dynamics of volatility. Forsberg and Ghysels (2007) proposed the use of realized absolute variation (RAV) which shows a more persistent dynamics than realized volatility being more robust to microstructure noise and jumps. The range has also been found to be significant by many authors, see e.g. Brandt and Jones (2006) and Engle and Gallo (2006). Motivated by the analysis of Bandi et al. (2008) who found liquidity to be a significant factor in asset pricing, we also compute the sum of squared tick-by-tick returns as a liquidity measure and employ it as a volatility factor. Recently, Barndorff-Nielsen et al. (2008) proposed the realized semivariance as the sum of square negative returns to capture the impact on volatility of downward price pressures. Visser (2008) combines RAV and semivariance by taking the sum of negative absolute squared returns.

In the spirit of Forsberg and Ghysels (2007), we compare the relative explanatory power of different volatility measures by estimating the following set of models (for space concerns we limit ourself to the one day horizon, thus we do not include positive returns which are significant at longer horizons only). First, we estimate a baseline LHAR-CJ model, containing as explanatory variables: heterogeneous continuous volatility, heterogeneous negative returns and daily jumps. Then we add the different volatility measures to the baseline LHAR-CJ model. Some of these measures turn out to be fairly related to the jump one (range and semivariance). For those measures we also estimate models where the daily jump regressor is removed so that a direct performance comparison with the LHAR-CJ is possible. Estimation results are reported in Table 2.

The liquidity proxy (LQ) and the realized absolute variation (RAV) computed at 5-minute frequency turn out to be not significant when included in the LHAR-CJ model. In particular, our result for the
Figure 2: t-statistics of daily jump coefficients for LHAR-CJ model estimated on S&P500 from January 1990 to December 2007 (4344 days) as a function of the forecasting horizon $h$.

Figure 3: RMSE of static in sample one-step ahead forecasts for realized volatility ranging from 1 day to 1 month of the S&P500 from January 1990 to December 2007 (4344 observations). The forecasting models are the standard HAR with only heterogeneous volatility, the HAR-CJ with heterogeneous jumps and the LHAR-CJ model.
RAV seems to contrast with findings in the literature of a higher explanatory power of RAV vs. RV due to a higher robustness of the RAV to microstructure noise and jumps. However, here the 5-minute RAV is confronted with the tick-by-tick two-scales measures cleaned from the jumps component by the C-Tz test, hence with a highly precise measure which (contrary to RV) is also robust to microstructure noise and jumps.

In line with previous literature, we find that the range has a significant impact on future volatility. However, it seems to be mainly a substitute for continuous volatility and jumps, which is not totally surprising since the range is an estimator (though noisy) of total quadratic variation. Indeed, when the range replaces the jumps (LHAR-Range model), the coefficients of daily continuous volatility almost halves. The $R^2$ of the two competing regressions (LHAR-Range and LHAR-CJ) is practically the same. When the range is inserted together with the jumps (LHAR-CJ-Range), both the coefficients of daily volatility and jumps decrease, although they remain highly significant. While, the significance of the heterogeneous leverage effect is untouched by the presence of the range. The $R^2$ of the encompassing regression increases marginally. We thus conclude that the range, while partially proxying for both volatility and jump, is also able to capture some other effect which is not captured by other variables, but it adds very little to the economic and statistical value of the LHAR-CJ model.

We found similar results for the realized semivariance of Barndorff-Nielsen et al. (2008) and the downward absolute power variation of Visser (2008) (being very similar, only the result for the realized semivariance are reported).\footnote{All results are available from the authors upon request.} As for the range, realized semivariance (LHAR-semiRV and LHAR-CJ-semiRV) is significant in explaining future volatility, and it seems very correlated with both the daily two-scales estimator and the jumps (typically depleting the significance of the corresponding coefficients without totally removing it), while unrelated with the leverage. The contribution of the realized semivariance to the model performance is very similar to that of the range. However, when they are included together in the LHAR-CJ model (LHAR-CJ-Range-semiRV) they both remain significant.

Summarizing, the results of this section show that when the other volatility measures proposed in the literature are inserted in the baseline LHAR-CJ model they either drop out or only marginally contribute to the performance of the model. Hence, the LHAR-CJ model seems to capture the main determinants of volatility dynamics.

### 3.3 Out-of-sample analysis

In this section, we appreciate the performance of the LHAR-CJ model on the basis of true out-of-sample analysis. For the out-of-sample forecast of $\hat{V}$ on the $[t, t+h]$ interval we keep the same forecasting horizons ranging from one day to one month and reestimate the model at each day $t$ on an increasing window of all the observation available up to time $t−1$. The out of sample forecasting performance for the square root
Figure 4: RMSE of out-of-sample forecasts for realized volatility ranging from 1 day to 1 month of the S&P500 from January 1994 to December 2007 (3344 observations, the first 1000 observation are used to initialize the models). The forecasting models are the standard HAR with only heterogeneous volatility, the HAR-CJ with heterogeneous jumps and the LHAR-CJ model.

The superiority of the LHAR-CJ model at all horizons is confirmed, validating the importance of including both the heterogeneous leverage effects and jumps in the forecasting model. It is interesting to note that the improvements in forecasting performance due to the inclusion of jumps (HAR-CJ vs HAR) is still valid out-of-sample, reinforcing the results in Corsi et al. (2008). For longer horizons the performance differences among the three models tend to decrease.

Finally, it is important to note that the inclusion of the jump component helps also in forecasting longer horizon volatility, which seems at odds with the short-lived nature of the impact of jumps on future volatility, as witnessed by the significance of the $\alpha$ coefficients in Table 1. To clarify this issue, we perform a Monte Carlo simulation analysis described in the following section.

### 4 A simulation study

We evaluate our empirical results for jumps through the lens of a Monte Carlo simulations. We simulate the stock index price with the flexible specification of Eraker et al. (2003), that is:

$$
\left( \begin{array}{c} dY_t \\ dV_t \end{array} \right) = \left( \begin{array}{c} \mu \\ \kappa (\theta - V_t) \end{array} \right) dt + \sqrt{V_t} \left( \begin{array}{cc} 1 & 0 \\ \sigma_v \rho & \sigma_v \sqrt{1 - \rho^2} \sigma_v \end{array} \right) dW_t + \left( \begin{array}{c} \xi^y dN^y_t \\ \xi^v dN^v_t \end{array} \right) \tag{4.1}
$$

where $W_t$ is a bidimensional Brownion motion and $dN^y$ and $dN^v$ are Poisson processes with intensity

...
λ^y and λ^V respectively; ξ^y is normally distributed, while ξ^V has an exponential law. As in Eraker et al. (2003), we consider two cases: the case in which dN^y is independent from dN^V (what they name the SVIJ model) and the case in which dN^y = dN^V (what they name the SVCJ model), and we hold their terminology. We use exactly the parameters estimated by Eraker et al. (2003) for the S&P500 time series.

Figure 5 and 6 report the results. Explicitly including the jump component has a direct benefit for both the SVIJ (independent jumps) and the SVCJ (contemporaneous jumps) specification. In the SVIJ case, jumps has no impact on future volatility, but there is still a benefit in removing the jump component. Indeed, in this model the persistence is conveyed only by the continuous volatility, while total quadratic variation (which is estimated by realized volatility) also contain the memoryless jumps. Thus, by separating the jumps from the persistent part in the explanatory variables, a better model specification is obtained. We conclude that, when the memory of volatility is mainly contained in the continuous part of quadratic variation, there is still a potential benefit in removing jumps even if they do not impact on future volatility. This benefit persist also for long horizon forecasts. Importantly, in this case, the jump component is found to be insignificant.

In the SVCJ the performance improvements of the models with jumps is also given by the direct link between jumps in price and volatility, which explains the overall lower RMSE (also for the HAR model, which incorporates jumps directly in quadratic variation estimated via realized volatility). When a jump occurs in price, it also occurs in volatility and it is positive. Thus, when there is a jump in price, volatility becomes higher and it stays higher because of its memory persistence. That is why jumps are (contrary to the SVIJ case) found to be significant in explaining future volatility in SVCJ models. Hence, our simulation results show that a possible mechanism explaining the significant impact of jumps on future volatility is given by contemporaneous jumps in price and volatility. Moreover, it confirms that the significance of the jumps coefficients on our empirical analysis provides an indication of the presence of a genuine forecasting power of jumps on future volatility. Hence, the similarity between the figures reporting the Newey-West corrected t-statistics of the daily jump coefficient estimated on the simulated SVCJ model (Figure 6 right panel) and on the empirical S&P500 (Figure 2), confirms the presence of a genuine forecasting power of jumps on future S&P volatility.

On the other hand, the heterogeneous leverage effect found in real data cannot be completely explained by model 4.1. Indeed, the presence of a negative coefficient ρ ≈ −0.5 (estimated on S&P 500 data) is able to explain only short-period leverage effect, by propagating negative returns into contemporaneous, and by memory persistence, future volatility. While, in the real data, we provided evidence for strong heterogeneous leverage effect, being also the weekly and monthly negative components highly significant. The model specification 4.1 is then insufficient to explain our results which demand for a more complicated continuous process with a richer specification.
5 Conclusions

This paper presents a new model for volatility forecasting which isolates three main determinants of volatility dynamics, namely heterogeneous past volatility, heterogeneous past negative returns and jumps. We find that each component plays a different role at different forecasting horizons, but all the three are highly significant and neglecting each one of them is detrimental to the forecasting performance of the model. Moreover, when other volatility measures proposed in the literature are inserted in the LHAR-CJ model they either drop out or only marginally contribute to the performance of the model confirming the ability of the LHAR-CJ model to capture the main determinants of volatility dynamics.
Explicitly modelling the jump component is important for two distinct reasons. First, it has a trimming effect on the dynamics of the persistent component of volatility which allows a better prediction of future volatility, confirming Andersen et al. (2007). Secondly, as suggested in Corsi et al. (2008), they have a direct positive and significant impact on future volatility. Moreover, there are evidences that this direct impact of jumps is of a short lived nature. If, as it seems reasonable, volatility is a measure of the uncertainty of the market about its fundamental values, our findings can be interpreted as follows: after a jump (usually a market crash) the market takes a longer period to reassess its fundamental value by dissipating the uncertainty created by the jumps; during this period residual uncertainty generate higher volatility. Our simulated experiments indicated that this mechanism can be statistically reproduced by a model having contemporaneous jumps in price and volatility.

On the other hand, while the mechanism of leverage effects on volatility dynamics is still not well understood, we find that not only daily but also weekly and monthly negative past returns are highly significant and have a remarkable forecasting power on future volatility. This novel effect seems to confirm the heterogeneous structure of the market and cannot be explained by continuous-time models, though flexible, as the ones specified so far in the literature. We also find that, at longer horizons, positive returns (price trends) have an impact on future volatility.

We conclude by noting that our model is very simple to implement, as it does not requires sofisticated computational technique. The estimation of the model parameters can be performed through a simple OLS regression, and the computation of the explanatory variables is trivial. We think that, for all the aforementioned reasons, this model may be effectively used for risk management.
Table 2: Estimated parameters, RMSE, and $R^2$ of alternative specifications of the baseline LHAR-CJ model; t-statistics are in parenthesis.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>0.586*</td>
<td>0.606*</td>
<td>0.365</td>
<td>0.617*</td>
<td>0.592*</td>
<td>0.709*</td>
<td>0.662*</td>
<td>0.669*</td>
</tr>
<tr>
<td></td>
<td>(7.735)</td>
<td>(7.044)</td>
<td>(1.113)</td>
<td>(8.120)</td>
<td>(7.829)</td>
<td>(9.142)</td>
<td>(8.281)</td>
<td>(8.406)</td>
</tr>
<tr>
<td>C</td>
<td>0.246*</td>
<td>0.265*</td>
<td>0.184*</td>
<td>0.124*</td>
<td>0.181*</td>
<td>0.083*</td>
<td>0.146*</td>
<td>0.077*</td>
</tr>
<tr>
<td>C(5)</td>
<td>0.352*</td>
<td>0.355*</td>
<td>0.352*</td>
<td>0.381*</td>
<td>0.360*</td>
<td>0.370*</td>
<td>0.356*</td>
<td>0.364*</td>
</tr>
<tr>
<td>C(22)</td>
<td>0.273*</td>
<td>0.278*</td>
<td>0.270*</td>
<td>0.283*</td>
<td>0.274*</td>
<td>0.277*</td>
<td>0.271*</td>
<td>0.271*</td>
</tr>
<tr>
<td></td>
<td>(10.257)</td>
<td>(10.364)</td>
<td>(10.397)</td>
<td>(10.398)</td>
<td>(10.312)</td>
<td>(10.196)</td>
<td>(10.183)</td>
<td>(10.221)</td>
</tr>
<tr>
<td>$p$</td>
<td>-0.008*</td>
<td>-0.008*</td>
<td>-0.008*</td>
<td>-0.008*</td>
<td>-0.007*</td>
<td>-0.006*</td>
<td>-0.007*</td>
<td>-0.006*</td>
</tr>
<tr>
<td></td>
<td>(-8.721)</td>
<td>(-8.798)</td>
<td>(-8.748)</td>
<td>(-8.007)</td>
<td>(-7.818)</td>
<td>(-6.248)</td>
<td>(-6.428)</td>
<td>(-5.519)</td>
</tr>
<tr>
<td>$p(5)$</td>
<td>-0.012*</td>
<td>-0.013*</td>
<td>-0.012*</td>
<td>-0.014*</td>
<td>-0.013*</td>
<td>-0.012*</td>
<td>-0.012*</td>
<td>-0.012*</td>
</tr>
<tr>
<td></td>
<td>(-5.349)</td>
<td>(-5.330)</td>
<td>(-5.102)</td>
<td>(-5.996)</td>
<td>(-5.504)</td>
<td>(-5.410)</td>
<td>(-5.187)</td>
<td>(-5.335)</td>
</tr>
<tr>
<td>$p(22)$</td>
<td>-0.010*</td>
<td>-0.010*</td>
<td>-0.009</td>
<td>-0.010*</td>
<td>-0.010*</td>
<td>-0.011*</td>
<td>-0.010*</td>
<td>-0.010*</td>
</tr>
<tr>
<td></td>
<td>(-2.012)</td>
<td>(-2.037)</td>
<td>(-1.955)</td>
<td>(-1.981)</td>
<td>(-2.010)</td>
<td>(-2.198)</td>
<td>(-2.157)</td>
<td>(-2.156)</td>
</tr>
<tr>
<td>J</td>
<td>0.048*</td>
<td>0.052*</td>
<td>0.039*</td>
<td>0.036*</td>
<td>0.030*</td>
<td>0.016</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.872)</td>
<td>(6.626)</td>
<td>(3.243)</td>
<td>(4.870)</td>
<td>(3.650)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RJ</td>
<td>-0.029</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.850)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAV</td>
<td>0.062</td>
<td></td>
<td>0.082*</td>
<td>0.059*</td>
<td></td>
<td>0.060*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(7.963)</td>
<td>(5.521)</td>
<td></td>
<td>(5.662)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SemiRV</td>
<td>0.142*</td>
<td>0.101*</td>
<td>0.104*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.806)</td>
<td>(4.746)</td>
<td>(4.944)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.7328</td>
<td>0.7328</td>
<td>0.7330</td>
<td>0.7326</td>
<td>0.7344</td>
<td>0.7334</td>
<td>0.7344</td>
<td>0.7361</td>
</tr>
</tbody>
</table>
### Table 3: OLS estimate for baseline LHAR-CJ model, for S&P500 futures from January 1990 to December 2007 (4344 observations). The LHAR-CJ model is estimated on 1-day, 1-week, 2-week and 1-month realized volatility. The significant jumps are computed using a critical value of $\alpha = 99.9\%$. Reported in parenthesis are $t$-statistics based on Newey-West correction.
References


Clements, M., A. Galvão, and J. Kim (2008). Quantile forecasts of daily exchange rate returns from


