Intraday LeBaron effects

Simone Bianco
William & Mary

Fulvio Corsi

Roberto Reno

Follow this and additional works at: https://scholarworks.wm.edu/aspubs

Recommended Citation
Intraday LeBaron effects

Simone Biancoa,1, Fulvio Corsib, and Roberto Renób

aDepartment of Applied Science, College of William and Mary, Williamsburg, VA 23187-8795; and bDipartimento di Economia Politica, Università di Siena, Piazza San Francesco 7, 53100 Siena, Italy

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved April 28, 2009 (received for review February 6, 2009)

We study the relation at intraday level between serial correlation and volatility of the Standard and Poor (S&P) 500 stock index futures returns. At daily and weekly levels, serial correlation and volatility forecasts have been found to be negatively correlated (LeBaron effect). After finding a significant attenuation of the original effect over time, we show that a similar but more pronounced effect holds by using intraday measures, by such as realized volatility and variance ratio. We also test the impact of unexpected volatility, defined as the part of volatility which cannot be forecasted, on the presence of intraday serial correlation in the time series by employing a model for realized volatility based on the heterogeneous market hypothesis. We find that intraday serial correlation is negatively correlated to volatility forecasts, whereas it is positively correlated to unexpected volatility.

Methodology and Results

The dataset we use is one of the most liquid financial assets in the world, that is the Standard and Poor (S&P) 500 stock index futures from 1993 to 2007, for a total of \( N = 4344 \) days. By using the futures instead of the cash index, we avoid the nonsynchronous trading bias (15). We have all high-frequency information, but to avoid microstructure effects we use a grid of \( n = 84 \) 5-minute logarithmic returns per day, interpolated according to the previous tick scheme (the price at time \( t \) is the last observed price before \( t \)). These choices are the standard ones in this kind of application.

Denote by \( R_t \) the close-to-close return at day \( t \). Let us assume to have \( r_{1,t}, \ldots, r_{n,t} \) intraday logarithmic returns. To quantify volatility, we construct daily realized variance measures defined as the cumulative sum of squared intraday 5-minute returns (16):

\[
RV_t = \sum_{i=1}^{n} r_{i,t}^2.
\]

Because volatility has been shown to be approximately log-normal (17), with power-law deviations in the tail events (9, 10), we use the logarithm of \( RV_t \) to obtain distributions which are close to normal.

The LeBaron effect (12) can be interpreted as the negative relation between volatility forecasts at time \( t \), obtained with observables up to time \( t-1 \), and the product \( R_t R_{t-1} \). We improve on the original LeBaron methodology in 2 ways. First, to obtain volatility forecasts, we borrow from recent advancements in financial econometrics, since we cannot ignore the fact that volatility is well known to display long-range dependence. One effective way to accommodate for this stylized fact without resorting to the estimation burden of a long memory model is the heterogeneous autoregressive (HAR) model of ref. 14. Following the heterogeneous market hypothesis of refs. 18–22, which recognizes the presence of heterogeneity in traders’ horizon and the asymmetric propagation of volatility cascade from long to short time periods (23) with respect to that from short to long time periods (24), the basic idea that emerges is that heterogeneous market structure generates an heterogeneous volatility cascade. Hence, ref. 14 proposed a stochastic additive cascade of 3 different realized variance components, which explains the long memory observed in the volatility as the superimposition of few processes operating at different time scales. These processes mirror the 3 typical time horizons operating in the financial market: daily, weekly, and monthly. This stochastic volatility cascade leads to a simple AR-type model in the realized variance with the feature of considering realized volatilities defined over heterogeneous time periods (the HAR model):

\[
\log RV_t = \beta_0 + \beta_{(d)} \log RV_{t-1} + \beta_{(w)} \log RV_{t-1}^{(w)} + \beta_{(m)} \log RV_{t-1}^{(m)} + \eta_t
\]

where \( \eta_t \) is a zero-mean estimation error and

\[1\] To whom correspondence should be addressed. E-mail: sbianco@wm.edu.

Author contributions: S.B., F.C., and R.R. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

www.pnas.org/cgi/doi/10.1073/pnas.0901165106
Fig. 1. The volatilities used in this study. (Upper) The time series estimate of \( \log RV_t \) used in this study (4,344 observations) is shown, together with the predictable volatility estimated by means of the HAR model. (Lower) The time series of unexpected volatility estimated as the residuals of the HAR model is shown.

\[
\log RV_{t}^{(w)} = \frac{1}{5} \sum_{k=1}^{5} \log RV_{t-k}, \log RV_{t}^{(m)} = \frac{1}{22} \sum_{k=1}^{22} \log RV_{t-k}. \quad [3]
\]

Although the HAR model does not formally belong to the class of long-memory models, it generates apparent power laws and long memory, i.e. it is able to reproduce a memory decay which is indistinguishable from that observed in the empirical data. It has been used in many applications in financial economics (25–28). We estimate the HAR model with ordinary least squares, and use the estimated coefficients \( \hat{\beta}_0, (d, w, m) \) to define the predictable volatility as:

\[
\sigma_p, t = \hat{\beta}_0 + \hat{\beta}_{(d)} \log RV_{t-1}^{(w)} + \hat{\beta}_{(w)} \log RV_{t-1}^{(m)} \quad [4]
\]

and the unexpected volatility as the residuals of the regression in Eq. 2:

\[
\sigma_u, t = \hat{\eta}_t. \quad [5]
\]

Fig. 1 shows the time series of \( \log RV_t, \sigma_p, \) and \( \sigma_u, \) in our sample. Note that the definition of unexpected volatility is model-dependent; however, the results presented here hold also with alternative prescriptions, such as simple autoregressive models for realized volatility (29, 30), with the HAR model providing the most clear-cut results.

Second, we test the LeBaron effect by measuring the dependence of serial correlation from volatility forecasts by using a Nadaraya-Watson estimator:

\[
\hat{\rho}(x) = \frac{\sum_{t=1}^{N-1} K \left( \frac{\sigma_p, t-x}{h} \right) R_t R_{t+1}}{\sum_{t=1}^{N-1} K \left( \frac{\sigma_p, t-x}{h} \right) R_t^2} \quad [6]
\]

with \( h = 3 \cdot \text{std}(\sigma_p) \cdot N^{-\frac{1}{4}} \) and \( K(y) = e^{-y^2/2} \) (31). Confidence intervals can be computed via simulation of uncorrelated replicates with the same variance. Fig. 2 shows the estimate in our sample: There is an inverse linear relation between volatility forecasts and serial correlation, which is however much weaker than that observed by LeBaron. This inverse linear relation is due to the relative smallness of our sample and to a likely increased market efficiency: serial correlation has almost disappeared, the \( AR(1) \) coefficient of \( R_t \) being just \(-0.0276\), whereas the mean value found by LeBaron in the period 1928–1990 was \( 0.0618\). Similar findings on increased market efficiency have been found in ref. 32, which use hourly returns to test the LeBaron effect.

Motivated by this finding, we investigate the presence of the LeBaron effect at intraday level (for data sampled at 5-minute frequencies) by studying the relation between realized volatility and high-frequency correlation. To measure the latter, we borrow from ref. 30 by using a modified overlapped variance ratio. Define

\[
\hat{\mu} \equiv \frac{1}{n} \sum_{k=1}^{n} r_k \quad [7]
\]

Fig. 2. Estimate of the averaged Nadaraya-Watson serial correlation \( \hat{\rho}(x) \) as in Eq. 6, as a function of HAR volatility forecasts. Confidence bands are computed by using 1,000 simulated runs of the HAR model with no serial correlation in the returns.
We define the variance ratio as follows:

\[ VR(q) = \frac{\sigma_q^2}{\sigma_0^2} \]

where

\[ \sigma_q^2 = \frac{1}{n-1} \sum_{k=1}^{n} (r_k - \mu)^2 \]

and

\[ \sigma_0^2 = \frac{1}{m} \sum_{k=m}^{n} \left( \sum_{j=k-q+1}^{k} r_j - q \mu \right)^2 \]

where

\[ m = q(n-q+1) \left( 1 - \frac{q}{n} \right) . \]

We define the variance ratio as follows:

\[ VR(q) = \left( \frac{\sigma_q^2}{\sigma_0^2} \right)^{\beta} . \]

The use of the power transformation \( f(x) = x^\beta \) makes the distribution closer to a normal one in small samples (33). The expression of Eq. 11 is, when the return process is a martingale difference with time-varying bounded variance (see ref. 33 for additional technical assumptions), asymptotically normal with mean 1 and given standard deviation, \( \beta \) is given by

\[ \beta = 1 - \frac{2}{3} \left( \frac{\sum_{j=1}^{(n-1)/2} W_k(\lambda_j)}{\sum_{j=1}^{(n-1)/2} W_k^2(\lambda_j)} \right)^2 . \]

where \( W_k \) is the Fejer kernel:

\[ W_k(\lambda) = \frac{1}{k} \sin^2(k\lambda/2) / \sin^2(\lambda/2) \]

and \( k = q - 1, \lambda_j = 2\pi j/n. \)

Intuitively, the variance ratio expresses the ratio of variances computed at 2 different frequencies whose ratio is given by \( q \). If there is no serial correlation in the data, \( VR(q) \) should be close to one. In the presence of positive serial correlation, the variance \( \sigma_q^2 \) is higher than \( \sigma_0^2 \) and \( VR(q) > 1 \). If instead there is negative serial correlation, this argument reverses and \( VR(q) < 1 \). We use \( q = 2, 3, 4, 5, 6 \). Higher values of \( q \) cannot be used without introducing possible distortions in the statistics behavior (34). The \( VR \) measure has been shown to be correct also for heteroskedastic data-generating processes (5), and it is defined with overlapping observations (34). This measure is a reliable measure of serial correlation both at daily (5) and intraday (29, 30) level.

We start by first studying the relation between intraday serial correlation and contemporaneous realized volatility by using the simple linear regression

\[ VR(q) = b_1 + c \log \text{RV}_t + \epsilon_t \]

and then inserting lagged volatility as well:

\[ VR(q)_t = b_1 + c_0 \log \text{RV}_t + c_1 \log \text{RV}_{t-1} + \epsilon_t . \]

As in ref. 30, when we use the variance ratio \( VR(q) \), as dependent variable, we also add as explanatory variable 5 lags of \( VR(q) \) to remove the autocorrelation of the residuals, that is:

\[ b_t = b + \sum_{j=1}^{5} a_j \text{VR(q)}_{t-j} . \]

Lagged volatility is, however, a very poor volatility forecast. Thus, we resort again to the HAR model by estimating the regression

\[ VR(q)_t = b_1 + c_0 \sigma_{p,2} + c_0 \sigma_{u,2} + \epsilon_t , \]

which fully takes into account heterogeneity, long memory, and heteroskedasticity of financial market volatility. We finally estimate the extension

\[ VR(q)_t = b_1 + c_0 \sigma_{p,2} + c_0 \sigma_{u,2} + \epsilon_t , \]

in which unexpected volatility is inserted as an additional explanatory variable. Note that in the regression in Eq. 17, we add a contemporaneous variable \( \sigma_{u,t} \), which cannot be used for prediction.

Estimation results with \( q = 2, 3, 4, 5, 6 \) are in Table 1, and Fig. 3 shows the estimated coefficients of regressions 14, 15, and 17 with \( q = 2, 6 \) on a rolling window with a length of 5 years.

**Discussion**

Estimates of the model in Eq. 14 depicted in the first row of Fig. 3 may look disappointing, showing no significant correlation between variance ratio and contemporaneous realized variance. Moreover, this correlation tends to be slightly positive (even if not significantly) instead of negative, especially in the first part of the sample.

However, the second row of Fig. 3 shows that the coefficient of lagged volatility on variance ratio is negative and significant across the entire sample. The same can be seen more clearly from the estimate of regressions 16 and 17, reported in Table 1 and the third row of Fig. 3.

Most interestingly, we find that contemporaneous volatility is significantly and positively correlated with the variance ratio. Hence, estimation results for Eq. 15 indicate a sharp difference in the relation between intraday serial correlation and volatility: strongly positive for contemporaneous volatility and strongly negative for lagged one. Such antithetical behavior of the relation is even more puzzling considering the well-known stylized fact of volatility to be highly persistent. How could we explain this result? By our heterogeneous “rotation” of the regressors,
Fig. 3. On a rolling window with a length of 5 years is shown the measures of coefficient $c$ of regression 14 (Top), the measures of coefficients $c_0$ and $c_1$ of regression 15 (Middle), and the measures of coefficients $c_{u}$ and $c_{p}$ of regression 17 (Bottom). In the left column, the variance ratios have been computed with $q = 2$ (5 minutes); on the right column, with $q = 6$ (25 minutes). Dashed lines represent confidence intervals at 95% confidence level. This figure shows that the correlation between serial correlation and predictable volatility is negative (LeBaron effect), whereas correlation between serial correlation and unexpected volatility is positive. Smaller standard errors which respect those reported in Table 1 are due to the fact that the estimates reported in this figure are made on subsamples.

we can rewrite Eq. 15 in the form of Eq. 17. This provides the separation between predictable and unexpected volatility illustrated in Fig. 1. The new specification greatly helps in shedding light on this result, providing a precise economic interpretation. Hence, as ref. 30 suggested, we can now provide an explanation in term of predictable and unexpected volatility: Because volatility is known to be predictable by market participants, it has a different impact with respect to its unpredictable component.

The third row of Fig. 3 shows indeed that the predictable volatility, now defined by means of the HAR model, is negatively correlated with the variance ratio (more with higher $q$) and that the unpredictable volatility is positively correlated with the variance ratio (more with higher $q$).

The full sample estimates in Table 1 corroborate this finding. We can then rephrase our results as follows: Intraday serial correlation is negatively correlated with the expected volatility. Moreover, we can conclude that serial correlation is instead positively correlated with unexpected volatility, which is a previously unrecognized empirical feature of financial returns. Our finding suggests that the usual explanation of the LeBaron effect in terms of feedback trading (35) is at least incomplete, advocating for a broader theory on the link between volatility and the way information is spread to heterogeneous market components. It is particularly interesting that a market anomaly like serial correlation is associated with higher unexpected volatility, typically due to unexpected news.

ACKNOWLEDGMENTS. We thank the Biomath Program at the College of William and Mary for financial support.