Intraday LeBaron effects

Simone Bianco[a,1], Fulvio Corsi[b], and Roberto Renò[b]

[a]Department of Applied Science, College of William and Mary, Williamsburg, VA 23187-8795; and [b]Dipartimento di Economia Politica, Università di Siena, Piazza San Francesco 7, 53100 Siena, Italy

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.

financial markets | volatility | serial correlation | variance ratio

Serial correlation of asset prices is one of the most elusive quantities of financial economics. According to the theory of efficient markets (1, 2), it should not exist at all, and, when it exists, it represents an anomaly of financial markets. Many economists and physicists devoted themselves to the study of stock return predictability (3, 4). Historical returns should prevent any forecasting technique, even if it has been shown, as in ref. 5, that the random walk hypothesis holds only weakly.

On the other hand, the variance of financial returns on a fixed time interval, which is more usually called volatility, is a highly predictable quantity (6, 7), with its probability distribution function showing fat tails (8–10). The natural association of volatility to financial risk forecast and control makes its analysis paramount in economics. To some extent, it seems obvious, therefore, to link volatility (or trading volume, as in ref. 11) to returns serial correlation. If anything else, the link between volatility and serial correlations can reveal basic properties of the price-formation mechanism.

A notable stylized fact on serial correlation is the LeBaron effect (12), according to which volatility forecasts are negatively correlated to serial correlation. In this work, we find milder evidence of such effect in the dataset we analyze which, being more recent than that used by LeBaron, suggests that market efficiency has increased. Most importantly, we improve on the existing literature by using measures of both volatility and serial correlation which are based on 5-minute returns. The forecasting model we use is directly based on realized volatility measures, and it is set in the framework of what can be termed the heterogeneous market hypothesis (13). In this model, volatility is consistently composed by a cascade of several time components (14). The model is particularly successful in recovering the volatility dynamics and mimicking the long-range dependence and fat tails which are observed in the realized volatility time series. To quantify serial correlation, we use instead a modification of the variance ratio statistics with overlapping observations.

Usage of suitable intraday measures allows us to test an intraday version of the LeBaron effect with a very large and liquid dataset. We provide evidence of a negative relation between volatility forecasts and intraday serial correlation. Moreover, we also refine this finding by showing that volatility can be split into 2 components: a predictable one and an unpredictable one, with the latter being positively correlated with serial correlation.

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):

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

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):

$$\log RV_t = \beta_0 + \beta_{(d)} \log RV_{t-1}$$
$$+ \beta_{(w)} \log RV_{t-1}^{(w)} + \beta_{(m)} \log RV_{t-1}^{(m)} + \eta_t \qquad [2]$$

where η_t is a zero-mean estimation error and

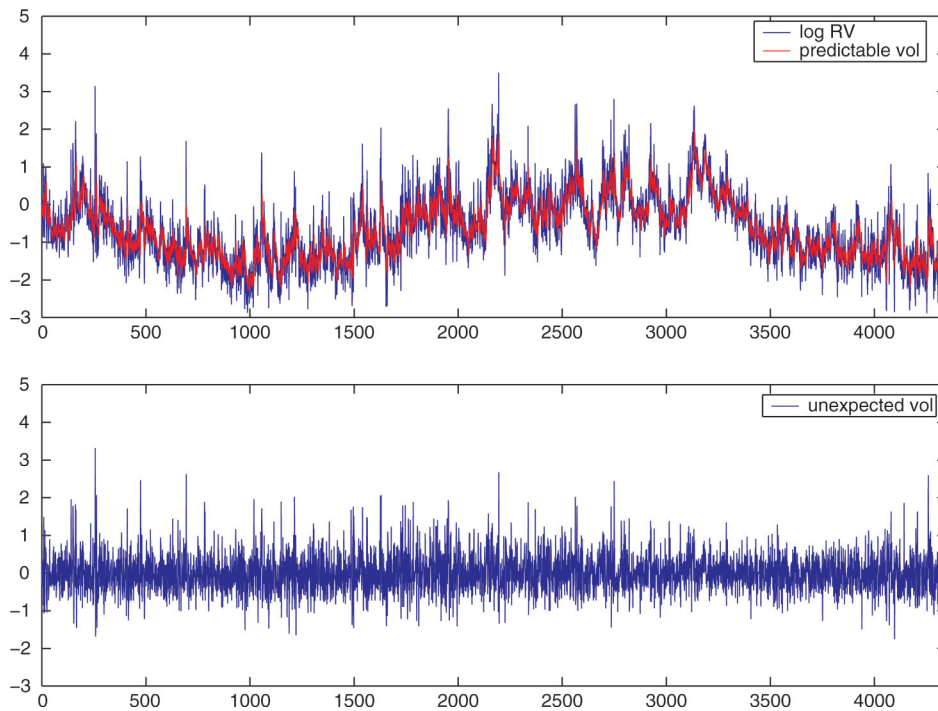ECONOMIC SCIENCES

APPLIED PHYSICAL SCIENCES

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.

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

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:

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

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 $\log RV_t$, $\sigma_{p,t}$ and $\sigma_{u,t}$ 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 $h = 3 \cdot std(\sigma_{p,t}) \cdot N^{-\frac{1}{5}}$ and $K(y) = e^{-y^2/2}$ (31). Confidence intervals can be computed via simulation of uncorrelated replicas 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

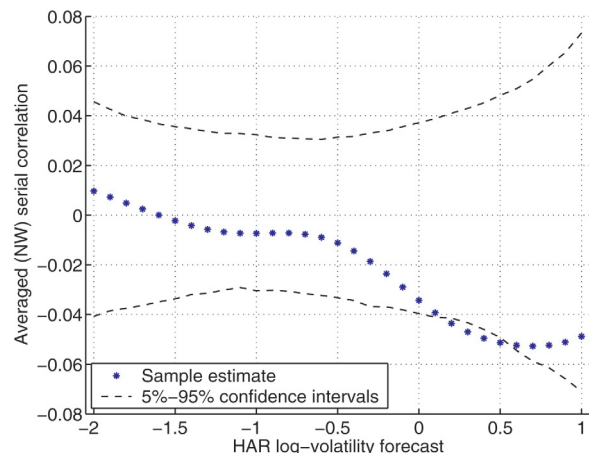$$\hat{\mu} \equiv \frac{1}{n} \sum_{k=1}^{n} r_k \quad [7]$$

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.

Table 1. On the full sample, estimated regression

	Regression **14**		Regression **15**			Regression **16**		Regression **17**		
q	c	$\overline{R}^2(\%)$	c_0	c_1	$\overline{R}^2(\%)$	c_p	$\overline{R}^2(\%)$	c_u	c_p	$\overline{R}^2(\%)$
2	−0.242	1.40	9.026	−11.851	3.82	−6.307	2.59	12.707	−6.376	4.85
	(0.805)		(1.506)	(1.461)		(0.988)		(1.702)	(0.958)	
3	−0.949	0.44	10.742	−14.950	3.59	−7.700	1.86	13.478	−7.780	3.93
	(0.933)		(1.754)	(1.656)		(1.174)		(1.998)	(1.156)	
4	−0.160	−0.01	12.838	−16.608	3.67	−7.016	1.14	14.645	−7.080	3.45
	(0.978)		(1.823)	(1.697)		(1.224)		(2.028)	(1.210)	
5	0.325	0.05	14.499	−18.108	3.95	−6.850	1.03	15.860	−6.903	3.45
	(1.051)		(1.969)	(1.777)		(1.218)		(2.154)	(1.205)	
6	0.810	0.23	16.477	−20.018	4.42	−6.941	1.10	17.603	−6.986	3.73
	(1.108)		(2.118)	(1.902)		(1.220)		(2.280)	(1.211)	

Standard errors are in brackets. The adjusted R^2 is indicated by $\overline{R}^2$ and is expressed in percentage form. All the coefficients are multiplied by 10^3.

$$\hat{\sigma}_a^2 \equiv \frac{1}{n-1}\sum_{k=1}^n (r_k - \hat{\mu})^2 \tag{8}$$

$$\hat{\sigma}_c^2(q) \equiv \frac{1}{m}\sum_{k=q}^n \left(\sum_{j=k-q+1}^k r_j - q\hat{\mu}\right)^2 \tag{9}$$

where

$$m = q(n-q+1)\left(1-\frac{q}{n}\right). \tag{10}$$

We define the variance ratio as follows:

$$VR(q) = \left(\frac{\hat{\sigma}_c^2(q)}{\hat{\sigma}_a^2}\right)^\beta. \tag{11}$$

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. β is given by

$$\beta = 1 - \frac{2}{3}\frac{\left(\sum_{j=1}^{(n-1)/2} W_k(\lambda_j)\right)\left(\sum_{j=1}^{(n-1)/2} W_k^3(\lambda_j)\right)}{\left(\sum_{j=1}^{(n-1)/2} W_k^2(\lambda_j)\right)^2}, \tag{12}$$

where W_k is the Fejer kernel:

$$W_k(\lambda) = \frac{1}{k}\frac{\sin^2(k\lambda/2)}{\sin^2(\lambda/2)} \tag{13}$$

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 $\hat{\sigma}_c^2(q)$ is higher than $\hat{\sigma}_a^2$, and $VR(q) > 1$. If instead there is negative serial correlation, this argument reverts 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

$$VR(q)_t = b_t + c\log RV_t + \varepsilon_t \tag{14}$$

and then inserting lagged volatility as well:

$$VR(q)_t = b_t + c_0\log RV_t + c_1\log RV_{t-1} + \varepsilon_t. \tag{15}$$

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

$$b_t = b + \sum_{j=1}^5 \alpha_j VR(q)_{t-j}.$$

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

$$VR(q)_t = b_t + c_p\sigma_{p,t} + \varepsilon_t, \tag{16}$$

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

$$VR(q)_t = b_t + c_p\sigma_{p,t} + c_u\sigma_{u,t} + \varepsilon_t, \tag{17}$$

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,
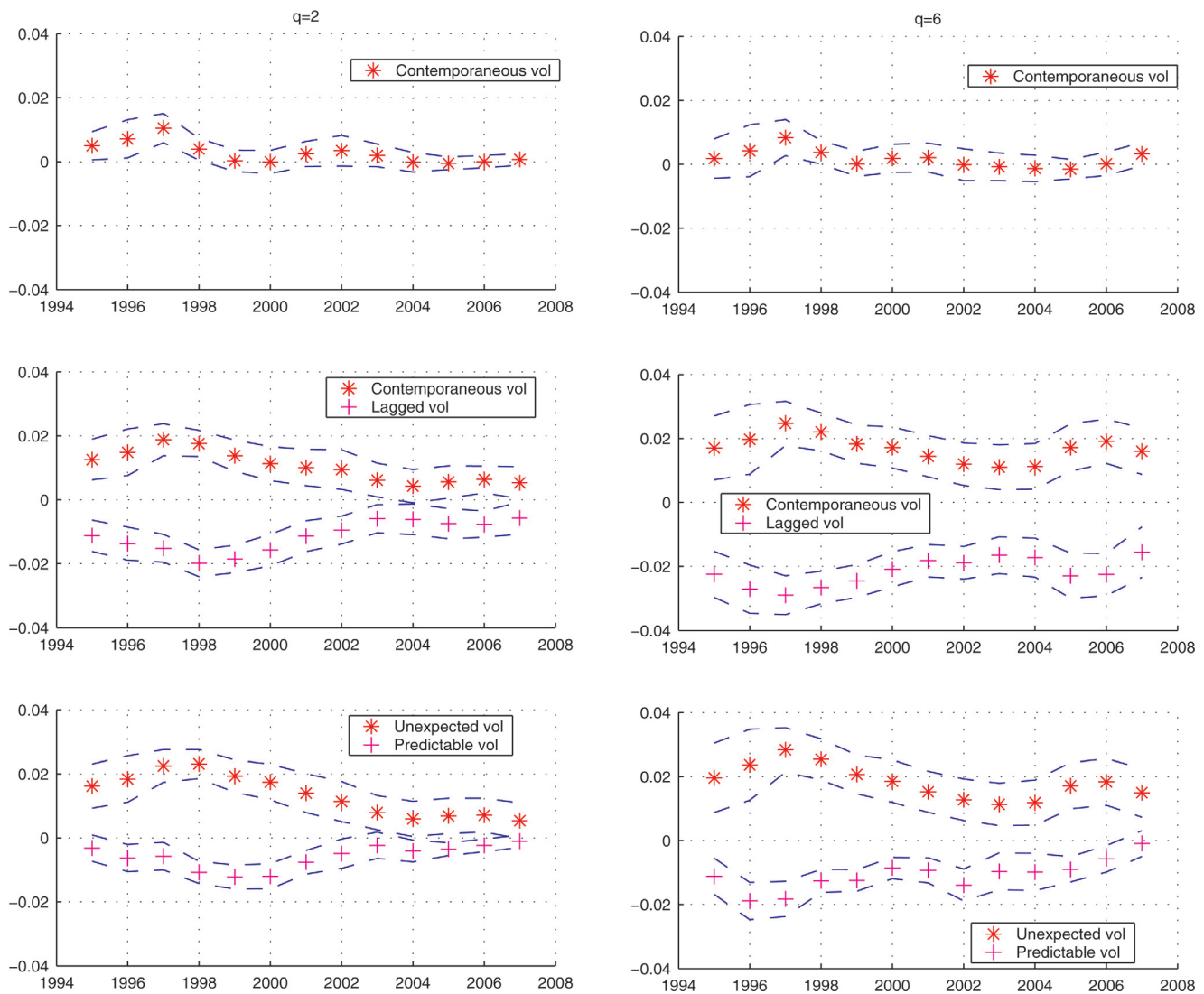
Bianco et al.

www.manaraa.com

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.

1. Fama E (1970) Efficient capital markets: A review of theory and empirical work. *J Finance* 25:383–417.
2. Farmer JD, Lo AW (1999) Frontiers of finance: Evolution and efficient markets. *Proc Natl Acad Sci USA* 96:9991–9992.
3. Farmer JD, Patelli P, Zovko II (2005) The predictive power of zero intelligence in financial markets. *Proc Natl Acad Sci USA* 102:2254–2259.
4. Pincus S, Kalman RE (2004) Irregularity, volatility, risk, and financial market time series. *Proc Natl Acad Sci USA* 101:13709–13714.
5. Lo AW, MacKinlay AC (1988) Stock market prices do not follow random walks: Evidence from a simple specification test. *Rev Financ Stud* 1:41–66.
6. Engle R (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50:987–1008.

Bianco et al.

7. Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econometrics* 31:307–327.
8. Liu Y, Cizeau P, Meyer M, Peng CK, Stanley HE (1997) Correlations in economic time series. *Physica A* 245:437–440.
9. Liu Y, et al. (1997) Statistical properties of the volatility of price fluctuations. *Phys Rev E* 60:1390–1400.
10. Plerou V, Gopikrishnan P, Nunes Amaral LA, Gabaix X, Stanley HE (1997) Economic fluctuations and anomalous diffusions. *Phys Rev E* 62:R3023–R3026.
11. Campbell J, Grossman SJ, Wang J (1993) Trading volume and serial correlation in stock returns. *Q J Econ* 108:905–940.
12. LeBaron B (1992) Some relations between volatility and serial correlations in stock market returns. *J Business* 65:199–219.
13. Muller U, et al. (1993) Fractals and intrinsic time: A challenge to econometricians. *XXXIXth International AEA Conference on Real Time Econometrics, 14-15 Oct 1994* (Luxembourg).
14. Corsi F (2009) A simple approximate long memory model of realized volatility. *J Financ Econometrics* 7:174–196.
15. Atchinson M, Butler K, Simonds R (1987) Nonsynchronous security trading and market index autocorrelation. *J Finance* 42:111–118.
16. Andersen T, Bollerslev T, Diebold FX, Labys P (2003) Modeling and forecasting realized volatility. *Econometrica* 71:579–625.
17. Cizeau P, Liu Y, Meyer M, Peng CK, Stanley HE (1997) Volatility distribution in the S&P500 stock index. *Physica A* 245:441–445.
18. Dacorogna MM, Gençay R, Müller UA, Olsen RB, Pictet OV (2001) *An Introduction to High-Frequency Finance*. (Academic, New York).
19. Lux T, Marchesi M (1999) Scaling and criticality in a stochastic multi-agent model of financial market. *Nature* 397:498–500.
20. LeBaron B (2002) Short-memory traders and their impact on group learning in financial markets. *Proc Natl Acad Sci USA* 99:7201–7206.
21. Weron A, Weron R (2000) Fractal market hypothesis and two power laws. *Chaos, Solitons and Fractals* 11:289–296.
22. McMillan DG, Speight AEH (2006) Volatility dynamics and heterogenous markets. *Int J Finance Econ* 11:115–121.
23. Ghashghaie S, Breymann W, Peinke J, Talkner P, Dodge Y (1996) Turbolent cascades in foreign exchange markets. *Nature* 381:767–770.
24. Lynch P, Zumbach G (2003) Market heterogeneities and the causal structure of volatility. *Quant Finance* 3:320–331.
25. Ait-Sahalia Y, Mancini L (2008) Out of sample forecasts of quadratic variation. *J Econometrics* 147:17–33
26. Andersen T, Bollerslev T, Diebold FX (2007) Roughing it up: Including jump components in the measurement modeling and forecasting of return volatility. *Rev Econ Stat* 89:701–720.
27. Clements MP, Galvão AB, Kim JH (2008) Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *J Empir Finance* 15:729–750.
28. Ghysels E, Santa-Clara P, Valkanov RS (2006) Predicting volatility: Getting the most out of return data sampled at different frequencies. *J Econometrics* 131:59–95.
29. Bianco S, Renò R (2006) Dynamics of intraday serial correlation in the Italian futures market. *J Futures Markets* 26:61–84.
30. Bianco S, Renò R (2008) Unexpected volatility and intraday serial correlation. *Quant Finance*, in press.
31. Fan J, Yao Q (2003) *Nonlinear time series*. (Springer, New York).
32. Chen CR, Su Y, Huang Y (2008) Hourly index return autocorrelation and conditional volatility in an EAR-GJR-GARCH model with generalized error dsitribution. *J Empir Finance* 15:789–798.
33. Chen W, Deo RS (2006) The variance ratio statistic at large horizon. *Econometric Theory* 22:206–234.
34. Deo RS, Richardson M (2003) On the asymptotic power of the variance ratio test. *Econometric Theory* 19:231–239.
35. Sentana E, Wadhwani S (1992) Feedback traders and stock return autocorrelation: Evidence from a century of daily data. *Econ J* 102:415–425.