

# Volatility

*Torben G. Andersen and Tim Bollerslev*

---

## Introduction

Volatility ranks among the most active and successful areas of research in econometrics and empirical asset pricing finance over the past three decades. This two-volume collection of papers comprises some of the most influential published works from the burgeoning volatility literature. As is the case with any such collection, our choice of papers to include is invariably somewhat subjective and colored by our own views. Nonetheless, it is our hope that this particular selection of previously published works offers a useful roadmap to the literature, starting with the first “volatility papers” more than three decades ago up to recent research, together with important milestones and contributions that helped pave the way for where we are today. We have organized the collection into seven separate parts, but many of the papers naturally span more than one of these parts.

In everyday language, the term volatility refers to the degree to which something fluctuates, or is volatile.<sup>1</sup> Within economics and finance, however, the term is generally used more formally to describe the variability of a single random variable, as quantified by its variance or standard deviation, or the co-variability among two or more random variables, as quantified by their covariance.<sup>2</sup> These volatilities are traditionally defined in terms of the variation over non-trivial discrete time intervals, like a day, month or year, but they may also be meaningfully defined over infinitesimal time intervals, corresponding to the instantaneous diffusion coefficients used in more abstract continuous-time representations. Regardless of the exact definition, the recognition that financial asset return volatilities change substantially through time dates back at least to the 1960s.

It is difficult to exactly pinpoint a single historical study that first highlights the importance of volatility clustering in financial markets. Mandelbrot (1963) and the quote from that paper about cotton prices stating that “large changes tend to be followed by large changes – of either sign – and small changes tend to be followed by small changes,” is often cited in support of time-varying volatility.<sup>3</sup> However, that quote appears in the very last section of the paper, which is otherwise devoted to characterizing the unconditional distribution of speculative prices by means of the Stable Paretian distribution. That same quote also appears in Fama (1965), which is another frequently cited study in support of volatility clustering. Still, Fama goes on to effectively dismiss the phenomenon as not being especially important empirically; quoting from the last section before the conclusion: “though there does seem to be more bunching of large values than would be predicted by a purely independent model, the tendency is not very strong.” Officer (1973) also explicitly points out that rolling sample variances for the returns on an aggregate market portfolio appears to change over time and informally relates these fluctuations to various business cycle indicators. Meanwhile, the paper makes no attempt to actually model, let alone forecast, the dynamic variation in the volatilities.

The short note by Black (1976, Volume I Chapter 1) does report the results from a set of simple regressions, both predictive and otherwise, involving logarithmic realized volatilities constructed from the summation of daily squared stock returns. Even though this paper is most often cited as the first to document the so-called leverage effect, whereby changes in financial leverage due to stock price changes leads to a direct negative link between the return on equity and volatility changes, the paper also reports the results from autoregressive type regressions, noting that the estimated coefficients suggest that “When stocks get more volatile, they tend to stay at a higher level of volatility than they had in the past.” We include this noteworthy, and not easily accessible, early paper as a prologue to the general collection of volatility papers. Rereading Black’s short note more than four decades after its publication, it is remarkable how his informal empirical-based observations foreshadowed many of the key developments that transpired in the volatility literature over the coming four decades, including: (1) volatility clusters in time, but eventually mean-reverts; (2) the negative return volatility relationship seems too large empirically to be solely explained by a “leverage effect,” and may instead at least in part be attributed to a risk-based volatility feedback channel; (3) commonalities in volatilities, as evidenced by the tendency for the prices of different stocks to “jump” together; (4) the use of finer sampled returns allow for more accurate *ex post* volatility measurements; (5) the Black–Scholes–Merton option pricing formula,<sup>4</sup> which assumes that volatility is constant, is wrong and in need of modification to allow for time-varying volatility. At the same time, however, Black is rather dismissive about the use of more formal volatility models and forecasting procedures, as underscored by his closing two sentences: “I don’t dare write down any sort of formal model of the process by which volatilities change. I’m not sure I ever will.”

Thus, even though the existence of volatility clustering had been noted by various authors, the fact that changes in volatility are highly predictable and that this predictability carries with it important economic implications, arguably was not fully recognized until the advent of the AutoRegressive Conditional Heteroskedastic (ARCH) and Generalized ARCH (GARCH) class of models,<sup>5</sup> and the first applications of these models in the late 1980s and early 1990s. Part II of this collection entitled GARCH Models includes some of the key papers that laid the foundation for the estimation and application of these first empirically realistic and statistically sound volatility models, along with several important applications and extensions of the basic model structures.

The papers on Stochastic Volatility Models included in Part III afford an alternative, and sometimes more explicitly economically founded, approach, in which volatility is treated as a separate latent stochastic process. This latency, however, comes at the cost of more complicated inference procedures than for GARCH models, which are specified in terms of conditional expectations of directly observable variables.

All of the papers included in the first two sections relate to univariate volatility models, and forecasts of variances or standard deviations. In finance, of course, the risk of an asset is typically quantified in terms of the covariance with some benchmark portfolio or sets of portfolios, as opposed to the variance of the asset itself. Part IV in Volume I on Multivariate Volatility Models includes key papers concerned with the formulation and estimation of empirically realistic models for time-varying covariances and correlations.

The pricing of financial options contracts are intimately related to volatility. In the celebrated Black–Scholes–Merton option pricing formula, the only unknown parameter is the volatility of the underlying asset, which in that formula is assumed to be constant over the life of the

options contract. As already noted by Black (1976) this assumption is untenable empirically. Part I in Volume II titled Options and Volatility, includes a series of papers that explore the implication of time-varying volatility from various option pricing perspectives.

In a related vein, the papers included in Part II Volume II relate directly to Volatility Forecasting and Evaluation, and a long-held common misconception in the finance literature that standard volatility models provide poor out-of-sample forecasts. Some of the procedures discussed in these papers rely explicitly on the use of high-frequency intraday data for more accurately measuring the volatility. The advent of reliable high-frequency data for a host of different financial assets has also spurred somewhat of a paradigm shift in the volatility literature more generally, with many of the new procedures relying on so-called realized volatility measures constructed from intraday data. The final Part III in Volume II titled High-Frequency Data and Realized Volatilities comprises some of the key papers that led to this paradigm shift.

As already noted, many of the seminal papers obviously span more than one of these seven different parts, so that our categorization of the papers is invariably somewhat arbitrary. As an organizing principle, we have tried to highlight the most important developments in close to chronological fashion within each category.

## GARCH Models

The volatility literature first really started with the publication of Engle's (1982, Volume I Chapter 2) Nobel Prize winning ARCH paper, which establishes the basic framework for modeling and forecasting volatility as a time-varying function of directly observable current information. The GARCH class of models, of which the GARCH(1,1) remains the workhorse, was subsequently introduced by Bollerslev (1986, Volume I Chapter 3).

To set out the key idea of the ARCH paper, consider the decomposition of the time series  $y_t$  into its conditional mean plus an innovation,

$$y_t = \mu_{t|t-1} + \sigma_{t|t-1} z_t \equiv \mu_{t|t-1} + \varepsilon_t \quad (1)$$

where  $z_t$  denotes an *i.i.d.* zero-mean, unit-variance innovation process.<sup>6</sup> Expressing not just the conditional mean  $\mu_{t|t-1}$ , but also  $\sigma_{t|t-1}^2$ , as an explicit function of information available at time  $t-1$ , immediately imbues  $\sigma_{t|t-1}^2$  with the interpretation as the conditional variance of  $y_t$ . Moreover, postulating a specific distribution for the  $z_t$  innovation sequence, the likelihood function is readily available in closed form from the product of recursive conditional distributions.

In particular, assuming that  $z_t$  is *i.i.d.* normally distributed, the log-likelihood function for the  $y_T, y_{T-1}, \dots, y_1$  sample may be expressed as,

$$\log L(\theta; y_T, \dots, y_1) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \left[ \log \sigma_{t|t-1}^2 - \sigma_{t|t-1}^{-2} (y_t - \mu_{t|t-1})^2 \right]. \quad (2)$$

This likelihood function obviously depends upon the parameters entering  $\mu_{t|t-1}$  and  $\sigma_{t|t-1}^2$  in a highly non-linear fashion. However, procedures for estimating all commonly used GARCH models are now available in the statistical software packages routinely used in economics.

Maybe the most important advance of the Engle (1982) and Bollerslev (1986) papers stems from the specification of empirically realistic and tractable functional forms for the way in which conditional variances evolve over time. In the original ARCH paper,  $\sigma_{it-1}^2$  is parameterized as a linear distributed lag of past squared innovations, akin to a rolling sample variance, or an AR type model for the conditional variance.<sup>7</sup> With this specification, long lags are typically called for empirically. The GARCH model “solves” this problem by including lagged values of the conditional variance itself as well. To illustrate, consider the popular GARCH(1,1) model,

$$\sigma_{it-1}^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1|t-2}^2, \quad (3)$$

where  $\omega > 0$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$ , to ensure that the conditional variance is always positive, and  $\alpha + \beta < 1$  to ensure covariance stationarity of the model and existence of the unconditional variance  $\omega/(1-\alpha-\beta)$ . Rearranging the equation, the model is readily interpreted as an ARMA(1,1) specification for the squared innovations. This simple recursive representation typically provides a good fit to the dynamic dependencies observed in speculative rates of return and other heteroskedastic economic time series.

The empirical applications in the original ARCH and GARCH papers relate to the uncertainty of aggregate inflation. Of course, as discussed above, the vast majority of the subsequent applications, and the primary reason for the great success of the GARCH class of models, stem from their use in finance. The Engle, Lilien and Robins (1987, Volume I Chapter 4) paper provides one of the first such finance applications. Motivated by the idea of a risk–return tradeoff, the paper estimates a series of ARCH models for the excess holding yield on US Treasuries, in which the expected return depends explicitly on the conditional variance, or functions thereof. The resulting ARCH-in-Mean, or ARCH-M, model has subsequently been used in numerous studies.

Another early “true” finance application is French, Schwert and Stambaugh (1987, Volume I Chapter 5), who explore the joint dynamics of aggregate stock market returns and volatility. While the paper finds it difficult to accurately estimate the GARCH-M coefficient for the conditional variance or standard deviation in the mean equation for the expected return, the estimates clearly suggests that unexpected returns are negatively related to unexpected changes in volatility. This latter finding, of course, is consistent with the “leverage effect” first noted by Black (1976), and indirectly supports the idea of a risk-based volatility feedback effect. A large body of literature now exists devoted to the estimation of this effusive risk–return tradeoff.

Schwert (1989, Volume I Chapter 6) further explores the economic forces behind the volatility clustering, including the leverage effect. Although this paper primarily relies on simple rolling sample variances and regression-based procedures, we include it in Part II together with the papers on GARCH Models, as it sets the stage for many subsequent GARCH-based studies seeking to enhance the explanatory power of the basic GARCH(1,1) model by including various macroeconomic variables and trading activity measures in the conditional variance equation.<sup>8</sup> Aside from the fact that aggregate stock market volatility tends to be counter-cyclical and volatility and trading volume are contemporaneously positively related, the general consensus emerging from these studies tend to echo Schwert (1989): “The puzzle highlighted by the results in this paper is that stock volatility is not more closely related to other measures of economic volatility.”

The original formulation of the ARCH and GARCH models and the likelihood function specified above were based on the assumption that the innovations are *conditionally* normally distributed. This assumption of conditional normality coupled with time-varying variances still generates model-implied *unconditional* distributions that have fatter tails than the normal.<sup>9</sup> As such, it is possible that GARCH models could simultaneously account for volatility clustering and the fact that unconditional return distributions tend to have heavy tails. However, as first demonstrated by Bollerslev (1987, Volume I Chapter 7), even though the standardized residuals from GARCH models generally are closer to being normally distributed than the raw returns, the volatility clustering within the models are not sufficient to fully account for the observed leptokurtosis. To remedy this, Bollerslev suggested the use of a Student  $t$ -distribution in place of the normal, with the degree of freedom of the  $t$ -distribution estimated jointly with the other model parameters based on maximizing the suitably modified likelihood function. Many alternative error distributions have been proposed subsequently, including the Generalized Error Distribution (GED) advocated by Nelson (1991, Volume I Chapter 12).

Meanwhile, as discussed by Bollerslev and Wooldridge (1992, Volume I Chapter 8), even if the assumption of conditional normality is violated empirically, as long as the conditional mean  $\mu_{t|t-1}$  and the conditional variance  $\sigma_{t|t-1}^2$  are both correctly specified, the parameter estimates for the mean and variance equations obtained under the (faulty) assumption of conditional normality are still consistent. However, the usual asymptotic covariance matrix for the Maximum Likelihood Estimates (MLE) of the parameters based on the inverse of Fisher's Information Matrix, must be modified to allow for valid Quasi-MLE (QMLE) inference. The resulting "robust," also known as Bollerslev–Wooldridge, standard errors are now included in many standard software packages as part of the GARCH toolbox. Of course, the QMLE-based estimates are not as efficient as the estimates obtained from a correctly specified conditional error distribution, such as the Student- $t$  or GED.

Still, the main advantage from the use of alternative conditionally non-normal error distributions lies not so much in the enhanced efficiency of parameter estimation, but rather in the ability to more accurately approximate the tails in the corresponding predictive distributions. This is especially true for financial risk measurement and management, where interest often centers on outliers. To illustrate, consider the popular Value-at-Risk (VaR) measure, corresponding to a specific quantile in the predictive return distribution. In the GARCH setup, the VaR is readily calculated as,

$$VaR_{t+1|t}^p = \mu_{t+1|t} + \sigma_{t+1|t} Z_p, \quad (4)$$

where  $Z_p$  refers to the desired  $p$ th quantile in the error distribution for the  $z_t$  innovations. This simple-to-implement procedure generally works well for estimating the VaR's within the main range of support of the distribution, say  $0.01 < p < 0.99$ . However, it often does not work so well for quantiles beyond that range. Instead, McNeil and Frey (2000, Volume I Chapter 9) show how the GARCH framework may be combined with Extreme Value Theory (EVT) for meaningfully estimating the VaR's in the extreme left or right tails of the distribution. Related EVT-based procedures have also been used more recently in the analysis of high-frequency intraday data as a way to sensibly extrapolate the behavior of the data to coarser, and economically more interesting, frequencies, as exemplified by Bollerslev and Todorov (2011a).

The basic GARCH model implies that positive and negative shocks of the same absolute magnitude have the identical impact on the conditional variance. As previously noted, changes in stock market volatility and returns appear to be negatively related, whether due to a direct “leverage effect,” an indirect volatility feedback effect as embodied with the GARCH-M model, or some other channel. The GJR-GARCH and Threshold GARCH (TGARCH) models, proposed independently by Glosten, Jagannathan and Runkle (1993, Volume I Chapter 10) and Zakoian (1994, Volume I Chapter 11), provide a particularly simple way of directly modeling this asymmetry by augmenting the standard GARCH model with an additional ARCH term conditioned on the sign of the past innovation,

$$\sigma_{i|t-1}^2 = \omega + \alpha \varepsilon_{i|t-1}^2 + \gamma \varepsilon_{i|t-1}^2 I(\varepsilon_{i-1} < 0) + \beta \sigma_{i|t-1|t-2}^2, \quad (5)$$

where  $I(\cdot)$  denotes the indicator function. Estimation of this model with aggregate stock market returns invariably generates  $\gamma > 0$ , implying that, indeed, past positive return shocks have a smaller effect on the future conditional variance than equally-sized past negative return shocks.

Many alternative asymmetric GARCH formulations have been explored in the literature. The Exponential GARCH (EGARCH) model of Nelson (1991, Volume I Chapter 12), which actually predates the GJR-GARCH and TGARCH models, is a well-known example. The EGARCH model is formulated in terms of the logarithm of the conditional variance,

$$\log(\sigma_{i|t-1}^2) = \omega + \alpha(|z_{t-1}| - E(|z_{t-1}|)) + \gamma z_{t-1} + \beta \log(\sigma_{i|t-1|t-2}^2), \quad (6)$$

where again  $z_t = \sigma_{i|t-1}^{-1} \varepsilon_t$ , so that  $\gamma \neq 0$  allows positive and negative innovations to impact the variance differently. By formulating the model in terms of the logarithm of the variance, the EGARCH model does not have to constrain the parameters to be positive. However, the transformation of forecasts for the logarithmic variance, as implied by the model, to forecasts for the variance, as is typically the object of interest, hinge importantly on additional distributional assumption. Because of this, coupled with the difficulties in estimating the model due to non-differentiability, the EGARCH model has fallen somewhat out of favor.

Estimates for GARCH type models with daily returns generally suggest highly persistent volatility dynamics. In particular, for the GARCH(1,1) model, the estimate of  $\alpha + \beta$  is typically very close to unity, akin to a unit root process for the conditional variance. This led Engle and Bollerslev (1986) to propose the Integrated GARCH (IGARCH) class of models, of which the GARCH(1,1) model with  $\alpha + \beta \equiv 1$  is a special case. Although this model has received a great deal of attention, the notion that shocks to the volatility are infinitely persistent has been largely discredited by now.<sup>10</sup> Instead, numerous studies, such as Ding, Granger and Engle (1993, Volume I Chapter 13), Andersen and Bollerslev (1997, Volume II Chapter 12), Andersen, Bollerslev, Diebold, and Ebens (2001), and Andersen, Bollerslev, Diebold and Labys (2001, Volume II Chapter 17), have argued that the way in which volatility shocks dissipate over long horizons is best described by a slow hyperbolic rate of decay. The Fractionally Integrated GARCH (FIGARCH) model first proposed by Baillie, Bollerslev and Mikkelsen (1996, Volume II Chapter 14) is directly motivated by this phenomenon. Several alternative long-memory, or fractionally integrated, volatility models have also been suggested in the literature to achieve this same goal, as exemplified by Comte and Renault (1998, Volume I Chapter 26) and Calvet and Fisher (2002, Volume I Chapter 27).

In spite of the proliferation of different GARCH models over the past three decades, the simple GARCH(1,1) remains the go-to model.<sup>11</sup> Quite remarkably, this simple recursive specification not only provides a good in-sample description of the volatility clustering for daily and coarser frequency returns but, as shown in the paper by Hansen and Lunde (2005, Volume I Chapter 15), the model typically also performs on par with, or even better in a forecasting sense, than the majority of the more complicated-to-implement GARCH models that have subsequently been proposed in the literature.

### Stochastic Volatility Models

Stochastic Volatility (SV) and GARCH models are closely related and, although there was much debate in the early literature about their relative merits, they are most fruitfully viewed as complementary rather than competitors. In fact, GARCH models can be classified as a specific type of SV model, possessing certain strengths and weaknesses relative to alternative SV type formulations.

The key distinction between SV and GARCH models, commonly adopted in the literature, is related to the specification of the volatility dynamics. GARCH models postulate an expression for the conditional variance as an explicit function of directly observable information, allowing for easy inference and practical use. SV models, on the other hand, include a separate innovation in the volatility equation.<sup>12</sup> This renders volatility latent, in turn necessitating the use of more complicated estimation and inference procedures. Importantly, whereas GARCH models are inherently discrete-time models, defining volatility over a fixed non-trivial time interval, SV models may be meaningfully specified both in discrete and continuous time. In fact, the primary advantage of the SV approach stems from its use in continuous-time settings and corresponding, often arbitrage based, asset pricing settings, as exemplified by some of the papers included in Part I of Volume II on Options and Volatility. The development of the realized volatility concept and the associated use of high-frequency intraday data for accurate volatility measurements, discussed Parts II and III, is also closely linked to the continuous-time SV framework.

In parallel to the difficulties of pinpointing a single study as first identifying volatility clustering per se, it is difficult to point to the exact origin of the SV literature. However, one major source of inspiration is Clark (1973, Volume I Chapter 16) and the so-called Mixture of Distributions Hypothesis (MDH). The MDH, in turn, is related to the notion of time-deformation, previously formalized more abstractly in statistics through time-changed Brownian motions, and the idea of an “economic time scale” operating at a different speed from regular calendar time.

More specifically, assume that the daily price change of some financial asset reflects the near continuous reaction to a myriad of different “news” arriving throughout the trading day. If the number of “news” arrivals is large, one may expect a central limit theory to apply, so the daily returns should be well approximated by a conditional normal distribution, with the conditioning variable being the number of relevant “news” events. More formally,

$$y_t \approx \mu_\delta s_t + \sigma_\delta s_t^{1/2} z_t, \quad (7)$$

where  $y_t$  denotes the daily return,  $s_t$  reflects the number of “news” arrivals on day  $t$ ,  $\mu_\delta$  and  $\sigma_\delta$  represent the typical response to a “news” event and the variation in that response, respectively, and  $z_t$  is *i.i.d.*  $N(0,1)$ . Intuitively, days with high information flow display more trading activity and, on average, larger price fluctuations than days with fewer “news” arrivals. Unconditionally, as long as  $s_t$  fluctuates through time, the MDH framework implies an unconditionally fat-tailed normal mixture distribution for returns.

In his empirical application of the MDH, Clark (1973) suggested the use of a log-normal distribution for the latent  $s_t$  process, thus implying a lognormal-normal mixture distribution for the returns.<sup>13</sup> This, of course, leaves unanswered the questions of what actually drives  $s_t$  and determines the within day price fluctuations. In an important subsequent paper, Tauchen and Pitts (1983, Volume I Chapter 17) explored the use of data on trading volume as a proxy for the  $s_t$  “news” arrival process, focusing their empirical investigations on unconditional features of the joint volume–volatility relationship implied by the model. Going one step further, Andersen (1996, Volume I Chapter 18) provides a more full-fledged dynamic account of the volume–volatility relationship within an extended MDH framework, focusing explicitly on the ability of the model to account for empirically realistic volatility clustering.

The expression for the returns underlying the MDH discussed above, in which the  $s_t$  process is not directly observable, has also motivated the use of more general SV models, in which the volatility is governed by its own separate stochastic process. Taylor (1982, Volume I Chapter 19), in particular, was the first to formulate and estimate a genuine SV model of the form,

$$\sigma_t^2 = \exp(h_t/2), \quad h_t = \mu + \varphi h_{t-1} + \eta_t, \quad (8)$$

where the  $\eta_t$  innovation is assumed to be *i.i.d.* normally distributed. Taylor completes the model by linking the latent volatility to the observed returns via the multiplicative specification,

$$y_t = x_t \sigma_t, \quad (9)$$

where  $x_t$  is a mean zero, unit variance process, independent of  $\sigma_t$ , although not necessarily *i.i.d.* Consistent with the basic premise of volatility clustering, the AR(1) structure for  $h_t$  readily translates into serial correlation in  $y^2$  and  $|y_t|$ , which Taylor (1982) corroborates empirically.<sup>14</sup>

The multiplicative specification adopted by Taylor (1982), in which  $x_t$  is not necessarily *i.i.d.*, bears some resemblance to the MDH framework, with a common latent  $s_t$  process affecting the mean and the variance. However, it complicates the interpretation of the model, as the dynamic dependencies in the  $\sigma_t$  “volatility process” are confounded by any predictability in the  $x_t$  “mean” process. The now common specification of SV models breaks this link by adopting a decomposition similar to that for the GARCH class of models discussed above,

$$y_t = \mu_{t|t-1} + \sigma_t z_t, \quad (10)$$

in which  $\mu_{t|t-1}$  denotes the conditional expectation, or mean, of the process,  $z_t$  is assumed to be *i.i.d.*, and  $\sigma_t$  refers to the stochastic volatility process of interest.

In a further parallel to the developments for GARCH models, numerous competing specifications for  $\sigma_t$  have been proposed within this setup. The Stochastic Autoregressive Volatility (SARV) class of models developed by Andersen (1994, Volume I Chapter 20),

$$v_t = \omega + \beta v_{t-1} + (\gamma + \alpha v_{t-1}) \eta_t, \quad (11)$$

where  $\eta_t$  denotes an *i.i.d.* sequence, and  $\sigma_t^2 = g(v_t)$  links the dynamic evolution of the latent state variable  $v_t$  to the stochastic volatility, naturally nests many of these specifications. For example, the popular log SV model is obtained for  $g(v_t) = \exp(v_t)$ , while SV generalizations of the GARCH(1,1) model are obtained for  $g(v_t) = v_t$ .

Irrespective of the exact functional form, the latency of  $\sigma_t^2$  renders the likelihood function unavailable in closed form, necessitating the use of specialized inference procedures for the estimation of SV models. Many of the earlier papers in the literature relied on state–space representations and accompanying Kalman filters for inferring  $\sigma_t^2$  from the data. However, these procedures can be highly inefficient in non-linear non-Gaussian settings. The indirect inference approach developed in the paper by Gouriéroux, Monfort and Renault (1993, Volume I Chapter 21), provides a possible solution. Intuitively, by simulating artificial data from the SV model of interest, which is easy to do, the indirect inference procedure works by finding the parameters of the model that provides the “closest” match between certain statistics calculated with the simulated data and the same statistics calculated with actual data. Gallant and Tauchen (1996, Volume I Chapter 22) go one step further, and formally show that by assessing “distance” via the score of a likelihood function from a GARCH-type model designed to provide an arbitrary close approximation to the SV model of interest as the sample size increases, asymptotically efficient parameter estimation may be achieved. This Efficient Method of Moments (EMM) approach is used in the study of interest rate dynamics by Andersen and Lund (1997, Volume I Chapter 23).

Although the likelihood function for SV models are not available in closed form, it is often possible to characterize the joint distribution through a series of conditional distributions, in which the distribution for a group of parameters is expressed conditional on the remaining parameters. Bayesian Markov Chain Monte Carlo procedures for estimation of specific discrete-time SV models based on this idea have been developed by Jacquier, Polson and Rossi (1994, Volume I Chapter 24).<sup>15</sup>

The GARCH and SV models estimated in the literature up until the mid-1990s were all discrete-time models, defining the volatility over specific time intervals, say a day, week or month. The fact that the models are not formally closed under temporal aggregation complicates any internally consistent model-based comparison of the volatility over alternate horizons. Building on earlier work related to temporal aggregation of GARCH models,<sup>16</sup> the paper by Meddahi and Renault (2004, Volume I Chapter 25) establishes conditions under which the general class of SARV models may be meaningfully defined across different sampling frequencies.

As noted earlier, one main advantage of the SV approach, compared to GARCH models, is their compatibility with the continuous-time framework.<sup>17</sup> Modeling volatility, as it evolves continuously through time, has the obvious advantage of allowing for any, possibly unevenly spaced, observation frequency. To illustrate, consider the representation for the instantaneous logarithmic return  $dp(t)$ ,

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \quad (12)$$

where  $\mu(t)$  is a continuous, locally bounded variation process, the diffusive volatility process  $\sigma(t)$  is strictly positive, and  $W(t)$  denotes a standard Brownian motion. The initial diffusion

models explored in the literature were not genuine SV models. Instead, the time-varying diffusive volatility was specified as a *deterministic* function of price, as in, for example, the Cox, Ingersoll and Ross (CIR) model for which  $\sigma(t) \equiv \sigma p(t)^{1/2}$ .

Meanwhile, as shown by Andersen and Lund (1997), the CIR model and related deterministic specifications for  $\sigma(t)$  are soundly rejected empirically. This led to the development of genuine continuous-time SV models, in which  $\sigma(t)$  is driven by a separate stochastic differential equation. The particular model estimated by Andersen and Lund (1997) takes the form,

$$d \log \sigma^2(t) = \beta(\alpha - \log \sigma^2(t))dt + vdB(t), \quad (13)$$

where  $B(t)$  denotes a separate standard Brownian motion, possibly correlated with the  $W(t)$  process that drives the price. This particular model is naturally interpreted as the continuous-time equivalent of Taylor's original discrete-time log SV model. Similarly, the diffusion limit of the GARCH(1,1) model, first derived and analyzed in a series of papers by Dan Nelson in the early 1990s, two of which are included in Part II of Volume II, takes the form,

$$d\sigma^2(t) = \beta(\alpha - \sigma^2(t))dt + v\sigma^2(t)dB(t). \quad (14)$$

A host of other continuous-time SV models have been explored in the literature. Many of these alternative specifications for  $\sigma(t)$  are motivated by the desire for more accurate option pricing models, and we will discuss these in connection with the papers included in Part I of Volume II on Options and Volatility.

Diffusive models for  $dp(t)$  portray the price as evolving continuously, or "smoothly," through time. Most asset prices occasionally exhibit large, seemingly discontinuous, increments, or "jumps."<sup>18</sup> In the context of continuous-time SV models, this may be accommodated by augmenting the representation for the diffusive price increments above with a jump term,<sup>19</sup>

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + J(t)dq(t), \quad (15)$$

where  $q(t)$  is a jump indicator taking the values zero (no jump) or unity (jump), and  $J(t)$  represents the size of the jump, if one occurs at time  $t$ . In this setting, the diffusive volatility and the jumps both contribute to the total overall price variation. Bates (1996, Volume II Chapter 5) and Eraker, Johannes and Polson (2003, Volume II Chapter 6), also argue forcefully for the importance of explicitly allowing for jumps in both the price and the volatility when using continuous-time SV models for the pricing of options. We will return to this and more general jump specifications, along with ways in which to empirically differentiate the continuous and jump variation, in our discussion of more recent papers included in Part III in Volume II on High-Frequency Data and Realized Volatilities.

The last two papers included in Part III relate to alternative continuous-time modeling procedures specifically designed to capture long-run dynamic dependencies in the volatility. These papers are partly motivated by the same empirical observations behind the FIGARCH model of Bollerslev, Baillie and Mikkelsen (1996). The paper by Comte and Renault (1998, Volume I Chapter 26), in particular, provides a framework for modeling and analyzing long-memory features in continuous-time SV models by using a fractional Brownian motion as the driving process for the diffusive volatility. The paper by Calvet and Fisher (2002, Volume I

Chapter 27) takes a very different tack by combining different Brownian motions subject to different time scale deformations. This same multifractal approach has subsequently been used in a series of papers by the same two authors.<sup>20</sup>

### Multivariate Volatility Models

The papers included in Parts II and III all concern univariate volatility models. In finance, of course, the risk of an individual asset is more appropriately measured by the systematic risk, typically defined by the covariance of the asset return with some benchmark portfolio, as opposed to the asset's individual volatility. Similarly, the non-diversifiable risk of an asset depends crucially on the covariances with other assets as opposed to its own variance.

To illustrate, consider the static one-period CAPM. The expected excess return on asset  $i$ ,  $r_i$ , equals the beta of the asset,  $\beta_i \equiv Cov(r_M, r_i)/Var(r_M)$ , times the expected excess return on the market,  $r_M$ . Hence, it is the covariance with the market, not the return volatility, that matters. More generally, absent arbitrage opportunities, the fundamental asset pricing theorem implies the existence of a stochastic discount factor, or pricing kernel,  $m$ , such that  $E[m(1 + r_i)] = 1$ . From this expression, one readily derives, for  $r_f$  denoting the risk-free interest rate, that,

$$E(r_i) = r_f - (1 - r_f)Cov(m, r_i), \quad (16)$$

underscoring the importance of covariance risk from an asset pricing perspective.

The unconditional “textbook representation” above obviously ignores any conditioning information.<sup>21</sup> The realization stemming from the univariate GARCH and SV papers, that the *conditional* variance of speculative returns are highly predictable, naturally prompted the question of whether similar procedures can be applied for modeling and forecasting time-varying conditional covariances? And if so, whether they may be employed for improved asset pricing and general financial decision-making, thus extending the traditional unconditional framework to a conditional setting?

Accordingly, Bollerslev, Engle and Wooldridge (1988, Volume I Chapter 28) introduced the multivariate GARCH model in the context of a dynamic CAPM. In analogy to the decomposition for univariate GARCH models, the multivariate version for the  $N \times 1$  vector process,  $Y_t$  is naturally decomposed as,

$$Y_t = M_{t|t-1} + \Omega_{t|t-1}^{1/2} Z_t \equiv M_{t|t-1} + e_t, \quad (17)$$

where  $M_{t|t-1}$  and  $\Omega_{t|t-1}$  denote the  $N \times 1$  conditional mean and  $N \times N$  conditional covariance matrix, respectively, and  $Z_t$  is an *i.i.d.* vector process with zero mean and identity covariance matrix. The so-called *vech* parameterization for  $\Omega_{t|t-1}$ , adopted by Bollerslev, Engle and Wooldridge (1988) closely mirrors that of the univariate GARCH(1,1) model,

$$vech(\Omega_{t|t-1}) = C + A vech(e_{t-1} e'_{t-1}) + B vech(\Omega_{t-1|t-2}), \quad (18)$$

where  $e_t = \Omega_{t|t-1}^{1/2} Z_t$ , and  $vech(\cdot)$  denotes the operator, that stacks the  $N(N+1)/2$  unique elements in the lower triangular part of a symmetric matrix into a  $N(N+1)/2 \times 1$  vector. General

conditions for this and related *vech* parameterizations to be well defined and for the  $\Omega_{t|t-1}$  conditional covariance matrix to be positive definite, have been derived by Engle and Kroner (1995, Volume I Chapter 29).

At first sight, the multivariate *vech* GARCH(1,1) parameterization may appear natural and straightforward to estimate. However, in its unrestricted form, in which each of the  $N(N+1)/2$  unique elements of  $\Omega_{t|t-1}$  may depend on lagged values of all the other elements, the model has a total of  $N^4/2 + N^3 + N^2 + N/2$  unique parameters, rendering it infeasible much beyond the bivariate case. Many subsequent papers in the multivariate volatility literature seek to resolve this issue through the formulation of empirically realistic, yet parsimonious, models that also guarantee positive definiteness of the implied conditional covariance matrices.

The multivariate factor ARCH models proposed by Diebold and Nerlove (1989, Volume I Chapter 30) affords one such approach. Factor structures are, of course, central to the field of finance, and the Arbitrage Pricing Theory (APT) in particular. To illustrate, consider a one-factor model of the form  $Y_t = a + bf_t + u_t$ , where  $u_t$  is assumed to be *i.i.d.* with covariance matrix  $\Lambda$ . The conditional covariance matrix for the observable  $Y_t$  process then takes the form,

$$\Omega_{t|t-1} = bb'\sigma_{f|t-1}^2 + \Lambda, \quad (19)$$

so that the temporal variation in the covariances is driven solely by shifts in the conditional factor variance,  $\sigma_{f|t-1}^2$ . This, and related multi-factor GARCH representations developed by Engle, Ng, and Rothschild (1990), greatly reduces the number of free parameters relative to the *vech* parameterization, while also ensuring that  $\Omega_{t|t-1}$  is positive definite. A drawback, however, is that with  $\Lambda$  being time-invariant, certain portfolios (with weights orthogonal to the loadings  $b$ ) have constant volatility, which runs counter to the volatility clustering observed empirically for practically all portfolios.

Another empirically successful class of multivariate models rely on the unique decomposition,

$$\Omega_{t|t-1} = D_{t|t-1} \Gamma_{t|t-1} D_{t|t-1}, \quad (20)$$

where  $D_{t|t-1}$  denotes the diagonal matrix of conditional standard deviations, and  $\Gamma_{t|t-1}$  denotes the matrix of conditional correlations. In particular, assuming that the temporal variation in the covariances are driven solely by the temporal variation in the conditional standard deviations, or  $\Gamma_{t|t-1} \equiv \Gamma$ , the Constant Conditional Correlation (CCC) GARCH model, proposed by Bollerslev (1990, Volume I Chapter 31), greatly reduces the number of parameters and effectively simplifies the estimation to  $N$  univariate GARCH models. The first genuine multivariate stochastic volatility model estimated by Harvey, Ruiz and Shephard (1994, Volume I Chapter 32) relies on a similar constant conditional correlation assumption. The assumption of constant conditional correlations may provide a reasonable simplification when modeling the volatility over a short time span but, arguably, it is too restrictive for longer time horizons. In response, Engle (2002, Volume I Chapter 33) introduced the Dynamic Conditional Correlation (DCC) class of multivariate GARCH models, in which  $\Gamma_{t|t-1}$  may vary across time.<sup>22</sup> Since the conditional correlation matrix contains  $N(N-1)/2$  unique elements, unrestricted modeling of  $\Gamma_{t|t-1}$  still presents formidable challenges, even for moderately sized  $N$ . However, by accounting for part of the temporal variation through the conditional standard deviations, the dynamic dependencies of the conditional correlations may be easier to model than the

dependencies in the conditional covariances. For instance, the DCC model employed by Engle (2002) relies on simple exponential smoothing. Many other more elaborate DCC models, including models explicitly designed to allow for asymmetric correlations during market up- and down-turns along the lines of Longin and Solnik (2001) and Ang and Chen (2002), have been explored in the subsequent literature.<sup>23</sup>

The CCC and DCC models tie together a set of univariate volatility models, using the conditional correlations as the “glue.” This linkage is typically augmented with the assumption of conditional normality to obtain a fully specified joint likelihood. Instead of postulating a specific joint distribution, copulas provide an alternative statistical framework for linking a set of marginal distributions.<sup>24</sup> Patton (2006, Volume I Chapter 34) was among the first to use copulas in the specification and estimation of multivariate GARCH models, highlighting the importance of asymmetric linkages. In a related context, several studies including Poon, Rockinger, and Tawn (2004), show that standard GARCH models do not satisfactorily account for the strong joint dependencies typically observed in the tails. The multivariate Extreme Value Theory (EVT) and dependency measures employed in that paper provide a natural extension of the univariate EVT approximations in the paper by McNeil and Frey (2000). Of course, just as the matrix of conditional correlations that links univariate GARCH models in the CCC setting may change over time, so may the conditional copulas that tie together a set of univariate volatility models. A number of recent studies have further explored the idea of dynamically varying conditional copulas.

## Options and Volatility

The pricing, trading and hedging of options constitute one of the most direct uses of volatility models and forecasts. Many important developments in the literature have been motivated by the desire to develop accurate option pricing models that conform to the observed volatility dynamics. The celebrated Black–Scholes–Merton<sup>25</sup> option pricing formula assumes that the returns of the underlying asset are *i.i.d.* with constant volatility  $\sigma$ , along with the feasibility of (costless) continuous trading and a constant risk-free interest rate. The formula expresses the price of an option as a function of  $\sigma$ , along with a set of observable variables and options characteristics (the price of the underlying asset, the risk-free rate, and the maturity and exercise price of the option). Conversely, it is possible to invert the formula to infer the volatility that would make an observed market price the “right” value for the option. This is commonly referred to as the Black–Scholes implied volatility, or simply the implied volatility. If the assumptions underlying the derivation of the formula were correct, the implied volatilities for different options would all coincide.

The constancy of the option implied volatilities is straightforward to test. Latané and Rendleman (1976, Volume II Chapter 1) provide one of the first systematic investigations of this hypothesis based on equity option prices. The results are unambiguous – the Black–Scholes implied volatilities are not constant.<sup>26</sup> Quoting the paper: “During some period, investors may feel that stocks in general are more risky than on other occasions.” Consistent with the presence of time-varying volatility, further empirical analyses document that the Black–Scholes–Merton model systematically misprices in- and out-of-the-money options, generating the so-called volatility smile or smirk.

At this point, the challenge was issued. How do we accommodate time-varying volatility within a tractable model of option prices,<sup>27</sup> and what is the suitable metric for volatility in that context? The paper by Hull and White (1987, Volume II Chapter 2) explicitly allows for an independent, time-varying, continuously evolving SV factor. Assuming that this additional volatility risk factor is not priced in equilibrium, the Hull–White option price simply equals the expected Black–Scholes–Merton price, where the expectation (under the actual or statistical probability measure,  $\mathbb{P}$ ), is taken with respect to the future integrated volatility (IV) over the life of the option,

$$E_t^{\mathbb{P}}[IV_{t,t+T}] = E_t^{\mathbb{P}} \left[ \int_t^{t+T} \sigma^2(u) du \right]. \quad (21)$$

This same statistic also plays a crucial role in the more recent literature on High-Frequency Data and Realized Volatilities discussed in Part III.

The option pricing refinements discussed above all rely on genuine SV models. Duan (1995, Volume II Chapter 24) provides the first discrete-time GARCH-based option pricing framework. The pricing principle is based on the assumption that the one-period returns of the underlying asset and the (latent) stochastic discount factor are jointly log-normally distributed, together with a local risk neutral valuation principle, which in turn results in an easy-to-implement simulation-based scheme involving a slightly modified GARCH model. This same valuation principle has subsequently been used in a number of other GARCH-based option pricing studies.

An important complication arising from the pricing of equity-index options, in particular, is the pronounced negative correlation between returns and volatility. This so-called leverage effect also motivated the GJR-GARCH and EGARCH models developed by Glosten, Jagannathan and Runkle (1993) and Nelson (1991), included in Volume I. Heston (1993, Volume II Chapter 3) first proposed an option pricing formula for a continuous-time stochastic volatility process, in which the innovations to returns and volatility are contemporaneously correlated. The model also allows for the possibility that volatility risk is priced, which further complicates the dynamic hedging arguments underlying the pricing of options vis-à-vis the Black–Scholes–Merton model with constant volatility.<sup>28</sup>

The Heston (1993) SV model stipulates that volatility follows a simple square-root process,

$$d\sigma^2(t) = \beta(\alpha - \sigma^2(t))dt + \nu\sigma(t)dB(t), \quad (22)$$

where the Brownian motion,  $B(t)$ , is contemporaneously correlated with the asset returns. Importantly, the model allows for an analytically tractable option pricing formula.

Even with the incorporation of stochastic volatility, it proved difficult to procure an acceptable fit to the cross-section of short-dated options. Consequently, Bates (1996, Volume II Chapter 5) argues for the importance of allowing for discontinuities, or jumps, in the price itself as an additional source of return variation, while maintaining the square-root model for  $\sigma(t)$  for analytical tractability. Bates (1996) assumes a constant jump intensity, but the subsequent literature demonstrates that tractable pricing formulas are valid in more general settings. This includes a jump intensity that is affine in the underlying state variables, such as factors associated with the spot variance, see, for example, Duffie, Pan and Singleton (2000). Eraker, Johannes and Polson (2003) go one step further, arguing it is important to also allow

for jumps in the volatility, especially during periods of market stress. They are also among the first to contemplate the possibility of co-jumps in returns and volatility, where both processes display a (correlated) discontinuity at the same moment. This feature has since become common among stochastic volatility, jump-diffusive specification for option pricing applications.

As the financial economics literature was developing tractable option pricing models, accommodating the stylized features of the volatility dynamics, a branch of the financial engineering literature was concerned with techniques for extracting volatility information directly from the option prices, without imposing the counterfactual Black–Scholes–Merton conditions. Much of this work emanated from the research departments of investment banks, seeking to generate instruments suitable for volatility trading and hedging, while avoiding strong assumptions regarding the volatility process. An important early contribution is Neuberger (1994), who documents the critical role of the payoff on a log-contract for spanning volatility risk. Among other significant early contributions, we have Dupire (1993), Demeterfi, Derman, Kamal, and Zou (1999), and Carr and Madan (1998). Ultimately, the notion of model-free volatility, was developed through the replication of the log contract via a large set of out-of-the-money (OTM) options, with directly observable prices. Among the first to introduce this concept into the financial economics literature is Britten-Jones and Neuberger (2000, Volume II Chapter 7). The point is that the expected integrated variance under the risk-neutral measure,  $\mathbb{Q}$ , effectively is well approximated by a specific linear combination of OTM options. That is, for any continuous-time stochastic volatility diffusion, the following quantity is effectively observed,

$$E_t^{\mathbb{Q}}[IV_{t,t+T}] = E_t^{\mathbb{Q}}\left[\int_t^{t+T} \sigma^2(u) du\right]. \quad (23)$$

The concept has numerous theoretical and empirical implications. It implies that we, under appropriate regularity conditions, may treat the price of expected future volatility (the actual expected volatility *plus* any volatility risk premium) as a traded object, even without having to specify a parametric model for the (diffusion) process governing the underlying asset returns. As a practical matter, this notion is now embodied in the widely-followed VIX volatility index, disseminated almost continuously during trading hours from the Chicago Board Options Exchange (CBOE).<sup>29</sup> Ultimately, the key to this development is the ability of linear combinations of option payoffs to span, and thus price, the payoffs of a wide range of different contracts, as originally argued by Ross (1976) and Breeden and Litzenberger (1973). Bakshi and Madan (2000) provide a striking extension of these ideas, documenting how conditional asset return moments under the risk-neutral measure may be extracted from a large cross-section of OTM options written on the asset.

Armed with the concept of model-free volatility, Carr and Wu (2009, Volume II Chapter 8) explore the empirical behavior of the variance risk premium, defined as the gap between the expected volatility under the actual or objective versus the risk-neutral probability measure,

$$VRP_{t,t+T} = E_t^{\mathbb{P}}[IV_{t,t+T}] - E_t^{\mathbb{Q}}[IV_{t,t+T}]. \quad (24)$$

The term on the far right is readily obtained from observed option prices, as stipulated in the construction of the VIX measure. Carr and Wu (2009) rely on a rolling sample variance estimate

to proxy for the first term on the right hand side. By averaging such measures over the full sample, this provides an estimate of the average expected return volatility. They find that the gap in equation (24), that is, the variance risk premium, is highly significant and negative. This implies that investors tolerate large losses to ensure positive payoffs in future states where the volatility spikes upward. Since returns often drop significantly in such states (the so-called leverage effect), this resembles an insurance premium that pays off in “bad” states when asset returns are poor.

Finally, Bollerslev, Tauchen and Zhou (2009, Volume II Chapter 9) takes the analysis an important step further by providing measures of the conditional variance risk premium. This is achieved by exploiting intraday return data for the underlying asset to obtain an improved forecast of the expected return volatility. The notion of “realized volatility” is critical for this purpose, and it is reviewed extensively in Volume II Part II. The ability to generate reliable measures of the variance risk premium at each observation date  $t$  enables them to identify time-variation in the premium. Importantly, they find that this premium has significant explanatory power for the future equity returns. It suggests that a substantial part of the equity risk premium represents compensation for the uncertain future return variation. This theme has inspired much subsequent work in the broader asset pricing literature.

### **Volatility Forecasting and Evaluation**

One primary objective behind the development of the GARCH and SV class of models is the construction of volatility forecasts. The evaluation of such forecasts and, in turn, the comparison of the quality of competing volatility models, is hampered by the latency of the true *ex post* volatility. Nonetheless, it is evident that the volatility process displays pronounced persistence, coupled with a tendency for mean reversion over longer horizons. As such, the performance of any predictor hinges on two facets of the associated forecast procedure. The first is the model’s accuracy in assessing the current level of (latent) volatility. Given the volatility persistence, a successful procedure must anchor the forecasts at an appropriate initial value to produce useful short-term forecasts. Second, for medium- and longer-term forecasts, the model must also approximate the volatility dynamics well. That is, we need a thorough model selection and estimation procedure.

In this context, it is useful to reflect on the properties that render GARCH models successful, even when they do not constitute the “true” data generating process (dgp). In theory, GARCH can serve as a consistent filter for volatility, in a wide range of setting, through an asymptotic scheme that involves gathering ever more finely sampled observations over a fixed or shrinking time interval. Limit results derived assuming this type of sampling scheme are obtained through so-called continuous record, or in-fill, asymptotics. This perspective was highlighted in a series of important papers by Dan Nelson. Specifically, Nelson (1992, Volume II Chapter 10) demonstrates that even badly misspecified GARCH models can generate consistent volatility estimates, if the underlying continuous-time process is well-approximated by a diffusion. The key property of diffusions is that, by picking a sufficiently short time interval, the change in the volatility can be made arbitrary small. In other words, locally, volatility is effectively constant. Now, given a large number of high-frequency returns over this short interval, it is natural to estimate the (near constant) volatility by the average squared returns, exploiting the

fact that the (diffusive) mean return is negligible over short horizons. For the GARCH(1,1) model (3), estimation of the parameters in this setting will result in  $\hat{\beta} \rightarrow 1$ ,  $\hat{\alpha} \rightarrow 0$ , and  $\hat{\omega} \rightarrow 0$ , so it generates a close approximation to an even weighting scheme of the squared returns. Formally, the asymptotic scheme lets the small time interval, just prior to time  $t$ , shrink, while also letting the number of return observations within this diminishing interval increase, as we “fill in” the interval. Assuming the diffusion is sampled without noise or measurement error, the volatility may be estimated arbitrarily well, in the limit, via this GARCH filter. In summary, a misspecified GARCH model generally provides sensible volatility measurement, even without having to specify the dgp for the underlying continuous-time model in detail.

The mainstream finance literature often adapts to the time-varying nature of variances and covariances by estimating the critical asset pricing relationship over rolling windows, implicitly assuming that the time-variation is negligible over windows covering a few years. Foster and Nelson (1996, Volume II Chapter 11) develop a general in-fill asymptotic theory for rolling-sample covariance matrix estimation, exploiting ideas similar to those above. However, the focus here is on deriving distributional approximations for the measurement error in the conditional variance and covariance estimators if the underlying diffusion is sampled at a moderate (daily or weekly) frequency only, generating discrete non-Gaussian innovations. The usual tradeoff applies, as reliance on longer local windows reduces the variance of the estimator, but increases the bias arising from the variability of the conditional covariance matrix. Armed with explicit solutions for the bias and variance terms, they determine the optimal window lengths and weighting schemes. In general, the commonly used flat-weight kernels are not optimal, as variability of the covariance matrix implies more weight should be assigned to nearby observations. Instead, exponentially declining weights arise naturally, mimicking features of a GARCH model. A natural application is estimation of time-varying betas for financial assets using rolling samples with optimized window length and weighting scheme.

Unfortunately, the double-asymptotic scheme justifying the GARCH filter properties is hard to mimic in practice, even at fairly high sampling frequencies. One issue is the pronounced intraday pattern in return volatility. In equity markets, for example, the average volatility at the open or close of the trading day is more than double the corresponding volatility in the middle of the day. This type of repetitive, or diurnal, variation in volatility is an order of magnitude higher than the typical shift in average volatility from one trading day to the next. In other words, when measuring concurrent volatility, the time-of-day effect is crucial – there is a dominant “periodic” component. The implication is that (GARCH) filters employing squared returns must be highly localized to ensure that the rapidly shifting volatility pattern does not seriously bias the measure. A second issue is that, for short time intervals, it is necessary to sample (squared) returns at a very high frequency to generate a reasonably precise measure, that is, reduce the error variance. But basic market microstructure features, such as the bid-ask spread (inducing return reversals), tick size (generating rounding errors), and varying market liquidity (causing irregularly spaced observations), imply that there is a limit to the effective sampling frequency that should be used. Hence, the suspicion is that the bias from time-varying intraday volatility and the excess variance generated by noise at the highest return frequencies may conspire to render the approximation offered by the double-asymptotic scheme of GARCH filters unreliable in practice.

In fact, Andersen and Bollerslev (1997, Volume II Chapter 12) document that the GARCH filter estimates are badly distorted by the diurnal volatility pattern, as the sampling shifts from

daily to intraday frequencies. In particular, the implied volatility persistence drops precipitously, as the signal from the mean-reverting intraday pattern overwhelms the more persistent dynamic dependencies at the daily level. On a more constructive note, they illustrate how direct modeling of the intraday pattern alleviates the issue. Controlling for the intraday periodicity, the parameter estimates are roughly consistent with the theoretical predictions based on the GARCH(1,1) filter for the intermediate intraday sampling frequencies. Moreover, it is straightforward to extend this style of empirical analysis to include scheduled news releases, following the approach in Andersen and Bollerslev (1998a). Nonetheless, the GARCH filter is clearly not perfectly specified, and the behavior at high frequencies continue to defy the theoretical predictions.

Andersen and Bollerslev (1998, Volume II Chapter 13) take the crucial step of establishing a formal basis for non-parametric volatility estimation via a single level in-fill asymptotic scheme. The key is to focus, not on spot volatility at any given point in time, but rather the average volatility over a non-trivial time interval, like one trading day. This allows for the procedure to utilize numerous non-overlapping return observations without sampling at extremely high frequencies. This notion of volatility is intimately related to fundamental properties of the class of stochastic processes labeled semimartingales. This is fortuitous, as standard no-arbitrage conditions imply that the asset price and associated return process must constitute a semimartingale.<sup>30</sup> Formally, if the returns follow a generic diffusion, as given by equation (15), then the theory of quadratic variation implies that,

$$RV_{t,t+T}^{(n)} = \sum_{i=1}^{n \cdot T} r_{t+i/n}^2 \rightarrow \int_t^{t+T} \sigma^2(u) du, \quad \text{as } n \rightarrow \infty, \quad (25)$$

where the returns are sampled at equidistant intervals within the  $[t, t+T]$  interval, so that  $r_{t+i/n} = p_{t+i/n} - p_{t+(i-1)/n}$ . More generally, the returns may be sampled at irregular times,  $t_i$ ,  $i = 1, \dots, n \cdot T$ , as long as  $\sup_i |t_i - t_{i-1}| \rightarrow 0$ .

The cumulative sum of squared returns in equation (25) is labeled the realized volatility, or alternatively the realized variance or the realized quadratic variation. It provides a fully nonparametric estimator of the underlying integrated volatility, as there are no model or tuning parameters, and it applies in a very general context subject only to the fundamental no-arbitrage condition. Notice that it implies we, under ideal circumstances, can measure the *ex post* realized volatility with almost arbitrary accuracy. In practice, microstructure frictions will thwart such ambitions, as we reach very high sampling frequencies.

Following this line of reasoning, it is now standard to utilize daily realized volatility measures rather than return variation measures based on daily (squared or absolute) returns as the basic metric for assessing alternative volatility forecasts. The choice of the daily RV measure circumvents complications arising due to the intraday volatility pattern. Since these are diurnal, integration over the full trading day effectively annihilates the impact of the periodic pattern. Once this feature is stripped out of the realized volatility series, analysis of one-day-ahead forecasts confirms the superiority of the realized volatility measure as benchmark for performance due to the vastly improved *ex post* measurement of the quadratic return variation. Additional results along these lines are provided by Andersen, Bollerslev, and Lange (1999), where the use of realized volatility measures also are incorporated into the GARCH filter forecasting procedure, generating improved forecast performance for longer weekly and monthly horizons as well as for daily forecasts.

More generally, daily  $RV$  measures lend themselves to an effective reduced-form forecast procedure, as the series are directly observable and represent good, albeit imperfect, proxies for the underlying quadratic return variation. Hence, it is natural to approximate their dynamics through ARMA style models, enabling forecast procedures through standard time series techniques. However, this does not directly speak to the efficacy of the reduced-form forecasts, as the true model, and hence the optimal forecasts, are unknown. Andersen, Bollerslev and Meddahi (2004, Volume II Chapter 14) exploit the fact that relevant moments for a broad family of popular stochastic volatility diffusions, belonging to the class of so-called eigenfunction stochastic volatility models, can be obtained in closed form. Thus, assuming that a specific model in this class constitutes the actual  $dgp$ , they are able to assess the deviation between the reduced form  $RV$  and optimal mean-squared-error (MSE) forecast. They find that the loss in efficiency of relying on the reduced-form approach is quite limited relative to the optimal, yet infeasible, forecasts that hinge on the true  $dgp$ .

The reliance on the MSE metric raises the question of whether it provides a useful forecast criterion in the context of volatility modeling. Patton (2011, Volume II Chapter 15) addresses this issue. He documents that the use of conditionally unbiased, yet imperfect, volatility proxies can generate problems for the comparison of competing forecasts. He derives a set of loss functions for which the relative ranking of forecasts are robust to “noise” in the volatility proxy. Using these robust loss functions, one ensures that alternative volatility forecasts can be compared meaningfully, even as an imperfect proxy is used in lieu of the true *ex post* (latent) realized volatility in the computation of the average loss. He shows that the MSE criterion is the only robust loss function that depends on the discrepancy between the forecast and realization, while the quasi-maximum likelihood (QLIKE) criterion is the only robust loss function that depends on the relative forecast error. It also follows that volatility forecasts can be ranked robustly on the basis of the  $R^2$  from the so-called Mincer-Zarnowitz (MZ) regression. For volatility forecasting, a typical specification of the latter regression takes the form,

$$\widehat{RV}_{t,t+T}^{(n)} = \alpha_0 + \beta_0 \hat{V}_{t,t+T} + \beta_1 X_{t,T} + u_{t,t+T}, \quad (26)$$

where  $\widehat{RV}_{t,t+T}^{(n)}$  represents an *ex post* (conditionally unbiased, albeit imperfect) realized volatility measure,  $\hat{V}_{t,t+T}$  denotes the volatility forecast of interest,  $X_{t,T}$  is an alternative volatility forecast, and  $u_{t,t+T}$  is the regression residual.

Standard tests for unbiasedness involves testing  $\hat{\alpha}_0 = 0$  and  $\hat{\beta}_0 = 1$  for univariate regressions (excluding the alternative forecast  $X_{t,T}$ ), while tests for superiority of one forecast versus another exploits the relative size and significance of the regression coefficients  $\hat{\beta}_0$  and  $\hat{\beta}_1$ . The robust loss function results imply that the regression  $R^2$  associated with different predictors in equation (26) also provide a consistent criterion for ranking alternative volatility forecasts.

The above discussion of the merits of alternative forecast procedures is based on purely statistical criteria. Of course, in practice, the assessment should be based on their ability to minimize an economic loss function that reflects the intended use of the forecast. Fleming, Kirby and Ostdiek (2003, Volume II Chapter 16) evaluate the benefits in switching from daily to intraday returns for estimation of the realized conditional covariance matrix through observed *ex post* gains in portfolio performance. In particular, they find investors willing to pay a substantial certainty-equivalent monthly fee to shift towards the strategy exploiting the intraday realized volatility measures.<sup>31</sup>

## High-Frequency Data and Realized Volatilities

This previous section stressed the significant advantages associated with the application of high-frequency data based procedures in generating volatility forecasts and, in particular, the gains stemming from the use of non-parametric realized volatility measures. In fact, equation (25) represents a very basic property of semimartingales, yet the recognition that this result has practical relevance for volatility measurement of return processes more generally has had truly dramatic implications. It also represents the key feature unifying research across a variety of disciplines. As discussed in the previous section, the observation originates from financial econometrics, inspired by issues arising from empirical volatility measurement and estimation. However, it provides a direct link to the continuous-time models and notions of no-arbitrage commonly used in mathematical finance and financial engineering. It also provides a bridge to the statistical literature in which alternative methods for estimation of diffusive volatility have been studied intensively. And importantly, it embeds the study of return volatility within the large probabilistic literature dealing with semimartingales, encompassing a broad class of Lévy processes and a wide set of tools for analyzing properties of quantities like the quadratic variation – which equals the integrated volatility in equation (25) in the diffusive setting. Correspondingly, the rapid evolution of the “realized volatility” literature stems from concurrent developments by applied econometricians and financial economists using the basic volatility measurement tools, financial engineers developing nonparametric pricing formulas for realized volatility measures, statisticians adapting their tools to handle stochastic volatility and jump processes, and martingale scholars deepening the theory related to the analysis of jump and diffusive components of general Lévy processes possibly contaminated by “noise.” The papers included in this part comprise some of the first that helped set the stage for these theoretical developments and mapped out some of the basic empirical facts that arose from applications of the new tools. Since this is still very much an active area of research, we purposely focus our discussion on the earliest work, leaving out more recent developments that have served to extend and refine the new tools in other important ways.

The initial studies spurred by the notion of realized volatility centered on empirical exploration of the distributional properties of volatility. In fact, the basic RV theory, reviewed below, implies that the realized volatility may be treated as a model-free observable quantity, albeit subject to random measurement error. In one of the first detailed studies, Andersen, Bollerslev, Diebold and Labys (2001, Volume II Chapter 17) document that foreign exchange (FX) volatilities and correlations are highly correlated. In addition, the log-realized volatilities are approximately unconditionally Gaussian and, consistent with prior evidence, the (realized) volatility and correlation series display long-memory style persistence. These empirical findings were further buttressed through a contemporaneous study of the thirty individual Dow-Jones stocks by Andersen, Bollerslev, Diebold, and Ebens (2001). Inspired by these initial studies, a number of subsequent studies have further collaborated and expanded on these basic distributional properties of the RV measures; see, for example, Bollerslev, Hood, Huss, and Pedersen (2017).

Of course, the amassing of empirical facts are valuable but, ultimately, the development of suitable models and forecast procedures is required to render the results useful as a basis for practical decision-making. Building on the early insights from the RV literature, Andersen, Bollerslev, Diebold and Labys (2003, Volume II Chapter 18) generate a unified framework for

measuring, modeling and forecasting realized volatility. The measurement step consists simply of constructing the realized covariation measures from high-frequency returns. The modeling is naturally guided by the pertinent distributional features. Specifically, since the log RV series are near homoskedastic and display long-range dependence, their dynamics may be captured through reduced-form fractionally integrated autoregressive (ARFIMA) processes. Standard procedures are then available for inference and forecasting of RV over both short and long horizons. Moreover, predictive densities may also be constructed directly from these ingredients, albeit under fairly restrictive assumptions, including the absence of a leverage effect; that is, no correlation between the return and volatility innovations. Correspondingly, as Andersen, Bollerslev, Diebold and Labys (2003) show, this approach works admirably in predicting Value-at-Risk for FX return series.

The strong (realized) volatility persistence has subsequently been corroborated by numerous other empirical studies, and the hyperbolic decay of the volatility autocorrelation pattern appears near universal.<sup>32</sup> A simple way to approximate this feature was proposed by Corsi (2009, Volume II Chapter 19), who advocated the use of a reduced-form Heterogeneous AR-representation (HAR), in which future realized volatility depends on past realized volatilities recorded over different horizons. For example,<sup>33</sup>

$$RV_{t,t+T}^{(n)} = \gamma_0 + \gamma_d RV_{t-1,t}^{(n)} + \gamma_w RV_{t-5,t-1}^{(n)} + \gamma_m RV_{t-21,t-5}^{(n)} + \gamma_q RV_{t-63,t-21}^{(n)} + \varepsilon_{t,t+T}^{(n)}, \quad (27)$$

where the non-overlapping realized volatilities on the right-hand-side reflect the impact from the past daily, weekly, monthly and quarterly trading day volatility measures. This simple structure closely mimics the hyperbolic decay in the autocorrelation structure, while retaining the simplicity of an autoregressive time series model. For this reason the model is often referred to as a “poor man’s long-memory model.” Of course, additional regressors could also be included on the right-hand-side. This flexibility and parsimony has rendered the Corsi HAR representation very popular, and it has arguably emerged as the go-to model for RV-based volatility modeling and forecasting; see, for example, Bollerslev, Patton, and Quaedvlieg (2016) for a discussion and comparisons of several such HAR-style formulations.

The reduced-form HAR model is based on realized variation measures that would naturally appear in a more structural dynamic representation of some underlying economic system. As such, the estimation results and subsequent forecasts may assist in developing theories that provide an economically-based causal rationale behind the observations. Alternatively, if the sole purpose of the predictive regression is tailored towards improved forecast performance, other less economically meaningful explanatory variables could be employed. In this spirit, Ghysels, Santa-Clara and Valkanov (2006, Volume II Chapter 20) propose a somewhat different framework, where the regressors constitute a diverse collection of volatility measures sampled at different frequencies. This renders the interpretation of the forecast regression harder, but provides rich opportunities for experimentation with alternative specifications and different weighting schemes for the lags of the included variables. An essential ingredient of their mixed-data-sampling (MiDaS) approach is the selection of an effective, yet parsimonious, parametric weighting function for the regressors that ensures parsimony. This approach has also proven popular in the literature.

The studies discussed in this section so far mostly assume, explicitly or implicitly, that the underlying price path is continuous. This is clearly counterfactual. As discussed above, when

significant economic news hits the market and induces a reassessment of asset values, discontinuities or “jumps” in prices typically occur. This is particularly striking for explicitly scheduled announcements, whether in the form of the release of corporate earnings or macroeconomic news announcements. The paper by Andersen, Bollerslev, Diebold and Vega (2003, Volume II Chapter 21) is among the first to illustrate how high-frequency returns can be used to separately analyze the response of the re-turn (jump) and volatility to regularly scheduled macroeconomic announcements. They document that the typical announcement leads to an almost immediate price jump, followed by a short, but sustained period of elevated volatility. The latter generally leads to a new equilibrium price level within an hour, but the direction of the price change is generally unpredictable. Hence, this is effectively a price discovery process, where the information content of the release is processed more carefully and portfolio reallocation takes place in response to the updated assessment of the future risk–return tradeoff.

Following the approach outlined above, combining high-frequency return data with information on news releases – their *ex ante* expected values and *ex post* realizations – one may directly explore the relative information content of “surprise” announcements and their relation to the general economic conditions. A number of empirical studies have further explored that idea, including Lee and Mykland (2008), Lahaye, Laurent, and Neely (2011) and Lee (2012) among others.

The rapid growth in empirical studies relying on realized volatility measures brought forth a couple of rather fundamental questions. First, how should one formally deal with indications of occasional discontinuities, or jumps, in the price path. If these generate an unusual price dynamic, their implications for future RV is likely different from the impact of a corresponding surge in diffusive volatility. Hence, improved forecasting procedures may be possible by treating the two components differently. Second, the empirical volatility measures can be sensitive to a variety of features associated with the sampling scheme for the high-frequency data, as market microstructure “noise” disrupts the martingale properties of the price path, especially over very fine time intervals. How can one procure less noisy realized volatility measures in the face of these complications? These dual questions quickly spurred an intense development of new theory and accompanying procedures.

In regards to the first question, jumps in the price path are readily incorporated into the realized volatility measures. From the theory of quadratic variation for semimartingales, it readily follows that,

$$RV_{t,t+T}^{(n)} = \sum_{i=1}^{nT} r_{t+i/n}^2 \xrightarrow{P} \int_t^{t+T} \sigma^2(u) du + \sum_{t \leq u \leq t+T} J^2(u), \quad \text{as } n \rightarrow \infty, \quad (28)$$

where the convergence is uniform in probability, and  $J(u)$  denotes the jump in the log-price at time  $u$ . In other words, the return variation reflects both the quadratic variation of the diffusive component and the cumulative squared jumps. This is a very clean result, yet it still begs the question of how to gauge the differential impact of jumps versus stochastic (diffusive) volatility. Is it possible to separate the contribution of the two components? Also, is it possible to identify the individual jumps in the price path?

The first approach for separating the impact of jumps versus diffusive volatility for the overall return variation was proposed by Barndorff-Nielsen and Shephard (2004, Volume II

Chapter 22). It utilizes the so-called bipower variation measure defined as the scaled cumulative sum of adjacent absolute high-frequency returns. In particular, under standard regularity conditions,

$$BV_{t,t+T}^{(n)} = \frac{\pi}{2} \sum_{i=2}^{n-T} |r_{t+(i+1)/n}| |r_{t+i/n}| \xrightarrow{p} \int_t^{t+T} \sigma^2(u) du, \quad \text{as } n \rightarrow \infty, \quad (29)$$

where again the convergence is uniform in probability. Hence, the BV measure annihilates the impact of jumps asymptotically. The critical feature behind this result is the fact that the individual jumps are dampened and eventually vanish in the limit.<sup>34</sup>

Equation (29) readily enables the consistent estimation of the contribution to the total variation stemming from jumps as the difference between the RV and BV realized volatility measures,

$$RV_{t,t+T}^{(n)} - BV_{t,t+T}^{(n)} \xrightarrow{p} \sum_{t \leq u \leq t+T} J^2(u), \quad \text{as } n \rightarrow \infty. \quad (30)$$

A number of alternative estimators that achieve the identical decomposition have subsequently been put forth in the literature, promising among other things, improved estimation efficiency, less finite sample distortion, and/or better “noise” robustness, including the quantile-based approach of Christensen, Oomen and Podolskij (2010), the nearest neighbor approach of Andersen, Dobrev and Schaumburg (2012), along with various truncation based procedures discussed further below. The emergence of the jump-robust volatility measures also inspired much new empirical work.

For example, it allows one to build reduced-form forecast procedures along the lines of equation (26) in which the past diffusive and jump volatilities appear as separate regressors. The study by Andersen, Bollerslev, and Diebold (2007, Volume II Chapter 23) documents that the diffusive component of the return variation does indeed provide the by far most important input for the prediction of the future return variation. In fact, the jump component appears largely unpredictable and approximately *i.i.d.* A number of subsequent studies have confirmed these same general findings, and the notion that the jump and diffusive return variation components of the overall return variation provide distinct signals about the current market conditions and, more generally, warrant different economic interpretations.

The integrated volatility estimators discussed above remove the contribution of jumps without generating any specific indication of the individual jump times and sizes. The main approach for direct identification of price jumps originates with the threshold estimator, detailed in Mancini (2009, Volume II Chapter 24), building on earlier contributions going back to Mancini (2001). Correspondingly, this approach also provides an intuitively simple alternative integrated volatility estimator based directly on the classical RV estimator, except that the jumps are explicitly excluded from the cumulative sum of squared returns, by eliminating the (absolute) returns that exceed a local threshold and thus are deemed (statistically) too large to constitute purely diffusive price moves. The design and empirical implementation of related threshold procedures explicitly accounting for the strong intraday pattern in volatility have been explored by Andersen, Bollerslev, and Dobrev (2007), Lee and Mykland (2008), and Bollerslev and Todorov (2011b), among others.

In a parallel development, the potential complications for RV estimation induced by market microstructure “noise” were becoming widely recognized and analyzed. In some respects, this was not a new realization. The “ahead-of-its-time” account of volatility estimation from high-frequency data in Zhou (1996, Volume II Chapter 26) touches on many themes that were addressed in the subsequent noise-robust volatility literature. Although Zhou provides no notion of in-fill asymptotic convergence, he explicitly advocates volatility estimation from high-frequency data, provided suitable adjustments are made to account for the distortions arising from the presence of *i.i.d.* noise.<sup>35</sup> Specifically, if the observed log prices are contaminated by a noise term, that is independent of the price process, say  $\tilde{p}_i = p_i + e_i$ , where  $e_i$  is an *i.i.d.*  $(0, \sigma_e^2)$  noise process, then the observed returns,  $\tilde{r}_i$ , will be excessively volatile relative to the underlying efficient return variance, and negatively serially correlated. Specifically,

$$\tilde{r}_{i+1} = \tilde{p}_{i+1} - \tilde{p}_i = r_{i+1} + (e_{i+1} - e_i). \quad (31)$$

Hence, assuming that the efficient return,  $r_i$ , is a martingale, it follows that,

$$\mathbb{E}(\tilde{r}_{i+1}^2) = \mathbb{E}(r_{i+1}^2) + 2\sigma_e^2, \quad \mathbb{E}(\tilde{r}_{i+1}\tilde{r}_i) = -\sigma_e^2, \quad (32)$$

while all higher order return autocorrelations are zero. Thus, under these assumptions, it is quite simple to obtain an (approximately) unbiased RV estimator; simply include the adjacent first order return products along with the squared returns,

$$RV_{t,T}^{(n,AC1)} = \sum_{i=2}^{nT-1} (r_{t+i/n}^2 + r_{t+(i-1)/n}r_{t+i/n} + r_{t+i/n}r_{t+(i+1)/n}). \quad (33)$$

Moreover, if one is only willing to impose that  $\mathbb{E}(\tilde{r}_{i+k+1}\tilde{r}_i) = 0$ , for some small integer  $k > 1$ , thus allowing for short-lived noise dependence, it is possible to obtain an effectively unbiased measure by only sampling every  $k^{\text{th}}$  return in equation (33). In addition, one may “subsample” and generate  $k$  versions of the estimator, where the first exploits the high-frequency observation 1,  $k + 1$ , ...,  $n(k + 1)$ , ..., the next utilizes observations 2,  $k + 2$ , ...,  $nk + 2$ , ..., and so forth, until the  $k^{\text{th}}$  version, constructed from observations  $k$ ,  $2k$ , ...,  $n(k + 1)$ , .... As such, this enables the use of essentially all available observations and, in turn, generates a more efficient estimator by averaging these  $k$  separate (highly dependent, but not identical) versions of the “skip  $k$ ” estimator.” The “volatility signature plot,” first proposed by Andersen, Bollerslev, Diebold, and Labys (2000), affords a simple informal diagnostic to gauge the sampling frequencies at which the impact of the noise begins to bias the RV estimates. This approach is now commonly used in practice, typically with  $k$  corresponding to a five-minute sampling interval; the comparison of several competing RV estimators in Liu, Patton, and Shephard (2015) also concludes that: “it is difficult to significantly beat 5-minute RV.” The bottom line is that even sparse sampling of high-frequency data provides a dramatic improvement over volatility measures based on daily or coarser observations.

The issues confronting the construction of RV estimators in the presence of noise ultimately reflects the standard bias-variance tradeoff in econometrics. Under Zhou’s (1996) specific assumptions, it is also possible to explicitly derive the “optimal sampling frequency” in an

MSE sense, balancing the bias arising from the noise and the variance reduction obtained from the use of an increasing number of squared returns over finer time intervals. Meanwhile, as the complexities of the dynamics in the noise process became more widely recognized, research into the development of more complicated consistent noise-robust realized volatility estimators flourished. The formalization of the fact that *i.i.d.* noise induces divergence at the rate of the number of intra-day observations increases, or  $n \rightarrow \infty$ , was formally laid out in Zhang, Mykland and Aït-Sahalia (2005, Volume II Chapter 28) and Bandi and Russell (2008). Both papers supplement the results in Zhou (1996) regarding the optimal sampling frequency that minimize the MSE, by suitably controlling the bias-variance tradeoff. By exploiting the rate of divergence with  $n$ , Zhang, Mykland and Aït-Sahalia (2005) bias-adjust their two-scale estimator by suitably combining two RV estimators based on different sampling frequencies. The study by Hansen and Lunde (2006, Volume II Chapter 25) surveys the main issues and provide comprehensive evidence based on a series of tick-by-tick data for individual US stocks. They conclude that the noise process display features that transcend the *i.i.d.* noise assumption. In particular, it is correlated with the efficient price; the noise is time-dependent; and the magnitude of the noise has decreased over time; the more recent study by Jacod, Li, and Zheng (2017) further elaborates on these features.

In addition to the paper by Zhang, Mykland and Aït-Sahalia (2005) on the two-scale estimator,<sup>36</sup> we also include the paper by Barndorff-Nielsen, Hansen, Lunde and Shephard (2008, Volume II Chapter 27) on the design of a noise-robust realized kernel estimator and the paper by Jacod, Li, Mykland, Podolskij and Vetter (2009, Volume II Chapter 29) which rely on pre-averaging techniques. The kernel estimator, in particular, adapts and extends the well-known heteroskedasticity and autocorrelation (HAC) consistent “long-run” covariance matrix estimators to the high-frequency return setting by including terms that account for the serial correlation induced by the noise process. One may naturally think of this as an estimator in the spirit of equation (33), employing a set of suitable weighted autocorrelation terms. The pre-averaging estimator, on the other hand, consists of averaging the observed prices before constructing the returns, followed by subsampling based on a moving window rather than non-overlapping pre-averaged returns. As it turns out, all of these three popular approaches are asymptotically (for  $n \rightarrow \infty$ ) equivalent to first order, but differ meaningfully in their treatment of “edge effects,” associated with the fact that the endpoint of the time interval of estimation does not allow the weighting scheme to include observations that fall outside the interval.

As a final theme, we also include two papers related to multivariate realized volatility and correlation estimation. This is obviously an important research topic, and the literature regarding this has grown tremendously in recent years. However, there is, as of yet, no consensus on the optimal approach, and a variety of alternative procedures are under development as we write. In particular, random and sparse matrix techniques, factor modeling, and big data inspired methodologies are now all being applied to the problem of estimating very large dimensional realized covariance matrices. As such, we confine our discussion to a couple of early contributions that align themselves well with the other papers in the volume.

One problem invariably complicates all multidimensional covariance matrix estimators, namely the issue of generating synchronous return observations at high sampling frequencies. Epps (1979, Volume II Chapter 30) in another “ahead-of-its-time” paper notes how information that naturally impact a set of stocks within the same industry is not reflected instantaneously in all prices. In turn, this induces a strong pattern in the return correlations as a function of

the return interval. Although modern liquid financial markets are unlikely to display the same pronounced adjustment or informational lags, the so-called “Epps effect” is ubiquitous in high-frequency return cross-correlations. That is, transaction costs and non-synchronous market activities across assets prevent simultaneous updating of prices and quotes, leading to positive lagged cross-correlations. These Epps effects present an additional obstacle for pushing the sampling frequencies used for realized covariance matrix estimation near the tick-by-tick level. This is particularly problematic in the high-frequency setting, as unrestricted estimates for the integrated covariance matrix over a given horizon involves a severe loss of degrees of freedom. If we want to reliably estimate the  $N(N+1)/2$  separate elements in the covariance matrix, we need a non-trivial multiple of  $N$  synchronized return observations. The development of new procedures and estimators to deal with these complications is an area of active current research.

Related to this, the final paper by Barndorff-Nielsen and Shephard (2004, Volume II Chapter 31) deliver the multi-variate limiting distribution for the standard RV estimator in the case of no noise or jumps. As detailed in Barndorff-Nielsen and Shephard (2002), for the univariate version corresponding to the stochastic volatility diffusion (12), conditional on the volatility path over  $[t, t+T]$ , the following mixed-Gaussian limit result obtains,

$$\sqrt{nT} \left( RV_{t,t+T}^{(n)} - \int_t^{t+T} \sigma^2(u) du \right) \xrightarrow{\mathcal{L}} N \left( 0, 2 \int_t^{t+T} \sigma^4(u) du \right) \quad \text{as } n \rightarrow \infty. \quad (34)$$

Since the volatility path is not observed, this does not provide a practical tool for inference regarding the integrated volatility. However, Barndorff-Nielsen and Shephard also show that the integrated quarticity,  $\int_t^{t+T} \sigma^4(u) du$ , can be estimated consistently for a diffusive price process by the realized quarticity,

$$\frac{nT}{3} \sum_{i=1}^{nT} r_{t+i/n}^4 \xrightarrow{p} \int_t^{t+T} \sigma^4(u) du \quad \text{as } n \rightarrow \infty. \quad (35)$$

Combining these two expressions, they further conclude that, if the innovations to the mean and volatility processes are independent of the Brownian motion  $W(t)$  generating the return innovations, then,

$$\frac{\left( RV_{t,t+T}^{(n)} - \int_t^{t+T} \sigma^2(u) du \right)}{\sqrt{\frac{2}{3} \sum_{i=1}^{nT} r_{t+i/n}^4}} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{as } n \rightarrow \infty. \quad (36)$$

Barndorff-Nielsen and Shephard (2004) extend these results to the multivariate setting. The covariance of asset returns with risk factors is, of course, a critical ingredient of empirical and theoretical asset pricing, and these tools provide entirely new opportunities for the study of time-variation in risk exposures, risk pricing, and hedge portfolios. It also remains an active area of research to modify the distributional results derived under stylized conditions in an optimal manner to render them applicable in the presence of jumps and market microstructure noise, using the type of techniques discussed previously in the univariate case.

These inference results also offer an excellent illustration of the strength the field has drawn through inspiration and input from different disciplines. For instance, as noted in Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006), the results in equations (34) and (36) may be strengthened to involve stable convergence in law. This implies that the estimation error in the numerator of equation (36) is independent of the realized quarticity statistic in the denominator asymptotically, irrespective of the actual dependency structure of the stochastic volatility diffusion (12). As such, the result remain valid in the presence of a leverage effect (correlation between the return innovations and the volatility process), as has been widely documented empirically in the financial economics and econometrics literature. Meanwhile, the concept of stable convergence originally stems from stochastic process and martingale theory, see, for example, Jacod and Shiryaev (2003) for an introduction.<sup>37</sup> The book by Aït-Sahalia and Jacod (2014) provides a much more thorough discussion of these issues and many of the key theoretical developments underlying the new flourishing field of high-frequency financial econometrics.

## Notes

1. The word traces back to the Latin verb *volare*, meaning “to fly.”
2. In chemistry and physics, volatility describes how easily a substance will vaporize.
3. Mandelbrot actually attributes this observation to personal communication with Hendrik S. Houthakker.
4. Black and Scholes (1973) and Merton (1973).
5. The one page note by McCulloch (1985) titled “On Heteros\*edasticity” makes it clear that the Greek root of the word dictates the correct spelling as heteroskedasticity with a “k” and not a “c.” Ironically, McCulloch’s note was published in *Econometrica*, with a “c.”
6. This representation is not entirely general as there could be higher-order conditional dependence in the innovations, but it serves to illustrate the main idea.
7. The working paper by Rosenberg (1972), unpublished until its 2005 inclusion in *Stochastic Volatility: Selective Readings*, edited by Neil Shephard, contains an expression reminiscent of an ARCH model. It also provides evidence that a weighted average of past squared price changes predicts future squared price changes. What is missing relative to the ARCH model, however, is an explicit link of the autoregressive model for volatility to the conditional variance, and correspondingly an expression for the likelihood function.
8. Notable examples are Lamoureux and Lastrapes (1990) and, more recently, Engle and Rangel (2008).
9. This well-established empirical feature dates back at least to the papers by Mandelbrot (1963) and Fama (1965) discussed above, which were primarily concerned with characterizing unconditional return distributions.
10. Even if shocks to volatility in the IGARCH(1,1) model are infinitely persistent in a forecasting sense, Nelson (1990) shows that the model is still strictly stationary, so that the analogue to unit root processes for the conditional mean is imperfect.
11. Bollerslev (2010) provides an extensive summary of the many different GARCH models that have been proposed in the literature.
12. In the terminology of Cox et al. (1981), GARCH models are “observation-driven,” whereas SV models are “parameter-driven.”
13. Incidentally, this same unconditional distribution has also been advocated more recently by Andersen, Bollerslev, Diebold and Labys (2001, Volume II Chapter 17), motivated by the empirical observations that the distribution of daily returns standardized by realized volatility appears close to normal, while the realized volatility appears approximately log-normally distributed.

14. The empirical analysis in Taylor (1982) is based on a long time series of daily returns on sugar futures contracts. The seminal book by Taylor (1986) provides a more comprehensive empirical analysis, including autocorrelograms for the absolute and squared returns for many other assets.
15. Related MCMC simulation-based procedures were explored in concurrent work by Shephard (1993).
16. Most notably Drost and Nijman (1993) and Drost and Werker (1996).
17. The modeling of asset returns in continuous time traces back to the early contributions by Merton (1969) and Black and Scholes (1973), which forcefully demonstrated the analytical tractability of the diffusive framework. There is also a remarkable early precedent in the 1900 PhD thesis by Louis Bachelier, "Théorie de la Spéculation," *Annales de l'Ecole Normale Supérieure*, 3, Paris: Gauthier-Villars.
18. As discussed further by Andersen, Bollerslev, Diebold and Vega (2003, Volume II Chapter 21), many of the largest (in an absolute sense) price increments are readily associated with specific macroeconomic, or other, news announcements.
19. Merton (1976) first explored the use of jump diffusion models in finance, albeit under the assumption of a constant diffusive volatility  $\sigma(t) = \sigma$ .
20. The book by Calvet and Fisher (2008) provides a comprehensive discussion of this alternative approach.
21. Hansen and Richard (1987) were among the first to explicitly highlight the role of conditioning information in theoretical dynamic asset pricing models and corresponding tests.
22. A closely related specification was proposed in the concurrent paper by Tse and Tsui (2002).
23. The Engle (2009) book provides an extensive review of DCC models.
24. By Sklar's Theorem, any multivariate joint distribution can be written in terms of the corresponding univariate marginal distributions and a copula function that describes the dependence between the variables.
25. Black and Scholes (1973) and Merton (1973).
26. Black and Scholes (1972) previously noted that the return variance likely varies through time.
27. The early references on stochastic volatility and option pricing include Johnson and Shanno (1987), Scott (1987), Wiggins (1987), Stein (1989), and Melino and Turnbull (1990).
28. The discrete-time GARCH option valuation approach in Heston and Nandi (2000) is generally tractable and incorporates the original Heston (1993) model as a special case in the continuous limit. It has proven particularly popular in practice.
29. The regular CBOE VIX index refers to the cumulative one-month-ahead volatility of the S&P 500 equity-index. There are now numerous alternative model-free volatility indices disseminated from exchanges all over the world, covering different equity-indices, individual stocks, commodities, exchange rates, energy products, etc.
30. See, e.g., the account in Back (1991).
31. Some authors choose option pricing as the metric for gauging the incremental value provided by the daily realized volatility measures relative to daily return data, thus also speaking to the issues addressed in Part I of this volume, see, e.g., Corsi, Fusari and La Vecchia (2013) and Christoffersen, Feunou, Jacobs and Meddahi (2014).
32. This, of course, also motivates the FIGARCH model in Baillie, Bollerslev and Mikkelsen (1996) included in Volume I Part II, as well as Comte and Renault (1998) and Calvet and Fisher (2002) included in Volume I Part III.
33. This same specification can also be cast in terms of the square-root or logarithm of the realized volatilities.
34. This is formally achieved by the multiplication of each jump term by an adjacent diffusive term that becomes infinitesimal as  $n \rightarrow \infty$ . Although jumps may cluster, they are invariably separated by a small time interval over which the price path is continuous, ensuring the validity of the limiting argument.
35. This same specification also has a precedent in Roll (1984), who exploits the negative return autocorrelation induced by transaction prices bouncing between bid and ask quotes to estimate the size of the bid-ask spread.
36. The two-scale estimator is readily extended to multiple scales, and may also be adapted to allow for dependent noise case, as demonstrated in Aït-Sahalia, Mykland and Zhang (2011).

37. This notion can also be adapted to settings beyond high-frequency returns. Andersen, Fusari, and Todorov (2015) apply the in-fill asymptotic scheme to the strike prices of options with the identical underlying asset and same time to maturity. This enables them to develop novel inference tools for derivatives pricing.

## References

- Aït-Sahalia, Y., and Jacod, J., 2014. *High-frequency Financial Econometrics*. Princeton University Press, Princeton NJ.
- Aït-Sahalia, Y., Mykland, P., and Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics* 160, 160–175.
- Andersen, T.G., and Bollerslev, T., 1998a. Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* 1 (53), 219–265.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., and Ebens, H., 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61 (1), 43–76.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P., 2000. Great realizations. *Risk* 13, 105–108.
- Andersen, T.G., Bollerslev, T., Dobrev, D., 2007. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics* 138, 125–180.
- Andersen, T.G., Bollerslev, T., Lange, S., 1999. Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance* 6 (5), 457–477.
- Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169, 75–93.
- Andersen, T.G., Fusari, N., and Todorov, V., 2015. Parametric inference and dynamic state recovery from option panels. *Econometrica* 83 (3), 1081–1145.
- Ang, A., and Chen, J., 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63 (3), 443–494.
- Back, K., 1991. Asset pricing for general processes. *Journal of Mathematical Economics* 20 (4), 371–395.
- Bakshi, G., and Madan, D., 2000. Spanning and derivative-security valuation. *Journal of Financial Economics* 55 (4), 205–238.
- Bandi, F., and Russell, J., 2008. Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* 75, 339–369.
- Barndorff-Nielsen, O., Graversen, S., Jacod, J., and Shephard, N., 2006. Limit theorems for bipower variation in financial econometrics. *Econometric Theory* 22, 677–719.
- Barndorff-Nielsen, O., and Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64 (2), 253–280.
- Black, F., and Scholes, M., 1972. The valuation of option contracts and a test of market efficiency. *Journal of Finance* 27 (2), 399–418.
- Black, F., and Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81 (3), 637–654.
- Bollerslev, T., 2010. Glossary to arch (garch). In: *Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle* (eds Tim Bollerslev, Jeffrey R. Russell and Mark Watson), Oxford University Press, Oxford, UK, 137–164.
- Bollerslev, T., Hood, B., Huss, J., and Pedersen, L.H., 2017. Risk everywhere: Modeling and managing volatility. Tech. rep., Working Paper, Duke University, AQR Capital Management and Copenhagen Business School.
- Bollerslev, T., Patton, A.J., and Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics* 192, 1–18.

- Bollerslev, T., and Todorov, V., 2011a. Estimation of jump tails. *Econometrica* 79, 1727–1783.
- Bollerslev, T., and Todorov, V., 2011b. Tails, fears and risk premia. *Journal of Finance* 66, 2165–2211.
- Breedon, D., and Litzenberger, R., 1973. Prices of state contingent claims implicit in options. *Journal of Business* 51, 621–652.
- Calvet, L.E., and Fisher, A.J., 2008. *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press, San Diego, CA.
- Carr, P., and Madan, D., 1998. Towards a theory of volatility trading. In: *Volatility*, (ed. R.A. Jarrow), Risk Books; London, UK, 417–427.
- Christensen, K., Oomen, R., and Podolskij, M., 2010. Realized quantile-based estimation of integrated variance. *Journal of Econometrics* 159, 74–98.
- Christoffersen, P., Feunou, B., Jacobs, K., and Meddahi, N., 2014. The economic value of realized volatility: Using high-frequency returns for option valuation. *Journal of Financial and Quantitative Analysis* 49 (3), 663–697.
- Corsi, F., Fusari, N., and La Vecchia, D., 2013. Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics* 107 (2), 284–304.
- Cox, D.R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S.L., 1981. Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, 93–115.
- Demeterfi, K., Derman, E., Kamal, M., and Zou, J., 1999. More than you ever wanted to know about volatility swaps. Tech. rep., Working Paper, Goldman Sachs.
- Drost, F.C., and Nijman, T.E., 1993. Temporal aggregation of garch processes. *Econometrica*, 909–927.
- Drost, F.C., and Werker, B.J., 1996. Closing the garch gap: Continuous time garch modeling. *Journal of Econometrics* 74 (1), 31–58.
- Duffie, D., Pan, J., and Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.
- Dupire, B., 1993. Model art. *Risk* September, 118–120.
- Engle, R., 2009. *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press, Princeton, NJ.
- Engle, R.F., and Bollerslev, T., 1986. Modelling the persistence of conditional variances. *Econometric Reviews* 5 (1), 1–50.
- Engle, R.F., Ng, V.K., and Rothschild, M., 1990. Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics* 45 (1–2), 213–238.
- Engle, R.F., and Rangel, J.G., 2008. The spline-garch model for low-frequency volatility and its macroeconomic causes. *Review of Financial Studies* 21 (3), 1187–1222.
- Fama, E.F., 1965. The behavior of stock-market prices. *Journal of Business* 38 (1), 34–105.
- Hansen, L.P., and Richard, S.F., 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* 55, 587–613.
- Heston, S.L., and Nandi, S., 2000. A closed-form garch option pricing model. *Review of Financial Studies* 13 (3), 585–626.
- Jacod, J., Li, Y., and Zheng, X., 2017. Statistical properties of microstructure noise. *Econometrica* 85, 1133–1174.
- Jacod, J., and Shiryaev, A., 2003. *Limit Theorems for Stochastic Processes*. Springer, New York and Berlin.
- Johnson, H., and Shanno, D., 1987. Option pricing when the variance is changing. *Journal of Financial and Quantitative Analysis* 22, 143–152.
- Lahaye, J., Laurent, S., and Neely, C.J., 2011. Jumps, cojumps and macro announcements. *Journal of Applied Econometrics* 26 (6), 893–921.
- Lamoureux, C.G., and Lastrapes, W.D., 1990. Heteroskedasticity in stock return data: Volume versus garch effects. *Journal of Finance* 45 (1), 221–229.
- Lee, S., and Mykland, P., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* 21 (6), 2535–2563.
- Lee, S.S., 2012. Jumps and information flow in financial markets. *Review of Financial Studies* 25 (2), 439–479.

- Liu, L., Patton, A., Shephard, K., 2015. Does anything beat 5-minute rv? A comparison of realized measures across multiple asset classes. *Journal of Econometrics* 187, 293–311.
- Longin, F., and Solnik, B., 2001. Extreme correlation of international equity markets. *Journal of Finance* 56 (2), 649–676.
- Mancini, C., 2001. Disentangling the jumps of the diffusion in a geometric jumping brownian motion. *Giornale Dell'Istituto Italiano Degli Attuari* 64, 19–47.
- Mandelbrot, B., 1963. The variation of certain speculative prices. *Journal of Business* 36 (4), 394–419.
- McCulloch, J.H., 1985. On heteros\*edasticity. *Econometrica* 53 (2), 483.
- Melino, A., and Turnbull, S., 1990. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239–265.
- Merton, R.C., 1969. Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics* 51, 247–257.
- Merton, R.C., 1973. Theory of rational option pricing. *Bell Journal of Economics and Management Science* 4, 141–183.
- Merton, R.C., 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3 (1–2), 125–144.
- Nelson, D.B., 1990. Stationarity and persistence in the garch(1,1) model. *Econometric Theory* 6 (3), 318–334.
- Neuberger, A. (1994), The log contract: A new instrument to hedge volatility. *Journal of Portfolio Management*, Winter, 74–80.
- Officer, R.R., 1973. The variability of the market factor of the New York stock exchange. *Journal of Business* 46 (3), 434–453.
- Poon, S.-H., Rockinger, M., and Tawn, J., 2004. Extreme value dependence in financial markets: Diagnostic, models and financial implications. *Review of Financial Studies* 17 (2), 581–610.
- Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 1127–1140.
- Rosenberg, B., 1972. The behavior of random variables with nonstationary variance and the distribution of security prices. Tech. rep., Working Paper No.11, University of California Berkeley, Institute of Business and Economic Research., reprinted in N. Shephard, *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford, UK., 2005.
- Ross, S.A., 1976. Options and efficiency. *Quarterly Journal of Economics* 90, 75–89.
- Scott, L., 1987. Option pricing when the variance changes randomly: Theory, estimation and an application. *Journal of Financial and Quantitative Analysis* 22, 419–438.
- Shephard, N., 1993. Fitting non-linear time series models, with applications to stochastic variance models. *Journal of Applied Econometrics* 8 (S1), S135–S152.
- Stein, J., 1989. Overreactions in the options market. *Journal of Finance* 44, 1011–1023.
- Taylor, S., 1986. *Modelling Financial Time Series*. John Wiley, Chichester, UK.
- Tse, Y.K., and Tsui, A.K.C., 2002. A multivariate garch model with time-varying correlations. *Journal of Business and Economic Statistics* 20 (3), 351–362.
- Wiggins, J., 1987. Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics* 19, 351–372.