Volatility Forecasting*

Torben G. Andersen[a], Tim Bollerslev[b],
Peter F. Christoffersen[c] and Francis X. Diebold[d]

May 17, 2005

* This paper is prepared for the forthcoming *Handbook of Economic Forecasting* edited by Graham Elliott, Clive W.J. Granger, and Allan Timmermann, Amsterdam: North Holland. The work of Andersen, Bollerslev and Diebold was supported by grants from the U.S. National Science Foundation, and the work of Christoffersen was supported by FQRSC, SSHRC and IFM2. The authors gratefully acknowledge extremely helpful comments from two anonymous referees. The ususal disclaimer applies.

[a] Kellogg School of Management, Northwestern University, Evanston, IL 60208, and NBER, phone: 847-467-1285,
e-mail: t-andersen@kellogg.northwestern.edu

[b] Department of Economics, Duke University, Durham, NC 27708, and NBER, phone: 919-660-1846,
e-mail: boller@econ.duke.edu

[c] Faculty of Management, McGill University, Montreal, Quebec, H3A 1G5, and CIRANO, phone: 514-398-2869,
e-mail: peter.christoffersen@mcgill.ca

[d] Department of Economics, University of Pennsylvania, Philadelphia, PA 19104, and NBER, phone: 215-898-1507
e-mail: fdiebold@sas.upenn.edu

ABSTRACT

Volatility has been one of the most active and successful areas of research in time series econometrics and economic forecasting in recent decades. This chapter provides a selective survey of the most important theoretical developments and empirical insights to emerge from this burgeoning literature, with a distinct focus on forecasting applications. Volatility is inherently latent, and Section 1 begins with a brief intuitive account of various key volatility concepts. Section 2 then discusses a series of different economic situations in which volatility plays a crucial role, ranging from the use of volatility forecasts in portfolio allocation to density forecasting in risk management. Sections 3, 4 and 5 present a variety of alternative procedures for univariate volatility modeling and forecasting based on the GARCH, stochastic volatility and realized volatility paradigms, respectively. Section 6 extends the discussion to the multivariate problem of forecasting conditional covariances and correlations, and Section 7 discusses volatility forecast evaluation methods in both univariate and multivariate cases. Section 8 concludes briefly.

Table of Contents

1. Introduction

2. Uses of volatility forecasts

3. GARCH volatility

4. Stochastic volatility

5. Realized volatility

6. Multivariate volatility

7. Evaluating volatility forecasts

8. Concluding remarks

References

Figures

1. Introduction

In everyday language, volatility refers to the fluctuations observed in some phenomenon over time. Within economics, it is used slightly more formally to describe, without a specific implied metric, the variability of the random (unforeseen) component of a time series. More precisely, or narrowly, in financial economics, volatility is often defined as the (instantaneous) standard deviation (or "sigma") of the random Wiener-driven component in a continuous-time diffusion model. Expressions such as the "implied volatility" from option prices rely on this terminology. In this chapter, we use the term volatility in the looser descriptive sense, characteristic of economics and econometrics, rather than the precise notion often implied in finance. Nonetheless, much of our discussion will be motivated by the need for forecasting the volatility of financial asset return series.

Return volatility is, of course, central to financial economics. Indeed, as noted by Campbell, Lo and MacKinlay (1997):

> " ... what distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation ... Indeed in the absence of uncertainty, the problems of financial economics reduce to exercises in basic microeconomics" (p. 3).

This departure of finance from standard microeconomics is even more striking once one recognizes that volatility is inherently unobserved, or latent, and evolves stochastically through time. Not only is there non-trivial uncertainty to deal with in financial markets, but the level of uncertainty is latent. This imbues financial decision making with complications rarely contemplated within models of optimizing behavior in other areas of economics.

Depending on the data availability as well as the intended use of the model estimates and associated forecasts, volatility models are cast either in discrete time or continuous time. It is clear, however, that the trading and pricing of securities in many of today's liquid financial asset markets is evolving in a near continuous fashion throughout the trading day. As such, it is natural to think of the price and return series of financial assets as arising through discrete observations from an underlying continuous-time process. It is, however, in many situations useful - and indeed standard practice - to formulate the underlying model directly in discrete time, and we shall consider both approaches in the chapter. Formally, there is also no necessary contradiction between the two strategies, as it is always, in principle, possible to deduce the distributional implications for a price series observed only discretely from an underlying continuous-time model. At the same time, formulation and estimation of empirically realistic continuous-time models often presents formidable challenges. Thus, even though many of the popular discrete-time models in current use are not formally consistent with an underlying continuous-time price processes, they are typically much easier to deal with from an inferential perspective, and as such, discrete-time models and forecasting procedures remain the method of choice in most practical applications.

1.1 Basic Notation and Notions of Volatility

We first introduce some notation that will allow us to formalize the discussion of the different models and volatility concepts considered throughout the chapter. As noted above, although it is often natural to think about the process being forecasted as evolving in continuous time, many of the key developments in volatility forecasting have been explicitly formulated in terms of models for discretely sampled observations. In the univariate case, with observations available at equally spaced discrete points in time, we shall refer to such a process as,

$$y_t \equiv y(t) \qquad t = 1, 2, \ldots \tag{1.1}$$

where $y(t)$ in turn may be thought of as the underlying continuously evolving process. We shall assume throughout that the *conditional* second moments of the y_t process exist, and refer to the corresponding conditional mean and variance as,

$$\mu_{t|t-1} = E[\, y_t \mid \mathscr{F}_{t-1} \,], \tag{1.2}$$

and,

$$\sigma^2_{t|t-1} = Var[\, y_t \mid \mathscr{F}_{t-1} \,] = E[(y_t - \mu_{t|t-1})^2 \mid \mathscr{F}_{t-1} \,], \tag{1.3}$$

respectively, where the information set, $\mathscr{F}_{t-1}$, is assumed to reflect all relevant information through time $t-1$. Just as the conditional mean may differ from the unconditional mean by effectively incorporating the most recent information into the one-step-ahead forecasts, $\mu_{t|t-1} \neq E(y_t)$, so will the conditional variance in many applications in macroeconomics and finance, $\sigma^2_{t|t-1} \neq Var(y_t)$. This difference between conditional and unconditional moments is, of course, what underlies the success of time series based forecasting procedures more generally. For notational simplicity we will focus our discussion on the univariate case, but many of the same ideas readily carry over to the multivariate case. In the case of vector processes, discussed in detail in Section 6, we shall use the notation Y_t, with the corresponding vector of conditional means denoted by $M_{t|t-1}$, and the conditional covariance matrix denote by $\Omega_{t|t-1}$.

As previously noted, most of the important developments and applications in volatility modeling and forecasting have come within financial economics. Thus, to help fix ideas, we focus on the case of return volatility modeling and forecasting in the remainder of this section. To facilitate subsequent discussions, it will sometimes prove convenient to refer to the corresponding "price" and "return" processes by the letters p and r, respectively. Specifically, let $p(t)$ denote the logarithmic price of an asset. The return over the discrete interval $[t-h,t]$, $h > 0$, is then given by,

$$r(t,h) = p(t) - p(t-h) . \tag{1.4}$$

When measuring the returns over one time unit, $h = 1$, indicating, say, daily returns, we will generally drop the second indicator, so that

$$r(t) \equiv r(t,1) = p(t) - p(t-1) . \tag{1.5}$$

Also, for discrete-time models and procedures, we shall follow the convention set out above, indicating the timing of the returns by subscripts in lieu of parentheses,

$$r_t = p_t - p_{t-1} . \tag{1.6}$$

Similarly, we shall refer to the multivariate case involving vectors of returns by the upper case letter, R_t.

Consider the discretely sampled return process, r_t. This one-period return is readily decomposed into an expected conditional mean return and an innovation, where the latter may be expressed as a standardized white noise process scaled by the time-varying conditional volatility. Specifically, using the notation in equations (1.2) and (1.3),

$$r_t = \mu_{t|t-1} + \varepsilon_t = \mu_{t|t-1} + \sigma_{t|t-1} z_t , \tag{1.7}$$

where z_t denotes a mean zero, variance one, serially uncorrelated disturbance (white noise) process. This is the decomposition and volatility concept underlying the popular, and empirically highly successful, ARCH and GARCH type models discussed in Section 3. One reason that this approach is very convenient and tractable is that - conditional on the null hypothesis that all relevant information is observed and the model correctly specified - the volatility is known, or predetermined, as of time $t-1$.

The assumption that all relevant information is observed and used in the formation of conditional expectations in accordance with the true model is obviously strong, but has powerful and very convenient implications. In contrast, if some relevant information is not directly observable, then it is only possible to exploit a genuine subset of the full information set, say $\mathcal{F}_{t-1} \subset \mathcal{F}_{t-1}$. Under this scenario, the "true" conditional variance will be unobservable, even under correct model specification, and the volatility process becomes genuinely latent,

$$E [(r_t - E [r_t \mid \mathcal{F}_{t-1}])^2 \mid \mathcal{F}_{t-1}] \neq \sigma^2_{t|t-1} \equiv E [\epsilon_t^2 \mid \mathcal{F}_{t-1}].$$

Treating the volatility process as latent effectively transforms the volatility estimation problem into a filtering problem in which the "true" volatility cannot be determined exactly, but only extracted with some degree of error. This general line of reasoning is relevant for our discussion of stochastic volatility models in Section 4, and for the relationship between continuous and discrete-time modeling and forecasting procedures.

For now, however, we proceed under the convenient assumption that we are dealing with correctly specified models and the associated full information sets, so that the conditional first and second moments are directly observable and well specified. In this situation, the one-period-ahead volatility defined in (1.3) provides an unbiased estimate of the subsequent squared return

innovation. Consequently, model specification and forecast evaluation tests can be constructed by comparing the realization of the squared return innovations to the corresponding one-step-ahead forecasts,

$$\epsilon_t^2 \;=\; \sigma^2_{t|t-1} z_t^2 \;=\; \sigma^2_{t|t-1} \;+\; \sigma^2_{t|t-1}(\, z_t^2 - 1\,). \tag{1.8}$$

The second term on the right-hand-side has mean zero, confirming the unbiasedness of the conditional variance. However, there is typically a large amount of noise in the one-period squared return innovations relative to the underlying volatility, as manifest by a large idiosyncratic error component governed by the variance of z_t^2. In fact, for daily or weekly return data, this variance term is an order of magnitude larger than the period-per-period variation in the volatility process. Hence, even if the conditional variance can be seen as the proper forecasts of the corresponding "realized volatility," as given by the squared return innovation, the latter provides a poor ex-post indicator of the actual volatility over the period, and would consequently not provide a very reliable way of judging the quality of the forecasts. We return to this point below.

Before doing so, however, it is useful to think of the returns as arising from an underlying continuous-time process. In particular, suppose that this underlying model involves a continuous sample path for the (logarithmic) price process. The return process may then, under general assumptions, be written in standard stochastic differential equation (sde) form as,

$$dp(t) \;=\; \mu(t)\, dt \;+\; \sigma(t)\, dW(t) \qquad\qquad t \geq 0\,, \tag{1.9}$$

where $\mu(t)$ denotes the drift, $\sigma(t)$ refers to the point-in-time or spot volatility, and $W(t)$ denotes a standard Brownian motion. We will be more specific regarding the additional properties of these processes later on in the chapter. Intuitively, over (infinitesimal) small time intervals, Δ,

$$r(t,\Delta) \;\equiv\; p(t) - p(t-\Delta) \;\simeq\; \mu(t-\Delta)\cdot\Delta \;+\; \sigma(t-\Delta)\,\Delta W(t),$$

where $\Delta W(t) \equiv W(t) - W(t-\Delta) \sim N(\,0\,,\Delta\,)$. Of course, for $\Delta = 1$, and constant drift, $\mu(\tau) \equiv \mu_{t|t-1}$, and volatility, $\sigma(\tau) \equiv \sigma_{t|t-1}$, for $t-1 < \tau \leq t$, this reduces to the discrete-time return decomposition in (1.7) with the additional assumption that z_t is *i.i.d.* N(0,1). Importantly, however, the drift, $\mu(t)$, and instantaneous volatility, $\sigma(t)$, for the continuous-time model in (1.9) need not be constant over the *[t-1,t]* time interval, resulting in the general expression for the one-period return,

$$r(t) \;=\; p(t) - p(t-1) \;=\; \int_{t-1}^{t} \mu(s)\, ds \;+\; \int_{t-1}^{t} \sigma(s)\, dW(s)\,. \tag{1.10}$$

The semblance between this representation and the previous one-period return for the discrete-time model in (1.7) is clear. The conditional mean and variance processes in the discrete

formulation are replaced by the corresponding integrated (averaged) realizations of the (potentially stochastically time-varying) mean and variance process over the following period, with the return innovations driven by the continuously evolving standard Brownian motion. For full generality, the above continuous-time model can be extended with a jump process allowing for discontinuities in the price path, as discussed further in Section 4.

Intuitively, the volatility for the continuous-time process in (1.9) over *[t-1,t]* is intimately related to the evolution of the diffusive coefficient, $\sigma(t)$, which is also known as the spot volatility. In fact, given the i.i.d. nature of the return innovations governed by the Brownian motion process, the return variation should be related to the cumulative (integrated) spot variance. It is, indeed, possible to formalize this intuition: the conditional return variation is linked closely and - under certain conditions in an ex-post sense - equal to the so-called integrated variance (volatility),

$$IV(t) \equiv \int_{t-1}^{t} \sigma^2(s)\,ds. \tag{1.11}$$

We provide more in-depth discussion and justification for this integrated volatility measure and its relationship to the conditional return distribution in Section 4. It is, however, straightforward to motivate the association through the approximate discrete return process, $r(t,\Delta)$, introduced above. If the variation in the drift is an order of magnitude less than the variation in the volatility over the *[t-1,t]* time interval - which holds empirically over daily or weekly horizons and is consistent with a no-arbitrage condition - it follows, for small (infinitesimal) time intervals, Δ,

$$Var(r(t)\,|\,\mathscr{F}_{t-1}) \simeq E\,[\sum_{j=1}^{1/\Delta} \sigma^2(t-j/\Delta)\cdot\Delta\,|\,\mathscr{F}_{t-1}] \simeq E\,[IV(t)\,|\,\mathscr{F}_{t-1}].$$

Hence, the integrated variance measure corresponds closely to the conditional variance, $\sigma^2_{t|t-1}$, for discretely sampled returns. It represents the realized volatility over the same one-period-ahead forecast horizon, and it simply reflects the cumulative impact of the spot volatility process over the return horizon. In other words, integrated variances are ex-post realizations that are directly comparable to ex-ante volatility forecasts. Moreover, in contrast to the one-period-ahead squared return innovations, which, as discussed in the context of (1.8), are plagued by large idiosyncratic errors, the integrated volatility measure is not distorted by error. As such, it serves as an ideal theoretical ex-post benchmark for assessing the quality of ex-ante volatility forecasts.

To more clearly illustrate these differences between the various volatility concepts, Figure 1.1 graphs the simulations from a continuous-time stochastic volatility process. The simulated model is designed to induce temporal dependencies consistent with the popular, and empirically

successful, discrete-time GARCH(1,1) model dis\cussed in Section 3.[1] The top left panel displays sample path realization of the spot volatility or variance, $\sigma^2(t)$, over the 2,500 "days" in the simulated sample. The top panel on the right shows the corresponding "daily" integrated volatility or variance, *IV(t)*. The two bottom panels show the "optimal" one-step-ahead discrete-time GARCH(1,1) forecasts, $\sigma^2_{t|t-1}$, along with the "daily" squared returns, r_t^2. A number of features in these displays are of interest. First, it is clear that even though the "daily" squared returns generally track the overall level of the volatility in the first two panels, as an unbiased measure should, it is an extremely noisy proxy. Hence, a naive assessment of the quality of the GARCH based forecasts in the third panel based on a comparison with the ex post squared returns in panel four invariable will suggest very poor forecast quality, despite the fact that by construction the GARCH based procedure is the "optimal" discrete-time forecast. We provide a much more detailed discussion of this issue in Section 7 below. Second, the integrated volatility provides a mildly smoothed version of the spot volatility process. Since the simulated series has a very persistent volatility component the differences are minor, but still readily identifiable. Third, the "optimal" discrete-time GARCH forecasts largely appear as smoothed versions of the spot and integrated volatility series. This is natural as forecasts, by construction, should be less variable than the corresponding ex-post realizations. Fourth, it is also transparent, however, that the GARCH based forecasts fail to perfectly capture the nature of the integrated volatility series. The largest spike in volatility (around the 700-750 "day" marks) is systematically underestimated by the GARCH forecasts while the last spike (around the 2300-2350 "day" marks) is exaggerated relative to the actual realizations. This reflects the fact that the volatility is not constant over the "day," and as such the (realized) integrated volatility is not equal to the (optimal) forecast from the discrete-time GARCH model which only utilizes the past "daily" return observations. Instead, there is a genuine random component to the volatility process as it evolves stochastically over the "trading day." As a result, the "daily" return observations do not convey all relevant information and the GARCH model simply cannot produce fully efficient forecasts compared to what is theoretically possible given higher frequency "intraday" data. At the same time, in practice it is not feasible to produce exact real-time measures of the integrated, let alone the spot, volatility, as the processes are latent and we only have a limited and discretely sampled set of return observations available, even for the most liquid asset markets. As such, an important theme addressed in more detail in Sections 4 and 5 below involves the construction of practical measures of ex-post realized volatility that mimic the properties of the integrated volatility series.

1.2 Final Introductory Remarks

This section has introduced some of the basic notation used in our subsequent discussion of the various volatility forecasting procedures and evaluation methods. Our initial account also

---

[1] The simulated continuous-time GARCH diffusion shown in Figure 1.1 is formally defined by $dp(t) = \sigma(t)dW_1(t)$ and $d\sigma^2(t) = 0.035[0.636 - \sigma^2(t)]dt + 0.144\sigma^2(t)dW_2(t)$, where $W_1(t)$ and $W_2(t)$ denote two independent Brownian motions. The same model has previously been analyzed in Andersen and Bollerslev (1998a), Andersen, Bollerslev and Meddahi (2004, 2005), among others.

emphasizes a few conceptual features and practical considerations. First, volatility forecasts and measurements are generally restricted to (non-trivial) discrete-time intervals, even if the underlying process may be thought of as evolving in continuous time. Second, differences between ARCH and stochastic volatility models may be seen as direct consequences of assumptions about the observable information set. Third, it is important to recognize the distinction between ex-ante forecasts and ex-post realizations. Only under simplifying - and unrealistic - assumptions are the two identical. Fourth, standard ex-post measurements of realized volatility are often hampered by large idiosyncratic components. The ideal measure is instead, in cases of general interest, given by the so-called integrated volatility. The relationships among the various concepts are clearly illustrated by the simulations in Figure 1.1.

The rest of the chapter unfolds as follows. Section 2 provides an initial motivating discussion of several practical uses of volatility forecasts. Section 3, 4 and 5 present a variety of alternative procedures for univariate volatility forecasting based on the GARCH, stochastic volatility and realized volatility paradigms, respectively. Section 6 extends the discussion to the multivariate problem of forecasting conditional covariances and correlations, and Section 7 discusses practical volatility forecast evaluation techniques. Section 8 concludes briefly.

2. Uses of Volatility Forecasts

This section surveys how volatility forecasts are used in practical applications along with applications in the academic literature. While the emphasis is on financial applications the discussion is kept at a general level. Thus, we do not yet assume a specific volatility forecasting model. The issues involved in specifying and estimating particular volatility forecasting models will be discussed in subsequent sections.

We will first discuss a number of general statistical forecasting applications where volatility dynamics are important. Then we will go into some detail on various applications in finance. Lastly we will briefly mention some applications in macroeconomics and in other disciplines.

2.1 Generic Forecasting Applications

For concreteness, assume that the future realization of the variable of interest can be written as a decomposition similar to the one already developed in equation (1.7),

$$y_{t+1} = \mu_{t+1|t} + \sigma_{t+1|t} z_{t+1}, \qquad z_{t+1} \sim i.i.d.\ F\ , \tag{2.1}$$

where $\{y_{t+1}\}$ denotes a discrete-time real-valued univariate stochastic process, and F refers to the distribution of the zero-mean, unit-variance innovation, z_{t+1}. This representation is not entirely general as there could be higher-order conditional dependence in the innovations. Such higher-moment dynamics would complicate some of the results, but the qualitative insights would remain the same. Thus, to facilitate the presentation we continue our discussion of the different forecast usages under slightly less than full generality.

2.1.1 Point Forecasting

We begin by defining the forecast loss function which maps the ex-ante forecasts $\hat{y}_{t+1|t}$ and the ex-post realization y_{t+1} into a loss value $L(y_{t+1},\hat{y}_{t+1|t})$, which by assumption increases with the discrepancy between the realization and the forecast. The exact form of the loss function depends, of course, directly on the use of the forecast. However, in many situations the loss function may reasonably be written in the form of an additive error, $e_{t+1} \equiv y_{t+1} - \hat{y}_{t+1}$, as the argument, so that $L(y_{t+1},\hat{y}_{t+1|t}) = L(e_{t+1})$. We will refer to this as the forecast error loss function.

In particular, under the symmetric quadratic forecast error loss function, which is implicitly used in many practical applications, the optimal point forecast is simply the conditional mean of the process, regardless of the shape of the conditional distribution. That is

$$\hat{y}_{t+1|t} \equiv \text{Arg min}_{\hat{y}} \; E[(y_{t+1} - \hat{y})^2 \,|\, \mathscr{F}_t] = \mu_{t+1|t} \,.$$

Volatility forecasting is therefore irrelevant for calculating the optimal point forecast, unless the conditional mean depends directly on the conditional volatility. However, this exception is often the rule in finance, where the expected return generally involves some function of the volatility of market wide risk factors. Of course, as discussed further below, even if the conditional mean does not explicitly depend on the conditional volatility, volatility dynamics are still relevant for assessing the uncertainty of the point forecasts.

In general, when allowing for asymmetric loss functions, the volatility forecast will be a key part of the optimal forecast. Consider for example the asymmetric linear loss function,

$$L(e_{t+1}) = a|e_{t+1}|I(e_{t+1} > 0) + b|e_{t+1}|I(e_{t+1} \le 0), \tag{2.2}$$

where $a, b > 0$, and $I(\cdot)$ denotes the indicator function equal to zero or one depending on the validity of its argument. In this case positive and negative forecast errors have different weights (a and b respectively) and thus different losses. Now the optimal forecast can be shown to be

$$\hat{y}_{t+1|t} = \mu_{t+1|t} + \sigma_{t+1|t} F^{-1}(a/(a+b)) \,, \tag{2.3}$$

which obviously depends on the relative size of a and b. Importantly, the volatility plays a key role even in the absence of conditional mean dynamics. Only if $F^{-1}(a/(a+b)) = 0$ does the optimal forecast equal the conditional mean.

This example is part of a general set of results in Granger (1969) who shows that if the conditional distribution is symmetric (so that $F^{-1}(1/2) = 0$) and if the forecast error loss function is also symmetric (so that $a/(a+b) = 1/2$) but not necessarily quadratic, then the conditional mean is the optimal point forecast.

2.1.2 Interval Forecasting

Constructing accurate interval forecasts around the conditional mean forecast for inflation was a leading application in Engle's (1982) seminal ARCH paper. An interval forecast consists of an upper and lower limit. One version of the interval forecast puts *p/2* probability mass below and above the lower and upper limit respectively. The interval forecast can then be written as

$$\hat{y}_{t+1|t} = \{\mu_{t+1|t} + \sigma_{t+1|t}F^{-1}(p/2)\,,\ \mu_{t+1|t} + \sigma_{t+1|t}F^{-1}(1-p/2)\}\,. \tag{2.4}$$

Notice that the volatility forecast plays a key role again. Note also the direct link between the interval forecast and the optimal point forecast for the asymmetric linear loss function in (2.3).

2.1.3 Probability Forecasting Including Sign Forecasting

A forecaster may care about the variable of interest falling above or below a certain threshold value. As an example, consider a portfolio manager who might be interested in forecasting whether the return on a stock index will be larger than the known risk-free bond return. Another example might be a rating agency forecasting if the value of a firm's assets will end up above or below the value of its liabilities and thus trigger bankruptcy. Yet another example would be a central bank forecasting the probability of inflation – or perhaps an exchange rate – falling outside its target band. In general terms, if the concern is about a variable y_{t+1} ending up above some fixed (known) threshold, c, the loss function may be expressed as

$$L(y_{t+1},\hat{y}_{t+1|t}) = (I(y_{t+1}>c) - \hat{y}_{t+1|t})^2\,. \tag{2.5}$$

Minimizing the expected loss by setting the first derivative equal to zero then readily yields

$$\hat{y}_{t+1|t} = E[I(y_{t+1}>c)\,|\,\mathscr{F}_t] = P(y_{t+1}>c\,|\,\mathscr{F}_t) = 1 - F((c-\mu_{t+1|t})/\sigma_{t+1|t})\,. \tag{2.6}$$

Thus, volatility dynamics are immediately important for these types of probability forecasts, even if the conditional mean is constant and not equal to c; i.e., $c - \mu_{t+1|t} \neq 0$.

The important special case where $c = 0$ is sometimes referred to as sign forecasting. In this situation,

$$\hat{y}_{t+1|t} = 1 - F(-\mu_{t+1|t}/\sigma_{t+1|t})\,. \tag{2.7}$$

Hence, the volatility dynamics will affect the forecast as long as the conditional mean is not zero, or the conditional mean is not directly proportional to the standard deviation.

2.1.4 Density Forecasting

In many applications the entire conditional density of the variable in question is of interest. That is, the forecast takes the form of a probability distribution function

$$\hat{y}_{t+1|t} = f_{t+1|t}(y) \equiv f(y_{t+1} = y \mid \mu_{t+1|t}, \sigma_{t+1|t}) = f(y_{t+1} = y \mid \mathscr{F}_t) \tag{2.8}$$

Of course, the probability density function may itself be time-varying, for example due to time-varying conditional skewness or kurtosis, but as noted earlier for simplicity we rule out these higher order effects here.

Figure 2.1 shows two stylized density forecasts corresponding to a high and low volatility day, respectively. Notice that the mean outcome is identical (and positive) on the two days. However, on the high volatility day the occurrence of a large negative (or large positive) outcome is more likely. Notice also that the probability of a positive outcome (of any size) is smaller on the high volatility day than on the low volatility day. Thus, as discussed in the preceding sections, provided that the level of the volatility is forecastable, the figure indicates some degree of sign predictability, despite the constant mean.

2.2 Financial Applications

The trade-off between risk and expected return, where risk is associated with some notion of price volatility, constitute one of the key concepts in modern finance. As such, measuring and forecasting volatility is arguably among the most important pursuits in empirical asset pricing finance and risk management.

2.2.1 Risk management: Value-at-Risk (VaR) and Expected Shortfall (ES)

Consider a portfolio of returns formed from a vector of N risky assets, R_{t+1}, with corresponding vector of portfolio weights, W_t. The portfolio return is defined as

$$r_{w,t+1} = \sum_{i=1}^{N} w_{i,t} r_{i,t+1} \equiv W_t' R_{t+1}, \tag{2.9}$$

where the w subscript refers to the fact that the portfolio distribution depends on the actual portfolio weights.

Financial risk managers often report the riskiness of the portfolio using the concept of Value-at-Risk (VaR) which is simply the quantile of the conditional portfolio distribution. If we model the portfolio returns directly as a univariate process,

$$r_{w,t+1} = \mu_{w,t+1|t} + \sigma_{w,t+1|t} z_{w,t+1} \qquad z_{w,t+1} \sim i.i.d.\ F_w, \tag{2.10}$$

then the VaR is simply

$$VaR_{t+1|t}^{p} = \mu_{w,t+1|t} + \sigma_{w,t+1|t} F_w^{-1}(p). \tag{2.11}$$

This, of course, corresponds directly to the lower part of the interval forecast previously defined in equation (2.4).

Figure 2.2 shows a typical simulated daily portfolio return time series with dynamic volatility (solid line). The short-dashed line, which tracks the lower range of the return, depicts the true 1-day, 1% VaR corresponding to the simulated portfolio return. Notice that the true VaR varies considerably over time and increases in magnitude during bursts in the portfolio volatility. The relatively sluggish long-dashed line calculates the VaR using the so-called Historical Simulation (HS) technique. This is a very popular approach in practice. Rather than explicitly modeling the volatility dynamics, the HS technique calculates the VaR as an empirical quantile based on a moving window of the most recent *250* or *500* days. The HS VaR in Figure 2.2 is calculated using a *500*-day window. Notice how this HS VaR reacts very sluggishly to changes in the volatility, and generally is too large (in absolute value) when the volatility is low, and more importantly too small (in absolute value) when the volatility is high. Historical simulation thus underestimates the risk when the risk is high. This is clearly not a prudent risk management practice. As such, these systematic errors in the HS VaR clearly highlight the value of explicitly modeling volatility dynamics in financial risk management.

The VaR depicted in Figure 2.2 is a very popular risk-reporting measure in practice, but it obviously only depicts a very specific aspect of the risk; that is with probability p the loss will be at least the VaR. Unfortunately, the VaR measure says nothing about the expected magnitude of the loss on the days the VaR is breached.

Alternatively, the Expected Shortfall (ES) risk measure was designed to provide additional information about the tail of the distribution. It is defined as the expected loss on the days when losses are larger than the VaR. Specifically,

$$ES^{p}_{t+1|t} \equiv E[r_{w,t+1} | r_{w,t+1} < VaR^{p}_{t+1|t}] = \mu_{w,t+1|t} + \sigma_{w,t+1|t} EF^{p}_{w} . \tag{2.12}$$

Again, it is possible to show that if $z_{w,t}$ is *i.i.d.*, the multiplicative factor, EF^{p}_{w}, is constant and depends only on the shape of the distribution, F_w. Thus, the volatility dynamics plays a similar role in the *ES* risk measure as in the VaR in equation (2.11).

The analysis above assumed a univariate portfolio return process specified as a function of the portfolio weights at any given time. Such an approach is useful for risk measurement but is not helpful, for example, for calculating optimal portfolio weights. If active risk management is warranted, say maximizing expected returns subject to a VaR constraint, then a multivariate model is needed. If we assume that each return is modeled separately then the vector of returns can be written as

$$R_{t+1} = \mathbf{M}_{t+1|t} + \Omega^{1/2}_{t+1|t} Z_{t+1} \qquad Z_{t+1} \sim i.i.d.\ F , \tag{2.13}$$

where $\mathbf{M}_{t+1|t}$ and $\Omega_{t+1|t}$ denote the vector of conditional mean returns and the covariance matrix for the returns, respectively, and all of the elements in the vector random process, Z_t, are independent with mean zero and variance one. Consequently, the mean and the variance of the portfolio returns, $W_t' R_{t+1}$, may be expressed as,

$$\mu_{w,t+1|t} = W_t' \mathbf{M}_{t+1|t} \qquad \sigma^2_{w,t+1|t} = W_t' \Omega_{t+1|t} W_t \ . \tag{2.14}$$

In the case of the normal distribution, $Z_{t+1} \sim N(0, I)$, linear combinations of multivariate normal variables are themselves normally distributed, so that $r_{w,t+1} \equiv W_t' R_{t+1} \sim N(\mu_{w,t+1|t}, \sigma^2_{w,t+1|t})$, but this aggregation property does not hold in general for other multivariate distributions. Hence, except in special cases, such as the multivariate normal, the VaR and ES measures are not known in closed form, and will have to be calculated using Monte Carlo simulation.

2.2.2 Covariance Risk: Time-varying Betas and Conditional Sharpe Ratios

The above discussion has focused on measuring the risk of a portfolio from purely statistical considerations. We now turn to a discussion of the more fundamental economic issue of the expected return on an asset given its risk profile. Assuming the absence of arbitrage opportunities a fundamental theorem in finance then proves the existence of a stochastic discount factor, say SDF_{t+1}, which can be used to price any asset, say I, via the conditional expectation

$$E[SDF_{t+1}(1 + r_{i,t+1}) \mid \mathscr{F}_t] \ = \ 1 . \tag{2.15}$$

In particular, the return on the risk free asset, which pays one dollar for sure the next period, must satisfy $1 + r_{f,t} = E[SDF_{t+1} \mid \mathscr{F}_t]^{-1}$. It follows also directly from (2.15) that the expected excess return on any risky asset must be proportional to its covariance with the stochastic discount factor,

$$E[r_{i,t+1} - r_{f,t} \mid \mathscr{F}_t] \ = \ -(1 + r_{f,t}) Cov(SDF_{t+1}, r_{i,t+1} \mid \mathscr{F}_t) . \tag{2.16}$$

Now, assuming that the stochastic discount factor is linearly related to the market return,

$$SDF_{t+1} \ = \ a_t \ - \ b_t(1 + r_{M,t+1}) \tag{2.17}$$

it follows from $E[SDF_{t+1}(1 + r_{M,t+1}) \mid \mathscr{F}_t] = 1$ and $1 + r_{f,t} = E[SDF_{t+1} \mid \mathscr{F}_t]^{-1}$ that

$$a_t \ = \ (1 + r_{f,t})^{-1} \ + \ b_t \mu_{M,t+1|t} \qquad b_t \ = \ (1 + r_{f,t})^{-1} (\mu_{M,t+1|t} - r_{f,t}) / \sigma^2_{M,t+1|t} \ , \tag{2.18}$$

where $\mu_{M,t+1|t} \equiv E[1 + r_{M,t+1} \mid \mathscr{F}_t]$ and $\sigma^2_{M,t+1|t} \equiv Var[r_{M,t+1} \mid \mathscr{F}_t]$. Notice that the dynamics in the

moments of the market return (along with any dynamics in the risk-free rate) render the coefficients in the *SDF* time varying. Also, in parallel to the classic one-period CAPM model of Markowitz (1952) and Sharpe (1964), the *conditional* expected excess returns must satisfy the relation,

$$E[r_{i,t+1} - r_{f,t} \mid \mathscr{F}_t] = \beta_{i,t}(\mu_{M,t+1|t} - r_{f,t}) , \tag{2.19}$$

where the *conditional* "beta" is defined by $\beta_{i,t} \equiv Cov(r_{M,t+1}, r_{i,t+1} \mid \mathscr{F}_t)/\sigma^2_{M,t+1|t}$. Moreover, the expected risk adjusted return, also know as the *conditional* Sharpe ratio, equals

$$SR_t \equiv E[r_{i,t+1} - r_{f,t} \mid \mathscr{F}_t]/Var(r_{i,t+1} \mid \mathscr{F}_t)^{1/2} = Corr(r_{M,t+1}, r_{i,t+1} \mid \mathscr{F}_t)/\sigma_{M,t+1|t}. \tag{2.20}$$

The simple asset pricing framework above illustrates how the expected return (raw and risk adjusted) on various assets will be driven by the mean and volatility dynamics of the overall market return as well as the dynamics of the covariance between the market and the individual assets. Covariance forecasting is thus at least as important as volatility forecasting in the context of financial asset pricing, and we discuss each in subsequent sections.

2.2.3 Asset Allocation with Time-varying Covariances

The above CAPM model imposes a very restrictive structure on the covariance matrix of asset returns. In this section we instead assume a generic dynamic covariance matrix and study the optimization problem of an investor who constructs a portfolio of *N* risky assets by minimizing the portfolio variance subject to achieving a certain target portfolio return, μ_p.

Formally, the investor chooses a vector of portfolio weights, W_t, by solving the quadratic programming problem

$$\min\ W_t' \Omega_{t+1|t} W_t \quad \text{s.t. } W_t' \mathrm{M}_{t+1|t} = \mu_p . \tag{2.21}$$

From the corresponding first order conditions, the resulting portfolio weights for the risky assets satisfy,

$$W_t^* = \frac{\Omega_{t+1|t}^{-1} \mathrm{M}_{t+1|t}}{\mathrm{M}_{t+1|t}' \Omega_{t+1|t}^{-1} \mathrm{M}_{t+1|t}} \mu_p , \tag{2.22}$$

with the optimal portfolio weight for the risk-free asset given by

$$w_{f,t}^* = 1 - \sum_{i=1}^{N} w_{i,t}^* . \tag{2.23}$$

Moreover, from (2.21) the portfolio Sharpe ratio equals,

$$SR_t = \mu_p \,/\, \sqrt{W_t^{*\prime}\,\Omega_{t+1|t}\,W_t^*}\ . \tag{2.24}$$

Just as in the CAPM pricing model discussed above, both volatility and covariance dynamics are clearly important for asset allocation. Notice also that even if we rule out exploitable conditional mean dynamics, the optimal portfolio weights would still be time-varying from the second moment dynamics alone.

2.2.4 Option Valuation with Dynamic Volatility

The above tools are useful for the analysis of primitive securities with linear payoffs such as stocks, bonds, foreign exchange and futures contracts. Consider now instead a European call option which gives the owner the right but not the obligation to buy the underlying asset (say a stock or currency) on a future date, T, at a strike price, K. The option to exercise creates a nonlinear payoff which in turn requires a special set of tools for pricing and risk management.

In the Black-Scholes-Merton (*BSM*) option pricing model the returns are assumed to be normally distributed with constant volatility, σ, along with the possibility of (costless) continuous trading and a constant risk free rate, r_f. In this situation, the call price of an option equals

$$c_t = BSM(s_t, \sigma^2, K, r_f, T) = s_t\Phi(d) - K\exp(-r_f T)\,\Phi(d - \sigma\sqrt{T}), \tag{2.25}$$

where s_t denotes the current price of the asset, $d = (\ln(s_t/K) + T(r_f + \sigma^2/2))/(\sigma\sqrt{T})$, and $\Phi(\cdot)$ refers to the cumulative normal distribution function.

Meanwhile, the constant volatility assumption in *BSM* causes systematic pricing errors when comparing the theoretical prices with actual market prices. One manifestation of this is the famous volatility-smiles which indicate systematic underpricing by the *BSM* model for in- or out-of-the-money options. The direction of these deviations, however, are readily explained by the presence of stochastic volatility, which creates fatter tails than the normal distribution, in turn increasing the value of in- and out-of-the-money options relative to the constant-volatility *BSM* model.

In response to this, Hull and White (1987) explicitly allow for an independent stochastic volatility factor in the process for the underlying asset return. Assuming that this additional volatility risk factor is not priced in equilibrium, the Hull-White call option price simply equals the expected *BSM* price, where the expectation is taken over the future integrated volatility. More specifically, defining the integrated volatility as the integral of the spot volatility during the remaining life of the option,

$$IV(T,t) = \int_t^T \sigma^2(u)du \ ,$$

where $IV(T,t) = IV(T) + IV(T-1) + ... + IV(t+1)$ generalizes the integrated variance concept from equation (1.11) to a multi-period horizon in straightforward fashion. The Hull-White option valuation formula may then be succinctly written as

$$C_t = E[BSM(IV(T,t)) \mid \mathscr{F}_t] \ . \tag{2.26}$$

In discrete time, the integrated volatility may be approximated by the sum of the corresponding one-period conditional variances,

$$IV(T,t) \approx \sum_{\tau=t}^{T-1} \sigma^2_{\tau+1|\tau} \ .$$

Several so-called realized volatility measures have also recently been proposed in the literature for (ex-post) approximating the integrated volatility. We will return to a much more detailed discussion of these measures in Sections 4 and 5 below.

Another related complication that arises in the pricing of equity options, in particular, stems from the apparent negative correlation between the returns and the volatility. This so-called leverage effect, as discussed further below, induces negative skewness in the return distribution and causes systematic asymmetric pricing errors in the *BSM* model.

Assuming a mean-reverting stochastic volatility process, Heston (1993) first developed an option pricing formula where the innovations to the returns and the volatility are correlated, and where the volatility risk is priced by the market. In contrast to the *BSM* setting, where an option can be hedged using a dynamically rebalanced stock and bond portfolio alone, in the Heston model an additional position must be taken in another option in order to hedge the volatility risk.

Relying on Heston's formulation, Fouque, Papanicolaou and Sircar (2000) show that the price may conveniently be expressed as

$$C_t = E[BSM(\xi_{t,T} s_t, (1-\rho^2)IV(T,t)) \mid \mathscr{F}_t] \ , \tag{2.27}$$

where ρ refers to the (instantaneous) correlation between the returns and the volatility, and $\xi_{t,T}$ denotes a stochastic scaling factor determined by the volatility risk premium, with the property that $E[\xi_{t,T} \mid \mathscr{F}_t] = 1$. Importantly, however, the integrated volatility remains the leading term as in the Hull-White valuation formula.

2.3 Volatility Forecasting in Fields Outside Finance

Although volatility modeling and forecasting has proved to be extremely useful in finance, the motivation behind Engle's (1982) original ARCH model was to provide a tool for measuring the dynamics of inflation uncertainty. Tools for modeling volatility dynamics have been applied in many other areas of economics and indeed in other areas of the social sciences, the natural sciences and even medicine. In the following we list a few recent papers in various fields showcasing the breath of current applications of volatility modeling and forecasting. It is by no means an exhaustive list but these papers can be consulted for further references.

Related to Engle's original work, the modeling of inflation uncertainty and its relationship with labor market variables has recently been studied by Rich and Tracy (2004). They corroborate earlier findings of an inverse relationship between desired labor contract durations and the level of inflation uncertainty. Analyzing the inflation and output forecasts from the Survey of Professional Forecasters, Giordani and Soderlind (2003) find that while each forecaster on average tends to underestimate uncertainty, the disagreement between forecasters provides a reasonable proxy for inflation and output uncertainty. The measurement of uncertainty also plays a crucial role in many microeconomic models. Meghir and Pistaferri (2004), for instance, estimate the conditional variance of income shocks at the micro level and find strong evidence of temporal variance dynamics.

Lastrapes (1989) first analyzed the relationship between exchange rate volatility and U.S. monetary policy. In a more recent study, Ruge-Murcia (2004) developed a model of a central bank with asymmetric preferences for unemployment above versus below the natural rate. The model implies an inflation bias proportional to the conditional variance of unemployment. Empirically, the conditional variance of unemployment is found to be positively related to the rate of inflation. In another central banking application, Tse and Yip (2003) use volatility models to study the effect on changes in the Hong Kong currency board on interbank market rates.

Volatility modeling and forecasting methods have also found several interesting uses in agricultural economics. Ramirez and Fadiga (2003), for instance, find evidence of asymmetric volatility patterns in U.S. soybean, sorghum and wheat prices. Building on the earlier volatility spill-over models used in analyzing international financial market linkages in the papers by Engle, Ito and Lin (1990) and King, Sentana and Wadhwani (1994), Buguk, Hudson and Hanson (2003) have recently used similar methods in documenting strong price volatility spillovers in the supply-chain of fish production. The volatility in feeding material prices (e.g. soybeans) affects the volatility of fish feed prices which in turn affect fish farm price volatility and finally wholesale price volatility. Also, Barrett (1999) uses a GARCH model to study the effect of real exchange rate depreciations on stochastic producer prices in low-income agriculture.

The recent deregulation in the utilities sector has also prompted many new applications of volatility modeling of gas and power prices. Shawky, Marathe and Barret (2003) use dynamic volatility models to determine the minimum variance hedge ratios for electricity futures. Linn and Zhu (2004) study the effect of natural gas storage report announcements on intraday volatility patterns in gas prices. They also find evidence of strong intraday patterns in natural gas price volatility. Batlle and Barquin (2004) use a multivariate GARCH model to simulate gas and

oil price paths, which in turn are shown to be useful for risk management in the wholesale electricity market.

In a related context, Taylor and Buizza (2003) use weather forecast uncertainty to model electricity demand uncertainty. The variability of wind measurements is found to be forecastable using GARCH models in Dripps and Dunsmuir (2003), while temperature forecasting with seasonal volatility dynamics is explored in Campbell and Diebold (2005). Marinova and McAleer (2003) model volatility dynamics in ecological patents.

In political science, Maestas and Preuhs (2000) suggest modeling political volatility broadly defined as periods of rapid and extreme change in political processes, while Gronke and Brehm (2002) use ARCH models to assess the dynamics of volatility in presidential approval ratings.

Volatility forecasting has recently found applications even in medicine. Ewing, Piette and Payne (2003) forecast time varying volatility in medical net discount rates which are in turn used to determine the present value of future medical costs. Also, Johnson, Elashoff and Harkema (2003) use a heteroskedastic time series process to model neuromuscular activation patterns in patients with spinal cord injuries, while Martin-Guerrero et al. (2003) use a dynamic volatility model to help determine the optimal EPO dosage for patients with secondary anemia.

2.4 Further Reading

Point forecasting under general loss functions when allowing for dynamic volatility has been analyzed by Christoffersen and Diebold (1996, 1997). Patton and Timmermann (2004) have recently shown that under general specifications of the loss function, the optimal forecast error will have a conditional expected value that is a function of the conditional variance of the underlying process. Methods for incorporating time-varying volatility into interval forecasts are suggested in Granger, White and Kamstra (1989). Financial applications of probability forecasting techniques are considered in Christoffersen and Diebold (2003).

Financial risk management using dynamic volatility models is surveyed in Christoffersen (2003) and Jorion (2000). Berkowitz and O'Brien (2002), Pritsker (2001), and Barone-Adesi, Giannopoulos and Vosper (1999) explicitly document the value added from incorporating volatility dynamics into daily financial risk management systems.

Volatility forecasting at horizons beyond a few weeks is found to be difficult by West and Cho (1995) and Christoffersen and Diebold (2000). However Brandt and Jones (2002) show that using intraday information improves the longer horizon forecasts considerably. Guidolin and Timmermann (2005a) uncover VaR dynamics at horizons of up to two years. Campbell (1987, 2003), Shanken (1990), Aït-Sahalia and Brandt (2001), Harvey (2001), Lettau and Ludvigsson (2003) and Marquering and Verbeek (2004) find that interest rate spreads and financial ratios help predict volatility at longer horizons.

A general framework for conditional asset pricing allowing for time-varying betas is laid out in Cochrane (2001). Market timing arising from time-varying Sharpe ratios is analyzed in Whitelaw (1997), while volatility timing has been explicitly explored by Johannes, Polson and Stroud (2004). The relationship between time-varying volatility and return has been studied in Engle, Lilien and Robbins (1987), French, Schwert and Stambaugh (1987), Bollerslev, Engle and Wooldridge (1988), Bollerslev, Chou and Kroner (1992), Glosten, Jagannathan and Runkle (1993), among many others.

The value of modeling volatility dynamics for asset allocation in a single-period setting have been highlighted in the series of papers by Fleming, Kirby and Oestdiek (2001, 2003), with multi-period extensions considered by Wang (2004). The general topic of asset allocation under predictable returns is surveyed in Brandt (2004). Brandt (1999) and Aït-Sahalia and Brandt (2001) suggest portfolio allocation methods which do not require the specification of conditional moment dynamics.

The literature on option valuation allowing for volatility dynamics is very large and active. In addition to some of the key theoretical contributions mentioned above, noteworthy empirical studies based on continuous-time methods include Bakshi, Cao and Chen (1997), Bates (1996), Chernov and Ghysels (2000), Eraker (2004), Melino and Turnbull (1990), and Pan (2002). Recent discrete-time applications, building on the theoretical work of Duan (1995) and Heston (1993), can be found in Christoffersen and Jacobs (2004), and Heston and Nandi (2000).

3. GARCH Volatility

The current interest in volatility modeling and forecasting was spurred by Engle's (1982) path breaking ARCH paper, which set out the basic idea of modeling and forecasting volatility as a time-varying function of current information. The GARCH class of models, of which the GARCH(1,1) remains the workhorse, were subsequently introduced by Bollerslev (1986), and also discussed independently by Taylor (1986). These models, including their use in volatility forecasting, have been extensively surveyed elsewhere and we will not attempt yet another exhaustive survey here. Instead we will try to highlight some of the key features of the models which help explain their dominant role in practical empirical applications. We will concentrate on univariate formulations in this section, with the extension to multivariate GARCH-based covariance and correlation forecasting discussed in Section 6.

3.1 Rolling Regressions and RiskMetrics

Rolling sample windows arguably provides the simplest way of incorporating actual data into the estimation of time-varying volatilities, or variances. In particular, consider the rolling sample variance based on the p most recent observations as of time t,

$$\hat{\sigma}_t^2 = p^{-1}\sum_{i=0}^{p-1}(y_{t-i} - \hat{\mu})^2 \equiv p^{-1}\sum_{i=0}^{p-1}\hat{\varepsilon}_{t-i}^2. \tag{3.1}$$

Interpreting $\hat{\sigma}_t^2$ as an estimate of the current variance of y_t, the value of p directly determines the variance-bias tradeoff of the estimator, with larger values of p reducing the variance but increasing the bias. For instance, in the empirical finance literature, it is quite common to rely on rolling samples of five-years of monthly data, corresponding to a value of $p=60$, in estimating time varying-variances, covariances, and CAPM betas.

Instead of weighting each of the most recent p observations the same, the bias of the estimator may be reduced by assigning more weights to the most recent observations. An exponentially weighted moving average filter is commonly applied in doing so,

$$\hat{\sigma}_t^2 = \gamma(y_t - \hat{\mu})^2 + (1 - \gamma)\hat{\sigma}_{t-1}^2 \equiv \gamma \sum_{i=1}^{\infty} (1 - \gamma)^{i-1} \hat{\varepsilon}_{t-i}^2. \tag{3.2}$$

In practice, the sum will, of course, have to be truncated at $I = t-1$. This is typically done by equating the pre-sample values to zero, and adjusting the finite sum by the corresponding multiplication factor $1/[1-(1-\gamma)^t]$. Of course, for large values of t and $(1-\gamma)<1$, the effect of this truncation is inconsequential. This approach is exemplified by RiskMetrics (J.P. Morgan, 1996), which rely on a value of $\gamma = 0.06$ and $\mu \equiv 0$ in their construction of daily (monthly) volatility measures for wide range of different financial rates of returns.

Although it is possible to write down explicit models for y_t which would justify the rolling window approach and the exponential weighted moving average as the optimal estimators for the time-varying variances in the models, the expressions in (3.1) and (3.2) are more appropriately interpreted as data-driven filters. In this regard, the theoretical properties of both filters as methods for extracting consistent estimates of the current volatility as the sampling frequencies of the underlying observations increases over fixed-length time intervals - or what is commonly referred to as continuous record, or fill-in, asymptotics - has been extensively studied in a series of influential papers by Dan Nelson (these papers are collected in the edited volume of readings by Rossi, 1996).

It is difficult to contemplate optimal volatility forecasting without the notion of a model, or data generating process. Of course, density or VaR forecasting, as discussed in Section 2, is even more problematic. Nonetheless, the filters described above are often used in place of more formal model building procedures in the construction of h-period-ahead volatility forecasts by simply equating the future volatility of interest with the current filtered estimate,

$$Var(y_{t+h} \mid \mathscr{F}_t) \equiv \sigma_{t+h|t}^2 \approx \hat{\sigma}_t^2. \tag{3.3}$$

In the context of forecasting the variance of multi-period returns, assuming that the corresponding one-period returns are serially uncorrelated so that the forecast equals the sum of the successive one-period variance forecasts, it follows then directly that

$$Var(y_{t+k} + y_{t+k-1} + \ldots + y_{t+1} \mid \mathscr{F}_t) \;\equiv\; \sigma^2_{t:t+k|t} \;\approx\; k\,\hat{\sigma}^2_t . \tag{3.4}$$

Hence, the multi-period return volatility scales with the forecast horizon, *k*. Although this approach is used quite frequently by finance practitioners it has, as discussed further below, a number of counterfactual implications. In contrast, the GARCH(1,1) model, to which we now turn, provides empirically realistic mean-reverting volatility forecasts within a coherent and internally consistent, yet simple, modeling framework.

3.2 GARCH(1,1)

In order to define the GARCH class of models, consider the decomposition of y_t into the one-step-ahead conditional mean, $\mu_{t|t-1} \equiv E(y_t \mid \mathscr{F}_{t-1})$, and variance, $\sigma^2_{t|t-1} \equiv Var(y_t \mid \mathscr{F}_{t-1})$, in parallel to the expression in equation (1.7) above,

$$y_t \;=\; \mu_{t|t-1} \;+\; \sigma_{t|t-1} z_t \qquad z_t \sim i.i.d. \quad E(z_t) = 0 \quad Var(z_t) = 1 . \tag{3.5}$$

The GARCH(1,1) model for the conditional variance is then defined by the recursive relationship,

$$\sigma^2_{t|t-1} \;=\; \omega \;+\; \alpha\,\varepsilon^2_{t-1} \;+\; \beta\,\sigma^2_{t-1|t-2}, \tag{3.6}$$

where $\varepsilon_t \equiv \sigma_{t|t-1} z_t$, and the parameters are restricted to be non-negative, $\omega > 0, \alpha \geq 0, \beta \geq 0$, in order to ensure that the conditional variance remains positive for all realizations of the z_t process. The model is readily extended to higher order GARCH(p,q) models by including additional lagged squared innovations and/or conditional variances on the right-hand-side of the equation.

By recursive substitution, the GARCH(1,1) model may alternatively be expressed as an ARCH(∞) model,

$$\sigma^2_{t|t-1} \;=\; \omega(1 - \beta)^{-1} \;+\; \alpha \sum_{i=1}^{\infty} \beta^{i-1} \varepsilon^2_{t-i} . \tag{3.7}$$

This obviously reduces to the exponentially weighted moving average filter in (3.2) for $\omega = 0$, $\alpha = \gamma$, and $\beta = 1-\gamma$. The corresponding GARCH model in which $\alpha + \beta = 1$ is also sometimes referred to as an Integrated GARCH, or IGARCH(1,1) model. Importantly, however, what sets the GARCH(1,1) model, and more generally the ARCH class of models, apart from the filters discussed above is the notion of a data generating process embedded in the distributional assumptions for z_t. This means that the construction of optimal variance forecasts is a well-posed question within the context of the model.

In particular, it follows directly from the formulation of the model that the optimal, in a mean-square-error sense, one-step ahead variance forecasts equals $\sigma^2_{t+1|t}$. Corresponding expressions for the longer run forecasts, $\sigma^2_{t+h|t}$ for $h>1$, are also easily constructed by recursive procedures. To facilitate the presentation, assume that the conditional mean is constant and equal to zero, or $\mu_{t|t-1} = 0$, and that $\alpha+\beta < 1$ so that the unconditional variance of the process exists,

$$\sigma^2 = \omega(1 - \alpha - \beta)^{-1}. \tag{3.8}$$

The h-step ahead forecast is then readily expressed as

$$\sigma^2_{t+h|t} = \sigma^2 + (\alpha + \beta)^{h-1}(\sigma^2_{t+1|t} - \sigma^2), \tag{3.9}$$

showing that the forecasts revert to the long-run unconditional variance at an exponential rate dictated by the value of $\alpha+\beta$.

Moreover, with serially uncorrelated returns, so that the conditional variance of the sum is equal to the sum of the conditional variances, the optimal forecast for the variance of the k-period return may be expressed as,

$$\sigma^2_{t:t+k|t} = k\sigma^2 + (\sigma^2_{t+1|t} - \sigma^2)(1 - (\alpha + \beta)^k)(1 - \alpha - \beta)^{-1}. \tag{3.10}$$

Thus, the longer the forecast horizon (the higher the value of k), the less variable will be the forecast per unit time-interval. That is, the term-structure-of-variance, or $k^{-1}\sigma^2_{t:t+k|t}$, flattens with k.

To illustrate, consider Figure 3.1. The left-hand panel plots the unconditional distribution of $\sigma^2_{t+1|t}$ for the same GARCH(1,1) model depicted in Figure 1.1. The mean of the distribution equals $\sigma^2 = 0.020(1 - 0.085 - 0.881)^{-1} \approx 0.588$, but there is obviously considerable variation around that value, with a much longer tail to the right. The panel on the right gives the corresponding term-structure for $k = 1, 2, ..., 250$, and $\sigma^2_{t+1|t}$ equal to the mean, five, and ninety-five percentiles in the unconditional distribution. The slope of the volatility-term-structure clearly flattens with the horizon. The figure also illustrates that the convergence to the long-run unconditional variance occurs much slower for a given percentage deviation of $\sigma^2_{t+1|t}$ above the median than for the same percentage deviation below the median.

To further illustrate the dynamics of the volatility-term structure, Figure 3.2 graphs $k^{-1}\sigma^2_{t:t+k|t}$ for $k = 1, 5, 22$ and 66, corresponding to daily, weekly, monthly and quarterly forecast horizons, for the same $t = 1, 2,..., 2,500$ GARCH(1,1) simulation sample depicted in Figure 1.1. Comparing

the four different panels, the volatility-of the-volatility clearly diminishes with the forecast horizon.

It is also informative to compare and contrast the optimal GARCH(1,1) volatility forecasts to the common empirical practice of horizon volatility scaling by *k*. In this regard, it follows directly from the expressions in (3.9) and (3.10) that

$$E(k\sigma^2_{t+1|t}) = k\sigma^2 = E(\sigma^2_{t:t+k|t}),$$

so that the level of the scaled volatility forecasts will be right on average. However, comparing the variance of the scaled *k*-period forecasts to the variance of the optimal forecast,

$$Var(k\sigma^2_{t+1|t}) = k^2 Var(\sigma^2_{t+1|t}) >$$

$$(1-(\alpha+\beta)^k)^2(1-\alpha-\beta)^{-2} Var(\sigma^2_{t+1|t}) = Var(\sigma^2_{t:t+k|t}),$$

it is obvious that by not accounting for the mean-reversion in the volatility, the scaled forecasts exaggerate the volatility-of-the-volatility relative to the true predictable variation. On tranquil days the scaled forecasts underestimate the true risk, while the risk is inflated on volatile days. Obviously not a very prudent risk management procedure.

This tendency for the horizon scaled forecasts to exhibit excessive variability is also directly evident from the term structure plots in Figure 3.2. Consider the optimal *k*-period ahead variance forecasts defined by *k* times the $k^{-1}\sigma^2_{t:t+k|t}$ series depicted in the last three panels. Contrasting these correct multi-step forecasts with their scaled counterparts defined by *k* times the $\sigma^2_{t+1|t}$ series in the first panel, it is obvious, that although both forecasts will be centered around the right unconditional value of $k\sigma^2$, the horizon scaled forecasts will result in too large "day-to-day" fluctuations. This is especially true for the longer run "monthly" (*k = 22*) and "quarterly" *(k = 66)* forecasts in the last two panels.

3.3 Asymmetries and "Leverage" Effects

The basic GARCH model discussed in the previous section assumes that positive and negative shocks of the same absolute magnitude will have the identical influence on the future conditional variances. In contrast, the volatility of aggregate equity index return, in particular, has been shown to respond asymmetrically to past negative and positive return shocks, with negative returns resulting in larger future volatilities. This asymmetry is generally referred to as a "leverage" effect, although it is now widely agreed that financial leverage alone cannot explain the magnitude of the effect, let alone the less pronounced asymmetry observed for individual equity returns. Alternatively, the asymmetry has also been attributed to a "volatility feedback" effect, whereby heightened volatility requires an increase in the future expected returns to

compensate for the increased risk, in turn necessitating a drop in the current price to go along with the initial increase in the volatility. Regardless of the underlying economic explanation for the phenomenon, the three most commonly used GARCH formulations for describing this type of asymmetry are the GJR or Threshold GARCH (TGARCH) models of Glosten, Jagannathan and Runkle (1993) and Zakoïan (1994), the Asymmetric GARCH (AGARCH) model of Engle and Ng (1993), and the Exponential GARCH (EGARCH) model of Nelson (1991).

The conditional variance in the GJR(1,1), or TGARCH(1,1), model simply augments the standard GARCH(1,1) formulation with an additional ARCH term conditional on the sign of the past innovation,

$$\sigma^2_{t|t-1} = \omega + \alpha\varepsilon^2_{t-1} + \gamma\varepsilon^2_{t-1}I(\varepsilon_{t-1}<0) + \beta\sigma^2_{t-1|t-2}, \tag{3.11}$$

where I(·) denotes the indicator function. It is immediately obvious that for $\gamma > 0$, past negative return shocks will have a larger impact on the future conditional variances. Mechanically, the calculation of multi-period variance forecast works exactly as for the standard symmetric GARCH model. In particular, assuming that $P(z_t \equiv \sigma^{-1}_{t|t-1}\varepsilon_t < 0) = 0.5$, it follows readily that

$$\sigma^2_{t+h|t} = \sigma^2 + (\alpha + 0.5\gamma + \beta)^{h-1}(\sigma^2_{t+1|t} - \sigma^2), \tag{3.12}$$

where the long-run, or unconditional variance, now equals,

$$\sigma^2 = \omega(1 - \alpha - 0.5\gamma - \beta)^{-1}. \tag{3.13}$$

Although the forecasting formula looks almost identical to the one for the GARCH(1,1) model in equation (3.9), the inclusion of the asymmetric term may materially affect the forecasts by importantly altering the value of the current conditional variance, $\sigma^2_{t+1|t}$.

The news impact curve, defined by the functional relationship between $\sigma^2_{t|t-1}$ and ε_{t-1} holding all other variables constant, provides a simple way of characterizing the influence of the most recent shock on next periods conditional variance. In the standard GARCH model this curve is obviously quadratic around $\varepsilon_{t-1} = 0$, while the GJR model with $\gamma > 0$ has steeper slopes for negative values of ε_{t-1}. In contrast, the Asymmetric GARCH, or AGARCH(1,1), model,

$$\sigma^2_{t|t-1} = \omega + \alpha(\varepsilon_{t-1} - \gamma)^2 + \beta\sigma^2_{t-1|t-2}, \tag{3.14}$$

shifts the center of the news impact curve from zero to γ, affording an alternative way of capturing asymmetric effects. The GJR and AGARCH model may also be combined to achieve even more flexible parametric formulations.

Instead of directly parameterizing the conditional variance, the EGARCH model is formulated in terms of the logarithm of the conditional variance, as in the EGARCH(1,1) model,

$$\log(\sigma^2_{t|t-1}) = \omega + \alpha(|z_{t-1}| - E(|z_{t-1}|)) + \gamma z_{t-1} + \beta \log(\sigma^2_{t-1|t-2}), \tag{3.15}$$

where as previously defined, $z_t \equiv \sigma^{-1}_{t|t-1}\varepsilon_t$. As for the GARCH model, the EGARCH model is readily extended to higher order models by including additional lags on the right-hand-side. The parameterization in terms of logarithms has the obvious advantage of avoiding non-negativity constraints on the parameters, as the variance implied by the exponentiated logarithmic variance from the model is guaranteed to be positive. As in the GJR and AGARCH models above, values of $\gamma > 0$ in the EGARCH model directly captures the asymmetric response, or "leverage" effect. Meanwhile, because of the non-differentiability with respect to z_{t-1} at zero, the EGARCH model is often somewhat more difficult to estimate and analyze numerically. From a forecasting perspective, the recursions defined by the EGARCH equation (3.15) readily deliver the optimal - in a mean-square-error sense - forecast for the future logarithmic conditional variances, $E(\log(\sigma^2_{t+h}) \mid \mathscr{F}_t)$. However, in most applications the interest centers on point forecasts for σ^2_{t+h}, as opposed to $\log(\sigma^2_{t+h})$. Unfortunately, the transformation of the $E(\log(\sigma^2_{t+h}) \mid \mathscr{F}_t)$ forecasts to $E(\sigma^2_{t+h} \mid \mathscr{F}_t)$ generally depends on the entire h-step ahead forecast distribution, $f(y_{t+h} \mid \mathscr{F}_t)$. As discussed further in Section 3.6 below, this distribution is generally not available in closed-form, but it may be approximated by Monte Carlo simulations from the convolution of the corresponding h one-step-ahead predictive distributions implied by the z_t innovation process using numerical techniques. In contrast, the expression for $\sigma^2_{t+h|t}$ in equation (3.12) for the GJR or TGARCH model is straightforward to implement, and only depends upon the assumption that $P(z_t<0) = 0.5$.

3.4 Long-Memory and Component Structures

The GARCH, TGARCH, AGARCH, and EGARCH models discussed in the previous sections all imply that shocks to the volatility decay at an exponential rate. To illustrate, consider the GARCH(1,1) model. It follows readily from equation (3.9) that the impulse effect of a time-t shock on the forecast of the variance h period into the future is given by $\partial\sigma^2_{t+h|t}/\partial\varepsilon^2_t = \alpha(\alpha + \beta)^{h-1}$, or more generally

$$\partial\sigma^2_{t+h|t}/\partial\varepsilon^2_t = \kappa\,\delta^h, \tag{3.16}$$

where $0 < \delta < 1$. This exponential decay typically works well when forecasting over short horizons. However, numerous studies, including Ding, Granger and Engle (1993) and Andersen and Bollerslev (1997), have argued that the autocorrelations of squared and absolute returns decay at a much slower hyperbolic rate over longer lags. In the context of volatility forecasting

using GARCH models parameterized in terms of ε_t^2, this suggests that better long term forecasts may be obtained by formulating the conditional variance in such a way that the impulse effect behaves as,

$$\partial\sigma^2_{t+h|t}/\partial\varepsilon_t^2 \;\approx\; \kappa\, h^{\delta}, \tag{3.17}$$

for large values of h, where again $0 < \delta < 1$. Several competing long-memory, or fractionally integrated, GARCH type models have been suggested in the literature to achieve this goal.

In the Fractionally Integrated FIGARCH(1,d,1) model proposed by Baillie, Bollerslev and Mikkelsen (1996) the conditional variance is defined by,

$$\sigma^2_{t|t-1} \;=\; \omega \;+\; \beta\sigma^2_{t-1|t-2} \;+\; [1 - \beta L - (1 - \alpha L - \beta L)(1 - L)^d]\varepsilon_t^2 \tag{3.18}$$

For $d = 0$ the model reduces to the standard GARCH(1,1) model, but for values of $0 < d < 1$ shocks to the point volatility forecasts from the model will decay at a slow hyperbolic rate. The actual forecasts are most easily constructed by recursive substitution in,

$$\sigma^2_{t+h|t+h-1} \;=\; \omega(1 - \beta)^{-1} \;+\; \lambda(L)\sigma^2_{t+h-1|t+h-2}, \tag{3.19}$$

with $\sigma^2_{t+h|t+h-1} \equiv \varepsilon_t^2$ for $h<0$, and the coefficients in $\lambda(L) \equiv 1 - (1 - \beta L)^{-1}(1 - \alpha L - \beta L)(1 - L)^d$ calculated from the recursions,

$$\lambda_1 = \alpha + d \qquad \lambda_j = \beta\lambda_{j-1} + [(j - 1 - d)j^{-1} - (\alpha + \beta)]\,\delta_{j-1} \quad j = 2, 3, \ldots$$

where $\delta_j \equiv \delta_{j-1}(j - 1 - d)j^{-1}$ refer to the coefficients in the Maclaurin series expansion of the fractional differencing operator, $(1 - L)^d$. Higher order FIGARCH models, or volatility forecast filters, may be defined in an analogous fashion. Asymmetries are also easily introduced into the recursions by allowing for separate influences of past positive and negative innovations as in the GJR or TGARCH model. Fractional Integrated EGARCH, or FIEGARCH, models may be similarly defined by parameterizing the logarithmic conditional variance as a fractionally integrated distributed lag of past values.

An alternative, and often simpler, approach for capturing longer-run dependencies involves the use of component type structures. Granger (1980) first showed that the superposition of an infinite number of stationary AR(1) processes may result in a true long-memory process. In fact, there is a long history in statistics and time series econometrics for approximating long-memory by the sum of a few individually short-memory components. This same idea has successfully been used in the context of volatility modeling by Engle and Lee (1999) among others.

In order to motivate the Component GARCH model of Engle and Lee (1999), rewrite the

standard GARCH(1,1) model in (3.6) as,

$$(\sigma^2_{t|t-1} - \sigma^2) = \alpha(\varepsilon^2_{t-1} - \sigma^2) + \beta(\sigma^2_{t-1|t-2} - \sigma^2), \tag{3.20}$$

where it is assumed that $\alpha + \beta < 1$, so that the model is covariance stationary and the long term forecasts converge to the long-run, or unconditional, variance $\sigma^2 = \omega(1 - \alpha - \beta)^{-1}$. The component model then extends the basic GARCH model by explicitly allowing the long-term level to be time-varying,

$$(\sigma^2_{t|t-1} - \zeta^2_t) = \alpha(\varepsilon^2_{t-1} - \zeta^2_t) + \beta(\sigma^2_{t-1|t-2} - \zeta^2_t), \tag{3.21}$$

with ζ^2_t parameterized by the separate equation,

$$\zeta^2_t = \omega + \rho\zeta^2_{t-1} + \phi(\varepsilon^2_{t-1} - \sigma^2_{t-1|t-2}). \tag{3.22}$$

Hence, the transitory dynamics is governed by $\alpha + \beta$, while the long-run dependencies are described by $\rho > 0$. It is possible to show that for the model to be covariance stationary, and the unconditional variance to exist, the parameters must satisfy $(\alpha + \beta)(1 - \rho) + \rho < 1$. Also, substituting the latter equation into the first, the model may be expressed as the restricted GARCH(2,2) model,

$$\sigma^2_{t|t-1} = \omega(1 - \alpha - \beta) + (\alpha + \phi)\varepsilon^2_{t-1} - (\phi(\alpha + \beta) + \rho\alpha)\varepsilon^2_{t-2}$$

$$+ (\rho + \beta + \phi)\sigma^2_{t-1|t-2} + (\phi(\alpha + \beta) - \rho\beta)\sigma^2_{t-2|t-3}.$$

As for the GARCH(1,1) model, volatility shocks therefore eventually dissipate at the exponential rate in equation (3.15). However, for intermediate forecast horizons and values of ρ close to unity, the volatility forecasts from the component GARCH model will display approximate long-memory.

To illustrate, consider Figure 3.3 which graphs the volatility impulse response function, $\partial\sigma^2_{t+h|t}/\partial\varepsilon^2_t$, *h = 1, 2, ..., 250*, for the RiskMetrics forecasts, the standard GARCH(1,1) model in (3.6), the FIGARCH(1,d,1) model in (3.18), and the component GARCH model defined by (3.21) and (3.22). The parameters for the different GARCH models are calibrated to match the volatilities depicted in Figure 1.1. To facilitate comparisons and exaggerate the differences across models, the right-hand-panel depicts the logarithm of the same impulse response coefficients. The RiskMetrics forecasts, corresponding to an IGARCH(1,1) model with $\alpha = 0.06$, $\beta = 1 - \alpha = 0.94$ and $\omega = 0$, obviously results in infinitely persistent volatility shocks. In contrast, the impulse response coefficients associated with the GARCH(1,1) forecasts die out at the exponential rate $(0.085 + 0.881)^h$, as manifest by the log-linear relationship in the right-hand-

panel. Although the component GARCH model also implies an exponential decay and therefore a log-linear relationship, it fairly closely matches the hyperbolic decay rate for the long-memory FIGARCH model for the first *125* steps. However, the two models clearly behave differently for forecasts further into the future. Whether these differences and potential gains in forecast accuracy over longer horizons are worth the extra complications associated with the implementation of a fractional integrated model obviously depends on the specific uses of the forecasts.

3.5 Parameter Estimation

The values of the parameters in the GARCH models are, of course, not known in practice and will have to be estimated. By far the most commonly employed approach for doing so is Maximum Likelihood Estimation (MLE) under the additional assumption that the standardized innovations in equation (3.5), $z_t \equiv \sigma_{t|t-1}^{-1}(y_t - \mu_{t|t-1})$, are *i.i.d.* normally distributed, or equivalently that the conditional density for y_t takes the form,

$$f(y_t \mid \mathscr{F}_{t-1}) = (2\pi)^{-1/2}\sigma_{t|t-1}^{-1}\exp(-1/2\,\sigma_{t|t-1}^{-2}(y_t - \mu_{t|t-1})^2). \tag{3.23}$$

In particular, let θ denote the vector of unknown parameters entering the conditional mean and variance functions to be estimated. By standard recursive conditioning arguments, the log-likelihood function for the $y_T, y_{T-1}, \ldots, y_1$ sample is then simply given by the sum of the corresponding T logarithmic conditional densities,

$$logL(\theta; y_T, \ldots, y_1) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\log\sigma_{t|t-1}^2(\theta) - \sigma_{t|t-1}^{-2}(\theta)\,(y_t - \mu_{t|t-1}(\theta))^2\right] \tag{3.24}$$

The likelihood function obviously depends upon the parameters in a highly non-linear fashion, and numerical optimization techniques are required in order to find the value of θ which maximizes the function, say $\hat{\theta}_T$. Also, to start up the recursions for calculating $\sigma_{t|t-1}^2(\theta)$, pre-sample values of the conditional variances and squared innovations are also generally required. If the model is stationary, these initial values may be fixed at their unconditional sample counter parts, without affecting the asymptotic distribution of the resulting estimates. Fortunately, there now exist numerous software packages for estimating all of the different GARCH formulations discussed above based upon this likelihood approach.

Importantly, provided that the model is correctly specified and satisfies a necessary set of technical regularity conditions, the estimates obtained by maximizing the function in (3.24) inherit the usual optimality properties associated with MLE, allowing for standard parameter inference based on an estimate of the corresponding information matrix. This same asymptotic distribution may also be used in incorporating the parameter estimation error uncertainty in the

distribution of the volatility forecasts from the underlying model. However, this effect is typically ignored in practice, instead relying on a simple plugin approach using $\hat{\theta}_T$ in place of the true unknown parameters in the forecasting formulas. Of course, in many financial applications the size of the sample used in the parameter estimation phase is often very large compared to the horizon of the forecasts, so that the additional influence of the parameter estimation error is likely to be relatively minor compared to the inherent uncertainty in the forecasts from the model. Bayesian inference procedures can, of course, also be used in directly incorporating the parameter estimation error uncertainty in the model forecasts.

More importantly from a practical perspective, the log-likelihood function in equation (3.24) employed in almost all software packages is based on the assumption that z_t is i.i.d. normally distributed. Although this assumption coupled with time-varying volatility implies that the unconditional distribution of y_t has fatter tails than the normal, this is typically not sufficient to account for all of the mass in the tails in the distributions of daily or weekly returns. Hence, the likelihood function is formally misspecified.

However, if the conditional mean and variance are correctly specified, the corresponding Quasi Maximum Likelihood Estimates (QMLE) obtained under this auxiliary assumption of conditional normality will generally be consistent for the true value of θ. Moreover, asymptotically valid robust standard errors may be calculated from the so-called "sandwich-form" of the covariance matrix estimator, defined by the outer product of the gradients post- and pre-multiplied by the inverse of the usual information matrix estimator. Since the expressions for the future conditional variances for most of the GARCH models discussed above do not depend upon the actual distribution of z_t, as long as $E(z_t \mid \mathscr{F}_{t-1}) = 0$ and $E(z_t^2 \mid \mathscr{F}_{t-1}) = 1$, this means that asymptotically valid point volatility forecasts may be constructed from the conditionally normal QMLE for θ without fully specifying the distribution of z_t.

Still, the efficiency of the parameter estimates, and therefore the accuracy of the resulting point volatility forecasts obtained by simply substituting $\hat{\theta}_T$ in place of the unknown parameters in the forecasting formulas, may be improved by employing the correct conditional distribution of z_t. A standardized student's t distribution with degrees of freedom $\nu > 2$ often provides a good approximation to this distribution. Specifically,

$$f(y_t \mid \mathscr{F}_{t-1}) = \Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{-1}((\nu-2)\sigma^2_{t|t-1})^{-1/2}(1 + (\nu-2)^{-1}\sigma^{-2}_{t|t-1}(y_t - \mu_{t|t-1})^2)^{-(\nu+1)/2} \tag{3.25}$$

with the log likelihood function given by the sum of the corresponding T logarithmic densities, and the degrees of freedom parameter ν estimated jointly with the other parameters of the model entering the conditional mean and variance functions. Note, that for $\nu \to \infty$ the distribution

converges to the conditional normal density in (3.23). Of course, more flexible distributions allowing for both fat tails and asymmetries could be, and have been, employed as well. Additionally, semi-nonparametric procedures in which the parameters in $\mu_{t|t-1}(\theta)$ and $\sigma^2_{t|t-1}(\theta)$ are estimated sequentially on the basis of nonparametric kernel type estimates for the distribution of $\hat{z}_t$ have also been developed to enhance the efficiency of the parameter estimates relative to the conditionally normal QMLEs. From a forecasting perspective, however, the main advantage of these more complicated conditionally non-normal estimation procedures lies not so much in the enhanced efficiency of the plugin point volatility forecasts, $\sigma^2_{T+h|T}(\hat{\theta}_T)$, but rather in their ability to better approximate the tails in the corresponding predictive distributions, $f(y_{T+h} \mid \mathscr{F}_T; \hat{\theta}_T)$. We next turn to a discussion of this type of density forecasting.

3.6 Fat Tails and Multi-Period Forecast Distributions

The ARCH class of models directly specifies the one-step-ahead conditional mean and variance, $\mu_{t|t-1}$ and $\sigma^2_{t|t-1}$, as functions of the time t-1 information set, $\mathscr{F}_{t-1}$. As such, the one-period-ahead predictive density for y_t is directly determined by the distribution of z_t. In particular, assuming that z_t is *i.i.d.* standard normal,

$$f_z(z_t) = (2\pi)^{-1/2}\exp(-1/2z_t),$$

the conditional density of y_t is then given by the expression in equation (3.23) above, where the $\sigma^{-1}_{t|t-1}$ term is associated with the Jacobian of the transformation from z_t to y_t. Thus, in this situation, the one-period-ahead VaR at level *p* is readily calculated by $VaR^p_{t+1|t} = \mu_{t+1|t} + \sigma_{t+1|t}F_z^{-1}(p)$, where $F_z^{-1}(p)$ equals the *p*'th quantile in the standard normal distribution.

Meanwhile, as noted above the distributions of the standardized GARCH innovations often have fatter tails than the normal distribution. To accommodate this feature alternative conditional error distributions, such as the student-t distribution in equation (3.25) discussed above, may be used in place of the normal density in equation (3.23) in the construction of empirically more realistic predictive densities. In the context of quantile predictions, or VaR's, this translates into multiplication factors, $F_z^{-1}(p)$, in excess of those for the normal distribution for small values of *p*. Of course, the exact value of $F_z^{-1}(p)$ will depend upon the specific parametric estimates for the distribution of z_t. Alternatively, the standardized in-sample residuals based on the simpler-to-implement QMLE for the parameters, say $\hat{z}_t \equiv \hat{\sigma}^{-1}_{t|t-1}(y_t - \hat{\mu}_{t|t-1})$, may be used in non-parametrically estimating the distribution of z_t, and in turn the quantiles, $\hat{F}_z^{-1}(p)$.

The procedures discussed above generally work well in approximating VaR's within the main range of support of the distribution, say $0.01 < p < 0.99$. However, for quantiles in the very far left or right tail, it is not possible to meaningfully estimate $F_z^{-1}(p)$ without imposing some additional structure on the problem. Extreme Value Theory (EVT) provides a framework for doing so. In particular, it follows from EVT that under general conditions the tails of any admissible distribution must behave like those of the Generalized Pareto class of distributions. Hence, provided that z_t is *i.i.d.*, the extreme quantiles in $f(y_{t+1} \mid \mathscr{F}_t)$ may be inferred exactly as above, using only the $[rT]$ smallest (largest) values of $\hat{z}_t$ in actually estimating the parameters of the corresponding extreme value distribution used in calculating $\hat{F}_z^{-1}(p)$. The fraction r of the full sample T used in this estimation dictates where the tails, and consequently the extreme value distribution, begin. In addition to standard MLE techniques, a number of simplified procedures, including the popular Hill estimator, are also available for estimating the required tail parameters.

The calculation of multi-period forecast distributions is more complicated. To facilitate the presentation, suppose that the information set defining the conditional one-step-ahead distribution, $f(y_{t+1} \mid \mathscr{F}_t)$, and consequently the conditional mean and variance, $\mu_{t+1|t}$ and $\sigma^2_{t+1|t}$ respectively, is restricted to current and past values of y_t. The multi-period-ahead predictive distribution is then formally defined by the convolution of the corresponding h one-step-ahead distributions,

$$f(y_{t+h} \mid \mathscr{F}_t) = \int\int \cdots \int f(y_{t+h} \mid \mathscr{F}_{t+h-1}) f(y_{t+h-1} \mid \mathscr{F}_{t+h-2}) \cdots f(y_{t+1} \mid \mathscr{F}_t)\, dy_{t+h-1}\, dy_{t+h-2} \cdots dy_{t+1}. \tag{3.26}$$

This multi-period mixture distribution generally has fatter tails than the underlying one-step-ahead distributions. In particular, assuming that the one-step-ahead distributions are conditionally normal as in (3.23) then, if the limiting value exists, the unconditional distribution, $f(y_t) = \lim_{h\to\infty} f(y_t \mid \mathscr{F}_{t-h})$, will be leptokurtic relative to the normal. This is, of course, entirely consistent with the unconditional distribution of most speculative returns having fatter tails than the normal. It is also worth noting that even though the conditional one-step-ahead predictive distributions, $f(y_{t+1} \mid \mathscr{F}_t)$, may be symmetric, if the conditional variance depends on the past values of y_t in an asymmetric fashion, as in the GJR, AGARCH or EGARCH models, the multi-step-ahead distribution, $f(y_{t+h} \mid \mathscr{F}_t)$ $h>1$, will generally be asymmetric. Again, this is directly in line with the negative skewness observed in the unconditional distribution of most equity index return series.

Despite these general results, analytical expressions for the multi-period predictive density in (3.26) are not available in closed-form. However, numerical techniques may be used in recursively building up an estimate for the predictive distribution, by repeatedly drawing future

values for $y_{t+j} = \mu_{t+j|t+j-1} + \sigma_{t+j|t+j-1} z_{t+j}$ based on the assumed parametric distribution $f_z(z_t)$, or by bootstrapping z_{t+j} from the in-sample distribution of the standardized residuals.

Alternatively, $f(y_{t+h} \mid \mathscr{F}_t)$ may be approximated by a time-invariant parametric or non-parametrically estimated distribution with conditional mean and variance, $\mu_{t+h|t} \equiv E(y_{t+j} \mid \mathscr{F}_t)$ and $\sigma^2_{t+h|t} \equiv Var(y_{t+j} \mid \mathscr{F}_t)$, respectively. The multi-step conditional variance is readily calculated along the lines of the recursive prediction formulas discussed in the preceding sections. This approach obviously neglects any higher order dependencies implied by the convolution in (3.26). However, in contrast to the common approach of scaling which, as illustrated in Figure 3.2, may greatly exaggerate the volatility-of-the-volatility, the use of the correct multi-period conditional variance means that this relatively simple-to-implement approach for calculating multi-period predictive distributions usually works very well in practice.

The preceding discussion has focused on one or multi-period forecast distributions spanning the identical unit time interval as in the underlying GARCH model. However, as previously noted, in financial applications the forecast distribution of interest often involves the sum of y_{t+j} over multiple periods corresponding to the distribution of continuously compounded multi-period returns, say $y_{t:t+h} \equiv y_{t+h} + y_{t+h-1} + \ldots + y_{t+1}$. The same numerical techniques used in approximating $f(y_{t+h} \mid \mathscr{F}_t)$ by Monte Carlo simulations discussed above may, of course, be used in approximating the corresponding distribution of the sum, $f(y_{t:t+h} \mid \mathscr{F}_t)$.

Alternatively, assuming that the y_{t+j}'s are serially uncorrelated, as would be approximately true for most speculative returns over daily or weekly horizons, the conditional variance of $y_{t:t+h}$ is simply equal to the sum of the corresponding h variance forecasts,

$$Var(y_{t:t+h} \mid \mathscr{F}_t) \equiv \sigma^2_{t:t+h|t} = \sigma^2_{t+h|t} + \sigma^2_{t+h-1|t} + \ldots + \sigma^2_{t+1|t}. \tag{3.27}$$

Thus, in this situation the conditional distribution of $y_{t:t+h}$ may be estimated on the basis of the corresponding in-sample standardized residuals, $\hat{z}_{t:t+h} \equiv \hat{\sigma}^{-1}_{t:t+h|t}(y_{t:t+h} - \hat{\mu}_{t:t+h|t})$. Now, if the underlying GARCH process for y_t is covariance stationary, we have $\lim_{h\to\infty} h^{-1}\mu_{t:t+h} = E(y_t)$ and $\lim_{h\to\infty} h^{-1}\sigma^2_{t:t+h} = Var(y_t)$. Moreover, as shown by Diebold (1988), it follows from a version of the standard Central Limit Theorem that $z_{t:t+h} \Rightarrow N(0,1)$. Thus, volatility clustering disappears under temporal aggregation, and the unconditional return distributions will be increasingly better approximated by a normal distribution the longer the return horizons. This suggests that for longer-run forecasts, or moderately large values of h, the distribution of $z_{t:t+h}$ will be approximately normal. Consequently, the calculation of longer-run multi-period VaR's may reasonably rely on the conventional quantiles from a standard normal probability table in place

of $F_z^{-1}(p)$ in the formula $VaR_{t:t+h|t}^{p} = \mu_{t:t+h|t} + \sigma_{t:t+h|t} F_z^{-1}(p)$.

3.7 Further Reading

The ARCH and GARCH class of models have been extensively surveyed elsewhere; see, e.g., review articles by Andersen and Bollerslev (1998b), Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and Nelson (1994), Diebold (2004), Diebold and Lopez (1995), Engle (2001, 2004), Engle and Patton (2001), Pagan (1996), Palm (1996), and Shephard (1996). The models have now also become part of the standard toolbox discussed in econometrics and empirical oriented finance textbooks; see e.g., Hamilton (1994), Mills (1993), Franses and van Dijk (2000), Gourieroux and Jasiak (2001), Alexander (2002), Brooks (2002), Chan (2002), Tsay (2002), Christoffersen (2003), Enders (2004), and Taylor (2004). A series of the most influential early ARCH papers have been collected in Engle (1995). A fairly comprehensive list as well as forecast comparison of the most important parametric formulations are also provided in Hansen and Lunde (2005).

Several different econometric and statistical software packages are available for estimating all of the most standard univariate GARCH models, including EViews, PC-GIVE, Limdep, Microfit, RATS, S+, SAS, SHAZAM, and TSP. The open-ended matrix programming environments GAUSS, Matlab, and Ox also offer easy add-ons for GARCH estimation, while the NAG library and the UCSD Department of Economics website provide various Fortran based procedures and programs. Partial surveys and comparisons of some of these estimation packages and procedures are given in Brooks (1997), Brooks, Burke and Persand (2001), and McCullough and Renfro (1998).

The asymmetry, or "leverage" effect, directly motivating a number of the alternative GARCH formulations were first documented empirically by Black (1976) and Christie (1982). In addition to the papers by Nelson (1991), Engle and Ng (1993), Glosten, Jagannathan and Runkle (1993), and Zakoian (1994) discussed in Section 3.3, other important studies on modeling and understanding the volatility asymmetry in the GARCH context include Campbell and Hentschel (1992), Hentschel (1995), and Bekaert and Wu (2000), while Engle (2001) provides an illustration of the importance of incorporating asymmetry in GARCH-based VaR calculations.

The long-memory FIGARCH model of Baillie, Bollerslev and Mikkelsen (1996) in Section 3.4 may be seen as a special case of the ARCH(∞) model in Robinson (1991). The FIGARCH model also encompasses the IGARCH model of Engle and Bollerslev (1986) for $d=1$. However, even though the approach discussed here affords a convenient framework for generating point forecasts with long-memory dependencies, when viewed as a model the unconditional variance does not exist, and the FIGARCH class of models has been criticized accordingly by Giraitis, Kokoszka and Leipus (2000), among others. An alternative formulation which breaks the link between the conditions for second-order stationarity and long-memory dependencies have been proposed by Davidson (2004). Alternative long-memory GARCH formulations include the FIEGARCH model of Bollerslev and Mikkelsen (1996), and the model in Ding and Granger (1996) based on the superposition of an infinite number of ARCH models. In contrast, the

component GARCH model in Engle and Lee (1999) and the related developments in Gallant, Hsu and Tauchen (1999) and Müller et al. (1997), is based on the mixture of only a few components; see also the earlier related results on modeling and forecasting long-run dynamic dependencies in the mean by O'Connell (1971) and Tiao and Tsay (1994). Meanwhile, Bollerslev and Mikkelsen (1999) have argued that when pricing very long-lived financial contracts, the fractionally integrated volatility approach can result in materially different prices from the ones implied by the more standard GARCH models with exponential decay. The multifractal models recently advocated by Calvet and Fisher (2002, 2004) afford another approach for incorporating long-memory into volatility forecasting.

Long-memory also has potential links to regimes and structural break in volatility. Diebold and Inoue (2001) argue that the apparent finding of long-memory could be due to the existence of regime switching. Mikosch and Starica (2004) explicitly uses nonstationarity as a source of long-memory in volatility. Structural breaks in volatility is considered by Andreou and Ghysels (2002), Lamoureux and Lastrapes (1990), Pastor and Stambaugh (2001), and Schwert (1989). Hamilton and Lin (1996) and Perez-Quiros and Timmermann (2000) study volatility across business cycle regimes. The connections between long-memory and structural breaks are reviewed in Banerjee and Urga (2005); see also the chapter by Clements and Hendry in this Handbook.

Early contributions concerning the probabilistic and statistical properties of GARCH models, as well as the MLE and QMLE techniques discussed in Section 3.5, include Bollerslev and Wooldridge (1992), Lee and Hansen (1994), Lumsdaine (1996), Nelson (1990), and Weiss (1986); for a survey of this literature see also Li, Ling and McAleer (2002). Bollerslev (1986) discusses conditions for existence of the second moment in the specific context of the GARCH model. Loretan and Phillips (1994) contains a more general discussion on the issue of covariance stationarity. Bayesian methods for estimating ARCH models were first implemented by Geweke (1989a) and they have since be developed further in Bauwens and Lubrano (1998, 1999). The GARCH-t model discussed in Section 3.5 was first introduced by Bollerslev (1987), while Nelson (1991) suggested the so-called Generalized Error Distribution (GED) for better approximating the distribution of the standardized innovations. Engle and Gonzalez-Rivera (1991) first proposed the use of kernel-based methods for non-parametrically estimating the conditional distribution, whereas McNeil and Frey (2000) relied on Extreme Value Theory (EVT) for estimating the uppermost tails in the conditional distribution; see also Embrechts, Klüppelberg and Mikosch (1997) for a general discussion of extreme value theory.

As discussed in Section 3.6, even if the one-step-ahead conditional distribution is known (by assumption), the corresponding multi-period distributions are not available in closed-form and are generally unknown. Some of the complications that arise in this situation have been discussed in Baillie and Bollerslev (1992), who also consider the use of a Cornish-Fisher expansion for approximating specific quantiles in the multi-step-ahead predictive distributions. Numerical techniques for calculating the predictive distributions based on importance sampling schemes were first implemented by Geweke (1989b). Other important results related to the distribution of temporally aggregated GARCH models include Drost and Nijman (1993), Drost

and Werker (1996), and Meddahi and Renault (2004).

4. Stochastic Volatility

This section introduces the general class of models labeled Stochastic Volatility (SV). In the widest sense of the term, SV models simply allow for a stochastic element in the time series evolution of the conditional variance process. For example, GARCH models are SV models. The more meaningful categorization, which we adopt here, is to contrast ARCH type models with *genuine* SV models. The latter explicitly includes an unobserved (non-measurable) shock to the return variance into the characterization of the volatility dynamics. In this scenario, the variance process becomes inherently latent so that - even conditional on all past information and perfect knowledge about the data generating process - we cannot recover the exact value of the current volatility state. The technical implication is that the volatility process is not measurable with respect to observable (past) information. Hence, the assessment of the volatility state at day t changes as contemporaneous or future information from days $t+j, j \geq 0$, is incorporated into the analysis. This perspective renders estimation of latent variables from past data alone (filtering) as well as from all available, including future, data (smoothing) useful. In contrast, GARCH models treat the conditional variance as observable given past information and, as discussed above, typically applies (quasi-) maximum likelihood techniques for inference, so smoothing has no role in that setting.

Despite these differences, the two model classes are closely related, and we consider them to be complementary rather than competitors. In fact, from a practical forecasting perspective it is hard to distinguish the performance of standard ARCH and SV models. Hence, even if one were to think that the SV framework is appealing, the fact that ARCH models typically are easier to estimate explains practitioners reliance on ARCH as the volatility forecasting tool of choice. Nonetheless, the development of powerful method of simulated moments, Markov Chain Monte Carlo (MCMC) and other simulation based procedures for estimation and forecasting of SV models may render them competitive with ARCH over time. Moreover, the development of the concept of realized volatility and the associated use of intraday data for volatility measurement, discussed in the next section, is naturally linked to the continuous-time SV framework of financial economics.

The literature on SV models is vast and rapidly growing, and excellent surveys are already available on the subject, e.g., Ghysels, Harvey and Renault (1996) and Shephard (1996, 2004). Consequently, we focus on providing an overview of the main approaches with particular emphasis on the generation of volatility forecasts within each type of model specification and inferential technique.

4.1 Model Specification

Roughly speaking, there are two main perspectives behind the SV paradigm when used in the context of modeling financial rate of returns. Although both may be adapted to either setting,

there are precedents for one type of reasoning to be implemented in discrete time and the other to be cast in continuous time. The first centers on the Mixture of Distributions Hypothesis (MDH), where returns are governed by an event time process that represents a transformation of the time clock in accordance with the intensity of price relevant news, dating back to Clark (1973). The second approach stems from financial economics where the price and volatility processes often are modeled separately via continuous sample path diffusions governed by stochastic differential equations. We briefly introduce these model classes and point out some of the similarities to ARCH models in terms of forecasting procedures. However, the presence of a latent volatility factor renders both the estimation and forecasting problem more complex for the SV models. We detail these issues in the following subsections.

4.1.1 The Mixture-of-Distributions Hypothesis

Adopting the rational perspective that asset prices reflect the discounted value of future expected cash flows, such prices should react almost continuously to the myriad of news that arrive on a given trading day. Assuming that the number of news arrival is large, one may expect a central limit theory to apply and financial returns should be well approximated by a conditional normal distribution with the conditioning variable corresponding to the number of relevant news events. More generally, a number of other variables associated with the overall activity of the financial market such as the daily number of trades, the daily cumulative trading volume or the number of quotes may well be similarly related to the information flow in the market. These considerations inspire the following type of representation,

$$y_t = \mu_y s_t + \sigma_y s_t^{1/2} z_t, \tag{4.1}$$

where y_t is the market "activity" variable under consideration, s_t is the strictly positive process reflecting the intensity of relevant news arrivals, μ_y represents the mean response of the variable per news event, σ_y is a scale parameter, and z_t is *i.i.d.* N(0,1). Equivalently, this relationship may be written,

$$y_t | s_t \sim N(\mu_y s_t, \sigma_y^2 s_t). \tag{4.2}$$

This formulation constitutes a normal mixture model. If the s_t process is time-varying it induces a fat-tailed unconditional distribution, consistent with stylized facts for most return and trading volume series. Intuitively, days with high information flow display more price fluctuations and activity than days with fewer news releases. Moreover, if the s_t process is positively correlated, then shocks to the conditional mean and variance process for y_t will be persistent. This is consistent with the observed activity clustering in financial markets, where return volatility, trading volume, the number of transactions and quotes, the number of limit orders submitted to the market, etc., all display pronounced serial dependence.

The specification in (4.1) is analogous to the one-step-ahead decomposition given in equation (3.5). The critical difference is that the formulation is endowed with a structural interpretation,

implying that the mean and variance components cannot be observed prior to the trading day as the number of news arrivals is inherently random. In fact, it is usually assumed that the s_t process is unobserved by the econometrician, even during period t, so that the true mean and variance series are both latent. From a technical perspective this implies that we must distinguish between the full information set ($s_t \in \mathscr{F}_t$) and observable information ($s_t \notin \mathfrak{I}_t$). The latter property is a defining feature of the genuine volatility class. The inability to observe this important component of the MDH model complicates inference and forecasting procedures as discussed below.

In the case of short horizon return series, μ_y is close to negligible and can reasonably be ignored or simply fixed at a small constant value. Furthermore, if the mixing variable s_t is latent then the scaling parameter, σ_y, is not separately identified and may be fixed at unity. This produces the following return (innovation) model,

$$r_t = s_t^{1/2} z_t, \tag{4.3}$$

implying a simple normal-mixture representation,

$$r_t | s_t \sim N(0, s_t). \tag{4.4}$$

Both univariate models for returns of the form (4.4) or multivariate systems including a return variable along with other related market activity variables, such as trading volume or the number of transactions, are referred to as derived from the Mixture-of-Distributions Hypothesis (MDH).

The representation in (4.3) is of course directly comparable to that for the return innovation in equation (3.5). It follows immediately that volatility forecasting is related to forecasts of the latent volatility factor given the observed information,

$$Var(r_{t+h} | \mathfrak{I}_t) = E(s_{t+h} | \mathfrak{I}_t). \tag{4.5}$$

If some relevant information is not observed and thus not included in $\mathfrak{I}_t$, then the expression in (4.5) will generally not represent the actual conditional return variance, $E(s_{t+h} | \mathscr{F}_t)$. This point is readily seen through a specific example.

In particular, Taylor (1986) first introduced the log SV model by adopting an autoregressive parameterization of the latent log-volatility (or information flow) variable,

$$\log s_{t+1} = \eta_0 + \eta_1 \log s_t + u_t, \qquad u_t \sim i.i.d.(0, \sigma_u^2), \tag{4.6}$$

where the disturbance term may be correlated with the innovation in the return equation, that is, $\rho = corr(u_t, z_t) \neq 0$. This particular representation, along with a Gaussian assumption on u_t, has been so widely adopted that it has come to be known as *the* stochastic volatility model. Note that, if ρ is negative, there is an asymmetric return-volatility relationship present in the model,

akin to the "leverage effect" in the GJR and EGARCH models discussed in Section 3.3, so that negative return shocks induce higher future volatility than similar positive shocks. In fact, it is readily seen that the log SV formulation in (4.6) generalizes the EGARCH(1,1) model by considering the case,

$$u_t = \alpha(|z_t| - E|z_t|) + \gamma z_t, \tag{4.7}$$

where the parameters η_0 and η_1 correspond to ω and β in equation (3.15) respectively. Under the null hypothesis of EGARCH(1,1), the information set, $\mathfrak{I}_t$, includes past asset returns, and the idiosyncratic return innovation series, z_t, is effectively observable so likelihood based analysis is straightforward. However, if u_t is not (only) a function of z_t, i.e., equation (4.7) no longer holds, then there are two sources of error in the system. In this more general case it is no longer possible to separately identify the underlying innovations to the return and volatility processes, nor the true underlying volatility state.

This above example illustrates both how any ARCH model may be seen as a special case of a corresponding SV model and how the defining feature of the genuine SV model may complicate forecasting, as the volatility state is unobserved. Obviously, in representations like (4.6), the current state of volatility is a critical ingredient for forecasts of future volatility. We expand on the tasks confronting estimation and volatility forecasting in this setting in Section 4.1.3.

There are, of course, an unlimited number of alternative specifications that may be entertained for the latent volatility process. However, Stochastic Autoregressive Volatility (SARV) of Andersen (1994) has proven particular convenient. The representation is again autoregressive,

$$v_t = \omega + \beta v_{t-1} + [\gamma + \alpha v_{t-1}]\, u_t, \tag{4.8}$$

where u_t denotes an *i.i.d.* sequence, and $s_t = g(v_t)$ links the dynamic evolution of the state variable to the stochastic variance factor in equation (4.3). For example, for the log SV model, $g(v_t) = \exp(v_t)$. Likewise, SV generalizations of the GARCH(1,1) may be obtained via $g(v_t) = v_t$ and an SV extension of a GARCH model for the conditional standard deviation is produced by letting $g(v_t) = v_t^{1/2}$. Depending upon the specific transformation g(·) it may be necessary to impose additional (positivity) constraints on the innovation sequence u_t, or the parameters in (4.8). Even if inference on parameters can be done, moment based procedures do not produce estimates of the latent volatility process, so from a forecasting perspective the analysis must necessarily be supplemented with some method of approximating the sample path realization of the underlying state variables.

4.1.2 Continuous-Time Stochastic Volatility Models

The modeling of asset returns in continuous time stems from the financial economics literature where early contributions to portfolio selection by Merton (1969) and option pricing by Black and Scholes (1973) demonstrated the analytical power of the diffusion framework in handling dynamic asset allocation and pricing problems. The idea of casting these problems in a

continuous-time diffusion context also has a remarkable precedent in Bachelier (1900).

Under weak regularity conditions, the general representation of an arbitrage-free asset price process is

$$dp(t) = \mu(t)\, dt + \sigma(t)\, dW(t) + j(t)\, dq(t), \quad t \in [0,T], \tag{4.9}$$

where $\mu(t)$ is a continuous, locally bounded variation process, the volatility process $\sigma(t)$ is strictly positive, $W(t)$ denotes a standard Brownian motion, q(t) is a jump indicator taking the values zero (no jump) or unity (jump) and, finally, the $j(t)$ represents the size of the jump if one occurs at time t. (See, e.g., Andersen, Bollerslev and Diebold (2003a) for further discussion.) The associated one-period return is

$$r(t) = p(t) - p(t-1) = \int_{t-1}^{t} \mu(\tau)\, d\tau + \int_{t-1}^{t} \sigma(\tau)\, dW(\tau) + \sum_{t-1 \le \tau < t} \kappa(\tau), \tag{4.10}$$

where the last sum simply cumulates the impact of the jumps occurring over the period, as we define $\kappa(t) = j(t) \cdot \mathbf{I}(q(t) = 1)$, so that $\kappa(t)$ is zero everywhere except when a discrete jump occurs.

In this setting a formal ex-post measure of the return variability, derived from the theory of quadratic variation for semi-martingales, may be defined,

$$QV(t) \equiv \int_{t-1}^{t} \sigma^2(s)\, ds + \sum_{t-1 < s \le t} \kappa^2(s). \tag{4.11}$$

In the special case of a pure SV diffusion, the corresponding quantity reduces to the integrated variance, as already defined in equation (1.11) in Section 1,

$$IV(t) \equiv \int_{t-1}^{t} \sigma^2(s)\, ds. \tag{4.12}$$

These return variability measures are naturally related to the return variance. In fact, for a pure SV diffusion (without jumps) where the volatility process, $\sigma(\tau)$, is independent of the Wiener process, $W(\tau)$, we have,

$$r(t) \mid \{\mu(\tau), \sigma(\tau);\ t-1 \le \tau \le t\} \sim N\left(\int_{t-1}^{t} \mu(\tau)\, d\tau,\ \int_{t-1}^{t} \sigma^2(\tau)\, d\tau \right), \tag{4.13}$$

so the integrated variance *is* the true measure of the actual (ex-post) return variance in this context. Of course, if the conditional variance and mean processes evolve stochastically we

cannot perfectly predict the future volatility, and we must instead form expectations based on the current information. For short horizons, the conditional mean variation is negligible and we may focus on the following type of forecasts, for a positive integer h,

$$Var(r(t+h)\,|\,\Im_t) \;\approx\; E[\int_{t+h-1}^{t+h} \sigma^2(\tau)d\tau\,|\,\Im_t] \;\equiv\; E[IV(t+h)\,|\,\Im_t]. \tag{4.14}$$

The expressions in (4.13) and (4.14) generalize the corresponding equations for discrete-time SV models in (4.4) and (4.5) respectively. Of course, the return variation arising from the conditional mean process may need to be accommodated as well over longer horizons. Nonetheless, the dominant term in the return variance forecast will invariably be associated with the expected integrated variance or, more generally, the expected quadratic variation. In simple continuous-time models, we may be able to derive closed-form expressions for these quantities, but in empirically realistic settings they are typically not available in analytic form and alternative procedures must be used. We discuss these issues in more detail below.

The initial diffusion models explored in the literature were not genuine SV diffusions but rather, with a view toward tractability, cast as special cases of the constant elasticity of variance (CEV) class of models,

$$dp(t) \;=\; (\,\mu - \phi\,[p(t) - \mu]\,)\,dt \;+\; \sigma\,p(t)^{\gamma}\,dW(t)\,, \qquad t \in [0,T], \tag{4.15}$$

where $\phi \geq 0$ determines the strength of mean reversion toward the unconditional mean μ in the log-price process, while $\gamma \geq 0$ allows for conditional heteroskedasticity in the return process. Popular representations are obtained by specific parameter restrictions, e.g., the Geometric Brownian motion for $\phi = 0$ and $\gamma = 0$, the Vasicek model for $\gamma = 0$, and the Cox-Ingersoll and Ross (CIR) or square-root model for $\gamma = ½$. These three special cases allow for a closed-form characterization of the likelihood, so the analysis is straightforward. Unfortunately, they are also typically inadequate in terms of capturing the volatility dynamics of asset returns. A useful class of extensions have been developed from the CIR model. In this model the instantaneous mean and variance processes are both affine functions of the log price. The affine model class extends the above representation with $\gamma = ½$ to a multivariate setting with general affine conditional mean and variance specifications. The advantage is that a great deal of analytic tractability is retained while allowing for more general and empirically realistic dynamic features.

Many genuine SV representations of empirical interest fall outside of the affine class, however. For example, Hull and White (1987) develop a theory for option pricing under stochastic volatility using a model much in the spirit of Taylor's discrete-time log SV in equation (4.6). With only a minor deviation from their representation, we may write it, for $t \in [0,T]$,

$$\begin{aligned} dp(t) \;&=\; \mu(t)\,dt \;+\; \sigma(t)\,dW(t)\,, \\ d\log\sigma^2(t) \;&=\; \beta\,(\alpha - \log\sigma^2(t))\,dt \;+\; \nu\,dW_{\sigma}(t). \end{aligned} \tag{4.16}$$

The strength of the mean reversion in (log-) volatility is given by β and the volatility is governed by ν. Positive but low values of β induces a pronounced volatility persistence, while large values of ν increase the idiosyncratic variation in the volatility series. Furthermore, the log transform implies that the volatility of volatility rises with the level of volatility, even if ν is time invariant. Finally, a negative correlation, $\rho < 0$, between the Wiener processes W(t) and $W_\sigma(t)$ will induce an asymmetric return-volatility relationship in line with the leverage effect discussed earlier. As such, these features allow the representation in (4.16) to capture a number of stylized facts about asset return series quite parsimoniously.

Another popular non-affine specification is the GARCH diffusion analyzed by Drost and Werker (1996). This representation can formally be shown to induce a GARCH type behavior for any discretely sampled price series and it is therefore a nice framework for eliciting and assessing information about the volatility process through data gathered at different sampling frequencies. This is also the process used in the construction of Figure 1.1. It takes the form,

$$\begin{aligned} dp(t) &= \mu\, dt + \sigma(t)\, dW(t)\,, \\ d\sigma^2(t) &= \beta\,(\alpha - \sigma^2(t))\, dt + \nu\, \sigma^2(t)\, dW_\sigma(t)\,, \end{aligned} \tag{4.17}$$

where the two Wiener processes are now independent.

The SV diffusions in (4.16) and (4.17) are but simple examples of the increasingly complex multi-factor (affine as well as non-affine) jump-diffusions considered in the literature. Such models are hard to estimate by standard likelihood or method of moments techniques. This renders their use in forecasting particularly precarious. There is a need for both reliable parameter estimates and reliable extraction of the values for the underlying state variables. In particular, the current value of the state vector (and thus volatility) constitutes critical conditioning information for volatility prediction. The usefulness of such specifications for volatility forecasting is therefore directly linked to the availability of efficient inference methods for these models.

4.1.3 Estimation and Forecasting Issues in SV Models

The incorporation of a latent volatility process in SV models has two main consequences. First, estimation cannot be performed through a direct application of maximum likelihood principles. Many alternative procedures will involve an efficiency loss relative to this benchmark so model parameter uncertainty may then be larger. Since forecasting is usually made conditional on point estimates for the parameters, this will tend to worsen the predictive ability of model based forecasts. Second, since the current state for volatility is not observed, there is an additional layer of uncertainty surrounding forecasts made conditional on the estimated state of volatility. We discuss these issues below and the following sections then review two alternative estimation and forecasting procedures developed, in part, to cope with these challenges.

Formally, the SV likelihood function is given as follows. Let the vector of return (innovations)

and volatilities over [0,T] be denoted by $\underline{r} = (r_1, \ldots, r_T)$ and $\underline{s} = (s_1, \ldots, s_T)$, respectively. Collecting the parameters in the vector θ, the probability density for the data given θ may then be written as,

$$f(\underline{r};\theta) = \int f(\underline{r},\underline{s};\theta)\,d\underline{s} = \prod_{t=1}^{T} f(r_t|\Im_{t-1};\theta) = \prod_{t=1}^{T} \int f(r_t|s_t;\theta) f(s_t|\Im_{t-1};\theta)\,ds_t \ . \tag{4.18}$$

For parametric discrete-time SV models, the conditional density $f(r_t|\, s_t, \theta)$ is typically known in closed form, but $f(s_t|\Im_{t-1};\theta)$ is not available. Without being able to utilize this decomposition, we face an integration over the full unobserved volatility vector which is a T-dimensional object and generally not practical to compute given the serial dependence in the latent volatility process.

The initial response to these problems was to apply alternative estimation procedures. In his original treatment Taylor (1986) uses moment matching. Later, Andersen (1994) shows that it is feasible to estimate a broad class of discrete-time SV models through standard GMM procedures. However, this is not particularly efficient as the unconditional moments that may be expressed in closed form are quite different from the (efficient) score moments associated with the (infeasible) likelihood function. Another issue with GMM estimates is the need to extract estimates of the state variables if it is to serve as a basis for volatility forecasting. GMM does not provide any direct identification of the state variables, so this must be addressed in a second step. In that regard, the Kalman filter was often used. This technique allows for sequential estimation of parameters and latent state variables. As such, it provides a conceptual basis for the analysis, even if the basic Kalman filter is inadequate for general nonlinear and non-Gaussian SV models.

Nelson (1988) first suggested casting the SV estimation problem in a state space setting. We illustrate the approach for the simplest version of the log SV model without a leverage effect, that is, $\rho = 0$ in (4.4) and (4.6). Now, squaring the expression in (4.3), takings logs and assuming Gaussian errors in the transition equation for the volatility state in equation (4.6), it follows that

$$\log r_t^2 = \log s_t + \log z_t^2, \quad z_t \sim i.i.d.N(0,1)$$

$$\log s_{t+1} = \eta_0 + \eta_1 \log s_t + u_t, \quad u_t \sim i.i.d.\ N(0, \sigma_u^2).$$

To conform with standard notation, it is useful to consolidate the constant from the transition equation into the measurement equation for the log squared return residual. Defining $h_t \equiv \log s_t$, we have

$$\begin{aligned} \log r_t^2 &= \omega + h_t + \xi_t, \quad \xi_t \sim i.i.d.\ (0, 4.93) \\ h_{t+1} &= \eta h_t + u_t, \quad u_t \sim i.i.d.\ N(0, \sigma_u^2), \end{aligned} \tag{4.19}$$

where $\omega = \eta_0 + E(\log z_t^2) = \eta_0 - 1.27$, $\eta = \eta_1$, and ξ_t is a demeaned *log* χ^2 distributed error term. The system in (4.19) is given in the standard linear state space format. The top equation provides the measurement equation where the squared return is linearly related to the latent underlying volatility state and an *i.i.d.* skewed and heavy tailed error term. The bottom equation provides the transition equation for the model and is given as a first-order Gaussian autoregression.

The Kalman filter applies directly to (4.19) by assuming Gaussian errors, see, e.g., Harvey (1989, 2004). However, the resultant estimators of the state variables and the future observations are only minimum mean squared error for estimators that are linear combinations of past log r_t^2. Moreover, the non-Gaussian errors in the measurement equation implies that the exact likelihood cannot be obtained from the associated prediction errors. Nonetheless, the Kalman filter may be used in the construction of QMLEs of the model parameters for which asymptotically valid inference is available, even if these estimates generally are fairly inefficient. Arguably, the most important insight from the state space representation is instead the inspiration it has provided for the development of more efficient estimation and forecasting procedures through nonlinear filtering techniques.

The state space representation directly focuses attention on the task of making inference regarding the latent state vector, i.e., for SV models the question of what we can learn about the current state of volatility. A comprehensive answer is provided by the solution to the *filtering problem*, i.e., the distribution of the state vector given the current information set, $f(s_t \mid \Im_t ; \theta)$. Typically, this distribution is critical in obtaining the one-step-ahead volatility forecast,

$$f(s_t|\Im_{t-1};\theta) = \int f(s_t|s_{t-1};\theta) f(s_{t-1}|\Im_{t-1};\theta)\, ds_{t-1}, \tag{4.20}$$

where the first term in the integral is obtained directly from the transition equation in the state space representation. Once the one-step-ahead distribution has been determined, the task of constructing multiple-step-ahead forecasts is analogous to the corresponding problem under ARCH models where multi-period forecasts also generally depend upon the full distributional characterization of the model. A unique feature of the SV model is instead the *smoothing problem*, related to ex-post inference regarding the in-sample volatility given the set of observed returns over the full sample, $f(s_t \mid \Im_T ; \theta)$, where $t \leq T$. At the end of the sample, either the filtering or smoothing solution can serve as the basis for out-of-sample volatility forecasts (for h a positive integer),

$$f(s_{T+h}|\Im_T;\theta) = \int f(s_{T+h}|s_T;\theta) f(s_T|\Im_T;\theta)\, ds_T, \tag{4.21}$$

where, again, given the solution for $h = 1$, the problem of determining the multi-period forecasts is analogous to the situation with multi-period ARCH-based forecasts discussed in Section 3.6.

As noted, all of these conditional volatility distributions may in theory be derived in closed form under the linear Gaussian state space representation via the Kalman filter. Unfortunately, even

the simplest SV model contains some non-Gaussian and/or nonlinear elements. Hence, standard filtering methods provide, at best, approximate solutions and they have generally been found to perform poorly in this setting, in turn necessitating alternative more specialized filtering and smoothing techniques. Moreover, we have deliberately focused on the discrete-time case above. For the continuous-time SV models, the complications are more profound as even the discrete one-period return distribution conditional on the initial volatility state typically is not known in closed form. Hence, not only is the last term on the extreme right of equation (4.18) unknown, but the first term is also intractable, further complicating likelihood-based analysis. We next review two recent approaches that promise efficient inference more generally and also provide ways of extracting reliable estimates of the latent volatility state needed for forecasting purposes.

4.2 Efficient Method of Simulated Moments Procedures for Inference and Forecasting

The Efficient Method of Moments (EMM) procedure is the prime example of a method of simulated moments (MSM) approach that has the potential to deliver efficient inference and produce credible volatility forecasting for general SV models. The intuition behind EMM is that, by traditional likelihood theory, the scores (the derivative of the log-likelihood with respect to the parameter vector) provide efficient estimating moments. In fact, maximum likelihood is simply a just-identified GMM estimator based on the score (moment) vector. Hence, intuitively, from an efficiency point of view, one would like to approximate the score vector when choosing the GMM moments. Since the likelihood of SV models is intractable, the approach is to utilize a semi-nonparametric approximation to the log-likelihood estimated in a first step to produce the moments. Next, one seeks to match the approximating score moments with the corresponding moments from a long simulation of the SV model. Thus, the main requirement for applicability of EMM is that the model can be simulated effectively and the system is stationary so that the requisite moments can be computed by simple averaging from a simulation of the system. Again, this idea, like the MCMC approach discussed in the next section, is, of course, applicable more generally, but for concreteness we will focus on estimation and forecasting with SV models for financial rate of returns.

More formally, let the sample of discretely observed returns be given by $\underline{r} = (r_1, r_2, \ldots, r_T)$. Moreover, let x_{t-1} denote the vector of relevant conditioning variables for the log-likelihood function at time t, and let $\underline{x} = (x_0, x_1, \ldots, x_{T-1})$. For simplicity, we assume a long string of prior return observations are the only components of $\underline{x}$, but other predetermined variables from an extended dynamic representation of the system may be incorporated as well. In the terminology of equation (4.18), the complication is that the likelihood contribution from the *t*'th return is not available, that is, $f(r_t | \mathcal{F}_{t-1}; \theta) \equiv f(r_t | x_{t-1}; \theta)$ is unknown. The proposal is to instead approximate this density by a flexible semi-nonparametric (SNP) estimate using the full data sample. Without going into specifics, an appropriate procedure may be developed to obtain a close approximation to this conditional density within a class of SNP densities which are analytically tractable and allow for explicit computation of the associated score vector. The leading term will typically consist of a GARCH type model. Essentially, the information regarding the probabilistic structure available from the data is being encoded into an empirically tractable SNP representation, so that, for a large enough sample, we have

$$g(r_t|x_{t-1};\hat{\eta}_T) \approx f(r_t|\Im_{t-1};\theta_0), \tag{4.22}$$

where $g(r_t|x_{t-1};\hat{\eta}_T)$ denotes the fitted SNP density evaluated at the (pseudo) maximum likelihood estimate $\hat{\eta}_T$, and θ_0 denotes the true (unknown) parameter vector of the model generating the data under the null hypothesis. In general, the functional form of g is entirely different from the unknown f, and hence there is no direct compatibility between the two parameter vectors η and θ, although we require that the dimension of η is at least as large as that of θ. Notice how this SNP representation sidesteps the lack of a tractable expression for the likelihood contribution as given by the middle term in the likelihood expression in (4.18). Although the SNP density is not used for formal likelihood estimation, it is used to approximate the "efficient" score moments.

By construction, $\hat{\eta}_T$ satisfies a set of first order conditions for the pseudo log-likelihood function under the empirical measure induced by the data, that is, letting $\underline{r}_t = (r_t, x_{t-1})$, it holds that,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\eta}\log g(r_t|x_{t-1};\hat{\eta}_T) \equiv \frac{1}{T}\sum_{t=1}^{T}\psi_T(\underline{r}_t) = 0. \tag{4.23}$$

It is clear that (4.23) takes the form of (pseudo) score moments. This representation of the data through a set of (efficient) moment conditions is the key part of the "projection step" of EMM. The data structure has effectively been projected onto an analytically tractable class of SNP densities augmented, as appropriate, by a leading dynamic (GARCH) term.

Since we are working under the assumption that we have a good approximation to the underlying true conditional density, we would intuitively expect that, for T large,

$$E_{\theta_0}[\psi_T(\tilde{r})] \approx \frac{1}{M}\sum_{i=1}^{M}\psi_T(\underline{\tilde{r}}_i) \approx \frac{1}{T}\sum_{t=1}^{T}\psi_T(\underline{r}_t) = 0, \tag{4.24}$$

and any large artificial sample, $\underline{\tilde{r}} = (\tilde{r}_1,\tilde{r}_2,\ldots,\tilde{r}_N,\tilde{x}_0,\tilde{x}_1,\ldots,\tilde{x}_{M-1})$, generated by the same assumed (true) data generating process, $f(r_t|\,\Im_{t-1};\,\theta_0)$, that is behind the observed return data, $\underline{r}$. These conjectures are formalized by Gallant and Tauchen (1996), who show how the pseudo score moments obtained in (4.23) by fixing $\hat{\eta}_T$ can serve as valid (and efficient) moment conditions for estimating the parameter of interest, θ. Since no analytic expression for the expectation on the extreme left in (4.24) is available, they propose a simulation estimator where the expectation is approximated arbitrarily well by a very large simulated sample moment ($M>>T$) from the true underlying model. The ability to practically eliminate the simulation error renders the EMM estimator (in theory) independent of simulation size, M, but the uncertainty associated with the projection step, for which the sample size is constrained by the actual data, remains and the estimator, $\hat{\theta}_T$, is asymptotically normal with standard errors that reflects the estimation uncertainty in (4.23).

An obvious attraction of the EMM technique, beyond the potential for efficient inference, is that there are almost no restrictions on the underlying parametric model apart from stationarity and the ability to be able to simulate effectively from the model. This implies that the procedure can be used for continuous-time processes, even if we only observe a set of discretely sampled data. A seemingly important drawback, however, is the lack of any implied estimates of the underlying latent state variables which are critical for successful forecasting. Gallant and Tauchen (1998) provides a solution within the EMM setting through the so-called reprojection technique, but the procedure can be used more widely in parametric dynamic latent variable model estimated by other means as well.

Reprojection takes the parameter estimate of the system as given, i.e., the EMM estimator for θ in the current context. It is then feasible to generate arbitrarily long simulated series of observable and latent variables. These simulated series can be used for estimation of the conditional density via a SNP density function approximation as under the projection step described above. In other words, the identical procedure is exploited but now for a long simulated series from the null model rather than for the observed data sample. For illustration, let $\tilde{\underline{r}} = (\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_N, \tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{M-1})$ be a long simulated series from the null model, $f(r_i \mid \mathscr{F}_{i-1}; \hat{\theta}_T)$, where we condition on the EMM estimate. We may then utilize the SNP density estimate based on the simulated sample, $g(\tilde{r}_t \mid \tilde{x}_{t-1}; \tilde{\eta})$, in lieu of the unknown density for practical calculations, where the point estimate, $\tilde{\eta}$, is treated as independent of the sample size M since the estimation error is negligible for a sufficiently large simulated sample. In effect, the simulations integrate out the latent variables in the representation (4.5). Given the tractability of the SNP densities, we can now evaluate the one-step-ahead conditional mean and variance (or any other moments of interest) directly as a function of any observed history x_{t-1} by simply plugging into the SNP density estimate and perform the integration analytically - this is the reprojection step of recombining the SNP density with the actual data. Clearly, the corresponding multi-step ahead conditional density estimates can be constructed in an analogous fashion. Moreover, since the simulations also generate contemporaneous values for the latent state vectors we may similarly represent the conditional distributions of future latent state variables given the current and past observable variables through the SNP density approximation strategy,

$$f(\tilde{s}_{t+j} \mid \tilde{x}_t; \hat{\theta}_T) \approx g(\tilde{s}_{t+j} \mid \tilde{x}_t; \tilde{\eta}), \qquad j \geq 0. \tag{4.25}$$

This allows for direct forecasts of conditional volatility and associated quantities in a genuine SV setting. As such, reprojection may be interpreted as a numerically intensive, simulation-based, nonlinear Kalman filtering technique, providing a practical solution to the filtering and forecasting problems in equations (4.20) and (4.21).

4.3 Markov Chain Monte Carlo (MCMC) Procedures for Inference and Forecasting

The MCMC method represents a Bayesian approach to the high-dimensional inference problem implicit in the expression for the likelihood given in equation (4.18). The approach was advocated as particularly well suited for analysis of the discrete SV model by Jacquier, Polson

and Rossi (1994). Beyond the standard Bayesian perspective of treating the model parameters as random variables rather than fixed coefficients, the main conceptual shift is that the entire latent state vector is treated as additional parameters. Hence, the main focus is on the joint distribution of the parameters and the vector of state variables, $\psi = (\theta, \underline{s})$, conditional on the data, $f(\psi \mid \underline{r})$, termed the posterior distribution. This density is extremely high-dimensional and analytically intractable. The MCMC approach instead exploits that the joint distribution can be characterized fully through a set of associated conditional distributions where the density for a group of parameters, or even a single parameter, is expressed conditional on the remaining parameters. Concretely, let ψ_i denote the i'th group of coefficients in ψ, and ψ_{-I} be the vector obtained from ψ by excluding the i'th group of coefficients. The so-called Clifford-Hammersley theorem then implies that the following set of conditional distributions determines $f(\psi \mid \underline{r})$,

$$f(\psi_1 \mid \psi_{-1}, \underline{r}), \quad f(\psi_2 \mid \psi_{-2}, \underline{r}), \quad \ldots, \quad f(\psi_k \mid \psi_{-k}, \underline{r}), \tag{4.26}$$

where, as described above, $\psi = (\psi_1, \psi_2, \ldots, \psi_k)$ is treated as k exclusive subsets of parameters.

The MCMC procedure starts by initializing $\psi = (\theta, \underline{s})$ through conditioning on the observed data, $\underline{r}$, and drawing ψ from the assumed prior distribution. Next, by combining the current draw for the parameter vector with the specified SV model dynamics and the observed returns, it is often feasible to draw the (group of) parameters sequentially conditional on the remainder of the system and cycle through the conditional densities in (4.26). A full run through the parameter vector is termed a sweep of the MCMC sampler. Some of these distributions may not be given in closed form and the draws may need to be extended through an accept-reject procedure termed a Metropolis-Hastings algorithm to ensure that the resulting Markov chain produces draws from the invariant joint posterior target distribution. If all the conditional distributions can be sampled directly we have a Gibbs sampler, but SV models often call for the two techniques to be used at different stages of the sweep, resulting in a hybrid MCMC algorithm. Typically, a large number of sweeps is necessary to overcome the serial dependence inherent in draws of any parameter from subsequent sweeps of the sampler. Once a long sample from the joint posterior distribution has been generated, inference on individual parameters and latent state variables can be done via the mode, mean and standard deviation of the posterior distribution, for example. Likewise, we can analyze properties of functions of the state variables directly using the posterior distribution.

A key advantage of the MCMC procedure is that the distribution of the latent state vector is obtained as an inherent part of the estimation. Moreover, the inference automatically accounts for the uncertainty regarding model parameters, θ. The resulting chain produces an elegant solution to the smoothing problem of determining $f(\underline{s} \mid \underline{r})$. Of course, from a forecasting perspective, the interest is in determining $f(s_{t+j} \mid \underline{x}_t)$, where the integer $j \geq 0$ and $\underline{x}_t = (r_1, r_2, \ldots, r_t)$, rather than $f(s_{t+j} \mid \underline{x}_T)$ which is generated by the MCMC procedure. Unfortunately, the filter related distribution, $f(s_{t+1} \mid \underline{x}_t)$, corresponds to the intractable term in equation (4.18) that renders the likelihood estimation impractical for genuine SV models. The MCMC inference procedure succeeds by sidestepping the need to compute this quantity. However, given the economic import of the issue, recent research is actively seeking new effective ways for better handling the filtering problem within the MCMC framework.

For a discrete-time SV model, the possibility of filtering as well as sequential one-step-ahead volatility forecasting is linked to the feasibility of providing an effective scheme to generate a random sample from $f(s_{t+1} \mid \underline{x}_t, \theta)$ given an existing set of draws (or *particles),* $s_t^1, s_t^2, \ldots, s_t^N$, from the preceding distribution $f(s_t \mid \underline{x}_{t-1}, \theta)$. Such an algorithm is termed a *particle filter*. In order to recognize the significance of the particle filter, note that by Bayes' law,

$$f(s_{t+1}|\underline{x}_{t+1},\theta) \propto f(r_{t+1}|s_{t+1},\theta)\, f(s_{t+1}|\underline{x}_t,\theta). \tag{4.27}$$

The first distribution on the right hand side is typically specified directly by the SV model, so the issue of determining the filtering distribution on the left hand side is essentially equivalent to the task of obtaining the predictive distribution of the state variable on the extreme right. But given a large set of particles we can approximate the latter term in straightforward fashion,

$$f(s_{t+1}|\underline{x}_t,\theta) = \int f(s_{t+1}|s_t,\theta)\, f(s_t|\underline{x}_t,\theta)\, ds_t \approx \frac{1}{M}\sum_{j=1}^{M} f(s_{t+1}|s_t^j,\theta). \tag{4.28}$$

This provides a direct solution to the latent state vector forecasting problem, that in turn can be plugged into (4.27) to provide a sequential update to the particle filter. This in essence is the MCMC answer to the filtering and out-of-sample forecasting problems in equations (4.20) and (4.21). The main substantive problem is how to best sample from the last distribution in (4.28), as schemes which may appear natural can be very inefficient, see, e.g., the discussion and suggestions in Kim, Shephard and Chib (1998).

In summary, the MCMC approach works well for many problems of significant interest, but there are serious issues under scrutiny concerning the use of the technique for more complex settings. When applicable, it has some unique advantages such as providing a complete solution to the smoothing problem and accounting for inherent parameter estimation uncertainty. On the other hand, there are systems that are more amenable to analysis under EMM and the associated diagnostic tools and general reprojection procedures under EMM render it a formidable contender. It is remarkable that the issues of efficient forecasting and filtering within genuine SV models now has two attractive, albeit computationally intensive, solutions whereas just a few years ago no serious approach to the problem existed.

4.4 Further Reading

The formal distinction between *genuine* stochastic volatility and ARCH models is developed in Andersen (1992); see also Fleming and Kirby(2003). An early advocate for the Mixture-of-Distributions-Hypothesis (MDH), beyond Clark (1973), is Praetz (1972) who shows that an *i.i.d.* mixture of a Gaussian term and an inverted Gamma distribution for the variance will produce Student-t distributed returns. However, if the variance mixture is not linked to observed variables, the *i.i.d.* mixture is indistinguishable from a standard fat-tailed error distribution and the associated volatility process is not part of the genuinely stochastic volatility class.

Many alternative representations of the driving process s_t have been proposed. Clark (1973) observes that trading volume is highly correlated with return volatility and suggest that volume may serve as a good proxy for the "activity variable," s_t . Moreover, he finds volume to be approximately lognormal (unconditionally), suggesting a lognormal-normal mixture for the return distribution. One drawback of this formulation is that daily trading volume is assumed *i.i.d.* Not only is this counterfactual for trading volume, but it also implies that the return process is *i.i.d.*. This is at odds with the strong empirical evidence of pronounced temporal dependence in return volatility. A number of natural extensions arise from the simple MDH. Tauchen and Pitts (1983) provide a more structural interpretation, as they develop a characterization of the joint distribution of the daily return and volume relationship governed by the underlying latent information flow s_t. However, partially for tractability, they retain the temporal independence of the information flow series. For early tests of the MDH model using high-frequency data, see, e.g., Harris (1986, 1987), while the early return-volume literature is surveyed by Karpoff (1987). Gallant, Rossi and Tauchen (1992) provides an extensive study of the joint conditional distribution without imposing any MDH restrictions. Direct studies of the MDH include Lamoureux and Lastrapes (1994) and Richardson and Smith (1994). While the latter strongly rejects the standard MDH formulation, Andersen (1996) develops an alternative structurally based version of the hypothesis and finds the "modified" MDH to perform much better. Further refinements in the specification have been pursued by, e.g., Liesenfeld (1998, 2001) and Bollerslev and Jubinsky (1999). In principle, the use of additional non-return variables along with return data should enhance estimation efficiency and allow for a better assessment of current market conditions. On the other hand, it is far from obvious that structural modeling of complicated multivariate models will prove useful in a prediction context as even minor mis-specification of the additional series in the system may impede forecast performance. In fact, there is no credible evidence yet that these models help improve volatility forecast performance, even if they have importantly enhanced our understanding of the qualitative functioning of financial markets.

SV diffusion models of the form analyzed by Hull and White (1987) were also proposed concurrently by Johnson and Shanno (1987), Scott (1987), and Wiggins (1987). An early specification and exploration of a pure jump continuous-time model is Merton (1976). Melino and Turnbull (1990) were among the first to estimate SV models via GMM. The log-SV model from (4.2)-(4.3) has emerged as a virtual testing ground for alternative inference procedures in this context. Andersen and Sørensen (1996) provide a systematic study of the choice of moments and weighting matrix for this particular model. The lack of efficiency is highlighted in Andersen, Chung and Sørensen (1999) where the identical model is estimated through the scores of an auxiliary model developed in accordance with the efficient method of moments (EMM) procedure. Another useful approach is to apply GMM to moment conditions in the spectral domain, see, e.g., Singleton (2001), Jiang and Knight (2002), and Chacko and Viceira (2003). Within the QMLE Kalman filter based approach, a leverage effect may be incorporated and allowance for the idiosyncratic return error to be conditionally Student-t distributed can be made, as demonstrated by Harvey, Ruiz and Shephard (1994) and Harvey and Shephard (1996). Andersen and Sørensen (1997) provides an extensive discussion of the relative efficiency of QMLE and GMM for estimation of the discrete-time log SV model. The issue of asymptotically

optimal moment selection for GMM estimation from among absolute or log squared returns in the log SV model has received a near definitive treatment in Dhaene and Vergote (2004). The standard log SV model has also been estimated through a number of other techniques by among others Danielsson and Richard (1993), Danielsson (1994), Fridman and Harris (1998), Monfardini (1998), and Sandman and Koopman (1998). Long-memory in volatility as discussed in Section 3.4 can be similarly accommodated within an SV setting, see, e.g., Breidt, Crato and de Lima (1998), Harvey (1998), Comte and Renault (1998), and Deo and Hurvich (2001). Duffie, Pan and Singleton (2000) is a good reference for a general treatment of modeling with the so-called affine class of models, while Piazzesi (2003) provides a recent survey of these models with a view toward term structure applications.

EMM may be seen as a refinement of the Method of Simulated Moments (MSM) of Duffie and Singleton (1993), representing a particular choice of indirect inference criterion, or binding function, in the terminology of Gouriéroux, Monfort and Renault (1993). The approach also has precedents in Smith (1990, 1993). An early application of EMM techniques to the discrete-time SV model is Gallant, Hsieh and Tauchen (1997). Among the earliest papers using EMM for stochastic volatility models are Andersen and Lund (1997) and Gallant and Tauchen (1997). Extensions of the EMM approach to SV jump-diffusions are found in Andersen, Benzoni and Lund (2002) and Chernov, Gallant, Ghysels and Tauchen (2003). As a starting point for implementations of the EMM procedure, one may access general purpose EMM and SNP code from a web site maintained by A. Ronald Gallant and George E. Tauchen at Duke University at the link *ftp.econ.duke.edu* in the directories *pub/get/emm* and *pub/arg/snp*, respectively. In practical applications, it is often advantageous to further refine the SNP density approximations through specifically designed leading GARCH terms which parsimoneously capture the dependency structure in the specific data under investigation. The benefits of doing so is further discussed in Andersen and Lund (1997) and Andersen, Benzoni and Lund (2002).

The particle filter discussed above for the generation of filter estimates for the latent variables of interest within the standard SV model arguably provides a more versatile approach than the alternative importance sampling methods described by, e.g., Danielsson (1994) and Sandmann and Koopman (1998). The extension of the MCMC inference technique to a continuous-time setting is discussed in Elerian, Chib and Shephard (2001) and Eraker (2001). The latter also provides one of the first examples of MCMC estimation of an SV diffusion model, while Eraker, Johannes and Polson (2003) further introduces jumps in both prices and volatility. Johannes and Polson (2003) offer a recent comprehensive survey of the still ongoing research on the use of the MCMC approach in the general nonlinear jump-diffusion SV setting.

5. Realized Volatility

The notion of realized volatility has at least two key distinct implications for practical volatility estimation and forecasting. The first relates to the measurement of realizations of the latent volatility process without the need to rely on an explicit model. As such, the realized volatility provides *the* natural benchmark for forecast evaluation purposes. The second relates to the

possibility of modeling volatility directly through standard time series techniques with discretely sampled daily observations, while effectively exploiting the information in intraday high-frequency data.

5.1 The Notion of Realized Volatility

The most fundamental feature of realized volatility is that it provides a consistent nonparametric estimate of the price variability that has transpired over a given discrete interval. Any log-price process subject to a no-arbitrage condition and weak auxiliary assumptions will constitute a semi-martingale that may be decomposed into a locally predictable mean component and a martingale with finite second moments. Within this class, there is a unique measure for the realized sample-path variation termed the quadratic variation. By construction the quadratic variation cumulates the intensity of the unexpected price changes over the specific horizon and it is thus a prime candidate for a formal volatility measure.

The intuition behind the use of realized volatility as a return variation measure is most readily conveyed within the popular continuous-time diffusion setting (4.9) obtained by ruling out jumps and thus reducing to the representation (1.7), reproduced here for convenience,

$$dp(t) \;=\; \mu(t)\,dt \;+\; \sigma(t)\,dW(t)\,, \qquad t \in [0,T]. \tag{5.1}$$

Applying a discretization of the process as in Section 1, we have for small $\Delta > 0$, that

$$r(t,\Delta) \;\equiv\; p(t) - p(t-\Delta) \;\simeq\; \mu(t-\Delta)\Delta \;+\; \sigma(t-\Delta)\Delta W(t), \tag{5.2}$$

where $\Delta W(t) \equiv W(t) - W(t-\Delta) \sim N(0\,,\Delta)$.

Over short intervals the squared return and the squared return innovation are closely related as both are largely determined by the idiosyncratic return component,

$$r^2(t,\Delta) \;\simeq\; \mu^2(t-\Delta)\Delta^2 \;+\; 2\,\Delta\,\mu(t-\Delta)\,\sigma(t-\Delta)\Delta W(t) + \;\sigma^2(t-\Delta)(\Delta W(t))^2. \tag{5.3}$$

In particular, the return variance is (approximately) equal to the expected squared return innovation,

$$Var[\,r(t,\Delta)|\mathscr{F}_{t-\Delta}] \;\simeq\; E[\,r^2(t,\Delta)|\mathscr{F}_{t-\Delta}] \;\simeq\; \sigma^2(t-\Delta)\Delta. \tag{5.4}$$

This suggests that we may be able to measure the return volatility directly from the squared return observations. However, this feature is not of much direct use as the high-frequency returns have a large idiosyncratic component that induces a sizeable measurement error into the actual squared return relative to the underlying variance. Up to the dominant order in Δ,

$$Var[\,r^2(t,\Delta)|\mathscr{F}_{t-\Delta}] \;\simeq\; 2\,\sigma^4(t-\Delta)\,\Delta^2, \tag{5.5}$$

where terms involving higher powers of Δ are ignored as they become negligible for small values of Δ. Thus, it follows that the "noise-to-signal" ratio in squared returns relative to the underlying volatility is of the same order as volatility itself,

$$\frac{Var[r^2(t,\Delta)|\mathscr{F}_{t-\Delta}]}{E[r^2(t,\Delta)|\mathscr{F}_{t-\Delta}]} \simeq 2\,E[r^2(t,\Delta)|\mathscr{F}_{t-\Delta}]. \tag{5.6}$$

This relationship cannot be circumvented when only one (squared) return observation is used as a volatility proxy. Instead, by exploiting the fact that return innovations, under a no-arbitrage (semi-martingale) assumption, are serially uncorrelated to construct volatility measures for lower frequency returns we find, to dominant order in Δ,

$$\sum_{j=1}^{1/\Delta} E[r^2(t-1+j\cdot\Delta,\Delta)|\mathscr{F}_{t-1+j\cdot\Delta}] \simeq \sum_{j=1}^{1/\Delta} \sigma^2(t-1+j\cdot\Delta)\cdot\Delta \simeq \int_{t-1}^{t} \sigma^2(s)\,ds, \tag{5.7}$$

where the last approximation stems from the sum converging to the corresponding integral as the size of Δ shrinks toward zero. Equation (5.7) generalizes (5.4) to the multi-period setting with the second approximation in (5.7) only being meaningful for Δ small.

The advantage of (5.7) is that the uncorrelated "measurement errors" have been effectively smoothed away to generate a much better noise-to-signal ratio. The expression in (5.5) may be extended in a similar manner to yield,

$$\sum_{j=1}^{1/\Delta} Var[r^2(t-1+j\cdot\Delta,\Delta)|\mathscr{F}_{t-1+j\cdot\Delta}] \simeq 2\sum_{j=1}^{1/\Delta} \sigma^4(t-1+j\cdot\Delta)\cdot\Delta^2 \simeq 2\Delta\int_{t-1}^{t} \sigma^4(s)\,ds. \tag{5.8}$$

Consequently,

$$\frac{\sum_{j=1}^{1/\Delta} Var[r^2(t-1+j\cdot\Delta,\Delta)|\mathscr{F}_{t-1+j\cdot\Delta}]}{\sum_{j=1}^{1/\Delta} E[r^2(t-1+j\cdot\Delta,\Delta)|\mathscr{F}_{t-1+j\cdot\Delta}]} \simeq 2\,\Delta\,\frac{\int_{t-1}^{t} \sigma^4(s)\,ds}{\int_{t-1}^{t} \sigma^2(s)\,ds} \equiv 2\,\Delta\,\frac{IQ(t)}{IV(t)}, \tag{5.9}$$

where the integrated quarticity is defined through the identity on the right hand side of (5.9), with the integrated variance, IV(t), having previously been defined in (4.12).

The fact that the "noise-to-signal" ratio in (5.9) shrinks to zero with Δ suggests that high-frequency returns may be very useful for estimation of the underlying (integrated) volatility process. The notion of realized volatility is designed to take advantage of these features. Formally, realized volatility is defined as,

$$RV(t,\Delta) \;=\; \sum_{j=1}^{1/\Delta} r^2(t-1+j\cdot\Delta,\Delta). \tag{5.10}$$

Equation (5.8) suggests that realized volatility is consistent for the integrated volatility in the sense that finer and finer sampling of the intraday returns, $\Delta \to 0$, ultimately will annihilate the measurement error and, in the limit, realized volatility measures the latent integrated volatility perfectly, that is,

$$RV(t,\Delta) \;\to\; IV(t)\,, \tag{5.11}$$

as $\Delta \to 0$. These arguments may indeed by formalized; see, e.g., the extended discussion in Andersen, Bollerslev and Diebold (2003a). In reality, there is a definite lower bound on the return horizon that can be used productively for computation of the realized volatility, both because we only observe discretely sampled returns and, more important, market microstructure features such as discreteness of the price grid and bid-ask spreads induce gross violations of the semi-martingale property at the very highest return frequencies. This implies that we typically will be sampling returns at an intraday frequency that leaves a non-negligible error term in the estimate of integrated volatility. It is natural to conjecture from (5.9) that asymptotically, as $\Delta \to 0$,

$$\sqrt{1/\Delta}\;[RV(t,\Delta)-IV(t)] \;\sim\; N(0,2\cdot IQ(t)), \tag{5.12}$$

which turns out to be true under quite general assumptions. Of course, the $IQ(t)$ measure must be estimated as well for the above result to provide a practical tool for inference. The distributional result in (5.12) and a feasible consistent estimator for $IQ(t)$ based purely on intraday data have been provided by Barndorff-Nielsen and Shephard (2002, 2004b). It may further be shown that these measurement errors are approximately uncorrelated across consecutive periods which has important simplifying implications for time series modeling.

The consistency result in (5.11) extends to the general semi-martingale setting where the price path may display discontinuities due to jumps, as specified in equation (4.9). The realized volatility will still converge in the continuous-record limit ($\Delta \to 0$) to the period-by-period quadratic variation of the semi-martingale. However, the quadratic variation is no longer identical to the integrated volatility but will also include the cumulative squared jumps,

$$RV(t,\Delta) \rightarrow QV(t) = \int_{t-1}^{t} \sigma^2(s)\,ds + \sum_{t-1<s\le t} \kappa^2(s). \tag{5.13}$$

A few comments are in order. First, *QV(t)* is best interpreted as the actual return variation that transpired over the period, and as such it is the natural target for realized volatility measurement. Second, *QV(t)* is the realization of a random variable which generally cannot be forecasted with certainty at time *t-1*. But it does represent the future realization that volatility forecasts for time *t* should be compared against. In other words, the quadratic variation constitutes *the* quantity of interest in volatility measurement and forecasting. Since the realizations of *QV(t)* are latent, it is natural to use the observed *RV(t,Δ)* as a feasible proxy. Third, financial decision making is concerned with forecasts of volatility (or quadratic variation) rather than the *QV(t)* directly. Fourth, the identification of forecasts of return volatility with forecasts of quadratic variation is only approximate as it ignores variation in the process induced by innovations in the conditional mean process. Over short horizons the distinction is negligible, but for longer run volatility prediction (quarterly or annual) one may need to pay some attention to the discrepancy between the two quantities, as discussed at length in Andersen, Bollerslev and Diebold (2003a).

The distribution theory for quadratic variation under the continuous sample path assumption has also been extended to cover cumulative absolute returns raised to an arbitrary power. The leading case involves cumulating the high-frequency absolute returns. These quantities display improved robustness properties relative to realized volatility as the impact of outliers are mitigated. Although the limiting quantity - the power variation - is not directly linked to the usual volatility measure of interest in finance, this concept has inspired further theoretical developments that has led to intriguing new nonparametric tests for the presence of jumps and the identification of the associated jump sizes, see, e.g., Barndorff-Nielsen and Shephard (2004a). Since the jumps may have very different intertemporal persistence characteristics than the diffusion volatility, explicit disentangling of the components of quadratic variation corresponding to jumps versus diffusion volatility can have important implications for volatility forecasting.

In summary, the notion of realized volatility represents a model-free approach to (continuous-record) consistent estimation of the quadratic return variation under general assumptions based primarily upon arbitrage-free financial markets. As such it allows us to harness the information inherent in high-frequency returns for assessment of lower frequency return volatility. It is thus *the* natural approach to measuring actual (ex post) realized return variation over a given horizon. This perspective has now gained widespread acceptance in the literature, where alternative volatility forecast models are routinely assessed in terms of their ability to explain the distribution of subsequent realized volatility, as defined above.

5.2 Realized Volatility Modeling

The realized volatility is by construction an observed proxy for the underlying quadratic variation and the associated (measurement) errors are uncorrelated. This suggests a

straightforward approach where the temporal features of the series are modeled through standard time series techniques, letting the data guide the choice of the appropriate distributional assumptions and the dynamic representation. This is akin to the standard procedure for modeling macroeconomic data where the underlying quantities are measured (most likely with a substantial degree of error) and then treated as directly observed variables.

The strategy of estimating time series models directly for realized volatility is advocated in a sequence of papers by Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys, henceforth ABDL, (2001, 2003). A striking finding is that the realized volatility series share fundamental statistical properties across different asset classes, time periods, and countries. The evidence points strongly toward a long memory type of dependency in volatility. Moreover, the logarithmic realized volatility series is typically much closer to being homoskedastic and approximately unconditionally Gaussian. These features are readily captured through an ARFIMA(p,d,0) representation of the logarithmic realized volatility,

$$\Phi(L)\,(1-L)^d\,(\log RV(t,\Delta) - \mu_0) \;=\; u_t\,, \qquad t = 1,2, \ldots T, \tag{5.14}$$

where $(1-L)^d$ denotes the fractional differencing operator, $\Phi(L)$ is a polynomial lag operator accounting for standard autoregressive structure, μ_0 represents the unconditional mean of the logarithmic realized volatility, and u_t is a white noise error term that is (approximately) Gaussian. The coefficient d usually takes a value around 0.40, consistent with a stationary but highly persistent volatility process for which shocks only decay at a slow hyperbolic rate rather than the geometric rate associated with standard ARMA models or GARCH models for the conditional variance. Finally, the volatility of volatility is strongly increasing in the level of volatility as log realized volatility is approximately homoskedastic. This is, of course, reminiscent of the log SV and the EGARCH models.

A number of practical modeling issues have been sidestepped above. One is the choice of the sampling frequency at which the realized volatility measures are constructed. The early literature focused primarily on determining the highest intraday frequency at which the underlying returns satisfy the maintained semi-martingale assumption of being approximately uncorrelated. An early diagnostic along these lines termed the "volatility signature plot" was developed by ABDL (1999, 2000), as discussed further in Section 7 below. A simple alternative is to apply standard ARMA filtering to the high-frequency returns in order to strip them of any "artificial" serial correlation induced by the market microstructure noise, and then proceed with the filtered uncorrelated returns in lieu of the raw high-frequency returns. While none of these procedures are optimal in a formal statistical sense, they both appear to work reasonable well in many practical situations. Meanwhile, a number of alternative more efficient sampling schemes under various assumptions about the market microstructure complications have recently been proposed in a series of interesting papers, and this is still very much ongoing research.

A second issue concerns the potential separation of jumps and diffusive volatility components in the realized volatility process. The theoretical basis for these procedures and some initial empirical work is presented in Barndorff-Nielsen and Shephard (2004a). The issue has been

pursued empirically by Andersen, Bollerslev and Diebold (2003b), who find compelling evidence that the diffusive volatility is much more persistent than the jump component. In fact, the jumps appear close to *i.i.d.*, although the jumps in equity indices display some clustering, especially in the size of the jumps. This points to potentially important improvements in modeling and forecasting from this type of separation of the realized volatility into sudden discrete shifts in prices versus more permanent fluctuations in the intensity of the regular price movements. Empirically, this is in line with the evidence favoring non-Gaussian fat-tailed return innovations in ARCH models.

A third issue is the approach used to best accommodate the indications of "long memory." An alternative to fractional integration is to introduce several autoregressive volatility components into the model. As discussed in the context of the GARCH class of models in Section 3.4, if the different components display strong, but varying, degrees of persistence they may combine to produce a volatility dependency structure that is indistinguishable from long memory over even relatively long horizons.

5.3 Realized Volatility Forecasting

Forecasting is straightforward once the realized volatility has been cast within the traditional time series framework and the model parameters have been estimated. Since the driving variable is the realized volatility we no longer face a latent variable issue. This implies that standard methods for forecasting a time series within the ARFIMA framework is available; see, e.g., Beran (1994) for an introduction to models incorporating long memory features. One-step-ahead minimum mean-squared error forecasts are readily produced, and within the linear Gaussian setting it is then legitimate to further condition on the forecast in order to iterate forward and produce multiple-step-ahead forecasts. There are a couple of caveats, however. First, as with most other volatility forecasting procedures, the forecasts are, of course, conditional on the point estimate for the model parameters. Second, if the model is formulated in terms of the logarithmic volatility then it is also log volatility that is being predicted through the usual forecast procedures. There is a practical problem of converting the forecast for log volatility into a "pure" volatility forecast as the expected value of the transformed variable depends not only on the expected log volatility, but on the entire multiple-step-ahead conditional distribution of log volatility. For short horizons this is not an issue as the requisite correction term usually is negligible, but for longer horizons adjustments may be necessary. This is similar to the issue that arise in the construction of forecast form the EGARCH model. As discussed in Section 3.6, the required correction term may be constructed by simulation based methods, but the preferred approach will depend on the application at hand and the distributional characteristics of the model. For additional inspiration on how to address such issues consult, e.g., the chapter on ARMA forecasting methods by Lütkepohl (2003).

A few additional comments are in order. First, the evidence in ABDL (2003a) indicates that the above approach has very good potential. The associated forecasts for foreign exchange rate volatility outperform a string of alternative candidate models from the literature. This is not a tautology as it should be preferable to generate the forecasts from the true underlying model

rather than an ad hoc time series model estimated from period-by-period observations of realized volatility. In other words, if a GARCH diffusion is the true model then optimal forecasts would incorporate the restrictions implied by this model. However, the high-frequency volatility process is truly complex, possessing several periodic components, erratic short run dynamics and longer run persistence features that combined appear beyond reach of simple parametric models. The empirical evidence suggests that daily realized volatility serves as a simple, yet effective, aggregator of the volatility information inherent in the intraday data.

Second, there is an issue of how to compute realized volatility for a calendar period when the trading day is limited by an official closing. This problem is minor for the over-the-counter foreign exchange market where 24-hour trading is observed, but this is often not the case for equity or bond markets. For example, for a one-month-ahead equity volatility forecast there may only be twenty-two trading days with about six-and-a-half hours of trading per day. But the underlying price process is not stalled while the markets are closed. Oftentimes there will be substantial changes in prices between one market close and the subsequent opening, reflecting return volatility overnight and over the weekend. One solution is to simply rely on the intraday returns for a realized volatility measure over the trading day and then scale this quantity up by a factor that reflects the average ratio of volatility over the calendar day versus the trading day. This may work quite satisfactorily in practice, but it obviously ignores the close-to-open return for a given day entirely in constructing the realized volatility for that calendar day. Alternatively, the volatility of the close-to-open return may be modeled by a conventional GARCH type model.

Third, we have not discussed the preferred sampling frequency of intraday returns in situations where the underlying asset is relatively illiquid. If updated price observations are only available intermittently throughout the trading day, many high-frequency returns may have to be computed from prices or quotes earlier in the day. This brings up a couple of issues. One, the effective sampling frequency is lower than the one that we are trying to use for the realized volatility computation. Two, illiquid price series also tend to have larger bid-ask spreads and be more sensitive to random fluctuations in order flow, implying that the associated return series will contain a relatively large amount of noise. A simple response that will help alleviate both issues is to lower the sampling frequency. However, with the use of less intraday returns comes a larger measurement error in realized volatility, as evidenced by equation (5.12). Nonetheless, for an illiquid asset it may only be possible to construct meaningful weekly rather than daily realized volatility measures from say half-hourly or hourly return observations rather than five-minute returns. Consequently, the intertemporal fluctuations are smoothed out so that the observed measure carries less information about the true state of the volatility at the end of the period. This, of course, can be critically important for accurate forecasting.

In sum, the use of the realized volatility measures for forecasting is still in its infancy and many issues must be explored in future work. However, it is clear that the use of intraday information has large potential to improve upon the performance of standard volatility forecast procedures based only on daily or lower frequency data. The realized volatility approach circumvents the need to model the intraday data directly and thus provides a great deal of simplification. Importantly, it seems to achieve this objective without sacrificing a lot of efficiency. For

example, Andersen, Bollerslev and Meddahi (2004) find the time series approach built directly from the realized volatility measures to be very good approximations to the theoretically optimal procedures in a broad class of SV diffusion models that can be analyzed analytically through newly developed tools associated with the so-called Eigenfunction SV models of Meddahi (2001). Nonetheless, if the objective exclusively is volatility forecasting, some very recent work suggests that alternative intraday measures may carry even more empirically relevant information regarding future volatility, including the power variation measures constructed from cumulative absolute returns; see, e.g., Ghysels, Santa-Clara and Valkanov (2004). This likely reflects superior robustness features of absolute versus squared intraday returns, but verification of such conjectures awaits future research. The confluence of compelling empirical performance, novel econometric theory, the availability of ever more high-frequency data and computational power, and the importance of forecast performance for decision making render this approach fertile ground for new research.

5.4 Further Reading

The realized volatility approach has a precedent in the use of cumulative daily squared returns as monthly volatility measures, see, e.g., French, Schwert and Stambaugh (1987) and Schwert (1989). Hsieh (1989) was among the first to informally apply this same procedure with high-frequency intraday returns, while Zhou (1996) provides one of the earliest formal assessments of the relationship between cumulative squared intraday returns and the underlying return variance, albeit in a highly stylized setting. The pioneering work by Olsen & Associates on the use of high-frequency data, as summarized in Dacorogna et al. (2001), also importantly paved the way for many of the more recent empirical developments in the realized volatility area.

The use of component structures and related autoregressive specifications for approximating long-memory dependencies within the realized volatility setting has been explored by Andersen, Bollerslev and Diebold (2003b), Barndorff-Nielsen and Shephard (2001), Bollerslev and Wright (2001), and Corsi (2003), among others. The finite sample performance of alternative non-parametric tests for jumps based on the bipower variation measure introduced by Barndorff-Nielsen and Shephard (2004a) have been extensively analyzed by Huang and Tauchen (2004). Andersen, Bollerslev and Diebold (2003b) demonstrate the importance of disentangling the components of quadratic variation corresponding to jumps versus diffusion volatility for volatility forecasting. The complexities involved in a direct high-frequency characterization of the volatility process is also illustrated by Andersen and Bollerslev (1998c).

Ways of incorporating noisy overnight returns into the daily realized volatility measure are discussed in Fleming, Kirby and Ostdiek (2003) and Hansen and Lunde (2004a). The related issue of measuring the integrated variance in the presence of market microstructure noise and how to best use all of the available high frequency data has been addressed in a rapidly growing recent literature. Corsi, Zumbach, Müller and Dacorogna (2001) argue for the use of exponential moving average filtering, similar to a standard MA(1) filter for the high-frequency returns, while other more recent procedures, including sub-sampling and ways of choosing the "optimal" sampling frequency, have been suggested and analyzed empirically by, e.g., Aït-Sahalia,

Mykland and Zhang (2005), Bandi and Russell (2004), Barucci and Reno (2002), Bollen and Inder (2002), Curci and Corsi (2004), and Hansen and Lunde (2004b), among others. Some of these issues are discussed further in Section 7 below, where we also consider the robust alternative range based volatility estimator recently explored by Alizadeh, Brandt and Diebold (2002) for dynamic volatility modeling and forecasting.

Implied volatility provides yet another forward looking volatility measure. Implied volatilities are based on the market's forecasts of future volatilities extracted from the prices of options written on the asset of interest. As discussed in Section 2.2.4 above, using a specific option pricing formula, one may infer the expected integrated volatility of the underlying asset over the remaining time-to-maturity of the option. The main complication associated with the use of these procedures lies in the fact that the option prices also generally reflect a volatility risk premium in the realistic scenario where the volatility risk cannot be perfectly hedged; see, e.g., the discussion in Bollerslev and Zhou (2005). Nonetheless, many studies find options implied volatilities to provide useful information regarding the future volatility of the underlying asset. At the same time, the results pertaining to the forecast performance of implied volatilities are somewhat mixed, and there is still only limited evidence regarding the relative predictive power of implied volatilities versus the realized volatility procedures discussed above. Another issue is that many assets of interest do not have sufficiently active options markets that reliable implied volatilities can be computed on, say, a daily basis.

6. Multivariate Volatility

The discussion in the preceding three sections has been focused almost exclusively on univariate forecasts. Yet, as discussed in Section 2, in many practical situations covariance and/or correlation forecasting plays an equal, if not even more important, role in the uses of volatility forecasts. Fortunately, many of the same ideas and procedures discussed in the context of univariate forecasts are easily adapted to the multivariate setting. However, two important complications arise in this setting, namely the imposition of sufficient conditions to ensure that the forecasts for the covariance matrix remain positive definite for all forecasting horizons, and, second, maintaining an empirically realistic yet parsimoniously parameterized model. We will organize our discussion of the various multivariate approaches with these key concerns in mind.

Before turning to this discussion, it is worth noting that in many situations, multivariate volatility modeling and forecasting may be conveniently sidestepped through the use of much-simpler-to-implement univariate procedures for appropriately transformed series. In particular, in the context of financial market volatility forecasting, consider the leading case involving the variance of a portfolio made up of N individual assets. In the notation of Section 2.2.1 above,

$$r_{w,t+1} = \sum_{i=1}^{N} w_{i,t} r_{i,t+1} \equiv w_t' R_{t+1} . \tag{6.1}$$

The conditional one-step-ahead variance of the portfolio equals,

$$\sigma^2_{w,t+1|t} = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{i,t}w_{j,t}\{\Omega_{t+1|t}\}_{i,j} = w_t^{\prime}\Omega_{t+1|t}w_t \ , \tag{6.2}$$

where $\Omega_{t+1|t}$ denotes the $N\times N$ covariance matrix for the returns. A forecast for the portfolio return variance based upon this representation therefore requires the construction of multivariate forecasts for the *½N(N+1)* unique elements in the covariance matrix for the assets in the portfolio. Alternatively, define the univariate time series of artificial historical portfolio returns constructed on the basis of the weights for the current portfolio in place,

$$r^{t}_{w,\tau} \equiv w_t^{\prime}R_\tau \qquad \tau = 1, 2, \ldots, t. \tag{6.3}$$

A univariate forecast for the variance of the returns on this artificially constructed portfolio indirectly ensures that the covariances among the individual assets receive exactly the same weight as in equation (6.2). Note, that unless the portfolio weights for the actual portfolio in place are constantly re-balanced, the returns on this artificially constructed portfolio will generally differ from the actual portfolio returns, that is $r^{t}_{w,\tau} \equiv w_t^{\prime}R_\tau \neq w_\tau^{\prime}R_\tau \equiv r_{w,\tau}$ for $\tau\neq t$. As such, the construction of the variance forecasts for $r^{t}_{w,\tau}$ requires the estimation of a new (univariate) model each period to properly reflect the relevant portfolio composition in place at time *t*. Nonetheless, univariate volatility models are generally much easier to implement than their multivariate counterparts, so that this approach will typically be much less computationally demanding than the formulation of a satisfactory full scale multivariate volatility model for $\Omega_{t+1|t}$, especially for large values of *N*. Moreover, since the relative changes in the actual portfolio weights from one period to the next are likely to be small, good starting values for the parameters in the period-by-period univariate models are readily available from the estimates obtained in the previous period. Of course, this simplified approach also requires that historical returns for the different assets in the portfolio are actually available. If that is not the case, artificial historical prices could be constructed from a pricing model, or by matching the returns to those of other assets with similar characteristics; see, e.g., Andersen, Bollerslev, Christoffersen and Diebold (2005) for further discussion along these lines.

Meanwhile, as discussed in Sections 2.2.2 and 2.2.3, there are, of course, many situations in which forecasts for the covariances and/or correlations play a direct and important role in properly assessing and comparing the risks of different decisions or investment opportunities. We next turn to a discussion of some of the multivariate models and forecasting procedures available for doing so.

6.1 Exponential Smoothing and RiskMetrics

The exponentially weighted moving average filter, championed by RiskMetrics, is arguable the most commonly applied approach among finance practitioners for estimating time-varying

covariance matrices. Specifically, let $Y_t \equiv R_t$ denote the $N \times 1$ vector of asset returns. The estimate for the current covariance matrix is then defined by,

$$\hat{\Omega}_t = \gamma Y_t Y_t' + (1 - \gamma)\hat{\Omega}_{t-1} \equiv \gamma \sum_{i=1}^{\infty} (1 - \gamma)^{i-1} Y_t Y_t'. \tag{6.4}$$

This directly parallels the earlier univariate definition in equation (3.2), with the additional assumption that the mean of all the elements in Y_t is equal to zero. As in the univariate case, practical implementation is typically done by truncating the sum at $I = t\text{-}1$, scaling the finite sum by $1/[1-(1-\gamma)^t]$. This approach is obviously very simple to implement in any dimension N, involving only a single tuning parameter, γ, or by appealing to the values advocated by RiskMetrics (*0.06* and *0.04* in the case of daily and monthly returns, respectively) no unknown parameters whatsoever. Moreover, the resulting covariance matrix estimates are guaranteed to be positive definite.

The simple one-parameter filter in (6.4) may, of course, be further refined by allowing for different decay rates for the different elements in $\hat{\Omega}_t$. Specifically, by using a smaller value of γ for the off-diagonal, or covariance, terms in $\hat{\Omega}_t$, the corresponding time-varying correlations,

$$\hat{\rho}_{ij,t} \equiv \frac{\{\hat{\Omega}_t\}_{ij}}{\{\hat{\Omega}_t\}_{ii}^{1/2} \{\hat{\Omega}_t\}_{jj}^{1/2}}, \tag{6.5}$$

will exhibit more persistent dynamic dependencies. This slower rate of decay for the correlations often provide a better characterization of the dependencies across assets.

Meanwhile, the h-period-ahead forecasts obtained by simply equating the future conditional covariance matrix with the current filtered estimate,

$$Var(Y_{t+h} \mid \mathscr{F}_t) \equiv \Omega_{t+h|t} \approx \hat{\Omega}_t, \tag{6.6}$$

are plagued by the same counterfactual implications highlighted in the context of the corresponding univariate filter in Sections 3.1 and 3.2. In particular, assuming that the one-period returns are serially uncorrelated so that the forecast for the covariance matrix of the multi-period returns equals the sum of the successive one-period covariance forecasts,

$$Var(Y_{t+k} + Y_{t+k-1} + \ldots + Y_{t+1} \mid \mathscr{F}_t) \equiv \Omega_{t:t+k|t} \approx k\hat{\Omega}_t, \tag{6.7}$$

the multi-period covariance matrix scales with the forecast horizon, k, rather than incorporating

empirically more realistic mean-reversion. Moreover, it is difficult to contemplate the choice of the tuning parameter(s), γ, for the various elements in $\hat{\Omega}_t$ without a formal model. The multivariate GARCH class of models provides an answer to these problems by formally characterizing the temporal dependencies in the forecasts for the individual variances and covariances within a coherent statistical framework.

6.2 Multivariate GARCH Models

The multivariate GARCH class of models was first introduced and estimated empirically by Bollerslev, Engle and Wooldridge (1998). Denoting the one-step-ahead conditional mean vector and covariance matrix for Y_t by $\mathbf{M}_{t|t-1} \equiv E(Y_t \mid \mathscr{F}_{t-1})$ and $\Omega_{t|t-1} \equiv Var(Y_t \mid \mathscr{F}_{t-1})$, respectively, the multivariate version of the decomposition in (3.5) may be expressed as,

$$Y_t = \mathbf{M}_{t|t-1} + \Omega_{t|t-1}^{1/2} Z_t \qquad Z_t \sim i.i.d. \quad E(Z_t) = 0 \quad Var(Z_t) = I, \tag{6.8}$$

where Z_t now denotes a vector white noise process with unit variances. The square root of the $\Omega_{t|t-1}$ matrix is not unique, but any operator satisfying the condition that $\Omega_{t|t-1}^{1/2} \cdot \Omega_{t|t-1}^{1/2} \equiv \Omega_{t|t-1}$ will give rise to the same conditional covariance matrix.

The multivariate counterpart to the successful univariate GARCH(1,1) model in (3.6) is now naturally defined by,

$$vech(\Omega_{t|t-1}) = C + A\,vech(e_{t-1}e_{t-1}') + B\,vech(\Omega_{t-1|t-2}), \tag{6.9}$$

where $e_t \equiv \Omega_{t|t-1}^{1/2} Z_t$, $vech(\cdot)$ denotes the operator that stacks the $\frac{1}{2}N(N+1)$ unique elements in the lower triangular part of a symmetric matrix into a $\frac{1}{2}N(N+1)\times 1$ vector, and the parameter matrices C, A, and B, are of dimensions $\frac{1}{2}N(N+1)\times 1$, $\frac{1}{2}N(N+1)\times\frac{1}{2}N(N+1)$, and $\frac{1}{2}N(N+1)\times\frac{1}{2}N(N+1)$, respectively. As in the univariate case, the GARCH(1,1) model in (6.9) is readily extended to higher order models by including additional lagged terms on the right-hand-side of the equation. Note, that for $N=1$ the model in (6.9) is identical to formulation in (3.6), but for $N>1$ each of the elements in the covariance matrix is allowed to depend (linearly) on all of the other lagged elements in the conditional covariance matrix as well as the cross products of all the lagged innovations.

The formulation in (6.9) could also easily be extended to allow for asymmetric influences of past negative and positive innovations, as in the GJR or TGARCH model in (3.11), by including the signed cross-products of the residuals on the right-hand-side. The most straightforward generalized would be to simply include $vech(\min\{e_{t-1},0\}\min\{e_{t-1},0\}')$, but other matrices involving the cross-products of $\max\{e_{t-1},0\}$ and/or $\min\{e_{t-1},0\}$ have proven important in some empirical applications. Of course, other exogenous explanatory variables could be included in a similar fashion.

Meanwhile, multi-step-ahead forecasts for the conditional variances and covariances from the linear model in (6.9) are readily generated by recursive substitution in the equation,

$$vech(\Omega_{t+h|t+h-1}) = C + A\,vech(F_{t+h-1|t+h-2}) + B\,vech(\Omega_{t+h-1|t+h-2}) \tag{6.10}$$

where by definition,

$$F_{t+h|t+h-1} \equiv e_{t+h}e_{t+h}^{/} \qquad h \le 0$$

and,

$$F_{t+h|t+h-1} \equiv \Omega_{t+h|t+h-1} \qquad h \ge 1 .$$

These recursions, and their extensions to higher order models, are, of course, easy to implement on a computer. Also, provided that the norm of all the eigenvalues of $A+B$ are less than unity, the long-run forecasts for $\Omega_{t+h|t}$ will converge to the "unconditional covariance matrix" implied by the model, $(I-A-B)^{-1}C$, at the exponential rate of decay dictated by $(A+B)^h$. Again, these results directly mirror the univariate expressions in equations (3.8) and (3.9).

Still, nothing guarantees that the "unconditional covariance matrix" implied by (6.9), $(I-A-B)^{-1}C$, is actually positive definite, nor that the recursion in (6.10) results in positive definite h-step ahead forecasts for the future covariance matrices. In fact, without imposing any additional restrictions on the C, A, and B parameter matrices, the forecasts for the covariance matrices will most likely not be positive definite. Also, the unrestricted GARCH(1,1) formulation in (6.9) involves a total of $\frac{1}{2}N^4+N^3+N^2+\frac{1}{2}N$ unique parameters. Thus, for $N=5$ the model has *465* parameters, whereas for $N=100$ there is a total of *51,010,050* parameters! Needless to say, estimation of this many free parameters isn't practically feasible. Thus, various simplifications designed to ensure positive definitness and a more manageable number of parameters have been developed in the literature.

In the diagonal vech model the A and B matrices are both assumed to be diagonal, so that a particular element in the conditional covariance matrix only depends on its own lagged value and the corresponding cross product of the innovations. This model may alternatively be written in terms of Hadamard products, or element-by-element multiplication, as

$$\Omega_{t|t-1} = C + A\circ(e_{t-1}e_{t-1}^{/}) + B\circ\Omega_{t-1|t-2} , \tag{6.11}$$

where C, A, and B now denote symmetric positive definite matrices of dimension $N\times N$. This model greatly reduces the number of free parameters to $3(N^2+N)/2$, and, importantly, covariance matrix forecasts generated from this model according to the recursions in (6.10) are guaranteed to be positive definite. However, the model remains prohibitively "expensive" in terms of parameters in large dimensions. For instance, for $N=100$ there are still *15,150* free parameters in the unrestricted diagonal vech model.

A further dramatic simplification is obtained by restricting all of the elements in the A and B matrices in (6.11) to be the same,

$$\Omega_{t|t-1} = C + \alpha(e_{t-1}e_{t-1}') + \beta\Omega_{t-1|t-2}. \tag{6.12}$$

This scalar diagonal multivariate GARCH representation mimics the RiskMetrics exponential smoother in equation (6.4), except for the positive definite C matrix intercept, and the one additional smoothing parameter. Importantly however, provided that $\alpha+\beta<1$, the unconditional covariance matrix implied by the model in (6.12) equals $\Omega = (1-\alpha-\beta)^{-1}C$, and in parallel to the expression for the univariate GARCH(1,1) model in equation (3.9), the h-period forecasts mean reverts to Ω according to the formula,

$$\Omega_{t+h|t} = \Omega + (\alpha + \beta)^{h-1}(\Omega_{t+1|t} - \Omega).$$

This contrasts sharply with the RiskMetrics forecasts, which as previously noted show no mean reversion, with the counter-factual implication that the multi-period covariance forecasts for (approximately) serially uncorrelated returns scale with the forecast horizon. Of course, the scalar model in (6.12) could easily be refined to allow for different (slower) decay rates for the covariances by adding just one or two additional parameters to describe the off-diagonal elements. Still, the model is arguably too simplistic from an empirical perspective, and we will discuss other practically feasible multivariate models and forecasting procedures in the subsequent sections. Before doing so, however, we briefly discuss some of the basic principles and ideas involved in the estimation of multivariate GARCH models.

6.3 Multivariate GARCH Estimation

Estimation and inference for multivariate GARCH models may formally proceed along the same lines as for the univariate models discussed in Section 3.5. In particular, assume that the conditional distribution of Y_t is multivariate normal with mean, $\mathbf{M}_{t|t-1}$, and covariance matrix, $\Omega_{t|t-1}$. The log-likelihood function is given by the sum of the corresponding T logarithmic conditional normal densities,

$$logL(\theta; Y_T, \ldots, Y_1) = \tag{6.13}$$

$$-\frac{TN}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\log\Omega_{t|t-1}(\theta) - (Y_t - \mathbf{M}_{t|t-1}(\theta))'\Omega_{t|t-1}(\theta)^{-1}(Y_t - \mathbf{M}_{t|t-1}(\theta))\right],$$

where we have highlighted the explicit dependence on the parameter vector, θ. Provided that the assumption of conditional normality is true and the parametric models for the mean and covariance matrices are correctly specified, the resulting estimates, say $\hat{\theta}_T$, will satisfy the usual

optimality conditions associated with maximum likelihood. Moreover, even if the conditional normality assumption is violated, the resulting estimates may still be given a QMLE interpretation, with robust parameter inference based on the "sandwich-form" of the covariance matrix estimator, as discussed in Section 3.5.

Meanwhile, as discussed in Section 2, when constructing interval or VaR type forecasts, the whole conditional distribution becomes important. Thus, in parallel to the discussion in Sections 3.5 and 3.6, other multivariate conditional distributions may be used in place of the multivariate normal distributions underlying the likelihood function in (6.13). Different multivariate generalizations of the univariate fat-tailed student t distribution in (3.24) have proved quite successful for many daily and weekly financial rate of returns.

The likelihood function in (6.13), or generalizations allowing for conditionally non-normal innovations, may in principle be maximized by any of a number of different numerical optimization techniques. However, even for moderate values of N, say $N \geq 5$, the dimensionality of the problem for the general model in (6.9) or the diagonal vech model in (6.11) renders the computations hopelessly demanding from a practical perspective. As previously noted, this lack of tractability motivates the more parsimonious parametric specifications discussed below.

An alternative approach for circumventing the curse-of-dimensionality within the context of the diagonal vech model has recently been advocated by Ledoit, Santa-Clara and Wolf (2003). Instead of estimating all of the elements in the C, A and B matrices jointly, inference in their Flex GARCH approach proceed by estimating separate bivariate models for all of the possible pairwise combinations of the N elements in Y_t. These individual matrix estimates are then "pasted" together to a full dimensional model in such a way that the resulting $N \times N$ matrices in (6.11) are ensured to be positive definite.

Another practical approach for achieving more parsimonious and empirically meaningful multivariate GARCH forecasting models rely on so-called variance targeting techniques. Specifically, consider the general multivariate formulation in (6.9) obtained by replacing C with

$$C = (I - A - B)vech(V), \tag{6.14}$$

where V denotes a positive definite matrix. Provided that the norm of all the eigenvalues for $A+B$ are less than unity, so that the inverse of $(I-A-B)$ exists, this re-parameterization implies that the long-run forecasts for $\Omega_{t+h|t}$ will converge to V for $h \to \infty$. As such, variance targeting can help ensure that the long-run forecasts are well behaved. Of course, this doesn't reduce the number of unknown parameters in the model per se, as the long-run covariance matrix, V, must now be determined. However, an often employed approach is to fix V at the unconditional sample covariance matrix,

$$\hat{V} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{M}_{t|t-1})(Y_t - \hat{M}_{t|t-1})',$$

where $\hat{\mathbf{M}}_{t|t-1}$ denotes some first-stage estimate for the conditional mean. This estimation of V obviously introduces an additional source of parameter estimation error uncertainty, although the impact of this is typically ignored in practice when conducting inference about the other parameters entering the equation for the conditional covariance matrix.

6.4 Dynamic Conditional Correlations

One commonly applied approach for large scale dynamic covariance matrix modeling and forecasting is the Constant Conditional Correlation (CCC) model of Bollerslev (1990). Specifically, let $D_{t|t-1}$ denote the $N \times N$ diagonal matrix with the conditional standard deviations, or the square root of the diagonal elements in $\mathbf{\Omega}_{t|t-1} \equiv Var(Y_t \,|\, \mathscr{F}_{t-1})$, along the diagonal. The conditional covariance matrix may then be uniquely expressed in terms of the decomposition,

$$\mathbf{\Omega}_{t|t-1} = D_{t|t-1} \mathbf{\Gamma}_{t|t-1} D_{t|t-1}, \tag{6.15}$$

where $\mathbf{\Gamma}_{t|t-1}$ denote the $N \times N$ matrix of conditional correlations. Of course, this decomposition does not result in any immediate simplifications from a modeling perspective, as the conditional correlation matrix must now be estimated. However, following Bollerslev (1990) and assuming that the temporal variation in the covariances are driven solely by the temporal variation in the corresponding conditional standard deviations, so that the conditional correlations are constant,

$$\mathbf{\Gamma}_{t|t-1} \equiv \mathbf{\Gamma}, \tag{6.16}$$

dramatically reduces the number of parameters in the model relative to the linear vech specifications discussed above. Moreover, this assumption also greatly simplifies the multivariate estimation problem, which may now proceed in two steps. In the first step N individual univariate GARCH models are estimated for each of the series in Y_t, resulting in an estimate for the diagonal matrix, $\hat{D}_{t|t-1}$. Then defining the $N \times 1$ vector of standardized residuals for each of the univariate series,

$$\hat{\boldsymbol{\varepsilon}}_t \equiv \hat{D}_{t|t-1}^{-1} (Y_t - \hat{M}_{t|t-1}), \tag{6.17}$$

the elements in Γ may simply be estimated by the corresponding sample analogue,

$$\hat{\mathbf{\Gamma}} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t^{\prime}. \tag{6.18}$$

Importantly, this estimate for Γ is guaranteed to be positive definite with ones along the diagonal and all of the other elements between minus one and one. In addition to being simple to implement, this approach therefore has the desirable feature that as long as the individual

variances in $\hat{D}_{t|t-1}$ are positive, the resulting covariance matrices defined by (6.15) are guaranteed to be positive definite.

While the assumption of constant conditional correlations may often be a reasonable simplification over shorter time periods, it is arguable too simplistic in many situations of practical interest. To circumvent this, while retaining the key features of the decomposition in (6.15), Engle (2002) and Tse and Tsui (2002) have recently proposed a convenient framework for directly modeling any temporal dependencies in the conditional correlations. In the most basic version of the Dynamic Conditional Correlation (DCC) model of Engle (2002), the temporal variation in the conditional correlation is characterized by a simple scalar GARCH(1,1) model, along the lines of (6.12), with the covariance matrix for the standardized residuals targeted at their unconditional value in (6.18). That is,

$$Q_{t|t-1} = (1 - \alpha - \beta)\hat{\Gamma} + \alpha(\hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-1}') + \beta Q_{t-1|t-2}. \tag{6.19}$$

Although this recursion guarantees that the $Q_{t|t-1}$ matrices are positive definite, the individual elements are not necessarily between minus one and one. Thus, in order to arrive at an estimate for the conditional correlation matrix, the elements in $Q_{t|t-1}$ must be standardized, resulting in the following estimate for the *ij*'th correlation,

$$\hat{\rho}_{ij,t} \equiv \{\hat{\Gamma}_{t|t-1}\}_{ij} = \frac{\{Q_t\}_{ij}}{\{Q_t\}_{ii}^{1/2}\{Q_t\}_{jj}^{1/2}}. \tag{6.20}$$

Like the CCC model, the DCC model is also relatively simple to implement in large dimensions, requiring only the estimation of *N* univariate models along with a choice of the two exponential smoothing parameters in (6.19).

Richer dynamic dependencies in the correlations could be incorporated in a similar manner, although this immediately raises some of the same complications involved in directly parameterizing $\Omega_{t|t-1}$. However, as formally shown in Engle and Sheppard (2001), the parameters in (6.19) characterizing the dynamic dependencies in $Q_{t|t-1}$, and in turn $\Gamma_{t|t-1}$, may be consistently estimated in a second step by maximizing the partial log likelihood function,

$$logL(\theta; Y_T, \ldots, Y_1)^* = -\frac{1}{2}\sum_{t=1}^{T}\left[\log|\Gamma_{t|t-1}(\theta)| - \hat{\varepsilon}_t'\Gamma_{t|t-1}(\theta)^{-1}\hat{\varepsilon}_t\right],$$

where $\hat{\varepsilon}_t$ refers to the first step estimates defined in (6.17). Of course, the standard errors for the resulting correlation parameter estimates must be adjusted to take account of the first stage estimation errors in $\hat{D}_{t|t-1}$. Extensions of the basic DCC structure in (6.19) and (6.20) along these lines allowing for greater flexibility in the dependencies in the correlations across different types of assets, asymmetries in the way in which the correlations respond to past negative and positive return innovations, regime switches in the correlations, to name but a few, are currently being actively explored by a number of researchers.

6.5 Multivariate Stochastic Volatility and Factor Models

An alternative approach for achieving a more manageable and parsimonious multivariate volatility forecasting model entails the use of factor structures. Factor structures are, of course, central to the field of finance, and the Arbitrage Pricing Theory (APT) in particular. Multivariate factor GARCH and stochastic volatility models were first analyzed by Diebold and Nerlove (1989) and Engle, Ng and Rothschild (1990). To illustrate, consider a simple one-factor model in which the commonality in the volatilities across the $N\times 1$ R_t vector of asset returns is driven by a single scalar factor, f_t,

$$R_t = a + bf_t + e_t , \tag{6.21}$$

where a and b denote $N\times 1$ parameter vectors, and e_t is assumed to be *i.i.d.* through time with covariance matrix Λ. This directly captures the idea that variances (and covariances) generally move together across assets. Now, assuming that the factor is conditionally heteroskedastic, with conditional variance denoted by $\sigma^2_{t|t-1} \equiv Var(f_t \mid \mathscr{F}_{t-1})$, the conditional covariance matrix for R_t takes the form,

$$\Omega_{t|t-1} \equiv Var(R_t \mid \mathscr{F}_{t-1}) = bb'\sigma^2_{t|t-1} + \Lambda . \tag{6.22}$$

Compared to the unrestricted GARCH models discussed in Section 6.2, the factor GARCH representation greatly reduces the number of free parameters. Moreover, the conditional covariance matrix in (6.22) is guaranteed to be positive definite.

To further appreciate the implications of the factor representation, let b_i and λ_{ij} denote the i'th and ij'th element in b and Λ, respectively. It follows then directly from the expression in (6.22) that the conditional correlation between the i'th and the j'th observation is given by,

$$\rho_{ij,t} \equiv \frac{b_i b_j \sigma^2_{t|t-1} + \lambda_{ij}}{(b_i^2 \sigma^2_{t|t-1} + \lambda_{ii})^{1/2} (b_j^2 \sigma^2_{t|t-1} + \lambda_{jj})^{1/2}} . \tag{6.23}$$

Thus, provided that the corresponding factor loadings are of the same sign, or $b_i b_j > 0$, the conditional correlation implied by the model will increase toward unity as the volatility of the factor increases. That is, there is an empirically realistic built-in volatility-in-correlation effect.

Importantly, multivariate conditional covariance matrix forecasts are also readily constructed from forecasts for the univariate factor variance. In particular, assuming that the vector of returns is serially uncorrelated, the conditional covariance matrix for the k-period continuously compounded returns is simply given by,

$$\Omega_{t:t+k|t} \equiv Var(R_{t+k} + \ldots + R_{t+1} \mid \mathscr{F}_t) = bb'\sigma^2_{t:t+k|t} + k\Lambda, \tag{6.24}$$

where $\sigma^2_{t:t+k|t} \equiv Var(f_{t+k} + \ldots + f_{t+1} \mid \mathscr{F}_t)$. Further assuming that the factor is directly observable and that the conditional variance for f_t is specified in terms of the observable information set, $\mathscr{F}_{t-1}$, the forecasts for $\sigma^2_{t:t+k|t}$ may be constructed along the lines of the univariate GARCH class of models discussed in Section 3. If, on the other hand, the factor is latent or if the conditional variance for f_t is formulated in terms of unobservable information, $\mathfrak{I}_{t-1}$, one of the more intensive numerical procedures for the univariate stochastic volatility class of models discussed in Section 4 must be applied in calculating $\sigma^2_{t:t+k|t}$. Of course, the one-factor model in (6.21) could easily be extended to allow for multiple factors, resulting in obvious generalizations of the expressions in (6.22) and (6.24). As long as the number of factors remain small, the same appealing simplifications hold true.

Meanwhile, an obvious drawback from an empirical perspective to the simple factor model in (6.21) with homoskedastic innovations concerns the lack of heteroskedasticity in certain portfolios. Specifically, let $\Psi = \{\psi \mid \psi' b = 0, \psi \neq 0\}$ denote the set of $N \times 1$ vectors orthogonal to the vector of factor loadings, b. Any portfolio constructed from the N original assets with portfolio weights, $w = \psi / (\psi_1 + \ldots + \psi_N)$ where $\psi \in \Psi$, will then be homoskedastic,

$$Var(r_{w,t} \mid \mathscr{F}_{t-1}) \equiv Var(w'R_t \mid \mathscr{F}_{t-1}) = w'bb'w\sigma^2_{t|t-1} + w'\Lambda w = w'\Lambda w.$$

Similarly, the corresponding multi-period forecasts defined in (6.24) will also be time invariant. Yet, in applications with daily or weekly returns it is almost always impossible to construct portfolios which are void of volatility clustering effects. The inclusion of additional factors does not formally solve the problem. As long as the number of factors is less than N, the corresponding null-set Ψ is not empty. Of course, allowing the covariance matrix of the idiosyncratic innovations to be heteroskedastic would remedy this problem, but that then raises the issue of how to model the temporal variation in the $N \times N$ dimensional Λ_t matrix. One approach would be to include enough factors so that the Λ_t matrix may be assumed to be diagonal, only requiring the estimation of N univariate volatility models for the elements in e_t.

Whether the rank deficiency in the forecasts of the conditional covariance matrices from the basic factor structure and the counterfactual implication of no volatility clustering in certain

portfolios discussed above should be a cause for concern ultimately depends upon the uses of the forecasts. However, it is clear that the reduction in the dimension of the problem to a few systematic risk factors may afford great computational simplifications in the context of large scale covariance matrix modeling and forecasting.

6.6 Realized Covariances and Correlations

The high-frequency data realized volatility approach for measuring, modeling and forecasting univariate volatilities outlined in Section 5 may be similarly adapted to modeling and forecasting covariances and correlations. To set out the basic idea, let $R(t,\Delta)$ denote the $N\times 1$ vector of logarithmic returns over the $[t-\Delta,t]$ time interval,

$$R(t,\Delta) \equiv P(t) - P(t-\Delta). \tag{6.26}$$

The $N\times N$ realized covariation matrix for the unit time interval, $[t-1,t]$, is then formally defined by,

$$RCOV(t,\Delta) = \sum_{j=1}^{1/\Delta} R(t-1+j\cdot\Delta,\Delta)R(t-1+j\cdot\Delta,\Delta)'. \tag{6.27}$$

This directly parallels the univariate definition in (5.10). Importantly, the realized covariation matrix is symmetric by construction, and as long as the returns are linearly independent and $N<1/\Delta$, the matrix is guaranteed to be positive definite.

In order to more formally justify the realized covariation measure, suppose that the evolution of the $N\times 1$ vector price process may be described by the N dimensional continuous-time diffusion,

$$dP(t) = M(t)\,dt + \Sigma(t)\,dW(t), \quad t \in [0,T], \tag{6.28}$$

where $M(t)$ denotes the $N\times 1$ instantaneous drifts, $\Sigma(t)$ refer to the $N\times N$ instantaneous diffusion matrix, and $W(t)$ now denotes an $N\times 1$ dimensional vector of independent standard Brownian motions. Intuitively, for small values of $\Delta > 0$,

$$R(t,\Delta) \equiv P(t) - P(t-\Delta) \simeq \mathrm{M}(t-\Delta)\Delta + \Sigma(t-\Delta)\Delta W(t), \tag{6.29}$$

where $\Delta W(t) \equiv W(t) - W(t-\Delta) \sim N(0, \Delta I_N)$. Of course, this latter expression directly mirrors the univariate equation (5.2). Now, using similar arguments to the ones in Section 5.1, it follows that the multivariate realized covariation in (6.27) will converge to the corresponding multivariate integrated covariation for finer and finer sampled high-frequency returns, or $\Delta\to 0$,

$$RCOV(t,\Delta) \to \int_{t-1}^{t} \Sigma(s)\Sigma(s)'ds \equiv ICOV(t). \tag{6.30}$$

Again, by similar arguments to the ones in Section 5.1, the multivariate integrated covariation defined by the right-hand-side of equation (6.30) provides the true measure for the actual return variation and *covariation* that transpired over the *[t-1,t]* time interval. Also, extending the univariate results in (5.12), Barndorff-Nielsen and Shephard (2004b) have recently shown that the multivariate realized volatility errors, $\sqrt{1/\Delta}\ [RCOV(t,\Delta) - ICOV(t)]$, are approximately serially uncorrelated and asymptotically (for $\Delta \to 0$) distributed as a mixed normal with a random covariance matrix that may be estimated. Moreover following (5.13), the consistency of the realized covariation measure for the true quadratic covariation caries over to situations in which the vector price process contains jumps. As such, these theoretical results set the stage for multivariate volatility modeling and forecasting based on the realized covariation measures along the same lines as the univariate discussion in Sections 5.2 and 5.3.

In particular, treating the $\frac{1}{2}N(N+1)\times 1$ vector, $vech[RCOV(t,\Delta)]$, as a direct observation (with uncorrelated measurement errors) on the unique elements in the covariation matrix of interest, standard multivariate time series techniques may be used in jointly modeling the variances and the off-diagonal covariance elements. For instance, a simple VAR(1) forecasting model, analogues to the GARCH(1,1) model in (6.9), may be specified as,

$$vech[RCOV(t,\Delta)] \;=\; C \;+\; A\,vech[RCOV(t-1,\Delta)] \;+\; u_t, \tag{6.31}$$

where u_t denotes an $N\times 1$ vector white noise process. Of course, higher order dynamic dependencies could be included in a similar manner. Indeed, the results in Andersen, Bollerslev, Diebold and Labys (2001, 2003), suggest that for long-run forecasting it may be important to incorporate long-memory type dependencies in both variances and covariances. This could be accomplished through the use of a true multivariate fractional integrated model, or as previously discussed an approximating component type structure.

Even though $RCOV(t,\Delta)$ is positive definite by construction, nothing guarantees that the forecasts from an unrestricted multivariate time series model along the lines of the VAR(1) in (6.31) will result in positive definite covariance matrix forecasts. Hence, it may be desirable to utilize some of the more restrictive parameterizations for the multivariate GARCH models discussed in Section 6.2, to ensure positive definite covariance matrix forecasts. Nonetheless, replacing $\Omega_{t|t-1}$ with the directly observable $RCOV(t,\Delta)$, means that the parameters in the corresponding models may be estimated in a straightforward fashion using simple least squares, or some other easy-to-implement estimation method, rather than the much more numerically intensive multivariate MLE or QMLE estimation schemes.

Alternatively, an unrestricted model for the $\frac{1}{2}N(N+1)$ non-zero elements in the Cholesky decomposition, or lower triangular matrix square-root, of $RCOV(t,\Delta)$, could also be estimated. Of course, the non-linear transformation involved in such a decomposition means that the corresponding matrix product of the forecasts from the model will generally not be unbiased for the elements in the covariation matrix itself. Following Andersen, Bollerslev, Diebold and Labys (2003), sometimes it might also be possible to infer the covariances of interest from the

variances of different cross-rates or portfolios through appropriately defined arbitrage conditions. In those situations forecasts for the covariances may therefore be constructed from a set of forecasting models for the corresponding variances, in turn avoiding directly modeling any covariances.

The realized covariation matrix in (6.27) may also be used in the construction of realized correlations, as in Andersen, Bollerslev, Diebold and Labys (2001) and Andersen, Bollerslev, Diebold and Ebens (2001). These realized correlations could be modeled directly using standard time series techniques. However, the correlations are, of course, restricted to lie between minus one and one. Thus, to ensure that this constraint is not violated, it might be desirable to use the Fisher transform, $z = 0.5 \cdot \log[(1+\rho)/(1-\rho)]$, or some other similar transformation, to convert the support of the distribution for the correlations from *[-1,1]* to the whole real line. This is akin to the log transformation for the univariate realized volatilities employed in equation (5.14). Meanwhile, there is some evidence that the dynamic dependencies in the correlations between many financial assets and markets are distinctly different from that of the corresponding variances and covariances, exhibiting occasional "correlation breakdowns." These types of dependencies may best be characterized by regime switching type models. Rather than modeling the correlations individually, the realized correlation matrix could also be used in place of $\hat{e}_t \hat{e}_t'$ in the DCC model in (6.19), or some generalization of that formulation, in jointly modeling all of the elements in the conditional correlation matrix.

The realized covariation and correlation measures discussed above are, of course, subject to the same market micro structure complications that plague the univariate realized volatility measures discussed in Section 5. In fact, some of the problems are accentuated with the possibility of non-synchronous observations in two or more markets. Research on this important issues is still very much ongoing, and it is too early to draw any firm conclusions about the preferred method or sampling scheme to employ in the practical implementation of the realized covariation measures. Nonetheless, it is clear that the realized volatility approach afford a very convenient and powerful approach for effectively incorporating high-frequency financial data into both univariate and multivariate volatility modeling and forecasting.

6.7 Further Reading

The use of historical pseudo returns as a convenient way of reducing the multivariate modeling problem to a univariate setting, as outlined in Section 6.1, is discussed at some length in Andersen, Bollerslev, Christoffersen and Diebold (2005). This same study also discusses the use of a smaller set of liquid base assets along with a factor structure as another computationally convenient way of reducing the dimension of time-varying covariance matrix forecasting for financial rate of returns.

The RiskMetrics, or exponential smoothing approach, for calculating covariances and associated Value-at-Risk measures is discussed extensively in Christoffersen (2003), Jorion (2000), and Zaffaroni (2004) among others. Following earlier work by DeSantis and Gerard (1997), empirically more realistic slower decay rates for the covariances in the context of exponential

smoothing has been successfully implemented by DeSantis, Litterman, Vesval and Winkelmann (2003).

In addition to the many ARCH and GARCH survey papers and book treatments listed in Section 3, the multivariate GARCH class of models has recently been surveyed by Bauwens, Laurent and Rombouts (2005). A comparison of some of the available commercial software packages for the estimation of multivariate GARCH models is available in Brooks, Burke and Persand (2003). Conditions for the covariance matrix forecasts for the linear formulations discussed in Section 6.2 to be positive definite was first established by Engle and Kroner (1995), who also introduced the so-called BEKK parameterization. Asymmetries, or leverage effects, within this same class of models were subsequently studied by Kroner and Ng (1998). The bivariate EGARCH model of Braun , Nelson and Sunier (1995) and the recent matrix EGARCH model of Kawakatsu (2005) offer alternative ways of doing so. The multivariate GARCH QMLE procedures outlined in Section 6.3 were first discussed by Bollerslev and Wooldridge (1992), while Ling and McAleer (2003) provide a more recent account of some of the subsequent important theoretical developments. The use of a fat tailed multivariate student t-distribution in the estimation of multivariate GARCH models was first considered by Harvey, Ruiz and Sentana (1992); see also Bauwens and Laurent (2005) and Fiorentini, Sentana and Calzolari (2003) for more recent applications of alternative multivariate non-normal distributions. Issues related to cross-sectional and temporal aggregation of multivariate GARCH and stochastic volatility models have been studied by Nijman and Sentana (1996) and Meddahi and Renault (2004).

Several empirical studies have documented important temporal dependencies in asset return correlations, including early contributions by Erb, Harvey and Viskanta (1994) and Longin and Solnik (1995) focusing on international equity returns. More recent work by Ang and Chen (2002) and Cappiello, Engle and Sheppard (2004) have emphasized the importance of explicitly incorporating assymmetries in the way in which the correlations respond to past negative and positive return shocks. Along these lines, Longin and Solnik (2001) report evidence in support of more pronounced dependencies following large (extreme) negative return innovations. A test for the assumption of constant conditional correlations underlying the CCC model discussed in Section 6.4 has been derived by Bera and Kim (2002). Recent work on extending the DCC model to allow for more flexible dynamic dependencies in the correlations, asymmetries in the responses to past negative and positive returns, as well as switches in the correlations across regimes, include Billio, Caporin and Gobbo (2003), Cappiello, Engle and Sheppard (2004), Franses and Hafner (2003), and Pelletier (2005). Guidolin and Timmermann (2005b) find large variations in the correlation between stock and bond returns across different market regimes defined as crash, slow growth, bull and recovery. Sheppard (2004) similarly finds evidence of business cycle frequency dynamics in conditional covariances.

The factor ARCH models proposed by Diebold and Nerlove (1989) and Engle, Ng and Rothschild (1990) have been used by Ng, Engle and Rothschild (1992) and Bollerslev and Engle (1993), among others, in modeling common persistence in conditional variances and covariances. Harvey, Ruiz and Shephard (1994) and King, Sentana and Wadhwani (1994) were among the first to estimate multivariate stochastic volatility models. More recent empirical

studies and numerically efficient algorithms for the estimation of latent multivariate volatility structures include Aguilar and West (2000), Fiorentini, Sentana and Shephard (2004) and Liesenfeld and Richard (2003). Issues related to identification within heteroskedastic factor models have been studied by Sentana and Fiorentini (2001). A recent insightful discussion of the basic features of multivariate stochastic volatility factor models, along with a discussion of their origins, is provided in Shephard (2004). The multivariate Markov-switching multifractal model of Calvet, Fisher and Thompson (2005) may also be interpreted as a latent factor stochastic volatility model with a closed form likelihood. Other related relatively easy-to-implement multivariate approaches include the two-step Orthogonal GARCH model of Alexander (2001), in which the conditional covariance matrix is determined by univariate models for a (small) set of the largest (unconditional) principal components.

The realized volatility approach discussed in Section 6.6 affords a simple practically feasible way for covariance and correlation forecasting in situations when high-frequency data is available. The formal theory underpinning this approach in the multivariate setting has been spelled out in Andersen, Bollerslev, Diebold and Labys (2003) and Barndorff-Nielsen and Shephard (2004b). A precursor to some of these results is provided by the alternative double asymptotic rolling regression based framework in Foster and Nelson (1996). The benefits of the realized volatility approach versus more conventional multivariate GARCH based forecasts in the context of asset allocation have been forcefully demonstrated by Fleming, Kirby and Ostdiek (2003). Meanwhile, the best way of actually implementing the realized covariation measures with high-frequency financial data subject to market micro structure frictions still remains very much of an open research question. In a very early paper, Epps (1979) first observed a dramatic drop in high-frequency based sample correlations among individual stock returns as the length of the return interval approached zero; see also Lundin, Dacorogna and Müller (1998). In addition to the many mostly univariate studies noted in Section 4, Martens (2003) provides a recent assessment and comparison of some of the existing ways for best alleviating the impact of market micro structure frictions in the multivariate setting, including the covariance matrix estimator of De Jong and Nijman (1997), the lead-lag adjustment of Scholes and Williams (1977), and the range-based covariance measure of Brandt and Diebold (2005).

The multivariate procedures discussed in this section are (obviously) not exhaustive of the literature. Other recent promising approaches for covariance and correlation forecasting include the use of copulas for conveniently linking univariate GARCH (e.g., Jondeau and Rockinger, 2005, and Patton, 2004) or realized volatility models; the use of shrinkage to ensure positive definitness in the estimation and forecasting of very large dimensional covariance matrices (e.g., Jagannathan and Ma, 2003, and Ledoit and Wolf, 2003); and forecast model averaging techniques (e.g., Pesaran and Zaffaroni, 2004). It remains to be seen which of these, if any, will be added to the standard multivariate volatility forecasting toolbox.

7. Evaluating Volatility Forecasts

The discussion up until this point has surveyed the most important univariate and multivariate

volatility models and forecasting procedures in current use. This section gives an overview of some of the most useful methods available for volatility forecast evaluation. The methods introduced here can either be used by an external evaluator or by the forecaster him/herself as a diagnostic tool on the forecasting model. A more general discussion and review of the forecast evaluation literature can be found in Diebold and Lopez (1996) and the chapter by West in this Handbook.

Below, we will first introduce a general loss function framework and then highlight the particular issues involved when forecasting volatility itself is the direct object of interest. We then discuss several other important forecasting situations where volatility dynamics are crucial, including Value-at-Risk, probability, and density forecasting.

7.1 Point Forecast Evaluation from General Loss Functions

Consider the general forecast loss function, $L(y_{t+1}, \hat{y}_{t+1|t})$, discussed in Section 2, in which the arguments are the univariate discrete-time real-valued stochastic variable, y_{t+1}, as well as its forecast, $\hat{y}_{t+1|t}$. From the optimization problem solved by the optimal forecast, $\hat{y}_{t+1|t}$ must solve the generic first order condition

$$E_t\left[\frac{\partial L(y_{t+1}, \hat{y}_{t+1|t})}{\partial \hat{y}}\right] = 0 . \tag{7.1}$$

The partial derivative of the loss function - the term inside the conditional expectation - is sometimes referred to as the generalized forecast error. Realizations of this partial derivative should fluctuate unpredictably around zero, directly in line with the standard optimality condition that regular forecasts display uncorrelated prediction errors.

Specifically, consider the situation in which we observe a sequence of out-of-sample forecasts and subsequent realizations, $\{y_{t+1}, \hat{y}_{t+1|t}\}_{t=1}^{T}$. A natural diagnostic on (7.1) is then given by the simple regression version of the conditional expectation, that is

$$\frac{\partial L(y_{t+1}, \hat{y}_{t+1|t})}{\partial \hat{y}} = a + b' x_t + \epsilon_{t+1} , \tag{7.2}$$

where x_t denotes a vector of candidate explanatory variables in the time t information set observed by the forecaster, $\mathscr{F}_t$, and b is a vector of regression coefficients. An appropriately calibrated forecast should then have $a = b = 0$, which can be tested using standard t- and F-tests properly robustified to allow for heteroskedasticity in the regression errors, ϵ_{t+1}. Intuitively, if a significant coefficient is obtained on a forecasting variable, which the forecaster should reasonably have known at time t, then the forecasting model is not optimal, as the variable in question could and should have been used to make the generalized forecast error variance lower

than it actually is.

If the forecasts arise from a known well-specified statistical model with estimated parameters then the inherent parameter estimation error should ideally be accounted for. This can be done using the asymptotic results in West and McCracken (1998) or the finite sample Monte Carlo tests in Dufour (2004). However, external forecast evaluators may not have knowledge of the details of the underlying forecasting model (if one exists) in which case parameter estimation error uncertainty is not easily accounted for. Furthermore, in most financial applications the estimation sample is typically fairly large rendering the parameter estimation error relatively small compared with other potentially more serious model specification errors. In this case standard (heteroskedasticity robust) t-tests and F-tests may work reasonably well. Note also that in the case of, say, *h*-day forecasts from a daily model, the horizon overlap implies that the first *h-1* autocorrelations will not be equal to zero, and this must be allowed for in the regression.

As an example of the general framework in (7.2), consider the case of quadratic loss, $L(y_{t+1},\hat{y}_{t+1|t}) = (y_{t+1} - \hat{y}_{t+1|t})^2$. In this situation

$$\frac{\partial L(y_{t+1},\hat{y}_{t+1|t})}{\partial \hat{y}} = -2(y_{t+1} - \hat{y}_{t+1|t}), \tag{7.3}$$

which suggests the forecast evaluation regression

$$(y_{t+1} - \hat{y}_{t+1|t}) = a + b'x_t + \epsilon_{t+1}. \tag{7.4}$$

While the choice of information variables to include in x_t is somewhat arbitrary, one obvious candidate does exist, namely the time *t* forecast itself. Following this idea and letting $x_t = \hat{y}_{t+1|t}$, results in the so-called Mincer and Zarnowitz (1969) regression, which can thus be viewed as a test of forecast optimality relative to a limited information set. We write

$$(y_{t+1} - \hat{y}_{t+1|t}) = a + b\hat{y}_{t+1|t} + \epsilon_{t+1},$$

or equivalently

$$y_{t+1} = a + (b+1)\hat{y}_{t+1|t} + \epsilon_{t+1} \quad . \tag{7.5}$$

Clearly the ex-ante forecast should not be able to explain the ex post forecast error. For example, if *b* is significantly negative, and thus $(b+1)<1$, then the forecast is too volatile relative to the subsequent realization and the forecast should be scaled down.

It is often of interest to compare forecasts from different models, or forecasters. This is easily done by letting $x_t = [\hat{y}_{t+1|t} \quad \hat{y}_{A,t+1|t}]$, where $\hat{y}_{A,t+1|t}$ denotes the alternative forecast. The forecast

evaluation regression then takes the form,

$$y_{t+1} = a + (b+1)\hat{y}_{t+1|t} + b_A \hat{y}_{A,t+1|t} + \epsilon_{t+1}, \tag{7.6}$$

where a failure to reject the hypothesis that $b_A = 0$ implies that the additional information provided by the alternative forecast is not significant. Or, in other words, the benchmark forecast encompasses the alternative forecast.

7.2 Volatility Forecast Evaluation

The above discussion was cast at a general level. We now turn to the case in which volatility itself is the forecasting object of interest. Hence, $y_{t+1} \equiv \sigma^2_{t:t+1}$ now refers to some form of ex-post volatility measure, while $y_{t+1|t} \equiv \hat{\sigma}^2_{t:t+1|t}$ denotes the corresponding ex-ante volatility forecast.

The regression-based framework from above then suggests the general volatility forecast evaluation regression

$$\sigma^2_{t:t+1} - \hat{\sigma}^2_{t:t+1|t} = a + b' x_t + \epsilon_{t+1}, \tag{7.7}$$

or as a special case the Mincer-Zarnowitz volatility regression

$$\sigma^2_{t:t+1} = a + (b+1)\hat{\sigma}^2_{t:t+1|t} + \epsilon_{t+1},$$

where an optimal forecast would satisfy $a = b = 0$. Immediately, however, the question arises of how to actually measure the ex-post variance? As discussed at some length in Sections 1 and 5, the "true" variance, or volatility, is inherently unobservable, and we are faced with the challenge of having to rely on a proxy in order to assess the forecast quality.

The simplest proxy is the squared observation of the underlying variable, y^2_{t+1}, which, when the mean is zero, has the property of being (conditionally) unbiasedness, or $E_t[y^2_{t+1}] = \sigma^2_{t:t+1}$. Thus, the accuracy of the volatility forecasts could be assessed by the following simple regression

$$y^2_{t+1} = a + (b+1)\hat{\sigma}^2_{t:t+1|t} + \epsilon_{t+1}. \tag{7.8}$$

However, as illustrated by Figure 1.1, the squared observation typically provides a very noisy proxy for the true (latent) volatility process of interest. We are essentially estimating the variance each period using just a single observation, and the corresponding regression fit is inevitably very low, even if the volatility forecast is accurate. For instance, regressions of the form (7.8), using daily or weekly squared returns as the left-hand-side independent variable, typically result in unspectacular R^2s of around 5-10%. We are seemingly stuck with an impossible task, namely

to precisely assess the forecastability of something which is itself not observed.

Fortunately, Figure 1.1 and the accompanying discussion in Sections 1 and 5 suggest a workable solution to this conundrum. In financial applications observations are often available at very high frequencies. For instance, even if the forecaster is only interested in predicting volatility over daily or longer horizons, observations on the asset prices are often available at much finer intradaily sampling frequencies, say $1/\Delta >> 1$ observations per "day" or unit time interval. Hence, in this situation following the discussion in Section 5.1, a proxy for the (latent) daily volatility may be calculated from the intradaily squared return as

$$RV(t+1,\Delta) \equiv \sum_{j=1}^{1/\Delta} [p(t+j\cdot\Delta) - p(t+(j-1)\cdot\Delta)]^2 .$$

The resulting forecast evaluation regression thus takes the form,

$$RV(t+1,\Delta) = a + (b+1)\hat{\sigma}^2_{t:t+1|t} + \epsilon_{t+1}, \tag{7.9}$$

which coincides with (7.8) for $\Delta = 1$. However, in contrast to the low R^2's associated with (7.8), Andersen and Bollerslev (1998a) find that in liquid markets the R^2 of the regression in (7.9) can be as high as 50% for the very same volatility forecast that produces an R^2 of only 5%-10% in the former regression! In other words, even a reasonably accurate volatility forecasting model will invariably appear to have a low degree of forecastability when evaluated on the basis of a noisy volatility proxy. Equally important, it will be difficult to detect a poor volatility forecasting model when a noisy volatility proxy is used.

Reliable high-frequency information is, of course, not available for all financial markets. Still, intra-day high and low prices, or quotes, are often available over long historical time periods. Under idealized conditions - a Geometric Brownian motion with a constant diffusive volatility σ - the expected value of the log range (the difference between the high and the low logarithmic price) over the unit time interval is directly related to volatility of the process by the equation

$$E[(\max\{p(\tau) \mid t \le \tau < t+1\} - \min\{p(\tau) \mid t \le \tau < t+1\})^2] = 4\log(2)\sigma^2 . \tag{7.10}$$

Hence, a range-based proxy for the per-period volatility is naturally defined by

$$\sigma^2_{r,t:t+1} = \frac{1}{4\log(2)}(\max\{p(\tau) \mid t \le \tau < t+1\} - \min\{p(\tau) \mid t \le \tau < t+1\})^2 . \tag{7.11}$$

It is readily seen that, under ideal conditions, this range-based volatility proxy is inferior to the realized variance measure constructed with a large number of intraday observations, or $\Delta << 1$. However, as previously discussed, a host of market microstructure and other complications often render practical situations less than ideal. Thus, even when high-frequency data are available, the range-based volatility forecast evaluation regression,

$$\sigma^2_{r,t:t+1} = a + (b+1)\hat{\sigma}^2_{t:t+1|t} + \epsilon_{t+1}, \tag{7.12}$$

may still provide a useful robust alternative, or complement, to the realized volatility regression in (7.9).

To illustrate, consider Figure 7.1, which graphs a simulated geometric Brownian motion price process during a "24 hour," or "288 five-minute," period. The "fundamental," but unobserved, price process is given by the dashed line. In practice, however, we only observe this fundamental price plus a random bid-ask spread, as indicated by the jagged solid line in the figure. The figure conveys several important insights. First, notice that the squared daily return is small (close to zero) even though there are large within-day price fluctuations. As such, the true but unobserved volatility is fairly high, and poorly estimated by the daily squared return. Second, the bid-ask bounces effect introduces artificial volatility in the observed prices. As a result, realized volatilities based on very finely sampled high-frequency squared returns produce upward biased volatility measures. As previously discussed, it is, of course, possible to adjust for this bias, and several procedures for doing so have recently been proposed in the literature. Nonetheless, the figure highlights the dangers of using too small a value for Δ in the realized volatility estimation without accounting for the bid-ask spread effect. Third, the bid-ask spread only affects the range-based measure (the difference between the two horizontal lines) twice as opposed to $1/\Delta$ times for every high-frequency return entering the realized volatility calculation. As such, the range affords a more robust (to market micro structure frictions) volatility measure. Meanwhile, an obvious drawback to the range-based volatility measure is that the multiplicative adjustment in equation (7.11) only provides an unbiased measure for integrated volatility under the ideal, and empirically unrealistic, assumption of a geometric Brownian motion, and the "right" multiplication factor is generally unknown. Moreover, extensions to multivariate settings and covariance estimation is difficult to contemplate in the context of the range.

The preceding discussion highlights the need for tools to help in choosing the value of Δ in the realized volatility measure. To this end Andersen, Bollerslev, Diebold and Labys (1999, 2000) first proposed the "volatility signature plot," as a simple indicative graphical tool. The signature plot provides a graphical representation of the realized volatility averaged over multiple days as a function of the sampling frequency, Δ, going from very high (say one-minute intervals) to low (say daily) frequencies. Recognizing that the bid-ask spread (and other frictions) generally bias the realized volatility measure, this suggests choosing the highest frequency possible for which the average realized volatility appears to have stabilized. To illustrate, Figure 7.2 shows a simulated example corresponding to the somewhat exaggerated market microstructure effects depicted in Figure 7.1. In this situation the plot suggests a sampling frequency of around "120 to 180 minutes," or "2 to 3 hours." Meanwhile, the actual empirical evidence for a host of actively traded assets indicate that fixing Δ somewhere between 5 and 15-minutes typically works well, but many other more refined procedures for eliminating the systematic bias in the simple realized volatility estimator are now also available.

7.3 Interval Forecast and Value-at-Risk Evaluation

We now discuss situations where the dynamic volatility constitutes an important part of the forecast, but the volatility itself is not the direct object of interest, leading examples of which include Value-at-Risk and probability forecasts. Specifically, consider the interval forecasts of the form discussion in Section 2,

$$\hat{y}_{t+1|t} \equiv \{\hat{y}^{L}_{t+1|t}, \hat{y}^{U}_{t+1|t}\}, \tag{7.13}$$

where the lower and upper parts of the interval forecast are defined so that there is a (1 - p/2) probability of the ex post realization falling below the lower interval and above the upper interval, respectively. In other words, the forecast promises that the ex post outcome, y_{t+1}, will fall inside the ex ante forecasted interval with conditional probability, p. This setting naturally suggests the definition of a zero-one indicator sequence taking the value one if the realization falls inside the predicted interval and zero otherwise. We denote this indicator by

$$I_{t+1} \equiv \boldsymbol{I}(\hat{y}^{L}_{t+1|t} < y_{t+1} < \hat{y}^{U}_{t+1|t}). \tag{7.14}$$

Thus, for a correctly specified conditional interval forecast the conditional probability satisfies,

$$P(I_{t+1} \mid \mathscr{F}_t) = p,$$

which also equals the conditional expectation of the zero-one indicator sequence,

$$E(I_{t+1} \mid \mathscr{F}_t) = p \cdot 1 + (1-p) \cdot 0 = p. \tag{7.15}$$

A general regression version of this conditional expectation is readily expressed as,

$$I_{t+1} - p = a + b' x_t + \epsilon_{t+1}, \tag{7.16}$$

where the joint hypothesis that $a = b = 0$ would be indicative of a correctly conditionally calibrated interval forecast series.

Since the construction of the interval forecast depends crucially on the forecasts for the underlying volatility, the set of information variables, x_t, could naturally include one or more volatility forecasts. The past value of the indicator sequence itself could also be included in the regression as an even easier and potentially effective choice of information variable. If the interval forecast ignores important volatility dynamics then the ex-post observations falling outside the ex-ante interval will cluster corresponding to periods of high volatility. In turn, this will induce serial dependence in the indicator sequence leading to a significantly positive b coefficient for $x_t = (I_t - p)$.

As noted in Section 2, the popular Value-at-Risk forecast corresponds directly to a one-sided interval forecast, and the regression in (7.16) can similarly be used to evaluate, or backtest,

VaR's. The indicator sequence in this case would simply be

$$I_{t+1} = \boldsymbol{I}(y_{t+1} < VaR^{p}_{t+1|t}), \tag{7.17}$$

where y_{t+1} now refers to the ex-post portfolio return. Capturing clustering in the indicator series (and thus clustered VaR violations) is particularly important within the context of financial risk management. The occurrence of, say, three VaR violations in one week is more likely to cause financial distress than three violations scattered randomly throughout the year. Recognizing that clusters in VaR violations likely are induced by neglected volatility dynamics again highlights the importance of volatility modeling and forecasting in financial risk management.

7.4 Probability Forecast Evaluation and Market Timing Tests

The interval and VaR forecasts discussed above correspond to quantiles (or thresholds) in the conditional distribution for a fixed and pre-specified probability of interest, p. In Section 2 we also considered probability forecasting in which the threshold of interest is pre-specified, with the probability of the random variable exceeding the threshold being forecasted. In this case the loss function is given by

$$L(y_{t+1}, \hat{y}_{t+1|t}) = (\boldsymbol{I}(y_{t+1} > c) - \hat{y}_{t+1|t})^2, \tag{7.18}$$

where c denotes the threshold, and the optimal forecast equals $\hat{y}_{t+1|t} = P(y_{t+1} > c \mid \mathscr{F}_t)$. The generalized forecast error follows directly from (7.3), $-2(\boldsymbol{I}(y_{t+1} > c) - \hat{y}_{t+1|t})$, resulting in the corresponding forecast evaluation regression

$$\boldsymbol{I}(y_{t+1} > c) - \hat{y}_{t+1|t} = a + b'x_t + \epsilon_{t+1}, \tag{7.19}$$

where the hypothesis of probability forecast unbiasedness corresponds to $a = 0$ and $b = 0$. Again, the volatility forecast as well as the probability forecast itself would both be natural candidates for the vector of information variables. Notice also the similarity between the probability forecast evaluation regression in (7.19) and the interval forecast and VaR evaluation regression in (7.16).

The probability forecast evaluation framework above is closely related to tests for market timing in empirical finance. In market timing tests, y_{t+1} is the excess return on a risky asset and interest centers on forecasting the probability of a positive excess return, thus $c = 0$. In this regard, money managers are often interested in the correspondence between ex ante probability forecasts which are larger than *0.5* and the occurrence of a positive excess return ex post. In particular, suppose that a probability forecast larger than *0.5* triggers a long position in the risky asset and vice versa. The following regression

$$\boldsymbol{I}(y_{t+1} > 0) = a + b\boldsymbol{I}(\hat{y}_{t+1|t} > 0.5) + \epsilon_{t+1}, \tag{7.20}$$

then provides a simple framework for evaluating the market timing ability in the forecasting model underlying the probability forecast, $\hat{y}_{t+1|t}$. Based on this regression it is also possible to show that $b = p_+ + p_- -1$, where p_+ and p_- denote the probabilities of a correctly forecasted positive and negative return, respectively. A significantly positive b thus implies that either p_+ or p_- or both are significantly larger than *0.5*.

7.5 Density Forecast Evaluation

The forecasts considered so far all predict certain aspects of the conditional distribution without necessarily fully specifying the distribution over the entire support. For many purposes, however, the entire predictive density is of interest, and tools for evaluating density forecasts are therefore needed. In Section 2 we explicitly defined the conditional density forecast as

$$\hat{y}_{t+1|t} = f_{t+1|t}(y) \equiv f(y_{t+1} = y \mid \mathscr{F}_t).$$

The Probability Integral Transform (PIT), defined as the probability of obtaining a value below the actual ex post realization according to the ex ante density forecast,

$$u_{t+1} \equiv \int_{-\infty}^{y_{t+1}} f_{t+1|t}(s)ds \ , \tag{7.21}$$

provides a general framework for evaluating the predictive distribution. As the PIT variable is a probability, its support is necessarily between zero and one. Furthermore, if the density forecast is correctly specified, u_{t+1} must be *i.i.d.* uniformly distributed,

$$u_{t+1} \sim i.i.d.\ \mathrm{U}(0,1). \tag{7.22}$$

Intuitively, if the density forecast on average puts too little weight, say, in the left extreme of the support then a simple histogram of the PIT variable would not be flat but rather have too many observations close to zero. Thus, the PIT variable should be uniformly distributed. Furthermore, one should not be able to forecast at time t where in the forecasted density the realization will fall at time t+1. If one could, then that part of the density forecast is assigned too little weight at time t. Thus, the PIT variable should also be independent over time.

These considerations show that it is not sufficient to test whether the PIT variable is uniformly distributed on average. We also need conditional tests to properly assess whether the u_{t+1}'s are *i.i.d.* Testing for an *i.i.d.* uniform distribution is somewhat cumbersome due to the bounded support. Alternatively, one may more conveniently test for normality of the transformed PIT variable,

$$\tilde{u}_{t+1} \equiv \Phi^{-1}(u_{t+1}) \sim i.i.d.\ N(0,1), \tag{7.23}$$

where $\Phi^{-1}(u)$ denotes the inverse cumulative density function of a standard normal variable.

In particular, the *i.i.d.* normal property in (7.23) implies that the conditional moment of any order j should equal the corresponding unconditional (constant) moment in the standard normal distribution, say μ_j. That is

$$E[\tilde{u}_{t+1}^{j} \mid \mathscr{F}_t] - \mu_j = 0 \cdot \tag{7.24}$$

This in turn suggests a simple density forecast evaluation system of regressions

$$\tilde{u}_{t+1}^{j} - \mu_j = a_j + b_j' x_{j,t} + \epsilon_{j,t+1}, \tag{7.25}$$

where j determines the order of the moment in question. For instance, testing the hypothesis that $a_j = b_j = 0$ for j = 1, 2, 3, 4 will assess if the first four conditional (non-central) moments are constant and equal to their standard normal values.

Consider now the case where the density forecast specification underlying the forecast supposedly is known,

$$y_{t+1} = \mu_{t+1|t} + \sigma_{t+1|t} z_{t+1}, \qquad z_{t+1} \sim i.i.d.\ F. \tag{7.26}$$

In this situation, it is possible to directly test the validity of the dynamic model specification for the innovations,

$$z_{t+1} = (y_{t+1} - \mu_{t+1|t})/\sigma_{t+1|t} \sim i.i.d.\ F. \tag{7.27}$$

The *i.i.d.* property is most directly and easily tested via the autocorrelations of various powers, j, of the standardized residuals, say $\mathrm{Corr}(z_t^j,\ z_{t-k}^j)$.

In particular, under the null hypothesis that the autocorrelations are zero at all lags, the Ljung-Box statistics for up to Kth order serial correlation,

$$LB^j(K) \equiv T(T+2)\sum_{k=1}^{K} \mathrm{Corr}^2(z_t^j, z_{t-k}^j)/(T-k), \tag{7.28}$$

should be the realization of a chi-square distribution with K degrees of freedom. Of course, this K degree of freedom test ignores the fact that the parameters in the density forecasting model typically will have to be estimated. As noted in Section 7.1, refined test statistics as well as simulation based techniques are available to formally deal with this issue.

As previously noted, in most financial applications involving daily or weekly returns, it is

reasonable to assume that $\mu_{t+1|t} \approx 0$, so that

$$z_{t+1}^2 \approx y_{t+1}^2 / \sigma_{t+1|t}^2.$$

Thus, a dynamic variance model can readily be thought of as removing the dynamics from the squared observations. Misspecified variance dynamics are thus likely to show up as significant autocorrelations in z_{t+1}^2. This therefore suggests setting $j = 2$ in (7.28) and calculating the Ljung-Box test based on the autocorrelations of the squared innovations, $\mathrm{Corr}\,(z_t^2,\ z_{t-k}^2)$. This same Ljung-Box test procedure can, of course, also be used in testing for the absence of dynamic dependencies in the moments of the density forecast evaluation variable from (7.23), $\tilde{u}_{t+1}$.

7.6 Further Reading

This section only scratches the surface on forecast evaluation. The properties and evaluation of point forecasts from general loss functions have recently been analyzed by Patton and Timmermann (2003, 2004). The statistical comparison of competing forecasts under general loss functions has been discussed by Diebold and Mariano (1995), Giacomini and White (2004), and West (1996). Forecast evaluation under mean squared error loss is discussed in detail by West in this Handbook. Interval, quantile and Value-at-Risk forecast evaluation is developed further in Christoffersen (1998), Christoffersen (2003), Christoffersen, Hahn and Inoue (2001), Christoffersen and Pelletier (2004), Engle and Manganelli (2004) and Giacomini and Komunjer (2005). The evaluation of probability forecasts, sign forecasts and market timing techniques is surveyed in Breen, Glosten and Jagannathan (1989), Campbell, Lo and MacKinlay (1997, chapter 2), and Christoffersen and Diebold (2003). Methods for density forecast evaluation are developed in Berkowitz (2001), Diebold, Gunther and Tay (1998), Giacomini (2002) and Hong (2000), as well as the chapter by Corradi and Swanson in this Handbook.

White (2000) provides a framework for assessing if the best forecasting model from a large set of potential models outperforms a given benchmark. Building on this idea, Hansen, Lunde and Nason (2003, 2005) develop a model confidence set approach for choosing the best volatility forecasting model.

Moreover, combining a number of volatility forecasts may be preferable to choosing a single best forecast. The general topic of forecast combination is discussed in detail in the chapter by Timmermann in this Handbook. Volatility forecast combination has been found to work well in practice by Hu and Tsoukalas (1999).

Further discussion of volatility forecasting and forecast evaluation based on realized volatility measures can be found in Andersen and Bollerslev (1998a), Andersen, Bollerslev and Meddahi (2004, 2005), and Patton (2005). Andersen, Bollerslev, Diebold and Labys (1999), Aït-Sahalia, Mykland, and Zhang (2005) and Bandi and Russel (2003, 2004), Bollen and Inder (2002), Curci and Corsi (2004), Hansen and Lunde (2004b), Martens (2003), and Zhang, Aït-Sahalia and

Mykland (2005) all analyze the important choice of sampling frequency and/or the use of various sub-sampling and other corrective procedures in the practical construction of unbiased (and efficient) realized volatility measures. Alizadeh, Brandt, and Diebold (2002) discuss the relative merits of realized and range-based volatility. For early work on the properties of range-based estimates, see Feller (1951) and Parkinson (1980).

Testing for normality of the transformed Probability Integral Transform (PIT) variable can be done in numerous ways. A couple of interesting recent procedures for testing dynamic models for correct distributional assumptions taking into account the parameter estimation error uncertainty are given by Bontemps and Meddahi (2005) and Duan (2003) .

Several important topics were not explicitly discussed in this section. In the general forecasting area they include covariance and correlation forecast evaluation (see, e.g., Brandt and Diebold, 2005), as well as related multivariate density forecast evaluation (see, e.g., Diebold, Hahn and Tay, 1999). In the area of financial forecast applications, we did not discuss the evaluation of time-varying betas (see, e.g., Ghysels, 1998), volatility-based asset allocation (see, e.g., Fleming, Kirby and Ostdiek, 2001 and 2003), and option valuation models (see, e.g., Bates, 2003, and Christoffersen and Jacobs, 2004a,b), to mention some. Nonetheless, the general volatility forecast evaluation framework set out above is flexible enough so that it may easily be adapted to each of these more specific situations.

8. Concluding Remarks

This chapter has focused on rigorous yet practical methods for volatility modeling and forecasting. The literature has obviously advanced rapidly and will almost surely continue to thrive for the foreseeable future, as key challenges remain at least partially open. Some of these, such as large-dimensional covariance matrix modeling and practical ways in which to best make use of the newly available ultra-high-frequency data have been touched upon.

Less obviously, and beyond the narrower realm of mathematical volatility models, the financial-econometric volatility literature has impacted the financial landscape in additional and important ways. Most notably, the newly-entrenched awareness of large time variation and high persistence in asset return volatility has led to the emergence of volatility as an asset class, with a variety of vehicles now available for taking positions exclusively in volatility. This contrasts with traditional options-based instruments, the value of which varies, for example, with the price of the underlying in addition to its volatility. The new vehicles include both exchange-traded products such as the Chicago Board Options Exchange's VIX volatility index, which depends directly on the one-month options implied volatility for the S&P500 aggregate market index, as well as more specialized over-the-counter volatility and covariance swaps, which are essentially futures contracts written on various realized volatility measures.

In addition to the obvious and traditional uses of such products, such as hedging volatility exposure associated with running an options book, important new uses in asset allocation

environments are emerging, as portfolio managers add volatility to otherwise-standard portfolios. While large positions in volatility may be undesirable, because volatility reverts to a fixed mean and hence has zero expected return in the long-run, small positions can provide a valuable hedge against crisis episodes in which simultaneously plunging prices cause both correlations and volatilities to increase. This type of hedge, of course, can be very appealing in both private and central bank asset-management environments.

Although it would be an exaggeration to claim that the mathematical volatility forecasting models reviewed here are solely responsible for the emergence and deepening of financial volatility markets, the development of the models nevertheless provided (and continue to provide) a major push, the effects of which we predict will continue to evolve and resonate with financial market practice for many years to come.

References

Aguilar, O. and M. West (2000), "Bayesian Dynamic Factor Models and Variance Matrix Discounting for Portfolio Allocation," *Journal of Business and Economic Statistics*, 18, 338-357.

Aït-Sahalia, Y. and M. Brandt (2001), "Variable Selection for Portfolio Choice," *Journal of Finance*, 56, 1297-1350.

Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2005), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, forthcoming.

Alexander, C. (2001) *Market Models: A Guide to Financial Data Analysis*. Chichester, UK: John Wiley and Sons, Ltd.

Alizadeh, S., M.W. Brandt, and F.X. Diebold (2002), "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047-1092.

Andersen, T.G. (1992), "Volatility," Working Paper, Finance Department, Kellogg School, Northwestern University.

Andersen, T.G. (1994), "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling," *Mathematical Finance*, 4, 75-102.

Andersen, T.G. (1996), "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 51, 169-204.

Andersen, T.G., L. Benzoni and J. Lund (2002), "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 57, 1239-1284.

Andersen, T.G. and T. Bollerslev (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *Journal of Finance*, 52, 975-1005.

Andersen, T.G. and T. Bollerslev (1998a), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.

Andersen, T.G. and T. Bollerslev (1998b), "ARCH and GARCH Models," in S. Kotz, C.B. Read and D.L. Banks (eds), *Encyclopedia of Statistical Sciences, Vol.II* . New York: John Wiley and Sons.

Andersen, T.G. and T. Bollerslev (1998c), "DM-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies," *Journal of Finance*,

53, 219-265.

Andersen, T.G., T. Bollerslev, P.F. Christoffersen and F.X. Diebold (2005), "Practical Volatility and Correlation Modeling for Financial Market Risk Management," in M. Carey and R. Stulz (eds.), *Risks of Financial Institutions*. Chicago: University of Chicago Press for National Bureau of Economic Research.

Andersen, T.G., T. Bollerslev and F.X. Diebold (2003a), "Parametric and nonparametric Measurements of Volatility," forthcoming in Y. Aït-Sahalia and L.P Hansen (eds.), *Handbook of Financial Econometrics,* North Holland.

Andersen, T.G., T. Bollerslev and F.X. Diebold (2003b), "Some Like It Smooth and Some Like It Rough: Untangling Continuous and Jump Components in Measuring, Modeling and Forecasting Asset Return Volatility," Working paper, Northwestern University, Duke University and University of Pennsylvania.

Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens (2001), "The Distribution of Stock Return Volatility," *Journal of Financial Economics*, 61, 43-76.

Andersen, T., T. Bollerslev, F.X. Diebold, and P. Labys (1999), "(Understanding, Optimizing, Using and Forecasting) Realized Volatility and Correlation ," Working Paper, Northwestern University, Duke University and University of Pennsylvania.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2000), "Great Realizations,"*Risk Magazine*, 18, 105-108.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42-55.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-625.

Andersen. T.G., T. Bollerslev, and N. Meddahi (2004), "Analytic Evaluation of Volatility Forecasts," *International Economic Review*, 45, 1079-1110.

Andersen. T.G., T. Bollerslev, and N. Meddahi (2005), "Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities," *Econometrica*, 73, 279-296.

Andersen, T.G. and J. Lund (1997), "Estimating Continuous-Time Stochastic Volatility Models of the Short term Interest Rate Diffusion," *Journal of Econometrics*, 77, 343-377.

Andersen, T.G. and B.E. Sørensen (1996), "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business & Economic Statistics*, 14, 328-352.

Andersen, T.G. and B.E. Sørensen (1997), "GMM and QML Asymptotic Standard Deviations in Stochastic Volatility Models: Comments on Ruiz (1994)," *Journal of Econometrics*, 76, 397-403.

Ang, A. and J. Chen (2002), "Asymmetric Correlation of Equity Portfolios," *Journal of Financial Economics*, 63, 443-494.

Andreou, E. and E. Ghysels (2002), "Detecting Multiple Breaks in Financial Market Volatility Dynamics," *Journal of Applied Econometrics*, 17, 579-600.

Bachelier, L. (1900), "Théorie de la Spéculation," *Annales de l'Ecole Normale Supérieure,* 3, Paris: Gauthier-Villars. English translation in Cootner, P.H. ed. (1964), *The Random Character of Stock Market Prices*, Cambridge, Massachusetts: MIT Press.

Baillie, R.T. and T. Bollerslev (1992), "Prediction in Dynamic Models with Time Dependent Conditional Variances," *Journal of Econometrics*, 52, 91-113.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3-30.

Bakshi, G, C. Cao and Z. Chen (1997), "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003-2049.

Bandi, F. and J. Russell (2003), "Volatility or Microstructure Noise," Working Paper, University of Chicago.

Bandi, F. and J.R. Russell (2004), "Microstructure Noise, Realized Variance, and Optimal Sampling," Working paper, Graduate School of Business, University of Chicago.

Banerjee, A. and G. Urga (2005), "Modelling Structural Breaks, Long Memory and Stock Market Volatility: An Overview," *Journal of Econometrics*, forthcoming.

Barndorff-Nielsen, O.E. and N. Shephard (2001), "Non-Gaussian Ornstein-Uhlenbeck-based Models and Some of their Uses in Financial Economics," *Journal of the Royal Statistical Society, Series B,* 63, 167-207.

Barndorff-Nielsen, O.E. and N. Shephard (2002), "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, 17, 457-477.

Barndorff-Nielsen, O.E. and N. Shephard (2004a), "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1-37.

Barndorff-Nielsen, O.E. and N. Shephard (2004b), "Econometric Analysis of Realized

Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics," *Econometrica*, 72, 885-925.

Barone-Adesi, G., K. Giannopoulos and L. Vosper (1999), "VaR without Correlations for Non-Linear Portfolios," *Journal of Futures Markets*, 19, 583-602.

Barrett, C. (1999), "The effects of real exchange rate depreciation on stochastic producer prices in low-income agriculture," Agricultural Economics 20, 215-230.

Barucci, E. and R. Reno (2002), "On Measuring Volatility and the GARCH Forecasting Performance," *Journal of International Financial Markets, Institutions and Money*, 12, 182-200.

Bates, D.S. (1996), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies*, 9, 69-107.

Bates, D.S. (2003) "Empirical Option Pricing: A Retrospection," *Journal of Econometrics*, 116, 387-404.

Battle, C. and J. Barquin (2004), "Fuel Prices Scenario Generation Based on a Multivariate Garch Model for Risk Analysis in a Wholesale Electricity Market," *International Journal of Electrical Power and Energy Systems*, 26, 273-280.

Bauwens, L. and S. Laurent (2005), "A New Class of Multivariate Skew Densities with Application to GARCH Models," *Journal of Business and Economic Statistics*, forthcoming.

Bauwens, L., S. Laurent and J.V.K. Rombouts (2005), "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics*, forthcoming.

Bauwens, L. and M. Lubrano (1998), "Bayesian Inference on GARCH models using the Gibbs Sampler," *The Econometrics Journal*, 1, 23-46.

Bauwens, L. and M. Lubrano (1999). *Bayesian Dynamic Econometrics*. Oxford University Press.

Bekaert, G. and G. Wu (2000), "Asymmetric Volatility and Risk in Equity Markets," *Review of Financial Studies*, 13, 1-42.

Bera, A.K. and S. Kim (2002), "Testing Constancy of Correlation and other Specifications of the BGARCH Model with an Application to International Equity Returns," *Journal of Empirical Finance*, 7, 305-362.

Beran, J. (1994) Statistics for long-memory processes. New York: Chapman & Hall.

Berkowitz, J. (2001), "Testing Density Forecasts with Applications to Risk Management" *Journal of Business and Economic Statistics*, 19, 465-474.

Berkowitz, J. and J. O'Brien (2002), "How Accurate are the Value-at-Risk Models at Commercial Banks?" *Journal of Finance*, 57, 1093-1112.

Billio, M., M. Caporin and M. Gobbo (2003), "Block Dynamic Conditional Correlation Multivariate GARCH Models," Working Paper, Università di Venezia.

Black, F. (1976), "Studies of Stock Market Volatility Changes," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.

Black, F. and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637-654.

Bollen, B. and B. Inder (2002), "Estimating Daily Volatility in Financial Markets Utilizing Intraday Data," *Journal of Empirical Finance*, 9, 551-562.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542-547.

Bollerslev, T. (1990), "Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics*, 72, 498-505.

Bollerslev, T., R.Y. Chou and K.F. Kroner (1992), "ARCH Modeling in Finance: A Selective Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5-59.

Bollerslev, T. and R.F. Engle (1993), "Common Persistence in Conditional Variances," *Econometrica*, 61, 167-186.

Bollerslev, T., R.F. Engle and D.B. Nelson (1994), "ARCH Models," in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics, Volume IV*, 2959-3038. Amsterdam: North-Holland.

Bollerslev, T., R.F. Engle and J.M. Wooldridge (1988), "A Capital Asset Pricing Model with Time Varying Covariances," *Journal of Political Economy*, 96, 116-131.

Bollerslev, T. and P.D. Jubinsky (1999), "Equity Trading Volume and Volatility: Latent Information Arrivals and Common Long-Run Dependencies," *Journal of Business &*

Economic Statistics, 17, 9-21.

Bollerslev, T. and H.O. Mikkelsen (1996), "Modeling and Pricing Long Memory in Stock Market Volatility," *Journal of Econometrics*, 73, 151-184.

Bollerslev, T. and H.O. Mikkelsen (1999), "Long-Term Equity Anticipation Securities and Stock Market Volatility Dynamics," *Journal of Econometrics*, 92, 75-99.

Bollerslev, T. and J.M. Wooldridge (1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, 143-172.

Bollerslev, T. and J.H. Wright (2001), "Volatility Forecasting, High-Frequency Data, and Frequency Domain Inference," *Review of Economic and Statistics*, 83, 596-602.

Bollerslev, T. and H. Zhou (2005), "Volatility Puzzles: A Simple Framework for Gauging Return-Volatility Regressions," *Journal of Econometrics*, forthcoming.

Bontemps, C. and N. Meddahi (2005), "Testing Normality: A GMM Approach," *Journal of Econometrics*, forthcoming.

Brandt, M.W. (1999), "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach," *Journal of Finance*, 54, 1609-1646.

Brandt, M.W. (2004), "Portfolio Choice Problems," in Y. Aït-Sahalia and L.P. Hansen (eds.), Handbook of Financial Econometrics, forthcoming.

Brandt, M.W. and F.X. Diebold (2005), "A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations," *Journal of Business*, forthcoming.

Brandt, M.W. and C. Jones (2002), "Volatility Forecasting with Range-based EGARCH Models," Working Paper, The Wharton School, University of Pennsylvania.

Braun, P.A., D.B. Nelson and A.M. Sunier (1995), "Good News, Bad News, Volatility, and Betas," *Journal of Finance*, 50, 1575-1603.

Breen, W., L.R. Glosten and R. Jagannathan (1989), "Economic significance of predictable variations in stock index returns," *Journal of Finance*, 44, 1177-1189.

Breidt, F.J. N. Crato and P.F.J. de Lima (1998), "On the Detection and Estimation of Long Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325-348.

Brooks, C. (1997), "GARCH Modelling in Finance: A Review of the Software Options," *The Economic Journal*, 107, 1271-1276.

Brooks, C. (2002) *Introductory Econometrics for Finance*. Cambridge, UK: Cambridge University Press.

Brooks, C., S.P. Burke and G. Persand (2001), "Benchmarks and the Accuracy of GARCH Model Estimation," *International Journal of Forecasting*, 17, 45-56.

Brooks, C., S.P. Burke and G. Persand (2003), "Multivariate GARCH Models: Software Choice and Estimation Issues," *Journal of Applied Econometrics*, 18, 725-734.

Buguk, C., D. Hudson and T. Hanson (2003), "Price Volatility Spillover in Agricultural Markets: an Examination of U.S. Catfish Markets," *Journal of Agricultural and Resource Economics,* 28, 86-99.

Calvet, L. and A. Fisher (2002), "Multifractality in Asset Returns: Theory and Evidence," *Review of Economics and Statistics*, 84, 381-406.

Calvet, L. and A. Fisher (2004), "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes" *Journal of Financial Econometrics*, 2, 49-83.

Calvet, L.E., A.J. Fisher and S.B. Thompson (2005), "Volatility Comovement: A Multifrequency Approach," *Journal of Econometrics*, forthcoming.

Campbell, S. and F.X. Diebold (2005), "Weather Forecasting for Weather Derivatives," *Journal of the American Statistical Association*, 100, 6-16.

Campbell, J.Y. (1987) "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373-399.

Campbell, J.Y. (2003) "Consumption-based Asset Pricing," in: Constantinides, G.M., Harris, M., and Stulz, R. eds., *Handbook of the Economics of Finance*, Vol. 1B (North-Holland, Amsterdam) 803-888.

Campbell, J.Y. and L. Hentschel (1992), "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31, 281-318.

Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1997) *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Cappiello, L., R.F. Engle and K. Sheppard (2004), "Asymmetric Dynamic in the Correlations of Global Equity and Bond Returns," Working Paper, NYU Stern School of Business.

Chacko, G. and L.M. Viceira (2003), "Spectral GMM Estimation of Continuous-Time Processes," *Journal of Econometrics*, 116, 259-292.

Chan, N.H. (2002) *Time Series: Applications to Finance*. New York: John Wiley and Sons, Inc.

Chernov, M., A.R. Gallant, E. Ghysels and G.E. Tauchen (2003), "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116, 225-257.

Chernov, M., and E. Ghysels (2000), "A Study Towards a Unified Framework for the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics*, 56, 407-458.

Christie, A.A. (1982), "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 10, 407-432.

Christoffersen, P. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841-862.

Christoffersen, P.F. (2003) *Elements of Financial Risk Management*. San Diego: Academic Press.

Christoffersen, P.F., and F.X. Diebold, (1996), "Further Results on Forecasting and Model Selection under Asymmetric Loss," *Journal of Applied Econometrics*, 11, 561-572.

Christoffersen, P.F. and F.X. Diebold (1997), "Optimal Prediction under Asymmetric Loss," *Econometric Theory*, 13, 808-817.

Christoffersen, P.F. and F.X. Diebold (2000) "How Relevant is Volatility Forecasting for Financial Risk Management?" *Review of Economics and Statistics*, 82, 12-22.

Christoffersen, P.F. and F.X. Diebold (2003), "Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics," Cambridge, Mass.: NBER Working Paper 10009.

Christoffersen, P., J. Hahn and A. Inoue (2001), "Testing and Comparing Value-at-Risk Measures," *Journal of Empirical Finance*, 8, 325-342.

Christoffersen, P. and K. Jacobs (2004a), "The Importance of the Loss Function in Option Valuation," *Journal of Financial Economics*, 72, 291-318.

Christoffersen, P. and K. Jacobs (2004b), "Which GARCH model for Option Valuation?" *Management Science*, 50, 1204-1221.

Christoffersen, P. and D. Pelletier (2004), "Backtesting Value-at-Risk: A Duration-Based Approach," *Journal of Financial Econometrics*, 2, 84-108.

Clark, P.K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41, 135-156.

Cochrane, J. (2001), *Asset Pricing*, Princeton: Princeton University Press.

Comte, F. and E. Renault (1998), "Long Memory in Continuous-Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291-323.

Corsi, F. (2003), "A Simple Long Memory Model of realized Volatility," Working paper, University of Southern Switzerland.

Corsi, F., G. Zumbach, U.A. Müller and M. Dacorogna (2001), "Consistent High-Precision Volatility From High-Frequency Data," *Economic Notes*, 30, 183-204.

Curci, G. and F. Corsi (2004), "A Discrete Sine Transform Approach for Realized Volatility Measurement," Working Paper, University of Pisa and University of Southern Switzerland.

Dacorogna, M.M., U.A. Müller, R.J. Nagler, R.B. Olsen and O.V. Pictet (2001) *An Introduction to High-Frequency Finance*. San Diego: Academic Press.

Danielsson, J. (1994), "Stochastic Volatility in Asset Prices: Estimation by Simulated Maximum Likelihood," *Journal of Econometrics*, 54, 375-400.

Danielsson, J. and J.F. Richard (1993), "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent variable Models," *Journal of Applied Econometrics*, 8, S153-S173.

Davidson, J. (2004), "Moment and Memory Properties of Linear Conditional Heteroskedasticity Models, and a New Model," *Journal of Business and Economic Statistics*, 22, 16-29.

De Jong, F. and T. Nijman (1997), "High Frequency Analysis of Lead-Lag Relationships between Financial Markets," *Journal of Empirical Finance*, 4, 259-277.

Deo, R. and C. Hurvich (2001), "On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models," *Econometric Theory*, 17, 686-710.

DeSantis, G. and B. Gerard (1997), "International Asset Pricing and Portfolio Diversification with Time-Varying Risk," *Journal of Finance*, 52, 1881-1912.

DeSantis, G., R. Litterman, A. Vesval and K. Winkelmann (2003), "Covariance Matrix Estimation," in R. Litterman (ed.), *Modern Investment Management: An Equilibrium Approach*. London, UK: John Wiley and Sons, Ltd.

Dhaene, G. and O. Vergote (2004), "Asymptotic Results for GMM Estimators of Stochastic

Volatility Models," Working Paper, Department of Economics, K.U. Leuven.

Diebold, F.X. (1988), *Empirical Modeling of Exchange Rate Dynamics*. New York: Springer-Verlag.

Diebold, F.X. (2004), "The Nobel Memorial Prize for Robert F. Engle," *Scandinavian Journal of Economics*, 106, 165-185.

Diebold, F.X., T. Gunther, T. and A. Tay (1998), "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review*, 39, 863-883.

Diebold, F.X., J. Hahn, J. and A. Tay (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange," *Review of Economics and Statistics*, 81, 661-673.

Diebold, F.X. and A. Inoue (2001), "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131-159.

Diebold, F.X. and J. Lopez (1995), "Modeling Volatility Dynamics," in K. Hoover (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, 427-472. Boston: Kluwer Academic Press.

Diebold, F.X. and J. Lopez (1996), "Forecast Evaluation and Combination," in G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics*, Amsterdam: North-Holland, 241-268.

Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-265.

Diebold, F.X. and M. Nerlove (1989), "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model," *Journal of Applied Econometrics*, 4, 1-21.

Ding, Z. and C.W.J. Granger (1996), "Modeling Volatility Persistence of Speculative Returns: A New Approach," *Journal of Econometrics*, 73, 185-215.

Ding, Z., C.W.J. Granger and R.F. Engle (1993), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83-106.

Drost, F.C. and T.E. Nijman (1993), "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909-927.

Drost, F.C. and B.J.M. Werker (1996), "Closing the GARCH Gap: Continuous Time GARCH Modeling," *Journal of Econometrics*, 74, 31-58.

Duan, J.-C. (1995), "The GARCH Option Pricing Model," *Mathematical Finance*, 5, 13-32.

Duan, J.-C. (2003), "A Specification Test for Time Series Models by a Normality Transformation," Working Paper, Rotman School of Management, University of Toronto.

Duffie, D. and K.J. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929-952.

Duffie, D., J. Pan and K.J. Singleton (2000), "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343-1376.

Dufour, J.M. (2004), "Monte Carlo Tests with Nuisance Parameters : A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," Working Paper, Université de Montréal.

Elerian, O., S. Chib and N. Shephard (2001), "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69, 959-994.

Embrechts, P, C. Klüppelberg and T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer Verlag.

Enders, W. (2004) *Applied Econometric Time Series*. Hoboken, NJ: John Wiley and Sons, Inc.

Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.

Engle, R.F. (1995) *ARCH: Selected Readings*. Oxford, UK: Oxford University Press.

Engle, R.F. (2001), "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics," *Journal of Economic Perspectives*, 15, 157-168.

Engle, R.F. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339-350.

Engle, R.F. (2004), "Nobel Lecture. Risk and Volatility: Econometric Models and Financial Practice," *American Economic Review*, 94, 405-420.

Engle, R.F. and T. Bollerslev (1986), "Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1-50.

Engle, R.F. and G. Gonzalez-Rivera (1991), "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9, 345-360.

Engle, R.F., T. Ito and W.L. Lin (1990), "Meteor Showers or Heat Waves? Heteroskedastic Intra-daily Volatility in the Foreign Exchange market" *Econometrica*, 58, 525-542.

Engle, R.F. and F.K. Kroner (1995), "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122-150.

Engle, R.F. and G.G.J. Lee (1999), "A Permanent and Transitory Component Model of Stock Return Volatility," in R.F. Engle and H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, 475-497. Oxford, UK: Oxford University Press.

Engle, R.F., D.M. Lilien and R.P. Robbins, (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391-407.

Engle, R.F. and S. Manganelli (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economic Statistics*, 22, 367-381.

Engle, R.F. and V.K. Ng (1993), "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749-1778.

Engle, R.F., V.K. Ng and M. Rothschild (1990), "Asset Pricing with a Factor-ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, 45, 213-238.

Engle, R.F. and A.J. Patton (2001), "What Good is a Volatility Model?" *Quantitative Finance*, 1, 237-245.

Engle, R.F. and K. Sheppard (2001), "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," Working Paper, NYU Stern School of Business.

Epps, T. (1979), "Comovements in Stock Prices in the Very Short Run," *Journal of the American Statistical Association*, 74, 291-298.

Eraker, B. (2001), "MCMC Analysis of Diffusions with Applications to Finance," *Journal of Business & Economic Statistics*, 19, 177-191.

Eraker, B. (2004), "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 59, 1367-1403.

Eraker, B., M. Johannes and N.G. Polson (2003), "The Impact of Jumps in Equity Index Volatility and Returns," *Journal of Finance*, 58, 1269-1300.

Erb, C., C. Harvey and T. Viskanta (1994), "Forecasting International Equity Correlations,"

Financial Analysts Journal, 50, 32-45.

Ewing, B.T., M.J. Piette and J.E. Payne (2003), "Forecasting Medical Net Discount Rates," *Journal of Risk and Insurance*, 70, 85-95.

Feller, W., (1951), "The Asymptotic Distribution of the Range of Sums of Random Variables," *Annals of Mathematical Statistics*, 22, 427-432.

Fiorentini, G. and E. Sentana (2001), "Identification and Testing of Conditionally Heteroskedastic Factor Models," *Journal of Econometrics*, 102, 143-164.

Fiorentini, G., E. Sentana and G. Calzolari (2003), "Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroskedastic Dynamic Regression Models with Student-t Innovations," *Journal of Business and Economic Statistics*, 21, 532-546.

Fiorentini, G., E. Sentana and N. Shephard (2004), "Likelihood-Based Estimation of Latent Generalized ARCH Structures," *Econometrica*, 72, 1481-1517.

Fleming, J. and C. Kirby (2003), "A Closer Look at the Relation between GARCH and Stochastic Autoregressive Volatility," *Journal of Financial Econometrics*, 1, 365-419.

Fleming, J. C. Kirby, and B. Ostdiek (2001) "The Economic Value of Volatility Timing," *Journal of Finance*, 56, 329-352.

Fleming, J. C. Kirby, and B. Ostdiek (2003) "The Economic Value of Volatility Timing Using "Realized" Volatility," *Journal of Financial Economics*, 67, 473-509.

Foster, D.P. and D.B. Nelson (1996), "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64, 139-174.

Fouque, J.-P., G. Papanicolaou, and K.R. Sircar (2000), *Derivatives in Financial Markets with Stochastic Volatility*. Princeton: Princeton University Press.

Franses, P.H. and C. Hafner (2003), "A Generalized Dynamic Conditional Correlation Model for Many Asset Returns," Working Paper, Erasmus University, Rotterdam.

Franses, P.H. and D. van Dijk (2000) *Non-Linear Time Series Models in Empirical Finance*. Cambridge, UK: Cambridge University Press.

French, K.R., G.W. Schwert and R.F. Stambaugh (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3-29.

Fridman, M. and L. Harris (1998), "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 16, 284-291.

Gallant, A.R., D.A. Hsieh and G.E. Tauchen (1997), "Estimation of Stochastic Volatility Models with Diagnostics," *Journal of Econometrics*, 81, 159-192.

Gallant, A.R., C.T. Hsu and G.E. Tauchen (1999), "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *Review of Economics and Statistics*, 81, 617-631.

Gallant, A.R., P.E. Rossi and G.E. Tauchen (1992), "Stock Prices and Volume," *Review of Financial Studies*, 5, 199-242.

Gallant, A.R. and G.E. Tauchen (1996), "Which Moments to Match," *Econometric Theory*, 12, 657-681.

Gallant, A.R. and G.E. Tauchen (1997), "Estimation of Continuous-Time Models for Stock Returns and Interest Rates," *Macroeconomic Dynamics*, 1, 135-168.

Gallant, A.R. and G. Tauchen (1998), "Reprojecting partially Observed Systems with Application to Interest Rate Diffusions," *Journal of the American Statistical Association*, 93, 10-24.

Geweke, J. (1989a), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317-1340.

Geweke, J. (1989b), "Exact Predictive Densities in Linear Models with ARCH Disturbances," *Journal of Econometrics*, 40, 63-86.

Ghysels, E. (1998), "On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt?" *Journal of Finance*, 53, 549-573.

Ghysels, E., A.C. Harvey and E. Renault (1996), "Stochastic Volatility," in *Handbook of Statistics, Volume 14;* G.S. Maddala and C.R. Rao, eds., Amsterdam: North Holland.

Ghysels, E., P. Santa-Clara and Valkanov (2004), "Predicting Volatility: Getting the Most out of Data Sampled at Different Frequencies," Working Paper, University of North Carolina and University of California, Los Angeles.

Giacomini, R. (2002), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests: Asymptotic and Bootstrap Methods" Boston College, Department of Economics Working Paper 583.

Giacomini, R. and I. Komunjer (2005), "Evaluation and Combination of Conditional Quantile Forecasts," *Journal of Business and Economic Statistics*, forthcoming.

Giacomini, R. and H. White (2004), "Tests of Conditional Predictive Ability," Working paper,

University of California, San Diego.

Giordani, P., and P. Soderlind (2003), "Inflation Forecast Uncertainty," *European Economic Review*, 47, 1037-1059.

Giraitis, L., P. Kokoszka and R. Leipus (2000), "Stationary ARCH Models: Dependence Structure and Central Limit Theorem," *Econometric Theory*, 16, 3-22.

Glosten, L.R., R. Jagannathan and D. Runkle (1993), "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779-1801.

Gourieroux, C. and J. Jasiak (2001) *Financial Econometrics*. Princeton, NJ: Princeton University Press.

Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, S85-S118.

Granger, C.W.J. (1980), "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227-238.

Granger, C.W.J. (1969), "Prediction with a Generalized Cost of Error Function," *Operational Research Quarterly*, 20, 199-207.

Granger, C.W.J., H. White and M. Kamstra, (1989), "Interval Forecasting: An Analysis Based on ARCH - Quantile Estimators," *Journal of Econometrics*, 40, 87-96.

Gronke, P. and J. Brehm (2002), "History, heterogeneity, and presidential approval: a modified ARCH approach," Electoral Studies 21, 425–452.

Guidolin, M. and A. Timmermann (2005a), "Term Structure of Risk under Alternative Specifications," *Journal of Econometrics*, forthcoming.

Guidolin, M. and A. Timmermann, (2005b), "An Econometric Model of Nonlinear Dynamics in the Joint Distribution of Stock and Bond Returns," *Journal of Applied Econometrics*, forthcoming.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamilton, J.D. and G. Lin (1996), "Stock Market Volatility and the Business Cycle," *Journal of Applied Econometrics*, 11, 573-593.

Hansen, P.R. and A. Lunde (2004a), "A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data," Working Paper, Department of Economics, Stanford

University.

Hansen, P.R. and A. Lunde (2004b), “An Unbiased Measure of Realized Variance,” Working Paper, Department of Economics, Stanford University.

Hansen, P.R. and A. Lunde (2005), “A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?” *Journal of Applied Econometrics*, forthcoming.

Hansen, P.R., A. Lunde and J.M. Nason (2003), “Choosing the best Volatility Models: a Model Confidence Set Approach,” *Oxford Bulletin of Economics and Statistics*, 65, 839-861.

Hansen, P.R., A. Lunde and J.M. Nason (2005), “Model Confidence Sets for Forecasting Models.” Working Paper, Stanford University.

Harris, L. (1986), "Cross-security Tests of the Mixture of Distributions Hypothesis," *Journal of Financial and Quantitative Analysis,* 21, 39-46.

Harris, L. (1987), "Transactions Data Tests of the Mixture of Distributions Hypothesis," *Journal of Financial and Quantitative Analysis*, 22, 127-141.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A.C. (1998), “Long Memory in Stochastic Volatility,” in eds. J. Knight and S. Satchell: *Forecasting Volatility in Financial Markets*, 307-320. Oxford: Butterworth-Heineman.

Harvey, C.R. (2001), “The Specification of Conditional Expectations,” *Journal of Empirical Finance*, 8, 573, 637.

Harvey, A.C. (2004), “Forecasting with Unobserved Components Time Series Models,” in eds. Elliott, G., C.W.J. Granger, and A. Timmermann: *Handbook of Economic Forecasting*. North Holland; forthcoming.

Harvey, A.C., E. Ruiz and E. Sentana (1992), “Unobserved Component Time Series Models with ARCH Disturbances,” *Journal of Econometrics*, 52, 129-157.

Harvey, A.C., E. Ruiz and E. Shephard (1994), “Multivariate Stochastic Variance Models,” *Review of Economic Studies*, 61, 247-264.

Harvey, A.C. and E. Shephard (1996), “Estimation of an Asymmetric Model of Asset Prices,” *Journal of Business & Economic Statistics*, 14, 429-434.

Hentschel, L. (1995), “All in the Family: Nesting Symmetric and Asymmetric GARCH Models,” *Journal of Financial Economics*, 39, 71-104.

Heston, S.L. (1993), "A Closed Form Solution for Options with Stochastic Volatility, with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327-343.

Heston, S.L., and S. Nandi (2000), "A Closed-Form GARCH Option Valuation Model," *Review of Financial Studies*, 13, 585-625.

Hong, Y., (2000), "Evaluation of Out-of-Sample Density Forecasts with Applications to Stock Prices," Working Paper, Department of Economics and Department of Statistical Science, Cornell University.

Hsieh, D.A. (1989), "Modeling Heteroskedasticity in Foreign Exchange Rates," *Journal of Business & Economic Statistics*, 7, 307-317.

Hu, M., and C. Tsoukalas (1999), "Combining Conditional Volatility Forecasts Using Neural Networks: An Application to the EMS Exchange Rates," *Journal of International Financial Markets, Institutions and Money*, 9, 407-422.

Huang, X. and G.E. Tauchen (2004), "The Relative Contribution of Jumps to Total Price Variance," Working Paper, Duke University.

Hull, J. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42, 281-300.

J. P. Morgan (1997) *RiskMetrics, Technical Documents*, 4th Edition. New York.

Jacquier, E., N.G. Polson and P.E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 12, 371-389.

Jagannathan, R. and T. Ma (2003), "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 1651-1684.

Jiang, G.J. and J.L. Knight (2002), "Efficient Estimation of the Continuous Time Stochastic Volatility Model via the Empirical Characteristic Function," *Journal of Business & Economic Statistics*, 20, 198-212.

Johannes, M. and N.G. Polson (2003), "MCMC Methods for Continuous-Time Financial Econometrics," forthcoming in *Handbook of Financial Econometrics;* Y. Aït-Sahalia and L.P. Hansen, eds., Amsterdam: North Holland.

Johannes, M., N. Polson and J. Stroud (2004), "Sequential Optimal Portfolio Performance: Market and Volatility Timing," Working Paper, Columbia University, University of Pennsylvania, and University of Chicago.

Johnson, T.D., R.M. Elashoff and S.J.A. Harkema (2003), "Bayesian Change Point Analysis of Electromyographic Data: Detecting Muscle Activation Patterns and Associated Applications," *Biostatistics*, 4, 143-164.

Johnson, H. and D. Shanno (1987), "Option Pricing when the Variance Is Changing," *Journal of Financial and Quantitative Analysis*, 22, 143-152.

Jondeau, E. and M. Rockinger (2005), "The Copula-GARCH Model of Conditional Dependence: An International Stock Market Application," *Journal of International Money and Finance*, forthcoming.

Jorion, P. (2000) *Value at Risk: The New Benchmark for Managing Financial Risk*. New York: McGraw-Hill.

Karpoff, J.M. (1987), "The Relation between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis*, 22, 109-126.

Kawakatsu, H. (2005), "Matrix Exponential GARCH," *Journal of Econometrics*, forthcoming.

Kim, S., N. Shephard and S. Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

King, M., E. Sentana, and S. Wadhwani (1994), "Volatility and Links Between National Stock Markets," *Econometrica*, 62, 901-933.

Kroner, K.F. and V.K. Ng (1998), "Modelling Asymmetric Comovements of Asset Returns," *Review of Financial Studies*, 11, 817-844.

Lamoureux, C.G. and W.D. Lastrapes (1990), "Persistence in Variance, Structural Change, and the GARCH Model," *Journal of Business and Economic Statistics*, 8, 225-234.

Lamoureux, C.G. and W.D. Lastrapes (1994), "Endogenous Trading Volume and Momentum in Stock-Return Volatility," *Journal of Business & Economic Statistics*, 14, 253-260.

Lastrapes, W.D. (1989), "Exchange Rate Volatility and US Monetary Policy: An ARCH Application," Journal of Money, Credit and Banking, 21, 66-77.

Ledoit O. and M. Wolf (2003), "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10, 603-621.

Ledoit O., P. Santa-Clara and M. Wolf (2003), "Flexible Multivariate GARCH Modeling with an Application to International Stock Markets," *Review of Economics and Statistics*, 85, 735-747.

Lee, S.W. and B.E. Hansen (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29-52.

Lettau, M. and S. Ludvigsson (2003), "Measuring and Modeling Variation in the Risk-Return Tradeoff," Working Paper, New York University and NBER.

Li, W.K., S. Ling and M. McAleer (2002), "Recent Theoretical Results for Time Series with GARCH Errors," *Journal of Economic Surveys*, 16, 245-269.

Liesenfeld, R. (1998), "Dynamic Bivariate Mixture Models: Modeling the Behavior of Prices and Trading Volume, *Journal of Business & Economic Statistics*, 16, 101-109.

Liesenfeld, R. (2001), "A Generalized Bivariate Mixture Model for Stock Price Volatility and Trading Volume," *Journal of Econometrics*, 104, 141-178.

Liesenfeld, R. and J.F. Richard (2003), "Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics," *Journal of Empirical Finance*, 10, 505-531.

Ling, S. and M. McAleer (2003), "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280-310.

Linn, S.C., and Z. Zhu (2004), "Natural Gas Prices and the Gas Storage Report: Public News and Volatility in Energy Futures Markets," *Journal of Futures Markets*, 24, 283-313.

Longin, F. and B. Solnik (1995), "Is the Correlation in International Equity Returns Constant: 1970-1990?" *Journal of International Money and Finance*, 14, 3-26.

Longin, F. and B. Solnik (2001), "Extreme Correlation of International Equity Markets," *Journal of Finance*, 56, 649-676.

Loretan, M. and P.C.B. Phillips (2004), "Testing Covariance Stationarity under Moment Condition Failure with an Application to Stock Returns," *Journal of Empirical finance*, 1, 211-248.

Lumsdaine, R.L. (1996), "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in GARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, 64, 575-596.

Lundin, M., M.M. Dacorogna and U.A. Müller (1998), "Correlation of High Frequency Financial Time Series," in P. Lequeux (ed.), *The Financial Markets Tick by Tick*. London: John Wiley & Sons.

Lütkepohl, H. (2004), "Forecasting with VARMA Models," in eds. Elliott, G., C.W.J. Granger, and A. Timmermann: *Handbook of Economic Forecasting*. North Holland; forthcoming.

Maestas, C. and R. Preuhs (2000), "Modeling volatility in political time series," Electoral Studies 19, 95–110.

Marinova, D. and M. McAleer (2003), "Modeling trends and volatility in ecological patents in the USA," Environmental Modelling & Software 18, 195–203.

Markowitz, H. (1952), "Portfolio Selection", *Journal of Finance*, 7, 77-91.

Marquering, W. and M. Verbeek (2004), "The Economic Value of Predicting Stock Index Returns and Volatility," *Journal of Financial and Quantitative Analysis*, 39, 407-429.

Martens, M. (2003), "Estimating Unbiased and Precise Realized Covariances," Working Paper, University of Rotterdam.

Martin-Guerrero, J.D., G. Camps-Valls, E. Soria-Olivas, A.J. Serrano-Lopez, J.J. Perez-Ruixo and N.V. Jimenez-Torres (2003), "Dosage Individualization of Erythropoietin Using a Profile Dependent Support Vector Regression," *IEEE Transactions on Biomedical Engineering*, 50, 1136-1142.

McCullough, B. and C. Renfro (1998), "Benchmarks and Software Standards: A Case Study of GARCH Procedures," *Journal of Economic and Social Measurement*, 25, 59-71.

McNeil, A.J. and R. Frey (2000), "Estimation of Tail-Related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach," *Journal of Empirical Finance*, 7, 271-300.

Meddahi, N. (2001), "An Eigenfunction Approach for Volatility Modeling," Working Paper, University of Montréal.

Meddahi, N. and E. Renault (2004), "Temporal Aggregation of Volatility Models," *Journal of Econometrics*, 119, 355-379.

Meghir, C. and L. Pistaferri (2004), "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72, 1-32.

Melino, A. and S.M. Turnbull (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. (1976), "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics*, 3, 125-144.

Mikosch, T. and C. Starica (2004), "Nonstationarities in Financial Time Series, the Long Range Dependence and the IGARCH Effects," *Review of Economics and Statistics*, 86, 378-390.

Mills, T.C. (1993) *The Econometric Modelling of Financial Time Series*. Cambridge, UK: Cambridge University Press.

Mincer, J. and V. Zarnowitz (1969), "The Evaluation of Economic Forecasts," in J. Mincer (ed.), *Economic Forecasts and Expectations*. New York: National Bureau of Economic Research.

Monfardini, C. (1998), "Estimating Stochastic Volatility Models through Indirect Inference," *The Econometrics Journal*, 1, C113-C128.

Müller, U.A., M.M. Dacorogna, R.D. Davé, R.B. Olsen, O.V. Puctet, and J. von Weizsäcker (1997), "Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components," *Journal of Empirical Finance*, 4, 213-239.

Nelson, D.B. (1988), "Time Series Behavior of Stock Market Volatility and Returns," Ph.D. dissertation, MIT.

Nelson, D.B. (1990), "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318-334.

Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.

Nijman, T. and E. Sentana (1996), "Marginalization and Contemporaneous Aggregation in Multivariate GARCH Processes," *Journal of Econometrics*, 71, 71-87.

Ng, V.K., R.F. Engle and M. Rothschild (1992), "A Multi-Dynamic-Factor Model for Stock Returns," *Journal of Econometrics*, 52, 245-266.

O'Connell, P.E. (1971), "A Simple Stochastic Modelling of Hurst's Law," *Proceedings of International Symposium on Mathematical Models in Hydrology*, 1, 169-187.

Pagan, A. (1996), "The Econometrics of Financial Markets," *Journal of Empirical Finance*, 3, 15-102.

Palm, F. (1996), "GARCH Models of Volatility," in C.R. Rao and G.S. Maddala (eds.) *Handbook of Statistics, Volume 14*, 209-240. Amsterdam: North-Holland.

Pan, J. (2002), "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3-50.

Parkinson, M. (1980), "The Extreme Value Method for Estimating the Variance of the Rate of Returns," *Journal of Business* 53, 61-65.

Pastor, L. and R. Stambaugh (2001), "The Equity Premium and Structural Breaks," *Journal of Finance*, 56, 1207-1245.

Patton, A. (2004), "Modeling Asymmetric Exchange Rate Dependence," Working Paper, London School of Economics.

Patton, A. (2005), "Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies," Working Paper, London School of Economics.

Patton, A. and A. Timmermann (2003), "Properties of Optimal Forecasts," CEPR Discussion Paper 4037.

Patton, A. and A. Timmermann (2004), "Testable Implications of Forecast Optimality," Working Paper, London School of Economics and University of California at San Diego.

Pelletier, D. (2005), "Regime Switching for Dynamic Correlations," *Journal of Econometrics*, forthcoming.

Perez-Quiros, G. and A. Timmermann (2000), "Firm Size and Cyclical Variations in Stock Returns," *Journal of Finance*, 55, 1229-1262.

Pesaran, M.H. and P. Zaffaroni (2004), "Model Averaging and Value-at-Risk based Evaluation of Large Multi Asset Volatility Models for Risk Management," Working Paper, Department of Economics, University of Cambridge.

Piazzesi, M. (2003), "Affine Term Structure Models," in L.P. Hansen and Y. Aït-Sahalia (eds.), *Handbook of Financial Econometrics*. Amsterdam: North-Holland, forthcoming.

Praetz, P.D. (1972), "The Distribution of Share Price Changes," *Journal of Business*, 45, 49-55.

Pritsker, M. (2001), "The Hidden Dangers of Historical Simulation," Working Paper, Federal Reserve Board.

Ramirez, O.A. and M. Fadiga (2003), "Forecasting Agricultural Commodity Prices with Asymmetric Error GARCH Models," *Journal of Agricultural and Resource Economics*, 28, 71-85.

Rich, R. and J. Tracy (2004), "Uncertainty and Labor Contract Durations," *Review of Economics and Statistics*, 86, 270-287.

Richardson, M. and T. Smith (1994), "A Direct Test of the Mixture of Distributions Hypothesis: Measuring the Daily Flow of Information," *Journal of Financial and Quantitative Analysis*, 29, 101-116.

Robinson, P.M. (1991), "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," *Journal of Econometrics*, 47, 67-84.

Rossi, P.E. (1996) *Modeling Stock Market Volatility: Bridging the Gap to Continuous Time*. San Diego: Academic Press, Inc.

Ruge Murcia, F.J. (2003), "Inflation Targeting under Asymmetric Preferences," *Journal of Money, Credit and Banking,* 35, 763-785.

Ruge Murcia, F.J. (2004), "The Inflation Bias When the Central Bank Targets the Natural Rate of Unemployment," *European Economic Review*, 48, 91-107.

Sandmann, G. and S.J. Koopman (1998), "Estimation of Stochastic Volatility models through Monte Carlo Maximum Likelihood," *Journal of Econometrics*, 87, 271-301.

Scholes, M. and J. Williams (1977), "Estimating Betas from Non-Synchronous Data," *Journal of Financial Economics*, 5, 309-327.

Schwert, G.W. (1989), "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance*, 44, 1115-1153.

Schwert, G.W. (1990), "Stock Volatility and the Crash of '87," *Review of Financial Studies*, 3, 77-102.

Scott, L.O. (1987), "Option Pricing when the Variance Changes Randomly: Theory, Estimation and an Application," *Journal of Financial and Quantitative Analysis*, 22, 419-438.

Sentana, E. and G. Fiorentini (2001), "Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models," *Journal of Econometrics*, 102, 143-164.

Shanken, J.A. (1990), "Intertemporal Asset Pricing: an Empirical Investigation," *Journal of Econometrics*, 45, 99-120.

Sharpe, W. (1964), "Capital Asset Prices - A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425-442.

Shawky, M.A., A. Marathe and C.L. Barrett (2003), "A First Look at the Empirical Relation Between Spot and Futures Electricity Prices in the United States," *Journal of Futures Markets*, 23, 931-955.

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility Models," in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (eds.) *Time Series Models in Econometrics, Finance and Other Fields*, 1-67. London: Chapman & Hall.

Shephard, N. (2004) *Stochastic Volatility: Selected Readings*. Oxford, UK: Oxford University Press.

Sheppard, K. (2004), "Economic Factors and the Covariance of Equity Returns," Working Paper, University of California, San Diego.

Singleton, K.J. (2001), "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111-141.

Smith, Jr., A.A. (1990), "Three Essays on the Solution and Estimation of Dynamic Macroeconomic Models," Ph.D. dissertation, Duke University.

Smith, Jr., A.A. (1993), "Estimating Nonlinear Time-series Models using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63-S84.

Tauchen, G. and M. Pitts (1983), "The Price Variability-Volume Relationship on Speculative Markets," *Econometrica*, 51, 485-505.

Taylor, S.J. (1986) *Modeling Financial Time Series*. Chichester, UK: John Wiley and Sons.

Taylor, S.J. (2004) *Asset Price Dynamics and Prediction*. Princeton, NJ: Princeton University Press.

Taylor, J.W. and R. Buizza (2003), "Using Weather Ensemble Predictions in Electricity Demand Forecasting," *International Journal of Forecasting*, 19, 57-70.

Tiao, G.C. and R.S. Tsay (1994), "Some Advances in Non-Linear and Adaptive Modeling in Time Series," *Journal of Forecasting*, 14, 109-131.

Tsay, R.S. (2002) *Analysis of Financial Time Series*. New York: John Wiley and Sons, Inc.

Tse, Y.K. and A.K.C. Tsui (2002), "A Multivariate GARCH Model with Time-Varying Correlations," *Journal of Business and Economic Statistics*, 20, 351-362.

Tse, Y.K. and P.S.L. Yip (2003), "The Impacts of Hong Kong's Currency Board Reforms on the Interbank Market," *Journal of Banking and Finance*, 27, 2273-2296.

Wang, L. (2004), "Investing when Volatility Fluctuates," Working Paper, the Wharton School and Singapore Management University.

Weiss, A.A. (1986), "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107-131.

West, K.D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067-1084.

West, K.D., and D. Cho (1995), "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics*, 69, 367-391.

West, K.D. and M.W. McCracken (1998) "Regression-Based Tests of Predictive Ability", *International Economic Review*, 39, 817-40.

White, H. (2000), "A Reality Check for Data Snooping," *Econometrica*, 68, 1097-1127.

Whitelaw, R.F. (1997), "Time-Varying Sharpe Ratios and Market Timing," Working Paper, NYU, Stern School of Business.

Wiggins, J.B. (1987), "Option Values under Stochastic Volatility: Theory and Empirical Estimates," *Journal of Financial Economics*, 19, 351-372.

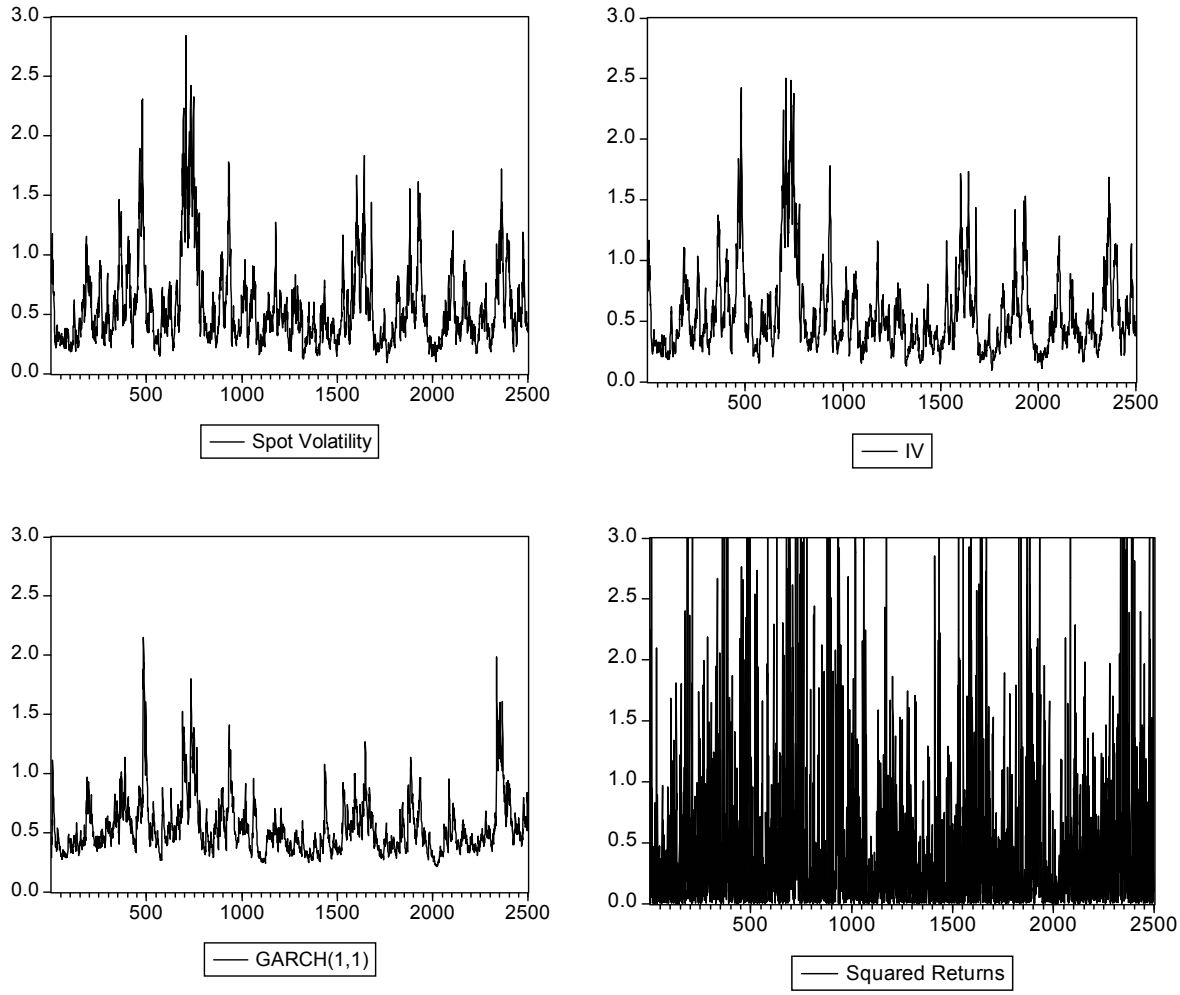
Zaffaroni, P. (2004), "Estimating and Forecasting Volatility with Large Scale Models: Theoretical Appraisal of Professional Practice," Working Paper, Banca d'Italia.

Zakoïan, J.-M. (1994), "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control*, 18, 931-955.

Zhang, L., Y. Aït-Sahalia and P.A. Mykland (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-frequency Data," *Journal of the American Statistical Association*, forthcoming.

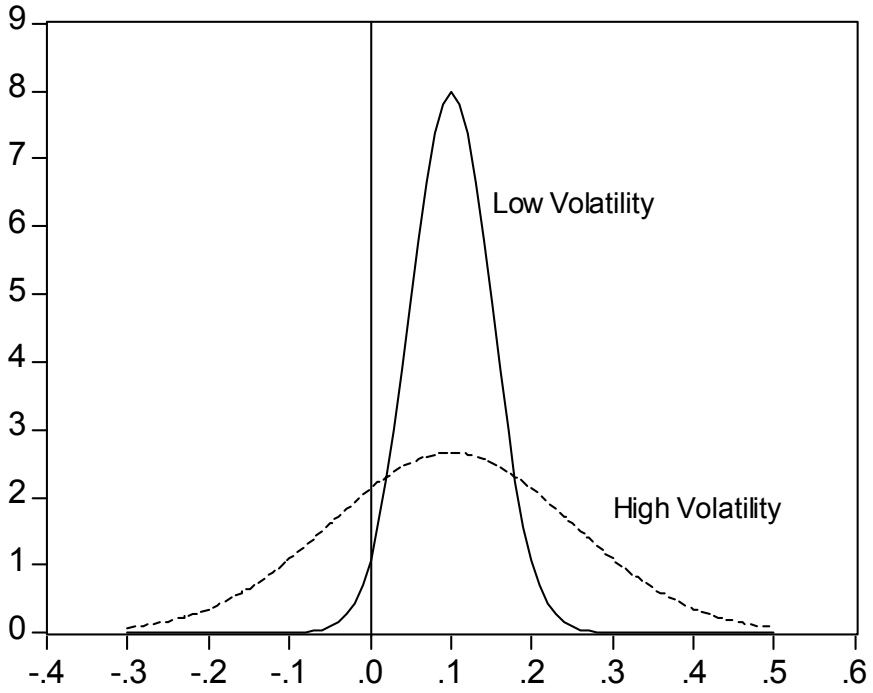
Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign Exchange Rates," *Journal of Business & Economic Statistics*, 14, 45-52.

Figure 1.1
Different Volatility Concepts



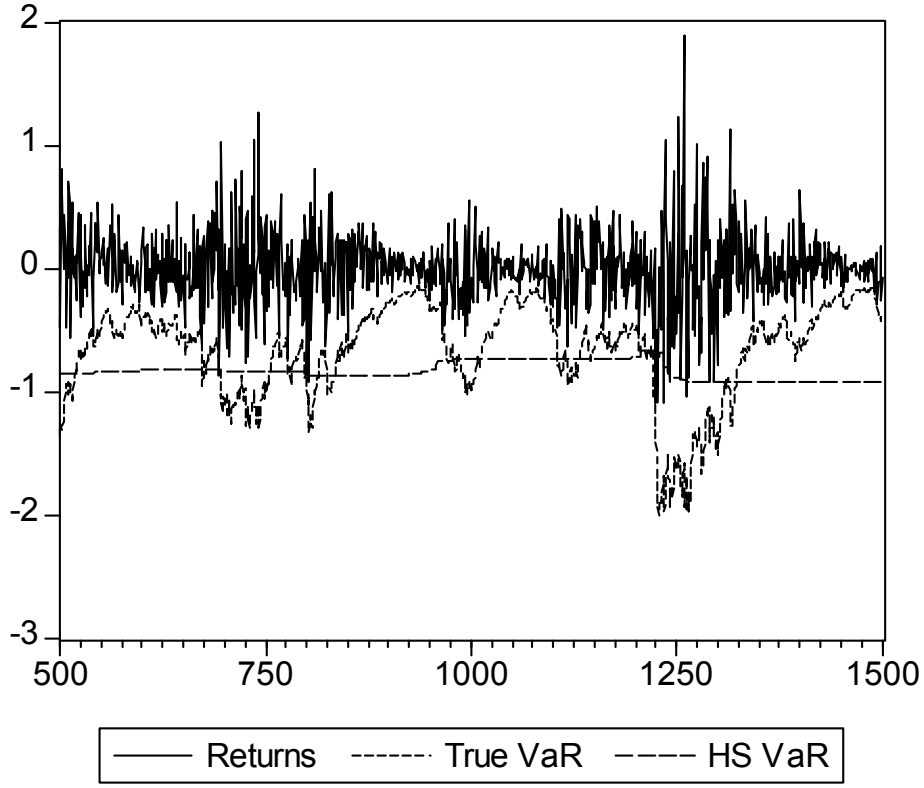
Notes to figure: We show the "daily" spot volatility, $\sigma^2(t)$, the integrated volatility, $IV(t)$, the discrete-time GARCH based volatility forecasts, $\sigma^2_{t|t-1}$, and the corresponding squared returns, r_t^2, from a simulated continuous-time GARCH diffusion.

Figure 2.1
Density Forecasts on High Volatility and Low Volatility Days



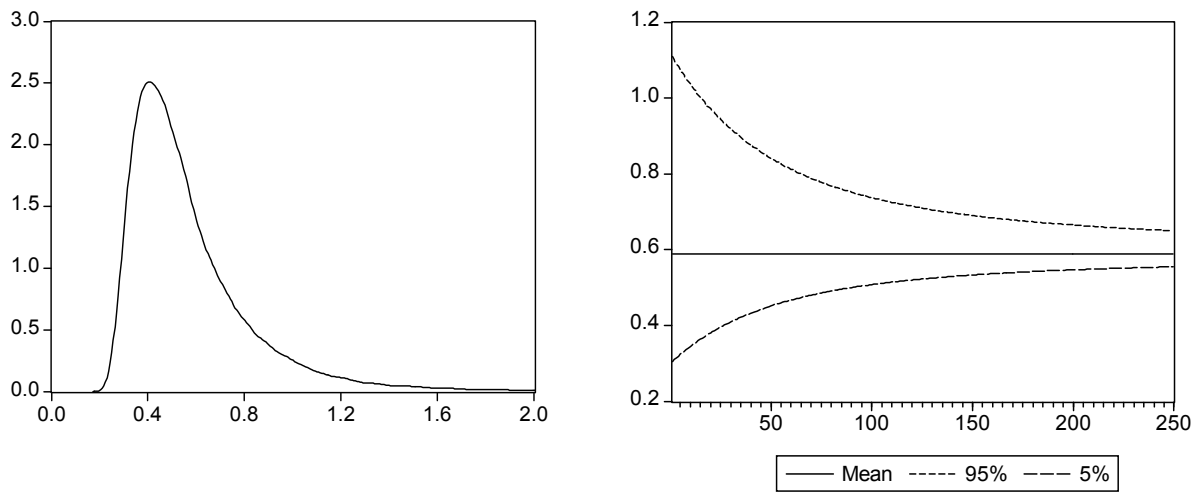
Notes to figure: The figure shows two hypothetical return distributions for a low volatility (solid line) and high volatility (dashed line) day. The areas to the left of the vertical line represent the probability of a negative return.

Figure 2.2
Simulated Portfolio Returns with Dynamic Volatility and Historical Simulation VaRs



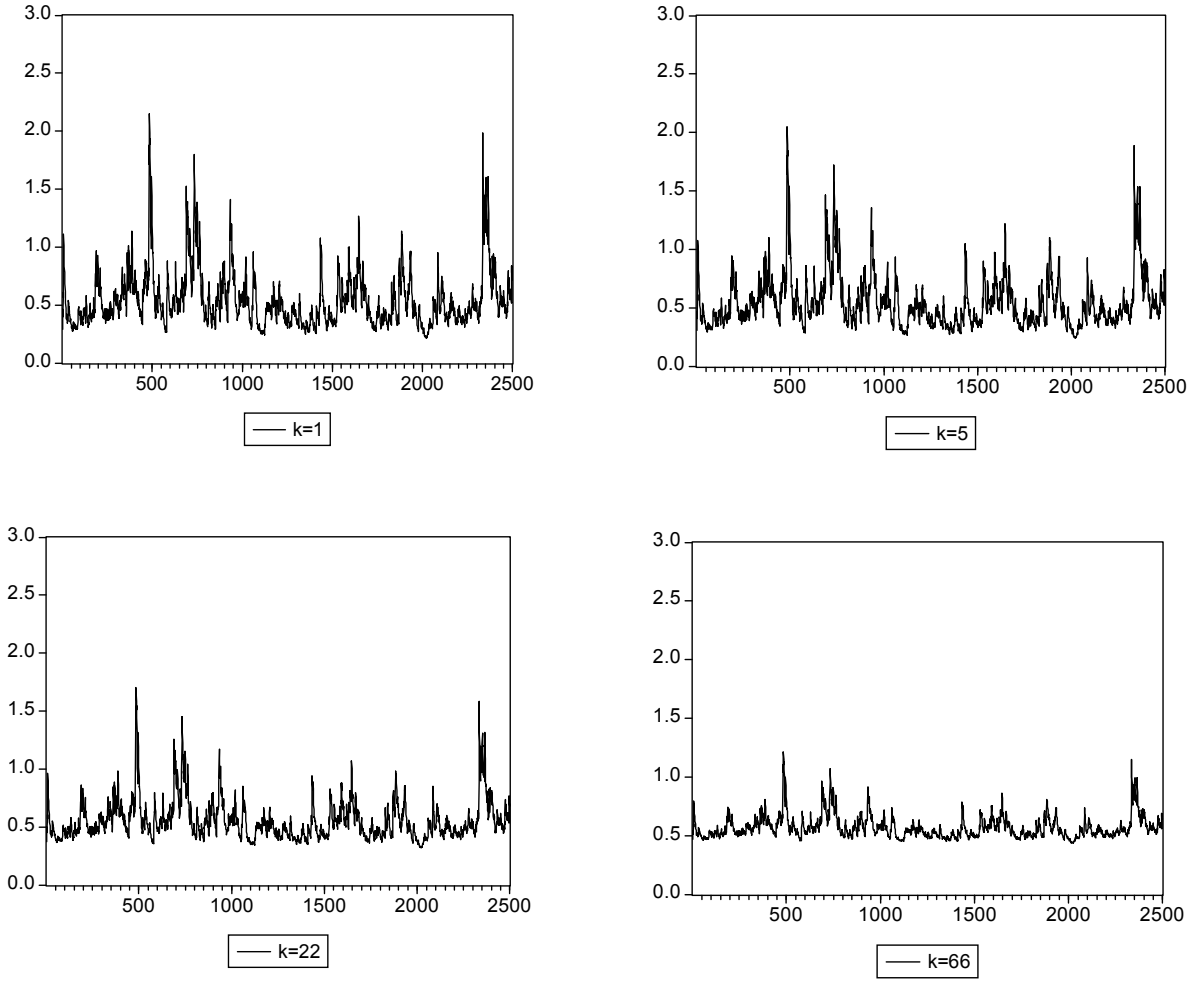
Notes to figure: The solid line shows a time series of typical simulated daily portfolio returns. The short-dashed line depicts the true one-day-ahead, 1% VaR. The long-dashed line gives the 1% VaR based on the so-called Historical Simulation (HS) technique and a 500-day moving window.

Figure 3.1
GARCH Volatility Term Structure



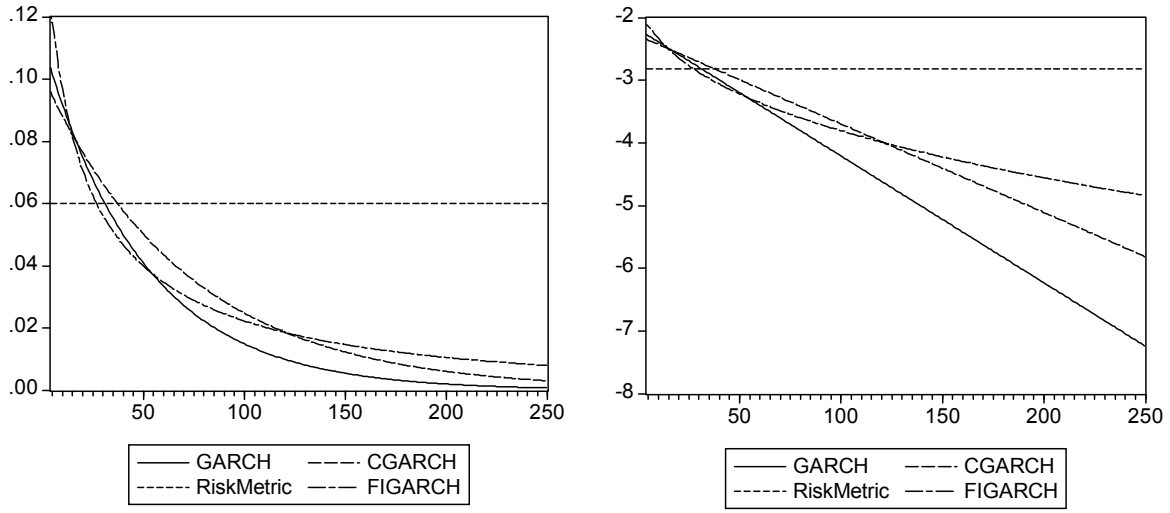
Notes to figure: The first panel shows the unconditional distribution of $\sigma^2_{t+1|t}$. The second panel shows the term-structure-of-variance, $k^{-1}\sigma^2_{t:t+k|t}$, for $\sigma^2_{t+1|t}$ equal to the mean, together with the fifth and the ninety-fifth percentiles in the unconditional distribution.

Figure 3.2
GARCH Volatility Forecasts and Horizons



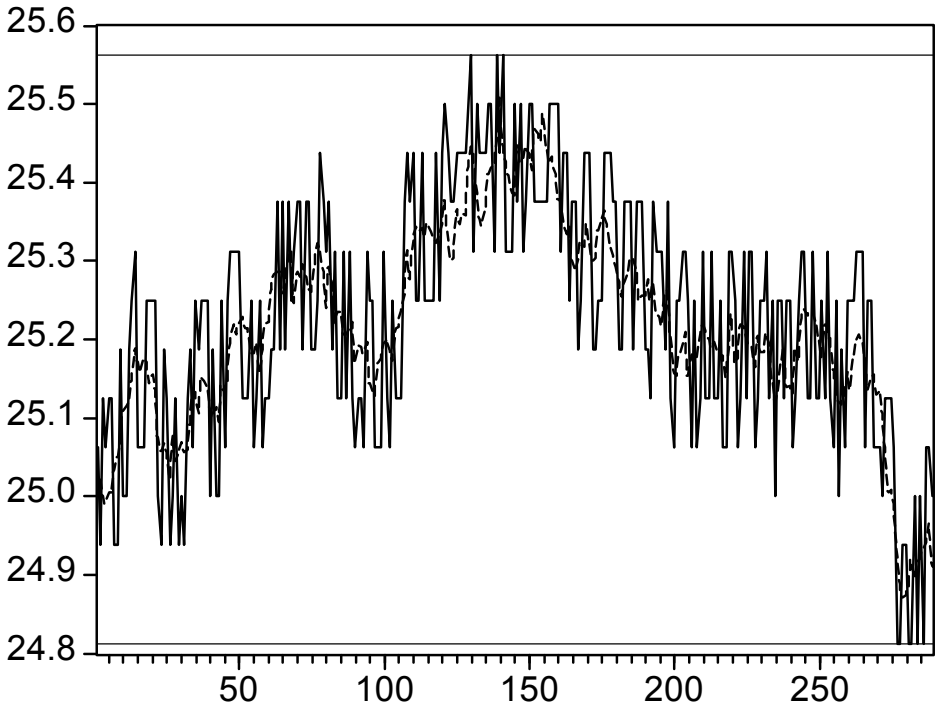
Notes to figure: The four panels show the standardized "daily" GARCH(1,1) volatility forecasts, $k^{-1}\sigma^2_{t:t+k|t}$, for horizons *k = 1, 5, 22, 66*.

Figure 3.3
Volatility Impulse Response Coefficients



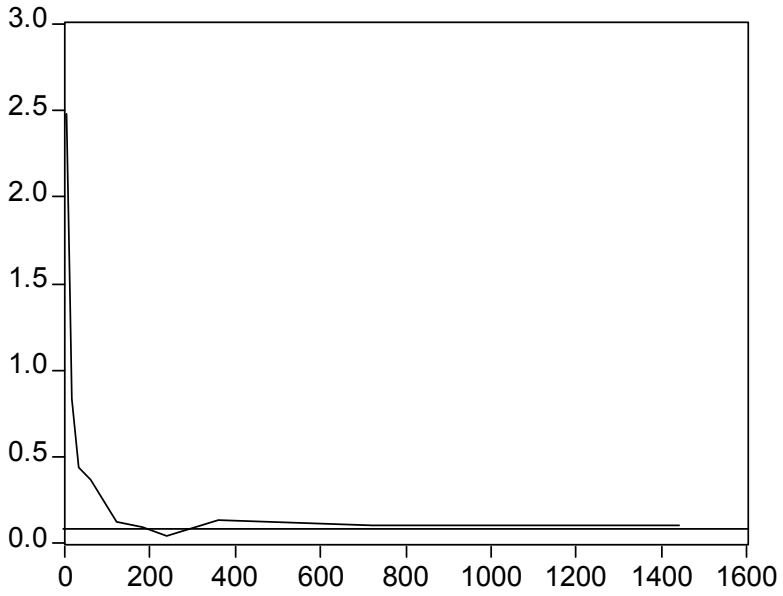
Notes to figure: The left panel graphs the volatility impulse response function, $\partial\sigma^2_{t+h|t}/\partial\varepsilon^2_t$, h = *1, 2, ..., 250*, for the RiskMetrics forecasts, the standard GARCH(1,1) model in (3.6), the FIGARCH(1,d,1) model in (3.18), and the component GARCH model in (3.21) and (3.22). The right panel plots the corresponding logarithmic values.

Figure 7.1
Simulated Fundamental and Observed Intraday Prices



Notes to figure: The smooth dashed line represents the fundamental, but unobserved, simulated asset price. The jagged solid line solid represents the observed transaction prices reflecting bid or ask driven transactions. The two horizontal lines denote the min and max prices observed within the day.

Figure 7.2
Volatility Signature Plot



Notes to figure: The figure depicts the impact of the bid-ask spread for measuring realized volatility by showing the unconditional sample means for the realized volatilities as a function of the length of the return interval for the high-frequency data underlying the calculations. The simulated prices are subject to bid-ask bounce effects shown in Figure 7.1.