Creating and Using Volatility Forecasts

KENNETH F. KRONER

Accurate forecasts of volatility are useful for several applications, such as risk management, derivatives pricing, options trading, and hedging. Yet there is no single model that is widely accepted by the industry as "the" way to forecast volatility. In fact, until the mid-1980s, the standard model was to use historical volatility as a forecast of future volatility.

Since then, several alternative models have been defended by their users, such as exponential smoothing (used by RiskMetrics), GARCH-based models (used by Salomon Brothers), and time series models on squared returns (used by many academics). The flavor of these models is the same, because they all model the way volatility tends to cluster through time.

This article compares and contrasts several of these popular forecasting models. It first argues that the intuition behind these models is the same; the only important difference is how much confidence the user has that historical data can give the true parameters in the volatility process.

Then, using daily historical data on exchange rates and their implied standard deviations, we demonstrate that this class of models can yield better volatility forecasts than market-based implied standard deviations. Finally, we present two potential volatility-based trading strategies, and demonstrate that this class of forecasting models can be used to successfully "beat the market."

KENNETH F. KRONER is a principal and research officer in the Advanced Strategies and Research Group at ~~BZW~~ Barclays Global Investors in San Francisco, California.

WHY VOLATILITY IS FORECASTABLE

Volatility can be defined as the magnitude of unexpected price changes. It can be high either because prices unexpectedly increase or because they unexpectedly decrease. The most common definition of volatility is the standard deviation of returns:

$$\sigma \equiv \sqrt{E(r - \mu_r)^2} \tag{1}$$

where r is the return and μ_r is the expected return. In practice, σ is unobservable, so it must be estimated. The usual way to estimate volatility is with the sample standard deviation

$$\tilde{\sigma} = \sqrt{\frac{\sum_{t=1}^{N}(r_t - \bar{r}_t)^2}{N - 1}} \tag{2}$$

where N is the sample size. If too large a data set is used to construct this estimate, there is a risk of clouding the estimate with stale data. On the other hand, if not enough observations are used, there is the risk of having a volatility estimate dominated by one or two observations, giving an imprecise estimate. To balance these two risks, moving (or rolling) standard deviations are frequently estimated, fixing N at an intermediate level and dropping an old observation whenever a new observation appears.

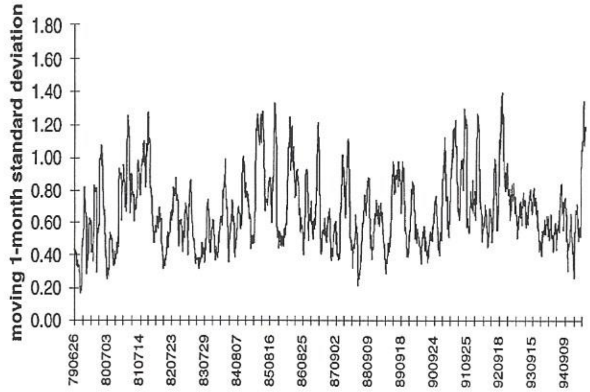
Exhibit 1 plots daily moving standard deviations for deutschemark returns, using a sixty-four-day (three-month) window since June 1979. Two observations are apparent.

First, this measure of volatility varies through time. It ranges from about 0.20 in late 1979 to about 1.20 in Summer 1985. Second, the more volatile periods tend to cluster together, as do the more tranquil periods.

For example, in 1981, volatility was typically about 1.00, while in 1983 it was about 0.60. This volatility clustering is also apparent in Exhibit 2, which plots the deutschemark realized returns surrounded by their 95% confidence intervals. Exhibit 2 also shows periods of tranquility (e.g., 1994), followed by periods of volatility (early 1995). Both these properties are common to most asset classes (including fixed incomes, equities, commodities, and currencies), and are especially apparent in data measured at the daily frequency or higher.

EXHIBIT 1

VOLATILITY IS TIME-VARYING — DEUTSCHEMARK MOVING STANDARD DEVIATIONS



Of course, the results in Exhibits 1 and 2 shouldn't surprise anyone, because the market perception of volatility is also changing through time. Exhibit 3 plots the daily implied standard deviations on thirty-day Swiss franc currency options since August 1990, as recorded by AIG International. This plot follows a similar pattern to Exhibits 1 and 2, with predicted volatility peaking during the breakdown of the ERM, and dropping steadily since then until early 1995.

The important conclusion from this discus-

EXHIBIT 2

VOLATILITY IS TIME-VARYING — DEUTSCHEMARK REALIZED RETURNS

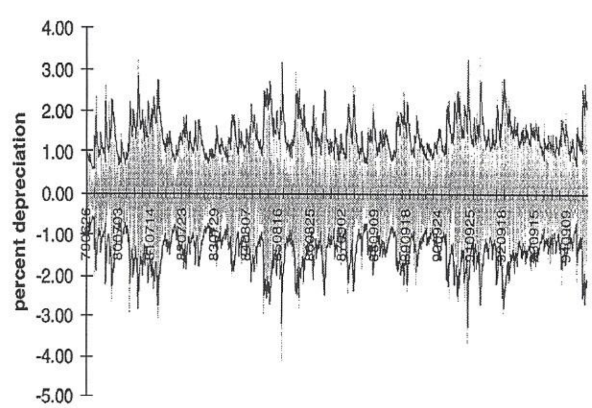
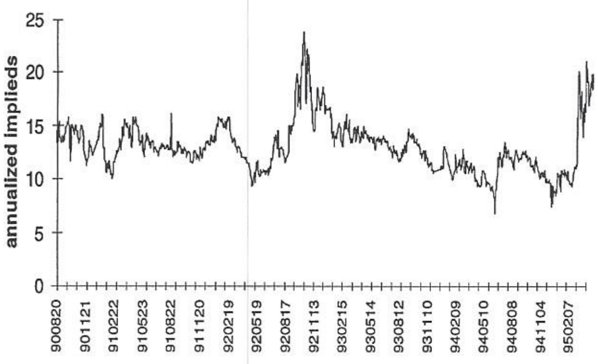


Exhibit 3

Expected Volatility is Time-Varying — Swiss Franc Implied Standard Deviations



sion is that volatility is not random, but clusters through time. Periods of volatility tend to follow periods of volatility, while periods of tranquility tend to follow periods of tranquility.

Several reasons have been proposed for volatility clustering in financial markets, such as serially correlated news arrival (Diebold [1988]), institutional trading rules (Bollerslev and Domowitz [1991]), market microstructure effects (Attiyeh [1995]), learning on the part of economic agents (Mizrach [1995]), and the dissemination of information across markets (Engle, Ito, and Lin [1990]). But whatever the reason, volatility clustering implies that if volatility is currently low, it will probably stay low for some time.

On the other hand, if it is currently high, it will probably stay high. This suggests that when creating volatility forecasting models, we should seek models that imply low (high) variance when past returns are unexpectedly small (large). Essentially, these kinds of models get inside the brain of a trader. Traders typically expect high volatility when recent price moves have been unusually large, and low volatility when recent price moves have been unusually small.

A straightforward way to capture this phenomenon is to use a time series model in which the driving variable is the size of past returns. For example,

$$\sigma_t = f\,(\delta_0 + \delta_1 s_{t-1} + \delta_2 s_{t-2} + \ldots) + \varepsilon_t \qquad (3)$$

where s_t is the size of the unexpected return at time $t - 1$ and $f(\cdot)$ is an increasing function, is a time series model that captures this phenomenon. If the parameters δ_i are all positive and declining to zero, volatility will tend to be high when recent shocks have been large, and low when recent shocks have been small.

COMPARING AND CONTRASTING THE MODELS

Consider the following four volatility models: historical volatility, an ARMA model on squared returns, GARCH, and exponential smoothing. We discuss each of these models, and demonstrate that each can be considered a special case of Equation (3). To simplify the notation, we define volatility as the variance of returns, $\sigma^2 \equiv E(r_t^2)$, in contrast to its usual definition as the standard deviation of returns.[1] If desired, readers can take square roots wherever required.

Historical Volatility

Historical volatility is perhaps the oldest and simplest volatility model. This model parameterizes current volatility as

$$\tilde{\sigma}^2 = \frac{1}{N-1}\sum_{j=1}^{N} r_{t-j}^2 \qquad (4)$$

Observations in this model get either equal or zero weight. Any observation inside the window of size N gets a weight of $1/(N-1)$, while any observation outside that window gets a weight of zero. The choice of N is arbitrary, but is frequently chosen to be about 60, corresponding to three months if daily data are used and five years if monthly data are used.

The forecast of volatility from this model is simply the current volatility. In other words, volatility is forecasted to be the same as it was over the last N periods. So if the sample variance over the last N observations is 13%, then the forecasted volatility (for any forecast horizon) is 13%.

If N is small enough, this model captures volatility clustering, because predicted volatility will be high if current volatility is high. In fact, rewriting this model as

$$\tilde{\sigma}_t^2 = \frac{1}{N-1} r_t^2 + \frac{1}{N-1} r_{t-1}^2 + \ldots + \frac{1}{N-1} r_{t-N}^2 \qquad (5)$$

and recognizing that r^2 is a measure of the size of a shock, reveals that the historical volatility model can be written as a special case of Equation (3) with $\delta_j = 1/(N-1)$ and $f(x) = x$.

The forecasts created from this model, however, do not exploit the volatility clustering property optimally. To illustrate, if N = twenty-two days, the forecasted volatility will be last month's variance, even though volatility might have been unusually high for the last five days. A better forecast would recognize that last week's volatility was high, and use this to forecast higher near-term volatility.[2]

Essentially, this model is backward-looking, chasing a moving target. We therefore view it as the strawman. Any new forecasting model must perform at least as well as historical volatility for it to be considered reasonable.

ARMA Models on Squared Returns

One class of models that better captures the dynamic properties of volatility is the ARMA (autoregressive moving average) model on squared returns. There are many models in this class, so for illustration purposes we focus on the ARMA (p, 0) model, or the autoregressive model of order p:

$$r_t^2 = \omega + \beta_1 r_{t-1}^2 + \ldots + \beta_p r_{t-p}^2 + \eta_t \qquad (6)$$

This model is simply a linear regression model of current squared returns on lagged squared returns, and can be estimated with ordinary least squares using any statistics package.

Taking expectations of Equation (6), conditional on information up to time t − 1, gives this model's estimate of current volatility:

$$\tilde{\sigma}_t^2 = \omega + \beta_1 r_{t-1}^2 + \ldots + \beta_p r_{t-p}^2 \qquad (7)$$

It is immediately obvious that Equation (7) is a generalization of the historical volatility model, reducing to the historical volatility model if $p = N$, $\omega = 0$, and $\beta_j = 1/(N-1)$. A key advantage of this model over the historical volatility model is that it is data-driven. The weights in this model depend on the data, and the lag length N can be selected from statistical criteria (for example, by maximizing the Akaike criterion), making it depend on the data as well. This model can also be written as a special case of Equation (3), suggesting that it captures volatility clustering.

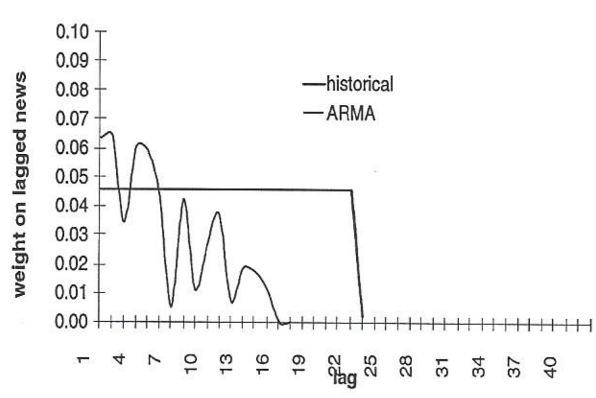
Exhibit 4 gives a comparison of the weights from the historical and ARMA models applied to deutschemark volatility. The estimation is done with ordinary least squares using daily deutschemark returns from 1975 to 1990. The lag length of fifteen is selected from statistical criteria.

The most striking feature in Exhibit 4 is that the ARMA weights drop off to zero. This is expected, because more distant returns should have a lower impact on current volatility than more recent returns. With a large enough data set, we would expect these weights to decline monotonically to zero. The lack of smoothness is caused by small sample problems.

To forecast volatility from the ARMA model, define $\tilde{\sigma}_{t+k|t}^2 \equiv E(r_{t+k}^2 \mid \Psi_t)$ as the forecasted volatility k periods hence, using only current information (Ψ_t). Then the forecast can be calculated recursively from the equation

Exhibit 4

Comparison of Coefficients from ARMA and Historical Models



$\beta_p \tilde{\sigma}^2_{t+k-p|t}$

$$\tilde{\sigma}^2_{t+k|t} = \omega + \beta_1 \tilde{\sigma}^2_{t+k-1|t} + \dots + \cancel{\beta_N} \tilde{\sigma}^2_{t+k-\cancel{N}|t} \qquad (8)$$

p p

β_p

In contrast to the forecasts from the historical volatility model, this forecast depends on the time series properties of volatility. Also, news from the distant past gets a different weight ($\cancel{\beta_N}$) than news from yesterday (β_1). And, finally, this model gives a term structure of volatility, in which the volatility forecast depends on the forecast horizon, k. To see this, suppose ~~N = 2~~ and compare the one-step-ahead forecast (k = 1),

$p = 2$

$$\tilde{\sigma}^2_{t+1|t} = \omega + \beta_1 r_t^2 + \beta_2 r_{t-1}^2$$

to the two-step-ahead forecast (k = 2),

β_p

$$\begin{aligned} \tilde{\sigma}^2_{t+2|t} &= \omega + \beta_1 \tilde{\sigma}^2_{t+1|t} + \dots + \cancel{\beta_N} \tilde{\sigma}^2_{t|t} \\ &= \dots \\ &= \omega(1 + \beta_1) + \\ &\quad (\beta_1^2 + \beta_2) r_t^2 + \beta_1 \beta_2 r_{t-1}^2 \end{aligned}$$

The weights on the squared returns are different for different forecast horizons, meaning that the forecasts will be different. It is straightforward to show that the forecasts eventually converge to

β_p

$$\sigma^2_{\infty|t} = \frac{\omega}{1 - \beta_1 - \dots - \cancel{\beta_N}}$$

GARCH

The GARCH (generalized autoregressive conditional heteroscedasticity) model is a time series model of volatility developed by Engle [1982] and generalized by Bollerslev [1986].[3] In this model, the measure of current volatility is

$$\tilde{\sigma}^2_t = \omega + \alpha r_{t-1}^2 + \beta \tilde{\sigma}^2_{t-1} \qquad (9)$$

The essential difference between the GARCH model and the ARMA model on squared returns is that the ARMA model parameterizes the dynamics of squared returns, while the GARCH model parameterizes the dynamics in the expected value of squared returns (i.e., volatility). Because of this difference, the estimation methods also differ. Least squares regression models do not work for GARCH models because the dependent variable is unobservable. Instead, maximum likelihood methods are usually used. See Bollerslev, Chou, and Kroner [1992] for a description of how to estimate GARCH models.

Rewriting this equation gives an alternative expression for the measure of current volatility from a GARCH model:

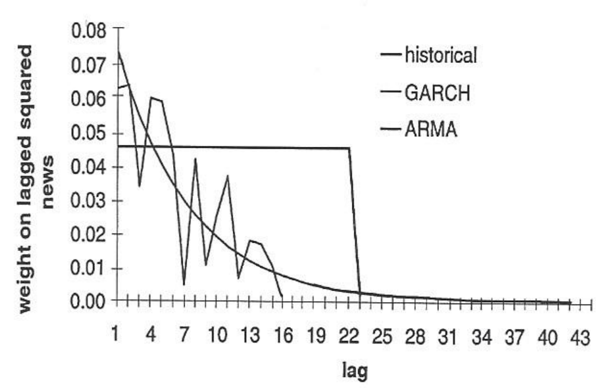
$$\tilde{\sigma}^2_t = \frac{\omega}{1 - \beta} + \alpha \sum_{j=1}^{\infty} \beta^{j-1} r_{t-j}^2 \qquad (10)$$

This equation shows that the GARCH model captures volatility clustering, because it is a special case of Equation (3), with $\delta_i = \alpha\beta^{i-1}$ and $f(x) = x$. This equation also emphasizes the similarity between the GARCH model and the ARMA model.

Notice, for example, that the GARCH measure of volatility is a special case of the ARMA measure. Both define current volatility as a linear function of lagged squared returns, but the GARCH model restricts the weights to decay geometrically, with β as the decay parameter. Exhibit 5 plots the GARCH model weights on top of the ARMA model weights

Exhibit 5

GARCH Model Weights



for the deutschemark. The GARCH model smooths out the weights of the ARMA model. So if the user thinks that there is too much noise in the data to estimate the ARMA weights accurately, then the GARCH model is a reasonable solution.

The equation for forecasting volatility from a GARCH model is

$$\tilde{\sigma}^2_{t+k|t} = \omega\frac{1-\rho^k}{1-\rho} + \rho^{k-1}\tilde{\sigma}^2_{t+1|t} \qquad (11)$$

where $\rho \equiv \alpha + \beta$. This model also gives a term structure of volatility because the forecast depends on the forecast horizon, k. To illustrate, see Exhibit 6, which presents the term structure of forecasted volatility for the deutschemark as of April 26, 1995. Analogous to the ARMA forecasts, this forecasted volatility converges to

$$\tilde{\sigma}^2_{t+\infty|t} = \frac{\omega}{1-\alpha-\beta}$$

as the forecast horizon increases.

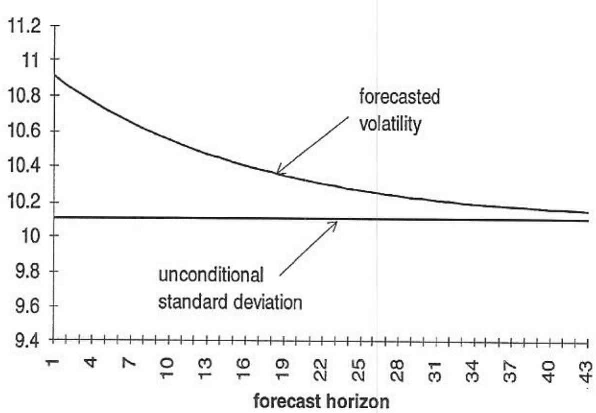
Several extensions of the GARCH model are possible. For example, it might not be realistic to assume that the long-horizon forecasts are always converging to the same constant. The components GARCH model allows this long-run mean to change through time according to the equation (capitalize)

$$\omega_t = \overline{\omega} + \varphi\omega_{t-1} + \theta(r^2_{t-1} - \sigma^2_{t-1})$$

Think of this as a long-run component that gradually adapts to market changes. Importantly, this modification does not change the essential characteristic of GARCH models, that they can be written as special cases of Equation (3) and are therefore models of volatility clustering.

EXHIBIT 6

VOLATILITY FORECASTS DEPEND ON FORECAST HORIZON — TERM STRUCTURE OF GARCH VOLATILITY — DEUTSCHEMARK APRIL 26, 1995



EXPONENTIAL SMOOTHING

The exponential smoothing model defines current volatility as

$$\tilde{\sigma}^2_t = \alpha r^2_{t-1} + (1-\alpha)\,\tilde{\sigma}^2_{t-1}$$

which can be rewritten as

$$\tilde{\sigma}^2_t = \alpha\sum_{j=1}^{\infty}(1-\alpha)^{j-1}r^2_{t-j}$$

Notice that this is a special case of the GARCH model, in which $\omega = 0$ and $\alpha + \beta = 1$. As such, it shares most of the properties of the GARCH model. Specifically, it can be written as a special case of Equation (3), and, as such, addresses the volatility clustering property. This is not necessarily, however, a good forecasting model. Plugging $\omega = 0$ and $\alpha + \beta = 1$ into Equation (11) reveals that the forecasts from this model are

$$\tilde{\sigma}^2_{t+k|t} = \tilde{\sigma}^2_{t|t} = r^2_t$$

The forecasts for any horizon are always equal to the current squared return. So this model overemphasizes volatility clustering, assuming that if volatility is high today, then it will be high forever. Long-horizon forecasts from this model do not revert to any mean, making them (perhaps) less reasonable.

COMMENTS ON THE MODELS

The models presented here are the simplest models in each of their classes, and therefore are easily improved. For example, as presented, all the models share the weakness that positive and negative returns are assumed to have the same impact on volatility. In currency markets, this might be reasonable. But in equity markets, it is commonly believed that negative returns have a greater impact on volatility than positive returns, due, e.g., to leverage-type effects. All these models can easily be modified to address this property by making the weights depend on whether the return is positive or negative. For example, the GARCH model can be generalized to

$$\tilde{\sigma}_t^2 = \omega + \alpha_1 r_{t-1}^2 + \alpha_2 n_{t-1}^2 + \beta \tilde{\sigma}_{t-1}^2$$

where n_{t-1} takes the value r_{t-1} if r_{t-1} is negative, and takes the value zero if r_{t-1} is positive. In this model, the weight on r_{t-1}^2 is $(\alpha_1 + \alpha_2)$ if r_{t-1} is negative, and α_1 if r_{t-1} is positive. We would expect that in equity markets, $\alpha_2 > 0$. This property is commonly referred to as "sign asymmetry." See Engle and Ng [1993] or Nelson [1990].

Another weakness of these models is that they ignore "size asymmetries." It is widely believed that large returns have a different impact on future volatility than normal-sized returns. Specifically, the impact of large returns is shorter-lived than the impact of small returns.

Anecdotally, recall the October 1987 Crash in the equity market. Market volatility returned to its pre-Crash levels very quickly after the Crash (within two weeks). To address this asymmetry, these models can be modified to allow the weights to depend on the size of the crash. There are several ways to do this; Rabemananjara and Zakoian [1993] and Glosten, Jagannathan, and Runkle [1993] present two of them.

A third weakness is that outliers generally have a noticeable effect on the estimated volatility process, so volatility models should minimize the weight of outliers. The statistical literature has several suggestions about how this can be done. The simplest is to give each observation a weight that is inversely proportional to its size when estimating the models.

Several other model improvements are possible. For example, other information (like volume) could be included as a regressor in the volatility models. This might be useful because several studies reveal that other information is a significant determinant of volatility (e.g., Lamoureux and Lastrapes [1990]). Perhaps more important, one can combine market expectations with time series models to improve forecasting performance. Kroner, Kneafsey, and Claessens [1995] show that the model

$$\tilde{\sigma}_t^2 = \omega + \alpha_1 r_{t-1}^2 + \beta \tilde{\sigma}_{t-1}^2 + \gamma ISD_{t-1}^2 ,$$

where ISD is the implied standard deviation from options prices, ~~This model~~ forecasts volatility better than either just the implied volatility or just the GARCH model.

We conclude this section with some cautions on creating volatility forecasting models. First, markets differ, implying that the appropriate volatility forecasting model is market-specific. For example, fixed-income volatility is correlated with yields, so interest rate volatility forecasting models should address this property (Brenner, Harjes, and Kroner [1996]).

Also, a long time span is needed to get an accurate measure of the mean to which long-run volatility forecasts revert, especially in the components GARCH model. But if the time span is too long, then potential structural breaks might appear in the data set. For example, fixed-income markets underwent a structural change between October 1979 and October 1982.

Finally, high-frequency data are required to get an accurate measure of current volatility (Nelson [1992]). But if the frequency is as high as intradaily, then seasonality becomes important (Anderson and Bollerslev [1995] and Ghose and Kroner [1996]).

EVALUATIONS OF FORECASTING ABILITY

The true test of a model's veracity is its ability to forecast. We therefore evaluate these models using several measures of forecasting ability, grouped into statistical evaluations and investment-based evaluations. In all cases, the models are estimated on four currencies (deutschemark, Swiss franc, Japanese yen, and British pound), using daily data from January 1975 to August 1990, and the estimated models are used to construct daily *out-of-sample* twenty-two-day volatility forecasts from September 1990 to April 1995.[4]

Statistical Evaluations

Two statistical properties are required of a good forecasting model. First, the model should be unbiased, meaning that on average it should give the correct forecast. Second, the typical forecast error should be small.

There are two ways to evaluate the bias of a forecasting model. One is to compare the average forecast with the average realized volatility. A more stringent test is to regress the true volatility on the forecasted volatility, and test if the intercept coefficient is zero and the slope coefficient is one. Specifically, one can regress

$$\sigma_t^2 = a + b\tilde{\sigma}_t^2 + \varepsilon_t$$

$$\cancel{\sigma_t^2 = \alpha_0 + \alpha_1\tilde{\sigma}_t^2 + \varepsilon_t}$$

and test if (a, b) $\cancel{(\alpha_0, \alpha_1)}$ = (0, 1). This test follows an F-distribution with (2, N) degrees of freedom.

The results of these evaluations are presented in Exhibit 7. The columns labeled "Average" give the average forecasted twenty-two-day volatility for the four models, along with the average market-based one-month implied volatility and the average realized volatility, where realized volatility is defined as the sample variance over the subsequent twenty-two days.

Generally speaking, the time series models understate realized volatility, while the market-

Exhibit 7

Statistical Evaluations

Deutschemark	MAFE	Average	$\sigma_t^2 = a + b\hat{\sigma}_t^2$: a	b	F
Historical	0.303	0.574	0.432 (0.018)	0.269 (0.027)	0.000
ARMA	0.235	0.458	0.073 (0.042)	1.121 (0.090)	0.405
GARCH	0.231	0.484	0.171 (0.033)	0.860 (0.066)	0.107
Implied	0.264	0.641	0.226 (0.027)	0.564 (0.039)	0.000
Truth		0.587			

British pound	MAFE	Average	$\sigma_t^2 = a + b\hat{\sigma}_t^2$: a	b	F
Historical	0.319	0.563	0.358 (0.019)	0.367 (0.027)	0.000
ARMA	0.259	0.470	0.033 (0.035)	1.133 (0.071)	0.177
GARCH	0.254	0.509	0.143 (0.027)	0.829 (0.048)	0.002
Implied	0.263	0.608	0.085 (0.025)	0.794 (0.038)	0.000
Truth		0.565			

Swiss Franc	MAFE	Average	$\sigma_t^2 = a + b\hat{\sigma}_t^2$: a	b	F
Historical	0.351	0.711	0.504 (0.020)	0.266 (0.024)	0.000
ARMA	0.252	0.615	0.161 (0.044)	0.866 (0.070)	0.163
GARCH	0.250	0.631	0.208 (0.038)	0.768 (0.058)	0.000
Implied	0.258	0.695	0.289 (0.028)	0.580 (0.038)	0.000
Truth		0.693			

Japanese Yen	MAFE	Average	$\sigma_t^2 = a + b\hat{\sigma}_t^2$: a	b	F
Historical	0.281	0.474	0.370 (0.016)	0.266 (0.028)	0.000
ARMA	0.225	0.417	0.163 (0.044)	0.797 (0.102)	0.140
GARCH	0.223	0.423	0.189 (0.037)	0.726 (0.085)	0.006
Implied	0.199	0.471	0.130 (0.021)	0.777 (0.041)	0.000
Truth		0.496			

based forecasts overstate it. This result is probably sample-specific, and need not hold for different time periods or for different asset classes. The columns labeled "F" give the p-value for the F-test of ~~(α_0, α_1)~~ (a, b) = (0, 1). For all four currencies, the historical and implied forecasts fail this test, while the ARMA model passes. The GARCH results are mixed.

The most common measure of the size of the typical forecast error is the MAFE (mean absolute forecasting error), defined as

$$\text{MAFE} = \sum_{i=1}^{T} \left| \tilde{\sigma}_t^2 - \sigma_t^2 \right|$$

The lower the MAFE, the more precise the volatility forecasts. These are presented in the columns labeled "MAFE" in Exhibit 7. For all currencies, the historical volatility gives the highest MAFE, and for all currencies except the yen, the time series models give the lowest MAFE.

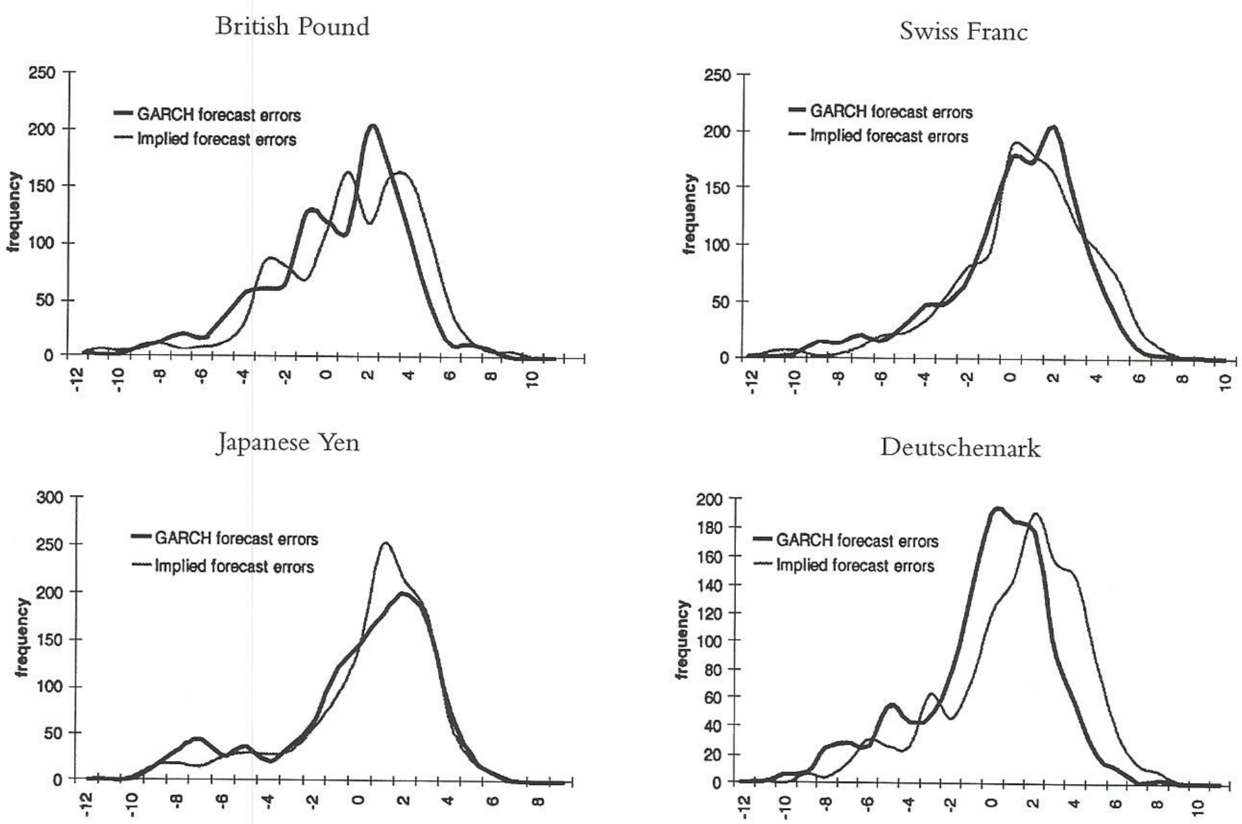
Another statistical way to evaluate the forecasting ability of the models is to examine the distribution of forecast errors. Exhibit 8 presents the distributions of the GARCH and ISD forecast errors for each currency. These distributions are skewed to the left, suggesting that the models are more likely to understate future volatility by large amounts than to overstate future volatility by large amounts. Also, all the models tend to make small positive forecast errors. Finally, for all currencies except the yen, the ISD has a higher probability of making large positive forecast errors than GARCH.

Investment-Based Evaluations

Statistical evaluations generally ignore potential investment opportunities. It is possible for a forecast to be consistently biased and have a large MAFE, even though it contains information that is useful to market participants.

Exhibit 8

Distributions of Forecast Errors



For example, the GARCH-based forecasts might consistently underpredict future volatility, but if the difference between the GARCH forecast and the market forecast (ISD) is much smaller than normal, this might suggest a trading opportunity. Therefore, as an alternative forecast evaluation method, we propose two investment-based evaluations that have positive expected value if the model's forecasting ability is better than the market-based implied volatility.

Specifically, we propose evaluation methods that involve buying and selling straddles and methods that involve buying and selling volatility swaps. We use the GARCH model as the representative time series model, because the evaluations using ARMA models are almost identical.

Consider the following investment-based evaluation. Each day, an investor buys \$1 worth of a thirty-day "historical versus implied volatility swap"[5] if the swap is cheap, and sells \$1 worth of the swap if it is expensive. The swap is cheap if the forecasted twenty-two-day volatility exceeds the implied volatility, and expensive if the forecasted volatility is less than the implied volatility.

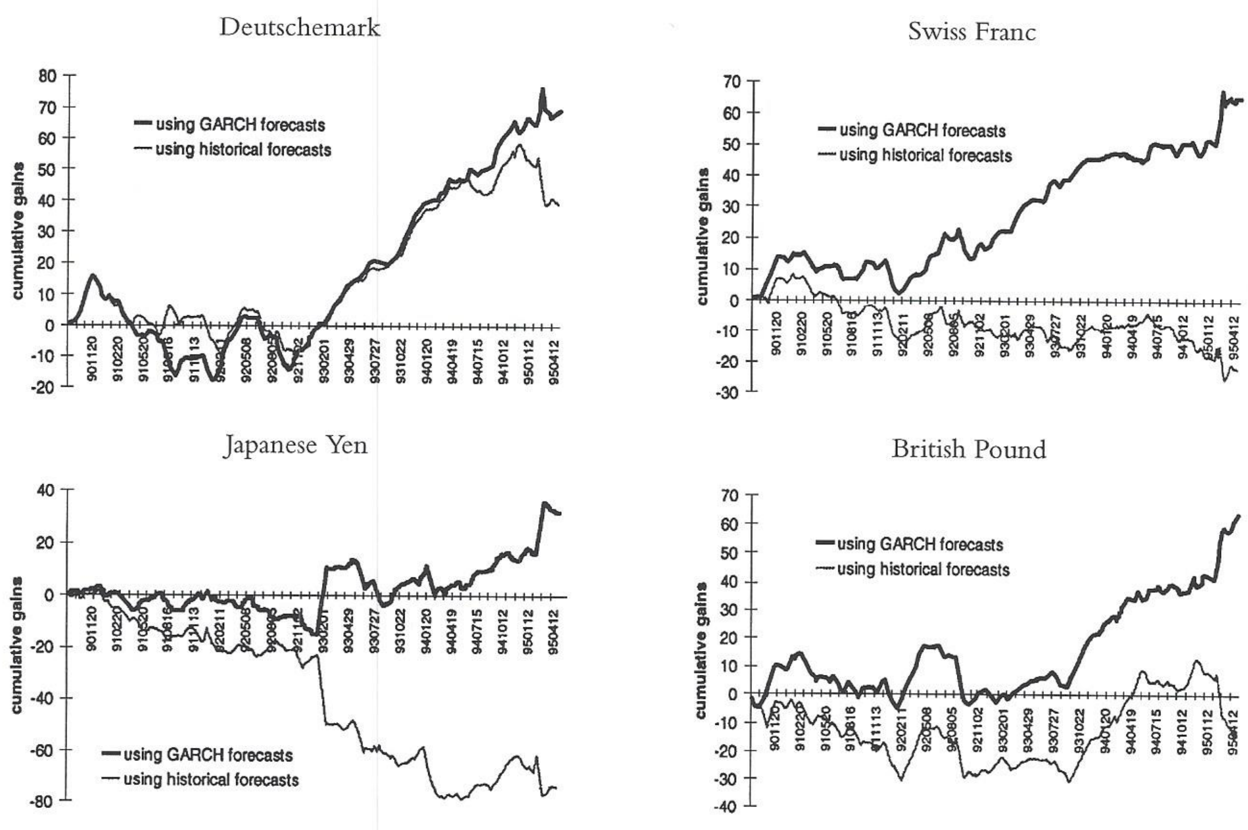
The investor holds this swap until expiration (one month), then puts the proceeds under the pillow. The cumulative "profits" under the pillow are a good evaluation of forecasting power, because any forecast that consistently has information about the future that is not already contained in the ISD should realize positive cumulative wealth.[6] Notice that this is not a trading strategy. Transaction costs are ignored, and the investor is engaging in daily activity no matter how strong or weak the signal is.

Exhibit 9 gives the cumulative wealth realized by this strategy from August 1990 to April 1995 using both the GARCH forecasts and the historical forecasts. Notice first that for all currencies, the GARCH model finishes with positive wealth, suggesting that the time series forecasts might contain information about future volatility that is not contained in ISDs.

Second, notice that, except for the

EXHIBIT 9

BUY (SELL) SWAP IF IT'S CHEAP (EXPENSIVE)



deutschemark, the historical-based volatility forecasts finish with negative wealth, suggesting that they do not contain information about mispricing of ISDs. In all cases, the GARCH-based evaluations outperform the historical-based evaluations.

Similar conclusions arise if the investment instrument is a one-month at-the-money straddle instead of a volatility swap. Exhibit 10 gives the cumulative wealth using straddles. Again, the time series forecasts outperform the historical forecasts for all four currencies. Also for all four currencies, the GARCH forecasts finish with positive wealth, although the evidence is weak for the British pound. Finally, for two of the four currencies (yen and pound), the historical forecasts finish with zero or negative wealth.

POTENTIAL APPLICATIONS

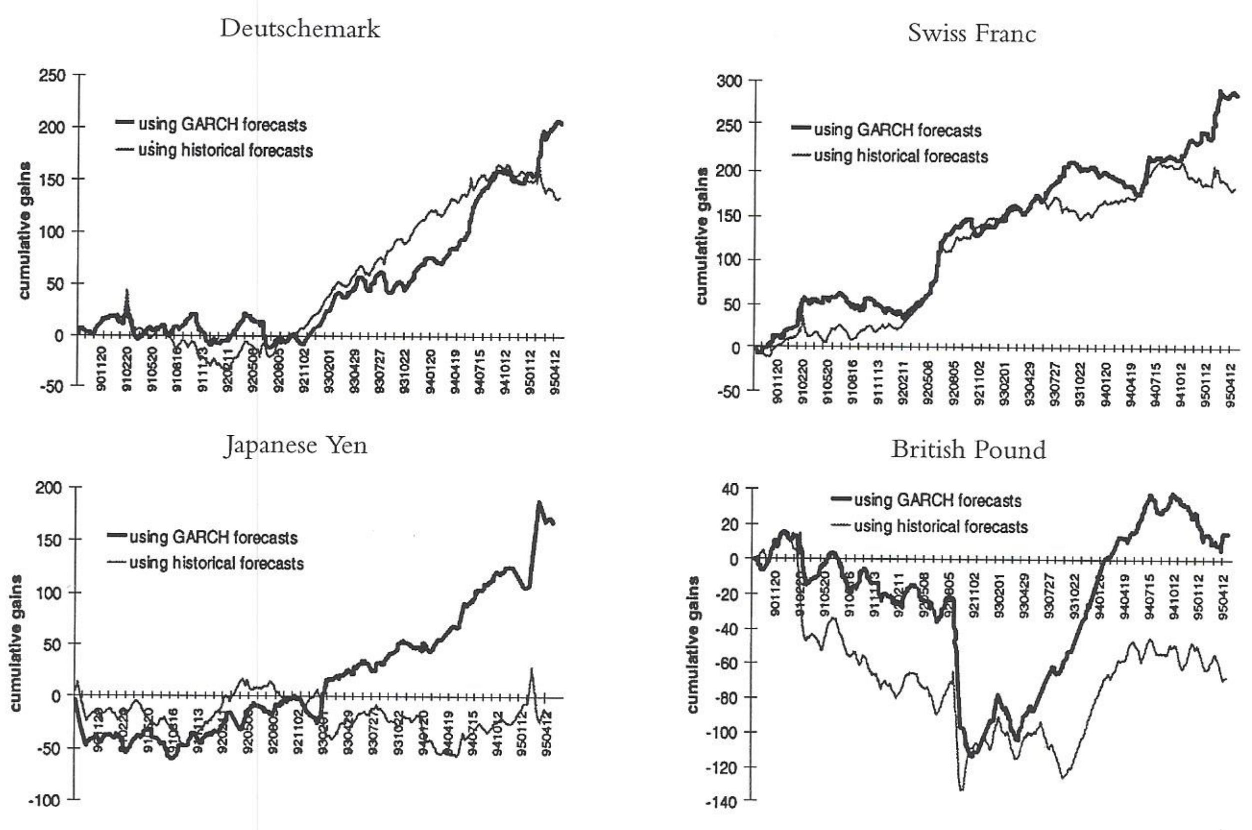
Both the statistical and investment-based evaluations suggest that time series models of volatility might provide better volatility forecasts than either historical forecasts or market-based implied volatility. Therefore, these forecasts should be useful for a wide range of potential applications.

For example, they could be used to price options (see Noh, Engle, and Kane [1994] or Mustafa [1988]). They are relevant for portfolio hedging because they can be used to determine whether to use real or synthetic options. By extending these models in the multivariate direction, they can be used to take correlation bets by trading basket options or spread options (see Engle and Kroner [1995]). They could be used to take term structure bets by trading forward volatility agreements or calendar spreads, because these models give a term structure of volatility. Asset volatility is correlated with asset returns, so they could be used as a potential return forecasting signal. And, as suggested earlier, they could be used to take volatility bets by trading historical versus implied volatility swaps, straddles, butterfly spreads, and so on.

Consider the following specific application,

Exhibit 10

Buy (Sell) Straddle If It's Cheap (Expensive)



which takes volatility bets by trading straddles. Suppose an investor buys one-month at-the-money straddles when they are very inexpensive, with the number of straddles purchased proportional to the forecasted profit opportunity. Specifically, she purchases an amount equal to the difference between the GARCH forecast and the implied, but only if this difference is at least 1%. The investor uses this threshold to avoid taking costly bets on a weak signal.

If GARCH volatility is forecasted to be 16%, and the implied is 13%, she buys \$3 worth of the straddle; if GARCH volatility is 13.5% and the implied is 13%, then she stays out of the market. Assume she never shorts the straddle because of the high probability of significant losses, and that she finances all purchases by borrowing at the risk-free rate. She holds the straddle until expiration, and puts her proceeds under the pillow. Assuming transaction costs of 1/2%, the light line in Exhibit 11 gives the cumulative profits under her pillow since August 1990.

The most obvious conclusion from Exhibit 11 is that the strategy is very effective. For all four currencies, she finishes with a significant amount of money under her pillow, even though her initial investment was zero. A second conclusion is that she is only in the market infrequently. Of the 1,204 days in the sample, she ranges from being in the market only 27 days for the deutschemark to 250 days for the yen. But when she is in the market, it usually pays off, as she tends to make gains and lock them in.

To better understand these results, the heavy lines in Exhibit 11 ~~also~~ plots the cumulative profits from the strategy of always buying \$1 worth of the straddle, no matter what the forecast says. Consider, for example, the Swiss franc. Notice that until about May 1992, owning straddles had minimal profit potential. Her strategy had her out of the market most of this time. The first real profit potential occurs in May and June 1992, and she is in the market buying straddles during these months.

Exhibit 11

Buy Straddle If It's Very Cheap and Put Proceeds Under Pillow

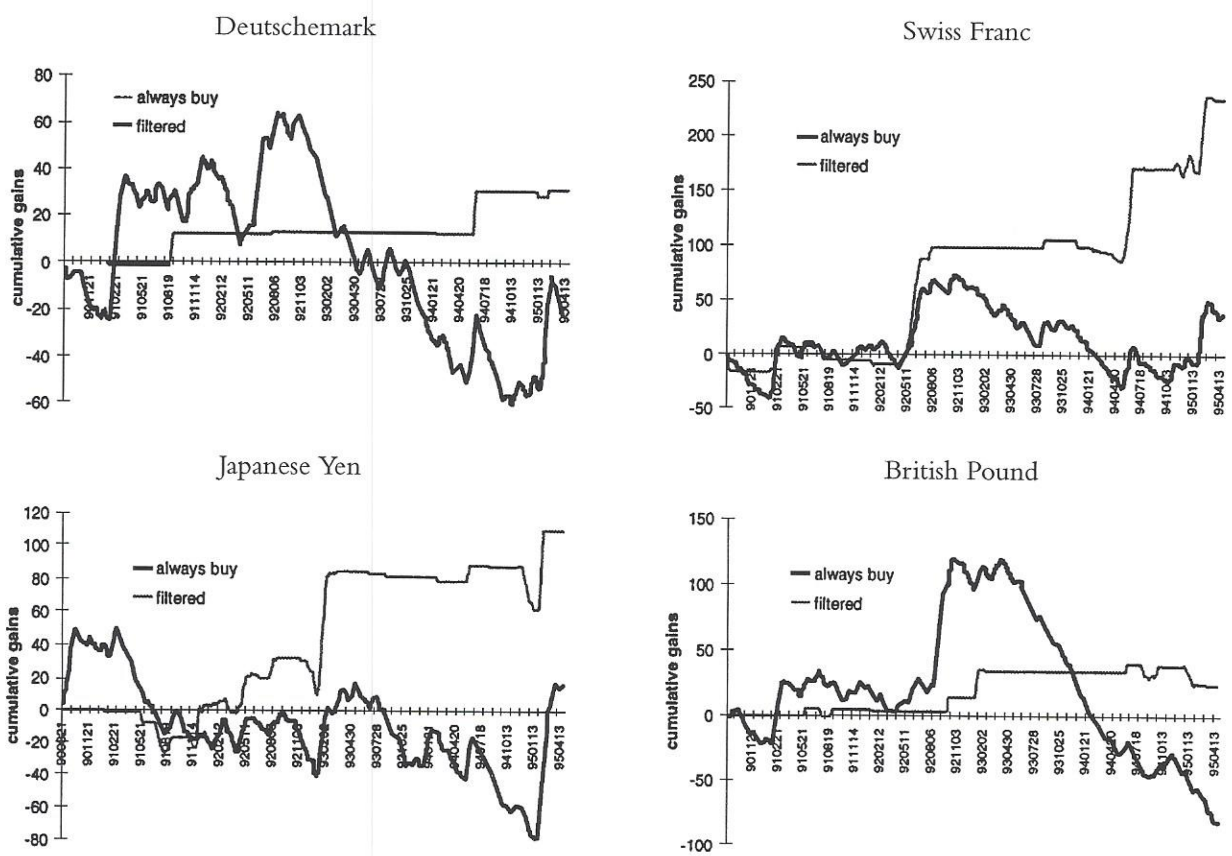
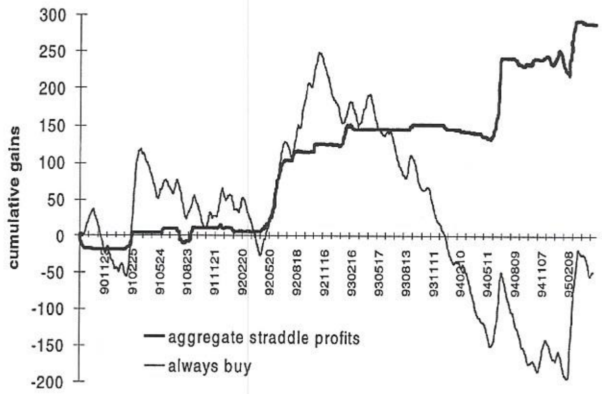


Exhibit 12

Total Money Under the Pillow Using Straddle Strategy



For the next two years, buying straddles is generally a losing proposition, and she is out of the market for most of this period. The next real profit opportunity is in June 1994, and she is then back in the market. She then locks in her gains, and reenters the market in time to catch the profit opportunity in early 1995.

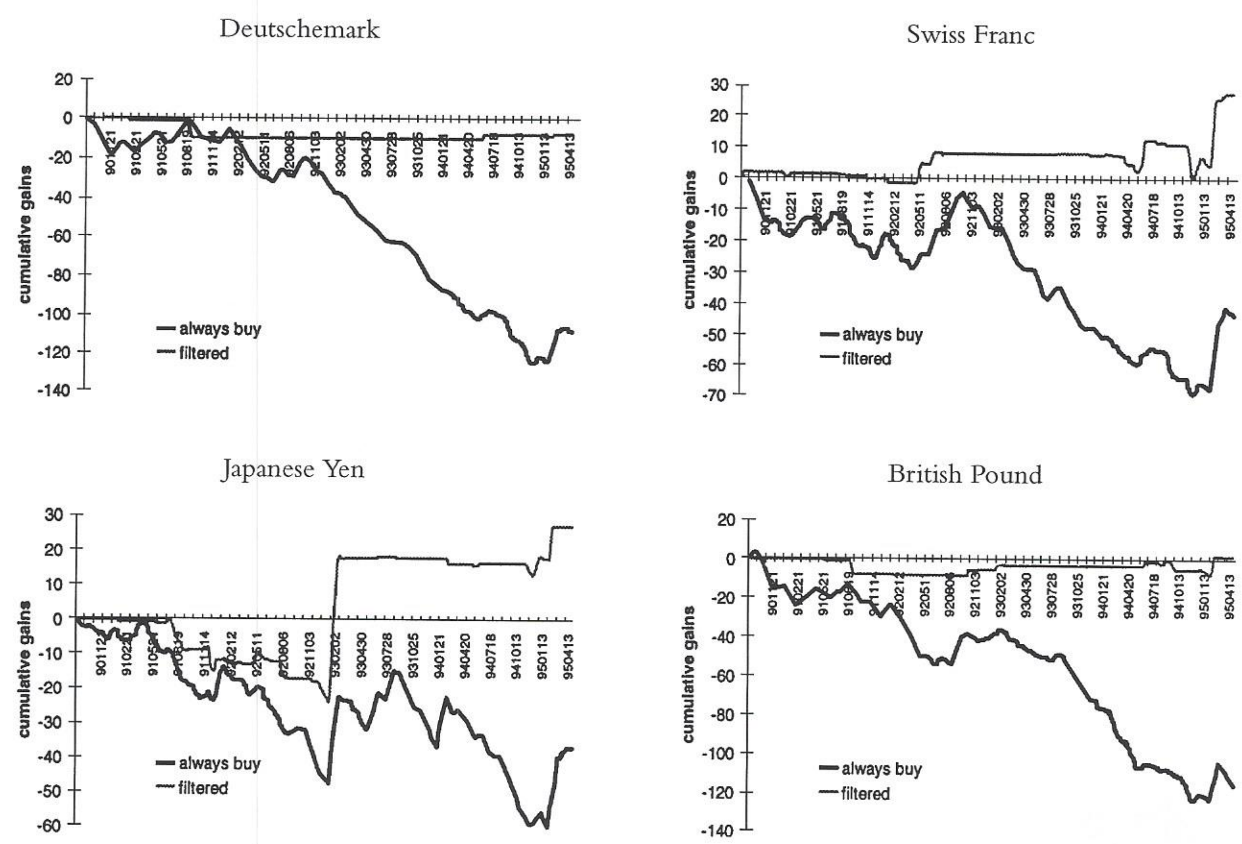
A similar analysis is possible for the other three currencies. In fact, perhaps her only major missed opportunity is the British pound during the currency crisis. Her strategy had her only testing the market, when she would have been better off investing heavily in straddles. Exhibit 12 plots the aggregate profits from all four currencies.[7]

A slightly different picture emerges if the investment instrument is a historical versus implied volatility swap instead of a straddle. Again, we assume the investor is in the market buying a notional value of the volatility swap proportional to the GARCH-implied spread, but only if the spread is at least 1%. The transaction costs are again assumed to be 1/2%.

Exhibit 13 plots the cumulative profits from this strategy. Our investor earns essentially zero profits from her deutschemark and pound invest-

Exhibit 13

Buy Swap If It's Very Cheap and Put Proceeds Under Pillow



ments, and earns only small profits from her yen and franc investments. It is obvious from the "always buy" lines, however, that purchasing the deutschemark and pound swaps is generally a bad idea for the entire sample period. Smartly, our investor was rarely in the market, and therefore earned almost zero profits. For the yen and the franc, there were only a few profit opportunities, and our investor was in the market for many of them. Exhibit 14 plots the aggregate profits from all four currencies.

CONCLUSIONS

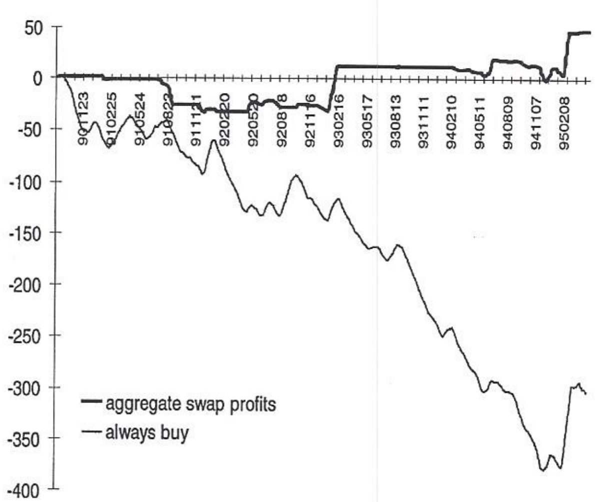
Several conclusions can be drawn from this study. First, short-term volatility is forecastable. Second, any model that captures volatility clustering has the potential to give reasonable volatility forecasts, because these models recognize that if current volatility is high, then volatility will probably remain high for the near future.

One class of models that does this is the time series models where the driving variable is the size of returns. Special cases of this include GARCH models, exponential smoothing, and ARMA models on squared returns.

Third, there are many potential applications of good volatility forecasts. Several are mentioned here, two of which are examined in detail. Specifically, we evaluate trading strategies that have the investor buying straddles or volatility swaps if the time series forecasting model indicates that they are inexpensive. Profit opportunities are available, even after accounting for transaction costs.

EXHIBIT 14

TOTAL MONEY UNDER THE PILLOW USING SWAP STRATEGY



ENDNOTES

The author thanks Tim Bollerslev, Jeff Hord, Asriel Levin, L. Sankarasubramanian, Vikas Srivastava, and members of the Advanced Strategies Group at ~~BZW~~ Barclays Global Investors for helpful comments.

[1]Also, to keep the notation simple, we assume that the expected, or average, return is zero. This assumption is reasonable for sufficiently high-frequency data (e.g., daily).

[2]This should not be viewed as a criticism of the model, but rather as a criticism of how the model is generally applied in forecasting exercises. This same model could be used to construct forecasts that do not suffer from this problem. When we discuss ARMA models in the next section, we will see how this can be done.

[3]For a survey of its applications in finance, see Bollerslev, Chou, and Kroner [1992].

[4]The forecasted twenty-two-day volatility is the sum of the forecasts from horizon k = 1 to horizon k = 22.

[5]A "historical versus implied volatility swap" has the buyer agreeing to pay the seller the implied standard deviation, and receiving the realized sample standard deviation. So if the (annualized) implied standard deviation on February 1 is 13%, and realized volatility in February was 15%, then on February 28 the buyer would receive from the seller $2 times the notional amount.

[6]An important weakness of this evaluation is that bad forecasts can still finish with positive profits. Any transaction for which the implied exceeds or falls short of both the realized and the GARCH volatility will make positive profits. To illustrate with an extreme situation, if the ISD always overstates realized volatility by a small amount, but the GARCH model always understates it dramatically, this strategy will always sell the swap, making money on every transaction, even though the ISD is a much better forecast than GARCH. Therefore, these evaluations should be interpreted as necessary but not sufficient criteria for the forecasts to meet.

[7]Further analysis suggests that these encouraging results cannot be attributed to market drift. Specifically, after transaction costs, delta-neutral straddles are even more profitable than the straddles reported here.

REFERENCES

Anderson, Torben, and Tim Bollerslev. "Intraday Seasonality and Volatility Persistence in Financial Markets." Kellogg School of Business, Northwestern University, 1995.

Attiyeh, Gregory. "Volatility Persistence in Strategic Models: A Theoretical and Experimental Investigation." Ph.D. dissertation, Department of Economics, University of Arizona, 1995.

Bollerslev, Tim, Ray T. Chou, and Kenneth F. Kroner.

"ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence." *Journal of Econometrics*, 52 (1992), pp. 5-59.

Bollerslev, Tim, and Ian Domowitz. "Price Volatility, Spread Volatility and the Role of Alternative Market Mechanisms." *Review of Futures Markets*, 10 (1991), pp. 78-102.

Brenner, Robin J., Richard H. Harjes, and Kenneth F. Kroner. "Another Look at Models of the Short-Term Interest Rate." *Journal of Financial and Quantitative Analysis*, 31 (1996), pp. 85-107.

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

Engle, Robert F., Takatoshi Ito, and Wen-Ling Lin. "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market." *Econometrica*, 58 (1990), pp. 525-542.

Engle, Robert F., and Kenneth F. Kroner. "Multivariate Simultaneous Generalized ARCH." *Econometric Theory*, 11 (1995), pp. 122-150.

Engle, Robert F., and Victor K. Ng. "Measuring and Testing the Impact of News on Volatility." *Journal of Finance*, 48 (1993), pp. 1749-1778.

Ghose, Devajyoti, and Kenneth F. Kroner. "Temporal Aggregation of High-Frequency Financial Data." Department of Economics, University of Arizona, 1996.

L.R. R. D.

Glosten, **X.X.**, **~~X~~X.** Jagannathan, and **~~X~~X.** Runkle. "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance*, 48 (1993), pp. 1779-1801.

Kroner, Kenneth F., Kevin P. Kneafsey, and Stijn Claessens. "Forecasting Volatility in Commodity Markets." *Journal of Forecasting*, 14 (1995), pp. 77-95.

Lamoureux, Chris, and William Lastrapes. "Heteroscedasticity in Stock Return Data: Volume versus GARCH Effects." *Journal of Finance*, 45 (1990), pp. 221-229.

Mizrach, Bruce. "Learning and Conditional Heteroscedasticity in Asset Returns." Working paper #95-26, Department of Economics, Rutgers University, 1995.

Mustafa, Choudhury. "The Persistence and Implied Persistence of Volatility of Stock Returns." Ph.D. dissertation, Department of Economics, University of California at San Diego, 1988.

Nelson, Daniel B. "Conditional Heteroscedasticity in Market Returns: A New Approach." *Econometrica*, 59 (1990), pp. 347-370.

——. "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model." *Journal of Econometrics*, 52 (1992), pp. 61-90.

Noh, Jason, Robert F. Engle, and Alex Kane. "Forecasting Volatility and Option Prices of the S&P 500 Index." *Journal of Derivatives*, 2 (1994), pp. 17-30.

Rabemananjara, R., and J.M. Zakoian. "Threshold ARCH Models and Asymmetries in Volatility." *Journal of Applied Econometrics*, 8 (1993).

Institutional Investor Journals
A Division of Institutional Investor, Inc.
488 Madison Avenue, New York, NY 10022
Phone (212) 224-3599 Fax (212) 224-3527

COPYRIGHT AGREEMENT FORM FOR FREELANCERS

The article on the subject of

Creating and Using Volatility Forecasts

is herewith submitted for publication to *Derivatives Quarterly*. It has not been published before. I, as an author of the article and to the extent of my contribution to it, hereby transfer to *Derivatives Quarterly* all my rights to the article under the existing copyright laws, except for the following rights which I as the author specifically retain:

1. The right to make copies of all or part of the published article for my own use; and

2. The right to use or adapt all or part of this material in articles, speeches, or a book of which I am the editor or principal author.

I agree that uses under the latter circumstance will carry the copyright notice that appears in the issue of *Derivatives Quarterly* in which the article appears.

Name Kenneth F Kroner

Signature Kenneth F Kroner

Date Dec 20/96

Address Barclays Global Investors
45 Fremont St
San Francisco, CA 94105

Phone # (415) 597-2827

Please fill out and mail or fax back immediately upon receipt to:

Noelle Schultz
Institutional Investor
488 Madison Avenue
New York, NY 10022
Phone: (212) 224-3193
Fax: (212) 224-3527

DERIVATIVES QUARTERLY

December 19, 1996

Re: *Derivatives Quarterly*
Winter 1996 issue

Dear Mr. Kroner:

I am sending you the galley of the following manuscript, for your review:

"Creating and Using Volatility Forecasts"

Please read these pages over carefully, make any necessary changes, and fax back to me only the pages needing correction no later than December 26, 1996, to ensure that we can meet our scheduled publication date. The date was moved forward one month, and this issue is scheduled for publication the last week of January 1997. All changes are accepted at the discretion of the copy editor.

I am also enclosing a copyright agreement form for freelancers. We are required, as part of our copyright registration agreement with the Library of Congress, to have a signed form on file for every article we publish. It is also necessary because Institutional Investor publications are listed with microfiche and microfilm services.

If you have any problems with this form, please contact me immediately, as we cannot proceed with production until we come to an agreement. Thank you for your prompt attention to this matter.

I would be happy to assist you with any questions or comments you may have. Please feel free to contact me at (212) 224-3193 or fax (212) 224-3527.

Sincerely,

NS

Noelle Schultz
Editorial Production Director

P.S. your article needed only very little copy editing for style. We thought it was one of the smoothest and ~~clear~~ clearest articles we have seen come through here!