ALTERNATIVE MODELS FOR CONDITIONAL STOCK VOLATILITY*

Adrian R. PAGAN

University of Rochester, Rochester, NY 14627, USA
Australian National University

G. William SCHWERT

University of Rochester, Rochester, NY 14627, USA
National Bureau of Economic Research

This paper compares several statistical models for monthly stock return volatility. The focus is on U.S. data from 1834–1925 because the post-1926 data have been analyzed in more detail by others. Also, the Great Depression had levels of stock volatility that are inconsistent with stationary models for conditional heteroskedasticity. We show the importance of nonlinearities in stock return behavior that are not captured by conventional ARCH or GARCH models. We also show the nonstationarity of stock volatility.

1. Introduction

Over the last decade several models of conditional volatility in economic time series have been proposed. Basic to all these suggestions is the notion that volatility can be decomposed into predictable and unpredictable components, and interest has largely centered on the determinants of the predictable part. For financial series this concern with the predictable component of volatility is motivated by the fact that, in many models, the risk premium is a function of it.

By definition, the predictable component of volatility in a series is the conditional variance of that series, σ_t^2. The different ways of modeling σ_t^2 reflect different answers to two basic questions. First, how does σ_t^2 vary with information available at time t; that is, what is the nature of the conditioning set F_t? Second, what does the mapping between information and σ_t^2 look like? Of these two questions, the first has to be dispensed with summarily.

Because of the large range of variables whose volatility has been measured, it is impossible to be precise about conditioning variables, other than to say that the history of the series being analyzed is the most popular choice. We also make that choice by focusing on univariate time series techniques. The debate over the mapping between σ_t^2 and conditioning variables can be more fruitfully analyzed out of the context of particular applications, and it is this question that we concentrate on in this paper.

Suppose we write the series y_t to be modeled as $y_t = x_t'\beta + u_t$, where x_t is a set of variables affecting the conditional mean of y_t, while u_t is an error term with zero mean and conditional variance $E(u_t^2|F_t) = \sigma_t^2$. Then Engle (1982) proposed that

$$\sigma_t^2 = \sigma^2 + \sum_{k=1}^{q} \alpha_k u_{t-k}^2, \tag{1}$$

the ARCH(q) model. Bollerslev (1986) generalized this to

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{k=1}^{q} \alpha_k u_{t-k}^2, \tag{2}$$

the GARCH(p, q) model, and Engle and Bollerslev (1986) extended GARCH to the class of integrated GARCH (IGARCH) models that have the restriction $\sum \beta_j + \sum \alpha_k = 1$. As Bollerslev (1988) records, the class of GARCH models has been extensively applied with some success. Nevertheless, several authors have felt that these models are too restrictive, because of their imposition of a quadratic mapping between the history of u_t and σ_t^2. Nelson (1988) argued that stylized facts associated with Christie (1982) and Black (1976) imply that σ_t^2 be an asymmetric function of the past data, and he modified the conditional variance to

$$\ln \sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2 + \sum_{k=1}^{p} \alpha_k \left[\theta \psi_{t-k} + \gamma \left(|\psi_{t-k}| - (2/\pi)^{1/2} \right) \right], \tag{3}$$

where $\psi_t = u_t/\sigma_t$. By modeling the logarithm of the variance $\ln \sigma_t^2$, it is not necessary to restrict parameter values to avoid negative variances as in the ARCH and GARCH models. An obvious name for the model in (3) is the exponential GARCH (EGARCH) model. To identify the parameters, γ is set to 1.

Hamilton (1988, 1989) proposed a bivariate state model in which σ_t^2 was a linear function of the conditional probability that the economy was in a state

$S_t = 1$, rather than the alternative $S_t = 0$. Because the conditional probability is a nonlinear function of F_t, once again this represents a departure from the GARCH class of volatility measures. The exact mapping between σ_t^2 and F_t induced by his two-state approach depends on the data, and this raises the broader issue of whether one might allow the data to determine the unknown function. Pagan and Ullah (1988) argued that nonparametric estimation methods could be used for this purpose, and Pagan and Hong (1988) gave some examples of where there seemed to be gains in doing nonparametric estimation rather than following the parametric formulations such as GARCH. Very little of a comparative nature has emerged about these methods. For this reason it is of interest to apply each of the techniques mentioned above to the same data set, with the aim of investigating the different implications each might have for the predictability of volatility. The following section selects a series on monthly stock returns from 1834 to 1925 as the basis for such a comparison.

2. Estimation of stock return volatility

We concentrate on monthly stock returns from 1834–1925, previously analyzed by Schwert (1989b). He gives details on the construction of the data and places it in an historical context. In fact, the series extends through 1987 but, because French, Schwert, and Stambaugh (1987) and Nelson (1988) have previously worked with the data from 1926 onward, it is useful to concentrate on a sample that has not received much attention. Furthermore, many of the models and estimators we consider impose *covariance stationarity* on the data. There is strong evidence that the stock return series is not covariance stationary when the period of the Great Depression (1929–1939) is included.[1] If this is true, models such as Hamilton's can be immediately rejected as inappropriate. Moreover, the assumptions underlying nonparametric estimators would also be violated, and one could not justify their usage based on asymptotic theory. Some assessment of whether the data are covariance stationary is therefore mandatory.

2.1. Recursive variance plots

Because covariance stationarity implies that the unconditional variance of the data is a constant over time, a simple graphical view of the likelihood of such constancy is available from a plot of the *recursive* estimates of the variance of the series against time [see Mandelbrot (1963)]. If $\hat{u}_t$ is the difference between the stock return and an estimate of its conditional mean

[1] Schwert (1989a) stresses that stock volatility was unusually high during the Depression relative to the volatility of other important macroeconomic series.
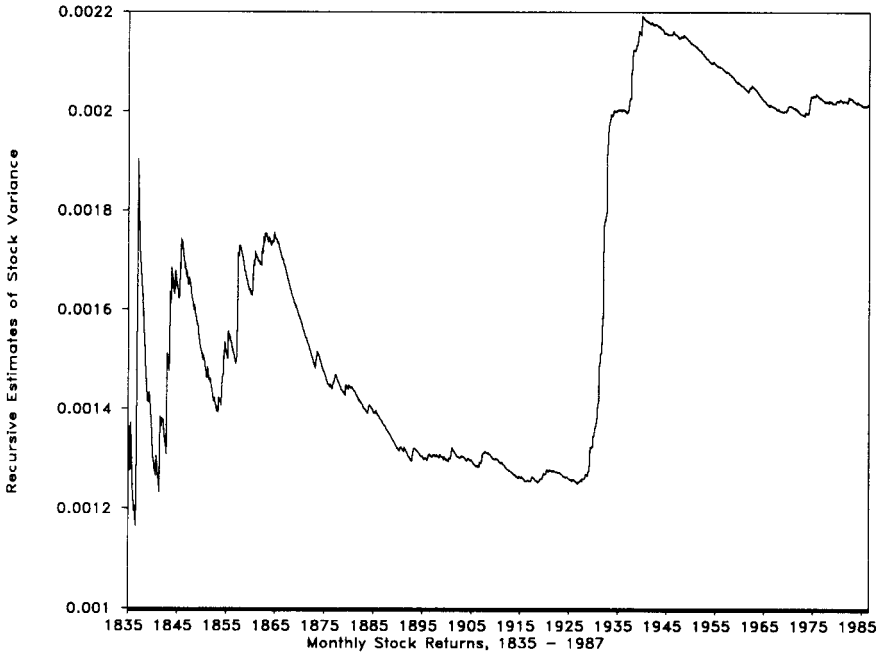
Fig. 1. Recursive estimates of the monthly stock return variance, 1835–1987.

described in section 2.2,

$$\hat{\mu}_2(t) = t^{-1} \sum_{k=1}^{t} \hat{u}_k^2 \tag{4}$$

is the recursive estimate of the unconditional variance at time t. Fig. 1 displays the plot of this against t for 1835 to 1987. There are three distinct phases. In the first, ending around 1866, the unconditional variance estimate is quite erratic. After that, the estimate is very stable until it jumps to a much higher level around 1930. It is this latter jump that is the most striking feature of the data and it suggests that data before 1930 has a different variance from that after 1930. One might argue that the pre- and post-1866 data are also different, although the switch from the Macaulay (1938) to the Cowles (1939) data occurring near that time could explain part of the aberrant behavior.

We have done a variety of tests for whether the unconditional variance is constant over the 1834–1987 period, and they all reject at conventional levels. For the 1834–1925 period, however, most of these tests do not reject

covariance stationarity at small significance levels. Hence, we use this sample period for the remainder of our analysis.

2.2. Modeling the conditional mean return

There is a long history of arguments in the analysis of stock returns that the mean return exhibits little predictability from the past. Qualifications to this conclusion are the existence of a possible moving average error term induced by nonsynchronous data and calendar effects. In the representation $y_t = x_t'\beta + u_t$, y_t being stock returns, x_t would be monthly dummies, and u_t would be an MA(1), $e_t + \theta e_{t-1}$. To account for these effects, we regressed out twelve monthly dummies to get $\hat{u}_t$, and then $\hat{u}_t$ was regressed against $\hat{u}_{t-1}, \ldots, \hat{u}_{t-12}$. Only lags 1, 2, 3 and 10 seemed to be significant. The point estimates for the first four lags are 0.27, -0.10, 0.07, -0.02. The alternating signs and size suggest that this is compatible with an MA(1) with parameter around 0.3. We decided to approximate this MA effect with an autoregression, so that $\hat{e}_t$ was computed as the residuals from the regression of $\hat{u}_t$ against $\hat{u}_{t-1}, \ldots, \hat{u}_{t-10}$. The $\hat{e}_t$ are then the raw data. Central to this procedure is the belief that there are no dependencies in the conditional mean other than linear ones. Nonparametric estimates of conditional mean functions reported later support this assumption.

The task is to model the conditional variance of the series $\hat{e}_t$. To do this, a set of conditioning variables F_t must be chosen and a decision made about how σ_t^2 relates to F_t. We decided to keep F_t as a function of the history of returns alone, and this meant that F_t could be constructed from either $\{\hat{u}_{t-j}\}$ or $\{\hat{e}_{t-j}\}$. If an infinite number of conditioning variables was possible there would be no difference between these, as they are just different linear combinations of y_{t-j}. Because we must restrict the lags to a finite set, differences can arise. We adopt $\{\hat{e}_{t-j}\}$ as the basis of the conditioning set, as this simplifies comparisons with GARCH models. Both $\{\hat{e}_{t-j}\}$ and $\{\hat{u}_{t-j}\}$ were always tried, however, and there were no important discrepancies in results. A finite number of lags was selected by considering the regression of $\hat{e}_t^2$ against $\hat{e}_{t-1}^2, \ldots, \hat{e}_{t-12}^2$. This regression yields the partial autocorrelation function of the $\hat{e}_t^2$. It is important to recognize that the error terms will be heteroskedastic and to adjust t-statistics with the method of White (1980). The difference in the ordinary and robust standard errors is dramatic, with t-statistics of the estimated coefficients of $\hat{e}_{t-1}^2$, $\hat{e}_{t-2}^2$, and $\hat{e}_{t-7}^2$ falling from (6.24, 4.63, and 3.15) to (2.16, 1.77, and 1.84), while that for $\hat{e}_{t-8}^2$ went from -1.45 to -2.03. The t-statistics for the remaining lags were small. Based on this evidence, we concluded that $F_t^4 = \{\hat{e}_{t-1}, \hat{e}_{t-2}, \hat{e}_{t-7}, \hat{e}_{t-8}\}$ should suffice as the broadest set of conditioning variables, but we also conducted experiments with $F_t^2 = \{\hat{e}_{t-1}, \hat{e}_{t-2}\}$ and $F_t^1 = \{\hat{e}_{t-1}\}$. To anticipate later develop-

ments, most of the information is in F_t^1, but the expansion to the larger set F_t^4 does improve the prediction of $\hat{e}_t^2$.

Having chosen F_t, it only remains to describe the set of methods employed to estimate σ_t^2. Because ten lags were used in constructing $\hat{u}_t$, and a further eight if F_t^4 was selected, the sample size was always July 1835 to December 1925, yielding 1086 observations. More observations were available when F_t^1 or F_t^2 are the conditioning sets, but working with a variable sample size makes it hard to compare the different results.

2.3. A two-step conditional variance estimator

Because $E(e_t^2|F_t) = \sigma_t^2$, a simple two-step estimator of σ_t^2 can be found as the predictions from the regression of $\hat{e}_t^2$ against $\{\hat{e}_{t-1}^2, \ldots, \hat{e}_{t-8}^2\}$ [see Davidian and Carroll (1987)]. The underlying model of volatility here is

$$\sigma_t^2 = \sigma^2 + \sum_{k=1}^{8} \alpha_k \hat{e}_{t-k}^2, \tag{5}$$

and all one does is replace σ_t^2 by $\hat{e}_t^2 + (\sigma_t^2 - e_t^2) + (e_t^2 - \hat{e}_t^2) = \hat{e}_t^2 + v_t$. It is easy to show that the term $(e_t^2 - \hat{e}_t^2)$ does not affect the limiting distribution of $\hat{\alpha}_{\text{OLS}}$, so v_t behaves like $(\sigma_t^2 - e_t^2)$, which is a martingale difference with respect to the sigma field generated by F_t. Ordinary least squares is therefore a consistent estimator, although not an efficient one. Efficiency could be improved by doing weighted least squares with $\hat{\sigma}_t^{-1}$ as weights, but the nonnormality of v_t also suggests that adaptive estimation of α might be preferable. The role of the two-step estimator is that of a benchmark, however, and the R^2 of 0.089 between $\hat{\sigma}_t^2$ and $\hat{e}_t^2$ sets a limit to which other models can be compared.

2.4. A GARCH model

The two-step estimator is effectively an eighth-order ARCH model and an obvious extension is to see if a GARCH specification would be superior. French, Schwert, and Stambaugh (1987) fitted a GARCH(1, 2) model to y_t over the period 1928–1984, although the second ARCH parameter α_2 was small. We estimated a GARCH(1, 2) model for $\hat{e}_t$ for 1835–1925. French, Schwert, and Stambaugh allowed for an MA(1), $u_t = e_t + \theta e_{t-1}$, and we did the same here, although since $\hat{e}_t$ has been purged of a tenth-order autoregression, the MA term was not significant. After estimation, the following model for σ_t^2 was found (t-values in parentheses):

$$\hat{\sigma}_t^2 = \underset{(3.65)}{0.000239} + \underset{(6.11)}{0.571}\ \hat{\sigma}_{t-1}^2 + \underset{(4.38)}{0.158}\ \hat{e}_{t-1}^2 + \underset{(1.35)}{0.064}\ \hat{e}_{t-2}^2. \tag{6}$$

A diagnostic test advocated by Pagan and Sabau (1987), involving the regression of $\hat{e}_t^2$ against unity and $\hat{\sigma}_t^2$, gave an estimated coefficient on $\hat{\sigma}_t^2$ of 0.827 in table 1, with a t-statistic of -0.60 for testing the null that the coefficient is unity [implied by the restriction $E(e_t^2|F_t) = \sigma_t^2$]. For this situation, however, where we are testing an ARCH rather than an ARCH-M model, results in Sabau (1988) show that the test is probably rather weak. A point to note is that the point estimates are compatible with the idea that σ_t^2 is generated by a GARCH rather than IGARCH process. The R^2 between $\hat{\sigma}_t^2$ and $\hat{e}_t^2$ is 0.067, which is less than the R^2 for the two-step method.[2]

2.5. An exponential GARCH model

The exponential GARCH(1, 2) model allows lagged shocks to have an asymmetric effect on conditional volatility. In particular, the evidence in Black (1976), Christie (1982), French, Schwert, and Stambaugh (1987), Nelson (1988), and Schwert (1990) suggests that negative stock returns lead to larger stock volatility than equivalent positive returns. We estimate an EGARCH(1, 2) model (t-values in parentheses):

$$\ln\hat{\sigma}_t^2 = \begin{array}{c} -1.73 \\ (-3.90) \end{array} + \begin{array}{c} 0.747 \\ (11.62) \end{array} \ln\hat{\sigma}_{t-1}^2 + \begin{array}{c} 0.262 \\ (5.21) \end{array} Z_{t-1} + \begin{array}{c} 0.124 \\ (2.06) \end{array} Z_{t-2}, \quad (7)$$

where

$$Z_{t-k} = \left[\left(|\hat{\psi}_{t-k}| - (2/\pi)^{1/2} \right) - \begin{array}{c} 0.352 \\ (-3.91) \end{array} \hat{\psi}_{t-k} \right],$$

and $\hat{\psi}_t = \hat{e}_t/\hat{\sigma}_t$. The log-likelihood for this model is 2198.2 versus 2191.8 for the GARCH(1, 2) model. Thus, the estimates of Nelson's EGARCH model confirm the previous evidence that conditional volatility increases more when return shocks are negative. The R^2 between $\hat{\sigma}_t^2$ and $\hat{e}_t^2$ is 0.118, which is a small improvement over the two-step method, but well above the GARCH(1, 2) model.[3]

---

[2] French, Schwert, and Stambaugh (1987) also estimated a GARCH-in-mean model, where the conditional mean return was a linear function of either the standard deviation or variance. We estimated such models for the 1835–1925 data, and the R^2 statistics were 0.076 and 0.077. Thus, the GARCH-in-mean results are essentially equivalent to the GARCH results reported in the text.

[3] The log-likelihood is for the *returns* $\hat{e}_t$, while the R^2 pertain to the explanation of the *squared returns* $\hat{e}_t^2$. Hence, although the two measures point in the same direction, they are not comparable.

2.6. Hamilton's two-state switching-regime model

Hamilton (1989) proposes a switching-regime Markov model for GNP growth rates as a model for recessions and expansions. Briefly, consider a variable y_t that follows an AR(m) process,

$$y_t - \mu(S_t) = \phi_1[y_{t-1} - \mu(S_{t-1})] + \phi_2[y_{t-2} - \mu(S_{t-2})] + \cdots$$
$$+ \phi_m[y_{t-m} - \mu(S_{t-m})] + \sigma(S_t)v_t, \quad (8)$$

where v_t is n.i.d.$(0, 1)$. The mean, $\mu(S_t)$, and the residual standard deviation, $\sigma(S_t)$, are functions of the regime in period t. The regimes are assumed to follow a two-state first-order Markov process,

$$
\begin{aligned}
&P(S_t = 1 | S_{t-1} = 1) = p, \\
&P(S_t = 0 | S_{t-1} = 1) = 1 - p, \\
&P(S_t = 1 | S_{t-1} = 0) = 1 - q, \\
&P(S_t = 0 | S_{t-1} = 0) = q,
\end{aligned}
\quad (9)
$$

and the parameters of (8) are modeled as

$$\mu(S_t) = \alpha_0 + \alpha_1 S_t, \qquad \sigma(S_t) = \omega_0 + \omega_1 S_t. \quad (10)$$

Finally, the errors v_t are assumed to be independent of all S_{t-j}. Given this structure, it is straightforward to use numerical procedures to maximize the likelihood as a function of the parameters $\{\phi_1, \ldots, \phi_m, p, q, \alpha_0, \alpha_1, \omega_0, \omega_1\}$.[4] Besides point estimates and asymptotic standard errors, Hamilton's algorithm estimates the probability that the variable is in regime 1 conditional on data available at data t. The estimates of Hamilton's model from July 1835 through December 1925 are

$$\hat{e}_t - \hat{\mu}(S_t) = \underset{(1.09)}{0.035} [\hat{e}_{t-1} - \hat{\mu}(S_{t-1})] - \underset{(-0.24)}{0.007} [\hat{e}_{t-2} - \hat{\mu}(S_{t-2})]$$

$$- \underset{(-0.24)}{0.007} [\hat{e}_{t-3} - \hat{\mu}(S_{t-3})]$$

$$\underset{(-0.04)}{-0.001} [\hat{e}_{t-4} - \hat{\mu}(S_{t-4})] + \hat{\sigma}(S_t)v_t, \quad (11)$$

$$\hat{\mu}(S_t) = \underset{(0.58)}{0.0006} - \underset{(-0.68)}{0.0025} S_t, \qquad \hat{\sigma}(S_t) = \underset{(26.02)}{0.0246} + \underset{(8.84)}{0.0253} S_t,$$

with t-statistics in parentheses. The estimates of the Markov probabilities are $\hat{q} = 0.9619$ (with a standard error of 0.0125) and $\hat{p} = 0.9034$ (with a standard error of 0.0328). Thus, these estimates imply that the high variance regime is less likely than the low variance regime, although both regimes are likely to persist once they occur.[5] Schwert (1989b) shows how to compute the conditional variance from this model. Briefly, if the variable was in regime 1 at $t-1$, the variance of the squared forecast error for period t is

$$E\{\sigma^2(S_t)|S_{t-1} = 1\} + \text{var}\{\mu(S_t)|S_{t-1} = 1\}$$

$$= \left[E\{\sigma(S_t)|S_{t-1} = 1\}\right]^2 + \text{var}\{\sigma(S_t)|S_{t-1} = 1\}$$

$$+ E\{\left[\mu(S_t) - E(\mu(S_t))\right]^2|S_{t-1} = 1\}$$

$$= \left[\omega_0 + \omega_1 p\right]^2 + \omega_1^2 p(1-p) + \alpha_1^2 p(1-p). \tag{12}$$

If the variable was in regime 0 at $t-1$, the variance of the squared forecast error for period t is

$$E\{\sigma^2(S_t)|S_{t-1} = 0\} + \text{var}\{\mu(S_t)|S_{t-1} = 0\}$$

$$= \left[E\{\sigma(S_t)|S_{t-1} = 0\}\right]^2 + \text{var}\{\sigma(S_t)|S_{t-1} = 0\}$$

$$+ E\{\left[\mu(S_t) - E(\mu(S_t))\right]^2|S_{t-1} = 0\}$$

$$= \left[\omega_0 + \omega_1(1-q)\right]^2 + \omega_1^2 q(1-q) + \alpha_1^2 q(1-q). \tag{13}$$

Multiplying (12) and (13) by the estimates of the conditional probabilities of being in each regime given data through $t-1$, $P(S_{t-1} = 1|\hat{e}_{t-1}, \ldots)$ and $P(S_{t-1} = 0|\hat{e}_{t-1}, \ldots)$ gives the estimate of the conditional variance of the forecast error at time t, $\hat{\sigma}_t^2$. The R^2 between $\hat{e}_t^2$ and $\hat{\sigma}_t^2$ is 0.057, which is the smallest among all the techniques we consider.

2.7. A nonparametric kernel estimator

Broadly there are two major philosophies in nonparametric estimation. The first is essentially a weighted average, that is

$$\hat{\sigma}_t^2 = \sum_{j=1}^{T} w_{jt}\hat{e}_j^2, \qquad \sum_{j=1}^{T} w_{jt} = 1, \tag{14}$$

[5]The expected durations of the regimes are $(1 - \hat{p})^{-1} = 10.4$ months and $(1 - \hat{q})^{-1} = 26.2$ months.

where T is the sample size. The weights w_{jt} are made to depend on F_j and F_t in such a way that, if F_j and F_t are 'far apart', w_{jt} is close to zero. What this does is make σ_t^2 equivalent to the sample variance of $\hat{e}_j$ using only those observations that are close to F_t. Since it is these observations that have variance σ_t^2, the method is analogous to the use of sample moments to estimate population moments. Many weighting schemes are possible. Letting z_t be the $r \times 1$ vector containing the elements in F_t, Nadaraya (1964) and Watson (1964) set

$$w_{jt} = K(z_t - z_j) \bigg/ \sum_{k=1}^{T} K(z_k - z_t), \tag{15}$$

where the kernel $K(\cdot)$ has the properties that it is nonzero, integrates to unity, and is symmetric. The kernel used in this paper was the Gaussian one,

$$K(z_t - z_j) = (2\pi)^{-1/2} |H|^{-1/2} \exp\left[-\tfrac{1}{2}(z_t - z_j)' H(z_t - z_j)\right]. \tag{16}$$

$H = \operatorname{diag}(h_1 \ldots h_r)$ contains the bandwidths, that were set to $\hat{\sigma}_k T^{-1/(4+r)}$, where $\hat{\sigma}_k$ is the sample standard deviation of z_{kt}, $k = 1, \ldots, r$. Silverman (1986) shows that the minimum mean square error choice of the bandwidth is proportional to $\hat{\sigma}_k T^{-1/(4+r)}$. No experimentation with the kernel or bandwidth was done, and we did not look at other weighting schemes. Partly this was due to our preference for the Fourier nonparametric estimator described later. One important modification that was employed was to leave out the tth observation when computing $\hat{\sigma}_t^2$,

$$\hat{\sigma}_t^2 = \sum_{\substack{j=1 \\ j \neq t}}^{T} w_{jt} \hat{e}_t^2. \tag{17}$$

Generally, it is important to adopt the 'leave-one-out' estimator to avoid the situation where 'outliers' in the data force w_{tt} to be unity, while all other w_{jt} are close to zero. In these circumstances, $\hat{e}_t^2$ becomes the estimator of σ_t^2 if all observations are used. While there is a sense in which this is the best estimate of σ_t^2, it tends to overstate the predictability of volatility by making a perfect prediction at time t. Based on F_t^1 the R^2 between $\hat{e}_t^2$ and $\hat{\sigma}_t^2$ is 0.126 if $\{\hat{e}_{t-1}\}$ is the conditioning variable. There is a major improvement over the GARCH and Hamilton models, and it is somewhat larger than for the EGARCH model. The R^2 between $\hat{e}_t^2$ and $\hat{\sigma}_t^2$ is lower for F_t^2 and F_t^4. This difference occurs because some of the observations on $\hat{e}_t^2$ for which $\hat{\sigma}_t^2$ was not computed were very large.
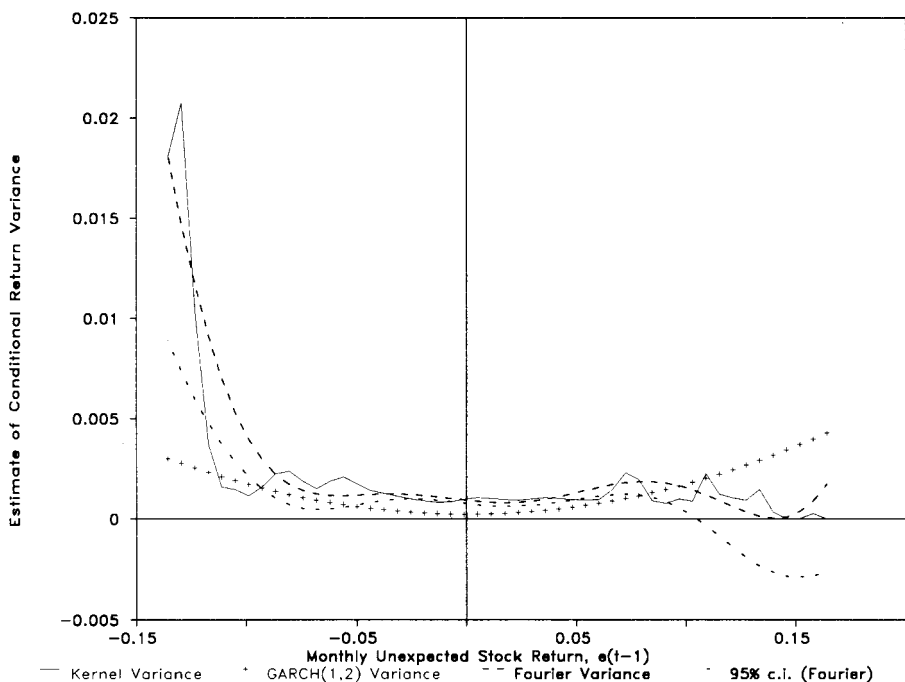
Fig. 2. Kernel, GARCH(1, 2), and Fourier estimates of the monthly stock return variance conditional on the lagged unexpected stock return, e_{t-1}, 1834–1925 (with the lower 95 percent confidence interval for the Fourier estimate).

It is worth noting that the estimate of conditional variance in (17) is a function of all the data, not just observations before time t. This is no different, however, than estimating the regression parameters in (5) using all the data. In both cases, the predictions of the conditional variance $\hat{\sigma}_t^2$ are a function of all the data. In section 3 we will discuss the results of a post-sample prediction experiment where the forecast models are estimated using data from 1835–1899, then forecasts are made for 1900–1925. We also use the estimates from 1835–1925 to forecast for 1926–1937.

It is not easy to summarize the mapping between σ_t^2 and $\{\hat{e}_{t-j}\}$ when the conditioning set is F_t^4. Some insight is available by computing the variance of $\hat{e}_t$ when the conditioning set is F_t^1. Fig. 2 displays the mapping of $\hat{\sigma}_t^2$ into a grid of fifty values of $\hat{e}_{t-1}$, located within the range of $\hat{e}_{t-1}$ found in the sample. An outstanding characteristic of fig. 2 is the difference in implied volatility for negative and positive values of $\hat{e}_{t-1}$, a stylized fact alluded to in the introduction. Fig. 2 is also similar to the equivalent mapping found by Pagan and Hong (1988) in their analysis of monthly stock returns from 1953

to 1984. Also in fig. 2 is the σ_t^2 implied by the GARCH(1, 2) model if one just took the lead term in the distributed lag connecting $\hat{\sigma}_t^2$ and $\hat{e}_{t-j}^2$. Comparing the GARCH and kernel functions it is clear that the GARCH model is likely to exhibit different volatility patterns when $|\hat{e}_{t-1}|$ is large. For small values of $|\hat{e}_{t-1}|$, the two predictions should be close. Unfortunately, this fact makes it hard to discriminate between the two methods, because large values of $|\hat{e}_{t-1}|$ are only a small fraction of the sample.

In addition to the conditional variance, one could compute the mean of $\hat{e}_t$ conditional on $\hat{e}_{t-1}$ to see if there are nonlinearities present. Both the kernel and Fourier estimators discussed later were used to estimate the conditional mean. There was very little dependence of the mean on $\hat{e}_{t-1}$. Thus, for this series it seems that the linear model used to estimate conditional means is an adequate representation of the data. This outcome is to be contrasted with the situation for the conditional variance.

2.8. A nonparametric flexible Fourier form estimator

An alternative nonparametric scheme involves a global approximation using a series expansion, followed by an evaluation of σ_t^2 using F_t. Many series expansions exist in the numerical approximation literature and could be adopted here, but the one used most extensively in economics has been the Flexible Fourier Form (FFF) [Gallant (1981)], where σ_t^2 is represented as the sum of a low-order polynomial and trigonometric terms constructed from the elements of F_t, $z_{tj} = \hat{e}_{t-j}$. Applying this idea to our context gives a model for volatility of the form

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^{L} \left\{ \left(\alpha_j z_{tj} + \beta_j z_{tj}^2 \right) + \sum_{k=1}^{2} \left[\gamma_{jk} \cos(k z_{tj}) + \delta_{jk} \sin(k z_{tj}) \right] \right\},$$

(18)

where $L = 1$, 2, or 4 depending on whether F_t^1, F_t^2, or F_t^4 was used. In theory, the number of trigonometric terms must tend to infinity, but in terms of significance it did not seem worthwhile going above order two.

A disadvantage of the FFF is the possibility that estimates of σ_t^2 can be negative, and indeed this happens for a few points in the sample. It has the advantage, however, that when few observations are available in a region of the sample space, the FFF will interpolate the function from other data points, whereas the kernel estimate is only based on the few observations. One must be ambivalent about this property. On the one hand, since 'difficult' points are often concentrated around the origin in multivariate problems [e.g., the 'empty space' phenomenon discussed in Silverman (1986)],

there is no 'extrapolation outside the sample', and the results should be reasonable. On the other hand, it is important to know that what we are seeing *is* just an interpolation. Joint viewing of output from both estimators is a prerequisite for an understanding of the behavior of nonparametric volatility measures.

Fig. 2 also shows the FFF estimates of $\hat{\sigma}_t^2$ as a function of $\hat{e}_{t-1}$ along with the lower part of the 95 percent confidence interval for the FFF estimates. The story of the mapping is much the same as for the kernel, except there is a larger estimate of volatility for large positive $\hat{e}_{t-1}$.[6] In this respect the FFF is closer to the GARCH estimate. Notice that across most of the range of $\hat{e}_{t-1}$, $\hat{\sigma}_t^2$ is constant, and it is only for large positive and negative values of $\hat{e}_{t-1}$ that any discrimination between the different ways of measuring σ_t^2 is possible. As there is only a small fraction of the sample featuring large $|\hat{e}_{t-1}|$, one must be sanguine about the possibility of differentiating between the techniques. Nevertheless, the F-statistic that the coefficients of the trigonometric terms in the FFF equal 0 is 6.47, compared with the 5% critical value of $F_{15,\infty} = 1.67$ (the actual degrees of freedom are 16 and 1061). Hence, the nonlinearities accounted for by the Fourier terms are important in explaining volatility. The R^2's between $\hat{e}_t^2$ and $\hat{\sigma}_t^2$ are 0.125 (F_t^1), 0.185 (F_t^2), and 0.205 (F_t^4). Because the EGARCH model has a conditioning set more like F_t^2 than F_t^1, it seems more appropriate to compare the fit of the different models with those R^2, and here the nonparametric estimator seems to represent a substantial improvement. Thus, it may be useful to consider extending the EGARCH model by the addition of Fourier terms in Z_{t-1} and Z_{t-2}.

2.9. Summary

Table 1 contains estimates of the regression

$$\hat{e}_t^2 = \alpha + \beta\hat{\sigma}_t^2 + v_t, \tag{19}$$

for 1835–1925, with heteroskedasticity-consistent standard errors in parentheses under the parameter estimates. If the forecasts are unbiased, $\alpha = 0$ and $\beta = 1$. For the two-step and the Fourier models, least squares estimation forces $\alpha = 0$ and $\beta = 1$. For the other methods, the estimates of α and β are within one standard error of their hypothesized values. The Box–Pierce (1970) statistics for twelve lags of the residual autocorrelations $Q(12)$, corrected for heteroskedasticity, are large for the Markov switching-regime model and the nonparametric kernel and Fourier (one lag) models, showing

---

[6] The close correspondence was another factor in deciding not to experiment with window width in kernel estimation.

Table 1

Comparison of within-sample predictive power for the conditional variance of stock returns, 1835–1925.[a]

$$\hat{e}_t^2 = \alpha + \beta\hat{\sigma}_t^2 + v_t$$

Model	α	β	R^2	$Q(12)$	R^2 for logs
1. Two-step	0.00000 (0.00036)	1.000 (0.3402)	0.089	0.69 (1.00)	0.022
2. GARCH(1,2)	0.00019 (0.00031)	0.8274 (0.2904)	0.067	17.3 (0.139)	0.033
3. EGARCH(1,2)	−0.00034 (0.00041)	1.318 (0.3911)	0.118	14.6 (0.265)	0.042
4. Markov switching-regime	−0.00009 (0.00021)	1.165 (0.2342)	0.057	23.3 (0.025)	0.034
5. Nonparametric kernel (1 lag)	0.00028 (0.00027)	0.7565 (0.2501)	0.126	29.2 (0.004)	0.012
6. Nonparametric Fourier (1 lag)	0.00000 (0.00033)	1.000 (0.3020)	0.125	29.0 (0.004)	0.015
7. Nonparametric Fourier (2 lags)	0.00000 (0.00018)	1.000 (0.1606)	0.185	13.2 (0.358)	0.028

[a]Standard errors using White's (1980) heteroskedasticity correction are in parentheses under the coefficient estimates. R^2 is the coefficient of determination. $Q(12)$ is the heteroskedasticity-corrected Box–Pierce (1970) statistic for twelve lags of the residual autocorrelations, with its p-value in parentheses below it. The corrected Box–Pierce statistic is calculated by comparing the sum of squared autocorrelation estimates, each divided by White's (1980) heteroskedasticity-consistent variance, and comparing this with a χ^2 distribution with twelve degrees of freedom. The R^2 for logs column shows the R^2 statistic from the regression of $\ln\hat{e}_t^2$ on $\ln\hat{\sigma}_t^2$.

serially correlated residuals v_t. The serial correlation shows there is additional persistence in volatility that is not captured by these models.

As a check on the criterion function we use to compare alternative models, we also ran the regression

$$\ln\hat{e}_t^2 = \alpha + \beta\ln\hat{\sigma}_t^2 + v_t \tag{20}$$

to compare the R^2 statistics from these regressions.[7] These statistics, labeled 'R^2 for logs' in table 1, are motivated by the idea of a proportional loss function, rather than the quadratic loss function implicit in (19). Mistakes in predicting small variances are given more weight in (20) than in (19). All the R^2 statistics for logs are smaller than the R^2's for the raw data. The nonparametric estimates are affected the most, showing that their apparent

[7]We are grateful to the referee and to John Campbell for suggesting this analysis.

advantage in predicting $\hat{e}_t^2$ is for very large values, which is consistent with the plots in fig. 2.

3. Post-sample prediction

The previous comparisons involve within-sample estimates of R^2 between $\hat{e}_t^2$ and $\hat{\sigma}_t^2$. Since some of the methods use a large number of parameters to model the data, there is the possibility that 'over-fitting' can occur. One way to evaluate this question is to estimate the model parameters with a subset of data and create out-of sample forecasts for the remainder.

3.1. Predictions for 1900–1925

Table 2 contains estimates of (19) and (20) for 1900–1925, where the model parameters were estimated using data from 1835–1899. For the two-step, GARCH(1,2), EGARCH(1,2), Hamilton, and kernel forecasts, the estimates of α and β are within one standard error of the hypothesized values ($\alpha = 0$, $\beta = 1$). For the Fourier forecasts, however, the estimates of α are more than two standard errors above 0, and the estimates of β are more than two standard errors below 1. The two-step model has the highest R^2 of 0.110, while the GARCH, EGARCH, and Hamilton forecasts have R^2's of about 0.07. The nonparametric forecasts have R^2's below 0.035. The Box–Pierce (1970) statistics for twelve lags of the residual autocorrelations $Q(12)$, corrected for heteroskedasticity, are large for all the forecast models, showing serially correlated forecast errors.

One might be tempted to conclude that the nonparametric methods of modeling conditional volatility suffer from over-fitting, since the R^2 statistics are so low for the out-of-sample forecasts. To check this possibility, we also calculated the R^2 statistics for 1900–1925 using the fitted values from the models estimated over the entire 1835–1925 sample period. If over-fitting is a serious problem, these R^2 statistics should be much higher than the out-of-sample prediction R^2's. Since the in-sample R^2's are only slightly higher than their out-of-sample counterparts, however, and the differences are similar for all the models in table 2, it seems that over-fitting or parameter instability is not a serious problem. Rather, the nonparametric forecasting methods work poorly in this sample because there are few large returns in the 1900–1925 period. Fig. 2 shows that the kernel and Fourier models obtain explanatory power from a few extreme returns and, as will be shown in section 4, many of these occur in the earlier part of the sample. It is well-known that nonparametric estimators are inefficient compared with parametric estimators of a *correctly specified* model. For this part of the data, the predictions of all the models would be for small values of σ_t^2, since the minimum value of $\hat{e}_t$ was

Table 2

Comparison of out-of-sample predictive power for the conditional variance of stock returns, 1900–1925.[a]

$$\hat{e}_t^2 = \alpha + \beta \hat{\sigma}_t^2 + v_t$$

Model	α	β	R^2	$Q(12)$	In-sample R^2	R^2 for logs
1. Two-step	−0.00020 (0.00045)	1.112 (0.4329)	0.110	22.2 (0.035)	0.137	0.023
2. GARCH(1,2)	0.00018 (0.00035)	0.7752 (0.3427)	0.075	30.4 (0.002)	0.077	0.027
3. EGARCH(1,2)	0.00007 (0.00033)	0.8771 (0.3310)	0.074	30.2 (0.003)	0.077	0.033
4. Markov switching-regime	−0.00003 (0.00032)	1.042 (0.3555)	0.070	30.4 (0.002)	0.086	0.026
5. Nonparametric kernel (1 lag)	0.00027 (0.00059)	0.7720 (0.5626)	0.013	21.8 (0.040)	0.019	0.019
6. Nonparametric Fourier (1 lag)	0.00090 (0.00024)	0.1416 (0.1964)	0.002	23.5 (0.024)	0.009	0.012
7. Nonparametric Fourier (2 lags)	0.00051 (0.00021)	0.4978 (0.2129)	0.032	22.2 (0.036)	0.046	0.018

[a]Standard errors using White's (1980) heteroskedasticity correction are in parentheses under the coefficient estimates. R^2 is the coefficient of determination. $Q(12)$ is the heteroskedasticity-corrected Box–Pierce (1970) statistic for twelve lags of the residual autocorrelations, with its p-value in parentheses below it. The corrected Box–Pierce statistic is calculated by comparing the sum of squared autocorrelation estimates, each divided by White's (1980) heteroskedasticity-consistent variance, and comparing this with a χ^2 distribution with twelve degrees of freedom. The parameters for these models are estimated using data from July 1835 through December 1899, then forecasts of conditional variances $\hat{\sigma}_t^2$ are made for the January 1900 through December 1925 period. The in-sample R^2 statistic in the next-to-last column measures the relation between fitted values from the model estimated over the entire 1835–1925 period with $\hat{e}_t^2$ over the 1900–1925 subsample. The R^2 for logs column shows the R^2 statistic from the regression of $\ln \hat{e}_t^2$ on $\ln \hat{\sigma}_t^2$ for the forecasts from 1900–1925.

−0.09. As seen from the slope coefficients, the nonparametric estimates have more variable $\hat{\sigma}_t^2$ than necessary, a sign of an inefficient estimator.

The R^2 statistics for logs from (20) are again smaller than for the raw data. The ranking of alternative methods is similar, however. Nelson's EGARCH model has the highest R^2 for $\ln \hat{e}_t^2$.

3.2. Predictions for 1926–1937

Table 3 contains estimates of (19) and (20) for 1926–1937, where the model parameters were estimated using data from 1835–1925. As mentioned earlier, the Great Depression from 1929–1939 was a period of unprecedented stock return volatility. The recursive variance estimates in fig. 1

Table 3

Comparison of out-of-sample predictive power for the conditional variance of stock returns, 1926–1937.[a]

$$\hat{e}_t^2 = \alpha + \beta\hat{\sigma}_t^2 + v_t$$

Model	α	β	R^2	$Q(12)$	R^2 for logs
1. Two-step	0.00373 (0.00170)	0.8146 (0.4409)	0.055	7.6 (0.813)	0.066
2. GARCH(1, 2)	0.00288 (0.00148)	0.9209 (0.3918)	0.078	8.4 (0.754)	0.091
3. EGARCH(1, 2)	0.00136 (0.00112)	1.895 (0.5478)	0.080	6.4 (0.893)	0.111
4. Markov switching-regime	− 0.00406 (0.00184)	5.644 (1.490)	0.045	7.0 (0.861)	0.026
5. Nonparametric kernel (1 lag)	0.00670 (0.00160)	− 0.0115 (0.2522)	0.000	24.0 (0.020)	0.025
6. Nonparametric Fourier (1 lag)	0.00631 (0.00120)	0.0074 (0.0077)	0.019	12.9 (0.374)	0.033
7. Nonparametric Fourier (2 lags)	0.00642 (0.00120)	0.0071 (0.0078)	0.016	15.8 (0.202)	0.047

[a]Standard errors using White's (1980) heteroskedasticity correction are in parentheses under the coefficient estimates. R^2 is the coefficient of determination. $Q(12)$ is the heteroskedasticity-corrected Box–Pierce (1970) statistic for twelve lags of the residual autocorrelations, with its p-value in parentheses below it. The corrected Box–Pierce statistic is calculated by comparing the sum of squared autocorrelation estimates, each divided by White's (1980) heteroskedasticity-consistent variance, and comparing this with a χ^2 distribution with twelve degrees of freedom. The R^2 for logs column shows the R^2 statistic from the regression of $\ln \hat{e}_t^2$ on $\ln \hat{\sigma}_t^2$.

strongly show that the unconditional variance is not constant between 1835–1925 and 1926–1937. Thus, we should expect that forecasting conditional variances in this sample period will be difficult. Nevertheless, the large changes in stock prices that occurred in this period provide an interesting out-of-sample experiment. If the 1900–1925 period was too quiet, the 1926–1937 period may be too volatile.

Indeed, the estimates (19) in table 3 show substantial bias in the forecasts. Most of the intercept estimates $\hat{\alpha}$ are more than two standard errors above 0, and the slope coefficient estimates $\hat{\beta}$ are more than two standard errors from 1. The two-step, GARCH(1, 2), and EGARCH(1, 2) models seem to work best, probably because they capture the persistence in volatility that was important in this period. Hamilton's model has an upper bound on the conditional variance that is too small, so the slope coefficient is over 5.5. While the Markov model correctly identifies periods of high variance, its estimate of volatility is too low. The nonparametric models also do poorly in

this period. The large positive and negative monthly returns that occurred in the 1926–1937 period have no precedent in the 1835–1925 sample, so the kernel has no basis for making predictions of conditional variance. The Fourier model also does poorly because it has to extrapolate outside of the range of the data.[8]

The R^2 statistics from the log regressions (20) are generally larger for this sample period. The large values of $\hat{e}_t^2$ are difficult for all the models to predict. Nevertheless, the relative ranking of the methods is similar: the EGARCH model does best, followed by the other parametric models, and the nonparametric methods do worst in this out-of-sample prediction experiment.

4. Analysis of important episodes of stock volatility

Another way to contrast the behavior of the alternative variance estimators is to analyze their behavior during important subperiods in the sample. Fig. 2 shows that the main difference between the GARCH(1, 2) model and the kernel or Fourier estimator occurs for large negative returns. These data also explain the difference between Nelson's EGARCH model and the GARCH model. Thus, it is worthwhile to plot some of the variance estimates around major drops in stock prices from 1835–1925. Schwert (1989b) notes that many of the stock market 'crashes' during the 19th century occurred at about the same time as banking panics. Therefore, we will use the dates of the bank panics and other major events to evaluate the different predictions of stock return volatility.

4.1. The banking crisis of 1837

There was a major banking crisis in May 1837. This is one of the cases where many banks refused to redeem demand deposits for currency. Stock prices fell in early 1837 as investors seeking liquidity sold stocks [see Sobel (1988, ch. 2) for an interesting history of this episode]. Fig. 3a plots the unexpected stock return $\hat{e}_t$ (E)[9] along with the one-lag Fourier (F), kernel (K), Hamilton (H), EGARCH(1, 2) (EG), and GARCH(1, 2) (G) estimates of the conditional standard deviation for 1837. Stock prices fell during early 1837, with monthly returns of -2, -5, -8, and -8 percent in February through May. On the other hand, the rise in stock prices in July 1837 of over

---

[8]Some of the forecasts of conditional standard deviation are over 200 percent per month from the Fourier models in this period.

[9]The unexpected stock returns $\hat{e}_t$ (E) in figs. 3a–3d are multiplied by 0.1 so they do not dominate the plots of the standard deviations. Thus, when E is -0.01 in one of these plots, the unexpected stock return was -10 percent that month.
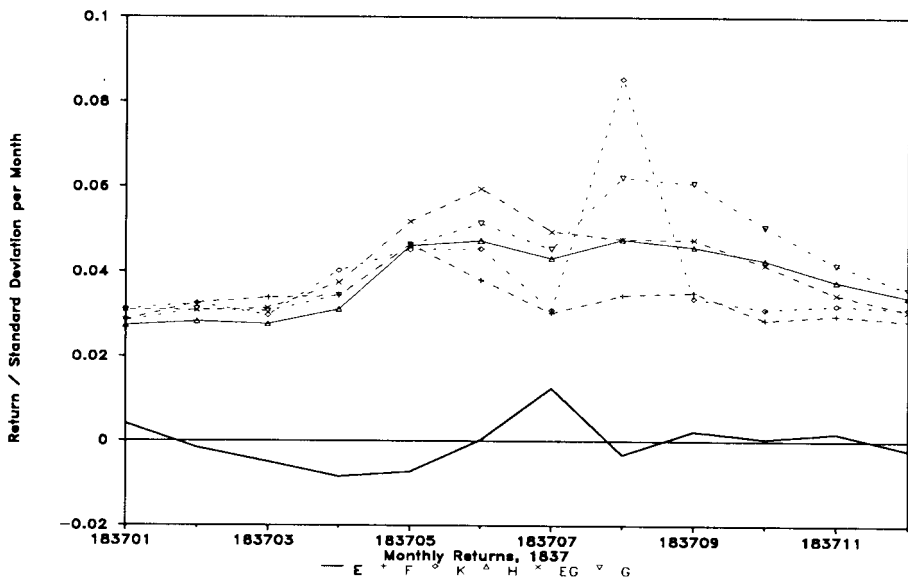
Fig. 3a. Unexpected stock returns, e_t, and estimates of conditional standard deviations from Fourier (F), kernel (K), Hamilton (H), EGARCH (EG), and GARCH (G) models, 1837.

12 percent is the third largest monthly return in the sample. This is characteristic of conditional heteroskedasticity – large returns follow large returns, with random signs. Among the volatility estimates, the Fourier estimate moves the least. The kernel estimate and the GARCH estimate increase in August 1837, following the erratic pattern of returns earlier in the year. The kernel estimate drops back to its previous level in September 1837, while the GARCH estimates gradually decay.

4.2. The banking panic of 1857

There was a major banking crisis in the Fall of 1857 [see Sobel (1988, ch. 3)]. Several major firms went bankrupt and there was a similar financial crisis in Europe. Fig. 3b plots the unexpected stock return $\hat{e}_t$ (E) along with the various conditional standard deviation estimates for the last half of 1857 and the first half of 1858. Stock prices fell 6, 14, and 13 percent in August, September, and October 1857. Then, in November 1857, prices rose by more than 16 percent. The returns for September–November 1857 are three of the four largest in absolute value for the 1835–1925 period. This episode is the best experiment to differentiate among the alternative variance estimators. Both the kernel and the Fourier estimates rise dramatically in October 1857,
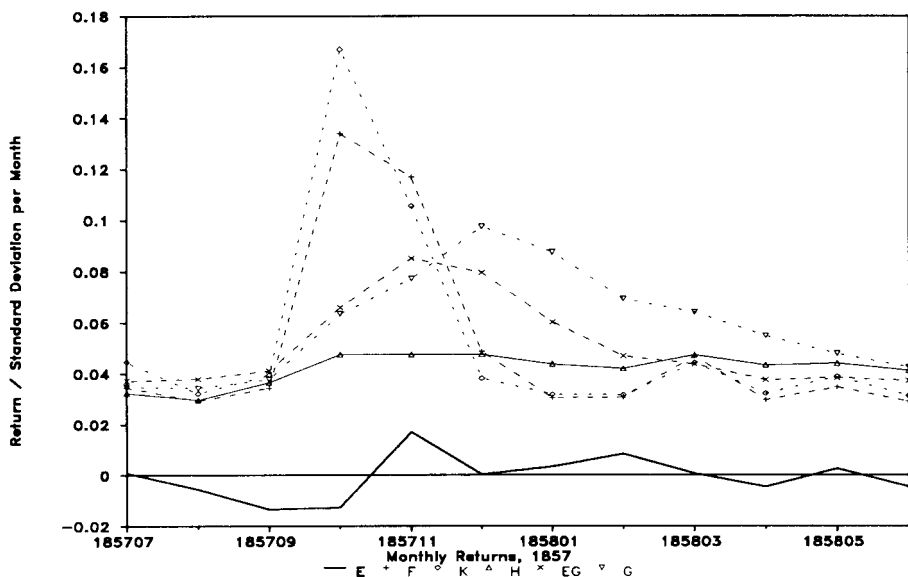
Fig. 3b. Unexpected stock returns, e_t, and estimates of conditional standard deviations from Fourier (F), kernel (K), Hamilton (H), EGARCH (EG), and GARCH (G) models, 1857–58.

and they decline sharply in December 1857. In contrast, the GARCH and EGARCH estimates rise gradually, peaking in December 1857 and gradually decaying after that. Hamilton's estimate rises and falls much less. Thus, the nonparametric estimates adapt more quickly to the fast increase in volatility and to its decrease when the panic subsided.

4.3. The start of the Civil War, 1860

It is not surprising that the beginning of the Civil War increased the volatility of stock returns. Fig. 3c plots the unexpected stock return $\hat{e}_t$ (E) along with the various conditional standard deviation estimates for the last half of 1860 and the first half of 1861. Stock prices fell 4, 10, and 5 percent in the last three months of 1860, rising about 10 percent in January 1861, only to fall 9 and 6 percent in April and May 1861. Again, the Fourier estimate of the conditional standard deviation rises the most in December 1860 and May 1861, returning to more normal levels in the next month. The other methods show a smaller increase in volatility in December 1860 and slight decay from that point.
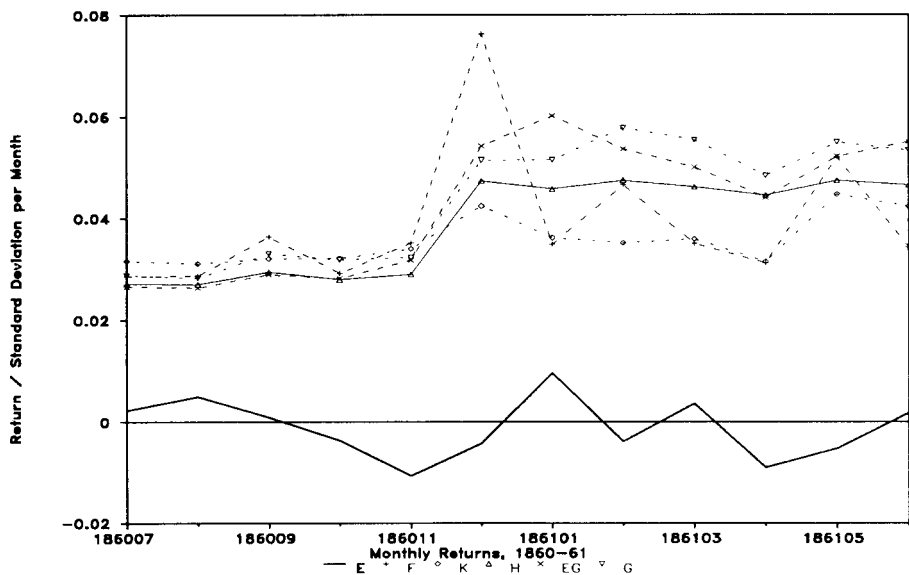
Fig. 3c. Unexpected stock returns, e_t, and estimates of conditional standard deviations from Fourier (F), kernel (K), Hamilton (H), EGARCH (EG), and GARCH (G) models, 1860-61.
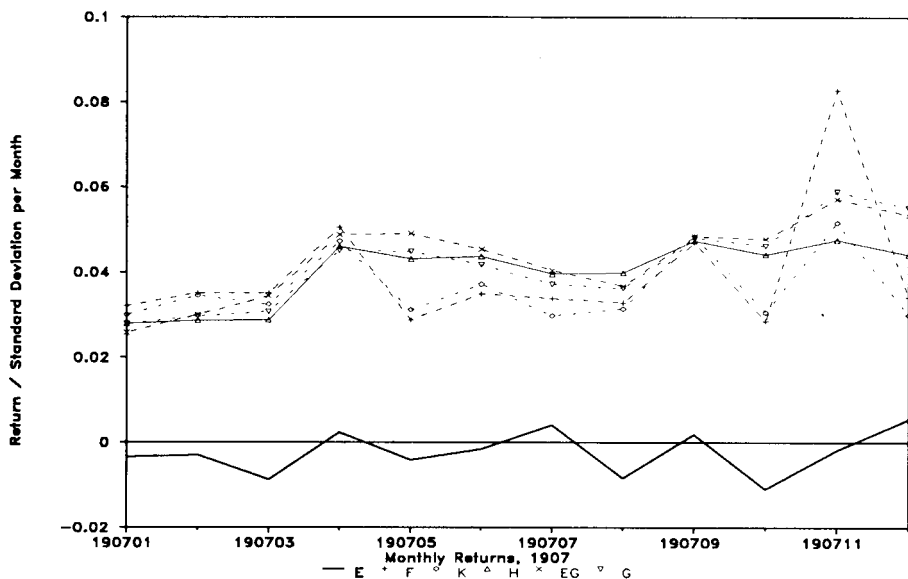

Fig. 3d. Unexpected stock returns, e_t, and estimates of conditional standard deviations from Fourier (F), kernel (K), Hamilton (H), EGARCH (EG), and GARCH (G) models, 1907.

4.4. The banking crisis of 1907

The banking crisis of 1907 is often credited with leading to the creation of the Federal Reserve System in 1914. Fig. 3d shows that stock prices fell by almost 9 percent in March, August, and October 1907. All the estimates of conditional standard deviations rose in April 1907, with the kernel and Fourier estimates dropping in May. The Fourier estimate jumps from October to November, then falls back to its previous level in December. The GARCH, EGARCH, and Hamilton estimates remain high throughout the second half of 1907.

4.5. Summary

The plots in figs. 3a–3d show that the nonparametric estimates of conditional volatility (kernel and Fourier) are different from the parametric estimates (GARCH, EGARCH, and Hamilton) in periods when stock prices fall. In particular, volatility rises fast after large negative unexpected returns. The parametric estimates all show slow adjustment to large volatility shocks, but the effects of these shocks persist after the crises subside. These plots reinforce the impression given by the goodness-of-fit regressions in tables 1, 2, and 3. The parametric and nonparametric methods of modeling conditional volatility capture different aspects of the data. The parametric methods use the persistent, smoother aspects of conditional volatility, while the nonparametric methods use the highly nonlinear response to large return shocks. Neither method subsumes the other.

5. Nesting parametric and nonparametric models

The previous evidence shows that parametric and nonparametric models for stock volatility capture different aspects of the data. One way to nest these models is to add Fourier terms to the parametric models. For example, we added the sine and cosine of lagged stock returns y_{t-1} to the GARCH$(1,2)$ model in (6) for 1835–1925 and the log-likelihood increased by 4.8. The likelihood ratio test statistic of 9.6 has a p-value of 0.008. When Fourier terms were added to the EGARCH$(1,2)$ model in (7), however, the log-likelihood increased by only 0.77, yielding a small test statistic with a p-value of 0.46. Thus, it seems that for the 1835–1925 period the EGARCH model captures the asymmetry in the relation between stock return and volatility.

We also added Fourier terms to GARCH and EGARCH models estimated over the entire 1835–1987 period. The χ^2 statistic for the Fourier terms is 16.8 in the GARCH model (with a p-value of 0.0002), which again shows the importance of asymmetries missed by the GARCH model. The χ^2 statistic is

11.8 for the EGARCH model, which has a p-value of 0.003. Thus, it seems there are important asymmetries missed by the EGARCH model over the longer sample period. It remains an open question whether the nonstationarity of the variance over this period affects the tests for the Fourier terms.

6. Conclusions and suggestions for future work

Our aim was to compare various measures of stock volatility. Taking the 1835–1925 period as the sample, it emerged that the nonparametric procedures tended to give a better explanation of the squared returns than any of the parametric models. Both Hamilton's and the GARCH model produced weak explanations of the data. Nelson's EGARCH model came closest to the explanatory power of the nonparametric models, because it reflects the asymmetric relation between volatility and past returns.

In out-of-sample prediction experiments, the nonparametric models fared worse than the parametric models. Nonparametric estimators of conditional moments are inefficient relative to parametric ones, and this is likely to show up in too much variability in the estimates of σ_t^2. An improved ability to capture the movements in σ_t^2 when returns decline therefore has to be set against this tendency, and it appears that even with a sample of the size used here nonparametric methods find it hard to overcome their inherent inefficiency. Previous uses of nonparametrics in this area, for example Pagan and Hong (1988), used the estimate of σ_t^2 in a regression to semiparametrically estimate risk coefficients, and hence the 'averaging' of $\hat{\sigma}_t^2$ makes the semiparametric and parametric estimators equivalent.

Our results imply that standard parametric models are not sufficiently extensive. Augmenting the GARCH and EGARCH models with terms suggested by nonparametric methods yields significant increases in explanatory power. This fact points to the need to merge the two traditions to capture a richer set of specifications than are currently employed. Nevertheless, our results emphasize that any extensions are best done in a parametric framework.

A secondary concern of the paper, which grew out of the data analysis, is that data taken over long periods cannot be assumed to be covariance stationary. Much work in this area ignores this question entirely, although the models proposed to fit the data imply covariance stationarity. A simple recursive variance test showed that the data could not be thought of as homogeneous before and after the Great Depression. This was illustrated by the fact that all the models performed poorly in predicting conditional variances in the 1926–1937 sample. If covariance nonstationarity is found to be a feature of many financial series, it forces us to examine what are likely to be good models of such data.

References

Black, Fischer, 1976, Studies of stock price volatility changes, Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association, 177–181.

Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–328.

Bollerslev, Tim, 1988, Integrated ARCH and cointegration in variance, Unpublished manuscript (Northwestern University, Evanston, IL).

Christie, Andrew A., 1982, The stochastic behavior of common stock variances: Value, leverage and interest rate effects, Journal of Financial Economics 10, 407–432.

Cowles, Alfred III and Associates, 1939, Common stock indexes, 2nd ed., Cowles Commission monograph no. 3 (Principia Press, Bloomington, IN).

Davidian, Marie and Raymond J. Carroll, 1987, Variance function estimation, Journal of the American Statistical Association 82, 1079–1091.

Engle, Robert F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50, 987–1007.

Engle, Robert F. and Tim Bollerslev, 1986, Modeling the persistence of conditional variances, Econometric Reviews 5, 1–50.

French, Kenneth R., G. William Schwert, and Robert F. Stambaugh, 1987, Expected stock returns and volatility, Journal of Financial Economics 19, 3–29.

Gallant, A. Ronald, 1981, On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form, Journal of Econometrics 15, 211–244.

Hamilton, James D., 1988, Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates, Journal of Economic Dynamics and Control 12, 385–423.

Hamilton, James D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57, 357–384.

Macaulay, Frederick R., 1938, The movements of interest rates, bond yields and stock prices in the United States since 1856 (National Bureau of Economic Research, New York, NY).

Mandelbrot, Benoit, 1963, The variation of certain speculative prices, Journal of Business 36, 394–419.

Nadaraya, E.A., 1964, On estimating regression, Theory of Probability and Its Applications 9, 141–142.

Nelson, Daniel B., 1988, Conditional heteroskedasticity in asset returns: A new approach, Unpublished manuscript (University of Chicago, Chicago, IL).

Pagan, Adrian R. and Aman Ullah, 1988, The econometric analysis of models with risk terms, Journal of Applied Econometrics 3, 87–105.

Pagan, Adrian R. and Yi-Soek Hong, 1989, Nonparametric estimation and the risk premium, in: W. Barnett, J. Powell, and G. Tauchen, eds., Semiparametric and nonparametric methods in econometrics and statistics (Cambridge University Press, Cambridge).

Pagan, Adrian R. and H. Sabau, 1987, Consistency tests for heteroskedastic and risk models, Unpublished manuscript (University of Rochester, Rochester, NY).

Sabau, H., 1988, Some theoretical aspects of econometric inference with heteroskedastic models, Unpublished doctoral dissertation (Australian National University, Canberra).

Schwert, G. William, 1989a, Why does stock market volatility change over time?, Journal of Finance 44, 1115–1153.

Schwert, G. William, 1989b, Business cycles, financial crises and stock volatility, Carnegie–Rochester Conference Series on Public Policy 31, 83–126.

Schwert, G. William, 1990, Stock volatility and the crash of '87, Review of Financial Studies 3, forthcoming.

Silverman, B.W., 1986, Density estimation for statistics and data analysis (Chapman and Hall, London).

Smith, Walter B. and Arthur H. Cole, 1935, Fluctuations in American business, 1790–1860 (Harvard University Press, Cambridge, MA).

Sobel, Robert, 1988, Panic on Wall Street, Rev. ed. (E.P. Dutton, New York, NY).

Watson, G.S., 1964, Smooth regression analysis, Sankhya A 26, 359–372.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817–838.