



ELSEVIER

Journal of Empirical Finance 3 (1996) 15–102

---

---

Journal of  
EMPIRICAL  
FINANCE

---

---

## The econometrics of financial markets

Adrian Pagan

*Economics Program, Research School Social Science, Australian National University, Canberra,  
A.C.T. 0200, Australia*

---

### Abstract

The paper provides a survey of the work that has been done in financial econometrics in the past decade. It proceeds by first establishing a set of stylized facts that are characteristics of financial series and then by detailing the range of techniques that have been developed to model series which possess these characteristics. Both univariate and multivariate models are considered.

*JEL classification:* G11; G12; G13; G14

*Keywords:* Portfolio choice; Asset pricing; Contingent pricing; Futures pricing; Information; Market efficiency

---

### 1. Introduction

Financial econometrics has emerged as one of the most vibrant areas of the discipline in the past decade, featuring an explosion of theoretical and applied work. Perhaps more so than in any other part of econometrics, the two have gone together, with the development of sophisticated techniques being driven by the special features seen in financial data. Understanding these characteristics, and the problems that they pose for modeling, must therefore be at the top of any agenda purporting to describe “financial econometrics”. It is worth thinking back to the econometrics of the fifties and sixties to realize the import of this. Then it would have been common to assume that any series being investigated could be regarded as stationary, independently, identically and (possibly) normally distributed, with

moments that existed.<sup>1</sup> To be sure, adjustments were sometimes made to allow for “complications”, such as deterministically trending variables, but these were all treated as departures from a base model. Such a perspective has been slowly jettisoned by those dealing with financial series. Non-stationarity, a lack of independence, and non-normality have become the characteristics of the standard model, with the earlier set of assumptions now being regarded as the *curiosa*. How and why this development came about is the concern of the first section of this paper. The “how” is a mechanical task; explaining “why” is the more interesting one, but we will have to content ourselves with the observation that the models of interest to financial economists depended very heavily upon the validity of these assumptions, so it is scarcely surprising that a major effort was mounted to assess their validity. As examples of this contention witness the fact that mean-variance portfolio models and optimal hedging formulae gave equal attention to the first and second moments of the data; option pricing formulae emphasized the need for a correct description of either the conditional or unconditional density of returns; while wealth choices based on maximizing expected utility  $E(U(y_t | F_t))$ , where  $F_t$  was some information set, focused attention upon the nature of higher order conditional moments.<sup>2</sup>

Having established a case for emphasizing features such as dependence and non-normality, Sections 2 and 3 seek to describe how the characteristics of the series have been captured by parametric models. Inevitably, this section is heavily oriented towards the construction, estimation and testing of models that explain the volatility of returns. Section 4 returns to the task of data description, now enquiring into what extra information is generated when moving from a univariate to a multivariate perspective. Our answer will be to concentrate upon locating the number and nature of the factors that are common to a set of returns. In Section 5, attention turns to economic models that have been proposed to account for the regularities seen in financial data. Because of the forward looking nature of financial markets, most of these models place heavy reliance on inter-temporal optimization, and this has led to an extensive literature which seeks to discover how accurate the description of agents’ behavior provided by these models is. Finally, Section 6 provides a short conclusion.

Financial data appears in many forms, each with its own idiosyncrasies. We will generally concentrate upon three representative series – stock prices, interest

<sup>1</sup> Hence assumptions like  $T^{-1}X'X$  is a constant as  $T \rightarrow \infty$  and the prevalence of  $F$ -statistics for testing hypotheses.

<sup>2</sup> In this review we will concentrate upon time series of financial data, ignoring the fact that data is sometimes available upon a cross section at a point in time or even on a panel. There are special econometric problems arising in the latter circumstances, but generally they would appear in any cross section or panel, and are not specific to financial series. It is in the time dimension that financial series are particularly distinct, and it is for this reason that a very special set of techniques has arisen to deal with these features.

rates, and exchange rates, ignoring other series such as futures and options prices. Nevertheless, much of what is said about the three series in question applies to others. The series are the log of the monthly CRSP value weighted price of shares on the NYSE over the period 1925/12 to 1989/12, the continuously compounded returns to one, three, six, and nine month zero coupon bonds over the period 1946(12) to 1987(12) (taken from McCulloch, 1989); and the log of the weekly exchange rate between the \$US and the Swiss Franc over the period July 1973 to August 1985 (used by Engle and Bollerslev, 1986). The choice of the series was largely governed by their accessibility. Sometimes we will also refer to two much longer series of daily observations, stock returns computed from the S&P Composite Price Index, 1928–1987, as adjusted in Gallant et al. (1992) and U.S. daily stock returns from 1885–1987, constructed by Schwert (1990).

## 2. Properties of univariate financial series

It is useful to begin by considering what type of models one might adopt for a series  $g(y_t)$  where  $g(\cdot)$  is some known transformation of a random variable  $y_t$ , e.g.,  $g(y_t) = y_t^2$  or  $\log(y_t^2)$ . Throughout, we will assume that  $g(y_t)$  has a zero expected value. Three simple models from time series analysis would be,<sup>3</sup> where  $L$  is the lag operator,

$$\text{ARMA}(1,1) : g(y_t) = \beta_1 g(y_{t-1}) + e_t + \alpha_1 e_{t-1} \quad (1)$$

$$\text{FI} : (1 - L)^d g(y_t) = e_t \quad (2)$$

$$\begin{aligned} \text{UC} : g(y_t) &= g_p(y_t) + g_T(y_t) \\ (1 - \beta_1 L) g_p(y_t) &= \eta_t \\ g_T(y_t) &= \xi_t, \end{aligned} \quad (3)$$

where  $e_t$ ,  $\eta_t$ , and  $\xi_t$ , all have conditional expectation zero and will be taken to be identically and independently distributed over time.<sup>4</sup>

The ARMA(1,1) model in (1) has been a work horse of time series econometrics since its popularization by Box and Jenkins in the late sixties. By setting  $\beta_1 = 1$  one can produce a process that is integrated of order one,  $I(1)$ , so rendering  $g(y_t)$  no longer covariance stationary. The second model in (2) represents the class of *fractionally integrated* processes; provided  $d < 0.5$  it is a covariance

---

<sup>3</sup> Of course these can all be generalized by allowing for higher order autoregressive and moving average processes, but there is little advantage to doing that in this survey.

<sup>4</sup> We will write this as  $E_{t-1}(v_t) = 0$ , where the subscript denotes that the expectation is conditional upon some set of past-dated random variables. Generally these will be the past history of  $v_t$ .

stationary process, but, if  $d = 1$ , it coincides with an  $I(1)$  series. Finally, (3) is an *unobserved components model*, composed of a permanent ( $g_p(y_t)$ ) and a transitory ( $g_T(y_t)$ ) part, each of which is generated as a stochastic process depending upon innovations  $\eta_t$  and  $\xi_t$ . Generally it is assumed that  $\eta_t$  and  $\xi_t$  are independent of one another, and this yields a representation for  $g(y_t)$  that is ARMA(1,1) but with a *negative* coefficient  $\alpha$ .<sup>5</sup> In some instances  $\eta_t$  and  $\xi_t$  are assumed proportional and this provides the type of decomposition set out in Beveridge and Nelson (1981).

### 2.1. Are financial series stationary?

Non-stationarity of a series could occur in many ways but, arising from the theory of efficient markets, it was natural for researchers to investigate whether there was a unit root associated with the log prices of financial assets, i.e., whether such a series defined as  $y_t$  was  $I(1)$ , or not. A huge literature has been spawned on the question of unit roots, and the resulting tests have been extensively applied to financial data. These tests can be classified according to their responses to four issues.

1. Whether the null hypothesis is that the series is  $I(1)$  or  $I(0)$ .
2. The model used to construct a test of this hypothesis.
3. Within a given model, which of the characteristics of an  $I(1)$  series is used to set up a test.
4. Whether the alternative distinguished in the testing procedure is composite or simple.

Table 1 summarizes the literature according to this four way classification, culminating in five types of tests that we will focus our attention upon. In the table

Table 1  
Classification of unit root tests

Hypoth	$H_0: I(1)$				$H_0: I(0)$			
Model	ARMA				FI	UC		
Based on	a.c.f.	Var		Degree( $d$ )		$\sigma_\xi^2$		
Nature $H_1$	comp	simp	comp	simp	comp	simp	comp	simp
Example	ADF	pt opt	var rat	?	FI tests	?	KPSS	?
Index	(i)	(ii)	(iii)		(iv)		(v)	

<sup>5</sup> The error term  $e_t$  is now a linear combination of  $\sum_{j=0}^{\infty} \phi_j \eta_{t-j}$  and  $\sum_{j=0}^{\infty} \psi_j \xi_{t-j}$  and so is no longer independently distributed even though  $\eta_t$  and  $\xi_t$  are (see McDonald and Darroch, 1983).

there are also question marks which represent what seem to be tests that are not in the literature to our knowledge.<sup>6</sup>

(i) Selecting (1) with  $g(y_t) = y_t$ , the process becomes

$$y_t = \beta_1 y_{t-1} + u_t, \quad (4)$$

and we might test  $H_0: \beta_1 = 1$  vs.  $H_1: \beta_1 < 1$  using the  $t$ -ratio from the regression of  $\Delta y_t$  on  $y_{t-1}$ . This leads to tests for a unit root of the Phillips–Perron, Dickey–Fuller (DF) type which feature in many econometric programs.

(ii) Instead of using the composite alternative one might put  $H_1: \beta_1 = \beta_1^*$ . In this situation the best test can be derived from the Neyman–Pearson Lemma as a likelihood ratio. This produces the class of point optimal tests examined in Dufour and King (1991) and Elliott et al. (1992). The latter also consider the case of near integration where  $\beta_1^* = 1 - c/T$ , so that, as  $T \rightarrow \infty$ , the alternative collapses towards the null.

(iii) The two tests just distinguished concentrate upon the autocorrelation coefficients of the series  $y_t$ ,  $\rho_j(y_t) = \text{cov}(y_t, y_{t-j})/\text{var}(y_t)$ , as these are all unity if the process is  $I(1)$ , so that it is natural to test if  $\beta_1$  is unity using  $\hat{\rho}_j(y_t) = (\sum_{t=j+1}^T y_{t-j}^2)^{-1} \sum_{t=j+1}^T y_t y_{t-j}$ . Instead, one might focus upon other implications of a unit root, in particular the nature of the “variance” of the series. If (1) is a correct representation of the data,  $\alpha_1 = 0$  and  $e_t$  is assumed i.i.d.  $(0, \sigma_e^2)$ , then

$$y_t = y_{t-k} + e_t + e_{t-1} + \dots + e_{t-k+1}, \quad (5)$$

and the *kth long difference* of  $y_t$ ,  $\Delta_k y_t = y_t - y_{t-k}$ , will have variance  $k \sigma_e^2$ .<sup>7</sup> Taking the ratio  $\lambda_1 = \text{var}(\Delta_k y_t)/\text{var}(\Delta_1 y_t)$  it would be  $k$ , and therefore a plot of  $\lambda_1$  against  $k$  should be an increasing straight line. Alternatively,  $\lambda_2 = k^{-1} \lambda_1$  should tend to unity as  $k$  increases. If, however,  $y_t$  does not have a unit root both  $\text{var}(\Delta_k y_t)$  and  $\text{var}(\Delta_1 y_t)$  will be constants, so that the ratio  $\lambda_1$  does not change after  $k$  becomes large enough, while  $\lambda_2$  will tend to zero. Cochrane (1988) formulated this idea and gave asymptotic standard errors for  $\hat{\lambda}_2$  when estimated from the data. Lo and MacKinlay (1989) give the asymptotic distribution of  $T^{1/2}(\hat{\lambda}_2 - 1)$  as  $N(0, 2(2k-1)(k-1)/3k)$  if  $e_t$  is  $N(0, \sigma^2)$ , but found in simulations that the approximation was poor when  $\delta = k/T$  is large. Richardson and Stock (1989) determine the asymptotic distribution when  $k \rightarrow \infty$ ,  $T \rightarrow \infty$ , in such a way that  $\delta$  is a constant. This turns out to involve functionals of Brownian motion, and the distribution has to be found by simulation. When  $\delta = 1/3$  they tabulate critical values of the statistic for various  $T$ . What is appealing about this test, from

<sup>6</sup> There is also a growing literature on Bayesian tests for unit roots that we will not summarize here, but which have sometimes used financial data to illustrate the techniques, e.g., Koop (1994) and Schotman and Van Dijk (1991).

<sup>7</sup> With some modification, the results that follow also hold if  $e_t$  is correlated. Cochrane (1988) deals with the general case.

the perspective of financial data, is that, if  $y_t$  is the log of the price, the  $k$ th long difference can be interpreted as the returns to an asset held for  $k$  periods (assuming only capital gains and no dividends). A disadvantage of the test is its presumption that  $e_t$  is normally distributed and that the  $\text{var}(\Delta_1 y_t)$  exists. Later in this survey both assumptions come under suspicion.

(iv) Instead of the alternative model used to construct tests being ARMA it could be the *fractionally integrated* process in (2). It is possible to estimate  $d$  – see Diebold and Rudebusch (1989) and Sowell (1992) – but probably not very precisely, although it has been done for various series, e.g., Baillie and Bollerslev (1994) with the forward premium. Another strategy is to derive tests such as an LM test or a one-sided locally best invariant test (LBI) that  $d$  takes a specified value, in particular whether  $d = 1$ . Writing the FI model as  $(1 - L)^{d-1} \Delta y_t = e_t$ , we can test this by checking if  $(d - 1)$  is zero. Wu (1992) derives the LBI test for this hypothesis as  $-2\sum_{k=1}^{T-1} (1/k)\hat{\rho}_k(\Delta y_t)$ . It is probably not a good idea to base a test on a  $\hat{\rho}_k$  that has  $k$  close to  $T$ , suggesting that we instead use  $\hat{\tau}_K = -\sum_{k=1}^K (1/k)\hat{\rho}_k(\Delta y_t)$ , with  $K$  set to a large value. Under the assumption that  $e_t$  is white noise,  $T^{1/2}s\hat{\tau}_K$  will be asymptotically  $N(0,1)$ , where  $s^2 = \sum_{k=1}^K 1/k^2$ .<sup>8</sup> One might wish to use a one-tailed test given that the test statistic inherits its LBI properties from its one-sided nature. Another way of interpreting the test is to observe that it looks at the correlation of  $\Delta y_t$  with  $\sum_{k=1}^K (1/k)\Delta y_{t-k}$ , and so differs from the DF test in that it seeks to add  $\sum_{k=1}^K (1/k)\Delta y_{t-k}$  to a regression with  $\Delta y_t$  as dependent variable rather than  $y_{t-1}$ .

(v) Changing the model to that of an unobserved components process (3), with  $g(y_t) = y_t$ , we might test if  $\beta_1$  is unity or, if the null hypothesis is taken to be that  $y_t$  is an  $I(0)$  process, by testing if the variance of  $\eta_t$  is zero. If the latter is zero then  $y_t = y_{Tt}$  and there is no unit root in the series,  $y_t$ . Tests for this case have been developed by Tanaka (1990) and Kwiatowski et al. (1992). They are based upon the summation of the squares of the partial sums of the mean corrected prices ( $S_t = \sum_{j=1}^t y_j$ ), and the statistic has the form  $\text{KPSS} = T^{-2}[\sum_{t=1}^T S_t^2]/v$ , where  $v$  is an estimate of the “long-run variance”,  $E[T^{-1} S_T^2]$ . As this term is essentially the spectral density for  $y_t$  at zero, it is necessary to specify a weighting function for the autocovariances as well as some lag truncation parameter in order to compute it. Note that  $\text{KPSS} = v^{-1} \sum_{t=1}^T (t/T)^2 \hat{\mu}_t^2$ , where  $\hat{\mu}_t = 1/t \sum_{j=1}^t y_j$  is the recursive estimate of the “mean” of  $y_t$ . Since an  $I(1)$  process does not have a mean,  $\hat{\mu}_t$  does not tend to zero as it would for a stationary random variable with  $E(y_t) = 0$ .

Table 2 presents the ADF, variance ratio (at lag length  $k = T/3$ ), KPSS and fractional integration tests for  $d = 1$  for the series mentioned earlier in the paper.

---

<sup>8</sup> The result comes from the fact that  $\hat{\rho}_j$  and  $\hat{\rho}_k$  are asymptotically uncorrelated with variance  $T^{-1}$ . The test can be made robust to heteroskedasticity in  $e_t$  by replacing  $T^{1/2}$  by a robust standard error, i.e., regress  $\Delta y_t$  against a constant and  $\Delta y_{t-1}$  and use White's standard errors.

**Table 2**  
Tests for a unit root in selected financial series

Series	$\hat{\rho}$	DF	ADF(4)	Var rat	KPSS	FI test
Log share pr	0.999	−0.007	0.07	0.73	1.00	−1.40
Int rate	0.975	−2.26	−1.98	0.23	1.02	0.43
\$logUS/SF	0.996	−1.318	−1.564	0.79	1.40	−1.77
Returns	0.111	−24.749	−11.751	0.004	0.08	

Series are described in Section 1. ADF(4) is the augmented Dickey–Fuller test with 4 lags. Variance ratio test is for lag  $k = T/3$ . 0.05 critical value is 0.11 from Richardson and Stock (1989), Table 1. Critical values for the KPSS test are 0.46(0.05) and 0.35(0.1).

To compute the KPSS test the weighting scheme was that used by Kwiatowski et al. (1992) and 48 lags were used in reflection of the large sample sizes. Conclusions are unaffected by this choice. The test for fractional integration is based on  $\hat{\tau}_{40}$  and robust standard errors are used. One item worth mentioning in connection with the latter test is that it is very sensitive to the inclusion of  $\hat{\rho}_1$ . In fact, one might wish to exclude this autocorrelation coefficient on the grounds that a single large autocorrelation coefficient should not be taken as evidence of fractional integration; the essence of fractional integration must be that the persistence only shows up in the *cumulated* autocorrelations. Doing so changes the test statistics to −0.06, 1.06 and −1.41 respectively. Overall, the evidence for a unit root in asset prices is very strong while returns do not seem to possess one.<sup>9</sup>

## 2.2. Are financial series independently distributed over time?

All of this does suggest that there is indeed a unit root in asset prices, but the question of dependence in “returns”,  $x_t = \Delta y_t$ , has not yet been addressed.<sup>10</sup> There are a number of ways that this question has been investigated.

(a) Since a series cannot be independently distributed if any of the  $\rho_j(x_t)$ ,  $j = 1, \dots, \infty$ , are non-zero, this points to computation of the autocorrelation function (a.c.f.) of  $x_t$ , followed by tests that the serial correlation coefficients are zero. Generally these are found to be zero, except if data has been measured such

<sup>9</sup> Many of the arguments made to account for unit roots involving “breaking trends” do not seem to be as relevant for asset prices where there are good theoretical reasons to expect a unit root.

<sup>10</sup> In the remainder of the paper we will refer to the changes in the logs of stock prices, foreign exchange rates and the level of bond yields as returns. As a one month bond yield is the log of the price of a zero coupon bond this usage seems consistent.

that there is an overlapping component.<sup>11</sup> As well as returns the same conclusion would be reached for certain “spreads”, e.g., those between bid and ask prices.

(b) A different viewpoint arises by thinking of the impact of the news  $\epsilon_t$  upon returns. With no dependence, i.e.,  $x_t = \epsilon_t$ , the short run and long run impact of news is the same, i.e.,  $\partial x_t / \partial \epsilon_t = 1$ ,  $\partial x_{t+\infty} / \partial \epsilon_t = 1$ . However, if the  $x_t$  process has dependence, e.g., is an MA(1),  $\epsilon_t + \alpha \epsilon_{t-1}$ , then  $\partial x_t / \partial \epsilon_t = 1$ ,  $\partial x_{t+\infty} / \partial \epsilon_t = (1 + \alpha)$ . In general, for  $x_t$  being an MA(q),  $\epsilon_t + \alpha_1 \epsilon_{t-1} + \dots + \alpha_q \epsilon_{t+q}$ ,  $\partial x_{t+\infty} / \partial \epsilon_t = (1 + \alpha_1 + \alpha_2 + \dots + \alpha_q)$ , leading to the idea of testing for dependence by testing if  $H_0: \alpha_1 + \alpha_2 + \dots + \alpha_q = 0$ . Testing if the sum of the  $\alpha$ 's is zero is likely to be more powerful than testing if the individual  $\alpha$ 's are zero, because only a scalar is being tested and it is more likely to be precisely estimated than any of its components. This has led some to estimate an MA(q) to get  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_q$  and then to test if the sum is zero – see the review in Christiano and Eichenbaum (1990). Obviously this method has the disadvantage of assuming that the alternative is a moving average process.

(c) Fama and French (1988a) work with the UC model (3). Their idea is to form  $y_{t+k} - y_t$  and  $y_t - y_{t-k}$ , i.e., the  $k$ th forward and backward *long differences* of the series  $y_t$ , and to regress the former on the latter. With  $y_t$  as the log of stock prices,  $y_{t+k} - y_t$  is the continuously compounded  $k$ th period return and it will be the sum of  $k$  one-period returns,  $r_{t+k} = y_{t+k} - y_{t+k-1}$ ,  $r_{t+k-1} = y_{t+k-1} - y_{t+k-2}$  etc., so that  $y_{t+k} - y_t = \sum_{j=1}^k r_{t+j}$ . In large samples the numerator of the regression coefficient will tend to  $E[(\sum_{j=1}^k r_{t+j}) (\sum_{j=0}^{k-1} r_{t-j})]$  and Fama and French are therefore testing if this is zero. To appreciate what is being tested set  $k = 1, 2, 3$ , giving

$$\begin{aligned} E(r_{t+1} r_t) &= \gamma_1 \quad (k = 1) \\ E[(r_{t+1} + r_{t+2})(r_t + r_{t-1})] &= \gamma_1 + 2\gamma_2 + \gamma_3 \quad (k = 2) \\ E[(r_{t+1} + r_{t+2} + r_{t+3})(r_t + r_{t-1} + r_{t-2})] \\ &= \gamma_1 + 2\gamma_2 + 3\gamma_3 + 2\gamma_4 + \gamma_5 \quad (k = 3), \end{aligned}$$

where  $\gamma_j$  are the autocovariances of  $r_t$ . Thus one is essentially testing if a weighted average of the autocovariances is zero, rather than whether the autocovariances themselves are.

One can think of Fama and French's test as a modification of the LM test for  $\text{var}(y_{Tt}) = 0$  in the UC model (3). This is done by observing that the composite error term for  $\Delta y_t$  in (3) (with  $\beta_1 = 1$  and  $g(y_t) = y_t$ ) is an MA(1),  $v_t + \alpha v_{t-1}$ , and, when  $\text{var}(y_{Tt}) = \text{var}(\xi_t) = 0$ , the true MA(1) coefficient  $\alpha$  equals 0. Inverting the MA to produce an infinite AR,  $(1 - \alpha L + \alpha^2 L^2 - \alpha^3 L^3 \dots) x_t = v_t$  creates a

<sup>11</sup> Many earnings series exhibit a weak first order dependence because of this measurement feature.

<sup>12</sup> Returns will include the dividend yield so that  $y_t$  needs to be re-defined as beginning of period price plus dividend. In fact, the stability of the dividend yield means that most variation in returns is from the capital gains component.

non-linear regression of the form  $x_t = (\alpha x_{t-1} - \alpha^2 x_{t-2} + \alpha^3 x_{t-3} - \dots) + \nu_t = g(z_t, \alpha) + \nu_t$ , and the LM test that  $\alpha = 0$  involves examining the covariance between  $x_t - g(z_t, \alpha = 0) = x_t$ , and  $[\partial g(z_t, \alpha)/\partial \alpha]_{\alpha=0} = \{\partial(\alpha L - \alpha^2 L^2 + \alpha^3 L^3 \dots) x_t\}/\partial \alpha|_{\alpha=0} = x_{t-1}$ , i.e., it examines the first order serial correlation coefficient of returns. Now, the LM test is a powerful test if the alternative is in the vicinity of the null, but may be dominated by others if the alternative is far from the null, i.e.,  $\alpha$  is not close to zero. This idea that one might use information that represents the alternative to improve the performance of the LM test has been exploited in the work of King (1985) on the design of point optimal tests. Under the null hypothesis, it is clear that any combination of past returns should be uncorrelated with current returns, and this indicates that we might form another test based on linear combinations of lagged returns that more closely approximates the alternative. Because the emphasis in Fama and French is upon the importance of the temporary component and, as the  $\text{var}(y_{T,t})$  becomes large,  $\alpha$  tends towards  $-1$ , this points to an examination of the covariance between  $[\partial g(z_t, \alpha)/\partial \alpha]_{\alpha=-1} = -(x_{t-1} + 2x_{t-2} + 3x_{t-3} + \dots)$  and  $x_t$ . As this is  $\gamma_1 + 2\gamma_2 + 3\gamma_3 + \dots$ , (ignoring the sign), the argument suggests that Fama and French are potentially improving on the LM test by utilizing information about the likely alternative. Ultimately, looked at from a variable addition perspective, one is trying to find suitable regressors to add to the equation describing returns  $x_t$ , and there are many possibilities, of which Fama and French's is just one set. Power considerations will ultimately determine which variable addition test is best.<sup>13</sup>

Let us term the coefficient estimated from the  $k$ th differenced regression,  $\hat{\beta}(k)$ . Because the denominator of  $\hat{\beta}(k)$  will tend to  $E[(\sum_{j=0}^{k-1} r_{t-j})^2] = k\sigma_\eta^2$  if there is no temporary component, it is clear that  $\hat{\beta}(k)$  will tend to zero as  $k$  rises. If there is a temporary component, the denominator still tends to infinity and the numerator remains bounded, so  $\hat{\beta}(k)$  still tends to zero. However, for small  $k$ ,  $\hat{\beta}(k)$  will generally be negative if there is a temporary component, owing to the fact that returns have  $\Delta_k y_{T,t}$  in them and  $\eta_t$  is uncorrelated. For example, if  $y_{T,t}$  was uncorrelated, then the first autocovariance of returns will be  $-\text{var}(y_{T,t})$ . Hence a temporary (mean-reverting) component makes  $\hat{\beta}(k)$  negative, and a plot of  $\hat{\beta}(k)$  against  $k$  should follow a U-shape with  $k$ . Indeed, Fama and French find that this is so for many stocks they examine (for market, decile and industrial portfolios).  $\hat{\beta}(k)$  can also be used as a measure of the importance of the temporary component. Because the  $R^2$  from a regression of  $w_t$  on  $z_t$  is given by  $\hat{\beta}^2 \text{var}(z_t)/\text{var}(w_t)$ , setting  $w_t = y_{t+k} - y_t$  and  $z_t = y_t - y_{t-k}$ , and using the fact that the variances of

---

<sup>13</sup> It needs to be emphasized that the LM test for  $\alpha = -1$  is the KPSS test described earlier – Tanaka (1990) – so that Fama and French are not performing an LM test that  $\alpha = -1$ . Rather, they use the information about what the model would look like if  $\alpha$  was  $-1$  in order to construct a test that  $\alpha = 0$ .

these two are equal due to stationarity, it is apparent that the  $R^2$  is identical to the square of the estimated coefficient from the regression. If there is no temporary component this should be zero in large samples, so that  $\hat{\beta}(k)$  gives a measure of the importance of the temporary component at the  $k$ th period horizon.

Because the  $\hat{\beta}(k)$  are the first order serial correlation coefficients of  $k$ -period returns, and it is known that there are small sample biases in this estimated coefficient, Fama and French provide some bias corrections that seem to be quite important. One should also note that, if  $\beta(k)$  is zero, then the error in the regression is just the dependent variable  $\Delta_k y_t$ , and this will be an MA( $k - 1$ ) when  $r_t = \Delta y_t$  is white noise, necessitating the use of an autocorrelation consistent covariance matrix. Alternatively, one can generate standard errors by simulation methods; since the null hypothesis is the absence of temporary components,  $\Delta y_t$  can be taken to be white noise and samples can be generated from this specification and the empirical distribution of  $\hat{\beta}(k)$  may be determined. Note that the test is very easily formed if we have available to us regression programs that do forward and backward lags.

Autocorrelation-consistent standard errors for  $\hat{\beta}(k)$  can be computed in a number of ways. For a regression model  $y_t = x'_t \beta + u_t$ , where  $E(x_t u_t) = 0$ , the asymptotic variance of  $T^{1/2}(\hat{\beta} - \beta)$  is

$$[E(x_t x'_t)]^{-1} \left[ E \frac{1}{T} \left( \sum_t x_t u_t \right) \left( \sum_t x_t u_t \right)' \right] [E(x_t x'_t)]^{-1}. \quad (6)$$

Defining  $\phi_t = x_t u_t$ , the middle term will be  $\gamma_0^\phi + 2 \sum_{j=1}^{k-1} \gamma_j^\phi$  when  $\phi_t$  is an MA( $k - 1$ ). Hansen and Hodrick (1980) used this approach, and it is the basis of Table 3, which presents  $\hat{\beta}(k)$  and  $t$ -ratios formed with these asymptotic standard

Table 3  
Long return tests for CRSP data

Return length ( $k$ )	FF ser. corr.	FF $t$ -stat	Bol/Hod $t$
12	-0.002	-0.018	-0.72
24	-0.159	-1.11	-1.78
36	-0.261	-1.67	-1.58
48	-0.200	-1.5	-0.79
60	-0.089	-0.60	-0.07
72	0.073	0.43	0.21
84	0.160	0.76	0.44
96	0.046	0.18	-0.36

Column 2 gives the first order serial correlation coefficient of the long returns of length given in the first column. Column 3 is the Fama–French (FF)  $t$ -statistics constructed with Hansen–Hodrick (1980) standard errors. Column 4 gives the  $t$ -ratio that the coefficient in the regression of one period returns  $r_t$  upon  $\sum_{j=1}^{2k-1} w_j r_{t+1-j}$  is zero.

errors. The evidence for predictability is not strong. Several authors have argued that the asymptotic theory may not be a good approximation, e.g., Nelson and Kim (1993) and Bollerslev and Hodrick (1992), finding that the use of *t*-ratios formed in this way results in a strong tendency to reject the null that  $\beta(k)$  is zero. Richardson and Stock (1989) examine the distribution of sums of  $\hat{\beta}(k)$  as well as that of a joint test that the  $\hat{\beta}(k)$  are zero when  $k/T$  tends to a constant as  $T \rightarrow \infty$ . The distribution is non-standard.

Bollerslev and Hodrick suggest an alternative way of measuring the predictability of long-horizon returns that exploits the presumed covariance stationarity of returns. As observed previously, the numerator of  $\hat{\beta}(k)$  involves the

$$\text{cov} \left[ \left( \sum_{j=1}^k r_{t+j} \right) \left( \sum_{j=0}^{k-1} r_{t-j} \right) \right] = \text{cov} \left[ r_{t+1}, \sum_{j=1}^{2k-1} \omega_j r_{t+1-j} \right], \quad (7)$$

where  $\omega_j = j$  for  $1 \leq j \leq k$  and  $\omega_j = 2k - j$  for  $(k+1) \leq j \leq (2k-1)$ . Thus, when  $k = 3$ , one is considering the covariance of  $r_{t+1}$  with  $(r_t + 2r_{t-1} + 3r_{t-2} + 2r_{t-3} + r_{t-4})$ , which equals  $\gamma_1 + 2\gamma_2 + 3\gamma_3 + 2\gamma_4 + \gamma_5$ , and, as was established earlier, represents the numerator of  $\hat{\beta}(k)$  in large samples. A simple way to see the above for  $k = 3$  is to write the LHS of (7) as

$$\begin{aligned} & \text{cov} \quad [ \{(1 + L^{-1} + L^{-2}) r_{t+1}\}, \{(1 + L + L^2) r_t\} ] \\ &= \text{cov} \quad [ r_{t+1}, (L^{-2} + 2L^{-1} + 3 + 2L + L^2) r_t ] \\ &= \text{cov} \quad [ L^{-2} r_{t+1}, \{1 + 2L + 3L^2 + 2L^3 + L^4\} r_t ]. \end{aligned}$$

Hence they advocate a regression of  $r_{t+1}$  upon  $\sum_{j=1}^{2k-1} \omega_j r_{t+1-j}$ , and a *t*-test on the coefficient of this regressor, observing that the error term in this regression is serially uncorrelated, and so no special formula needs to be used to take account of the serial correlation. Simulations in their paper reveal this test to be much better behaved in “small” samples. Table 3 gives *t*-statistics that the coefficient in the regression of  $r_{t+1}$  against  $\sum_{j=1}^{2k-1} \omega_j r_{t+1-j}$  is zero for various values of  $k$ . The pattern turns out to be very similar to that observed with Fama and French’s method and one would draw essentially the same conclusions regardless of which type of test one adopted.

Even if our series on (say) earnings had passed all the tests described above, it would still not follow that the series would be independent. The reason is that they are essentially tests of a lack of *correlation* in returns across time, and independence is a broader notion, referring to the ability to express a joint density as the product of marginals. One way to think of the relation of independence and correlation is to refer to “independence in the *r*th moment”. Thus what we are investigating above is independence in the first moment, whereas complete independence requires that this be true for all other moments as well. In particular, let  $x_t$  and  $x_{t-j}$  ( $j \neq 0$ ) be independent with zero means. Then independence

Table 4  
Measures of non-linearity in series

ACF	STK-M	<i>t</i>	BOND	<i>t</i>	\$US/SF	<i>t</i>	STK-D	<i>t</i>
Lag 1	0.292	8.1	0.298	6.5	0.114	2.9	0.189	24.0
Lag 2	0.159	4.4	0.308	6.8	0.118	3.0	0.174	22.1
Lag 3	0.225	6.3	0.207	4.6	0.169	4.3	0.140	17.8
Lag 4	0.122	3.4	0.275	6.1	0.212	5.4	0.092	11.7
Lag 5	0.060	1.7	0.200	4.4	0.083	2.1	0.142	18.0
Lag 6	0.093	2.6	0.251	5.6	0.110	2.8	0.113	14.4
Lag 7	0.100	2.8	0.285	6.3	0.161	4.1	0.076	9.6
Lag 8	0.339	9.5	0.212	4.7	0.012	0.3	0.083	10.5
Lag 9	0.335	9.4	0.173	3.9	0.102	2.6	0.076	9.6
Lag 10	0.201	5.6	0.158	3.6	0.102	0.1	0.074	9.0
LM test	429.5		301.6		111.9			
BDS test	8.88		22.74		6.88			
Tsay test	3.35		9.45		1.02			

The BDS test was constructed by setting  $\gamma$  to 1.25 times the standard deviation of the series and using  $N = 4$  histories. The test should be referred to an  $N(0,1)$  random variable. The Tsay test uses  $p = 5$  lags and is referred to an  $F$  distribution with  $pm = p(p+1)/2$  and  $T-pm-p-1$  degrees of freedom. For  $m = 5$  the 5% critical value is approximately 2.07 ( $T = \infty$ ). The LM test is  $T$  times the sum of the squared acf value for 12 lags. The 5% critical value for  $\chi^2(12)$  is 21.02.

between  $x_t$  and  $x_{t-j}$  means that  $E[g(x_t)h(x_{t-j})] = E[g(x_t)]E[h(x_{t-j})]$ , implying that  $\text{cov}(g(x_t)h(x_{t-j})) = 0$  for any measurable functions  $g(\cdot)$  and  $h(\cdot)$ .

Of course the number of functions is vast, making it impossible to truly test independence with a finite sample. One could replace  $g(\cdot)$  and  $h(\cdot)$  by polynomial approximations, e.g.,  $g(x_t) = \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2 + \dots$ ,  $h(x_{t-j}) = \delta_0 + \delta_1 x_{t-j} + \delta_2 x_{t-j}^2 + \dots$  and then test if all the pairwise terms,  $\text{cov}(x_t^k x_{t-j}^\ell)$ , are zero, allowing the order of approximating polynomials to tend to infinity with the sample size. Such non-parametric tests have been suggested in the literature but rarely used – see Cameron and Trivedi (1993) for such an approach. Instead, attention has focused upon whether  $\text{cov}(x_t^k x_{t-j}^\ell)$  is zero for certain specific values of  $k$  and  $\ell$ . Of particular interest have been the choices of  $k = \ell = 2$  and  $k = 2$ ,  $\ell = 1$ . These hypotheses could be tested in one of two ways. The first involves regressing  $x_t^2$  against a constant and one of  $x_{t-j}^2$  and  $x_{t-j}$ , while the second forms either  $x_t^2 x_{t-j}^2$  or  $x_t^2 x_{t-j}$  and regresses these against a constant, testing if the intercept is zero. Table 4 presents these quantities for our financial series and it is clear from this that there is higher order dependence (STK-M and STK-D are the monthly and daily stock returns data respectively).

Generally, the argument has been that the lack of independence arises from the presence of “nonlinearities” in the series. This is a somewhat imprecise description and one that arises more by default than anything else. Since the a.c.f. of a series  $x_t = \Delta y_t$  tests if  $x_t$  has a linear structure, such as  $x_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j}$ , any

dependence in  $x_t$  manifest in higher order moments might cause us to say that  $x_t$  has a non-linear structure. Attempts have been made to construct various tests for general non-linear dependence. The best known of these are Tsay's test (Tsay, 1986) and the BDS test – Brock et al. (1987) – but many others could be designed.

Tsay's test implicitly involves testing if the coefficients of  $w_t' = (x_{t-1}^2, x_{t-1} x_{t-2}, \dots, x_{t-p}^2)$  are zero in the regression of  $x_t$  upon  $z_t' = (1, x_{t-1}, \dots, x_{t-p})$  and  $w_t$ . His test is not expressed in this way but the interpretation just given may be reconciled with his format by starting with

$$x_t = z_t' \alpha + w_t' \gamma + e_t, \quad (8)$$

taking expectations with respect to  $z_t$ ,

$$E(x_t | z_t) = z_t' \alpha + E(w_t | z_t)' \gamma, \quad (9)$$

and then subtracting (9) from (8) to leave

$$x_t - E(x_t | z_t) = (w_t - E(w_t | z_t))' \gamma + e_t. \quad (10)$$

If one assumes that the conditional expectations in (10) are linear in  $z_t$  and are estimated as the predictions,  $(\hat{x}_t, \hat{w}_t)$ , from the regression of  $x_t$  and  $w_t$  upon  $z_t$ , respectively, one gets Tsay's test, expressed as a regression of the residuals  $x_t - \hat{x}_t$  against  $w_t - \hat{w}_t$ .<sup>14</sup> Looking at Tsay's test in this way emphasizes that it is a test for non-linearities in the conditional mean.

The BDS test can be shown to be testing for independence by focusing upon estimated marginal and joint densities. Consider two random variables  $X$  and  $Z$ . These are independent if

$$f_x(x)f_z(z) = f_{xz}(x, z), \quad (11)$$

where  $f(\cdot)$  are the density functions of the random variables. Consider replacing  $x$  and  $z$  with the random variables  $X$  and  $Z$  themselves, and, after that, taking the expectation. Then, due to independence,

$$E[f_x(X)]E[f_z(Z)] = E(f_{xz}(X, Z)), \quad (12)$$

and we will now show that the BDS test involves testing this condition.

Constructing the test based on the sample moments gives

$$\left[ T^{-1} \sum_{t=1}^T f_x(x_t) \right] \left[ T^{-1} \sum_{t=1}^T f_z(z_t) \right] - T^{-1} \sum_{t=1}^T f_{xz}(x_t, z_t).$$

The problem is that we don't know the densities and therefore we need to estimate them. Suppose we do this non-parametrically with symmetric kernels  $K_x$

---

<sup>14</sup> The fact that the conditional expectations may not be linear is unimportant under the null hypothesis, since then  $E(x_t | z_t) = 0$  and  $\gamma = 0$ .

and  $K_z$ , widow widths  $h_x$  and  $h_z$ , using the leave-one-out estimator, and a product kernel for the estimation of the joint density.<sup>15</sup> Then the test would be based on

$$\begin{aligned}
 & \left[ T^{-1} \sum_{t=1}^T \sum_{j=1, j \neq t}^T \frac{1}{h_x(T-1)} K_x((x_t - x_j)/h_x) \right] \\
 & \times \left[ T^{-1} \sum_{t=1}^T \frac{1}{h_z(T-1)} \sum_{j=1, j \neq t}^T K_z((z_t - z_j)/h_z) \right] \\
 & - T^{-1} \sum_{t=1}^T \sum_{j=1, j \neq t}^T \frac{1}{h_x h_z(T-1)} K_x((x_t - x_j)/h_x) K_z((z_t - z_j)/h_z) \\
 & = \left[ 2 \frac{1}{h_x T(T-1)} \sum_{1 \leq t < j}^T K_x((x_t - x_j)/h_x) \right] \\
 & \times \left[ 2 \frac{1}{h_z T(T-1)} \sum_{1 \leq t < j}^T K_z((z_t - z_j)/h_z) \right] \\
 & - 2 \frac{1}{h_x h_z T(T-1)} \sum_{1 \leq t < j}^T K_x((x_t - x_j)/h_x) K_z((z_t - z_j)/h_z).
 \end{aligned}$$

In the BDS case  $Z$  is simply the lagged value of  $X$  and the marginal density will be the same for both variables, so it makes sense to set  $\hat{f}_x(x) = \hat{f}_z(z)$  as well as making  $K_x = K_z = K$ , and  $h_x = h_z = h$ . The effective sample size is now  $T_1 = T - 1$  and  $t = 1$  corresponds to the second observation of the original sample. Hence the test simplifies to

$$\begin{aligned}
 & \left[ 2 \frac{1}{h T_1(T_1-1)} \sum_{1 \leq t < j}^{T_1} K((x_t - x_j)/h) \right]^2 \\
 & - 2 \frac{1}{h^2 T_1(T_1-1)} \sum_{1 \leq t < j}^{T_1} K((x_t - x_j)/h) K((z_t - z_j)/h),
 \end{aligned}$$

which can be written as

$$\frac{C_1^2}{h^2} - \frac{C_2}{h^2},$$

where  $C_k = 2(1/T_1(T_1-1)) \sum_{1 \leq t < j}^{T_1} \prod_{l=1}^k K((x_{tl} - x_{jl})/h)$  and  $x_{t1} = x_t$ ,  $x_{t2} = z_t$ . Eliminating the constant of proportionality  $h^2$ , and replacing the kernel by

---

<sup>15</sup> For all these concepts see Silverman (1986) and Härdle (1990).

$I(h/2 - |x_{it} - x_{jt}|)$ , where  $I(\cdot)$  is an indicator function, gives the numerator of the BDS test when only  $x_t$  and its lag is used.<sup>16</sup> Since the denominator is just the standard deviation of the numerator, the two test statistics are identical. Extension to higher order lags is straightforward.

A choice of  $\gamma$ , the window width, and  $N$  (the maximum lag in  $x_t$  used to form  $z_t$  in constructing the test) needs to be made for the BDS test to be computed. In our situation we choose  $N = 4$  and  $\gamma$  as 1.25 times the sample standard deviation of the variable under investigation. Applying each of these tests to the series supports the lack of independence found from the a.c.f. of squares (see Table 4). It should be mentioned that non-linearities in stock returns, exchange rates and interest rates have been documented before with these tests by Scheinkman and LeBaron (1989), Hsieh (1989b) and Kearns (1990b) respectively.

The connection between method of moments testing and the BDS test illustrated above is very useful for understanding many of the properties of the test. Defining  $v_t = E(f_x(X_t))E(f_z(Z_t)) - f_{xz}(X_t, Z_t)$  the BDS test can be regarded as testing if the sample mean of  $v_t$  is zero, i.e., whether the intercept in the regression of  $v_t$  against unity is zero. Clearly it is possible that  $E(v_t) = 0$  but that the random variables are not independent: although (11) implies (12) the converse is not true. It is also interesting to observe that the  $t$ -statistic from such a regression is robust to heteroskedasticity in  $v_t$ . To see why, specialize (6) to the case where there is only heteroskedasticity to get White's heteroskedastic consistent standard errors (White, 1980). These are  $(T^{-1}\sum x_t^2)^{-1}(T^{-1}\sum x_t x_t' \hat{u}_t^2)(T^{-1}\sum x_t^2)^{-1}$ , where  $\hat{u}_t$  are the residuals from the OLS regression of  $v_t$  on  $x_t$ . These are to be contrasted with the "formula variance"  $\hat{\sigma}_u^2(T^{-1}\sum x_t^2)^{-1}$  associated with OLS computer programs, where  $\hat{\sigma}_u^2 = T^{-1}\sum_{t=1}^T \hat{u}_t^2$ . In the special case that  $x_t = 1$  the two estimators coincide, and this is the situation here. Hence, this suggests that the BDS test is likely to be robust to heteroskedasticity (but not to serial correlation).

The implications of a lack of dependence between  $x_t$  and its past history raises the issue of how to summarize such dependence in a useful way. One way is to treat functions of  $x_t$  as being determined by models such as (1), (2) and (3). Thus, defining the expectation of a random variable conditioned upon its past history as  $E_{t-1}$ , these models make  $E_{t-1}(g(x_t))$  a function of  $x_{t-j}$  ( $j > 0$ ). When  $g(x_t) = x_t^2$ , and  $E_{t-1}(x_t) = 0$ ,  $\sigma_t^2 = E_{t-1}(x_t^2)$  is the conditional variance of  $x_t$  and having  $x_t^2$  generated by such models makes  $\sigma_t^2$  potentially dependent upon the past. The distinction between a conditional,  $\sigma_t^2$ , and an unconditional variance,  $E(x_t^2)$ , was emphasized in Engle (1982), and it is a distinction that has proved to be more important in finance than in any other field of econometrics. Prior to this work,

---

<sup>16</sup> The numerator of the BDS test is  $C_N - C_1^N$  where  $C_N = 2(1/T_N(T_N - 1))\sum_{1 \leq t < j}^{T_N} \prod_{k=0}^{N-1} I(\gamma - |x_{t+k} - x_{j+k}|)$  and  $T_N = T - N + 1$ . When  $N = 1$  we get the formula in the text.

many finance models were understood in terms of the unconditional moments of the series, e.g., the CAPM was stated in terms of the covariance of the return on an asset with the market return, but it soon became evident from the theoretical derivations that the proper items of investigation were actually the conditional moments, unless assets were being held for long periods of time, whereupon the relevant conditioning set for decisions is far in the past and the distinction between conditional and unconditional moments evaporates, i.e., for stationary series in which  $E(g(x_t))$  exists  $\lim_{k \rightarrow \infty} E_{t-k}(g(x_t)) = E(g(x_t))$ .

Table 4 not only reveals that the squares of returns are correlated, but the slow decline in the autocorrelation coefficients may be used to argue that the correlation is very persistent.<sup>17</sup> Brockwell and Davis (1991) defined a *short memory* series as one whose autocorrelation coefficients were bounded like  $|\rho_j| \leq Ca^j$ , as contrasted with a *long memory* process in which  $\rho_j$  declined like  $C_j^{2d-1}$ , where  $d < .5$  and  $C$  is a non-zero constant. Using this definition, the squares of the returns seem to possess long memory. Findings of long memory in the squares of stock returns are legion – examples being Ding et al. (1993) and De Lima and Crato (1994) – but it has also been found in other series such as the forward premium, Baillie and Bollerslev (1994). Many techniques have been used to investigate this property – see Baillie (1996) for a survey. By far the simplest would be to fit models (1), (2) and (3) to suitably transformed  $x_t$ , e.g.,  $x_t^2$  or  $|x_t|$ , and then to test for unit roots, fractional integration etc. as outlined earlier. Sometimes a ‘‘model independent’’ way of assessing long-range dependence is advocated in the form of the modified re-scaled range statistic,  $s^{-1}\{\max_{1 \leq t \leq T} \sum_{i=1}^t (x_i - \bar{x}) - \min_{1 \leq t \leq T} \sum_{i=1}^t (x_i - \bar{x})\}$ , where  $s^2$  is an estimate of the spectrum of  $x_t$  using  $q$  autocovariances and  $\bar{x}$  is the sample mean of  $x_t$  – see Lo (1991) and De Lima et al. (1994). The problem with this measure is the choice of  $q$ . As  $q$  grows the test statistic is made completely robust to serial correlation of *any order*, including long-range dependence, so that the power of the test eventually equals its size. Hence any conclusions obtained simply reflect the choice of  $q$ . If you want to find long range dependence keep  $q$  short; if you don’t, make  $q$  long.

Table 5 looks at the cross correlation between  $x_t^2$  and  $x_{t-j}$  ( $j > 0$ ), a relationship sometimes encapsulated as the ‘‘leverage effect’’ that volatility depends upon the sign of returns. There is evidence here of such a negative effect in stock returns and exchange rates, while for interest rates it is not as clear. If there is some relationship for interest rates it would be a positive rather than negative one. The magnitude of the effect may not be large but what is striking is its persistence, with many negative values for the correlation coefficients and a slow dying away of its values.

<sup>17</sup> The same would be true if absolute returns were examined, see Ding et al. (1993).

Table 5  
Cross correlations between squared and level returns

CROSS	STK-M	<i>t</i>	BOND	<i>t</i>	\$US/SF	<i>t</i>	STK-D	<i>t</i>
Lag 1	0.02	0.6	0.04	0.9	-0.10	2.5	-0.07	8.7
Lag 2	-0.09	2.4	0.12	2.7	-0.04	1.0	-0.08	10.4
Lag 3	-0.13	3.6	0.08	1.7	-0.13	3.2	-0.06	7.2
Lag 4	-0.12	3.3	-0.02	0.5	-0.10	2.5	-0.05	6.4
Lag 5	-0.12	3.2	-0.07	1.6	-0.00	0.0	-0.04	5.6
Lag 6	-0.12	3.2	0.15	3.4	-0.03	0.8	-0.05	6.4
Lag 7	-0.10	2.8	-0.02	0.4	-0.02	0.5	-0.05	5.7
Lag 8	0.01	0.2	0.06	1.2	-0.08	2.1	-0.05	6.0
Lag 9	0.1	2.7	0.00	0.1	-0.08	2.1	-0.05	5.8
Lag 10	-0.06	1.5	0.06	3.6	0.02	0.5	-0.03	4.3
Lag 11	-0.18	5.0	-0.02	0.3	0.02	0.5	-0.03	3.5
Lag 12	-0.14	4.0	0.05	1.1	-0.05	1.3	-0.02	2.9

### 2.3. The existence of moments

Very early in the history of financial econometrics Mandlebrot (1963) raised the spectre that asset returns had second (unconditional) moments that did not exist. To assess that possibility he recommended that an examination of the recursive variance be made, and this has been done by Hols and De Vries (1991) for exchange rates, Pagan and Schwert (1990b) for stock returns, and Loretan and Phillips (1994) for a wide variety of returns. If the process  $x_t$  was covariance stationary then the sequence of *recursive variances*,  $\mu_{2,\tau} = \tau^{-1} \sum_{j=1}^{\tau} x_j^2$ ,  $\tau = 1, 2, 3, \dots$  formed by adding on one observation at a time, should converge to  $E(x_t^2)$  as  $\tau \rightarrow \infty$ . If  $E(x_t^2)$  does not exist then there will be no convergence, and in fact one tends to observe “jumps” in the sequence  $\mu_{2,\tau}$ . Pagan and Schwert (1990b) describe a number of possible test statistics to formally test this feature. One involves testing if the mean of the sample variance over one part of the sample is the same as over the remainder. Another is based directly upon  $\psi(t) = (T\hat{\sigma})^{-1/2} \sum_{j=1}^t (x_j^2 - \hat{\mu}_{2,T})$ , where  $\hat{\sigma}$  is an estimate of the asymptotic variance of  $T^{-1/2} \sum (x_j^2 - \hat{\mu}_{2,T})$  (in applications below set to  $\hat{\gamma}_0 + 2\sum_{j=1}^8 \hat{\gamma}_j(1-(j/9))$ , where  $\hat{\gamma}_j$  is the  $j$ th estimated autocovariance of  $x_j^2$ ). The logic of the test is seen by replacing  $\hat{\mu}_{2,T}$  by  $\mu_2$ , which it estimates. Then the random variable  $T^{-1/2} \sum (x_j^2 - \mu_2)$  should have variance  $v$ .  $\psi(t)$  is an  $N(0, (t/T)(1-t/T))$  random variable under the null hypothesis of a constant variance and Fig. 1 shows a plot of  $\psi(t)$  along with the 99% confidence intervals for the monthly stock returns series. Stock returns do not seem to be covariance stationary. Loretan and Phillips (1994) examine a very wide range of series, all of which have this characteristic.

As Loretan and Phillips (1994) point out, what is at issue here is the shape of the density in the tails. Unless it declines sufficiently swiftly moments will not exist, as should be familiar from analysis of the Cauchy density, e.g. see Spanos

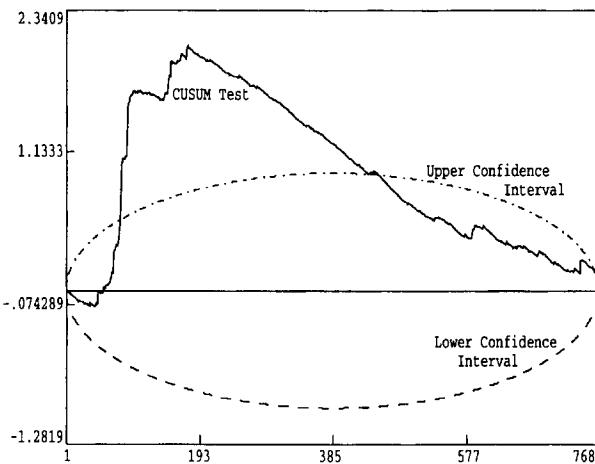


Fig. 1. CUSUM of squares test, CRSP returns, 99% confidence intervals.

(1986, p. 70). More precisely, suppose that there are Pareto-type tails in the density. This means that  $\Pr(X > x) = kx^{-\alpha}$ , so that  $\log[\Pr(X > x)] = \log k - \alpha \log x$ . If  $|\alpha| > c$  then moments of order  $c$  exist, so that what we are interested in is the question of the value of  $\alpha$  for our data. After ordering the data to produce  $x_{(1)} \dots x_{(T)}$ , where  $x_{(1)} > x_{(2)} > \dots > x_{(T)}$  they estimate  $\Pr(X > x_{(j)})$  by the empirical survivor function  $[\#(x_i \geq x_{(j)})/T] = j/T$ , and give a plot of the log  $\hat{P}(X > x_{(j)})$

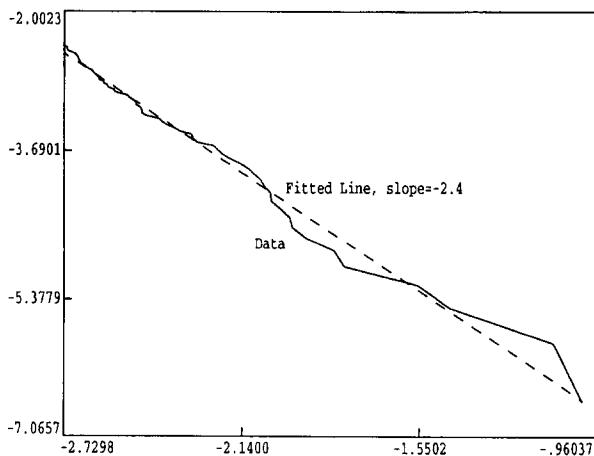


Fig. 2. Right tail of empirical survivor function of CRSP returns.

Table 6  
Estimates of the tail index for various financial series

	SRM	SRD	BR	SF/\$
$\hat{\alpha}$ (left)	2.03	2.29	1.49	2.70
$\hat{\alpha}$ (right)	2.81	2.52	1.50	3.26
$m$	60	1600	40	50

SRM = Monthly stock returns (Fig. 1.5).

SRD = Daily stock returns.

BR = Change in bond rate.

SF/\$ = Change in log SF/\$ exchange rate.

against  $\log x_{(j)}$ . In Fig. 2 we show a plot of these quantities for the monthly stock return data as well as the regression line fitted to it over the last 68 data points (after data has been re-ordered). The slope of this line is  $-2.4$ , which points to the fact that the variance might exist, but no higher order moments.

Loretan and Phillips also utilize more formal methods for estimating the inverse of the tail index,  $\gamma = \alpha^{-1}$ , and thereby  $\alpha$  itself, of which the most popular is the proposal of Hill (1975) that

$$\hat{\gamma} = \left[ (m-1) \sum_{i=1}^{m-1} \log x_{(i)} \right] - \log x_{(m)}.$$

A problem is to determine  $m$ . Asymptotic analysis with  $x_t$  being i.i.d. yields  $m^{1/2}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \alpha^2)$ , showing that  $m$  needs to be large. But clearly the larger  $m$  is the greater the extent to which one is sampling out of the range of the "tail" of the density, and thereby the larger the bias found in the estimator. The situation is made even more complex by the fact that  $x_t$  is not i.i.d.: in Kearns and Pagan (1992), where  $x_t$  is generated from heteroskedastic processes typical of daily stock returns, it is found that the standard deviation of  $\hat{\alpha}$  is around seven times larger than predicted from the i.i.d. case. All of this points to the fact that very large data sets are needed to be able to accurately compute a tail index. Table 6 gives the tail index computation for the three monthly series we are working with, as well as that for U.S. daily stock data from 1885 to 1987. The evidence points to the existence of the second but probably not the fourth moment for all series except interest rates, where the situation is clouded. Increasing  $m$  to 100 for this series produced estimates of 1.03 and 1.47 for  $\alpha$ , so that it may be that the second moment does not exist for that series.<sup>18</sup> Of course it may be that it is higher order moments that fail to exist rather than the second; Loretan and Phillips point to the

<sup>18</sup> Although one might argue that all moments are finite, albeit large, the point is that the series is behaving as if the moments do not exist.

possibility that the fourth moment may behave in this way.<sup>19</sup> Since the variance of the estimated second moment depends upon the fourth moment the failure to exist of the latter will generally mean that occasional large outliers will be found for estimates of the second moment and this will show up as jumps in the recursive estimates of the second moment.

One might ask the importance of this observation that moments might not exist for financial analysis. There are a number of responses that can be made. First, if true, it rules out certain models as candidates for the analysis of financial time series, in particular those that imply covariance stationarity for returns. Second, the inputs into many financial models pre-suppose the existence of certain moments, for example the optimal hedging formula depends upon the variance of asset returns. If moments do not exist, attempts to estimate quantities that depend upon them will generally exhibit a lot of volatility, and it may be that this explains the observed fact that many estimates made of the optimal hedge display quite a bit of instability. More generally, this fact may be important to many econometric procedures, e.g., tests for unit roots such as the variance ratio test and the Dickey–Fuller test. The effects are likely to be particularly devastating for the former.

#### 2.4. Are returns normally distributed?

Most tests of normality focus upon higher order moments of  $x_t$ , in particular inquiring if  $E(x_t^3) = 0$ ,  $E(x_t^4) = 3[E(x_t^2)]^2$ , i.e., whether there is skewness or excess kurtosis. This leads to the two most widely used tests,  $\hat{\tau}_1 = T^{-1} \sum_{t=1}^T (1/\sqrt{6\hat{\sigma}^6})\hat{u}_t^3$  and  $\hat{\tau}_2 = \sum_{t=1}^T (1/\sqrt{24\hat{\sigma}^8})(\hat{u}_t^4 - 3\hat{\sigma}^2\hat{u}_t^2)$ , where  $\hat{u}_t = x_t - \bar{x}$ . There are some difficulties with the use of these statistics to assess normality stemming from the fact that the quantities  $x_t^k$  exhibit dependence. Accordingly, one might want to make these tests robust to this dependence. Perhaps the simplest way to do this is to recognize that there are three moment conditions being used in the construction of each of the tests, e.g., taking the test for skewness these would be  $E((1/\sqrt{6\hat{\sigma}^6})x_t^3 - \tau_1) = 0$ ;  $E(x_t) - \mu = 0$ ; and  $E((x_t - \mu)^2 - \sigma^2) = 0$ . Defining these as  $E(\zeta_i(x_t; \theta)) = 0$ , where  $\theta' = [\tau_1, \mu, \sigma^2]$ , the test for skewness is based

<sup>19</sup> Loretan and Phillips observe that the failure of a fourth moment to exist can be a problem with the cusum test  $\psi(t)$ . In particular, they allow the errors to be realizations from a stable law, in which case the quantity  $v$  in the denominator of the test may not exist. They propose modifications of the test that are robust to this feature. Mandlebrot (1963) also argued for stable processes as a way of inducing infinite variance into financial time series. The problem with this approach is that it fails to produce series that are dependent, and some additions need to be made to make the series exhibit the type of dependence seen in the squares of returns. From some simulations that Phil Kearns and I have run, it does not seem as if the assumption of stable laws is a good predictor of what the behavior of test statistics would be when moments fail to exist because of an IGARCH structure – see also Ghose and Kroner (1994).

Table 7  
Testing departures from normality for returns data

	STK ret-M	STK ret-D	Bond rates	Ex rate
Skewness t				
Non-robust	3.04	-13.1	-11.36	-1.14
Robust	0.4	-0.78	-0.96	-0.46
Kurtosis t				
Non-robust	44.28	213.0	55.1	7.54
Robust	2.11	2.17	1.65	2.04
Fraction around 0	0.14	0.14	0.27	0.14
$\hat{f}(0)$	0.52	0.56	0.99	0.52

The skewness and kurtosis tests should be referred to an  $N(0,1)$  random variable.

Fraction is the fraction of standardized  $x_t$  lying between -0.1257 and 0.1257. This is 0.1 for an  $N(0,1)$  random variable. NP estimate of density of standardized  $x(t)$  at origin.  $f(0)$  for an  $N(0,1)$  density is 0.40.

upon the first element of  $T^{-1} \sum_{t=1}^T \zeta_1(x_t; \hat{\theta})$ , where  $\hat{\theta}$  is the method of moments estimator of  $\theta$ , i.e., upon  $\hat{\tau}_1$ , and it is well known from the theory of GMM estimators how to make tests concerning  $\tau$  robust to dependence in  $\zeta_1$ . The tests  $\hat{\tau}_j$  ( $j = 1, 2$ ) given above then correspond to the non-robust “ $t$ -statistic” for whether the first element of  $T^{-1} \sum_{t=1}^T \zeta_j(x_t; \hat{\theta})$  is zero. Table 7 presents both robust and non-robust tests of skewness and excess kurtosis. What is striking from this table is how the conclusions reached vary according to whether one has allowed for the dependence in  $x_t^k$ . If no allowance is made there are very emphatic rejections, but the converse tends to be true once some adjustment is performed. Only in the case of the kurtosis tests do the conclusions generally match up, and the presence of excess kurtosis is frequently interpreted as the density of returns possessing “fat tails”.

Rather than focusing upon the higher order moments of returns, it is sometimes more useful to obtain a picture of the density for  $x_t$  by using non-parametric estimation methods, and to concentrate upon certain of the characteristics that stand out from such an inspection. An easy to way to estimate the density non-parametrically is to use a kernel based estimator  $\hat{f}(x) = (1/Th) \sum_{t=1}^T K((x_t - x)/h)$ . Fig. 3, Fig. 4, and Fig. 5, computed with a Gaussian kernel and  $c = 0.9\hat{\sigma}_x T^{-1/5}$  for the window width, present non-parametric estimates of the densities of stock returns and the changes in the one month bond yield and the log of \$US/SF respectively, *standardized* as  $\hat{\sigma}_x^{-1}(x_t - \bar{x})$ . For comparison purposes the figures also have the density of an  $N(0,1)$  density. The densities for stock returns and exchange rates tend to be fatter tailed than the normal and to possess marked peaks. It is sometimes difficult to see the fatter tails, but the range of the horizontal axis coincides with the minimum and maximum values observed in the data, events which would be highly improbable with a normal density.

An alternative to kernel based non-parametric density estimation is the “semi

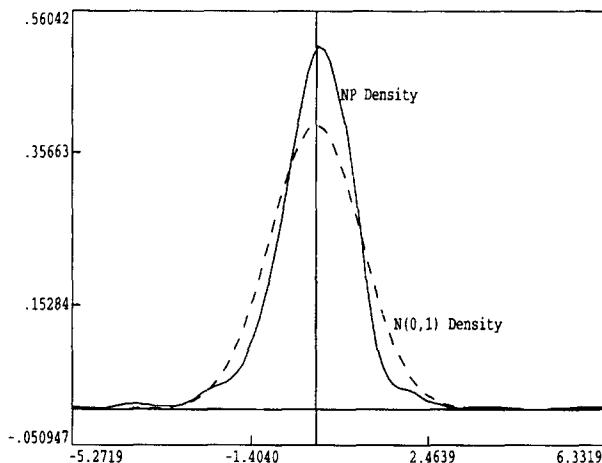


Fig. 3. NP estimate of density of standardized monthly CRSP returns.

non-parametric'' (SNP) method advanced in Gallant et al. (1991) and other articles. Suppose that  $z$  is a random variable with zero expectation and unit variance. The classical approach to the approximation of the density for  $z$  would replace it by the product of the standard normal density  $\phi(z) = (2\pi)^{-1/2} \exp(-z^2)$  and a polynomial  $[P(z)]^2 = [1 + \alpha_1 z + \dots + \alpha_p z^p]^2$ . To force this approximation to integrate to unity the approximating density must therefore be

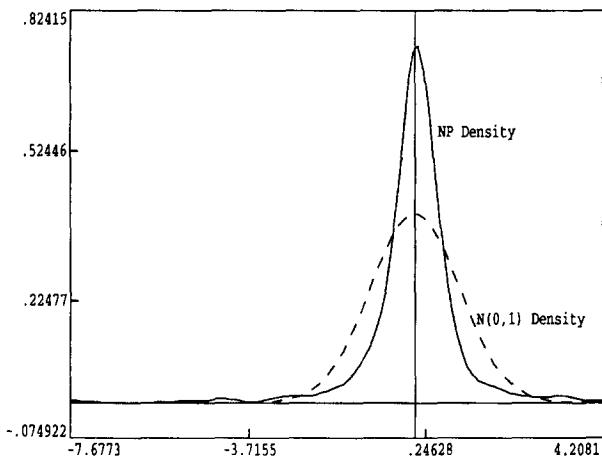


Fig. 4. NP density of standardized returns on one month bond yields.

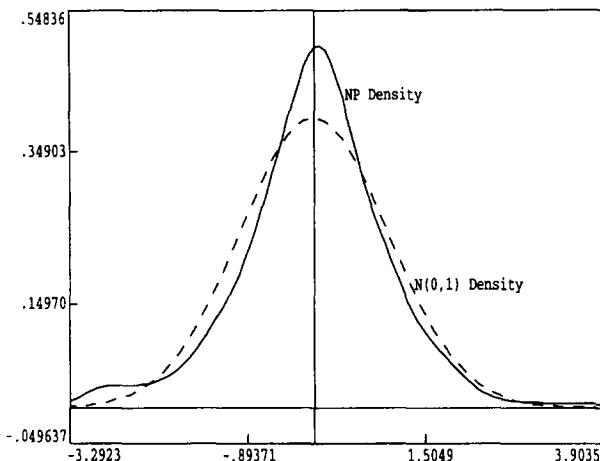


Fig. 5. NP density of returns to \$US/SF.

$\tilde{f}(z) = P^2(z)\phi(z)/\int P^2(\psi)\phi(\psi)d\psi$ .<sup>20</sup> If the density of  $x$  is to be approximated, and  $x$  has mean  $\mu$  and standard deviation  $\sigma$ , Gallant finds the density of  $x$  as  $\sigma^{-1}\tilde{f}(z)$ , where  $z = \sigma^{-1}(x - \mu)$ . After these substitutions all unknown parameters,  $\alpha_1, \dots, \alpha_p$ ,  $\mu$  and  $\sigma$  are chosen by maximizing  $\sum_{j=1}^T \log \tilde{f}(x_j)$ . Computational details are available in Gallant et al. (1991) with an application to exchange rates, while Gallant et al. (1992) explore stock returns and the volume of trading.<sup>21</sup> All of these studies show that financial time series exhibit densities which have tails that are fatter than the normal, and have much higher peaks than the normal around zero, i.e., one has too many small and large returns to have come from a normal density.

It is worth commenting here on the conventional wisdom that fat tails are the cause of excess kurtosis. Deleting the two largest and smallest observations in the data on one-month bond yields reduces  $\hat{\tau}_2$  from 55.1 to 28.2, demonstrating the influence of "fat tails". However, if one further reduces the sample by eliminating the 200 smallest and largest observations, essentially leaving the observations around the mean, the statistic becomes  $-8.39$ , showing that the coefficient also

<sup>20</sup> In the classical approach  $P^2(z)$  are the sum of Hermite polynomials – see Spanos (1986, p. 204) – but this simply represents a reparametrization.

<sup>21</sup> In most of the SNP applications the density being estimated is the density of  $x_t$  conditional upon  $x_{t-1}, \dots, x_{t-p}$  and the volume of trading rather than the unconditional density, and the latter is derived from the former. The procedure described for estimation remains the same however, except that  $\sigma$  and  $\mu$  are taken to be simple functions of  $x_{t-1}, \dots, x_{t-p}$  whose parameters are to be estimated along with the  $\alpha_j$ .

responds to the strong peak or leptokurtosis in the density.<sup>22</sup> Because the leptokurtosis is so dramatic it is useful to have some measures of it that are easily computed, and we have found two of these to be useful. The first compares the fraction of observations of the standardized data that lie between  $\pm 0.1257$ ; this should be 0.1 if the density was  $N(0,1)$ . The second is the non-parametric density estimate of the standardized data at the origin,  $\hat{f}(0)$ . Table 7 shows these, and they replicate the visual impression.

### 3. Models for univariate series

#### 3.1. Formulating models of volatility

Models need to be devised to reproduce some (or all) of the characteristics just listed.<sup>23</sup> A useful strategy is to view returns as being the product of two random variables, i.e.,  $x_t = \sigma_t \epsilon_t$ , where  $\epsilon_t$  is i.i.d. (0,1), and  $\sigma_t$  is a random variable independent of  $\epsilon_t$  whose properties will determine the nature of  $x_t$ . If  $\sigma_t^2$  is assumed to be i.i.d. ( $\sigma^2, v$ ), then  $E(x_t^k) = E(\sigma_t^k)E(\epsilon_t^k)$  and, if  $\epsilon_t$  is n.i.d. (0,1),  $E(x_t^4) = 3E(\sigma_t^4) \geq 3[E(\sigma_t^2)]^2 = 3\sigma^4$ , which illustrates the fact that this model of returns produces the desired feature of excess kurtosis. Historically, such “mixture models” appealed for other reasons, principally the ability to interpret  $\sigma_t$  as related to the quantity of “news” items arriving at a point in time. Such an interpretation means that the variance of returns at a point in time will be an increasing function of the amount of “news” arriving in the market. Volatility of returns will then be high when the market is exposed to large amounts of information. Studies have also appeared in which  $\sigma_t$  was taken to be an increasing Poisson process while  $\epsilon_t$  is n.i.d. ( $\mu_\epsilon, \sigma_\epsilon^2$ ). Such “jump” processes have been applied to stock returns, Friedman and Laibson (1989) and Jorion (1988), and to exchange rates by Nieuwland et al. (1994) and Feinstone (1987).

Perhaps the major defect of this literature was its inability to account for autocorrelation in  $x_t^2$ , as the assumptions make  $x_t$  (and hence  $x_t^2$ ) i.i.d. This need not lead one to jettison the model, but, instead, should concentrate attention upon the fact that  $\sigma_t^2$  has to be made dependent upon the past in some way, and choosing suitable relations for  $\sigma_t^2$  has been a major item of interest in financial

<sup>22</sup> This symmetric trimming strategy is strictly valid only if the data density is symmetric and Table 7 does indicate that there may be a mild departure from it.

<sup>23</sup> In an interesting paper Harrison (1994) shows that the characteristics of 18th century financial asset returns are the same as the 20th century ones. This is not to say that all series have these characteristics; Diebold and Lopez (1995) point to the fact that there is considerable variety in the behaviour of series. Nevertheless, whilst there are always exceptions to “stylized facts”, they are useful as summary measures of what is distinctive about many financial series and therefore what must be generally accounted for.

econometrics during the past decade.<sup>24</sup> Broadly, there are two ways of doing this. One is to recognize that  $\sigma_t^2$  is a random variable and to regard it as being generated from one of the processes (1), (2) and (3). An alternative is to allow  $x_t^2$  to follow the same set of processes, i.e.,  $g(x_t) = x_t^2$ , and then derive the imputed model for  $\sigma_t^2$  that generates the requisite outcome. Both approaches are featured in the literature. As well as this division, there is a further one, in that, when following the first strategy, it is possible to distinguish between cases in which  $\sigma_t^2$  is explicitly related to past returns, and those when it is treated as a random variable with no specific reference to past returns as a determinant of its evolutionary path. The latter is generally referred to as “stochastic volatility” and will be treated later in this section.<sup>25</sup>

### 3.1.1. The GARCH class of models

Engle (1982) introduced the Autoregressive Conditional Heteroskedasticity model of order one (ARCH(1))<sup>26</sup>

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2,$$

which, using the identity  $x_t^2 \equiv E_{t-1}(x_t^2) + v_t = \sigma_t^2 + v_t$ , where  $E_{t-1}(v_t) = 0$ , becomes an AR(1) in the squared returns,

$$x_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + v_t,$$

and this produced a simple way of capturing the dependence in  $x_t^2$ , or “volatility clustering”. A quick glance at the implied a.c.f. for  $x_t^2$  however shows that it is unlikely that such a simple model has managed to capture all the characteristics. The a.c.f. has both height and shape dimensions, and a minimum of two parameters might be expected to be needed to capture both features. In the ARCH(1) model a single parameter  $\alpha_1$  has to account for both. As the autocorrelation function of an AR(1) is  $\rho_j = \alpha_1^j$ , a small estimated value of  $\rho_1$  implies a low value for  $\alpha_1$ , while the observed persistence in the autocorrelation function of squared returns demands that  $\alpha_1$  be set close to unity. Inevitably, it is impossible for the ARCH(1) model to reconcile these two features.

<sup>24</sup> Another way of inducing some dependence, used by Nieuwland et al. (1994) is to have a jump process driving an AR(1) in  $x_t$ . This does not seem a very satisfactory solution as the dependence in  $x_t$  itself is weak. In fact Nieuwland et al. find that they need  $\sigma_t$  to be dependent as well.

<sup>25</sup> Such a distinction does not bear up under close examination. Suppose that  $\epsilon_t$  was distributed as Student's t and  $\sigma_t$  was a function of past  $x_{t-j}$ . Now  $\epsilon_t$  could be written as  $x_t \eta_t$ , where  $x_t$  is an inverted gamma random variable and  $\eta_t$  is  $N(0,1)$ . Accordingly,  $x_t = \sigma_t x_t \eta_t = \bar{\sigma}_t \eta_t$ , and now  $\bar{\sigma}_t$  is “unobservable”. Paradoxes such as these require precise definitions to resolve them, and these have been supplied by Andersen (1992).

<sup>26</sup> Engle actually has  $x_t = w_t' \delta + e_t$  and  $\sigma_t^2$  is related to  $e_{t-1}^2$ , but we will assume that there are no variables in  $w_t$ . The generalization is easily made.

The introduction of the Generalized ARCH (GARCH(1,1)) process by Bollerslev (1986) improved matters, being a two parameter model

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 x_{t-1}^2, \quad (13)$$

and implying an ARMA(1,1) process for  $x_t^2$  of

$$x_t^2 = \alpha_0 + (\alpha_1 + \beta_1) x_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1}. \quad (14)$$

In this model  $\rho_1 = [1 - \beta_1(\alpha_1 + \beta_1)]\alpha_1^2(1 - 2\beta_1(\alpha_1 + \beta_1) + \beta_1^2)$  and autocorrelations thereafter die away like  $\rho_j = (\beta_1 + \alpha_1)^j \rho_1 (j \geq 2)$ . Thus one could have a small initial autocorrelation coefficient, and yet have the others slowly dying away, provided that  $\alpha_1$  was small and  $\beta_1 + \alpha_1$  large; such a combination effectively produces “root cancellation” in the AR and MA polynomials of (14) and the resulting acf of  $x_t^2$  looks like that of white noise, i.e., flat.

Another way of introducing persistence into the autocorrelation function of squares is to invoke a components model as in (3). Engle and Lee (1994) do this, selecting the following model for  $\sigma_t^2$ ,

$$\sigma_t^2 = \sigma_{pt}^2 + \sigma_{Tt}^2$$

$$\Delta \sigma_{pt}^2 = \phi \nu_{t-1}$$

$$(1 - (\alpha_1 + \beta_1)L)\sigma_{Tt}^2 = -(\alpha_1 + \beta_1)\omega + [\alpha_1 - (\alpha_1 + \beta_1)\phi]\nu_{t-1}.$$

Because of the presence of the same error in both components this is the Beveridge–Nelson model of permanent and transitory effects (Beveridge and Nelson, 1981). In terms of observables it becomes,

$$\begin{aligned} (1 - (\alpha_1 + \beta_1)L)\Delta x_t^2 \\ = (1 - \alpha_1 - \beta_1)\omega + [1 - (1 + \beta_1 - (1 - \alpha_1 - \beta_1)\phi)L + \beta_1 L^2]\nu_t. \end{aligned}$$

It is clear from this expression that, if  $\alpha_1 + \beta_1 \approx 1$  and  $\phi \approx 0$ , there will be a unit root that effectively cancels, leaving an ARMA(1,1) process of the same form as the GARCH(1,1) model. Engle and Lee conclude that this model actually gives a better representation of the data than the GARCH(1,1) model. There are quite different long-run implications however. Unless  $\alpha_1 + \beta_1 = 1$ , i.e., the GARCH process is an Integrated GARCH (IGARCH) process, the ARMA(1,1) model implied by GARCH does not have a unit root, while the components model always does, so that a fraction of shocks into volatility persists.

An integrated process is one with very long memory, whereas a non-integrated ARMA process has only short memory. If it is held that the evidence is actually in favor of long memory in  $x_t^2$  it is natural to try to directly model it. The UC model just discussed does so, at the expense of introducing a unit root into the  $x_t^2$  series, but it might be felt that “treading more softly” would be advantageous. It seems natural therefore to seek to apply the fractional integration model to volatility. Baillie et al. (1996) have done this, naming the model FIGARCH.

The second characteristic of the data which models may need to account for is the correlation between  $x_t^2$  and  $x_{t-j}$ . Using the basic model,  $E(x_t^2 x_{t-j}) =$

$E(\epsilon_t^2 \sigma_t^2 \epsilon_{t-j} \sigma_{t-j}) = E(\sigma_t^2 \epsilon_{t-j} \sigma_{t-j})$ . If  $\sigma_t^2$  is an even function of  $\epsilon_{t-j}$  it follows that this covariance is zero, indicating that any model seeking to explain the cross correlation must specify the conditional variance as an asymmetric function of  $\epsilon_{t-j}$ . Most importantly, GARCH models will be incapable of replicating the feature, and this has led to a search for modifications of the GARCH model that would be satisfactory. The first of these was the exponential GARCH (EGARCH) model of Nelson (1991), set up to incorporate the asymmetric effects of returns on volatility noted by Black (1976) and Christie (1982),

$$\ln \sigma_t^2 = \alpha_0 + \beta_1 \ln \sigma_{t-1}^2 + \alpha_1 z_{t-1}, \quad (15)$$

where  $z_t = [|\epsilon_t| - (2/\pi)^{1/2}] + \delta \epsilon_t$ . In this expression,  $\epsilon_t$  is n.i.d. (0,1) so that  $E|\epsilon_t| = (2/\pi)^{1/2}$ , accounting for that choice of centering factor. Because of the use of both  $\epsilon_t$  and  $|\epsilon_t|$ ,  $\sigma_t^2$  will be non-symmetric in  $\epsilon_t$  and, for negative  $\delta$ , will exhibit higher volatility for large negative  $\epsilon_t$ .

The GARCH/EGARCH class of models make  $\sigma_t^2$  specific functions of  $x_{t-j}$ , but perhaps greater flexibility in these functional forms is needed. A useful way of characterizing the possibilities is given by Engle and Ng (1993). Fixing  $x_{t-j}$ ,  $j = 2, \dots$ , they consider the mapping between  $\sigma_t^2$  and  $x_{t-1}$ , terming this the “news impact curve”. They point out that two broad decisions need to be made: about the *shape* and the *position* of such a curve. In particular, one needs to indicate whether it is symmetric or not and, if it is, what value for  $x_{t-1}$  it is symmetric about. Thus a GARCH process has an impact curve that is symmetric around zero, whereas an EGARCH model’s curve is asymmetric around zero.

Other attempts have been made to produce models with generalized news impact curves. Hentschel (1991) advocates the absolute GARCH (AGARCH) model,

$$\sigma_t = \alpha_0 + \alpha_1 |x_{t-1}| + \beta \sigma_{t-1},$$

or, if one wants to allow for asymmetry and a non-zero centering,

$$\sigma_t = \alpha_0 + \alpha_1 [|x_{t-1} - b| - c(x_{t-1} - b)] + \beta \sigma_{t-1}.$$

The parameter  $c$  accounts for the asymmetry in the news impact curve, and the value of  $b$  determines what point it is asymmetric about. A related model is the QGARCH model of Sentana (1991):

$$\sigma_t^2 = \alpha_0 + \alpha_1 (x_{t-1} - b)^2 + \beta \sigma_{t-1}^2.$$

Strictly speaking this produces a symmetric curve around  $b$ , but not around zero, which has been the traditional point of reference. Some very general formulations have appeared which use the Box–Cox transformation as a way of producing a variety of functional forms. Higgins and Bera (1992) originally wrote

$$\sigma_t^2 = \left[ \alpha_0 + \alpha_1 (x_{t-1}^2)^{1/\lambda} \right]^{1/\lambda}, \quad (16)$$

which could be regarded as a Box–Cox transformation applied to  $x_{t-1}$ . They refer to this specification as NARCH. Ding et al. (1993) apply the Box–Cox transformation to GARCH models, adding in an asymmetry, and designating the model A-PARCH (Asymmetric Power ARCH). One cannot be too optimistic about the popularity of this latter approach since Box–Cox transformations have rarely been applied to the conditional mean, as there seem numerical difficulties in the optimization. Perhaps its main use would be as a specification test.

In all the above formulations the non-linearity was taken to be continuous. Friedman and Laibson (1989) are also concerned with a non-linear response of  $\sigma_t^2$  to past  $x_t$ , concentrating upon whether  $\sigma_t^2$  varies continuously with  $x_{t-1}^2$  up to a certain threshold but is constant thereafter. In their application they estimate the model

$$\sigma_t^2 = \alpha_0 + \alpha_1 g(x_{t-1}^2) + \beta_1 \sigma_{t-1}^2, \quad (17)$$

where  $g(x_{t-1}^2) = \sin [ax_{t-1}^2]$  if  $ax_{t-1}^2 < (\pi/2)$  and is unity if  $ax_{t-1}^2 \geq (\pi/2)$ . Such a modified ARCH (MARCH) model has the characteristic of subdued volatility changes in response to sharp movement in returns. By inspection it is clear that the distributed lag connecting  $\sigma_t^2$  to  $x_{t-1}^2$  will tend to be shorter for large changes than for small ones, so that the effect of events like the crash of October 1987 disappear relatively quickly. When applied to the excess stock yield – the S&P returns less the Treasury Bill return – they find  $\alpha_0 = 0.0003$ ,  $\alpha_1 = 0.0084$ ,  $\beta_1 = 0.6147$ ,  $a = 67$ . Neither  $a$  nor  $\alpha_1$  were significantly different from zero. Because the MARCH formulation is symmetric in  $x_{t-1}$ , one way to test for it is to estimate a GARCH model trimming out returns whose absolute value exceeds some specified number  $\delta$ . If there is no threshold effect the estimated parameters  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_1$  should remain much the same since the GARCH model would then be correctly specified. However, if there is a MARCH effect the GARCH estimates for  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_1$  should be different below the threshold than above. Kearns and Pagan (1993) applied this symmetric trimming idea to Australian stock return data, but found no evidence of a threshold as  $\delta$  was varied through a wide range of data. Threshold GARCH (TARCH) models, in which  $\alpha_1 g(x_{t-1}) = \alpha_1^+ I(x_{t-1} > c) + \alpha_1^- I(x_{t-1} < c)$ , where  $I(\cdot)$  is an indicator function taking the value unity if the event in brackets is true and zero otherwise, have become increasingly popular – Glosten et al. (1993) and Zakoian (1994) being good examples. Hentschel (1994) allows for very general functional forms with the relation

$$\sigma_t^{\lambda_1} = \alpha_0 + \beta_1 \sigma_{t-1}^{\lambda_2} + \alpha_1 \sigma_{t-1}^{\lambda_3} g(\epsilon_{t-1}),$$

and it is easily seen that most of the parametric forms described above can be located within this specification by suitable choice of  $\lambda_j$  and  $g(\cdot)$ .

Non-parametric approaches are also to be found in the literature. Gouriéroux and Monfort (1992) suggest that the unknown relation between  $\sigma_t^2$  and  $x_{t-1}$  might be approximated by a series of steps, i.e., dummy variables are added that

take the value unity within regions of possible values that  $x_{t-1}$  might take, but are zero outside it. Pagan and Schwert (1990a) use non-parametric methods to directly estimate the news impact curve, i.e., they compute  $E(x_t^2 | x_{t-1})$ . The major restriction on this approach is that the number of conditioning variables has to be small, and cannot reasonably be restricted to a few lags of  $x_t$  due to the long memory in asset returns. Hence, in the concluding section of their paper they formulate and estimate a model in which  $\sigma_t^2$  is a GARCH model augmented with Fourier terms  $\cos(ke_{t-j})$ ,  $\sin(ke_{t-j})$ ,  $k = 1, 2, 3$ . This formulation is Gallant's (1981) Flexible Fourier Form, which has been used in the production function literature to model producer behavior. The presence of sine terms means that  $\sigma_t^2$  will be an asymmetric function of  $x_{t-j}$ .<sup>27</sup> Other types of approximating functions would be possible, e.g., one might use Laguerre polynomials, since these are always positive.

The third main characteristic of asset prices is their leptokurtosis, as reflected in the fact that their density has a very large peak around zero, i.e., there are far too many small returns compared to what would be found with a normal density. It seems as if GARCH and EGARCH models allied with the assumption that  $\epsilon_t$  is  $N(0,1)$  fail to adequately capture this leptokurtosis. As an example, we fitted such models to the monthly stock return series, and then computed  $f(0)$  with data simulated from them, producing values of 0.46 (GARCH) and 0.43 (EGARCH) versus the 0.52 seen in the data. There is an obvious failure to match this feature. Another way of addressing this correspondence has been to examine the skewness and excess kurtosis properties of the  $x_t/\hat{\sigma}_t$ . For the CRSP equity returns the normalized residuals have excess kurtosis of 1.46, while for the \$US/SF rate Engle and Bollerslev (1986) found that the normalized residuals had excess kurtosis of 0.74.

The main response to this failure has been to change the density of  $\epsilon_t$  from an  $N(0,1)$  to something else. Engle and Bollerslev made  $\epsilon_t$  follow a Student's  $t$ -density rather than a standard normal, i.e.,

$$f(\epsilon_t | F_{t-1}) = \Gamma\left(\frac{1}{2}(\nu + 1)\right) \left[ \Gamma\left(\frac{1}{2}\nu\right) \right]^{-1} [(\nu - 2)]^{-1/2} \left[ 1 + \epsilon_t^2 (\nu - 2)^{-1} \right]^{-(\nu + 1)/2}.$$

This adds an additional parameter,  $\nu$ , the degrees of freedom associated with Student's  $t$ , which can either be pre-set or estimated along with the other parameters of the model. When Engle and Bollerslev do this they find it to be around ten, and that the normalized residuals from this fitted model did not seem

---

<sup>27</sup> Sine terms raise the possibility of negative estimates of  $\sigma_t^2$ . One could avoid parameter regions that cause these by penalizing the log likelihood whenever parameter estimates are such that a negative  $\sigma_t^2$  was observed.

to have excess kurtosis. Other suggestions along these lines have also been made. Hsieh (1989a) and Nelson (1991) use the GED density

$$f(\epsilon_t) = \frac{\nu \exp(-\frac{1}{2}|\epsilon/\lambda|^\nu)}{\lambda 2^{(1+\nu^{-1})}\Gamma(\nu^{-1})},$$

where  $\Gamma$  is the gamma function and  $\lambda = [2^{-(2/\nu)}\Gamma(1/\nu)/(3/\nu)]^{1/2}$  (when  $\nu = 2$  this produces a normal density while  $\nu > (<)2$  is more thin (fat) tailed than a normal). Nelson estimates models up to an EGARCH(4,4) for stock returns and finds  $\nu = 1.576$  for an EGARCH(2,1) process. Lye and Martin (1991) use the “generalized  $t$  distribution”  $f(\epsilon_t) = \exp[\theta_1 \tan^{-1}(\epsilon_t/\gamma) + \theta_2 \log(\gamma^2 + \epsilon_t^2) + \sum_{i=3}^M \theta_i \epsilon_t^i - z]$ , where  $z$  is a normalizing constant.

It is useful to look at the potential for success of this solution. To do that we will derive an expression for the value of the *unconditional* density of  $x_t$  at the origin,  $f_x(0)$ .

$$\begin{aligned} f_x(x) &= \int f(x_t, \sigma_t) d\sigma_t \\ &= \int f(x_t | \sigma_t) f(\sigma_t) d\sigma_t \\ &= \int f_\epsilon(\epsilon = \sigma_t^{-1} x_t) \sigma_t^{-1} f(\sigma_t) d\sigma_t, \end{aligned}$$

so that  $f_x(0) = f_\epsilon(0)E[\sigma_t^{-1}]$ . This result shows that manipulation of the density for  $\epsilon_t$  will certainly enable one to control to some degree for the leptokurtosis in  $f_x$ , but that it is necessary that  $E[\sigma_t^{-1}]$  does not change as a result. However, estimators of the parameters in the density  $f_\epsilon$ , and those in  $\sigma_t^2$ , will generally be correlated, and therefore there may be offsetting effects. Although it is hard to find analytical expressions for how  $E(1/\sigma_t)$  changes with the values of the parameters  $\theta$  contained in the function determining  $\sigma_t^2$ , numerical experiments with a GARCH(1,1) model point to  $\partial E(1/\sigma_t)/\partial\theta$  being negative, so that any increase in these parameters will tend to depress the peak of the density of  $x_t$  produced by the model. In particular, there would be a conflict between the need to make  $\beta_1 + \alpha_1$  high to account for the strong dependence in the series, and the need to make it low to account for the leptokurtosis. If there is a positive correlation between  $\hat{f}_\epsilon(0)$  and  $\hat{\theta}$ , where these are estimated quantities from some data set, then it may be impossible to reproduce the observed degree of leptokurtosis simply by varying the nature of  $f_\epsilon$ . To illustrate this we fitted a GARCH(1,1) model to the monthly stock return data using both an  $\mathcal{N}(0,1)$  and a  $t$ -density for the errors.<sup>28</sup> The estimated degrees of freedom for the  $t$ -density was 7, and the parameters of the conditional variance function were very similar, except for the intercept, which was  $0.8 \times 10^{-4}$  with a normal density but  $1.3 \times 10^{-4}$  with a  $t$ -density. Thus, whilst the use of a  $t$ -density has raised  $f_\epsilon(0)$ , the expected value of  $\sigma_t$  has declined, and these interact to produce a value of  $f(0)$  that is very close in both models.

<sup>28</sup> The program used was a beta test version of MICROFIT, Version 4.

The conflicts just detailed hint that, even after allowing the density for  $\epsilon_t$  to deviate from normality and a use of general specifications for  $\sigma_t^2$ , existing models may not be rich enough for the statistical modeling of actual data. Both Gallant et al. (1991) and Hansen (1994) find that there is evidence that the higher order moments of  $\epsilon_t$  also depend upon the past. Both propose models for this dependence. Gallant et al. do so by using their SNP approach to density estimation, which effectively makes the density of  $\epsilon_t$  a polynomial in the past history of returns, while Hansen's ARCD (Autoregressive Conditional Density) models make some parameter of the density of  $\epsilon_t$ , say the degrees of freedom parameter in Student's  $t$ -density, change over time. He finds that the degrees of freedom coefficient exhibits quite a deal of variation for some financial series, including the excess holding yield on US Treasury Bills.

### 3.1.2. The stochastic volatility class of models

Although the GARCH class of models has met with substantial empirical success, one item hindering their widespread acceptance within financial circles has been the difficulty that their specification departs from the type of models encountered in finance texts. To some extent this is a consequence of the shift from continuous to discrete time, a subject discussed later, but it also reflects the fact that finance models see volatility as being driven by a process separate from returns *per se*. ‘‘Stochastic volatility’’ (SV) models have therefore arisen in which the EGARCH functional form is retained but the process driving the conditional variance is made an n.i.d. (0,1) process that is independent of  $x_t$ . Such a model has the format

$$\log \sigma_t^2 = \alpha_0 + \beta_1 \log \sigma_{t-1}^2 + \eta_t \quad (18)$$

Because it is a ‘‘two parameter’’ model the stochastic volatility model (ST) has the potential for replicating the type of autocorrelation function for  $x_t^2$  seen in the data. Defining  $\mu = \alpha_0/(1 - \beta_1)$  and  $v = \sigma_\eta^2/(1 - \beta_1^2)$  as respectively the mean and variance of  $\log \sigma_t^2$ , the autocorrelation function of  $x_t^2$  is  $\rho_j = \exp(2\mu) \exp(v(1 + \beta_1^j))$  (see Jacquier et al. (1994) for a very clear derivation of this result). It tends to decline a little slower than the GARCH models. However, the SV model does not fare as well on the correlation between  $x_t^2$  and  $x_{t-j}$ . Since

$$\begin{aligned} \text{cov}(x_t^2 x_{t-j}) &= E(\sigma_t^2 \sigma_{t-j} \epsilon_t^2 \epsilon_{t-j}) \\ &= E(\sigma_t^2 \sigma_{t-j}) E(\epsilon_t^2) E(\epsilon_{t-j}), \end{aligned}$$

from the independence assumptions made about  $\epsilon_t$  and  $\sigma_t^2$ , the SV model predicts a zero correlation between  $x_t^2$  and  $x_{t-j}$ . No transformation of either  $\sigma_t^2$  or  $\eta_t$  would change this situation, and the standard way to produce a correlation is to allow  $\eta_t$  to be correlated with  $\epsilon_t$ . If  $\epsilon_t$  and  $\eta_t$  are jointly normal with correlation coefficient  $\gamma$  we can write  $\eta_t = \gamma \epsilon_t + \zeta_t$ , where  $\zeta_t$  is independent of  $\epsilon_t$ , so that

$\gamma \neq 0$  means that  $\sigma_t^2$  becomes a function of  $\epsilon_{t-j}$  and so it yields the requisite odd function.<sup>29</sup>

A variant of the SV model that has been extensively used by Hamilton in a variety of applications, e.g., to exchange rates – Engel and Hamilton (1990) – and interest rates – Hamilton (1988) – is to relate  $\sigma_t^2$  to an unobserved state variable  $z_t$  that can take either the value of zero or unity, with this variable evolving according to the first-order Markov process

$$\begin{aligned} \text{Prob} [ z_t = 1 | z_{t-1} = 1 ] &= p \\ \text{Prob} [ z_t = 0 | z_{t-1} = 1 ] &= 1 - p \\ \text{Prob} [ z_t = 0 | z_{t-1} = 0 ] &= q \\ \text{Prob} [ z_t = 1 | z_{t-1} = 0 ] &= 1 - q \end{aligned} \quad (19)$$

It can be shown by substitution that this scheme implies that  $z_t$  is an AR(1) process

$$z_t = (1 - q) + p z_{t-1} + \eta_t, \quad (20)$$

where  $\rho = p + q - 1$  and, conditional upon  $z_{t-1} = 1$ ,

$$\begin{aligned} \eta_t &= (1 - p) && \text{with probability } p \\ &= -p && \text{with probability } (1 - p), \end{aligned} \quad (21)$$

while, conditional upon  $z_{t-1} = 0$ ,

$$\begin{aligned} \eta_t &= -(1 - q) && \text{with probability } q \\ &= q && \text{with probability } (1 - q). \end{aligned} \quad (22)$$

Defining  $\sigma_t^2 = \phi_1 + \phi_2 z_t$  we may write

$$\sigma_t^2 = \rho \sigma_{t-1}^2 + (1 - \rho) \phi_1 + (1 - q) \phi_2 + \phi_2 \eta_t, \quad (23)$$

and, from (21) and (22),

$$\begin{aligned} E(\eta_t^2 | z_{t-1}) &= q(1 - q) + \delta z_{t-1} \\ &= q(1 - q) - \delta \phi_1 \phi_2^{-1} + \phi_2^{-1} \delta \sigma_{t-1}^2, \end{aligned}$$

where  $\delta = p(1 - p) - q(1 - q)$ , making this model differ from a standard SV model in two ways: the innovations driving the conditional variance are a discrete random variable with four states and the conditional variance of the innovations depends upon  $\sigma_{t-1}^2$ . The ability to replicate standard characteristics is similar to the standard SV model. Most importantly, persistence in the autocorrelation

---

<sup>29</sup> But at the expense of inducing a non-zero mean into  $x_t$  since  $E_{t-1}(x_t) = E_{t-1}(\sigma_t \epsilon_t) = E_{t-1}(\sigma_t \eta_t \epsilon_t) \neq 0$ .

function of the squares is governed by the size of  $p$  and, and to ensure that it is close to unity, one needs to have both  $p$  and  $q$  close to one, i.e., once a state is entered into there is little tendency to leave it.

Hamilton's model has been applied to a variety of financial time series – to interest rates in Hamilton (1988) and to stock returns in Pagan and Schwert (1990a) and Sola and Timmermann (1994). The first of these is a particularly interesting application; there turn out to be two regimes, corresponding to the 1979–1982 years and the remainder of the sample, respectively. The three year period 1979–1982 coincides with the change in the Fed's operating procedures which led to very high volatility in interest rates. The other two papers allow for both shifts in the mean and variance of returns.

Instead of viewing the ARCH and latent state models as separate one might attempt to link them together. A motivation for doing this comes from the high degree of estimated persistence in volatility observed after fitting either GARCH or SV models, a concomitant of the behavior of the acf of squares. Another explanation for this phenomenon might be that there have been shifts in the parameters of the underlying processes. It is well known that shifts in the mean of a random variable will tend to show up as a unit root when an autoregressive process is fitted, and one might extend this argument to  $x_t^2$  as well. Diebold (1986) and Lamoureux and Lastrapes (1990) maintained this interpretation and echoes of the theme may be found in Ding et al. (1993) and Diebold and Lopez (1995). As mentioned above, Hamilton's model can be used to allow the intercept in the variance equation to shift, but one might expect some extra autoregressive elements in it, even after allowing for that effect. Accordingly, Hamilton and Susmel (1994) and Cai (1994) consider switching ARCH (not GARCH) models in which the conditional variance is selected from a number of possible ARCH processes depending upon the state that eventuates. Such SWARCH models have been applied to stock returns by Hamilton and Susmel (1994) and to Treasury Bill yields by Cai (1994).

### *3.2. Estimating models of volatility*

Five methods of estimation for the above models figure prominently in the financial econometrics literature.

1. Generalized method of moments.
2. Maximum likelihood
3. Quasi maximum likelihood
4. Indirect estimation.
5. Bayesian estimation.

Closely associated with the decline in computational costs has been a shift towards the use of the final two methods, and it seems likely that this will continue, although computer packages may well continue to feature the first three.

### 3.2.1. GMM estimation

GMM estimation of ARCH type models has been implemented by Glosten et al. (1993), Rich et al. (1991) among others. To find GMM estimates one must have orthogonality conditions and Rich et al. use  $E\{z_t[x_t^2 - \sigma_t^2]\} = 0$ , where  $z_t$  are elements of  $F_{t-1}$ , e.g.,  $x_{t-1}$ ,  $x_{t-1}^2$ . However, there seems little advantage to using this estimator rather than the MLE. SV models were first estimated by method of moments – see Taylor (1986), Melino and Turnbull (1990) and Duffie and Singleton (1993) (the latter simulate the moments rather than deriving them analytically) – but after these first studies there has been declining interest in this way of doing estimation. The moments used typically involved the first four moments of  $x_t$  as well as the a.c.f.'s of  $x_t^2$  and  $|x_t|$ . In a recent Monte Carlo study Andersen and Sørensen (1994a) conclude that the results are sensitive to the number of moments used; if too many moments are used the sampling distributions of the estimators are adversely affected, even with quite large numbers of observations.

Under certain circumstances difficulties can be experienced with the GMM estimator. To appreciate these suppose that there is a single parameter  $\theta$  to be estimated and the moment condition used for this is  $E(z_t v_t(x_t; \theta)) = 0$ , where  $z_t$  is an instrument. In the case of an ARCH(1) model one might set  $z_t = x_{t-1}^2$  and  $v_t = x_t^2 - \alpha_0 - \alpha_1 x_{t-1}^2$ , and assume that only one of  $\alpha_0$  and  $\alpha_1$  is unknown. The GMM estimator of  $\theta$  solves  $\sum z_t v_t(x_t; \hat{\theta}) = 0$  and  $\hat{\theta} - \theta_0 \approx -(\sum z_t v_{\theta_0 t})^{-1} (\sum z_t v_t)$ , where  $\approx$  means that only the linear terms from the expansion of  $v_t(x_t; \hat{\theta})$  around  $\theta_0$  are retained, as is common in asymptotic analysis, and  $v_{\theta t} = \partial v_t / \partial \theta$  at  $\theta = \theta_0$ .<sup>30</sup> What makes for complications with the GMM estimator is clear from this expression: it is the ratio of two random variables that are generally not independent. In standard analysis  $T^{1/2}(\hat{\theta} - \theta_0)$  is asymptotically normal since  $T^{-1} \sum z_t v_{\theta t} \rightarrow^P E(z_t v_{\theta_0 t})$ , which is a constant, and so the distribution is determined by the distribution of the numerator,  $T^{1/2} \sum z_t v_t$ . But what if  $E(z_t v_{\theta_0 t}) = 0$ ? Then we would have  $\hat{\theta} - \theta_0 \approx -(T^{-1/2} \sum z_t v_{\theta_0 t})^{-1} (T^{-1/2} \sum z_t v_t)$  and both numerator and denominator would converge in distribution to normally distributed random variables with zero means. If the numerator and denominator were independent, this distribution would actually be Cauchy. Hence, the case when  $E(z_t v_{\theta_0 t}) = 0$ , which corresponds to the situation where the instrument has a zero correlation with  $v_{\theta t}$ , would be expected to produce a GMM estimator whose distribution departs substantially from that predicted by asymptotic theory. Of course, this polar case is too extreme, even if it is suggestive. Accordingly, Staiger and Stock (1993) have performed an analysis when  $E(z_t v_{\theta t})$  is “local to zero” of the form  $\delta T^{-1/2}$ . This allows  $T^{1/2}(\hat{\theta} - \theta_0)$  to have a limiting distribution determined by the ratio of  $N(0, \text{var}(z_t v_t))$  to  $N(\delta, \text{var}(z_t v_{\theta t}))$  random variables, which can also

<sup>30</sup> If  $v_t$  was linear in  $\theta$ , say  $v_t = x_{2t} - x_{1t}\theta$ , then  $v_{\theta t} = -x_{1t}$  and the GMM estimator would be the simple instrumental variables estimator.

depart substantially from a normal distribution unless  $\delta$  is large relative to  $\text{std}(z, v_{\theta_t})$ . Simulation studies by Fuhrer et al. (1995), Kocherlakota (1990), Pagan and Jung (1993), Mao (1990) and Staiger and Stock (1993) have all found that the asymptotic approximation can be very poor when instruments are “weak”.

All of this suggests that great care needs to be exercised when applying GMM estimators. Some guidelines for detecting cases in which there are likely to be problems are available, see Staiger and Stock (1993) and Pagan and Jung (1993), of which the most useful involves an examination of the “concentration ratio”, a quantity related to the expected value of the denominator, and which can be computed from the  $R^2$  of the regression of  $v_{\theta_t}$  against the instruments  $z_t$  if  $v_{\theta_t}$  was linear in  $\theta$ . What to do about the problem is a much harder question. One might consider performing an iterated bootstrap in which the parameter estimates found with the data are used to simulate realizations which can then be used to study the properties of the GMM estimator. However, this requires one to be able to simulate from a known model for  $x_t$  and that may not always be available. Fundamentally, one would like to make the denominator non-stochastic, i.e., to replace  $\sum z_t v_{\theta_t}$  by  $TE(z_t v_{\theta_t})$ . For some applications of GMM which match moments, e.g., in SV models, this expectation can be found either analytically or numerically (since  $z_t$  is actually unity). What makes it difficult to follow this strategy in many instances however is that the number of moments available,  $E(\phi(x_t; \theta)) = 0$ , exceeds the number of parameters in  $\theta$ . To use all available information, the larger number of moments are reduced to the correct number  $E(\zeta_t(x_t; \theta)) = 0$  by forming  $\zeta_t = W\phi_t$ , where  $W$  is a  $(\dim(\theta), \dim(\phi))$  matrix. Hansen (1982) showed that the best choice of  $W$  was  $W = [E(T^{-1} \sum \partial \phi_t / \partial \theta)] [T^{-1} \text{var}(\sum \phi_t)]^{-1}$ . It is clear that the resulting GMM estimator has the form  $\hat{\theta} - \theta_0 = -(\sum W \partial \phi_t / \partial \theta)^{-1} (\sum W \phi_t)$  and, even if  $\partial \phi_t / \partial \theta$  was non-stochastic, the denominator may become random if  $W$  has to be estimated from data.<sup>31</sup>

### 3.2.2. Maximum and quasi maximum likelihood

Estimation of the unknown parameters can be done by MLE using  $f(x_t | X_{t-1})$ , where  $X_{t-1} = \{x_0, x_1, \dots, x_{t-1}\}$ . To see how this item determines the log likelihood, write the joint density of returns as the product of a conditional and a marginal density,  $f(x_t, X_{t-1}) = f(x_t | X_{t-1})f(X_{t-1})$ . Building this up for all  $T$  observations, gives the joint density

$$f(x_1, \dots, x_T) = f(x_T | X_{T-1})f(x_{T-1} | X_{T-2}) \dots f(x_0),$$

---

<sup>31</sup> If the problem is such that artificial series can be generated allowing the estimation of  $W$  to any desired degree of accuracy, that approach might be used. In Monte Carlo studies such an option is available and Andersen and Sørensen (1994a) used the technique in their study of the GMM estimator of the SV model to show that its properties are greatly improved if the randomness due to  $W$  can be eliminated.

making the log likelihood of  $x_1, \dots, x_T$  equal to

$$L = \sum_{t=1}^T \log f(x_t | X_{t-1}) + \log f(x_0), \quad (24)$$

and this may be maximized with respect to the unknown parameters.

To estimate the unknown parameters of the ARCH process it is necessary to find  $f(x_t | X_{t-1})$ . But, if  $\epsilon_t$  is n.i.d. (0,1), it follows that the density of  $x_t$  conditional upon  $X_{t-1}$  is  $\mathcal{N}(0, \sigma_t^2)$  and (24) would be

$$L = \sum_{t=1}^T \left\{ -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma_t^2 - \frac{1}{2} \sigma_t^{-2} x_t^2 \right\} + \log f(x_0). \quad (25)$$

After replacing  $\sigma_t^2$  in (25) by any of the GARCH type specifications discussed earlier, one can proceed to maximize  $L$  with respect to the unknown parameters in  $\sigma_t^2$ . The main difficulty here is the final term in (25). But, because it is a single term, it is dominated (for large  $T$ ) by the summation, and so it has generally been neglected. In fact little is known about the density  $f(x_0)$  – if the  $x_t$  are assumed to be strictly stationary, then  $f(x_0)$  is the unconditional density for the series  $x_t$ . To date no analytic expression for this is available, but we can determine it experimentally once particular values of the GARCH parameters are specified.<sup>32</sup>

As mentioned earlier models have also been estimated when  $f_\epsilon$  is not normal. A useful way to characterize this search for flexibility in the density of  $\epsilon_t$  is to observe that the “density score”,  $d\log f(\epsilon) = f'(\epsilon)/f(\epsilon)$ , is  $-\epsilon$  for an  $N(0,1)$  density, i.e., it is linear in  $\epsilon$ . Different choices of the density for  $\epsilon_t$  amount to making the “density score” a non-linear function of  $\epsilon_t$ . To appreciate why this is of interest observe that the log likelihood for any ARCH model is

$$-\frac{1}{2} \sum_{t=1}^T \log \sigma_t^2 - \sum_{t=1}^T \log f(\epsilon_t), \quad (26)$$

and therefore the scores for  $\theta$  will be

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2} \sum_{t=1}^T \left( \frac{\partial \sigma_t^2}{\partial \theta} \right) \sigma_t^{-2} \left[ \left( \frac{f'(\epsilon_t)}{f(\epsilon_t)} \right) \epsilon_t + 1 \right]. \quad (27)$$

Notice that this involves the unknown density score so that the various parametric approaches can just be viewed as alternative ways of estimating this quantity.<sup>33</sup> Another approach, used by Engle and Gonzalez-Rivera (1991), is to

<sup>32</sup> Diebold and Schuermann (1996) have included  $\log f(x_0)$  in the maximization by forming non-parametric estimates of  $f(\cdot)$ . For samples of 100 the difference is minor.

<sup>33</sup> Nelson (1990b) argues that one needs to be cautious in following this strategy since there will be an interaction between the assumed density and the specification of  $\sigma_t^2$ . In particular, it is possible that the estimated  $\sigma_t^2$  might imply thinner tails than the true  $\sigma_t^2$  would, since some of the thickness will be absorbed by the assumed density. We have encountered this phenomenon earlier when setting out an expression for  $f(0)$ .

try to produce estimates of  $\theta$  that would be “robust” to the density function by replacing  $f'(\epsilon)/f(\epsilon)$  in (27) with a non-parametric estimator of that quantity, i.e., to find a *semi-parametric* (SP) estimator of  $\theta$ .

Within the SP literature a distinction is made between the parameters of interest  $\theta$  and a set of “nuisance” parameters  $\eta$  that may be thought of as indexing the possible set of densities,  $f(\theta_0, \eta)$ , with  $f(\theta_0, \eta_0)$  being the true density. Because  $\eta$  is infinite dimensional analysis of the situation is more complex than in parametric problems where  $\eta$  might (say) be the degrees of freedom parameter for the  $t$ -density. Nevertheless, it is the case that familiar propositions from the parametric literature tend to generalize to the semi-parametric case. In particular, it is well known that, to estimate  $\theta$  without knowledge of  $\eta$ , it is necessary that  $E[\partial \log f(\theta_0, \eta)/\partial \theta] = 0$  for all possible values of  $\eta$ , where the expectation is taken with respect to the *true* density. In the SP estimation context, if this condition holds  $\theta$  is said to be *adaptively estimable*. What are the prospects for this? It is clear from the definition of the score in (27) that adaption requires  $E[\sigma_i^{-2}(\partial \sigma_i^2/\partial \theta)] = 0$ . Taking the ARCH(1) model for illustrative purposes, and following the approach in Linton (1993), re-parameterize  $\sigma_i^2$  as  $\sigma_i^2 = e^\delta (1 + \omega(x_{i-1}^2 - a))$ , where  $e^\delta = \alpha_0 + \omega a$ ,  $\omega = e^{-\delta} \alpha_1$  and  $a = E[\sigma_i^{-2} x_{i-1}^2]/E[\sigma_i^{-2}]$ . It is then clear that

$$E[\sigma_i^{-2}(\partial \sigma_i^2/\partial \omega)] = E[\sigma_i^{-2}(x_{i-1}^2 - a)] = 0.$$

Hence, the above results point to the conclusion reached by Linton (1993) and Steigerwald (1992) that neither the intercept  $\alpha_0$  nor the slope  $\alpha_1$  are adaptively estimable. These results parallel what is known about the possibility of adaptively estimating the intercept and slope coefficients in a linear conditional expectation, see Pagan and Ullah (1995, Theorem 5.5). As discussed there it is also necessary that the covariance of the scores of  $\delta$  and  $\omega$  be zero, but this can be shown to be satisfied by the choice of  $a$  in the text. However, there is one important difference to the conditional mean context: rarely is there much interest in the intercept in a conditional mean, but most uses of the conditional variance require  $\sigma_i^2$  in its totality, i.e., the constant term as well as the ARCH part. Consequently, the quantity of interest,  $\sigma_i^2$ , cannot be adaptively estimated. Again this is a familiar result from the SP literature; it is impossible to adaptively estimate a scale parameter.

It is unwise to become too preoccupied with the possibilities of adaption. A more useful question is to ask whether a given SP estimator is “efficient”. To answer this requires some definition of efficiency. In the parametric case this is provided by the Cramer–Rao lower bound for  $\theta$  of  $(I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta})^{-1}$ , where  $I$  is the information matrix. There is an analogue to this in the SP case that is termed the SP efficiency bound. Steigerwald (1992) and Drost and Klassen (1994) derive this bound for GARCH models. Provided the density of  $\epsilon_i$  is symmetric the estimator proposed by Engle and Gonzalez-Rivera attains this bound.

Instead of trying to find the MLE when  $f_\epsilon$  is unknown, it has been suggested

that one proceed as if  $\epsilon_t$  was n.i.d. (0,1), i.e., maximize the function  $-(1/2)\sum \log \sigma_{t|t-1}^2 - (1/2)\sum \sigma_{t|t-1}^{-2} [x_t - E(x_t | X_{t-1})]^2$  where  $\sigma_{t|t-1}^2 = \text{var}(x_t | X_{t-1})$ . The resulting *quasi MLE* of  $\theta, \hat{\theta}_{\text{QMLE}}$ , will be consistent and asymptotically normally distributed, provided the expected value of the scores under the normality assumption have zero expected value under the true density, i.e.,  $\frac{1}{2}E[\sum_{t=1}^T (\partial \sigma_t^2 / \partial \theta) \sigma_t^{-2} (\epsilon_t^2 - 1)] = 0$ . Clearly, this holds for any density. The major difficulty with the quasi MLE is that, whilst  $\hat{\theta}$  is consistent, one has to be careful about the construction of the estimator of the covariance matrix. This should be  $\text{var}(\hat{\theta}) = H^{-1}VH^{-1}$ , where  $H = -E(\partial^2 L / \partial \theta \partial \theta')$  and  $V = \text{var}(\partial L / \partial \theta)$ , which differs from the value,  $H^{-1}$ , that would obtain if the assumed density for  $\epsilon_t$  coincides with the true one. One can replace  $V$  with the “outer product” estimate  $\sum_{t=1}^T (\partial L_t / \partial \theta)(\partial L_t / \partial \theta)' = \frac{1}{4}\sum \sigma_t^{-4} \left( \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta} \right) (\epsilon_t^2 - 1)^2$  so that, in large samples,  $V = \frac{1}{2}H[E(\epsilon_t^2 - 1)^2]$ , showing that the true covariance matrix is a multiple of the “formula” variance,  $H^{-1}$ , determined by the fourth moment of  $\epsilon_t$ . The quasi-MLE standard errors have become widely used. In the discussion above it is assumed that the density of the innovations generating the QMLE is normal, but the widespread use of the *t*-density in estimation prompts the question of the properties of QMLE’s based on other densities than the normal. Newey and Steigerwald (1996) deal with this question showing that, if both assumed and true densities are symmetric around zero, the parameters of the conditional mean are consistently estimated, while those in the conditional variance are consistently estimated only up to a scale factor. They also discuss how to modify the QMLE if symmetry does not hold.

Estimation of  $f(x_t | X_{t-1})$  for SV models is very demanding computationally. To see why, observe that, since the SV model has  $\sigma_t^2$  following a Markovian process,

$$f(x_t | X_{t-1}) = f_{\sigma_t} f_{\sigma_{t-1}} f(\sigma_t, \sigma_{t-1}, x_t | X_{t-1}), \quad (28)$$

and it becomes necessary to perform numerical integration at each observation in order to evaluate the likelihood. In the case of Hamilton’s specification the burden is lessened by the fact that  $\sigma_t^2$  is discrete, allowing the bivariate integration to be carried out by a dual sum over the four possible state values. Of course, this integration first requires that the density  $f(\sigma_t, \sigma_{t-1}, x_t | X_{t-1})$  be found, but this can be done recursively from the following relations between conditional and unconditional densities:

$$f(\sigma_t, \sigma_{t-1}, x_t | X_{t-1}) = f(x_t | \sigma_t, \sigma_{t-1}, X_{t-1}) f(\sigma_t, \sigma_{t-1} | X_{t-1}). \quad (29)$$

$$f(\sigma_t, \sigma_{t-1} | X_{t-1}) = f(\sigma_t | \sigma_{t-1}, X_{t-1}) f(\sigma_{t-1}, \sigma_{t-2} | X_{t-1}). \quad (30)$$

$$f(\sigma_t, \sigma_{t-1} | X_t) = f(\sigma_t, \sigma_{t-1}, x_t | X_{t-1}) / f(x_t | X_{t-1}). \quad (31)$$

Beginning with some initial density  $f(\sigma_1, \sigma_0 | x_1)$  we can apply (30), (29), (28) and

(31) in that order to find  $f(\sigma_2, \sigma_1 | X_2)$ , whereupon the process is repeated.<sup>34</sup> Since  $f(x_t | \sigma_t, \sigma_{t-1}, X_{t-1})$  and  $f(\sigma_t | \sigma_{t-1}, X_{t-1})$  derive from the model the circle is closed.

The problems of having to perform numerical integration at each observation has dampened the enthusiasm of investigators for exact MLE (except in the case of Hamilton's model). Some progress has been made in designing efficient ways of performing the numerical integrations, of which the best method seems to be the accelerated Gaussian importance sampler of Danielsson (1994) and Danielsson and Richard (1993), and more can be expected. Until recently therefore, the major way of estimating the SV model was by QMLE. To implement this method, square  $x_t$  to get  $\sigma_t^2 \epsilon_t^2$ , and then take logs, thereby summarizing the SV model (18) in the state space form,

$$\begin{aligned}\log x_t^2 &= k + \log \sigma_t^2 + \zeta_t \\ \log \sigma_t^2 &= \alpha_0 + \beta_1 \log \sigma_{t-1}^2 + \sigma_\eta \eta_t,\end{aligned}$$

where  $k = E(\log \epsilon_t^2)$  and  $\zeta_t = \log \epsilon_t^2 - k$ . If  $\zeta_t$  was normally distributed, and  $\epsilon_t$  was independent of  $\eta_t$ , then  $f(x_t | X_{t-1})$  could be found directly with the Kalman filter.<sup>35</sup> We can ignore the fact that  $\zeta_t$  is not normal and proceed as if it was so, i.e., use a quasi MLE, and this approach was suggested by Nelson (1988) and implemented by Harvey and Shephard (1993) and Harvey et al. (1994).<sup>36</sup> Its major defect is that it can be very inefficient, since the distribution of  $\zeta_t$  lies far from a normal density, having a very long left hand tail. Computationally, problems can arise when  $\epsilon_t$  becomes small, as this makes  $\log \epsilon_t^2$  very large, and this "inlier" problem was dealt with in Harvey and Shephard by effectively trimming out these observations.

Drawing on ideas in Shephard (1994), Kim and Shephard (1994) show how the formulation above may be adapted to perform exact maximum likelihood. The idea is to replace the  $\zeta_t$  with a mixture of  $m N(\mu_i, v_i)$ ,  $i = 1, \dots, m$ , random variables whose probability of occurrence is  $\pi_i$ . Stochastically this can be thought of as making  $\zeta_t$  a function of a discrete random variable  $\omega_t$  taking values  $i = 1, \dots, m$ , so that  $f(\zeta_t) = \sum_{i=1}^m \pi_i f(\zeta_t | \omega_i) = \sum_{i=1}^m \pi_i N(\mu_i, v_i)$ , e.g., if  $\omega_t = 2$  is realized then we would draw  $\zeta_t$  from an  $N(\mu_2, \sigma_2)$  random variable. As Kim and Shephard point out the weights  $\pi_i$ ,  $\mu_i$  and  $v_i$  can be determined once  $m$  is

<sup>34</sup> One possibility is to set  $f(\sigma_1, \sigma_0 | x_1)$  to the unconditional density  $f(\sigma_0, \sigma_1)$ ; another is to estimate it. Hamilton (1989) uses the former whereas Hamilton (1990) proceeds under the latter assumption.

<sup>35</sup> In the case of Hamilton's model the error term in the state dynamics equation has a time varying conditional variance that depends on unobserved quantities, as  $E(\eta_t^2 | z_{t-1})$  depends on  $\sigma_{t-1}^2$ . While the Kalman filter allows for the conditional variances of the errors to vary in a known way with the past history of  $x$ , it does not allow them to depend on the past unobserved states.

<sup>36</sup> Ruiz (1994) suggests that the QMLE is more efficient than the GMM estimator of the parameters of SV models but Andersen and Sørensen (1994b) demonstrate that this conclusion is incorrect.

selected by finding what parameter values would make  $\sum_{i=1}^m \pi_i N(\mu_i, v_i)$  as close a fit to the density of the log of a  $\chi^2(1)$  as possible, and these weights are given in their paper for  $m = 7$ .<sup>37</sup> What makes this idea ingenious is that, *conditional upon a realization*  $\tilde{\omega} = (\omega_1, \dots, \omega_T)$ , the random variable in the observation equation is  $\zeta_t | \tilde{\omega}$ , and this will be normal. Hence the Kalman filter can be applied. Of course, this immediately raises the issue of where  $\tilde{\omega}$  comes from. Defining  $\tilde{\sigma}^2 = (\sigma_1^2, \dots, \sigma_T^2)$  we could obtain a set of realizations on  $\tilde{\omega}$  if we could simulate from  $f(\tilde{\sigma}^2, \tilde{\omega} | X_T)$ , but that can be done with the Gibbs sampler, for which we will need  $f(\tilde{\sigma}^2 | \tilde{\omega}, X_T)$  and  $\Pr(\tilde{\omega} | \tilde{\sigma}^2, X_T)$ . The first of these, sampling from the Kalman smoother, is accomplished by the algorithm in De Jong and Shephard (1993), leaving the task of getting realizations from  $\Pr(\tilde{\omega} | \tilde{\sigma}^2, X_T) = \prod_{t=1}^T \Pr(\omega_t | \sigma_t^2, \log x_t^2)$ , and this can be done using Bayes theorem, culminating in the requisite density as being proportional to  $\prod_{t=1}^T f(\log x_t^2 | \omega_t, \sigma_t^2) \Pr(\omega_t)$ .<sup>38</sup> Kim and Shephard report that the method works very quickly.<sup>39</sup>

### 3.2.3. Indirect estimation

A method of estimation whose popularity has grown in near exponential fashion since its introduction in 1992 has been that of indirect estimation. To understand the nature of this method it is useful to consider the simple example of estimating the MA(1) coefficient  $\alpha$

$$y_t = e_t + \alpha e_{t-1}, \quad (32)$$

where  $e_t$  is n.i.d. (0,1), by using the AR(1)

$$y_t = \beta y_{t-1} + u_t.$$

We assume that  $u_t$  is n.i.d. (0,  $\sigma^2$ ) and write down the sample scores for  $\beta$  as

$$T^{-1} \sigma^{-2} \sum y_{t-1} (y_t - \beta y_{t-1}).$$

Equating these to zero gives the OLS estimator  $\hat{\beta}$ . Since this model is mis-specified it is known from the theory of estimation in mis-specified models that the estimator  $\hat{\beta}$  converges to the pseudo-true value  $\beta^*$  defined by

$$E_{MA} [ y_{t-1} (y_t - \beta^* y_{t-1}) ] = 0, \quad (33)$$

<sup>37</sup> It is not necessary that  $e_t$  be taken as an  $N(0,1)$  random variable as one can approximate the distribution of  $e_t^2 - E(e_t^2)$  by a mixture of normals if  $e_t$  follows other densities. Mahieu and Schotman (1994a), in estimating an SV model, allow  $\mu_i$  and  $v_i$  to be free parameters.

<sup>38</sup> The constant of proportionality is found by summing  $f(\log x_t^2 | \omega_t = i, \sigma_t^2) \Pr(\omega_t = i)$  over all  $i = 1, \dots, m$ .

<sup>39</sup> Both the quasi and exact MLE's use the state space format for squared returns but this will only appropriate when  $e_t$  and  $\eta_t$  are independent, and it has already been argued that the cross correlations between  $x_t^2$  and  $x_{t-j}^2$  make this assumption suspect. When one allows for the correlation it will be necessary to estimate that parameter, but there is no information in  $x_t^2$  to do so. Harvey and Shephard suggest that one use both  $x_t^2$  and the  $\text{sgn}(x_t)$  for estimation.

where  $E_{MA}$  means that the expectation is taken with respect to the true model (the DGP). In this instance it is easy to solve for  $\beta^*$ , producing

$$\beta^* = E_{MA}(y_t y_{t-1}) / E_{MA}(y_{t-1}^2) = \alpha / (1 + \alpha^2),$$

and this shows the dependence of  $\beta^*$  upon  $\alpha$ . As the value of  $\alpha$  varies so will  $\beta^*$ , and we might write this dependence as  $\beta = \phi(\alpha)$ , a quantity that has been termed the “binding function”. This further suggests that we might estimate  $\alpha$  by using such a relation, i.e., we could solve

$$T^{-1} \sum_{t=1}^T y_{t-1} \left( y_t - \frac{\hat{\alpha}}{1 + \hat{\alpha}^2} y_{t-1} \right) = 0,$$

to get an estimate of  $\alpha$ . Standard properties for  $\hat{\alpha}$  of consistency and asymptotic normality follow from  $\hat{\beta} = \phi(\hat{\alpha})$  and the fact that  $\hat{\beta}$  is consistent and asymptotically normal around  $\beta^*$ . In this instance we know that  $\hat{\alpha}$  has these properties from the times series literature, as factorizing the autocovariance function was one of the earliest ways of estimating an MA(1) – see Durbin (1959).

In this simple example we could find the binding function analytically, whereas in most instances this will be hard to do. However, all we need to be able to do is find the value of  $\beta^*, \tilde{\beta}^*(\bar{\alpha})$ , that is associated with any given value  $\bar{\alpha}$  through the binding function, and to then choose a value of  $\alpha$  that implies a  $\beta^*$  that equals  $\hat{\beta}$ , i.e., we wish to choose a value of  $\alpha, \hat{\alpha}$ , such that  $\hat{\beta} = \hat{\alpha} / (1 + \hat{\alpha}^2)$ . Now, because  $\beta^*$  is defined from (33), it is clear that the value of  $\beta, \tilde{\beta}^*$ , given by  $1/M \sum_{m=1}^M 1/T \sum_{t=1}^T \bar{y}_{mt-1} (\bar{y}_{mt} - \tilde{\beta}^* \bar{y}_{mt-1}) = 0$ , where the  $\bar{y}_{mt}$  are values simulated from the MA(1) model at the  $m$ th replication, with  $\alpha$  set to  $\bar{\alpha}$ ,  $m = 1, \dots, M$ , will converge to  $\tilde{\beta}^*(\bar{\alpha})$  as  $M \rightarrow \infty$ . Thus we might  $\min_{\alpha} (\hat{\beta} - \tilde{\beta}^*)^2$ ; the value of  $\tilde{\beta}^*$  changing as new values for  $\alpha$  are tried. At the minimum value we have the “indirect estimator”,  $\hat{\alpha}$ . It will not exactly equal the  $\hat{\alpha}$  found from  $\hat{\beta} = \phi(\hat{\alpha})$ , unless  $M$  was infinite, but we might expect it to be close. If there are more elements in  $\beta$  than in  $\alpha$  the distance function between  $\hat{\beta}$  and  $\tilde{\beta}^*$  needs to be modified to  $(\hat{\beta} - \tilde{\beta}^*)' V (\hat{\beta} - \tilde{\beta}^*)$ , where  $V$  is some weighting matrix. This describes the original proposal by Gouriéroux et al. (1993). A related method of indirect estimation is due to Gallant and Tauchen (1992), who would find the value of  $\alpha$  that minimizes  $[1/M \sum_{m=1}^M 1/T \sum_{t=1}^T \bar{y}_{mt-1} (\bar{y}_{mt} - \hat{\beta} \bar{y}_{mt-1})]^2$ ; this works since, as  $M \rightarrow \infty$ , the function tends to  $[(\bar{\alpha} - \beta(1 + \bar{\alpha}^2))]^2$ , which is minimized by setting  $\hat{\alpha} = \hat{\beta}(1 + \hat{\alpha}^2)$ .<sup>40</sup> Therefore, computer simulation of the model we are ultimately interested in estimating, along with actual estimation of an entirely different model, will produce consistent and asymptotically normal estimators of the parameters of interest.

<sup>40</sup> Although Gallant et al. (1994) refer to their method as “efficient method of moments” we feel that the descriptor used by Gouriéroux et al. is sufficiently evocative to justify calling all methods operating with this principle by that name.

The ideas described above generalize. Think of the model to be estimated as having some density for the random variable  $y_t$ ,  $g(y_t; \alpha)$ , that is characterized by unknown parameters  $\alpha$ , while the *auxiliary model* used for estimation purposes satisfies a set of moment conditions  $1/T \sum_{t=1}^T \xi(y_t; \hat{\beta}) = 0$ . Then the pseudo-true value of  $\beta$  solves  $E_g(1/T \sum_{t=1}^T \xi(y_t; \beta^*)) = 0$ , and the estimator of  $\alpha$  that we seek is what solves this set of equations. To make this operational, we replace  $E_g$  by an average using observations simulated from the model with a given value of  $\alpha$ , and replace  $\beta^*$  with a consistent estimator of it,  $\hat{\beta}$ , leading to the indirect estimates of  $\alpha$  being determined by

$$\min_{\alpha} \left[ 1/M \sum_{m=1}^M 1/T \sum_{t=1}^T \xi(\bar{y}_{mt}; \hat{\beta}) \right]' V \left[ 1/M \sum_{m=1}^M 1/T \sum_{t=1}^T \xi(\bar{y}_{mt}; \hat{\beta}) \right].$$

In this formula  $\bar{y}_{mt}$  are simulated values from the model described by  $g(y_t; \alpha)$  for given values of  $\alpha$ , and  $V$  is some weighting matrix. As this can be thought of as a GMM estimator, the literature on ways of selecting  $V$  is relevant; for a good review of these see Den Haan and Levin (1994).

Obviously, the success of the method depends on a number of factors. First, there must be some connection between  $\beta^*$  and  $\alpha$ , i.e., varying values for  $\alpha$  must show up in terms of changed values of  $\beta^*$ . Consequently, there would be no sense in trying to estimate parameters relating to a conditional variance of  $y_t$  from the conditional mean, unless these were connected in some way. Second, the choice of the auxiliary model  $\xi(y_t; \beta)$  (or score generator in Gallant and Tauchen's terminology, since they normally select  $\xi(y_t; \beta)$  as the scores of some model) is important for the efficiency of the indirect estimator of  $\alpha$ . In terms of our simple example, the estimator of the MA(1) coefficient using an AR(1) as the auxiliary model is known to be very inefficient. Using a higher order AR, e.g., an AR(3), should improve the properties of the estimator, a result established many years ago in time series analysis. Clearly, what we are seeking in  $\xi(y_t; \beta)$  should be the best statistical representation of the data that is possible. It is important to stress in this connection that the auxiliary model may be a mis-specified one.

Although it is computationally very demanding, the procedure is extremely attractive, particularly if it is hard to find MLE's of  $\alpha$  but it is possible to find some representation of the data that may be easy to estimate. A good example is the stochastic volatility model. It is simple to generate data from the model once parameter values are selected, and it is known that GARCH models can be found that give reasonable fits to the data, making it natural to select these as candidates for an auxiliary model. Engle and Lee (1994), in estimating an SV model with non-zero correlation between  $\epsilon_t$  and  $\eta_t$ , found indirect estimates using a GARCH(1,1) model with an asymmetric effect of returns on volatility as the auxiliary model. One might think that the scores of an EGARCH model would be a better candidate for  $\xi(y_t; \beta)$  in this instance, since it has a similar functional form to the SV model. In perhaps the most comprehensive study of the indirect estimation of the SV model, Gallant et al. choose their SNP method of approxi-

mating densities to derive  $\xi(\cdot)$ . Because  $\dim(\beta)$  therefore greatly exceeds  $\dim(\alpha)$  it is possible for them to assess the adequacy of the SV model as a representation of the daily stock data used in this survey by testing how close the scores  $\sum_{t=1}^T \xi(y_t; \beta^*(\hat{\alpha}))$  are to zero. They find that the SV model is soundly rejected unless the density of  $\eta_t$  is allowed to take a very general form and there is also a very flexible specification of the way in which  $\sigma_t^2$  maps into  $\eta_t$ .

### 3.2.4. Bayesian estimation methods

Bayesian methods of estimation are becoming more common in financial econometrics. One reason is that the parameters  $\theta$  are regarded as random variables. Therefore, they can be treated as extending the set of latent variables  $\sigma_t$ , and posterior densities for both will be simultaneously obtained. Some see this as a major advantage, i.e., instead of first determining  $\theta$ , and then using this to construct a point estimate of  $\sigma_t$ , the complete density of the latter is found. As is well known it is easy to write down the posterior for  $\theta$ , given the data, in terms of the product of the likelihood and the prior density, but rather harder to give this an analytic expression. Fortunately, methods have become available that enable simulation of realizations from the posterior without explicitly knowing its form. Generally, these methods fall under the heading of Markov Chain Monte Carlo methods, of which leading examples are the Gibbs sampler and Metropolis–Hastings algorithm. Surveys of this technology are available in Chib and Greenberg (1994a) and Chib and Greenberg (1994b). Jacquier et al. (1994) have estimated the SV model with a variant of the Metropolis algorithm, while Geweke (1994) has estimated both GARCH and SV models with the same approach. Geweke also shows how the prediction error decomposition (24) for the log likelihood may be used to easily compute posterior odds recursively, and, based on an examination of these, concludes that the SV model is superior to a GARCH model for an exchange rate data set. Albert and Chib (1993) estimate Hamilton's model by these Bayesian methods.

### 3.3. Parametric models and moment existence

Estimation experience with GARCH models is now quite extensive. One of the most striking features has been the fact, documented in Bollerslev et al. (1992), that the model parameters  $\beta_j, \alpha_j$  frequently sum to unity, i.e., for a GARCH(1,1) process  $\beta_1 + \alpha_1 = 1$ . For EGARCH models the sum of the autoregressive coefficients on lagged values of  $\log \sigma_t^2$  can also be close to unity, e.g., Nelson (1989), Pagan and Schwert (1990a) and Kearns and Pagan (1993). One must therefore enquire into the consequences of this feature for estimation.

Let us begin with an EGARCH process that has the autoregressive coefficients summing to unity. This means that  $\log \sigma_t^2$  has a unit root. Without an intercept therefore  $\log \sigma_t^2$  would evolve as an integrated process; with an intercept it would be integrated with drift. In the former case it is known that  $T^{-2} \sum_{t=1}^T \log \sigma_t^2$  would

tend in law to a random variable, so that the log likelihood for such a process must eventually tend to  $T^{-2} \sum_{t=1}^T \log \sigma_t^2$ , since the other term  $T^{-2} \sum \sigma_t^{-2} x_t^2 = T^{-2} \sum \epsilon_t^2 = T^{-1} (T^{-1} \sum \epsilon_t^2) \xrightarrow{P} 0$  as  $T^{-1} \sum \epsilon_t^2 \rightarrow 1$ . Thus the estimation problem is not well defined and we could not expect that MLE's obtained by maximizing (25) would have standard properties. If there is an intercept,  $T^{-2} \sum \log \sigma_t^2$  would tend in probability to a constant that would not depend on the EGARCH parameters, other than the intercept, and again this would create estimation problems. Hence, it appears that an EGARCH process with unit roots would have to be avoided. Therefore the finding of near unit root behavior has to be very disturbing.

Turning to the GARCH(1,1) process in which  $\alpha_1 + \beta_1 = 1$ , termed an Integrated GARCH(1,1), or IGARCH(1,1) process by Engle and Bollerslev (1986), the situation is much less clearer. Assume that  $\epsilon_t$  is i.i.d. with a symmetric unimodal density, monotonically increasing from  $-\infty$  to zero and monotonically decreasing from zero to infinite, with zero third moment and finite moments up to order six. Nelson (1990a) showed that, if  $E[\ln(\alpha_1 \epsilon_t^2 + \beta_1)] < 0$ ,

1. For the IGARCH(1,1) model, when the intercept in the equation for  $\sigma_t^2$  was zero,  $\sigma_t^2 \xrightarrow{a.s.} 0$ .
2. For the IGARCH(1,1) model, with a non-zero intercept in the equation for  $\sigma_t^2$ ,  $\sigma_t^2$  has a strictly stationary and ergodic limiting distribution.
3. For the IGARCH(1,1) model,  $E(\sigma_t^{2k}) < \infty$  if  $E[(\alpha_1 \epsilon_t^2 + \beta_1)^k] < 1$ .<sup>41</sup>

Now, from (3), putting  $k = 1$ , gives  $E(\alpha_1 \epsilon_t^2 + \beta_1) = \alpha_1 + \beta_1$ , and therefore when the process is IGARCH ( $\alpha_1 + \beta_1 = 1$ ),  $E(\sigma_t^2)$  does not exist (this is not surprising given the solution for  $E(\sigma_t^2)$  in the GARCH(1,1) case). Now  $E(\sigma_t) < \{E(\sigma_t^2)\}^{1/2}$  by Jensen's inequality, i.e.,  $E[(\alpha_1 \epsilon_t^2 + \beta_1)^{1/2}] < [E(\alpha_1 \epsilon_t^2 + \beta_1)]^{1/2} = 1$  when  $\alpha_1 + \beta_1 = 1$ , and by (3) this guarantees the existence of the fractional moment  $E(\sigma_t)$ . Consequently, when one looks at the log likelihood for an IGARCH process,  $T^{-1} \sum \log \sigma_t$  will converge to  $E[\log \sigma_t]$  by the ergodic theorem, while  $T^{-1} \sum \sigma_t^{-2} \epsilon_t^2 = T^{-1} \sum \epsilon_t^2$  tends to unity. Therefore, unlike the EGARCH process, the log likelihood of an IGARCH process is well behaved asymptotically, and it is not surprising that the MLE estimators of the IGARCH parameters will be asymptotically normal, converging at the standard rate of  $T^{1/2}$ . Lumsdaine (1991) and Lee and Hansen (1994) give formal proofs of this fact; the latter under the conditions just cited, the former under the assumption that moments of  $\epsilon_t$ , up to the 32nd order exists.

The IGARCH process is therefore one for which the variance of returns does not exist but for which returns are a strictly stationary process (the latter follows from  $x_t = \sigma_t \epsilon_t$ , and this is the product of two strictly stationary processes), i.e., the IGARCH(1,1) process is *strictly stationary* but not *covariance stationary*, whereas an IEGARCH(1,1) process would be neither covariance nor strictly

<sup>41</sup> Results for GARCH ( $p, q$ ) models are available in Bougerol and Picard (1992).

stationary. Accordingly, because an IGARCH process is strictly stationary, it is somewhat misleading to attach the label “integrated” to it. This is not to say that it does not possess some of the characteristics of an integrated series. Using  $x_t = \epsilon_t \sigma_t$ , it is possible to re-write (13) as

$$\sigma_t^2 = \alpha_0 + [\beta_1 + \alpha_1 E_{t-1}(\epsilon_{t-1}^2)]\sigma_{t-1}^2 + \alpha_1[\epsilon_{t-1}^2 - E_{t-1}(\epsilon_{t-1}^2)]\sigma_{t-1}^2,$$

whereupon  $E_{t-1}(\sigma_{t-1}^2) = \alpha_0 + (\beta_1 + \alpha_1)\sigma_{t-1}^2$  and  $E_{t-1}(\sigma_t^2) = \alpha_0 + \sigma_{t-1}^2$  if the process is IGARCH, and this would also be true of an  $I(1)$  series with drift.

Another important feature of Nelson’s results is that there must be an intercept in the variance equation, otherwise  $\sigma_t^2$  will be degenerate for large  $t$ . Thus removal of an “insignificant” intercept would be rather unwise if the IGARCH process was to be used for forecasting. It is also important to observe that one cannot “start up”  $\sigma_t^2$  with  $\alpha_0/(1 - \alpha_1 - \beta_1)$ , as in the GARCH(1,1) case, any longer. Nevertheless, setting the initial value  $\sigma_0^2$  to zero will mean that it is effectively incorporated into the intercept, so that the estimated intercept really has two components. It may well be that a small intercept is found if an IGARCH process is estimated with a GARCH program because the initial value will be set to larger and larger numbers as the  $\beta_1 + \alpha_1$  sum tends to unity.

### 3.4. Specification tests for ARCH models

A large range of models have been mentioned above as being useful for the modeling of financial series. Each model differs from others in being able to replicate some particular characteristic of financial data. Consequently, it seems clear that one would want to develop some methods for assessing *ex ante* which models would be most useful in capturing the behavior of any particular series, as well as determining *ex post* how successful they have been. This is the realm of specification tests.

There are two philosophies that might be adopted when addressing issues of the appropriateness of any given model. One possibility is that of matching “stylized facts”, i.e., a comparison is made of certain characteristics of the series being investigated with what is implied by the fitted model. At first glance this approach to model evaluation may not appear to be very effective, but there are many examples in which a simple stylized fact has proven remarkably effective in weeding out poor models. We have already encountered an example of this in the form of the asymmetric relation between stock volatility and the level of returns. As mentioned previously, this is a stylized fact that has proven to be very influential in delineating the class of models needed to fully describe stock returns. A second example is the “jump” behavior of recursive variances. Unless a model is capable of replicating this feature one would be concerned about adopting it. A third example would be the ability of models to replicate the strong peak seen in the density of returns. If U.S. daily stock return data from 1835 to 1987 is used, the estimates of  $f(0)$  are 0.63 (data); 0.46 (GARCH); 0.40 (EGARCH); and 0.45

(AGARCH); all of the models producing substantial understatements of the  $f(0)$  seen in the data.

A more conventional approach to the evaluation question is to postulate an alternative specification to the one being examined and to construct Lagrange Multiplier (LM) tests for whether the alternative model is to be preferred. Of course, such a test can also reveal inadequacies in the maintained model even if the alternative is not correct, i.e., the diagnostic test can have power against alternatives other than the one specified. Let  $\theta$  be the set of parameters in both the alternative and null models, where a sub-set of  $\theta$  being set to zero produces the latter. When  $\epsilon_t$  is  $N(0,1)$  it is easily seen that the scores for  $\theta$  are  $\frac{1}{2} \sum \sigma_t^{-2} \partial \sigma_t^2 / \partial \theta (\epsilon_t^2 - 1)$  (or  $\frac{1}{2} \sum \sigma_t^{-2} \partial \sigma_t^2 / \partial \theta [-(f'(\epsilon_t)/f(\epsilon_t))\epsilon_t - 1]$ , when  $\epsilon_t$  has density  $f(\epsilon_t)$ ). Hence the LM test that the maintained model is correct is just based on  $TR^2$  from the regression of  $(\hat{\epsilon}_t^2 - 1)$  upon  $\hat{\sigma}^{-2} \partial \hat{\sigma}_t^2 / \partial \theta$ , where the hats indicate that  $\theta$  is replaced by the MLE of  $\theta$  under the null hypothesis. In the special case when the null hypothesis is that there is no ARCH,  $\hat{\sigma}_t^2 = \hat{\sigma}^2$  and  $\hat{\epsilon} = x_t/\hat{\sigma}$ . Thus, if this is the null, and the alternative is an ARCH(1) process,  $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2$ , then  $(\hat{\epsilon}_t^2 - 1) = (x_t^2/\hat{\sigma}^2 - 1)$  is regressed against  $\hat{\sigma}^{-2} \partial \hat{\sigma}_t^2 / \partial \alpha_0 (= \hat{\sigma}^{-2})$  and  $\hat{\sigma}^{-2} \partial \hat{\sigma}_t^2 / \partial \alpha_1 (= \hat{\sigma}^{-2} x_{t-1}^2)$ , which is the same as the regression of  $x_t^2$  upon an intercept and  $x_{t-1}^2$ , i.e., the test used previously to assess dependence in returns.<sup>42</sup> If instead we are checking that the model is ARCH(2), when ARCH(1) is maintained, then  $\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_{t-1}^2 + \hat{\alpha}_2 x_{t-2}^2$ ,  $\partial \hat{\sigma}_t^2 / \partial \alpha_0 = 1$ ,  $\partial \hat{\sigma}_t^2 / \partial \alpha_1 = x_{t-1}^2$ ,  $\partial \hat{\sigma}_t^2 / \partial \alpha_2 = x_{t-2}^2$  and the regression is now  $\hat{\epsilon}_t^2 = \hat{\sigma}_t^{-2} x_t^2$  against unity,  $\hat{\epsilon}_{t-1}^2$  and  $\hat{\epsilon}_{t-2}^2$ .

The central problem therefore is to determine a suitable alternative model. Mostly, the proposals have related to different models for the conditional variance. Engle and Ng (1993) suggest that the extra terms be dummy variables of either an additive or a multiplicative nature, e.g., a dummy variable that takes the value unity only if  $x_{t-1}$  is negative would imply the possibility of an asymmetric response if we try to add the product of it with  $x_{t-1}$  to the existing model. Pagan and Schwert (1990a) argued for a non-parametric approach, approximating  $\sigma_t^2$  by a GARCH process along with Fourier terms. They then tested the significance of the Fourier terms. There are clearly many alternative ways of generating diagnostic tests, varying by what we choose to add.

Instead of trying different types of non-parametric approximations, an alternative is to derive tests based on a very general parametric model. If one derives an LM test for the presence of the  $j$ th lag in  $\sigma_t^2$  from the AGARCH model discussed earlier, then it involves the regression of  $x_t^2$  against unity,  $x_{t-j}^2$ ,  $|x_{t-j}|$ ,  $x_{t-j}$ , and  $x_t |x_{t-j}|$ , followed by a test that  $T$  times the  $R^2$  from this regression is zero. This result is easily established by looking at the derivatives of  $\sigma_t^2$  with respect to the

<sup>42</sup> Lee and King (1993) point out that, since  $\alpha_1 \geq 0$ , the LM test can be improved upon by taking account of this sign information.

unknown parameters of the AGARCH model. A different perspective on such test procedures is to be had by recognising that the LM test is testing if  $E[\sigma_t^{-2}(\partial\sigma_t^2/\partial\theta)(\epsilon_t^2 - 1)] = 0$ , and this moment restriction holds under the null hypothesis because  $(\epsilon_t^2 - 1)$  has a zero expectation with respect to any function of past dated random variables. One of these is that selected by the LM test viz.  $\sigma_t^{-2}(\partial\sigma_t^2/\partial\theta)$ . Pagan and Sabau (1992) proposed that other functions might be used, in particular those that were related to  $\sigma_t^2$ , since an estimate of this variable was available under the null hypothesis. They applied the test to a number of applications and found that the fitted models were rejected. Nelson (1989) also provides tests of this nature.

It should be noted that all of the tests described above require that  $\epsilon_t$  be  $N(0,1)$ . If the density  $f(\epsilon)$  has some other form, then  $\epsilon_t^2 - 1$  is replaced by  $\xi_t = -[f'(\epsilon)/f(\epsilon)]\epsilon_t - 1$ . However, since  $f(\epsilon)$  will typically involve some parameters that need to be estimated, e.g., the degrees of freedom parameter if  $f(\epsilon)$  is Student's t, it is not generally true that one can regress  $\xi_t$  against quantities such as  $\sigma_t^{-2}(\partial\sigma_t^2/\partial\theta)$  to produce the requisite test statistic, since the fact that the density parameters have been estimated means that some allowance needs to be made for this effect. To analyze this problem it is useful to note that all the tests discussed in this section are testing a moment condition of the form  $E[z_t\phi_t] = 0$ , where  $\phi_t$  is  $\epsilon_t^2 - 1$ , etc. and are therefore based upon testing whether  $\sum z_t\hat{\phi}_t$  is zero. In order that one can ignore the fact that  $\hat{\phi}_t$  involves estimated parameters, it is necessary that  $E[z_t(\partial\phi_t/\partial\theta)] = 0$ , and this will rarely be the case. To account for this dependence one needs to follow Newey (1985) and Tauchen (1985), regressing  $z_t\hat{\phi}_t$  against an intercept and the scores for  $\theta$  and testing if the intercept is zero.<sup>43</sup> The procedure is explained in Pagan and Vella (1989). Note that  $\phi_t$  may also depend on estimated parameters if the conditional mean depends on ARCH parameters, as occurs with the ARCH-M models discussed later, so that a similar adjustment needs to be made for tests in that context.

If one just considers whether the model might be GARCH or EGARCH, it is clear that the two formulations are not nested, owing to the use of the logarithmic transformation and the fact that  $|\epsilon_t| = |\sigma_t^{-1}x_t|$  and  $\epsilon_t = \sigma_t^{-1}x_t$  are the driving forces in the EGARCH model. Pagan and Schwert (1990a) compared the log likelihoods from both estimated models, and one could modify this criterion along the lines of Schwarz (1978) to reflect the extra parameters in EGARCH. Going beyond such simple comparisons, there are a number of ways that one might effect a non-nested comparison. The simplest is through "artificial nesting". To perform this, define  $\sigma_t^{2G}$  and  $\sigma_t^{2E}$  as the GARCH and EGARCH conditional variance specifications respectively, and set up the expanded model  $\sigma_t^2 = \sigma_t^{2G} + \delta\sigma_t^{2E}$ , leading to an LM test that  $\delta$  is zero (or unity). Testing if  $\delta$  is zero, i.e., testing for

<sup>43</sup> Typically  $z_t$  will involve estimated parameters so that the scores for these would also need to be added to the regression.

the superiority of the GARCH model, would involve taking  $T$  times the  $R^2$  from the regression of the standardized GARCH errors  $\hat{\epsilon}_t^2$  upon unity,  $\hat{\sigma}_t^{-2G} \partial \sigma_t^{2G} / \partial \theta$ , and  $\hat{\sigma}_t^{-2G} \hat{\sigma}_t^{2E}$ . To fully account for the non-nested character of the models is much more difficult, as it is hard to evaluate what the expectation of a quantity such as the likelihood for one model would be if the other is correct. One way of doing this is by computer simulation, just as for the choice of linear and log linear models in Pesaran and Pesaran (1993), i.e., one could estimate both models from the data, simulate  $x_t$  from (say) the estimated GARCH model, compute whatever criterion was being used to compare the models, and find its empirical distribution by performing enough replications.<sup>44</sup> Kearns (1993) reports an application of this idea.

Apart from the issues of appropriate specification of the conditional variance function there has also been some concern over the stability of the conditional variance coefficients. Lamoureux and Lastrapes (1990) argued that the fact that parameter estimates seemed to indicate an IGARCH process might simply be symptomatic of instability in GARCH parameters, and they split the sample to test this hypothesis. Lee and Hansen (1991) provide a test of stability of the GARCH coefficients using Nyblom's approach (Nyblom, 1989). This is essentially an LM test of the hypothesis that the coefficients follow a random walk versus the alternative that they are constant, and involves the cumulative sums of the scores with respect to the GARCH coefficients thought to vary. The statistic used is  $L^* = \sum_{t=1}^T (\Sigma_{j=1}^t \hat{d}_{\theta,j})' V^{-1} (\Sigma_{j=1}^t \hat{d}_{\theta,j})$ , where  $d_{\theta,t}$  are the scores for the GARCH parameters, and  $V$  is an estimate of the variance of  $d_{\theta,t}$ . Hansen (1990) gives tables for the distribution of  $L^*$  under the null hypothesis that the coefficients are constant. Hansen (1994) also applied the test to whether the degrees of freedom parameter of the  $t$ -density varies as a way of determining if an ARCD model is necessary. Chu (1995) also provides a test for the stability of GARCH parameters based upon the summation of the scores over sub-samples.

### 3.5. ARCH models and diffusion processes

GARCH models have been fitted to data with a wide range of frequencies, and it has been noticed that, while the GARCH effects seem to disappear the longer the sampling interval, they become more intense for very short intervals, in the sense that  $\hat{\beta}_1 + \hat{\alpha}_1 \rightarrow 1$ . Diebold (1988) showed that, if one aggregated GARCH processes over time, as the sampling interval became large  $y_t$  tended to resemble a normally distributed random variable – see also Diebold and Lopez (1995). Nelson (1990c) and Drost and Nijman (1993) represent the most comprehensive

---

<sup>44</sup> In an interesting paper West et al. (1993) compare different models for the conditional variance of exchange rate returns through the utility gained by an investor as the different models are used to determine portfolio allocations.

work on the general topic of the effects of changing the sampling interval upon the nature of models for conditional volatility. Drost and Nijman focus upon the ARMA A(1,1) process in (14) and study what happens to it as one aggregates over time. They distinguish between three types of GARCH processes, depending upon the nature of the errors  $\epsilon_t$ , and  $v_t$  – strong GARCH if  $\epsilon_t$  is i.i.d.; semi-strong GARCH if  $\epsilon_t$  is a martingale difference; and weak GARCH in which  $E(v_t | x_{t-j}^r) = 0$  for  $j \geq 1$ ,  $r = 0, 1, 2$ . Only weak GARCH aggregates to a process of the same type. In particular, if one tries to aggregate strong GARCH, the equivalent innovations to  $\epsilon_t$  in the aggregated form will not be i.i.d.<sup>45</sup>

Once one thinks of questions of aggregation of GARCH processes in terms of aggregation of an ARMA process it is clear why one gets some of the outcomes noted above. Consider an AR(1) for a stock variable  $y_t$  with  $y_t = \beta y_{t-1} + e_t$ , where  $e_t$  are uncorrelated. If this is observed at intervals  $mt$  ( $m \geq 1$  being an integer), then  $y_{mt} = \beta^m y_{m(t-1)} + \xi_{mt}$ , with  $\xi_{mt}$  being uncorrelated with  $y_{m(t-1)}, \dots$ , leading to the results that, as  $m \rightarrow \infty$ ,  $y_t$  becomes uncorrelated, while, as  $m \rightarrow 0$ ,  $\beta^m \rightarrow 1$  and so a unit root tends to appear in the process.

In the analysis above, the GARCH model is assumed to hold for some time interval and a study is made of what happens as the time intervals of observation expand or contract. Many models developed in the theoretical finance literature express the evolution of financial variables in continuous rather than discrete time and sometimes this is very important to the solution of those models, e.g., in the Black–Scholes options pricing model. There are also some conceptual advantages to working in continuous time. For example, information about the prospects of companies continues to arrive when trading in shares is suspended, and this has implications for the modeling of volatility on a daily basis. For example, one would expect share trading on a Monday to be different to a Tuesday, because a much longer interval of time has elapsed since the last trade, and hence there has been a longer interval for information to accumulate. Thus an investigation of the relationship between continuous and discrete time models seems important.

Most of the continuous time processes used in finance are diffusions with the format

$$dy_t = adt + b y_t dt + \sigma_t y_t^\gamma dW_{1t}, \quad (34)$$

$$d \log \sigma_t^2 = \alpha_0 dt + (\beta_1 - 1) \log \sigma_t^2 dt + c dW_{2t}, \quad (35)$$

where  $dW_{1t}$  and  $dW_{2t}$  are correlated Brownian motions. If the diffusion process had non-stochastic constant volatility, i.e., it was (34) only, it was possible to find a form for the density function that could be used to set up the likelihood and hence produce MLE's of the unknown parameters. Numerical integration is

---

<sup>45</sup> This suggests that it would be illegal to perform MLE on a series with the same assumptions about  $\epsilon_t$  for a number of different sampling intervals. However, based on simulation studies, Drost and Nijman conclude that "the asymptotic bias of the QMLE, if there is any, is small" (p. 922).

involved to get the density in most instances however. Moreover, it seems unlikely that one can assume that volatility is non-stochastic with financial data, so that other ways of estimating the parameters need to be explored. There are many ways to approximate stochastic differential equations – see Kloeden and Platten (1992) – of which the simplest is the Euler scheme, which discretizes the equation using an interval of length  $ht$  ( $h \leq 1$ ) to get

$$\Delta y_{th} = ah + bhy_{(t-1)h} + \sigma_{th} y_{th}^\gamma h^{\frac{1}{2}} \epsilon_t \quad (36)$$

$$\Delta \log \sigma_{th}^2 = \alpha_0 h + (\beta_1 - 1) h \log \sigma_{(t-1)h}^2 + ch^{\frac{1}{2}} \eta_t,$$

where  $\epsilon_t$  and  $\eta_t$  are n.i.d. (0,1) random variables.

As  $h \rightarrow 0$  this approximation should tend to the continuous time one. Of course, data on  $y_{th}$  is not observable unless  $h = 1$ , and setting  $h = 1$  produces a very coarse approximation that can result in substantial biases in the estimators of the parameters of the diffusion. However, as Duffie and Singleton (1993) pointed out, one can simulate from (34) and (35) by an approximation such as (36), and thereby compute the moments of the continuous processes, which in turn may be used to match the moments of the data via a GMM estimator.<sup>46</sup> Alternatively, one could use indirect estimation methods, simulating from the continuous time process, but estimating using a discrete time auxiliary model. Gouriéroux et al. (1993) use the Euler approximation with  $h = 1$  as the auxiliary model and  $h = 1/10$  to simulate the data, while Engle and Lee (1994) use GARCH as the auxiliary model.<sup>47</sup>

An interesting question that arises when one attempts to estimate models closer to the finance literature is the treatment of the parameter  $\gamma$ . Most investigators have set this to zero, although the literature on the term structure frequently works with models in which  $\gamma = 0.5$ , see Cox et al. (1985). If  $\gamma \neq 0$ , there is a “levels” effect on volatility, which is over and above whatever specification is adopted for  $\sigma_t^2$ . There is some interest therefore in estimating  $\gamma$ . Chan et al. (1992) use (36) with  $h = 1$ ,  $\sigma_t = \sigma$ , since then  $E(u_t^2 - \sigma^2 y_t^{2\gamma}) = 0$ , where  $u_t = (\Delta y_t - a - b y_{t-1})^2$ , provides a moment condition to estimate  $\gamma$ . Brenner et al. (1994) generalize this to allow for  $\sigma_t^2$  to be a GARCH model, since it is possible for the levels effect on volatility to simply reflect mis-specification in  $\sigma_t^2$ . They find

<sup>46</sup> One might wish to use as accurate an approximation as possible; second order schemes such as the Mihlstein (1974) approximation might therefore be preferred.

<sup>47</sup> There are many other questions focussing on the relation between GARCH and continuous time processes that have been examined in the literature. Nelson (1990c) shows how to formulate a continuous time SV model that is the limit of a GARCH process. Also of interest has been the question of whether a GARCH process might be used to consistently estimate continuous time volatility. Because GARCH models average  $\epsilon_{t-j}^2$ , Nelson and Foster (1992) interpreted them as a filter and showed that the answer was in the affirmative. A good summary of this literature is available in Bollerslev et al. (1994).

strong evidence of a levels effect. A similar conclusion is reached by Koedijk et al. (1993) who have  $\sigma_t^2$  as driven by  $\epsilon_{t-1}^2$  rather than  $\sigma_{t-1}^2 \epsilon_{t-1}^2$ . Pagan et al. (1995) and Broze et al. (1993) use indirect estimation to estimate  $\gamma$  keeping  $\sigma_t^2$  as a constant. In the first paper the auxiliary model was either the discrete Euler approximation with  $h = 1$  or an EGARCH(1,1) model; in either case the value of  $\gamma$  was around 0.7 for short-term interest rates.<sup>48</sup>

### 3.6. Are means and variances related? The GARCH-M model

When modeling returns it is conventional to take these as the excess over some risk free rate, so that it is really a risk premium that is being explained. Similarly, the difference between spot and forward rates is a risk premium. Theoretical models, such as Merton (1973), make the market return a function of volatility, i.e., the risk premium should be larger when the asset return is more volatile. Of course, for an agent making decisions at a point in time  $t$ , the appropriate concept of volatility is what the conditional variance of the asset return,  $\sigma_t^2$ , would be over the holding period for the asset, leading to the relation

$$x_t = \delta g(\sigma_t^2) + e_t. \quad (37)$$

In Merton's model,  $g(\sigma_t^2) = \beta \sigma_t^2$ ,  $\beta > 0$ , but other functional forms for  $g(\cdot)$  could emerge, possibly allowing the response to depend upon the sign and level of  $\sigma_t^2$ . Higher order conditional moments might also enter into the relation in certain types of utility maximizing models, although, in an interesting experiment pricing stocks in an economy with a stochastic dividend process, Gennotte and Marsh (1992) found that  $g$  was linear.

Models such as (37) feature a conditional mean for returns that depend on higher order conditional moments. Early work on this topic, e.g., Pindyck (1984), made  $\sigma_t^2$  a distributed lag of the squares of returns, i.e.,  $\sigma_t^2 = \sum_{k=0}^K w_k x_{t-k}^2$ , and proceeded to regress returns on this constructed variable.<sup>49</sup> Many criticisms can be made about this practice see Pagan and Ullah (1988), but an important one is that the assumption implies a very arbitrary model for the conditional variance. This leads to the idea of expressing  $\sigma_t^2$  as some parametric model in the GARCH class, and Engle et al. (1987) accomplished this with their ARCH in mean (ARCH-M) model, wherein (37) was augmented by an ARCH model for  $\sigma_t^2$  such as

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2. \quad (38)$$

Maximum likelihood was used to estimate the unknown parameters (once a form for  $g(\cdot)$  was specified). Assuming that  $e_t$  was conditionally  $\mathcal{N}(0, \sigma_t^2)$ , it is easily

<sup>48</sup> Some of the volatility may be due to jump processes driving the diffusion. Das (1993) estimates such a combined model for bond yields.

<sup>49</sup> More precisely a weighted average of  $(x_t - \bar{x}_t)^2$  was employed, where  $\bar{x}_t$  was a rolling sample average.

seen that this means  $x_t$  is conditionally normal ( $\delta g(\alpha_0 + \alpha_1 e_{t-1}^2), \sigma_t^2$ ), allowing a likelihood to be easily written down and maximized. Extensions to allow  $\sigma_t^2$  to follow GARCH or EGARCH forms are quite straightforward.

Analysis of ARCH-M models is much more complex than was true of pure ARCH models. In the latter, whatever variables impinged upon the conditional mean could be regressed out, and then the residuals from such a regression,  $\hat{e}_t$ , might be used for diagnostic checking and specification. Previously, the autocorrelation function of the squares of the residuals was a primary device for specification analysis. Now, however, it is impossible to estimate  $e_t$  without first specifying a valid model for  $\sigma_t^2$  (and  $g(\cdot)$ ), so that pre-estimation specification analysis is now very difficult, making post-estimation investigation very important. It was for ARCH-M models that the specification tests set out by Pagan and Sabau (1992) were developed. Since estimation is by MLE, specification tests involve setting up some moment restriction that is true under the null,  $E(m_i(\theta)) = 0$ , regressing  $m_i(\hat{\theta})$  against unity and the scores for  $\theta$ , and then testing if the intercept in such a regression is zero.

The ARCH-M structure can also be troublesome when trying to establish sampling theory for parametric estimators. Clearly, if  $\sigma_t^2$  was to be made IGARCH, a very common finding in the applied literature, there would be a regressor in (37) that would have an infinite variance. If  $g(\cdot)$  was the square root function then the regressor would be  $\sigma_t$ , and one might weight the data with this term, as was done in the proofs for IGARCH, but then difficulties would arise in estimating an intercept in the equation. A theoretical analysis of this problem is important, given the prevalence of the GARCH-M model in the analysis of risk premia – see Bollerslev et al. (1992) for a list of applications. Recently, Lee (1992) has established sufficient conditions for the asymptotic normality of the MLE of the GARCH-M (1,1) model when  $g(\sigma_t^2) = \sigma_t$ . An important condition is that  $E[(\beta_1 - \alpha_1 \delta \epsilon_t / \beta_1 + \alpha_1 \epsilon_t^2 | \mathcal{F}_{t-1})] < 1$ . When  $\epsilon_t$  is n.i.d. (0,1) this becomes an unconditional expectation and, with  $\alpha_1 + \beta_1 \leq 1$ , the condition effectively implies that  $\delta$  has to be in the range  $-3 < \delta < 3$ .

An alternative approach that can sometimes pay dividends is to engage in a non-parametric analysis, estimating  $E_{t-1}(x_t)$  and  $\sigma_t^2$  in this way rather than through parametric models. Pagan and Hong (1991) had some success with this procedure for the series on the excess yield on Treasury Bills used in Engle et al. (1987), but very little for the return to equity on the NYSE. In fact, the conditional variance of the latter was well estimated, and displayed the typical skewed response to the sign of returns mentioned earlier, but no relation was found between the conditional mean and variance of returns. Indeed, I think that this is a fair summary of the outcome of parametric investigations as well. In a marked departure from the consensus of the literature, Linton (1992), using semi-parametric techniques, finds that the  $g(\cdot)$  function in (37) should not be linear, and that taking account of the non-linearity is crucial for finding an important impact of  $\sigma_t^2$  upon the excess holding yield.

A note of caution needs to be sounded regarding the use of non-parametrics in this context, as the presence of  $\sigma_t^2$  in (37) means that  $E_{t-1}(x_t)$  depends upon all past returns, even if  $\sigma_t^2$  is just an ARCH(1) model. To see this let  $g(\sigma_t^2) = \sigma_t^2$ , and  $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$ , so that (37) becomes

$$x_t = \delta(\alpha_0 + \alpha_1 e_{t-1}^2) + e_t \quad (39)$$

$$= \delta[\alpha_0 + \alpha_1(x_{t-1} - \delta(\alpha_0 + \alpha_1 e_{t-2}^2))]^2 + e_t \quad (40)$$

By continued iteration it is evident that all past  $x_t$  appear in the determination of  $x_t$ . In fact, the dependence is a very non-linear one, and conditions for  $x_t$  to be covariance stationary have yet to be worked out. From this simple model it is apparent that ARCH-M models imply that the conditional mean of  $x_t$  depends on an infinite number of conditioning elements, whereas the methods of non-parametric analysis are restricted to a finite number. Hence, a non-parametric analysis could not exactly replicate an ARCH-M model structure, and the quality of its approximation will be a function of the magnitude of coefficients such as  $\alpha_1$  and  $\beta_1$ .

#### 4. Statistical representations of multivariate data

In an earlier section we described the properties of univariate data and statistical models that were useful descriptions of such data. However, financial analysis is largely concerned with the relationship between series, e.g., options prices and volatility, rates of return on different assets, forcing a proper analysis of the multivariate characteristics of data. By far the most important summary measures of how series relate to one another have been those concerned with the number of factors that are common to the data, and we will organize this section along those lines.

##### 4.1. The common trend factor

Given that financial series are frequently  $I(1)$ , determining if series are cointegrated is a first step in much financial analysis. For example, as seen in Baillie (1989), the expectations theory of the term structure for  $n$  maturities would imply that there are  $(n - 1)$  co-integrating vectors among these bond yields. There are a variety of methods for testing for the presence of co-integration and the number of co-integrating vectors. Johansen's technique is perhaps the one most in use as it has been programmed into many microcomputer packages (Johansen, 1988). Testing if the one month, two month, three month, four month and five year zero coupon bond yields in the McCulloch data set are co-integrated, Johansen's maximal eigenvalue test statistic values were (with 5% significance critical values in brackets): 72.89 (34.4) for  $r = 0$  vs.  $r = 1$ ; 51.94 (28.13) for

$r = 1$  vs.  $r = 2$ ; 39.89 (22.00) for  $r = 2$  vs.  $r = 3$ ; 31.28 (15.67) for  $r = 3$  vs.  $r = 4$ ; 2.90 (9.24) for  $r = 4$  vs.  $r = 5$ . Consequently, one could be comfortable with the conclusion that there are indeed four co-integrating vectors within this data.

However, in many applications, the co-integrating vectors are frequently predicted to be certain values by theory, so that one might want to impose these and then test if the series formed from such linear combinations are integrated or not. An example is the term structure case which predicts that any two yields are linked by a co-integrating vector [1 – 1]. Thus, we could form the vector of series such as the yield differential between yields at all maturities and one-period yields, and then test if there are unit roots in these series. We might in fact just run a sequence of unit root tests, such as were discussed earlier, for each of these series separately, rather than explicitly looking at it in a multivariate framework. In any case, the techniques for testing for co-integration, and the rationale for them, requires a separate treatment. It is probably worth mentioning that care has to be taken when estimating co-integrating vectors, if more than one is expected, as their lack of uniqueness means that reduced form estimators such as Johansen's may produce linear combinations of the vectors rather than the vectors themselves. To illustrate this, Johansen's estimator was applied to a data set consisting of one month, 10 year and 20 year bond yields from the McCulloch data set. The expected co-integrating vectors were [1 – 1 0] and [0 1 – 1], i.e., the interest differentials are  $I(0)$ , but the actual estimates found (by MFIT386) were [-1 2.8 – 2.0] and [-1 – 16.3 17.5]. Approximating these by [-1 3 – 2] and [0 – 1 1], it is seen that the estimates are the linear combination  $-1 \times [1 - 1 0] - 2 \times [0 - 1 1]$ . In this instance it is easily seen how to unscramble the final estimates to get back to a set of basic co-integrating vectors, but in those cases where there are unknown elements in the vectors that would not be possible.<sup>50</sup> For example, if one was estimating a demand elasticity for financial asset holdings using observations (say) on interest rates and an aggregate portfolio, there is a risk that the estimates found could actually be linear combinations of the demand and supply elasticities. The solution to such problems has to be to formulate and estimate structural models and not to attempt to recover structural parameters from a reduced form.

Because of the emphasis upon factors in financial analysis, in most instances it is useful to re-formulate co-integrating relations as the common trends idea in Stock and Watson (1988). Let  $y_t$  be an  $n \times 1$  vector of  $I(1)$  variables with  $r$  cointegrating vectors. Then

$$\Delta y_t = C(L) e_t \quad (41)$$

---

<sup>50</sup> For the McCulloch data the hypothesis that the co-integrating errors are the spreads is rejected. Pagan et al. (1995) discuss what might be the cause of this outcome. Possible reasons are non-linearities in the underlying processes, the presence of levels effects in volatility, and the effect of having near-integrated rather than strictly integrated series.

would be an MA representation of the  $I(0)$  variable  $\Delta y_t$ , and this may be written as

$$\Delta y_t = C(1)e_t + C^*(L)\Delta e_t, \quad (42)$$

using the result that  $C(L) \equiv C(1) + \Delta C^*(L)$  ( $C^*(L)$  is actually defined by this relation). If there are  $r$  co-integrating relations among the  $y_t$ , defining a vector of stochastic trends  $\tau_t$  with dimension  $(n - r)$  which evolve as

$$\tau_t = \tau_{t-1} + \phi_t, \quad (43)$$

where  $\tau_0 = 0$ , we get the “common trends” representation

$$y_t = J\tau_t + C^*(L)e_t. \quad (44)$$

In (44)  $y_t$  is driven by  $(n - r)$  stochastic trends and  $\phi_t$  in (43) is white noise.<sup>51</sup> If there were no co-integrating vectors ( $r = 0$ ) there will be  $n$  trends driving the  $n$  variables, i.e., there are *no* common trends and each variable exhibits separate trending behavior. For the term structure case  $r = n - 1$  and there should be a single common trend that drives all interest rates; for a small country this is probably the world interest rate. Many studies have been performed on the number and nature of common trends in various asset price series, e.g., Baillie and Bollerslev (1989) for exchange rates and Kasa (1992) for stock prices.

#### 4.2. Common factors in detrended series

It is possible that there may be common factors in series after these have been detrended in some way. There are two ways in which these common factors might be manifest. In the first, the  $\text{cov}(u_t)$ ,  $u_t = C^*(L)e_t$ , has full rank but with restrictions upon it induced by a factor structure, while the second features the case in which  $u_t = C^*(L)e_t$  has a rank-deficient covariance matrix that can be further decomposed into  $J\xi_t$ , where  $\rho(\text{cov}(u_t)) = \rho(\text{cov}(\xi_t)) = \rho(J) < n$ , and  $\rho(\cdot)$  denotes the rank of the matrix in brackets. This latter situation leads to what Vahid and Engle (1993) have called the “common-trend, common-cycle” representation

$$y_t = J\tau_t + J\xi_t, \quad (45)$$

and, cast into our taxonomy, the series can be regarded as being composed of two types of factors, the “common trends”,  $\tau_t$  and the non-trend factors,  $\xi_t$ .<sup>52</sup> This type of representation is very popular in the term structure literature, e.g., Cox et al. (1985), Chen and Scott (1993) and Duffie and Kan (1993). Two different views

<sup>51</sup> We have suppressed the dependence of  $y_t$  upon  $y_0$ .

<sup>52</sup> Although we have used the terminology of Vahid and Engle the situation we are interested in is different to that of those authors. The  $\text{var}(u_t)$  could be rank-deficient because either all elements in  $C^*(L)$  or  $\text{var}(e_t)$  are singular. Our focus is upon the latter; Vahid and Engle concentrate upon the former.

of the non-trend factors can now be elicited, depending upon how  $y_t$  is “detrended” – the first concentrates upon returns  $\Delta y_t$ , while the second deals with the co-integrating errors  $\zeta_t = \alpha' y_t$ , where  $\alpha$  is the  $(r \times n)$  matrix of co-integrating vectors. The latter will generally be the “spreads” between asset prices.

#### 4.2.1. Factors in returns

The most famous factor model of returns  $r_t = \Delta y_t$  is of course the “market model” in which  $r_t$  is a vector of returns.<sup>53</sup> Suppose that the objective is to explain the return on the  $j$ th portfolio of assets  $r_{jt}$ . If there is only one asset per portfolio this will simply be the return on that asset. “Portfolios” can be interpreted in diverse ways. For example, Harvey (1991) has  $r_{jt}$  as the return on equity in the  $j$ th country. In the market model  $r_{jt}$  is related to the return on the market or aggregate portfolio,  $r_{mt}$ . The latter may be a simple average of the returns to all stocks in the economy or perhaps is a weighted average, with weights depending on the value of the portfolio accounted for by each stock.

A popular way to derive the factor structure is to impose the restriction that  $r_{jt}$  and  $r_{mt}$  are multivariate normal, from which it follows that

$$E(r_{jt} | r_{mt}) = \alpha_j + \beta_j r_{mt}, \quad (46)$$

where  $\beta_j = \text{cov}(r_{jt}, r_{mt})/\text{var}(r_{mt})$  and  $\alpha_j = E(r_{jt}) - \beta_j E(r_{mt})$ . Therefore,

$$r_{jt} = \alpha_j + \beta_j r_{mt} + e_{jt}. \quad (47)$$

Hence,  $r_{jt}$  could be predicted once  $r_{mt}$ ,  $\alpha_j$  and  $\beta_j$  are known. The object of interest therefore is the determination of  $\beta_j$ , as that parameter summarizes the relationship of the return on the  $j$ th asset to that of the market portfolio.

Traditionally  $\beta_j$  has been estimated by regression, meaning that the range of problems encountered with any linear regression model recur here as well. Examples would be that outliers in the data can affect the point estimate of  $\beta_j$ , which encourages robust estimation of the parameters, while the presence of heteroskedasticity in returns can also be true of the errors in (47), thereby demanding the computation of heteroskedastic robust standard errors in order to make valid inferences. Some have argued that a linear relationship might be inappropriate. Tracing the argument back to the source of linearity however, the question becomes one concerning the underlying multivariate density. Linearity is associated with normality but not uniquely so, for example Spanos (1986, pp. 122–125) points out that the Student's  $t$ -density yields a linear conditional expectation, while densities such as the bivariate Pareto result in non-linear conditional means. Moreover, models formed from densities such as the bivariate  $t$  also have implications for conditional variances. Non-parametric estimation methods can be usefully employed here to shed light on the conditional moments of an asset return given the market return.

---

<sup>53</sup> This is an example of a factor model in which  $\text{cov}(e_t)$  is not singular.

It is possible to produce multiple factors by concentrating upon the nature of  $\beta_j$  in (47). There it was a constant equal to the ratio of the  $\text{cov}(r_{jt}, r_{mt})$  to the variance of  $r_{mt}$ . However, in line with the distinction between conditional and unconditional moments, one might wish to consider models for  $r_{jt}$  in which the conditional rather than the unconditional density of returns is utilized. To this end let  $F_{t-1}$  be a set of conditioning variables, not including  $r_{jt}$  and  $r_{mt}$  but containing their past histories. Then, the conditional capital asset pricing model has

$$E[r_{jt} | F_{t-1}] = \beta_{jt} E[r_{mt} | F_{t-1}] \quad (48)$$

where  $\beta_{jt} = \text{cov}(r_{jt}, r_{mt} | F_{t-1}) / \text{var}(r_{mt} | F_{t-1})$ .

Because the coefficients of the conditional market model are functions of the conditional moments for  $r_{jt}$  and  $r_{mt}$  it is necessary to model these in some way. Assuming that

$$r_{mt} = \sum_{j=1}^K r_{jt} = r'_t,$$

where  $r'_t$  is the  $(I \times K)$  vector of returns  $(r_{1,t}, \dots, r_{K,t})$ , it is clear that  $\text{var}(r_{mt} | F_{t-1}) = \text{var}(r'_t | F_{t-1}) = \text{cov}(r_t | F_{t-1})$ , while  $\text{cov}(r_{jt}, r_{mt} | F_{t-1}) = r'_t \text{cov}(r_t | F_{t-1})$ , showing that the conditional moments appearing in (48) are determined by the conditional means and variances of the returns  $r_{jt}$ , so that multiple factors are involved, albeit in a non-linear fashion. This is also true for the literature on optimal hedging where the conditional variance of the optimal hedged portfolio is minimized by choosing as weights the ratio of the conditional covariance of spot and futures prices divided by the conditional variance of the futures price. Modeling these constituent moments was the concern of earlier sections. A natural way to proceed is to allow the conditional mean for  $r_{jt}$  to depend on its conditional variance, as in GARCH-M models; and to subsequently model  $\text{cov}(r_t | F_{t-1})$  by a multivariate GARCH or factor GARCH structure described later. Bollerslev et al. (1988) estimate such a model for stocks, bonds and bills. Estimation is just as for the GARCH-M model described earlier. A number of other applications of this model to financial data have been made and these are described in Bollerslev et al. (1992). Baillie and Myers (1991) apply the GARCH framework to the determination of optimal hedge ratios.

Although the linearity of (48) seems to be intimately connected with a conditional normal density for returns, models have been suggested that take the linearity as a prior constraint and then proceed to specify other models for the conditional moments than those in the GARCH class. Harvey (1991) makes  $E_t(r_{jt}) = z'_t \delta_j$  and  $E_t(r_{mt}) = z'_t \delta_m$ , where  $z_t$  are elements in  $F_{t-1}$ , so that  $e_{jt} = r_{jt} - z'_t \delta_j$  and  $e_{mt} = r_{mt} - z'_t \delta_m$ , allowing (48) to be written as

$$z'_t \delta_j = (z'_t \delta_m) E[e_{jt} e_{mt} | z_t] / E[e_{mt}^2 | z_t], \quad (49)$$

from which

$$E[e_{mt}^2 z'_t \delta_j | z_t] - E[e_{jt} e_{mt} z'_t \delta_m | z_t] = 0. \quad (50)$$

By definition  $E[z_t e_{mt}] = E[z_t e_{jt}] = 0$ , and these two moment conditions, along with (50), may be used to estimate  $\delta_j$  and  $\delta_m$  by GMM. No estimate of  $\beta_{jt}$  can be made without being more specific about the conditional moments, but it is possible to test various hypotheses about the conditional CAPM using the GMM estimates, as well as to assess how adequate an explanation it provides of the data.

There are many variants of this idea of estimating  $\beta$  from its conditioning set, e.g., Schwert and Seguin's (1990) single index model of the conditional covariance between the  $i$ th and  $j$ th portfolio (see (63)) implies a covariance between the  $j$ th portfolio and the market portfolio that is a linear function of the market conditional variance, so that  $\beta_{jt}$  can be estimated quite simply from their model (Schwert and Seguin, 1990).

Braun et al. (1992) set up a model, initially based on (48), treating  $r_{mt}$  and  $r_{jt}$  as coming from a recursive bivariate EGARCH process in which the innovations driving  $\sigma_{mt}^2$ ,  $z_{mt}$ , are "aggregate shocks", while  $\sigma_{jt}^2$  depends on specific shocks to the  $j$ th portfolio,  $z_{jt}$ , as well as to the market shocks. Subsequently, however, they adopt an ad-hoc specification of  $\beta_t$  as in (51) (the definition of  $z$ 's is given below (15))

$$\beta_{jt} = \lambda_0 + \lambda_4(\beta_{jt-1} - \lambda_0) + \lambda_1 z_{m,t-1} z_{j,t-1} + \lambda_2 z_{m,t-1} + \lambda_3 z_{j,t-1}, \quad (51)$$

where the last two terms allow for leverage effects on conditional betas; if both are negative conditional betas rise when returns fall.

The distinctive feature of the approach just mentioned was the formulation of the conditional beta as an autoregression. Representing  $\beta_{jt}$  as an autoregression seems to have been successful in practice, Rosenberg (1985), and one might think of it as an approximation to whatever the true process for  $\beta_{jt}$  is, hoping that by choosing the order of autoregression large enough a good approximation would be achieved. Rosenberg formulates his model as

$$r_{jt} = \alpha + \beta_{jt} r_{mt} + e_{jt} \quad (52)$$

$$\beta_{jt} = \rho_j \beta_{jt-1} + v_{jt}, \quad (53)$$

and estimates the unknown parameters  $\rho_j$ ,  $\sigma_{v,j}^2$  by varying coefficient regression techniques. Generally this is done by maximum likelihood. Defining  $F_t$  as a conditioning set based on past data and  $r_{mt}$ , if  $r_{jt}$  is  $N(\mu_t, \sigma_t^2)$  conditional on  $F_t$ , it follows that the log likelihood of  $r_{j1}, \dots, r_{JT}$  is (ignoring terms that are asymptotically negligible)

$$-\frac{1}{2} \sum \log \sigma_t^2 - \frac{1}{2} \sum \sigma_t^{-2} (r_{jt} - \mu_t)^2. \quad (54)$$

Generally,  $\mu_t$  and  $\sigma_t^2$  can be computed by the Kalman filter for any given values of the unknown parameters, and so (54) can be maximized. There is obviously some difficulty in allowing  $r_{mt}$  to appear in the information set as it is a linear combination of all the portfolios, although for a large enough number of

assets, and small portfolios, we might think of the market as solely reflecting the macro shock and it is therefore this that is being conditioned upon; the error term  $e_{ji}$  then just reflects the micro shocks.

#### 4.2.2. Factors in spreads

The dichotomy in the ways that factors manifest themselves is most important when the detrending is done via the construction of “spreads”. Mahieu and Schotman (1994b) consider a model of bilateral exchange rates in which the changes in the rates reflect differences in the news in each country. Taking a system of bilateral rates it is clear that there is a common factor – the news occurring in the country whose currency is being used to express the rates. Thus, if we consider the yen/\$, mark/\$ and franc/\$ rates, news about the U.S. will represent the common factor, and this imposes a restriction upon the covariance matrix of the returns which can be used to estimate the characteristics of the factors.

Factors giving rise to a singularity in the  $\text{cov}(u_t)$  need to be given a separate treatment. A slightly different interpretation of the factors  $\xi_t$  in (45) is available by multiplying (45) by the transposed matrix of co-integrating vectors,  $\alpha'$ , and imposing  $\alpha' J = 0$ , to get  $\zeta_t = \alpha' y_t = \alpha' \bar{J} \xi_t$ . Hence,

$$\begin{aligned}\rho(\text{cov}(\zeta_t)) &= \min(\rho(\alpha' \bar{J}), \rho(\text{cov}(\xi_t))) \\ &= \min(r, \dim(\xi_t)).\end{aligned}$$

Since the total number of factors must be less than  $n$ ,  $\dim(\xi_t) \leq n - (n - r) = r$ , and it is apparent that  $\rho(\text{cov}(\zeta_t)) = \rho(\text{cov}(\xi_t))$ , i.e., the common factors  $\xi_t$  show up in the cointegrating errors  $\zeta_t$ . Generally, the co-integrating errors have a clear interpretation in financial data as “spreads”, e.g., in the case of the term structure they are the spreads between yields of different maturities, while in relations between a spot and a forward market they would be the forward premium. Consequently, this result directs attention to the spreads in the search for common factors.<sup>54</sup>

Spreads between asset prices have been extensively studied in recent years. Dybvig (1989) applies principal components to  $\zeta_t$  to conclude that the number of factors in the term structure is much lower than the number of rates. Applying principal components to the data set composed of  $sp_2$ ,  $sp_3$ ,  $sp_4$ ,  $sp_5$ ,  $sp_6$ , and  $sp_9$ , where  $sp_j$  is the spread between the  $j$ -month and one month yields, the eigenval-

<sup>54</sup> Vahid and Engle (1993) noted that, if there were common cycles in  $y_t$ , there exists a  $\delta$  such that  $\delta' \bar{J} = 0$ , meaning that  $\delta' \Delta y_t = \delta' J \phi_t$  would be white noise and a test for common cycles could be regarded as a test for serial correlation in the residuals  $\hat{\delta}' \Delta y_t$ , where  $\hat{\delta}$  is an estimator of  $\delta$ . Obviously determining the  $\rho(\text{cov}(\zeta_t))$  represents another way of finding the number of common cycles. There may be advantages to it as this method does not require the use of instrumental variables as their test does.

ues of the covariance matrix of these variables are 1.8, 0.05, 0.005, 0.002,  $7.6 \times 10^{-5}$  and  $1.3 \times 10^{-5}$ , pointing to the fact that these six spreads can be summarized very well by two components (at most).<sup>55</sup> Restricting attention to  $sp_3$ ,  $sp_6$  and  $sp_9$ , the eigenvalues are 1.2, 0.03 and 0.002, telling the same story. Knez et al. (1989) estimate a factor model for the excess yields over the repo rate by maximum likelihood. Other factor models, based on the Cox–Ingersoll–Ross model of the term structure, have been estimated by Chen and Scott (1993) and Pearson and Sun (1994).

A difficulty in directly applying principal component methods is that  $\zeta_t$  will rarely be i.i.d. and some allowance should be made for that fact. To describe the autocorrelation structure of  $\zeta_t$ , return to the levels  $y_t$ , and observe that the presence of co-integration among the series means that they follow a vector ECM (taking a first order system for simplicity)

$$\Delta y_t = \gamma \zeta_{t-1} + v_t,$$

so that

$$\alpha' \Delta y_t = \alpha' \gamma \zeta_{t-1} + \alpha' v_t. \quad (55)$$

It is worth dwelling on some of the implications of (55). In particular, the matrix  $\gamma$  represents the influence of past spreads upon the returns  $\Delta y_t$ . Given the earlier evidence that it was hard to predict  $\Delta y_t$  using past information, it is likely that  $\gamma$  will be close to zero, and therefore  $\zeta_t$  will exhibit “near unit root” behavior. Strong persistence in spreads has indeed been observed – see Evans and Lewis (1994), Pagan et al. (1995) and Hejazi (1994) for the term structure and Baillie and Bollerslev (1994) and Crowder (1994) for the forward premium in exchange rates. Baillie and Bollerslev argue that the persistence seen in the premia is best represented as a fractionally integrated process rather than the  $I(1)$  structure proposed by Crowder. The same point may also apply to the term structure; Hejazi found that there was no evidence of a unit root in the excess holding yields but that, if one regressed them on a forward rate which was  $I(1)$ , the resulting regression coefficient was non-zero. This suggests that the type of persistence apparent in holding yields is not of the  $I(1)$  variety.

#### 4.3. Relations between financial series

If expected returns remain constant, efficient markets demands that all information has been used in pricing an asset so that returns should be uncorrelated with information from other sources than just the past history of returns. Thus the autocorrelation function does not provide a sufficiently wide perspective to judge

---

<sup>55</sup> It is interesting to observe that the covariance matrix of bilateral returns given in Mahieu and Schotman (1994b, p. 282) does not reveal any singularity, emphasizing the need to distinguish between the two impacts a factor structure might have.

questions regarding efficient markets, and this has led to research demonstrating that returns are influenced by a wide variety of other series, e.g., future stock returns are explained by the dividend yield ( $D_t/P_t$ ).<sup>56</sup> Much of this work involves just simple regressions and little in the way of new econometric technique is needed. One relatively new development has been Fama and French's extension of their long horizon returns work (Fama and French, 1988b). They consider the regression of  $k$  period returns on stocks,  $r_{t+k,t}$ , against the dividend yield and test if this coefficient is zero or not. They sample data so as to avoid over-lapping errors, and therefore work with traditional OLS standard errors when testing the hypothesis of no relation. This reduces the effective sample size quite significantly for large  $k$ .

Hodrick (1991) points out an alternative approach to Fama and French. Because a  $k$  period continuously compounded return is the sum of one period returns, i.e.,  $r_{t+k,t} = r_{t+1} + \dots + r_{t+k}$ , the regression coefficient is proportional (in large samples) to

$$E[\{r_{t+1} + \dots + r_{t+k}\}(D_t/P_t)],$$

which is the probability limit of the numerator when  $r_t$  are stationary ergodic processes. The denominator just represents a positive scaling factor so a test of whether the parameter is zero is a test of whether the numerator is zero. But, under stationarity, this is identical to  $E[r_{t+1}\{(D_t/P_t) + \dots + (D_{t-k+1}/P_{t-k+1})\}]$ , so that we could interpret Fama and French's methodology as regressing  $r_t$  against  $\sum_{j=0}^{k-1} [D_{t-j}/P_{t-j}]$  and testing if the coefficient of the regressor is zero. This last regression has the advantage that it uses all observations and that there is no overlapping data problem. In practice Hodrick multiplies the sum of lagged dividends by  $k^{-1}$  so that the slope coefficient measures the response of annualized expected returns over a given horizon to a change in the ex ante dividend yield. He uses heteroskedastic and autocorrelation consistent standard errors, since there is no certainty that returns are i.i.d. under the null (and we actually know that this is extremely unlikely).

Bekaert and Hodrick (1992) mention an alternative procedure that reduces the loss of observations owing to the need to lag  $(D/P)_t$   $k$  times. They put the log of the first period return, the dividend yield, and the one month Treasury Bill yield ( $rb_t$ ) into a vector  $z_t$ , and then fit a first order VAR

$$z_t = Az_{t-1} + \eta_t, \quad (56)$$

to this trivariate system. The  $j$ th autocovariance of  $z_t$  is given by  $C(j) = A^j C(0)$ , where  $C(0)$  is the variance of  $z_t$  and is a function of  $A$  and  $\text{var}(\eta_t)$ . Hence the autocovariance between the log of the first period return and the  $j$ th lagged

<sup>56</sup> As Gennette and Marsh (1992) point out one should be careful in giving such correlations a causal interpretation. They simulated an artificial economy with fully rational actors but the generated data exhibited a return-dividend correlation since the stochastic output process driving this economy affects both dividends and stock prices.

dividend yield is  $e'_1 C(j) e_2$ , where  $e_\ell$  is a  $(3 \times 1)$  vector with unity in the  $\ell$ th place and zero elsewhere. Thereupon

$$\begin{aligned} E[r_{t+1}\{(D_t/P_t) + \dots + (D_{t-k+1}/P_{t-k+1})\}] &= e'_1 C(1) e_2 + \dots + e'_1 C(k-1) e_2 \\ &= e'_1 A C(0) e_2 + e'_1 A^2 C(0) e_2 + \dots + e'_1 A^{k-1} C(0) e_2 \\ &= e'_1 A [I + A + \dots + A^{k-1}] C(0) e_2, \end{aligned}$$

and we can test if this is zero by finding

$$\text{var}\{e'_1 \hat{A}(I + \hat{A} + \dots + \hat{A}^{k-1}) C(0) e_2\}.$$

This variance can be found using the  $\delta$  method. It would seem that the only advantage this method has is that it economizes on observations, since only one observation is lost in estimating the VAR, and the autocovariances are all then derived ones. For small numbers of observations, or for very large  $k$ , this may be a saving, but it is not so clear what the advantage is otherwise. Moreover, there is a potential cost if the process is not well represented by a linear VAR; in particular if there are non-linearities affecting various autocovariances.

Just as was true for univariate series one might seek non-linear relations between series rather than linear ones. Suppose we want to estimate the  $E(y_t | x_t)$ , without imposing any functional form upon this relation. By definition

$$y_t = m(x_t) + u_t, \quad (57)$$

where  $m(x_t)$  is some unknown function of  $x_t$  and  $u_t$  is taken to be i.i.d.  $(0, \sigma^2)$ . An example that has attracted some interest would be if  $y_t$  is the price of a derivative while  $x_t$  contains the price of the derivative asset ( $p_t$ ), the strike price ( $s_t$ ), volatility ( $\sigma$ ), the risk free interest rate ( $r_t$ ) and the maturity of the derivative ( $\tau$ ). In Black and Scholes' formula  $u_t = 0$  and a specific function linking these quantities is provided,

$$m^{\text{BS}}(x_t) = p_t \Phi(d_1) - s_t e^{-r_t \tau} \Phi(d_2),$$

where

$$\begin{aligned} d_1 &= \frac{\log(p_t/s_t) + (r_t + \sigma^2/2)\tau}{\sigma\tau} \\ d_2 &= d_1 - \sigma\tau, \end{aligned}$$

and  $\Phi(\cdot)$  is the cumulative normal density. However, it is known that this formula tends to break down for out-of-the money and short-maturity options, where the degree of non-linearity seems more pronounced. Moreover, the derivation of the Black–Scholes formula assumes that volatility is a constant, whereas implicit volatilities computed from options prices using the formula change over time and also seem to be a function of  $p_t/s_t$  – the latter constituting the well known “smile” effect in volatility. Consequently, it is of interest to estimate  $m(x_t)$  non-parametrically in an attempt to correct for these problems.

The basic idea behind nonparametric estimation of  $E(y_t | x_t)$  is to approximate  $m(x_t)$  arbitrarily closely by combining a set of basis functions  $\sum_{j=1}^M \beta_j \varphi_j(x)$  and to then estimate the  $\beta_j$ , leading to  $\hat{m}(x_t) = \sum_{j=1}^M \hat{\beta}_j \varphi_j(x_t)$ . Hutchinson et al. (1994) fitted the options model above using neural networks in which  $\varphi_{jt} = (1 + \exp(-x'_t \delta))$  to approximate  $m(x_t)$ . Kearns (1993) simulated data from an options pricing model with stochastic volatility and non-parametrically estimated the non-linear function implied by a particular pricing model using the flexible Fourier form of Gallant (1981), which has as  $\varphi_{jt}$  equal to 1,  $x_t$ ,  $x_t^2$ ,  $\cos(jx_t)$ ,  $\sin(jx_t)$  ( $j = 1, 2, \dots$ ). After this model has been calibrated, it can be used with actual data to produce predictions of observed options prices. The prediction error may be used as a diagnostic device. The method seems to work quite well. In Kearns' case 98% of the variation in options prices was captured.

Each of the approaches above has approximated the unknown function globally and then predicted what value it would have at given points  $x_t = x$ . An alternative method is to estimate  $\alpha = m(x_t = x)$  by using only those observations whose value is close to  $x$ . Since we can think of  $m(x)$  as a constant, the appropriate estimate is found by choosing  $\alpha$  to minimize  $\sum_{t=1}^T (y_t - \alpha)^2 K(x_t - x/h)$ , where  $K(\cdot)$  is a kernel or weighting function that gives low weight to observations for which  $x_t$  is far from  $x$ . The window-width parameter,  $h$ , determines exactly how far away the observation can be in order to be included in the computation. From standard least squares theory this gives

$$\hat{m}(x_t = x) = \hat{\alpha} = \sum_{t=1}^T K(\psi_t) y_t / \sum_{t=1}^T K(\psi_t), \quad (58)$$

where  $\psi_t = (x_t - x)/h$ . Diebold and Nason (1990) consider predicting exchange rates using a conditional mean estimated non-parametrically in this way, the conditioning variables being the lagged value of the rate. Huang and Lin (1990) have analyzed non-linearities in the term premiums and forward rates with these methods. The Cox et al. (1985) model of the term structure implies a linear relation between expected term premiums on Treasury securities and forward interest rates. However, the linearity stems from the assumptions on the stochastic processes driving the model, which make the state variables linear Markov processes. If the forcing processes were non-linear, term premiums and forward rates would be related in a non-linear way. Huang and Lin apply kernel based methods to the data, but discover that the linearity assumption seems quite good. Pagan and Hong (1991) look at the conditional mean for excess holding yields on Treasury bills as a function of past yields and the yield differential, finding that the relation seems to be non-quadratic.

The principal disadvantages of such local analysis are that there are well known biases in the estimator of  $m(x)$  and, when the  $\dim(x_t)$  is large, very few observations will be used to determine each point. In recent years an attempt has been made to combine the two approaches by assuming that the function can be parametrically approximated around the point  $x$  by, say, a linear polynomial, so

that the optimization problem becomes one of choosing  $\alpha$  and  $\beta$  such that  $\sum_{t=1}^T (y_t - \alpha - \beta(x_t - x))^2 K(\psi_t)$  is minimized, e.g. see Fan et al. (1994), and it seems as if these *locally parametric* methods can produce big improvements in the properties of the local estimator. Of course, there is no reason to chose a polynomial in  $x_t$ . In fact, if one thought that the global function was likely to have a specific shape, then it would make sense to use that. Thus, for options pricing, one might use  $m^{BS}(x_t - x)$ , and this idea has been exploited by Bossaerts and Hillion (1994). Another important application in finance of non-parametrics has been to estimate the yield curve. McCulloch (1971) pioneered the application of spline functions to this – here  $y_t$  are observed yields,  $x_t$  is the maturity and  $\phi_t$  are regression splines – while Fisher et al. (1994) use smoothing splines, which add extra constraints to penalize “roughness” in the estimated yield curve. Gouriéroux and Scaillet (1994) use the local parametric idea with the candidate parametric functions being factor models taken from the term structure literature, such as Vasicek's (Vasicek, 1977).

Another way of proceeding is to model the joint density of  $y_t$  and  $x_t$ , and to then derive the conditional moments from that. If one uses the SNP method, the joint density would be approximated as the product of a polynomial and a multivariate normal. Gallant et al. (1992) have used this approach to examine the relation between returns and the volume of trading. They show that the variance of returns conditional on both past returns and the volume of trading shows no asymmetry in returns, so that it appears that most of the asymmetry is due to the fact that the volume of trading is different in a bear than a bull market.

#### 4.4. Modeling multivariate volatility

It seems very likely that the conditional variance of the return on an asset would be related not only to the past history of its own return but also to those on other assets. Indeed if one non-parametrically computes the variance of stock returns conditional upon the past history of returns as well as the volume of trading, it appears that the asymmetry between the variance and past returns noted previously disappears, emphasizing the need to look at multivariate relations between the variances of series.

One trend has been to engage in multivariate extensions of the univariate GARCH, EGARCH etc. models. Defining the conditional covariance matrix of an  $n \times 1$  vector of returns  $x_t$  as  $\Omega_t$ , Bollerslev et al. (1988) defined the multivariate GARCH model to be

$$\begin{aligned} x_t &= w_t' \beta + e_t \\ \text{vech}(\Omega_t) &= A_o + \sum_{j=1}^q B_j \text{vech}(\Omega_{t-j}) \end{aligned} \quad (59)$$

$$+ \sum_{j=1}^p A_j \text{vech}(e_{t-j} e_{t-j}'), \quad (60)$$

where  $e_t = \Omega_t^{1/2} \epsilon_t$ , and  $\epsilon_t$  is n.i.d.  $(0, I_n)$ . The log likelihood is easily written down as

$$-\frac{T_n}{2} \log 2\pi - \frac{1}{2} \sum \log |\Omega_t| - \frac{1}{2} \sum (x_t - w_t' \beta)' \Omega_t^{-1} (x_t - w_t' \beta), \quad (61)$$

but maximizing (61) is a formidable challenge owing to the huge number of parameters entering it – from (60) alone there are  $n(n+1)/2 + (p+q)n^2/(n+1)^2/4$ , and for (say)  $n = 3$ ,  $p = q = 1$ , there are 78 GARCH parameters to be estimated. For this reason most applications have concentrated upon ways of restricting (60) in some sensible fashion to reduce the number of unknown parameters.

A first suggestion was by Bollerslev et al. (1988) who set  $p = q = 1$  and made the matrices  $A_1$ ,  $B_1$  diagonal, giving the simple model for the elements of  $\Omega_t$ ,

$$\sigma_{ij,t} = \alpha_{0,ij} + \alpha_{1,ij} e_{it-1} e_{jt-1} + \beta_{1,ij} \sigma_{ij,t-1} \quad (62)$$

$$i, j = 1, \dots, n,$$

for a total of  $(p+q+1)(n(n+1)/2)$  parameters: when  $n = 3$ , there are 18 parameters. Later Bollerslev (1990) proposed that  $\sigma_{ij,t} = \rho_{ij} \sigma_{ii,t}^{1/2} \sigma_{jj,t}^{1/2} (i \neq j)$ . In this model there are  $n(n+1)/2$  correlation coefficients  $\rho_{ij}$  as well as the parameters of the  $N$  conditional variance equations to estimate. If each one of these variances was univariate GARCH(1,1), then the total number of parameters would be reduced to  $3n + (n(n+1)/2)$ , i.e., 12 if  $n = 3$ . Such a specification certainly looks attractive as it does not seem unreasonable to restrict the conditional covariances to vary in line with the conditional variances. Nevertheless, there are still a very large number of parameters and one might also wish to further reduce the number of unknowns by imposing some structure upon the  $\rho_{ij}$ . Various other formulations have been tried; one which has the advantage of being both parsimonious and ensuring that the estimated  $\Omega_t$  is positive definite is that given in Engle and Kroner (1995),

$$\Omega_t = C + \sum_{m=1}^M \sum_{j=1}^q B_{mj} \Omega_{t-j} B_{mj}' + \sum_{m=1}^M \sum_{j=1}^p A_{mj} e_{t-j} e_{t-j}' A_{mj}',$$

where the  $B_{mj}$  and  $A_{mj}$  each possess  $n^2$  unknown parameters.

A different approach to reducing the number of unknown parameters is by changing the number of forcing variables, i.e., rather than operate upon  $A_j$  and  $B_j$ , the idea is to replace  $\text{vech}(e_{t-j} e_{t-j}')$  by a smaller dimensional vector. Generally, this is done by concentrating upon a small number of “factors” that are thought to drive the conditional variances. One perspective on this is that it is derivative from Ross’s arbitrage pricing model in which a vector of returns  $x_t$  is written as a linear combination of  $K < N$  factors  $f_t$ , i.e.,  $x_t = f_t' \beta + e_t$ , and this idea is just being shifted to the second moment (Ross, 1976). The difficulty is to translate this idea into an operational model, since factors are generally unob-

served. One strategy that does so is Diebold and Nerlove's formulation which has  $\Omega_t$  a function of a single latent factor,  $f_t$ , and the conditional variance of the factor,  $\sigma_{f_t}^2$ , following a GARCH process, thereby endowing  $\Omega_t$  with the structure  $\Omega_t = C + \delta\delta'\sigma_{f_t}^2$  (Diebold and Nerlove, 1989). Since  $E(\Omega_t) = C + \delta\delta' \text{var}(f_t)$ , they used factor analysis on the unconditional covariance matrix of returns (they have no  $w_t$ 's) to determine estimates of  $\hat{f}_t$  and then applied these to estimate the GARCH model, producing  $\hat{\sigma}_t^2$ . After the latter quantity is determined, MLE can proceed in the standard way. Harvey et al. (1992) point out that it is possible to estimate  $E_{t-1}(x_t)$  and  $\text{var}_{t-1}(x_t)$  under this model and so the quasi-MLE can be applied. Actually, it would seem that the simplest way to estimate the model would be to perform simulated MLE, since  $f_t$  can be generated for any given values of the GARCH parameters, and this would avoid the two step procedure.

Although not explicitly taking the viewpoint adopted here, one can view Schwert and Seguin's work in this way (Schwert and Seguin, 1990). In their case  $x_t$  was a vector of monthly returns to different portfolios of stocks, and the common factor was taken as the market return, i.e., they set

$$\sigma_{ijt} = \alpha_{0ij} + \alpha_{1ij}\sigma_t + \alpha_{2ij}\sigma_t^2, \quad (63)$$

where  $\sigma_t^2$  is the conditional variance of the market return, measured in their case by the volatility of daily returns to the S&P composite portfolio from 1928 to 1986. Thus, in this instance, the factor is rendered observable by the use of daily data, although it could have also been estimated by (say) making  $\sigma_t^2$  a GARCH(1,1) and using the monthly data on the aggregate portfolio. Schwert and Seguin also proposed that  $\sigma_{ijt}$  be replaced by  $\rho\sigma_{ii,t}^{1/2}\sigma_{jj,t}^{1/2}$ , i.e., the equivalent of what Bollerslev does in the general GARCH model. One can estimate (63) by OLS replacing  $\sigma_{ijt}$  by  $x_{it}x_{jt}$ , although one needs to correct for heteroskedasticity in the errors when making inferences about the  $a$ 's.<sup>57</sup> They find that small firm portfolios are four times more sensitive to market volatility changes than are large firm portfolios.

A generalization of this idea of making  $\Omega_t$  a function of the variances of factors which are portfolios constructed from the assets in  $x_t$  is available in Engle et al. (1990b), where the model is termed Factor-GARCH. Defining  $f_{kt} = \gamma_k' x_t$  as the  $k$ th portfolio, let  $\sigma_{f_{kt}}^2$  be its conditional variance. Engle et al. assume that  $\Omega_t = \sum_{k=1}^K \delta_k \delta_k' \sigma_{f_{kt}}^2 + \Omega$ , an obvious extension of the APT, and then allow each of the factor conditional variances to be GARCH(1,1), i.e.,

$$\sigma_{f_{kt}}^2 = \alpha_{0k} + \alpha_{1k}(\gamma_k' e_{t-1})^2 + \beta_{1k}\sigma_{f_{kt-1}}^2. \quad (64)$$

---

<sup>57</sup>  $\sigma_{ijt}$  is replaced by the residuals  $\hat{e}_{it}\hat{e}_{jt}$  if there are variables  $w_t$  determining  $x_t$ .

After substituting out  $\sigma_{fki}^2$  in the expression for  $\Omega_i$ , one is left with

$$\Omega_i = C + \sum_{k=1}^K \delta_k \delta'_k \alpha_{1k} (\gamma'_k e_{t-1})^2 + \sum_{k=1}^K \beta_{1k} \delta_k \delta'_k \gamma'_k \Omega_{t-1} \gamma_k. \quad (65)$$

In their terms this would be a “univariate portfolio representation” since  $\sigma_{fki}^2$  depends only upon its own lagged value and a combination  $(\gamma'_k e_{t-1})^2$  of the errors associated with its own portfolio. A “recursive portfolio representation” would allow some dependence across portfolios; as the name suggests, the conditional variances of portfolios are assumed to be ordered so that the dependence is upon information about other portfolios further down in the chain but not upon information above that point.

Engle et al. apply the idea to excess returns on stocks and Treasury Bills with maturities ranging from one to twelve months. Principal component analysis of the data led them to the idea of having two portfolios as the factors – one comprising stocks, and the other a linear combination of Treasury Bills (equally weighted). The recursive portfolio representation in which stocks affect bills but not conversely seemed superior to the univariate one. In this work the portfolio weights are given and not estimated. Lin (1992) considers full maximum likelihood estimation with unknown weights.

In all the above analysis, it was assumed that the object of interest was to model the complete conditional covariance matrix. Sometimes, however, one might only be interested in the diagonal elements (or at least be prepared to concentrate upon them alone). This occurs in the Engle et al. (1990a) study of the effects of news upon return volatility as it is transmitted around the world's stock markets. Thus, the vector of returns might be returns on the New York, Milan, London, and Tokyo markets. Here a subtle modification needs to be made to the specification of ARCH models to reflect the fact that, when New York opens, London and Milan have already been open for a number of hours. Hence, if the model is one of intra-day volatility,  $\sigma_t^2$  for New York should be a function of not only past returns, but also current returns in London and Milan, i.e., one would have an ARCH specification for the  $i$ th market (out of  $n$ ) as

$$\sigma_{ii}^2 = \sigma_i^2 + \sum_{j=1}^{i-1} \alpha_{ij} e_{ji}^2 + \sum_{j=i}^n \alpha_{ij} e_{ji-1}^2 + \beta_{ii} \sigma_{ii-1}^2. \quad (66)$$

In addition, one might also be interested in the behavior of covariances, particularly given the belief that stock markets are now much more closely related than they were previously, but it would be a formidable task to allow these to vary as well.

All of the above represents generalizations of the simplest GARCH processes to multivariate models. There have been few attempts to estimate more complex models. Braun et al. (1992) consider a bivariate EGARCH model when estimating betas for stocks (as described earlier) but that is a rare exception. Multivariate

stochastic volatility models have been proposed and estimated. The extension to the multivariate case is conceptually quite easy since one is simply dealing with a VAR rather than an AR process in the conditional variance. Harvey and Shephard (1993) point out that the quasi-MLE can be applied easily as the Kalman filter is already formulated for multivariate problems. Danielsson (1993) extends the accelerated importance sampler to perform simulated MLE and applies the method to a bivariate system of exchange rates. Gallant et al. (1994) estimate a stochastic volatility trivariate system with an indirect estimator and an SNP multivariate density as the auxiliary model.

#### 4.5. Are variances co-persistent?

Many financial series possess the property of integration and are also co-integrated. Integration can be interpreted as that characteristic wherein shocks to a series are permanent, or in which a perturbation in the initial condition never dies out of the series. Thus, for the AR(1) in (4),  $\partial y_t / \partial y_0 = \beta_1'$  for any given sequence of shocks  $\{u_j\}_{j=1}^t$ . When  $\beta_1 = 1$  any changes to the initial condition are permanently embedded in  $y_t$ . In contrast, for an  $I(0)$  series, shocks are not persistent. Therefore, if a co-integrating error is formed by weighting each series by their parameter value in the cointegrating vector, this error is  $I(0)$  by definition, and shocks to it will not be persistent even though they are to the individual series from which it is constituted.

Effectively, cointegration is a statement about the conditional means of variables. Hence it is not surprising that a similar distinction regarding the impact of shocks has been proposed by Bollerslev and Engle (1993) for variances. Thus, suppose that  $y_t$  was GARCH(1,1), i.e., its conditional variance  $\sigma_t^2$  is that given in (13). Iterating that equation backward in time

$$\sigma_t^2 = [\prod_{j=1}^{t-1} (\beta_1 + \alpha_1 \epsilon_{t-j}^2)] \sigma_0^2 + (\beta_1 + \alpha_1)' \alpha_0, \quad (67)$$

and this relation enables one to assess the impact of a change in  $\sigma_0^2$  on  $\sigma_t^2$ . One difficulty is that the impact is stochastic and so there is no unique way of measuring it. A simple measure would be just to consider  $E(\partial \sigma_t^2 / \partial \sigma_0^2) = E[\prod_{j=1}^{t-1} (\beta_1 + \alpha_1 \epsilon_{t-j}^2)] = [\prod_{j=1}^{t-1} (\beta_1 + \alpha_1)]$ .<sup>58</sup> When the process is IGARCH,  $\beta_1 + \alpha_1 = 1$ , and the shock is deemed persistent into  $\sigma_t^2$ , whereas if it is just GARCH there is no such persistence.

With the ideas relating to persistence in variance just stated, it is natural to define co-persistence in variance as occurring if individual series are IGARCH but if there exists some linear combination of them which is GARCH. The terminology is a useful one, as there do seem to be instances in which this phenomenon

<sup>58</sup> Bollerslev and Engle adopt a different definition involving the stochastic behavior of predictions of  $\sigma_t^2$  as the prediction horizon lengthens, and the current definition was chosen for its simplicity.

occurs, e.g., spot and forward exchange rates may each have IGARCH in the variance, but the difference between them, the risk premium, may just display GARCH effects. It cannot be emphasized too strongly that there may be no relation between the vectors that co-integrate the levels of series and those that “co-integrate” the variances. Indeed it is quite possible for two financial series to be both  $I(0)$ , each featuring IGARCH variances, and yet for the difference (say) to display only GARCH. Moreover, although co-persistence in variance may be an interesting empirical phenomenon, it does not seem to have any connection with theoretical models in the same way that co-integration does. There we have the notion that the co-integrating vector reflects an equilibrium relation from static theory; this idea deriving from the belief that deviations from an equilibrium should be stationary. However, even if we had a model indicating an equilibrium relation between variances, the fact that  $\sigma_t^2$  remains an  $I(0)$  process even if it is IGARCH means that no argument can be made in favor of the combined variances being GARCH.

## 5. Economic models of multivariate financial data

In the previous section statistical models to describe financial data have been outlined. Ultimately, however, one seeks some economic explanation of the inter-relationships, and various models have been proposed for this task. In some instances, these involve considering what the first order conditions for optimizing consumers and producers would be either in a partial or a general equilibrium framework, followed by a derivation of the implications of those conditions for relationships between financial series. In others, the optimality conditions come from general arguments considering the relation between the price of an asset and the flow of income that is associated with it. It is the latter which is the focus of the next sub-section, and the former is the concern of the following one. Mostly, the literature in this area has been concerned with testing whether the conditions for inter-temporal optimization hold rather than trying to calibrate the underlying models, and that explains the emphasis given to this question in what follows. In those cases where there is interest in estimating the parameters of an underlying model, estimation has proceeded with one of the techniques outlined earlier, e.g., Chen and Scott (1993) apply maximum likelihood to estimate a multifactor model of the term structure. Perhaps the most interesting development has been the use of indirect estimation for this purpose – see Bansal et al. (1994). There are many models in economics that are easy to simulate but for which it is difficult to write down the likelihood, e.g., models of exchange rates with “bands”, see Ball and Roma (1994), and these are obvious candidates for indirect estimation methods, using as auxiliary models the statistical structures described in the previous section.

### 5.1. Inter-temporal relations from equilibrium conditions

From discounted present value theory it would be expected that the *beginning of period* stock prices ( $P_t$ ) would be related to dividends ( $D_t$ ) by

$$P_t = \sum_{j=0}^{\infty} \beta^j E_t(D_{t+j}), \quad (68)$$

where  $\beta$  is an ex-ante constant real discount rate and  $E_t$  indicates that the expectation is taken conditional on information available at time  $t$  (this does not include  $D_t$ ). Defining  $S_t = P_t - \theta D_t = \sum_{j=0}^{\infty} \beta^j (E_t(D_{t+j}) - D_t)$ , where  $\theta = 1/1 - \beta$ ,  $S_t$  will be  $I(0)$  if  $D_t$  is an integrated process, and therefore  $(1 - \theta)$  would be a co-integrating vector between  $P_t$  and  $D_t$ . When tests for co-integration between  $P_t$  and  $D_t$  are applied, one frequently sees acceptance of the null hypothesis of no-cointegration. Timmermann (1995) notes that the discount rate may vary over time, and therefore  $\theta_t = \bar{\theta} + \nu_t$  would be a random variable. The relationship between stock prices and dividends would then be  $P_t - \bar{\theta} D_t$ , and this error would equal  $\phi_t = \nu_t D_t$ . As  $D_t$  is an  $I(1)$  process the  $\phi_t$  will have a conditional (upon  $D_t$ ) variance that depends upon  $D_t$ , i.e., there is a “levels” effect similar to that seen with yields. Timmermann constructs simulation experiments to show that the rejection of co-integration might simply stem from the fact that  $\phi_t$  has substantial persistence in it due to the documented fact that  $\nu_t$  depends on the dividend/price ratio, which tends to be quite persistent.

Earlier work by Shiller (1981) sought to determine the validity of (68) by considering its implications for sample moments. In particular, he defined an ex-post rational price  $P_t^*$  as

$$P_t^* = \sum_{j=0}^{\infty} \beta^j D_{t+j}. \quad (69)$$

Then

$$P_t^* = \sum_{j=0}^{\infty} \beta^j E_t(D_{t+j}) + \sum_{j=0}^{\infty} \beta^j (D_{t+j} - E_t(D_{t+j})) \quad (70)$$

$$= P_t + \sum_{j=0}^{\infty} \beta^j V_{t+j}, \quad (71)$$

and, following from the fact that  $E_t(V_{t+j}) = 0$ ,  $E(P_t V_{t+j}) = 0$ , producing

$$\text{var}(P_t^*) \geq \text{var}(P_t). \quad (72)$$

Shiller took this variance bounds inequality and estimated both sides from sample

data. The LHS poses difficulties since it depends upon dividends an infinite distance into the future, so he truncated this as  $P_t^{**} = \sum_{j=0}^{T-1} \beta^j D_{t+j} + \beta^T P_T$ , i.e., instead of  $P_t^*$  being used to summarize future information he replaced it with the observed quantity  $P_T$ . He then discovered that the sample variance of  $P_t$  exceeded that of  $P_t^{**}$ .

An immediate problem with this test, noted by Kleidon (1986), was that  $P_t$  was very likely to be an integrated process so that the “variance” does not exist. Hence, a comparison such as (72) can only be done with  $P_t$  being a non-integrated process. This has led to other proposals for checking the validity of (68). One of these is to recognize that  $P_t$  and  $D_t$  should be co-integrated with vector  $(1 - \theta)$  (see the discussion in (a) above) so that if we formed

$$S_t^* = P_t^* - \theta D_t = \sum_{j=0}^{\infty} \beta^j D_{t+j} - \theta D_t \quad (73)$$

$$= \sum \beta^j (D_{t+j} - D_t), \quad (74)$$

$S_t^*$  will be an  $I(0)$  process when  $D_t$  is integrated. Comparing this to  $S_t = \sum_{j=0}^{\infty} \beta^j (E_t(D_{t+j}) - D_t)$ , shows that a prediction of the present value model is that

$$\text{var}(S_t^*) \geq \text{var}(S_t), \quad (75)$$

and a comparison can be made using the sample variances owing to the fact that  $S_t^*$  and  $S_t$  are both  $I(0)$ .<sup>59</sup> However, even if these variables are  $I(1)$  there are a number of factors that make an empirical comparison of the population quantities in (75) much more clouded. Flavin (1983) pointed out that there are small sample biases in estimating variances if the series have substantial serial correlation. Indeed that must be the case for  $S_t^*$  as it involves an infinite combination of  $\Delta^j D_{t+j}$ . Moreover, as Shea (1989) shows in some simulation experiments, the comparison can be a problem if  $\Delta D_t$  is close to being integrated: For such near integrated processes statistics such as sample variances frequently behave more like  $I(1)$  than  $I(0)$  series if the sample is small. Shea presents some Monte Carlo evidence to show that the sample variance of  $S_t$  can easily exceed that of  $S_t^*$  even if the present value model is correct. Actually, what Shea examines is the properties of the Mankiw et al. (1985) statistic which involves comparing  $\text{var}(P_t^* - P_t^0)$  to  $\text{var}(P_t - P_t^0)$ , where  $P_t^0$  is a “naive” forecast. However, as they set  $P_t^0$  to  $\theta D_t$ , the experiment is relevant. Another variant is LeRoy and Parke (1987) who compare the  $\text{var}(\bar{S}_t^*)$  to the  $\text{var}(\bar{S}_t)$ , where  $\bar{S}_t^* = S_t^*/D_t$  and  $\bar{S}_t = S_t/D_t$ .

<sup>59</sup> Of course  $S_t^*$  is replaced by  $S_t^{**}$  but, as  $S_t^{**} = S_t^* + \beta^T (P_t - P_t^*)$ , and  $P_t - P_t^*$  should be  $I(0)$  when the present value model holds, that substitution should not be of concern, although the actual computed sample variance can be different if we decide to compute the ex-post rational price using a terminal date that is within the sample. In the remaining discussion we focus on  $S_t^*$  only.

Because  $\bar{S}_t^* = (P_t^*/D_t) - \theta$  and  $\bar{S}_t = (S_t/D_t) - \theta$ ,  $\theta$  does need not to be estimated in their comparison. Generally, one has to conclude that these relative variance tests are very inefficient ways of examining the validity of a present value relation, and that direct testing of the model seems a more sensible strategy.

Another variance bounds test that does not suffer from problems arising out of integrated processes has been developed by West (1988). He modifies (68) to

$$P_t = E_t \left[ \sum_{j=1}^{\infty} \beta^j D_{t+j} \right], \quad (76)$$

where  $P_t$  is now an *end-of-period* price rather than a beginning-of-period price. (76) implies that

$$P_t = \beta E_t(D_{t+1} + P_{t+1}), \quad (77)$$

where the appropriate transversality condition is assumed satisfied. Now, suppose that the information set used in (76) is  $I_t$  and that there is a sub-set of it designated  $G_t$ . Constructing the hypothetical quantities

$$x_{I_t} = E \left[ \sum_{j=0}^{\infty} \beta^j D_{t+j} \mid I_t \right], \quad x_{G_t} = E \left[ \sum_{j=0}^{\infty} \beta^j D_{t+j} \mid G_t \right], \quad (78)$$

he shows that, under the present value model,

$$\text{var}(x_{G_t} - E(x_{G_t} \mid G_{t-1})) \geq \text{var}(x_{I_t} - E(x_{I_t} \mid I_{t-1})). \quad (79)$$

He proposes that both sides of the inequality be estimated from sample data and that the present value model be rejected if the inequality fails to hold.

The main issue is how to construct sample estimates of both variances. From inspection of (78) and (76),  $x_{I_t} = D_t + P_t$ , and so the RHS of (79) is  $\text{var}(D_t + P_t - E(D_t + P_t \mid I_{t-1}))$ . From (77) we can write

$$P_t = \beta(D_{t+1} + P_{t+1}) - \eta_t, \quad (80)$$

where  $\eta_t = \beta(D_{t+1} + P_{t+1} - E(D_{t+1} + P_{t+1} \mid I_t))$  has the property that  $E(\eta_t \mid I_t) = 0$ . Lagging  $\eta_t$  once shows that its variance is the variance on the RHS of (79). Hence, it is simply a matter of determining the variance of  $\eta_t$  from (80). By definition  $E(\eta_t \mid I_t) = 0$

$$\therefore E[P_t - \beta(D_{t+1} + P_{t+1}) \mid I_t] = 0 \quad (81)$$

$$\text{or } E[(1 - \beta(D_{t+1} + P_{t+1})) / P_t \mid I_t] = 0 \quad (82)$$

as  $I_t$  includes  $P_t$  in the information set.

The method of moments estimator of  $\beta$  from (82) is just  $\hat{\beta}$  such that

$$\Sigma(1 - \hat{\beta}[(D_{t+1} + P_{t+1})/P_t]) = 0 \quad (83)$$

or  $\hat{\beta} = 1/T^{-1}\Sigma[(D_{t+1} + P_{t+1})/P_t]$ . As the denominator is one plus the sample mean of  $\{(D_{t+1} + \Delta P_{t+1})/P_t\}$ , this will give an estimated discount rate approximately equal to 1 minus the sample mean of  $(D_{t+1} + \Delta P_{t+1})/P_t$ , i.e., one minus the mean of ex-post returns.<sup>60</sup> Since  $(D_{t+1} + \Delta P_{t+1})/P_t$  will be  $I(0)$ ,  $\hat{\beta}$  should therefore be consistently estimated in this way. Thereupon the variance of  $\eta_t$  can be estimated as  $T^{-1}\Sigma(P_t - \hat{\beta}(D_{t+1} + P_{t+1}))^2$ .

To determine the LHS of (79) requires that the information set  $G_t$  be specified and a model for the dividend process be set up. Suppose  $\Delta D_t = \epsilon_t$  is white noise and  $G_t$  contains  $D_{t-1}$ . Then  $x_{G_t} = (1 - \beta)^{-1} D_t$ ,  $E(x_{G_t} | G_{t-1}) = (1 - \beta)^{-1} D_{t-1}$  and the LHS is just  $\text{var}((1 - \beta)^{-1}(D_t - D_{t-1})) = (1 - \beta)^{-2} \text{var}\Delta D_t$ , from which an estimate can be constructed using  $\hat{\beta}$  and the sample variance of  $\Delta D_t$ . Since both quantities involve estimating  $\beta$ ,  $\text{var}(\Delta D_t)$  and  $\text{var}(\eta_t)$  by method of moments, a standard error can be placed on the difference between the estimated  $\text{var}(x_{G_t} - E(x_{G_t} | G_{t-1}))$  and  $\text{var}(x_{I_t} - E(x_{I_t} | I_{t-1}))$ , although this does not provide a solution to the problem of testing an inequality. With data like Shiller's on stocks and dividends the inequality was clearly rejected.

Restricted VAR's have been used to test other aspects of models of financial data. For example, if we return to the question of the present value pricing model for stocks or the term structure of interest rates, we have observed that the relationship should exhibit a co-integrating vector. Concentrating on stocks and dividends,  $S_t = P_t - \theta D_t = \sum_{j=0}^{\infty} \beta^j E_t \Delta D_{t+j}$ , and if we form a VAR from  $S_t$  and  $\Delta D_t$ , restrictions are implied upon its parameters. In particular, writing it as (56), where  $z'_t = (S_t \ \Delta D_t)$ , the restrictions on a  $p$ th order VAR would be that  $e'_{p+1} = e'_1 \beta A(I - \beta A)^{-1}$ , where  $e_j$  is a vector with unity in the  $j$ th position and zeros elsewhere. One can test these restrictions by doing a Wald test. A better way to test the restrictions would be to test  $e'_{p+1}(I - \beta A) = e'_{p+1} \beta A$ , as this is testing a set of linear rather than non-linear restrictions and Wald tests have better sampling properties when restrictions are close to linear – see Gregory and Veall (1985) and Phillips and Park (1988). Such tests have been used previously by Campbell and Shiller (1987) and Kearns (1990a) in studying the term structure relations. They find that the restrictions are generally rejected.

<sup>60</sup> West suggests using other instruments,  $w_t$ , to estimate  $\beta$ , but if they are to be instruments they must be uncorrelated with  $1 - \beta[(D_{t+1} + P_{t+1})/P_t]$ , i.e.,  $E[w_t(1 - \beta[D_{t+1} + P_{t+1}/P_t])] = 0$ . Since there is a constant being used,  $w_t$  can be assumed to have  $E(w_t) = 0$ , so that this condition requires  $\beta E[w_t(D_{t+1} + P_{t+1})/P_t] = 0$ . However, this means there is no information in  $w_t$  for  $\beta$  as the derivative of  $[1 - \beta(D_{t+1} + P_{t+1})/P_t]$  with respect to  $\beta$  has to be correlated with the instrument for the latter to be relevant. As discussed earlier in connection with the GMM estimator the use of such weak instruments can have deleterious effects upon the properties of the instrumental variables estimator.

If one observes rejection the question of what caused the rejection arises. One candidate is that there are speculative bubbles.<sup>61</sup> A second is that the approximations employed to get relations such as the expectations hypothesis  $R_t = k^{-1} \sum_{j=0}^{k-1} E_t r_{t+j}$ , where  $R_t$  and  $r_t$  are respectively the long and short term interest rates, are invalid; in particular they involve the linearization around  $(1 + \bar{R})^{-1}$ , where  $\bar{R}$  is taken to be the monthly average yield on long bonds. When yields follow an integrated process  $\bar{R}$  will no longer converge to a constant so that the linearization of the holding yield of an  $n$ -period bond paying a coupon  $C$ ,

$$h_t^n = \left\{ C + \frac{C}{R_{t+1}^{n+1}} + \frac{R_{t+1}^{n-1} - C}{R_{t+1}^{n-1} [1 + R_{t+1}^{n-1}]^{n-1}} \right\}^{-1} \left\{ \frac{C}{R_t^n} + \frac{R_t^n - C}{R_t^n [1 + R_t^n]^n} \right\} - 1$$

– see Shiller (1979) – is being performed around a stochastic quantity. If the process is close to being integrated the same problem is likely to arise unless the sample size is very large. Even if  $\bar{R}$  can be taken as a constant the linearization would fail if bond yields deviated substantially from  $\bar{R}$ . It has generally been argued that the linearization is a good approximation, but this is not sufficient for estimation purposes. In a regression, any approximation error in the conditional mean goes into the regression error and generally causes a correlation between regressors and errors, thereby inducing inconsistencies into parameter estimators. In turn, these may translate into rejections of hypotheses. Indeed this seems to be true here. Both Kearns (1990a) and Shea (1992) simulate data from models in which the present value model holds but find substantial rejections of it. Whether this is due to poor sampling properties of the test statistics or the linearization error remains to be determined. However, it seems more likely to be the latter. Following the Abel–Mishkin approach, Campbell and Shiller (1987, p. 1068) show that the linear restrictions on the VAR can be tested by regressing  $z_i = (\Delta r_i + S_i - (S_{i-1}/\beta))$  on  $\Delta r_{i-1}$  and  $S_{i-1}$ , and testing if the effects of these variables are zero. Consequently, the Wald test effectively involves adding variables to a

---

<sup>61</sup> There is another literature that treats the presence of bubbles as arising from a failure of the transversality condition  $\lim_{j \rightarrow \infty} \beta^j E_t(P_{t+j+1}) = 0$  to hold, so that the solution to the forward difference equation underlying (68),  $P_t = \beta E_t(P_{t+1}) + D_t$ , is not unique and can be augmented with any process satisfying  $B_{t+1} = \beta^{-1} B_t + \xi_{t+1}$ , where  $E_t(\xi_t) = 0$ . This results in a failure of co-integration between  $P_t$  and  $D_t$  – see e.g., Hamilton and Whiteman (1985). In fact, if the bubble is an integrated process, the Wald test based on the VAR would not be a consistent test for a rational bubble as the test statistic can be shown to have a limiting distribution under the alternative and so will not reject the null with probability one. The main complication with this view is the difficulty of discriminating between a bubble and a mis-specification of the fundamentals process as the source of a lack of co-integration. Durlauf and Hall (1989) have made some progress in choosing between the two possibilities.

linear regression, albeit with predetermined rather than strongly exogenous regressors, and therefore one would not expect such a test to exhibit the level of biases observed in the simulations (rejection rates of 25.3% when the nominal 10% significance levels are used).<sup>62</sup>

### 5.2. Intertemporal models with varying IMRS

Perhaps the most distinctive feature of financial economics has been the development of theories of asset prices that are capable of being tested with data. Of these the most famous would have to be the Sharpe/Lintner Capital Asset Pricing Model. Stochastic intertemporal asset pricing theories have been developed by a number of authors, and a particularly influential one was Breeden (1979). The central implication of these theories was that the first order conditions for an intertemporally optimizing consumer were

$$E_t[(\beta u'(C_{t+1})/u'(C_t))R_t] = 1, \quad (84)$$

where  $u'(C_t)$  is the marginal utility from consumption,  $R_t$  is the gross return  $(p_{t+1} + d_t)/p_t$ , where  $p_t$  is the real price of the asset,  $d_t$  are real dividends and  $E_t$  denotes that the expectation is taken with respect to information at time  $t$ . This model is also sometimes referred to as the consumption – CAPM as it emphasizes that returns are a function of the covariance of the intertemporal rate of substitution (IMRS) – “stochastic discount factor” –  $\psi_t = \beta u'(C_{t+1})/u'(C_t)$  with the return  $R_t$ .<sup>63</sup> To see this write  $E(\psi_t R_t) = \text{cov}(\psi_t R_t) - E(R_t)E(\psi_t)$ , so that (84) implies  $E(R_t) = \text{cov}(R_t \psi_t)/E(\psi_t)$ , and it is the covariance with  $\psi_t$  rather than the market portfolio which is important for asset prices (although if one follows Campbell (1992) and log-linearizes the consumer's budget constraint to find  $E_t(\Delta c_{t+1}) = \mu_m + E_t(r_{m,t+1})$ , where  $c_t = \log(C_t)$ , it is clear that there is implicitly a relation with the market return,  $r_{m,t}$ ).

Under a special form for the utility function it is possible to write out a specific set of restrictions that are implied by (84). For example, if  $u(C_t)$  has the constant relative risk aversion (CRAA) form  $\gamma^{-1} (C_t^\gamma - 1)$ , then (84) becomes

$$E_t[\beta(C_{t+1}/C_t)^\alpha R_t] = 1, \quad (85)$$

where  $\alpha = \gamma - 1$ . Hansen and Singleton (1982) suggested that one estimate the

<sup>62</sup> Of course if  $z_t$ ,  $S_t$  and  $\Delta r_t$  were near integrated, such over-rejection would be predicted.

<sup>63</sup> If  $\psi_t$  is a constant and  $d_t = 0$  one gets the implication that returns are unpredictable.

parameters of (85) by GMM, using as “instruments”  $z'_t = (1 \ w'_t)$ , the information being conditioned upon in  $E_t$ , leading to the moment restrictions

$$E[\beta(C_{t+1}/C_t)^\alpha R_t] - 1 = 0 \quad (86)$$

$$E[\beta w_t(C_{t+1}/C_t)^\alpha] = 0, \quad (87)$$

where we can assume without loss of generality that  $E(w_t) = 0$  due to the presence of unity in  $z_t$ . Replacing the population mean in (86) by its sample value produces a method of moments estimator of  $\beta$  of  $\hat{\beta} = 1/(T^{-1}\sum(C_{t+1}/C_t)^\alpha R_t)$ , where  $\hat{\alpha}$  is some estimator of  $\alpha$  from (87), making  $\hat{\beta}$  just the inverse of a sample mean.

This estimation method has been applied to a wide variety of asset prices, both domestic and international. Because, at heart, it involves a GMM estimator, there are potential problems associated with it, which were outlined in an earlier section. As explained there, these stem from the possibility that  $w_t$  is a poor instrument for  $\partial v_t/\partial\theta$ , where  $v_t = \beta(C_{t+1}/C_t)^\alpha R_t - 1$  and  $\theta' = (\beta \ \alpha)$ , and this would be assessed by regressing  $\partial v_t/\partial\theta$  against  $w_t$ . A small value of the  $R^2$  from this regression signals likely poor performance of the GMM estimator of  $\theta$ . Unfortunately,  $\partial v_t/\partial\theta$  depends upon  $\theta$ , leading to a circularity in the argument, in that accurate computation of the  $R^2$  etc. depends upon a good estimate of  $\theta$ , and this will be poor if the  $R^2$  is low (the situation one is attempting to detect!). Nevertheless, it would seem useful to compute measures such as the  $R^2$  using the point estimates of  $\theta$ , as that should provide some insights into possible problems with the GMM estimator.

Returning to the asset pricing model in (86) and (87), Hansen and Singleton (1982) fitted this with various instruments chosen from lags of consumption growth and gross returns. As an exercise, we set  $w_t$  to unity,  $(C_t/C_{t-1})$  and  $(C_{t-1}/C_{t-2})$  and take  $R_t$  to be the gross market return on equities. This produces GMM estimates for  $\beta$  and  $\alpha$  of  $\hat{\beta} = 0.998$  and  $\hat{\alpha} = -1.02$  (assuming that  $v_t$  is an MA(2)). Now  $\partial v_t/\partial\beta = (C_{t+1}/C_t)^\alpha R_t$ , and  $\partial v_t/\partial\alpha = \beta \log(C_{t+1}/C_t)(C_{t+1}/C_t)^\alpha R_t$ , and these derivatives (evaluated at the GMM estimates of  $\beta$  and  $\alpha$ ) can be regressed against the instruments, resulting in  $R^2$  of 0.007 and 0.13 respectively. These results demonstrate that one would need to be very careful when using the GMM estimates, although the very tight distribution of the squared instrument for  $\beta$  around its mean (the mean is 1.0015 and the standard deviation is 0.007) suggests that this parameter should be quite accurately estimated, a result in accord with existing simulation work – see Mao (1990).

If the  $\dim(w_t) > \dim(\theta)$  it is possible to test whether the moment conditions for inter-temporal optimization are satisfied using the J-test set out in Hansen (1982). Mostly, applications of this test have led to rejection of the model. There are a number of possible reasons for this. One is that the over-rejection is caused by the poor sampling distributions of the GMM estimator of  $\theta$ , as this has been observed in some simulation work, e.g., Mao (1990). Another is that the specifications used for the utility function are incorrect, and this has led to a literature exploring

generalized forms, e.g., the introduction of inertia into the utility function as in Constantinides (1990).

Instead of performing a specification test directly upon the data a popular way of assessing the quality of a model has been to utilize “Hansen–Jagannathan bounds”. Defining  $\rho$  as the correlation between  $R_t$  and  $\psi_t$ , we have

$$\begin{aligned}\rho^2 &= [\text{cov}(R_t, \psi_t)]^2 / \sigma_R^2 \sigma_\psi^2 \\ &= [E(R_t \psi_t) - E(R_t) E(\psi_t)]^2 / \sigma_R^2 \sigma_\psi^2 \\ &= [1 - E(R_t) E(\psi_t)]^2 / \sigma_R^2 \sigma_\psi^2,\end{aligned}$$

if the asset pricing condition (84) holds.<sup>64</sup> Inverting this relation and recognizing that the maximum value of  $\rho^2$  is unity gives

$$\sigma_\psi^2 \geq [1 - E(R_t) E(\psi_t)]^2 / \sigma_R^2, \quad (88)$$

and this is the bound provided by Hansen and Jagannathan (1991) (they work with a vector of gross returns so that a matrix representation of the above is needed). Using the equality one can trace out a relation between  $E(\psi_t)$  and the minimum value of  $\sigma_\psi^2$ . For any given candidate for  $\psi_t$ , e.g.,  $\psi_t = \beta(C_{t+1}/C_t)^\alpha$ , it is possible to derive estimates of  $E(\psi_t)$  and  $\sigma_\psi^2$  and to see if the bounds are satisfied. One nice feature of this approach is that it enables one to perform a sensitivity analysis with respect to parameters such as  $\alpha$ , i.e., one asks what value  $\alpha$  would have to take in order to make the bound hold. If one does not wish to perform such a sensitivity analysis then it would be necessary to take account of the fact that the point estimates of the left and right hand sides of (88) are both random variables and one might therefore wish to attach some probability statement to whether the bounds are violated. Burnside (1994) and Cecchetti et al. (1994) do this using the  $\delta$  method of deriving asymptotic distributions of non-linear functions of random variables. One might well ask what the advantage of using these bounds as a formal specification test is over the direct test of the over-identifying restrictions. Sometimes the latter may not be available as  $\dim(w_t) = \dim(\theta)$  but, when it is, the simulation studies in Burnside (1994) suggest that it is the best test, although that it could over-reject in some cases as the true values of  $\beta$  and  $\alpha$  changed, and this is most likely due to the changing quality of the instruments.

## 6. Conclusion

A striking feature from this survey is how a purely statistical modeling approach has been dominant in the analysis of financial data, particularly when the

---

<sup>64</sup> The conditional expectation has been replaced by an unconditional expectation using the law of iterated expectations. Gallant et al. (1990) use the conditional expectations, estimating these with SNP densities.

focus is a univariate one. One cannot help but feel that these statistical approaches to the modeling of financial series have possibly reached the limits of their usefulness. The models are having to be made increasingly complex so as to capture the nature of the conditional density of returns. Initially, it looked as if the ARCH class of processes would provide a simple explanation, but, as documented in these notes, it is becoming clear that this is not so. Ultimately one must pay more attention to whether simple economic models are capable of generating the complex behavior that is evident, i.e., is it possible to construct economic models that might be useful in explaining what is observed? To date progress in this area has been slow. Few models are capable of generating the type of ARCH one sees in the data. The same can be said about present value models. To find the type of dependence observed in returns one would need a very strong dependence in the dividends process (or endowments if one is looking at general equilibrium methods of generating asset prices), and the evidence on this is weak. Most of these studies are best summarized with the adage that “to get GARCH you need to begin with GARCH”. Nevertheless, the search for models that will take as inputs moderate degrees of volatility and accentuate it has to be accorded a high priority if we are to make further progress in the modeling of this aspect of financial series. An interesting paper that does just this is Den Haan and Spear (1994) who manage to produce GARCH in interest rates by allowing for income distribution effects changing with the cycle.

The situation with multivariate modeling differs in that theoretical models have informed decisions about how to link the mean behavior of variables, although the explanation of co-movements in volatility is still the province of statistical models. To some extent this emphasis comes from the motivation for much empirical work in finance; if it works and one can make money from it then the lack of a theoretical base is not accorded much importance. For many economists however it is desirable to be able to understand the phenomena one is witnessing, and this is generally best done through theoretical models. Of course this desire does not mean that the search for statistical models which fit the data should be abandoned. One of the nice features of the indirect estimation approach discussed in Section 3.2 is the emphasis placed upon the use of a statistical model that fits data well as a way of estimating the parameters of an underlying theoretical model. It is this interplay between statistics and theoretical work in economics which needs to become dominant in financial econometrics in the next decade.

## 7. Other sources

Abel and Mishkin (1983, Chesney and Scott (1989, Clark (1973, Engle and Lee (1992, Fama (1976, French et al. (1987, Gallant and Tauchen (1990, Hansen and Singleton (1990, Kearns (1992, Nelson and Startz (1990, Newey and West (1987, Pagan (1975, Summers (1986, Tauchen (1986, Tauchen and Pitts (1983, Taylor (1990) and Watson (1964).

## Acknowledgements

I would like to thank Torben Andersen, Richard Baillie, Tim Bollerslev, Colin Cameron, Mardi Dungey, Frank Diebold, Phil Kearns, Walter Kramer, John Robertson, Allan Timmermann and anonymous referees for their comments on earlier versions of this survey.

## References

- Abel, A.B. and F.S. Mishkin, 1983, An integrated view of tests of rationality, market efficiency and short-run neutrality of monetary policy, *Journal of Monetary Economics* 11, 3–24.
- Albert, J. and S. Chib, 1993, Bayesian analysis via Gibbs sampling of autoregressive time series subject to mean and variance Markov shifts, *Journal of Business and Economic Statistics* 11, 1–15.
- Andersen, T.G., 1992, Volatility, Working paper No. 144, Kellogg Graduate School of Business, Northwestern University.
- Andersen, T. and B.E. Sørensen, 1994a, GMM estimation of a stochastic volatility model: A Monte Carlo study, Working paper No. 94–6, Brown University.
- Anderson, T.G. and B.E. Sørensen, 1994b, A note on Ruiz (1994): Quasi-maximum likelihood estimation of stochastic volatility models, Working paper No. 189, Kellogg Graduate School of Management, Northwestern University.
- Baillie, R.T., 1989, Tests of rational expectations and market efficiency, *Econometric Reviews* 8, 151–186.
- Baillie, R.T., 1996, Long memory processes and fractional integration in economics and finance, *Journal of Econometrics* (forthcoming).
- Baillie, R. and T. Bollerslev, 1989, Common stochastic trends in a system of exchange rates, *Journal of Finance* 44, 167–182.
- Baillie, R. and T. Bollerslev, 1994, The long memory of the forward premium, *Journal of International Money and Finance* 13, 565–572.
- Baillie, R.T., T. Bollerslev and H.O. Mikkelsen, 1996, Fractionally integrated autoregressive conditional heteroskedasticity, *Journal of Econometrics* (forthcoming).
- Baillie, R.T. and R.J. Myers, 1991, Bivariate GARCH estimation of the optimal commodity futures hedge, *Journal of Applied Econometrics* 6, 109–124.
- Ball, C.A. and A. Roma, 1994, Target zone modelling and estimation for European monetary system exchange rates, *Journal of Empirical Finance* 1, 385–420.
- Bansal, R., A.R. Gallant, R. Hussey and G. Tauchen, 1994, Nonparametric estimation of structural models for high frequency market data, *Journal of Econometrics* (forthcoming).
- Bekaert, G. and R.J. Hodrick, 1992, Characterizing predictable components in excess returns on equity and foreign exchange markets, *Journal of Finance* 47, 467–509.
- Beveridge, S. and C.R. Nelson, 1981, A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle, *Journal of Monetary Economics* 7, 151–174.
- Black, F., 1976, Studies in stock volatility changes, Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association, 177–181.
- Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., 1990, Modelling the coherence in short-run nominal exchange rates: A generalized ARCH model, *Review of Economic Statistics* 72, 498–505.
- Bollerslev, T. and R.F. Engle, 1993, Common persistence in conditional variances: Definition and representation, *Econometrica* 61, 167–186.

- Bollerslev, T. and R.J. Hodrick, 1992, Financial market efficiency tests, Working paper No. 132, Kellogg Graduate School of Management, Northwestern University.
- Bollerslev, T., R.Y. Chou and K.F. Kroner, 1992, ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* 52, 5–59.
- Bollerslev, T., R.F. Engle and D.B. Nelson, 1994, ARCH models, in: R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4 (North-Holland).
- Bollerslev, T.R., F. Engle and J. Wooldridge, 1988, A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116–131.
- Bossaerts, P. and P. Hillion, 1994, Local parametric analysis of hedging in discrete time (mimeo, Tilburg University).
- Bougerol, P. and N. Picard, 1992, Stationarity of GARCH processes and of some non-negative time series, *Journal of Econometrics* 52, 115–127.
- Braun, P.A., D.B. Nelson and A.M. Sunier, 1992, Good news, bad news, volatility and betas (mimeo, University of Chicago).
- Breeden, D.T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.
- Brenner, R.J., R.H. Harjes and K. Kroner, 1994, Another look at alternative models of the short-term interest rate (mimeo, University of Arizona).
- Brock, W., W.D. Dechert and J.A. Scheinkman, 1987, A test for independence based on the correlation dimension (mimeo, University of Wisconsin, Madison).
- Brockwell, P.J. and R.A. Davis, 1991, *Time series: Theory and methods*, 2nd ed. (Springer-Verlag, New York).
- Broze, L., O. Scaillet and J.M. Zakoian, 1993, Testing for continuous time models of the short-term interest rates, Discussion paper No. 9331, CORE.
- Burnside, C., 1994, Hansen–Jagannathan bounds as classical tests of asset pricing models, *Journal of Business and Economic Statistics* 12, 57–79.
- Cai, J., 1994, A Markov model of unconditional variance in ARCH, *Journal of Business and Economic Statistics* 12, 309–316.
- Cameron, A.C. and P.K. Trivedi, 1993, Tests of independence in parametric models with applications and illustrations, *Journal of Business and Economic Statistics* 11, 29–43.
- Campbell, J.Y., 1992, Intertemporal asset pricing without consumption data (mimeo, Princeton University).
- Campbell, J. and R.J. Shiller, 1987, Cointegration and tests of present value models, *Journal of Political Economy* 95, 1062–1088.
- Cecchetti, S.G., P-S Lam and N.C. Mark, 1994, Testing volatility restrictions on intertemporal marginal rates of substitution implied by Euler equations and asset returns, *Journal of Finance*, XLIX, 123–152.
- Chan, K.C., G.A. Karolyi, F.A. Longstaff and A.B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance*, XLVII, 1209–1227.
- Chen, R.R. and L. Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14–31.
- Chesney, M. and L.O. Scott, 1989, Pricing European currency options: A comparison of the modified Black–Scholes model and a random variance model, *Journal of Financial and Quantitative Analysis* 24, 267–284.
- Chib, S. and E. Greenberg, 1994a, Markov chain Monte Carlo simulation methods in econometrics (mimeo, Washington University).
- Chib, S. and E. Greenberg, 1994b, Understanding the Metropolis–Hastings algorithm (mimeo, Washington University).
- Christie, A.A., 1982, The stochastic behaviour of common stock variances: Value, leverage and interest rate effects, *Journal of Financial Economics* 10, 407–432.

- Christiano, L.J. and M. Eichenbaum, 1990, Unit roots in GNP: Do we know and do we care?, *Carnegie-Rochester series on public policy*, 32, Spring, 7–61.
- Chu, C-S.J., 1995, Detecting parameter shift in GARCH models, *Econometric Reviews* 14, 241–266.
- Clark, P.K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–155.
- Cochrane, J.H., 1988, How big is the random walk in GNP, *Journal of Political Economy* 96, 893–920.
- Constantinides, G.M., 1990, Habit formation: A resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519–543.
- Cox, J., J. Ingersoll and S. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.
- Crowder, W.J., 1994, Foreign exchange market efficiency and common stochastic trends, *Journal of International Money and Finance* 13, 551–564.
- Danielsson, J., 1993, Multivariate stochastic volatility (mimeo, University of Iceland).
- Danielsson, J., 1994, Stochastic volatility in asset prices: Estimation with simulated maximum likelihood, *Journal of Econometrics* 64, 375–400.
- Danielsson, J. and J.F. Richard, 1993, Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics* 8, S153–S174.
- Das, S.R., 1993, Jump-hunting interest rates (mimeo, New York University).
- De Jong and N. Shephard, 1993, Efficient sampling from the smoothing density in time series models (mimeo, Nuffield College).
- De Lima, P. and N. Crato, 1994, Long range dependence in the conditional variance of stock returns, *Economic Letters* 45, 281–285.
- De Lima, P. and F.J. Breidt and N. Crato, 1994, Modelling long-memory stochastic volatility (mimeo, Johns Hopkins University).
- Den Haan, W.J. and S. Spear, 1994, Credit conditions, cross-sectional dispersion and ARCH effects in a dynamic equilibrium model (mimeo, University of California, San Diego).
- Den Haan, W.J. and A. Levin, 1994, Inferences from parametric and non-parametric covariance matrix estimation procedures (mimeo, University of California, San Diego).
- Diebold, F.X., 1986, Modeling the persistence of conditional variances: A comment, *Econometric Reviews* 5, 51–56.
- Diebold, F.X., 1988, Empirical modeling of exchange rate dynamics, (Springer-Verlag, New York).
- Diebold, F.X. and J.A. Lopez, 1995, ARCH models, in: K. Hoover (ed.), *Macroeconomics: Developments, tensions and prospects* (Kluwer Publishing Co).
- Diebold, F.X. and M. Nerlove, 1989, The dynamics of exchange rate volatility: A multivariate latent-factor ARCH model, *Journal of Applied Econometrics* 4, 1–22.
- Diebold, F.X. and J. Nason, 1990, Nonparametric exchange rate prediction, *Journal of International Economics* 28, 315–322.
- Diebold, F.X. and G.D. Rudebusch, 1989, Long memory and persistence in aggregate output, *Journal of Monetary Economics* 24, 189–209.
- Diebold, F.X. and T. Schuermann, 1996, Exact maximum likelihood estimation of observation-driven econometric models, in: R.S. Mariano, M. Weeks and T. Schuermann (eds.), *Simulation-based inference in econometrics: Methods and applications*, (Cambridge University Press, Cambridge) (forthcoming).
- Ding, Z., C.W.J. Granger and R.F. Engle, 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83–105.
- Drost, F.C., and C.A.J. Klaassen, 1994, Adaptivity in semiparametric GARCH models (mimeo, University of Tilburg).
- Drost, F.C. and T.E. Nijman, 1993, Temporal aggregation of GARCH processes, *Econometrica* 61, 909–927.
- Duffie, D. and R. Kar, 1993, A yield-factor model of interest rates (mimeo, Graduate School of Business, Stanford University).

- Duffie, P. and K.J. Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929–952.
- Dufour, J.M. and M.L. King, 1991, Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or non-stationary AR(1) errors, *Journal of Econometrics* 47, 115–143.
- Durbin, J., 1959, Efficient estimation of parameters in moving-average models, *Biometrika* 46, 306–316.
- Durlauf, S.N. and R.E. Hall, 1989, A signal extraction approach to recovering noise in expectations based models (mimeo, Stanford University).
- Dybvig, P.H., 1989, Bonds and bond option pricing based on the current term structure, Working paper, Washington University (St. Louis).
- Elliott, G., T.J. Rothenberg and J.H. Stock, 1992, Efficient tests for an autoregressive unit root (mimeo, Harvard University).
- Engel, C. and J.D. Hamilton, 1990, Long swings in the exchange rate: Are they in the data and do markets know it, *American Economic Review* 80, 689–713.
- Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987–1008.
- Engle, R.F. and T. Bollerslev, 1986, Modelling the persistence of conditional variances, *Econometric Reviews* 5, 1–50.
- Engle, R.F. and G. Gonzalez-Rivera, 1991, Semiparametric ARCH models, *Journal of Business and Economic Statistics* 9, 345–360.
- Engle, R.F. and K.F. Kroner, 1995, Multivariate simultaneous generalized GARCH, *Econometric Theory* 11, 122–150.
- Engle, R.F. and G.G.J. Lee, 1992, A permanent and transitory component model of stock return volatility (mimeo, University of California at San Diego).
- Engle, R.F. and G.G.J. Lee, 1994, Estimating diffusion models of stochastic volatility (mimeo, University of California at San Diego).
- Engle, R.F. and V.K. Ng, 1993, Measuring and testing the impact of news on volatility, *Journal of Finance* 48, 1749–1778.
- Engle, R.F., T. Ito and W.L. Lin, 1990a, Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market, *Econometrica* 58 525–542.
- Engle, R.F., D.M. Lillien and R.P. Robins, 1987, Estimating time varying risk premia in the term structure, *Econometrica* 55, 391–407.
- Engle, R.F., V.K. Ng and M. Rothschild, 1990b, Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills, *Journal of Econometrics* 45, 213–237.
- Evans, M.D.D. and K.L. Lewis, 1994, Do stationary risk premia explain it all? Evidence from the term structure, *Journal of Monetary Economics* 33, 285–318.
- Fama, E.F., 1976, Foundations of Finance (Basic Books, New York).
- Fama, E.F. and K.R. French, 1988a, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246–273.
- Fama, E.F. and K.R. French, 1988b, Dividend yields and expected stock returns, *Journal of Political Economy* 96, 246–273.
- Fan, J., N.E. Heckman and M.P. Wand, 1994, Local polynomial kernel regression for generalized linear models and quasi-likelihood functions, *Annals of Statistics* 22, 1346–1370.
- Feinstone, L.J., 1987, Minute by minute: Efficiency, normality and randomness in intra-daily asset prices, *Journal of Applied Econometrics* 2, 193–214.
- Fisher, M., D. Nychka and D. Zervos, 1994, Fitting the term structure of interest rates with smoothing splines (mimeo, Board of Governors of the Federal Reserve System).
- Flavin, M., 1983, Excess volatility in the financial markets: A reassessment of the empirical evidence, *Journal of Political Economy* 91, 89–111.
- French, K.R., G.W. Schwert and R. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3–30.

- Friedman, B.M. and D.I. Laibson, 1989, Economic implications of extraordinary movements in stock prices, *Brookings Papers on Economic Activity*, 2/89, 137–189.
- Führer, J.G. Moore and S. Schuh, 1995, Maximum likelihood versus generalized method of moments estimation of the linear-quadratic inventory model, *Journal of Monetary Economics* 35, 115–157.
- Gallant, R., 1981, On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form, *Journal of Econometrics* 15, 211–244.
- Gallant, A.R. and G. Tauchen, 1990, A nonparametric approach to nonlinear time series analysis: Estimation and simulation, *IMA Volumes in Mathematics and its Application* (Springer-Verlag).
- Gallant, A.R. and G. Tauchen, 1992, Which moments to match? (mimeo, Duke University).
- Gallant, A.R., L.P. Hansen and G. Tauchen, 1990, Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution, *Journal of Econometrics* 45, 141–179.
- Gallant, A.R., D.A. Hsieh and G. Tauchen, 1991, On fitting a recalcitrant series: The pound/dollar exchange rate, 1974–83, in: W.A. Barnett, J. Powell and G.E. Tauchen (eds.), *Nonparametric and semiparametric methods in econometrics and statistics* (Cambridge University Press, Cambridge).
- Gallant, A.R., D. Hsieh and G. Tauchen, 1994, Estimation of stochastic volatility models with diagnostics (mimeo, Duke University).
- Gallant, A.R., J. Rossi and G. Tauchen, 1992, Stock prices and volumes, *Review of Financial Studies* 5, 199–242.
- Gennette, G. and T.A. Marsh, 1992, Variations in economic uncertainty and risk premiums on capital markets (mimeo, University of California at Berkeley).
- Geweke, J., 1994, Bayesian comparison of econometric models (mimeo, University of Minnesota).
- Ghose, D. and K.F. Kroner, 1994, The relationship between GARCH and symmetric stable processes: Finding the source of fat tails in financial data, *Journal of Empirical Finance* (forthcoming).
- Glosten, L.R., R. Jagannathan and D. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1802.
- Gouriéroux, C. and A. Monfort, 1992, Qualitative threshold ARCH models, *Journal of Econometrics* 52, 159–199.
- Gouriéroux, C., A. Monfort and E. Renault, 1993, Indirect inference, *Journal of Applied Econometrics* 8, S85–S118.
- Gouriéroux, C. and O. Scaillet, 1994, Estimation of the term structure from bond data, Working paper No. 9415, CEPREMAP.
- Gregory, A. and M. Veall, 1985, Formulating tests of non-linear restrictions, *Econometrica* 53, 1465–1468.
- Hamilton, J.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, J.D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357–384.
- Hamilton, J.D., 1990, Analysis of time series subject to changes in regime, *Journal of Econometrics* 45, 39–70.
- Hamilton, J.D. and R. Susmel, 1994, Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* 64, 307–333.
- Hamilton, J.D. and C.H. Whiteman, 1985, The observable implications of self-fulfilling expectations, *Journal of Monetary Economics* 16, 353–373.
- Hansen, B.E., 1990, Lagrange multiplier tests for parameter instability in non-linear models (mimeo, University of Rochester).
- Hansen, B.E., 1994, Autoregressive conditional density estimation, *International Economic Review* 35, 705–730.
- Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

- Hansen, L.P. and R.J. Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric investigation, *Journal of Political Economy* 88, 829–853.
- Hansen, L.P. and R. Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.
- Hansen, L.P. and K.J. Singleton, 1982, Generalized instrumental variables estimation of non-linear rational expectations models, *Econometrica* 50, 1269–1285.
- Hansen, L.P. and K.J. Singleton, 1990, Efficient estimation of linear asset pricing models with moving-average errors, NBER Technical paper No. 86.
- Härdle, W., 1990, *Applied non-parametric regression* (Cambridge University Press, Cambridge).
- Harrison, P., 1994, Are all financial time-series alike (mimeo, Duke University).
- Harvey, A.C. and N. Shephard, 1993, The econometrics of stochastic volatility (mimeo, London School of Economics).
- Harvey, A., E. Ruiz and E. Sentana, 1992, Unobserved component time series models with ARCH disturbances, *Journal of Econometrics* 52, 129–157.
- Harvey, A., E. Ruiz and N. Shephard, 1994, Multivariate stochastic variance models, *Review of Economic Studies* 61, 247–264.
- Harvey, C., 1991, The world price of covariance risk, *Journal of Finance* 46, 111–157.
- Hejazi, W., 1994, Are term premia stationary? (mimeo, University of Toronto).
- Hentschel, L., 1991, The absolute value GARCH model and the volatility of U.S. stock returns (mimeo, Princeton University).
- Hentschel, L., 1994, All in the family: Nesting asymmetric GARCH models (paper given to the Econometric Society Winter Meeting, Washington DC).
- Higgins, M.L. and A.K. Bera, 1992, A class of nonlinear ARCH models: Properties, testing and applications, *International Economic Review* 33, 137–158.
- Hill, B.M., 1975, A simple general approach to inference about the tail of a distribution, *Annals of Statistics* 3, 1163–1174.
- Hodrick, R.J., 1991, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, Finance Department working paper No. 88, Northwestern University.
- Hols, M.C.A.B. and C. de Vries, 1991, The limiting distribution of extremal exchange rate changes, *Journal of Applied Econometrics* 6, 287–302.
- Hsieh, D.A., 1989a, Modeling heteroskedasticity in daily foreign exchange rates, *Journal of Business and Economic Statistics* 7, 307–317.
- Hsieh, D.A., 1989b, Testing for nonlinear dependence in daily foreign exchange rates, *Journal of Business* 62, 339–368.
- Huang, R.D. and C.S.Y. Lin, 1990, An analysis of nonlinearities in term premiums and forward rates (mimeo, Vanderbilt University).
- Hutchinson, J.M., A.W. Lo and T. Poggio, 1994, A nonparametric approach to pricing and hedging derivative securities via learning networks, *Journal of Finance*, XLIX, 851–889.
- Jacquier, E., N.G. Polson and P.E. Rossi, 1994, Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* 12, 57–80.
- Johansen, S., 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231–254.
- Jorion, P., 1988, On jump processes in the foreign exchange and stock markets, *Review of Financial Studies* 1, 427–445.
- Kasa, K., 1992, Common stochastic trends in international stock markets, *Journal of Monetary Economics* 29, 95–124.
- Kearns, P., 1990a, Testing speculative bubbles in interest rates with the present value model (mimeo, University of Rochester).
- Kearns, P., 1990b, Non-linearities in the term structure (mimeo, University of Rochester).
- Kearns, P., 1992, Pricing interest rate derivative securities when volatility is stochastic (mimeo, University of Rochester).

- Kearns, P., 1993, Volatility and the pricing of interest rate derivative claims (unpublished Ph.D. thesis, University of Rochester).
- Kearns, P. and A.R. Pagan, 1992, Estimating the density tail index for financial series (mimeo, Australian National University).
- Kearns, P. and A.R. Pagan, 1993, Australian stock market volatility: 1875–1987, *Economic Record* 69, 163–178.
- Kim, S. and N. Shephard, 1994, Stochastic volatility: Likelihood inference and comparison with ARCH models (mimeo, Nuffield College, Oxford).
- King, M.L., 1985, A point optimal test for autoregressive disturbances, *Journal of Econometrics* 27, 21–37.
- Kleidon, A.W., 1986, Variance bounds tests and stock price valuation models, *Journal of Political Economy* 96, 953–1001.
- Kloeden, P. and E. Platten, 1992, *The numerical solution of stochastic differential equations* (Springer-Verlag).
- Knez, P., R. Litterman and J. Scheinkman, 1989, Explorations into factors explaining money market returns, Discussion paper No. 6, Goldman Sachs and Co.
- Kocherlakota, N., 1990, On tests of representative consumer asset pricing models, *Journal of Monetary Economics* 26, 285–304.
- Koedijk, K.G., F.G.J.A. Nissen, P.C. Schotman and C.P. Wolff, 1993, The dynamics of short-term interest rate volatility reconsidered (mimeo, Limburg Institute of Financial Economics).
- Koop, G., 1994, An objective Bayesian analysis of common stochastic trends in international stock prices and exchange rates, *Journal of Empirical Finance* 1, 343–364.
- Kwiatowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* 54, 159–178.
- Lamoureux, C.G. and W.D. Lastrapes, 1990, Persistence in variance, structural change and the GARCH model, *Journal of Business and Economic Statistics* 8, 225–234.
- Lee, J. and M.L. King, 1993, A locally most mean powerful based score test for ARCH and GARCH regression disturbances, *Journal of Business and Economic Statistics* 11, 17–27.
- Lee, S.W., 1992, Asymptotic properties of the maximum likelihood estimator of the GARCH-M and IGARCH-M models (mimeo, University of Rochester).
- Lee, S.W. and B.E. Hansen, 1991, Asymptotic properties of the maximum likelihood estimator and test of the stability of parameters of the GARCH and IGARCH models (mimeo, University of Rochester).
- Lee, S.W. and B.E. Hansen, 1994, Asymptotic theory for the GARCH (1,1) quasi maximum likelihood estimator, *Econometric Theory* 10, 29–52.
- LeRoy, S. and W.R. Parke, 1987, Stock price volatility: A test based on the geometric random walk (mimeo, University of California at Santa Barbara).
- Lin, W.L., 1992, Alternative estimators for factor GARCH models – A Monte Carlo comparison, *Journal of Applied Econometrics* 7, 259–279.
- Linton, O., 1992, The shape of the risk premium: Evidence from a semi-parametric, mean-exponential GARCH Model (mimeo, Nuffield College, Oxford).
- Linton, O., 1993, Adaptive estimation in ARCH models, *Econometric Theory* 9, 539–569.
- Lo, A.W., 1991, Long-term memory in stock market prices, *Econometrica* 59, 1279–1313.
- Lo, A.W. and A.C. MacKinlay, 1989, The size and power of the variance ratio test in finite samples: A Monte Carlo investigation, *Journal of Econometrics* 40, 203–238.
- Loretan, M. and P.C.B. Phillips, 1994, Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial data sets, *Journal of Empirical Finance* 1, 211–248.
- Lumsdaine, R.L., 1991, Asymptotic properties of the maximum likelihood estimator in GARCH (1,1) and IGARCH (1,1) models (mimeo, Princeton University).

- Lye, J.N. and V.L. Martin, 1991, Modelling and testing stationary exchange rate distributions: The case of the generalized Student's t distribution (mimeo, University of Melbourne).
- McCulloch, J.H., 1971, Measuring the term structure of interest rates, *Journal of Business* 44, 19–31.
- McCulloch, J.H., 1989, U.S. term structure data, 1946–1987, *Handbook of monetary economics*, 672–715.
- McDonald, J. and J. Darroch, 1983, Consistent estimation of equations with composite moving average disturbance terms, *Journal of Econometrics* 23, 253–267.
- Mahieu, R. and P. Schotman, 1994a, Stochastic volatility and the distribution of exchange rate news, (mimeo, University of Limburg, Maastricht).
- Mahieu, R. and P. Schotman, 1994b, Neglected common factors in exchange rate volatility, *Journal of Empirical Finance* 1, 279–311.
- Mandelbrot, B., 1963, The variation of certain speculative prices, *Journal of Business* 36, 394–419.
- Mankiw, N.G., D. Romer and M.D. Shapiro, 1985, An unbiased re-examination of stock market volatility, *Journal of Finance* 40, 677–687.
- Mao, C.S., 1990, Hypothesis testing and finite sample properties of generalized method of moments estimators: A Monte Carlo study (mimeo, Federal Reserve Bank of Richmond).
- Melino, A. and S. Turnbull, 1990, Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* 45, 239–265.
- Merton, R., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–888.
- Mihalstein, G.N., 1974, Approximate integration of stochastic differential equations, *Theory of Probability and its Applications* 19, 557–562.
- Nelson, C.R. and M.J. Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.
- Nelson, C.R. and R. Startz, 1990, The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one, *Journal of Business* 63, S125–140.
- Nelson, D.B., 1988, The time series behaviour of stock market volatility and returns (unpublished Ph.D. thesis, M.I.T.).
- Nelson, D.B., 1989, Modeling stock market volatility changes, *Proceedings of the American Statistical Association, Business and Economic Statistics Section*.
- Nelson, D.B., 1990a, Stationarity and persistence in the GARCH (1,1) model, *Econometric Theory* 6, 318–334.
- Nelson, D.B., 1990b, A note on the normalized residuals from ARCH and stochastic volatility models (mimeo, University of Chicago).
- Nelson, D.B., 1990c, ARCH models as diffusion approximations, *Journal of Econometrics* 45, 7–38.
- Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.
- Nelson, D.B. and D.P. Foster, 1992, Filtering and forecasting with misspecified ARCH models II: Making the right forecast with the wrong model (mimeo, Graduate School of Business, University of Chicago).
- Newey, W.K., 1985, Maximum likelihood specification testing and conditional moment tests, *Econometrica* 53, 1047–1070.
- Newey, W.K. and D.G. Steigerwald, 1996, Consistency of the quasi-MLE for models with conditional heteroskedasticity, *Econometric Theory* (forthcoming).
- Newey, W.K. and K.D. West, 1987, A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Nieuwland, F.G.M.C., W.F.C. Verschoor and C.C.P. Wolff, 1994, Stochastic trends and jumps in EMS exchange rates, *Journal of International Money and Finance* 13, 699–727.
- Nyblom, J., 1989, Testing the constancy of parameters over time, *Journal of the American Statistical Association* 84, 223–230.
- Pagan, A.R., 1975, A note on the extraction of components from time series, *Econometrica* 43, 163–168.

- Pagan, A.R. and Y.S. Hong, 1991, Non-parametric 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, A.R. and Y. Jung, 1993, Understanding the failure of some instrumental variables estimators (mimeo, Australian National University).
- Pagan, A.R. and H. Sabau, 1992, Consistency tests for heteroskedasticity and risk models, *Estudios Económicos* 7, 3–30.
- Pagan, A.R. and G.W. Schwert, 1990a, Alternative models for conditional stock volatility, *Journal of Econometrics* 45, 267–290.
- Pagan, A.R. and G.W. Schwert, 1990b, Testing for covariance stationarity in stock market data, *Economics Letters* 3, 165–170.
- Pagan, A.R. and A. Ullah, 1988, The econometric analysis of models with risk terms, *Journal of Applied Econometrics* 3, 87–105.
- Pagan, A.R. and A. Ullah, 1995, Non-parametric econometrics (unpublished manuscript, Australian National University).
- Pagan, A.R. and F.X. Vella, 1989, Diagnostic tests for models based on individual data, *Journal of Applied Econometrics* 45, 429–452.
- Pagan, A.R., V. Martin and A.D. Hall, 1995, Modelling the term structure, Working paper in Economics and Econometrics No. 284, Australian National University.
- Pearson, N.D. and T.-S. Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll and Ross model, *Journal of Finance*, XLIX, 1279–1304.
- Pesaran, M.H. and B. Pesaran, 1993, A simulation approach to the problem of computing Cox's statistic for testing non-nested models, *Journal of Econometrics* 57, 377–392.
- Phillips, P.C.B. and J.Y. Park, 1988, On the formulation of Wald tests of nonlinear restrictions, *Econometrica* 56, 1065–1083.
- Pindyck, R.S., 1984, Risk, inflation and the stock market, *American Economic Review* 74, 335–351.
- Rich, R., J. Raymond and J.S. Butler, 1991, Generalized instrumental variables estimation of autoregressive conditional heteroskedastic models, *Economics Letters* 35, 179–185.
- Richardson, M. and J.H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323–348.
- Rosenberg, B., 1985, Prediction of common stock betas, *Journal of Portfolio Management* 11, 5–14.
- Ross, S.A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.
- Ruiz, E., 1994, Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics* 63, 289–306.
- Scheinkman, J. and B. LeBaron, 1989, Nonlinear dynamics and stock returns, *Journal of Business* 62, 311–337.
- Schotman, P. and H.K. van Dijk, 1991, A Bayesian analysis of the unit root in real exchange rates, *Journal of Econometrics* 49, 195–238.
- Schwarz, G., 1978, Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.
- Schwert, G.W., 1990, Indexes of stock prices from 1802 to 1897, *Journal of Business* 63, 399–426.
- Schwert, G.W. and P.J. Seguin, 1990, Heteroskedasticity in stock returns, *Journal of Finance* 45, 1129–1155.
- Sentana, E., 1991, Quadratic ARCH models: A potential re-interpretation of ARCH models (mimeo, London School of Economics).
- Shea, G.S., 1989, Testing stock market efficiency with volatility statistics: Some exact finite sample results (mimeo, Pennsylvania State University).
- Shea, G.S., 1992, Benchmarking the expectations hypothesis of the interest rate term structure: An analysis of cointegration vectors, *Journal of Business and Economic Statistics* 10, 347–366.
- Shephard, H., 1994, Partial non-Gaussian state space, *Biometrika* 81, 115–131.

- Shiller, R.J., 1979, The volatility of long-term interest rates and expectations models of the term structure, *Journal of Political Economy* 87, 1190–1219.
- Shiller, R.J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421–436.
- Silverman, B.W., 1986, Density estimation for statistics and data analysis (Chapman Hall, New York).
- Sola, M. and A. Timmermann, 1994, Fitting the moments: A comparison of ARCH and regime switching models for daily stock returns (mimeo, Birkbeck College).
- Sowell, F., 1992, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics* 53, 165–188.
- Spanos, A., 1986, Statistical foundations of econometric modelling (Cambridge University Press, Cambridge).
- Staiger, D. and J.H. Stock, 1993, Instrumental variables regression with weak instruments, NBER Technical working paper No. 151.
- Steigerwald, D.G., 1992, Semi-parametric estimation in financial models with time varying variances (mimeo, University of California, Santa-Barbara).
- Stock, J.H. and M.W. Watson, 1988, Testing for common trends, *Journal of the American Statistical Association* 83, 1097–1107.
- Summers, L.H., 1986, Does the stock market rationally reflect fundamental values?, *Journal of Finance* 41, 591–601.
- Tanaka, K., 1990, Testing for a moving average unit root, *Econometric Theory* 6, 445–458.
- Tauchen, G., 1985, Diagnostic testing and evaluation of maximum likelihood models, *Journal of Econometrics* 30, 415–443.
- Tauchen, G., 1986, Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data, *Journal of Business and Economic Statistics* 4, 397–425.
- Tauchen, G. and M. Pitts, 1983, The price variability-volume relationship on speculative markets, *Econometrica* 51, 485–505.
- Taylor, S.J., 1986, Modelling financial time series (John Wiley, Chichester).
- Taylor, S.J., 1990, Modelling stochastic volatility (mimeo, University of Lancaster).
- Timmermann, A., 1995, Cointegration tests of present value models with a time-varying discount factor, *Journal of Applied Econometrics* 10.
- Tsay, R., 1986, Nonlinearity tests for time series, *Biometrika* 73, 461–466.
- Vahid, F. and R. Engle, 1993, Common trends and common cycles, *Journal of Applied Econometrics* 8, 341–360.
- Vasicek, O., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.
- Watson, G.S., 1964, Smooth regression analysis, *Sankhya, Series A*, 26, 359–372.
- West, K.D., 1988, Dividend innovations and stock price volatility, *Econometrica* 56, 37–61.
- West, K.D., H.J. Edison and D. Cho, 1993, A utility-based comparison of some models of exchange rate volatility, *Journal of International Economics* 35, 23–45.
- White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.
- Wu, P., 1992, Testing fractionally integrated time series, Working paper, Victoria University of Wellington.
- Zakoian, J.M. 1994, Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* 18, 931–955.