ANNALES D'ÉCONOMIE ET DE STATISTIQUE. - N° 40 — 1995

Some Properties of Absolute Return An Alternative Measure of Risk

C. W. J. GRANGER, Zhuanxin DING *

ABSTRACT – The expected absolute return belongs to a class of risk measure derived by Luce (1980) from axioms. The paper considers the time series properties of $E[|r|^{\theta}]$ and also the marginal distribution properties, for various properties of θ. Using a long daily stock index series it is found that the autocorrelations decline slowly for all positive θ but this "long-memory" property is strongest for $\theta = 1$, the absolute return. The moments of absolute returns, after removal of a few outliers, suggest that an exponential distribution is appropriate.

Quelques propriétés du rendement total : une mesure alternative du risque

RÉSUMÉ – Le rendement total espéré appartient à une classe de mesure du risque définie par Luce (1980) à partir d'axiomes. L'article considère les propriétés des séries temporelles de $E[|r|^{\theta}]$ et aussi les propriétés de distribution marginale pour des propriétés différentes de θ. En utilisant un grand nombre de séries, il est apparu que les autocorrélations déclinent lentement pour tout θ positif, mais cette propriété est plus forte pour $\theta = 1$, le rendement total. Les périodes de rendement total, après suppression des points aberrants, suggèrent qu'une distribution exponentielle est appropriée.

* C. W. J. GRANGER, Z. DING: University of California, San Diego, USA.

1 Introduction

The returns on a speculative asset over some future period is uncertain and so can only be described fully by the use of a distribution function $F(x)$. Rather than trying to determine the full distribution it would be convenient if it could be summarized in terms of just a few parameters. It is usual to concentrate on just two, a measure of location for the distribution, m, and a measure of dispersion, d. It is also usual, and apparently uncontroversial, to use the mean return

$$m = E[x] = \int x \, dF(x)$$

as the measure of location. It has also become common place to use the variance σ^2, or better the standard deviation σ, as the measure of dispersion, where

$$\sigma^2 = \int (x-m)^2 \, dF(x).$$

Most of optimum portfolio theory is based on (m, σ) or (m, σ^2) considerations, for example. Although variance is certainly mathematically convenient and is the correct measure of dispersion for some distributions, such as the Gausssian, there is no compelling theoretical reason why it should be used. BORCH [1969] has produced convincing criticisms of the use of (m, σ^2) theories but similar criticisms will apply if σ^2 is replaced by any other particular measure of dispersion.

There may not exist a perfect measure of dispersion, definable from a distribution and thus possible to estimate from data. However, MEYER [1987] has shown that if an investor does decide on a pair of measures (m, d) and considers distributions of the class $F\left(\frac{x-m}{d}\right)$ for some given $F(\quad)$ then a number of strong theoretical consequences follow about investor behavior.

In this paper the measure of dispersion considered is

$$\mu_a = E(|r - m|),$$

the expected absolute value of the deviation of the return from its mean, and also related measures which raise $|r - m|$ to fractional powers or subtract quantities other than m. The next section considers the concept of risk, the next links μ_a and σ^2 and the remaining sections generally discuss empirical and theoretical properties of μ_a.

2 Considerations of Risk

There is no risk without uncertainty but uncertainty is not just risk. Earthquake and fire insurance are used to remove rare but large downside uncertainty (risk) corresponding to negative returns. But very rarely does anyone take out "insurance" to deal with unexpected large positive returns, such as come from winning a lottery or the Nobel Prize. These types of uncertainty are generally not thought of as risk. A useful and concise discussion of risk as seen from the viewpoint of economic theory, is provided by MACHINA and ROTHSCHILD [1987]. Much of the discussion concerns the attitudes of a risk averse investor to returns that are drawn from alternative distribution functions $F(x)$, $G(x)$. A risk averse investor always experiences a decline in expected utility when moving from an asset with a certain return m to an asset with uncertain r having $E[r] = m$. For this discussion it will be assumed that all returns have zero mean, so that any non-zero constant mean has been subtracted. Particular attention is paid in the theory to "mean preserving spreads", in which probability masses have their locations moved on the distribution without altering the mean of the distribution. If this mean preserving spread moves mass from the center of a distribution towards the tails, it will result in an increased risk.

LUCE [1980] has considered the consequences of a pair of assumptions concerning a measure of risk. Suppose that one is considering a pair of returns r_1, r_2 with zero means, $r_1 = r$ has distribution $F(x)$ and $r_2 = \alpha r$ has distribution $G_\alpha(x) = F(x/\alpha)$, where $\alpha > 0$ is a scale factor. Let $R(F)$ be a measure of risk associated with F, then assume that there is an $S(\alpha)$ such that

$$(1) \qquad R(G_\alpha(x)) = S(\alpha)\, R(F(x)),$$

so that the effect on R of a scale change is a multiplicative one. Further, assume that there exists a function $T(x)$ so that, for any F,

$$(2) \qquad R(F) = E[T(x)] = \int T(x)\, dF(x).$$

(1) and (2) are clearly based on assumptions and so need not be true, in particular there need not exist a function $T(x)$ so that (2) holds, but neither assumption is clearly unreasonable. Luce proves from these assumptions that

$$(3) \qquad S(\alpha) = \alpha^\theta, \qquad \theta > 0$$

and

$$(4) \qquad R(F) = AE[|r|^\theta], \qquad \theta > 0,\ A > 0.$$

A reformulation of Luce's proof is given in the Appendix. The value of A is essentially of no importance and it is convenient to put $A = 1$. If one accepts the assumptions, this theorem provides a limited class of potential measures of risk depending on a value of θ to be determined by the individual investor, but which can possibly be estimated from data if there is a consensus about this value. It might be noted that if $F(x)$ is Gaussian, with mean zero and variance σ^2, then $R(F)$ is proportional to σ^θ, and is thus proportional to the standard deviation if $\theta = 1$ and to the variance if $\theta = 2$. Which is the more reasonable depends on whether you think that the scale effect $S(\alpha)$ is better expressed as α or α^2. For non-Gaussian distributions, there is not necessarily any simple relationship between $R(F)$ and σ except when $\theta = 2$.

The two measures of risk which have the greatest intuitive appeal, within the above class, because of their simplicity are $\mu_a = E(|r - m|)$ and $\sigma^2 = E[(r-m)^2]$, although other measures from (4) corresponding to other values of θ will be analyzed in the empirical sections. Consideration of mean preserving spreads does not help to decide if either μ_a or σ^2 is a superior measure of risk. If a spread occurs with all changes occurring on one side of the mean then μ_a will be unaltered but σ^2 will change. However, one can construct other spreads where the reverse occurs. Yet others can either change both σ^2 and μ_a or leave both unchanged. What these results indicate is that either μ_a or σ^2 will, at most, only be an approximate characterization of risk and cannot capture all of its complexities.

3 Two Alternative Representations of a Return

The following two representations will be considered for a return (possibly after mean removal)

(5) $$\text{(i)} \quad r_t = s_t\, a_t$$

$$\text{where } s_t = \text{sign}(r_t) = \begin{cases} 1 & \text{if } r_t > 0 \\ 0 & \text{if } r_t = 0 \\ -1 & \text{if } r_t < 0 \end{cases}$$

and $a_t = |r_t|$, and

(6) $$\text{(ii)} \quad r_t = \sigma_t\, e_t$$

where $\sigma_t^2 = E[r_t^2]$, so that σ_t captures any heteroskedasticity in the returns and $E[e_t^2] = 1$. If σ_t^2 is the unconditional variance it will, at most, be a deterministic function of time and so any stochastic temporal properties of r_t will be embedded in e_t. For example, if r_t is an uncorrelated sequence, then for $k \neq 0$,

(7) $$\rho_k = \text{corr}(r_{t+k}, r_t) = 0 = \text{corr}(e_{t+k}, e_t).$$

The two representations are merely identities and so have no impact unless further constraints are imposed, and if these constraints, or assumptions, closely match the actual empirical properties of returns. For example, suppose that r_t has zero autocorrelations, as in (7), and that s_t, a_t are independent in (5). Remember that r_t has zero mean, it follows that under these assumptions

$$\begin{aligned} \text{cov}(r_{t+k}, r_t) &= E[r_{t+k}\, r_t] \\ &= E[s_{t+k}\, s_t] \cdot E[a_{t+k}\, a_t] = 0 \end{aligned}$$

so that, as a_t is positive, necessarily

(8) $$E[s_{t+k}\, s_t] = 0, \quad k \neq 0.$$

Of greater interest are conditonal properties of the return at time $t+1$ based on an information set I_t available at time t. For example, using (6), put

$$\sigma_{t+1}^2 = E[r_{t+1}^2 \mid I_t]$$

and now $E[e_{t+1}^2 \mid I_t] = 1$ so that e_t is conditinally homoskedastic. Particular models for σ_t^2 will correspond to the large variety of ARCH-type models that now exist. It should be noted that it does not follow that $E[g(e_{t+1}) \mid I_t]$ is not a function of I_t for any $g(\ )$. In particular $E[| e_{t+1} |^\delta \mid I_t]$ may be a function of the information set for $\delta \neq 2$. If r_t is conditionally normally distributed or if σ_t is a sufficient statistic for measuring the volatility of the conditional distribution of r_t given I_{t-1}, then these expectations will not be functions of I_{t-1}, but in other cases they generally will be.

Using the representation (5), regardless of whether or not s_t and a_t are independent, a possible simple relation between the conditional forms of (5) and (6) is easily found as follows.

Denoting

$$\mu_{tk}^a = E[a_{t+k} \mid I_t]$$

so that

$$a_{t+k} = \mu_{t,k}^a + e_{t,k}^a$$

with

$$E[e_{tk}^a \mid I_t] = 0,$$

consider

(9) $$\begin{aligned} \sigma_{t,k}^2 = E[r_{t+k}^2 \mid I_t] &= E[s_{t+k}^2\, a_{t+k}^2 \mid I_t] \\ &= E[(\mu_{t,k}^a + e_{t,k}^a)^2 \mid I_t] \\ &= (\mu_{t,k}^a)^2 + E[(e_{t,k}^a)^2 \mid I_t]. \end{aligned}$$

Thus, if the special case

$$\begin{aligned} \mu_{t,k}^a &= b_{0k} + b_{1k}\, a_t \\ E[(e_{t,k}^a)^2 \mid I_t] &= \text{constant} \end{aligned}$$

occurs, it follows that $\sigma_{t,k}^2$ will be a quadratic function of a_t.

4 Properties to be Considered

As this paper may be thought of as exploratory, only a few of the many possible properties of the series will be considered. Initially two effects are discussed. Using the notation $a_t = |r_t|$, $\rho_k(\delta, \theta) = \text{corr}(a_{t+k}^{\delta}, a_t^{\theta})$ and assuming that a_t is stationary so that these auto or cross-correlations can be estimated from an observed series, the two effects are

(i) *The Taylor Effect*

Hypothesis: $\rho_k(1, 1) > \rho_k(\delta, \delta)$ for any $\delta \neq 1$ and that a_t is long memory, so that $\rho_k(1, 1)$ declines slowly. The property $\rho_k(1, 1) > \rho_k(2, 2)$ was noted in Tayor [1986, chapter 2] for a variety of speculative price series, which suggested the above hypothesis, and,

(ii) *The Machina Effect*

Hypothesis $\rho_k(\delta, 1) > \rho_k(\delta, \theta)$ for any $\theta \neq 1$, and any $\delta > 0$, named after a suggestion by Mark Machina.

The data used in this study is a daily price index of the S&P 500 (New York) stocks, for the period January 3, 1928 to April 30, 1991, kindly provided to use by William Schwert. The length of the series is 17,055 observations. If p_t is the price series, the return is defined as $r_t = \log p_t - \log p_{t-1}$. a_t can be defined as either $|r_t|$ of $|r_t - m|$ where m is the mean return for the time period covered by the sample being used. For the whole data set the mean of r_t is 0.00018 and calculations using either form of a_t generally gives very similar results. If a table indicates that the data is "mean-removed" then $a_t = |r_t - m|$ is used. A second very real possibility is that with such a long time-span there will have been structural shifts in the generating mechanism, just as there has clearly been institutional shifts in the stock markets and structural shifts in the encompassing economy. To give an indication of the stability, or otherwise, of the results the full sample was arbitrarily sub-divided into ten equal components, each with about 1,700 terms and many of the calculations are repeated for each sub-sample. A practical problem of some substance is whether or not there exist outliers in the data and what to do about them. It can be argued that extraordinary price changes, as occurred around October 1987 for the example, are something that should be expected to occur occassionally in a speculative market, and such events are merely an outcome of the generating mechanism and not a break-down of the usual mechanism. Rather than taking a formal position about outliers, much of the analysis was performed on the original data and then repeated on a data set with the outliers downweighted. This "outlier-reduced" set was formed by putting any a_t that was greater than four standard deviations from the rest of the sample at the four standard deviation value. The resulting values were still large but no longer dominating. The number of outliers reduced is shown in Table 2B.

The Taylor effect was examined in Ding, Granger, and Engle [1993]. There, a long memory property for a_t^{θ} was established and found to be stronger for $\theta = 1$ than for $\theta = 0.25, 0.5, 0.75, 1.25, 1.5, 1.75$ and 2. An alternative way to explore a possible long memory property is by fitting a fractionally integrated type model to the series. Special care should be taken here since in our case the series themselves do not have long memory property but their absolute values do. The long memory models that can be found so far in the literature (*see* e.g. Mandelbrot and van Ness [1968], Granger and Joyeux [1980]) only deal with long memory property of the series themselves. A slightly different long memory model is needed for our purpose. It is clear that the following model has the desired property,

$$r_t = \sigma_t\, e_t, \quad e_t - i.i.d.\,(0,1),$$
$$\sigma_t = \left(1-(1-B)^d\right)\frac{a_t}{\lambda} = \frac{1}{\lambda}\sum_{j=1}^{\infty}\frac{d}{\Gamma(1-d)}\cdot\frac{\Gamma(j-d)}{\Gamma(j+1)}\,a_{t-j},$$

where $a_t = |\,r_t\,|$, $\lambda = E\,|\,e_t\,|$ and $0 < d \leq 1/2$.

For this model one has corr $(r_t,\, r_{t-k}) = 0$ for all $k > 0$ but corr $(|\,r_t\,|,\, |\,r_{t-k}\,|) =$ corr $(a_t,\, a_{t-k}) = \dfrac{\Gamma(1-d)}{\Gamma(d)}\,\dfrac{\Gamma(k+d)}{\Gamma(k+1-d)}$ which is the same as the correlation function of fractionally integrated series. This can also be seeen by by rewriting the conditional standard deviation equation in the form

$$(1-B)^d\, a_t = \sigma_t\,(\,|\,e_t\,| - \lambda) = \eta_t$$

where $\eta_t = \sigma_t\,(|\,e_t\,| - \lambda)$ is a mean zero short memory stationary process with conditional heteroskedasticity. Thus. a_t is essentially a fractionally integrated process.

This kind of model can arise from aggregation of micro level economic variables as discussed in GRANGER [1980]. We plan to discuss various aspect of the aggregation in a subsequent paper.

To estimate the fractional differencing parameter d in this model, the method proposed by GEWEKE and PORTER-HUDAK [1983] is used since the conditions for using this method are quite weak and our model satisfies the requirements. Table 1 A shows some basic sample characteristics, including the estimate of d, for a_t^{θ} for various values of θ, where JB is the Jarques-Bera normality test statistic.

Using an optimization package, it is found that the normality test statistic is minimized at $\theta = 0.191$ and the maximum estimated d of 0.5085 is reached when $\theta = 0.214$. For this sample size, the approximate 95%

TABLE 1A

Sample Statistics for Full Sample, Mean Subtracted, of a_t^{θ}

θ	mean	st.dev.	skewness	kurtosis	JB	d
.10	.580	.068	–.38	3.42	533	.507
.20	.342	.078	.025	3.25	45	.508
.30	.204	.068	.415	3.64	777	.507
.50	.075	.041	1.24	6.34	12257	.499
.75	.023	.020	2.55	17.10	15500	.481
1.00	.007	.009	4.64	51.44	$17E^5$	.474
1.25	.003	.004	8.08	156	$17E^6$	.464
1.50	.001	.002	14.5	482	$16E^7$	.452
1.75	.000	.001	23.9	1174	$97E^7$	.440
2.00	.000	.001	38	2518	$45E^8$	.427

TABLE 1B

Estimated* d *values

	Original Data		Outlier reduced	
	$\hat{d}$	Standard Error	$\hat{d}$	Stantard Error
Full Sample	0.474	.058	0.491	.060
Sub-sample 1	0.358	.133	0.365	.122
2	0.405	.131	0.402	.134
3	0.438	.097	0.426	.109
4	0.336	.144	0.353	.131
5	0.156	.106	0.193	.102
6	0.445	.113	0.464	.108
7	0.518	.088	0.517	.098
8	0.714	.105	0.709	.114
9	0.436	.110	0.448	.112
10	0.352	.070	0.452	.082
Average over Sub-samples	0.416		0.433	

confidence interval of $\hat{d}$ around a true value of 0.5 is 0.375 to 0.625. These results appear not to agree with the Taylor effect as $\hat{d}$ is not a maximum at $\theta = 1$. However the autocorrelations of an $I(d_1)$ process are very similar to an $I(d_2)$ process with $d_1 = 0.498$ and $d_2 = 0.474$, for example. The Taylor effect can still occur if a_t is $I(d) + x_t$, a_t^θ is $I(d) + \chi_{\theta_t}$, where variance $\chi_{\theta_t} >$ variance x_t, $\theta \neq 1$, and the x's are stationary series. It is interesting to note that CAO and TSAY [1992], when considering absolute values of monthly S&P returns, formed Box-Cox transformations a_t^θ and chose $\theta = 0.25$ as the maximum likelihood estimate, which is remarkably close to the value found here that makes a_t^θ nearest to normal with daily data. Cao and Tsay find evidence of nonlinearities in their data, but this possibility is not considered here and they failed to take into account the possible long-memory property of this data when performing a linearity test.

Table 1 B shows estimates of d, with the associated standard errors, for a_t but for subsamples as well as the full sample and for both the original data and outlier-reduced data.

It is seen in Table 1 B that if the $I(d)$ model is accepted as a worthwhile approximation, then all the estimated d values are significantly different from both zero (except sub-sample 5, which corresponds to the early 1960's) and from one, under the assumption that $\hat{d}$ is normally distributed. Generally, going from the original data to the outlier reduced data does not affect the value of $\hat{d}$, although the values are somewhat higher for the second data set as might be expected as outliers probably add a high frequency component to the data. The value of d does seem to change significantly through time, with a low value in Sample 5, the early 60's and a high one in Sample 8, the late 1970's. Overall, a value of d near, but just below, 0.5 seems to be the most likely. It might be noted that if $d = 1/2$, one is

TABLE 2A

Original Dada

	mean ($\times E^3$)	st.dev.($\times E^3$)	var ($\times E^5$)	skewness	kurtosis
Full Sample	7	9	8	4.6	51.4
Sub Sample 1	15	16	27	2.6	13.3
2	11	11	12	2.2	10.2
3	6	7	5	4.1	33.1
4	6	6	4	2.9	17.4
5	5	5	2	3.0	27.6
6	5	5	2	3.7	31.2
7	5	4	2	2.3	12.6
8	7	6	3	1.8	7.8
9	7	6	3	1.6	7.2
10	7	9	8	10.0	207.5

TABLE 2B

Outlier Reduced Data

	Number of positive	Ouliers Negative	mean $\times E^3$	st.dev $\times E^3$	variance $\times E^5$	skewness	kurtosis
Full Sample	64	74	7	8	6	2.4	10.3
Sub Sample 1	5	5	15	16	24	2.1	8.4
2	3	3	11	11	11	1.9	7.4
3	7	8	6	6	4	2.3	9.9
4	2	9	6	6	4	2.1	8.8
5	1	3	5	4	2	1.7	6.8
6	7	4	5	4	2	2.1	9.5
7	3	3	5	4	2	1.8	7.3
8	4	1	7	6	3	1.6	6.6
9	3	1	7	6	3	1.5	6.0
10	3	6	7	7	5	2.2	10.3

If $\mu_k = E[a^k]$, *then skewness* $= \mu_3/\mu_2^{3/2}$ *and kurtosis* $= \mu_4/\mu_3^2$.

at the borderline between stationary and "non-sationary" series. An $I(1/2)$ process has a variance that increases with time, proportional to $\log t$ as shown in GRANGER [1988]. If the inputs are homoskedastic, which seems unlikely, the ratio of the variance of the full outlier-reduced sample to any subsample would be

$$\log(17,000)/\log(1,700) \approx 1.31.$$

As seen in Table 2, the ratio of the variance actually lies between 0.3 (for subsample 1, which includes the Great Crash) and 4, so the evidence is inconclusive. Table 2 also shows the number of outliers reduced in each sample corresponding to $s_t > 0$, plus estimates of various unconditional moments.

TABLE 3A

Contemporaneous Correlation at Lag $k = 0$

δ, θ	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
0.25	1.00	.98	.92	.84	.74	.63	.52	.42
0.50		1.00	.98	.93	.85	.75	.64	.53
0.75			1.00	.98	.93	.85	.75	.65
1.00				1.00	.98	.93	.85	.76
1.25					1.00	.98	.93	.86
1.50	(symmetric)					1.00	.98	.94
1.75							1.00	.99
2.00								1.00

TABLE 3B

Correlation at Lag $k = 1$

δ, θ	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
0.25	.22	.24	.25	.25	.24	.22	.20	.17
0.50	.24	.26	.28	.28	.28	.26	.23	.20
0.75	.25	.28	.30	.31	.30	.29	.26	.23
1.00	.25	.28	.31	.32	.32	.30	.28	.25
1.25	.23	.27	.30	.32	.32	.31	.29	.26
1.50	.21	.25	.28	.30	31	.30	.28	.25
1.75	.19	.22	.25	.27	.28	.28	.26	.24
2.00	.16	.19	.22	.24	.25	.25	.24	.22

As might be expected, the biggest differences between Tables 2 A, 2 B are for the measures of skewness and kurtosis.

Turning to the Machina effect, Table 3 shows

$$\rho_k(\delta, \theta) = \text{corr}(a_{t+k}^{\delta}, a_t^{\theta}).$$

The power of the current variable is shown horizonally, those of the future variable vertically. All values are for returns without mean removed and without outlier reduction. Table 3 A shows the contemporaneous correlations ($k = 0$) so that, for example, the correlation between a_t and a_t^2 is 0.76. As one would expect $\rho_0(\delta, \theta)$ is largest when δ, θ are nearly equal.

It is seen that $\rho_0(\delta, \delta + .25) = .97$, $\rho_0(\delta, \delta + .50) = .93$, $\rho_0(\delta, \delta + .75) = .85$, $\rho_0(\delta, \delta + 1) = 0.75$ to a close degree of approximation for all δ. The other sections of Table 3 show the correlations for $k = 1, 2, 5$, and 10 to illustrate the patterns observed.

The pattern is clear. For most columns the highest correlation is reached at $\theta = 1$ or at a neighbouring value. Figures 1-4 shows these values diagramatically. The numbers 1, 2, ..., 8 at x and y axes corresponds to θ, $\delta = 0.25, 0.5, \ldots, 2$ respectively. It is seen clearly that the autocorrelation function $\rho_k(\theta, \delta)$ is a smooth changing concave surface in three dimensional

TABLE 3C

Correlation at Lag $k = 2$

δ, θ	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
0.25	.21	.23	.25	.25	.24	.23	.20	.17
0.50	.23	.26	.28	.29	.28	.27	.24	.21
0.75	.24	.28	.30	.31	.31	.30	.27	.24
1.00	.24	.28	.31	.32	.33	.32	.29	.27
1.25	.23	.27	.30	.32	.33	.32	.30	.28
1.50	.21	.25	.28	.30	.31	.31	.30	.28
1.75	.19	.22	.25	.27	.28	.28	.28	.26
2.00	.16	.19	.21	.23	.25	.25	.25	.23

TABLE 3D

Correlation at Lag $k = 5$

δ, θ	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
0.25	.22	.24	.25	.24	.23	.21	.18	.16
0.50	.24	.26	.28	.28	.27	.25	.22	.19
0.75	.25	.28	.29	.30	.29	.27	.24	.22
1.00	.25	.28	.30	.30	.30	.28	.26	.23
1.25	.23	.26	.28	.30	.30	.28	.27	.24
1.50	.21	.24	.26	.27	.28	.27	.26	.24
1.75	.18	.20	.23	.24	.25	.24	.23	.22
2.00	.15	.17	.19	.20	.21	.21	.20	.19

TABLE 3E

Correlation at Lag $k = 10$

δ, θ	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
0.25	.19	.21	.21	.21	.20	.18	.15	.13
0.50	.21	.23	.24	.24	.22	.20	.18	.15
0.75	.21	.24	.25	.25	.24	.22	.19	.16
1.00	.21	.23	.25	.25	.24	.22	.19	.16
1.25	.19	.22	.23	.24	.23	.21	.18	.16
1.50	.17	.19	.21	.21	.21	.19	.17	.14
1.75	.14	.16	.18	.18	.18	.17	.15	.13
2.00	.11	.13	.14	.15	.15	.14	.13	.11

space. For each Figure, it is true the $\rho_k(\theta, \delta)$ increases with θ, δ when they are less than 1 and then decreases after they become bigger than 1.

Thus on this evidence, the Machina effect is supported, but later results are less supportive.

5 A Parsimonious Model

In this section a parsimonius model is proposed for the representation $r_t = s_t a_t$. Although it is able to capture, or at least closely approximate, several of the observed empirical features of the data, and is very simple, it certainly is *not* perfectly correct. The suggested model is:

(*a*) s_t, a_t are independent.

(*b*) s_t is iid. No dependence.

(*c*) prob $(s_t < 0)$ = prob $(s_t > 0)$. Symmetry.

(*d*) $a_t \sim I(d)$, $d = 1/2$. Long memory.

(*e*) The marginal distribution of a_t is exponential, with a parameter that changes through time.

If (*a*) were true, a consequence is that the marginal distribution of a_t when $s_t > 0$ is the same as the marginal distribution of a_t when $s_t < 0$. Using data with means subtracted and without outliers reduced, for the full sample one finds prob $(s_t > 0) = 0.51$ and the two distribution have moments

	mean	st.dev	skewness	kurtosis
$a_t \mid s_t > 0$	.007	.008	4.352	38.53
$a_t \mid s_t < 0$	.007	.008	4.841	59.82

A Kolmogorov-Smirnov test of the hypothesis H_0 : the two distributions are identical is rejected at a very high level of significance. However, for an extremely large sample, virtually every simple hypothesis will be rejected. The correlation between a_t, s_t is 0.02 and altogether (*a*) is seen to be at least a reasonable approximation to the truth. This is also found to hold for all of the sub samples, the means and variances of the two distributions for $s_t < 0$, $s_t > 0$ are very similar, as shown in Table 5, and the Kolmogorov-Smirnov test rejects the null for only one of the ten sub-samples at the 5% significance level. The details are not shown.

Table 4 shows the first three autocorrelatons of s_t and also estimates of $P = \text{prob}(s_t > 0)$.

Note, if a true probability is $p = 1/2$, with a sample size of 1,700, the 95% confidence interval for $\hat{P}$ is .488 to .512. Similarly $2/\sqrt{N}$ is approximately 0.05 for N=1,700 and 0.015 for $N = 17,000$, so several of the autocorrelations are significantly different from zero, particularly for lag one, according to the usual criterion.

Other autocorrelations up to the lag ten, not shown, rarely reach significant values. Overall, the Table indicates that statements (*b*) and (*c*) of the parsimonious model are not correct but are still providing a reasonable first approximation. When the outlier reduced data is used very similar results

TABLE 4

Original Data

	Prob $(s_t > 0)$	Autocorrelations s_t		
		ρ_1	ρ_2	ρ_3
Full Sample	0.51	.087	-.042	-.008
Sub-sample 1	.52	.016	−.022	.06
2	.50	−.047	−.046	.019
3	.50	.069	−.068	−.040
4	.52	.097	−.112	−.066
5	.52	.150	−.060	−.028
6	.52	.141	−.031	.019
7	.53	.202	−.007	.015
8	.49	.152	−.030	.015
9	.49	.067	−.011	−.027
10	.52	.015	−.041	−.049

concerning (*a*), (*b*), (*c*) are obtained and are not shown here. The only differences are that the estimates of prob$(s_t < 0)$ is now often clearly smaller than a half and, of course, the kurtosis estimate is much reduced.

Property (*d*) of the parsimonious model was discussed in the previous section.

Concerning property (*e*), that the marginal distribution of a_t is exponential, it should be noted that this possibility was suggested because this distribution has the property that mean = standard deviation. As seen in Table 2, this property is often observed, both when using the original data and also the outlier reduced data. If the marginal distribution is $f(x) = \frac{1}{\alpha} e^{-\frac{x}{\alpha}}, x \geq 0$ then the mean is α. It is seen from Table 2 that for much of the subperiods α is estimated by the mean to fall in the region 0.005 to 0.007, but it is substantially larger in the first two sub-periods. A further property of the exponential distribution is that skewness = 2, kurtosis = 9 regardless of the values of α. It is seen from Table 2B that, with the outlier reduced data, skewness lies in the range 1.5 to 2.4 and kurtosis in the range 6.0 to 10.3 These empirical regularities certainly suggest that the exponential distribution is a plausible one for a_t for all sub-periods as well as for the full sample.

When the distribution of a_t is considered separately for $s_t > 0$ and for $s_t < 0$ the conclusion changes little, with the range of kurtosis now widening but is also based on a sample of half the size. The results are shown in Table 5.

To summarize, the parsimonious model is a very simple one that is not correct but still approximates many of the observed properties of the data.

TABLE 5

Mean Subtracted and Outlier Reduced Data Marginal Distributions of a_t for $s_t > 0$ and < 0

		mean	st.dev	skewness	kurtosis
Full Sample	> 0	.007	.007	2.49	11.20
	< 0	.007	.008	2.31	9.50
Sub Sample 1	> 0	0.14	.016	2.313	9.20
	< 0	0.16	.015	1.884	7.50
2	> 0	.011	.010	1.781	7.00
	< 0	.011	.011	1.923	7.50
3	> 0	.006	.006	2.302	10.40
	< 0	.006	.007	2.242	9.20
4	> 0	.006	.005	1.819	7.90
	< 0	.006	.007	2.097	8.00
5	> 0	.005	.004	1.591	6.50
	< 0	.005	.005	1.695	6.50
6	> 0	.004	.004	2.230	11.10
	< 0	.005	.005	1.977	8.10
7	> 0	.004	.004	2.115	8.90
	< 0	.005	.004	1.565	5.90
8	> 0	.007	.006	1.784	7.20
	< 0	.007	.006	1.430	5.50
9	> 0	.007	.006	1.601	6.30
	< 0	.007	.006	1.318	5.20
10	> 0	.007	.006	1.950	8.90
	< 0	.007	.007	2.322	10.60

In the following table it is compared to a less simple but probably superior model:

Parsimonious Model	Improved Model
(*a*) s_t, a_t independent	same
(*b*) s_t *i.i.d.*	s_t obeys Markov process
(*c*) symmetry	not symmetric. Prob $(s_t > 0) = 0.52$, Prob $(s_t < 0) = 0.46$
(*d*) $a_t \sim I(d)$, $d = 1/2$	$I(d)$ but $d = 0.47$
(*e*) marginal distribution of a_t is exponential	same

However, coefficients of the model may change through time.

The question of why such properties arise in the data will be discussed in a later paper, where aggregation is considered.

6 Temporal Properties of Risk Measures

To ease exposition from now only two measures of risk will be considered,

$$\mu^a_{t,k} = E\left[a_{t+k} \mid I_t\right]$$
$$\text{and } \sigma^2_{t,k} = E\left[a^2_{t+k} \mid I_t\right]$$
$$\text{where } a_t = \mid r_t - m \mid .$$

The following Tables explore the extent and type of forecastability achieved by a simple model of the form suggested by (8) and the following results. The regressions run are

(*a*) a_t^2 on const., a^2_{t-k}, a_{t-k},

(*b*) a_t on const., a^2_{t-k}, a_{t-k}, for k = 1, 2, 5, 10, 20, 50 and 100 in each case.

Table 6 A1 shows the results from the full sample case, for regression (*a*) using the original data and 6 A2 shows results for the same regression using

TABLE 6A

Regression Results for Full Sample $r_t^2 = \beta_0 + \beta_1 r^2_{t-k} + \beta_2 \mid r_{t-k} \mid$.

TABLE 6A1

Original Data Mean Substracted

k	β_0	(t)	β_1	(t)	β_2	(t)	R^2	D-K
1	.00003	(3.70)	.0830	(7.27)	.0131	(15.55)	.061	2.09
2	.00004	(6.20)	.1325	(11.61)	.0098	(11.64)	.062	1.74
5	.00005	(6.85)	.0924	(8.02)	.0098	(11.59)	.045	1.68
10	.00005	(6.49)	-.0174	(−1.49)	.0121	(14.09)	.023	1.64
20	.00004	(5.64)	−.0580	(−4.97)	.0137	(15.91)	.021	1.62
50	.00007	(9.84)	−.0271	(−2.30)	.0089	(10.31)	.010	1.60
100	.00007	(10.19)	−.0471	(−4.00)	.0089	(10.28)	.008	1.59

TABLE 6A2

Outlier Reduced Data

k	β_0	(t)	β_1	(t)	β_2	(t)	R^2	D-W
1	.00007	(19.18)	.2578	(14.36)	.0020	(3.10)	.096	2.15
2	.00007	(18.95)	.2664	(14.88)	.0020	(3.12)	.101	1.69
5	.00006	(16.24)	.1678	(9.30)	.0048	(7.42)	.087	1.67
10	.00006	(17.00)	.1043	(5.70)	.0053	(8.06)	.061	1.57
20	.00006	(16.78)	.0972	(5.31)	.0055	(8.38)	.060	1.54
50	.00007	(19.51)	.0950	(5.13)	.0041	(6.12)	.042	1.49
100	.00008	(20.88)	.0770	(4.12)	.0036	(5.32)	.030	1.47

the outlier reduced data. Table 6B1 and 6B2 shows the corresponding results for regressions (*b*).

TABLE 6B

Regression Results for Full Sample $|r_t| = \beta_0 + \beta_1 r_{t-k}^2 + \beta_2 |r_{t-k}|$.

TABLE 6B1

Original Data Mean Substracted

k	β_0	(t)	β_1	(t)	β_2	(t)	R^2	D-W
1	.00501	(53.61)	.1386	(.91)	.3114	(27.88)	.102	2.16
2	.00511	(54.83)	.6369	(4.20)	.2881	(25.85)	.106	1.68
5	.00510	(54.31)	.0791	(.52)	.2997	(26.68)	.092	1.63
10	.00524	(54.92)	−.8923	(−5.75)	.2982	(26.13)	.063	1.56
20	.00517	(54.11)	−1.4086	(−9.07)	.3170	(27.73)	.061	1.52
50	.00564	(58.26)	−.9292	(−5.91)	.2445	(21.12)	.039	1.47
100	.00575	(58.97)	−1.3122	(−8.29)	.2372	(20.37)	.031	1.45

TABLE 6B2

Outlier Reducted Data

k	β_0	(t)	β_1	(t)	β_2	(t)	R^2	D-W
1	.00554	(56.21)	4.4143	(8.84)	.1556	(8.66)	.095	2.15
2	.00559	(56.76)	5.0186	(10.07)	.1402	(7.82)	.098	1.70
5	.00528	(53.40)	2.2290	(4.45)	.2267	(12.59)	.091	1.69
10	.00548	(54.80)	1.6225	(3.20)	.2072	(11.36)	.068	1.61
20	.00546	(54.55)	1.3807	(2.72)	.2138	(11.71)	.068	1.58
50	.00581	(57.25)	1.5328	(2.98)	.1640	(8.88)	.046	1.52
100	.00590	(57.78)	.9568	(1.85)	.1598	(8.58)	.037	1.50

There are several clear features of these tables:

(*a*) R^2 values generally increase in going from the original data to the outlier reduced data.

(*b*) R^2 values are generally higher for $|r_t|$ than for $|r_t^2|$ for corresponding k values. In terms of the summary statistics (R^2 and Durbin-Watson) the equations for r^2 with outlier-reduction are similar to $|r|$ using the original data.

(*c*) There can be substantial changes in coefficient values, and corresponding t's, in going from the original data to outlier reduced data, this being particularly true for β_1. When interpreting β_1, it should be remembered that daily returns are typically very small, so that $|r|$( ] is generally much larger in magnitude (with mean .007) compared to r^2 (with mean about .0003).

For the original data both r^2 and $|r|$ are better explained by lagged $|r|$, according to the larger t-value, than by lagged r^2, although both terms are usually significant. For the outlier reduced data r^2 is explained equally by lagged $|r|$ and r^2, in terms of t-values, for $k \geq 5$ and $|r|$ is better explained by lagged $|r|$ for $k \geq 5$. Thus, for the full sample, the Machina effect still seems to be relevant.

However, turning to the ten sub-samples leads to a less clear picture. Table 7 summarizes some of the evidence, showing the regression results as before with r_t^2 and $|r_t|$ as the dependent variables, using just $k = 5$ and 20 and the outlier reduced, mean subtracted data. Table 7A shows the results for $k = 5$, Table 7B for $k = 20$. It is seen that periods 3, 4, 5, 9 and 10 often produce very low R^2 values, corresponding roughly to the years 1940 to 1959 and 1980 to 1991. For the whole data set and for half of the subsamples, these simple regressions appear to find some forecastability for $|r_t|$ and r_t^2. For the sub-samples there is little evidence for the Machina

TABLE 7A

Regression Results for Sub-Samples, Outlier Reduced, $k = 5$

TABLE 7A1

$r_t^2 = \beta_0 + \beta_1 r_{t-5}^2 + \beta_2 |r_{t-5}|$

	β_0	(t)	β_1	(t)	β_2	(t)	R^2	D-W
Sample 1	.00020	(4.44)	-.0751	(-1.24)	.0202	(4.92)	.053	1.61
2	.00019	(9.11)	.1989	(3.38)	-.0008	(-.31)	.035	1.82
3	.00005	(6.51)	-.0409	(-.71)	.0047	(2.90)	.017	1.56
4	.00006	(8.37)	.0848	(1.45)	.0012	(.77)	.016	1.79
5	.00004	(10.91)	.2539	(4.22)	-.0019	(-1.74)	.027	1.65
6	.00003	(7.67)	.1266	(2.31)	.0010	(.91)	.031	1.40
7	.00003	(9.33)	.3172	(5.36)	-.0014	(-1.25)	.063	1.54
8	.00005	(8.12)	.2290	(3.91)	.0009	(.60)	.067	1.77
9	.00007	(10.21)	.0981	(1.62)	.0005	(.32)	.013	1.86
10	.00008	(8.13)	.1995	(3.73)	.0000	(.00)	.040	1.60
Full Sample	.00006	(16.24)	.1700	(9.30)	.0048	(7.42)	.087	1.67

TABLE 7A2

$|r_t| = \beta_0 + \beta_1 r_{t-5}^2 + \beta_2 |r_{t-5}|$

	β_0	(t)	β_1	(t)	β_2	(t)	R^2	D-W
Sub-Sample								
1	.01004	(15.41)	-2.1805	(-2.48)	.3956	(6.62)	.069	1.69
2	.01033	(22.17)	4.9120	(3.72)	-.0506	(-.85)	.032	1.85
3	.00486	(19.07)	-1.2293	(-.60)	.1700	(2.96)	.019	1.65
4	.00538	(20.45)	1.9207	(.89)	.0874	(1.49)	.019	1.76
5	.00461	(22.69)	8.8103	(2.69)	-.0281	(-.47)	.018	1.69
6	.00406	(20.80)	4.2196	(1.58)	.0876	(1.59)	.028	1.49
7	.00402	(20.82)	11.9680	(3.76)	.0220	(.37)	.059	1.56
8	.00542	(20.19)	6.9389	(2.95)	.1014	(1.73)	.072	1.76
9	.00625	(22.45)	3.5146	(1.40)	.0382	(.63)	.015	1.92
10	.00636	(21.36)	3.6956	(2.24)	.0507	(.94)	.028	1.83
Full Sample	.00528	(53.40)	2.2300	(4.45)	.2267	(12.59)	.091	1.69

TABLE 7B

Regression Results for Sub-Samples, Outlier Reduced, $K = 20$

TABLE 7B1

$r_t^2 = \beta_0 + \beta_1 r_{t-20}^2 + \beta_2 |r_{t-20}|$

	β_0	(t)	β_1	(t)	β_2	(t)	R^2	D-W
Sub-Sample								
1	.00025	(5.52)	−.1521	(-2.47)	.0194	(4.61)	.024	1.49
2	.00019	(9.29)	.0760	(1.30)	.0009	(.35)	.010	1.76
3	.00007	(9.11)	.0763	(1.32)	−.0002	(−.13)	.005	1.53
4	.00007	(10.30)	.0803	(1.36)	−.0011	(−.72)	.002	1.74
5	.00004	(9.66)	−.0285	(−.46)	.0013	(1.17)	.001	1.63
6	.00003	(7.40)	.0896	(1.69)	.0015	(1.41)	.029	1.31
7	.00003	(7.85)	.0791	(1.31)	.0017	(1.52)	.028	1.45
8	.00006	(8.47)	.0925	(1.54)	.0019	(1.27)	.027	1.72
9	.00007	(10.56)	.1230	(2.02)	−.0002	(−.14)	.013	1.85
10	.00007	(7.18)	−.0434	(−.80)	.0045	(2.50)	.010	1.48
Full Sample	.00006	(16.78)	.0970	(5.00)	.0055	(8.38)	.060	1.54

TABLE 7B2

$|r_t| = \beta_0 + \beta_1 r_{t-20}^2 + \beta_2 |r_{t-20}|$

	β_0	(t)	β_1	(t)	β_2	(t)	R^2	D-W
Sub-Sample								
1	.01080	(16.26)	−2.8966	(−3.23)	.3735	(6.12)	.042	1.53
2	.00999	(21.41)	1.6893	(1.28)	.0361	(.60)	.013	1.80
3	.00555	(21.63)	2.9470	(1.43)	−.0062	(−.11)	.006	1.64
4	.00600	(22.54)	2.8975	(1.33)	−.0262	(−.44)	.003	1.71
5	.00459	(22.29)	−.6071	(−.18)	.0569	(.93)	.001	1.66
6	.00405	(21.04)	5.3201	(2.02)	.0691	(1.27)	.032	1.46
7	.00403	(20.44)	3.5002	(1.08)	.0942	(1.56)	.024	1.47
8	.00584	(21.08)	3.1859	(1.32)	.0848	(1.41)	.026	1.70
9	.00629	(22.39)	1.4591	(.58)	.0587	(.96)	.008	1.91
10	.00625	(20.69)	−.9752	(−.58)	.1312	(2.40)	.011	1.76
Full Sample	.00546	(54.55)	1.3800	(2.72)	.2138	(11.71)	.065	1.58

effect, with no clear difference between the forecastability of $|r|$ or r^2 at horizons of 5 and 20 days.

7 Taylor and Machina Effects from the Bivariate Exponential Distribution

If it is agreed that a_t is stationary and that a plausible marginal distribution is the exponential it follows that a plausible distribution for a pair a_t, a_s,

$t \neq s$ is a bivariate exponential. NAGAO and KADOYA [1971] consider such a distribution taking the form

$$p(x_1, x_2) = \frac{1}{\alpha^2 (1-\rho)} \exp\left[\frac{x_1 + x_2}{\alpha(1-\rho)}\right] \cdot I_o\left(\frac{2\sqrt{\rho}}{\alpha(1-\rho)} \sqrt{x_1 \cdot x_2}\right)$$

where $I_o(y)$ is a modified Bessel function of the first kind as defined as follows: $I_o(Z) = \sum_{n=0}^{\infty} \frac{Z^{2n}}{2^{2n}(n!)^2}$. Each marginal has an exponential distribution with mean and standard deviation α and the correlation between the pair of variables is ρ.

For this family of bivariate distributions, one has, by tedious but straightforward algebra, that,

$$\begin{aligned} cov(a_t^\delta, a_s^\theta) &= E(a_t^\delta - E a_t^\delta)(a_s^\theta - E a_s^\theta) \\ &= F(1+\delta), 1+\theta, 1, \rho)\Gamma(1+\delta)\Gamma(1+\theta) \\ &\quad - (1-\rho) F(1+\delta, 1, 1, \rho)\Gamma(1+\delta) \\ &\quad \times F(1, 1+\theta, 1, \rho)\Gamma(1+\theta) \\ \text{var}(a_t^\delta) &= E(a_t^\delta - E a_t^\delta)^2 \\ &= F(1+2\delta, 1, 1, \rho)\Gamma(1+2\delta) \\ &\quad - (1-\rho)[F(1+\delta, 1, 1, \rho)\Gamma(1+\delta)]^2, \end{aligned}$$

where $\Gamma(x)$ is the gamma function and $F(a, b, c, \rho)$ is the hypergeometric function defined as follows:

$$F(a, b, c, \rho) = \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)\Gamma(c)\rho^n}{\Gamma(a)\Gamma(b)\Gamma(c+n)\Gamma(n+1)}.$$

When a, b, c in the above formula are positive integers, it is possible to get a simple expression. Since the correlation between a_t^δ and a_s^θ is just

$$\rho(a_t^\delta, a_s^\theta) = \text{cov}(a_t^\delta, a_s^\theta) / \sqrt{\text{var}(a_t^\delta)\,\text{var}(a_s^\theta)},$$

one can also get a very simple formula for $\rho(a_t^\delta, a_s^\theta)$. Specifically, when $\delta = \theta = 2$ and $\delta = \theta = 3$, one has

$$\text{corr}(a_t^2, a_s^2) = \rho - \frac{1}{5}\rho(1-\rho),$$

and

$$\text{corr}(a_t^3, a_s^3) = \rho - \frac{1}{19}\rho(1-\rho)(10+\rho)$$

where ρ is the correlation coefficient of a_t and a_s. It is readily seen from the above relations that when $0 < \rho < 1$

$$\rho > \rho - \frac{1}{5}\rho(1-\rho) > \rho - \frac{1}{19}\rho(1-\rho)(10+\rho),$$

that is

$$\text{corr}(a_t, a_s) > \text{corr}(a_t^2, a_s^2) > \text{corr}(a_t^3, a_s^3).$$

So the propety we have called Taylor effect exists for this distribution when $\delta = \theta = 2$ and $\delta = \theta = 3$.

TABLE 8A

Correlations of a_t^δ ***and*** a_s^θ ***when*** $\mathrm{corr}(a_t,\ a_s) = 0.3.$

δ, θ	.25	.50	.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
.25	.2495	.2636	.2687	.2673	.2612	.2516	.2395	.2257	.2108	.1954	.1798	.1645
.50	.2636	.2801	.2870	.2870	.2818	.2728	.2609	.2470	.2318	.2159	.1996	.1834
.75	.2687	.2870	.2956	.2970	.2931	.2851	.2740	.2607	.2458	.2299	.2136	.1971
1.00	.2673	.2870	.2970	.3000	.2975	.2908	.2808	.2683	.2541	.2388	.2228	.2065
1.25	.2612	.2818	.2931	.2975	.2964	.2911	.2824	.2711	.2579	.2434	.2280	.2122
1.50	.2516	.2728	.2851	.2908	.2911	.2871	.2798	.2698	.2578	.2443	.2299	.2149
1.75	.2395	.2609	.2740	.2808	.2824	.2798	.2739	.2652	.2545	.2423	.2289	.2148
2.00	.2257	.2470	.2607	.2683	.2711	.2698	.2652	.2580	.2486	.2376	.2254	.2124
2.25	.2108	.2318	.2458	.2541	.2579	.2578	.2545	.2486	.2406	.2309	.2199	.2080
2.50	.1954	.2159	.2299	.2388	.2434	.2443	.2423	.2376	.2309	.2224	.2127	.2019
2.75	.1798	.1996	.2136	.2228	.2280	.2299	.2289	.2254	.2199	.2127	.2041	.1946
3.00	.1645	.1834	.1971	.2065	.2122	.2149	.2148	.2124	.2080	.2019	.1946	.1862

It can also be shown that $\text{corr}(a_t, a_s^2) = \frac{2}{\sqrt{5}}\rho$, so $\text{corr}(a_t, a_s^2) > \text{corr}(a_t^2, a_s^2)$ if $\rho < 2\sqrt{5} - 4 = 0.47$ and ρ lies in the range for the data encountered here.

Further, one gets

$$E[a_t^2 | a_s] = \beta_0 + \beta_1 a_s + \beta_2 a_s^2$$

which is the form of equation used in Tables 6 and 7 and suggests that the very simple analysis following (7) is, in fact, quite general if the bivariate exponential distributions, at least approximately, correct.

When δ, θ take other noninteger positive values, a numerical method has to be used to evaluate $\text{corr}(a_t^\delta, a_s^\theta)$. Table 8A shows values of $\text{corr}(a_t^\delta, a_s^\theta)$ for a variety of values of δ, θ with $\rho = 0.3$ (it might be noted that $\rho = 0.3$ is found in Table 3C). Similar tables were also formed for $\rho = 0.5$, 0.4, 0.2 and 0.1 but are not shown. The values on the main diagonal are for $\text{corr}(a_t^\delta, a_s^\delta \mid \rho = 0.3)$ and here, and also in the tables not shown, it is seen that this correlation takes its maximum at $\delta = 1$, which confirms that the Taylor effect is true for all the δ value we considered. Thus, the Taylor effect would seem to be a consequence of the a_t series having joint exponential distributions, if that occurs.

TABLE 8B

θ Which Maximizes $\text{corr}(a_t^\delta, a_s^\theta)$

δ	$\rho = 0.5$	$\rho = 0.4$	$\rho = 0.3$	$\rho = 0.2$	$\rho = 0.1$
0.25	.75	.75	.75	1.00	1.00
0.50	.75	.75	.75	1.00	1.00
0.75	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00
1.25	1.00	1.00	1.00	1.00	1.00
1.50	1.25	1.25	1.25	1.00	1.00
1.75	1.25	1.25	1.25	1.25	1.00
2.00	1.50	1.50	1.25	1.25	1.00
2.25	1.75	1.50	1.25	1.25	1.00
2.50	1.75	1.75	1.50	1.25	1.25
2.75	2.00	1.75	1.50	1.25	1.25
3.00	2.00	1.75	1.50	1.50	1.25

Looking across rows, the Machina effect would suggest that in each row $\text{corr}(a_t^\delta, a_t^1)$ is the maximum, but this is seen not to be correct. Table 8B shows the value θ which maximizes $\text{corr}(a_t^\delta, a_s^\theta)$ for each δ and for each of the ρ values considered. It is seen that for $\rho \leq 0.3$ θ is equal or near to one (rage 0.75 to 1.25) for most of the δ values considered. Thus, although the Machina effect is not theoretically perfectly correct, it is seen to be a good approximation to the truth for a wide range of realistic values. If one wants a simple model relating a_{t+n} to the past, a_t is a sensible explanatory variable.

8 Alternative ARCH Models under Different Distributional Assumptions

In the previous paper by DING, GRANGER, and ENGLE [1993], various ARCH models were estimated and analyzed for the same return series used here under the assumption that the conditional distribution is normal. It was shown by simulation that the Taylor effect does not necessarily imply that the correct model should be Taylor and Schwert's (TAYLOR [1986], SCHWERT [1990]) ARCH model in absolute returns. A nested test also showed that both Engle and Bollerslev's (ENGLE [1982] and BOLLERSLEV [1986]) ARCH in squared returns model and Taylor and Schwert's ARCH is absolute returns model can be rejected if the alternative is the asymmetric power ARCH model. In this section a further comparison will be presented for various ARCH models under different distributional assumptions. Four different specifications for the conditional standard deviation (variance) equation will be considered. They are:

(1) Taylor and Schwert's ARCH in absolute return model denoted as TS:

$$\sigma_t = \alpha_0 + \alpha_1 \,|\, \varepsilon_{t-1} \,|\, + \beta_1 \, \sigma_{t-1}$$

(2) Engle and Bollerslev's ARCH in squared return model denoted as EB:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \, \varepsilon_{t-1}^2 + \beta_1 \, \sigma_{t-1}^2$$

(3) Ding, Granger, and Engle's asymmetric power ARCH model denoted as DGE:

$$\sigma_t^\delta = \alpha_0 + \alpha_1 \left(|\, \varepsilon_{t-1} \,| + \gamma \varepsilon_{t-1} \right)^\delta + \beta_1 \, \sigma_{t-1}^\delta$$

(4) Nelson's exponential ARCH [NELSON (1991)] model denoted as NEL:

$$\log \sigma_t = \alpha_0 + \alpha_1 \left(\frac{|\, \varepsilon_{t-1} \,| + \gamma \varepsilon_{t-1}}{\sigma_{t-1}} \right) + \beta_1 \log \sigma_{t-1}.$$

The mean equation is specified as follows:

$$\Gamma_t = c_o + c_1 \, \varepsilon_{t-1} + \varepsilon_t$$

$$\varepsilon_t = \sigma_t \, e_t, \; e_t \sim i.i.d. \, (0, \, 1)$$

which is the same as in DING, GRANGER, and ENGLE [1993] except that two different distributional assumptions will be considered here,

(1) $e_t \sim i.i.d. \; N(0, \, 1)$

(2) $e_t \sim i.i.d.$ double exponential with mean zero and variance 1.

Table 9 shows the estimated parameters, their t-statistics (in parentheses) and also the likelihood values for different models. It is seen that, for all the four different specifications for the variance equation, the conditional double exponential distribution yields higher likelihood values than the conditional normal distribution, although they all have the same number of parameters. The differences under the alternative distributional assumption for the TS, EB, DGE, and NEL models are 374, 420, 309, and 308 respectively. It is also seen that both the Taylor and Schwert model and the Engle and Bollerslev model can be rejected if the alternative is the asymmetric power

ARCH no matter what conditional distribution is used. It is also interesting to note that the asymmetric power ARCH model, with one more parameter, also has higher likelihood values than Nelson's Exponential ARCH model. From this discussion, it seems reasonable to conclude that the conditional double exponential distribution is a better approximation to the data than a conditional normal distribution.

TABLE 9

Estimated Results for various ARCH

model		normal distribution		double exponential	
	likelihood	56776.6		571198.0	
TS	c_0	.401E-3	(7.0)	.587E-3	(12.6)
	c_1	.138	(19.6)	.113	(18.3)
	α_0	.990E-4	(12.9)	.894E-4	(5.7)
	α_1	.103	(67.3)	.988E-1	(18.3)
	β_1	.913	(517.0)	.921	(202.3)
	likelihood	56822.4		57198.1	
EB	c_0	.401E-3	(7.2)	.598E-3	(12.8)
	c_1	.144	(18.4)	.115	(18.3)
	α_0	.807E-6	(12.7)	.820E-6	(5.9)
	α_1	.934E-1	(50.1)	.962E-1	(15.5)
	β_1	.904	(420.0)	.910	(170.7)
	likelihood	56974.1		57284.9	
	c_0	.210E-3	(3.2)	.484E-3	(10.5)
	c_1	.145	(18.9)	.116	(18.6)
DGE	α_0	.146E-4	(4.5)	.411E-4	(2.1)
	α_1	.866E-1	(32.4)	.914E-1	(14.6)
	β_1	.917	(453.0)	.922	(190.6)
	γ_1	–.372	(–21.1)	–.412	(–8.7)
	δ_1	1.428	(33.3)	1.197	(12.8)
	likelihood	56960.5		57268.5	
	c_0	.163E-3	(2.6)	.466E-3	(9.9)
NEL	c_1	.142	(19.1)	.115	(18.4)
	α_0	–.122	(–31.1)	–.155	(–13.1)
	α_1	.808E-1	(46.6)	.920E-1	(16.1)
	β_1	.987	(1387.0)	.981	(482.6)
	γ_1	–.419	(–23.7)	–.457	(–9.7)

APPENDIX

Proof of Luce's [1980] Results

If r has probability density function $f(x)$ then on re-scaling, $\alpha.r$ has $p.d.f.$

$$f_x(x) = \alpha^{-1} f(x/\alpha).$$

By the obvious iteration rule

$$(f_\alpha)_\beta(x) = \beta^{-1} f_\alpha(x/\beta) = (\alpha\beta)^{-1} f(x/\alpha\beta) = f_{\alpha\beta}(x).$$

Under the multiplicative assumption for risk on scaling

$$R(f_\alpha) = S(\alpha) R(f)$$

so

$$\begin{aligned} R(f_{\alpha\beta}) &= S(\alpha\beta) R(f) \\ &= S(\beta) R(f_\alpha) \\ &= S(\alpha) S(\beta) R(f) \end{aligned}$$

then

$$S(\alpha\beta) = S(\alpha) S(\beta).$$

Taking logs gives the functional equation

$$\log S(\alpha\beta) = \log S(\alpha) + \log S(\beta)$$

which has the solution

$$\log S(\alpha) = \theta \log \alpha,\ \theta > 0,$$

or equivalently,

$$S(\alpha) = \alpha^\theta,\ \theta > 0. \quad (\star)$$

From the definition $R(f) = E[T(x)] = \int T(x) f(x)$ and

$$\begin{aligned} R(f_\alpha) &= \alpha^{-1} \int T(x) f(x/\alpha)\, dx \\ &= \int T(\alpha x) f(x)\, dx \end{aligned}$$

so, from $(\star)$, $\alpha^\theta E[T(x)] = E[T(\alpha x)]$, *i.e.*

$$\int [\alpha^\theta T(x) - T(\alpha x)] f(x)\, dx = 0$$

and, as $f(x) \geq 0$, this give the functional equation

$$\alpha^\theta T(x) = T(\alpha x)$$

whose solution is $T(x) = A|x|^\theta$, $A > 0$, $\theta > 0$ (as the risk measure R is to be positive and real).

• References

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

Borch, K. (1969). – "A Note on Uncertainty and Indifferences Curves", *Review of Economic Studies*, 36, pp. 1-4.

Cao, C. Q. Tsay, R. S. (1992). – "Non-Linear Time-Series Analysis of Stock Volatilities", *Journal of Applied Econometrics,* 7, pp. 5161-5185.

Ding, Z., Granger, C. W. J., Engle, R. F. (1993). – "A Long-Memory Property of Stock Market Returns and a New Model", to appear, *Journal of Empirical Finance.*

Engle, R. F. (1982). – "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica,* 50, pp. 987-1008.

Geweke, J., Porter-Hudak, S. (1983). – "The Estimation and Application of Long-memory Time Series Models", *Journal of Time Series Analysis,* 4, pp. 221-238.

Granger, C. W. J. (1980). – "Long Memory Relationships and the Aggregation of Dynamic Models", *J. of Econometrics,* 14, pp. 227-238.

Granger, C. W. J. (1988). – "Models that Generate Trends", *Journal of Time Series Analysis,* 9, pp. 329-343.

Granger, C. W. J., Joyeux, R. (1980). – "An Introduction to Long-Memory Time Series Models and Fractional Differencing", *Journal of Time Series Analysis,* 1, pp. 15-29.

Luce, R. D. (1980). – "Several Possible Measures of Risk", *Theory and Decision,* 12, pp. 217-228.

Machina, M. J., Rothschild, M. (1987). – "Risk", *New Palgrave a Dictionary of Economics.* Ed. by J. Eatwell, M. Milgate and P. Newman, MacMillan, London.

Mandelbrot, B. B., van Ness, J. W. (1968). – "Fractional Brownian Motions, Fractional Noise and Applications", SIAM Rev. 10, pp. 422-437.

Nagao, M., Kadoya, M. (1971). – "Two-Variate Exponential Distribution and Its Numerical Table for Engineering Application", *Bulletin of the Disaster Prevention Research Institute,* 33, pp. 477-480.

Nelson, D. B. (1991). – "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica,* 59, No. 2, pp. 347-370.

Schwert, W. (1990). – "Stock Volatility and the Crash of 87", *Review of Financial Studies,* 3, No. 1, pp. 77-102.

Taylor, S. (1986). – *Modelling Financial Time Series,* J. Wiley and Sons, Chichester.