CHAPTER 13

Models for financial series

Because that's where the money is.

attributed to Willie Sutton

13.1 Financial series I: thinly traded shares on the Johannesburg Stock Exchange

One of the difficulties encountered in modelling the price series of shares listed on the Johannesburg Stock Exchange is that many of the shares are only thinly traded. The market is heavily dominated by institutional investors, and if for any reason a share happens not to be an 'institutional favourite' there will very likely be days, or even weeks, during which no trading of that share takes place. One approach is to model the presence or absence of trading quite separately from the modelling of the price achieved when trading does take place. This is analogous to the modelling of the sequence of wet and dry days separately from the modelling of the amounts of precipitation occurring on the wet days. It is therefore natural to consider, as models for the trading pattern of one or several shares, HMMs of the kind discussed by Zucchini and Guttorp (1991), who used them to represent the presence or absence of precipitation on successive days, at one or several sites.

In order to assess whether such models can be used successfully to represent trading patterns, data for six thinly traded shares were obtained from Dr D.C. Bowie, then of the University of Cape Town, and various models, including two-state HMMs, were fitted and compared. Of the six shares, three are from the coal sector and three from the diamonds sector. The coal shares are Amcoal, Vierfontein and Wankie, and the diamond shares Anamint, Broadacres and Carrigs. For all six shares the data cover the period from 5 October 1987 to 3 June 1991 (inclusive), during which time there were 910 days on which trading could take place. The data are therefore a multivariate binary time series of length 910.

13.1.1 Univariate models

The first two univariate models fitted to each of the six shares were a model assuming independence of successive observations and a Markov

Table 13.1 *Six thinly traded shares: minus log-likelihood values and BIC values achieved by five types of univariate model.*

Values of $-l$:

model	Amcoal	Vierf'n	Wankie	Anamint	Broadac	Carrigs
independence	543.51	629.04	385.53	612.03	599.81	626.88
Markov chain	540.89	611.07	384.57	582.64	585.76	570.25
second-order M. chain	539.89	606.86	383.87	576.99	580.06	555.67
2-state HMM, no trend	533.38	588.08	382.51	572.55	562.96	533.89
2-state HMM, single trend	528.07	577.51	381.28	562.55	556.21	533.88

Values of BIC:

model	Amcoal	Vierf'n	Wankie	Anamint	Broadac	Carrigs
independence	1093.83	1264.89	**777.88**	1230.88	1206.43	1260.58
Markov chain	1095.41	1235.77	782.77	1178.91	1185.15	1154.13
second-order M. chain	1107.03	1240.97	794.99	1181.23	1187.37	1138.59
2-state HMM, no trend	1094.01	1203.41	792.27	1172.35	1153.17	**1095.03**
2-state HMM, single trend	**1090.22**	**1189.08**	796.63	**1159.17**	**1146.49**	1101.83

chain (the latter fitted by conditional maximum likelihood). In all six cases, however, the sample ACF bore little resemblance to the ACF of the Markov chain model, and the Markov chain was therefore considered unsatisfactory. Two-state Bernoulli–HMMs, with and without time trend, and second-order Markov chains were also fitted, the second-order Markov chains by conditional maximum likelihood and the HMMs by unconditional. In the HMMs with trend, the probability of trading taking place on day t in state i is ${}_tp_i$, where

$$\operatorname{logit} {}_tp_i = a_i + bt;$$

the trend parameter b is taken to be constant over states. Such a model has five parameters. In the HMM without trend, b is zero, and there are four parameters. The resulting log-likelihood and BIC values are shown in Table 13.1.

From that table we see that, of the five univariate models considered, the two-state HMM with a time trend fares best for four of the six shares: Amcoal, Vierfontein, Anamint and Broadacres. Of these four shares, all but Anamint show a negative trend in the probability of trading taking place, and Anamint a positive trend. In the case of Wankie, the model assuming independence of successive observations is chosen by BIC, and in the case of Carrigs an HMM without time trend is chosen.

Since a stationary HMM is chosen for Carrigs, it is interesting to compare the ACF of that model with the sample ACF and with the

Table 13.2 *Trading of Carrigs Diamonds: first eight terms of the sample ACF, compared with the autocorrelations of two possible models.*

ACF of Markov chain	0.350	0.122	0.043	0.015	0.005	0.002	0.001	0.000
sample ACF	0.349	0.271	0.281	0.237	0.230	0.202	0.177	0.200
ACF of HM model	0.321	0.293	0.267	0.244	0.223	0.203	0.186	0.169

ACF of the competing Markov chain model. For the HMM the ACF is $\rho(k) = 0.3517 \times 0.9127^k$, and for the Markov chain it is $\rho(k) = 0.3499^k$. Table 13.2 displays the first eight terms in each case. It is clear that the HMM comes much closer to matching the sample ACF than does the Markov chain model; a two-state HMM can model slow decay in $\rho(k)$ from any starting value $\rho(1)$, but a two-state Markov chain cannot.

13.1.2 Multivariate models

Two-state multivariate HMMs of two kinds were then fitted to each of the two groups of three shares: a model without time trend, and one which has a single (logit-linear) time trend common to the two states

Table 13.3 *Comparison of several multivariate models for the three coal shares and the three diamond shares.*

Coal shares

model	k	$-l$	BIC
3 'independence' models	3	1558.08	3136.60
3 univariate HMMs, no trend	12	1503.97	3089.69
3 univariate HMMs with trend	15	1486.86	**3075.93**
multivariate HMM, no trend	8	1554.01	3162.52
multivariate HMM, single trend	9	1538.14	3137.60

Diamond shares

model	k	$-l$	BIC
3 'independence' models	3	1838.72	3697.88
3 univariate HMMs, no trend	12	1669.40	3420.56
3 univariate HMMs with trend	15	1652.64	3407.48
multivariate HMM, no trend	8	1590.63	3235.77
multivariate HMM, single trend	9	1543.95	**3149.22**

Table 13.4 *Coal shares: univariate HMMs with trend.*

share	t.p.m.	a_1	a_2	b
Amcoal	$\begin{pmatrix} 0.774 & 0.226 \\ 0.019 & 0.981 \end{pmatrix}$	-0.332	1.826	-0.001488
Vierfontein	$\begin{pmatrix} 0.980 & 0.020 \\ 0.091 & 0.909 \end{pmatrix}$	0.606	3.358	-0.001792
Wankie	$\begin{pmatrix} 0.807 & 0.193 \\ 0.096 & 0.904 \end{pmatrix}$	-5.028	-0.943	-0.000681

and to the three shares in the group. The first type of model has eight parameters, the second has nine. These models were then compared with each other and with the 'product models' obtained by combining independent univariate models for the individual shares. The three types of product model considered were those based on independence of successive observations and those obtained by using the univariate HMMs with and without trend. The results are displayed in Table 13.3.

It is clear that, for the coal shares, the multivariate modelling has not been a success; the model consisting of three independent univariate hidden Markov models with trend is 'best'. We therefore give these three univariate models in Table 13.4. In each of these models ${}_tp_i$ is the probability that the relevant share is traded on day t if the state of the underlying Markov chain is i, and $\text{logit}\, {}_tp_i = a_i + bt$.

For the diamond shares, the best model of those considered is the multivariate HMM with trend. In this model $\text{logit}\, {}_tp_{ij} = a_{ij} + bt$, where ${}_tp_{ij}$ is the probability that share j is traded on day t if the state is i. The transition probability matrix is

$$\begin{pmatrix} 0.998 & 0.002 \\ 0.001 & 0.999 \end{pmatrix},$$

the trend parameter b is -0.003160, and the other parameters a_{ij} are as follows:

share	a_{1j}	a_{2j}
Anamint	1.756	1.647
Broadacres	0.364	0.694
Carrigs	1.920	-0.965

Table 13.5 *Coal shares: means, medians and standard deviations of bootstrap sample of estimators of parameters of two-state HMMs with time trend.*

share		$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}$
Amcoal	mean	0.251	0.048	−1.40	2.61	−0.00180
	median	0.228	0.024	−0.20	1.95	−0.00163
	s.d.	0.154	0.063	4.95	3.18	0.00108
Vierfontein	mean	0.023	0.102	0.599	4.06	−0.00187
	median	0.020	0.097	0.624	3.39	−0.00190
	s.d.	0.015	0.046	0.204	3.70	0.00041
Wankie	mean	0.145	0.148	−14.3	0.09	−0.00081
	median	0.141	0.089	−20.9	−0.87	−0.00074
	s.d.	0.085	0.167	10.3	4.17	0.00070

13.1.3 Discussion

The parametric bootstrap, with a sample size of 100, was used to investigate the distribution of the estimators in the models for the three coal shares which are displayed in Table 13.4. In this case the estimators show much more variability than do the estimators of the HMM used for the binary version of the geyser eruptions series; see p. 146. Table 13.5 gives for each of the three coal shares the bootstrap sample means, medians and standard deviations for the estimators of the five parameters. It will be noted that the estimators of a_1 and a_2 seem particularly variable. It is, however, true that, except in the middle of the range, very large differences on a logit scale correspond to small ones on a probability scale; two models with very different values of a_1 (for instance) may therefore produce almost identical distributions for the observations. For all three shares the trend parameter b seems to be more reliably estimated than the other parameters. If one is interested in particular in whether trading is becoming more or less frequent, this will be the parameter of most substantive interest.

As regards the multivariate HMM for the three diamond shares, it is perhaps surprising that the model is so much improved by the inclusion of a single (negative) trend parameter. In the corresponding univariate models the time trend was positive for one share, negative and of similar magnitude for another share, and negligible for the remaining share. Another criticism to which this multivariate model is open is that the off-diagonal elements of the transition probability matrix of its underlying Markov chain are so close to zero as to be almost negligible; on average only one or two changes of state would take place during a sequence of 910 observations. Furthermore, it is not possible to interpret the two states as conditions in which trading is in general more likely and conditions in which it is in general less so. This is because the probability of trading (${}_tp_{ij}$) is not consistently higher for one state i than the other.

In view of the relative lack of success of multivariate HMMs in this application, these models are not pursued here. The above discussion does, we hope, serve as an illustration of the methodology, and suggests that such multivariate models are potentially useful in studies of this sort. They could, for instance, be used to model occurrences other than the presence or absence of trading, e.g. the price (or volume) rising above a given level.

13.2 Financial series II: a multivariate normal–HMM for returns on four shares

It is in practice often the case that a time series of share returns displays kurtosis in excess of 3 and little or no autocorrelation, even though the series of absolute or squared returns does display autocorrelation. These — and some other — phenomena are so common that they are termed 'stylized facts'; for discussion see for instance Rydén, Teräsvirta and Åsbrink (1998) or Bulla and Bulla (2007).

The model we shall discuss here is that the daily returns on four shares have a multivariate normal distribution selected from one of m such distributions by the underlying Markov chain, and that, conditional on the underlying state, the returns on a share at a given time are independent of those at any other time; that is, we assume longitudinal conditional independence. We do not, however, assume contemporaneous conditional independence. Indeed we impose here no structure on the m 4×4 variance-covariance matrices, one for each possible state, nor on the transition probability matrix $\mathbf{\Gamma}$. A model for p shares has in all $m\{m - 1 + p(p + 3)/2\}$ parameters: $m^2 - m$ to determine $\mathbf{\Gamma}$, mp state-dependent means, and $mp(p+1)/2$ state-dependent variances and covariances.

We evaluate the (log)-likelihood in the usual way and use `nlm` to minimize minus the log-likelihood, but in order to do so we need an unconstrained parametrization of the model. This is by now routine in respect of the transition probabilities (e.g. via the generalized logit transform). In addition, each of the m variance-covariance matrices $\mathbf{\Sigma}$ is parametrized in terms of $\mathbf{T}$, its unique Cholesky upper-triangular 'square root' such that the diagonal elements of $\mathbf{T}$ are positive; the relation between $\mathbf{\Sigma}$ and $\mathbf{T}$ is that $\mathbf{\Sigma} = \mathbf{T}'\mathbf{T}$.

We fitted models with two and three states to 500 consecutive daily

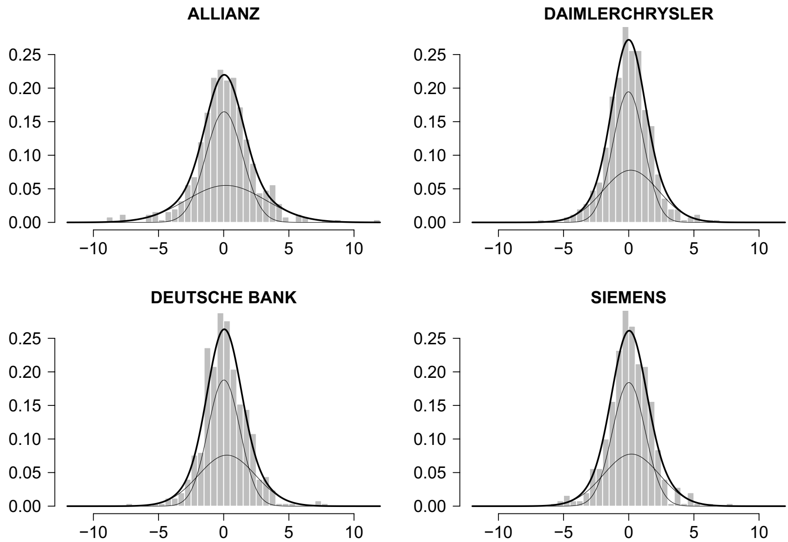

Figure 13.1 *Two-state HMM for four share returns: marginal distributions compared with histograms of the returns. The state-dependent distributions, multiplied by their mixing probabilities, are also shown.*

returns on the following four shares: Allianz (ALV), Deutsche Bank (DBK), DaimlerChrysler (DCX), and Siemens (SIE). These are all major components of the DAX30 index, with weights (as of 31 March 2006), of 8.71%, 7.55%, 6.97% and 10.04%. The 501 trading days used were 4 March 2003 to 17 February 2005, both inclusive. We computed the daily returns as $100 \log(s_t/s_{t-1})$, where s_t is the price on day t.

Of these two models the one that fitted better by BIC (but not AIC) was that with two states. In that case $-l = 3255.157$, there are 30 parameters, AIC = 6570.3 and BIC = 6696.8. For the three-state model the corresponding figures are $-l = 3208.915$, 48 parameters, AIC = 6513.8 and BIC = 6716.1. The details of the two-state model, which is depicted in Figure 13.1, are as follows.

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0.918 & 0.082 \\ 0.059 & 0.941 \end{pmatrix} \qquad \boldsymbol{\delta} = (0.418, 0.582)$$

state-dependent means:

	ALV	DBK	DCX	SIE
state 1	0.172	0.154	0.252	0.212
state 2	0.034	−0.021	0.021	0.011

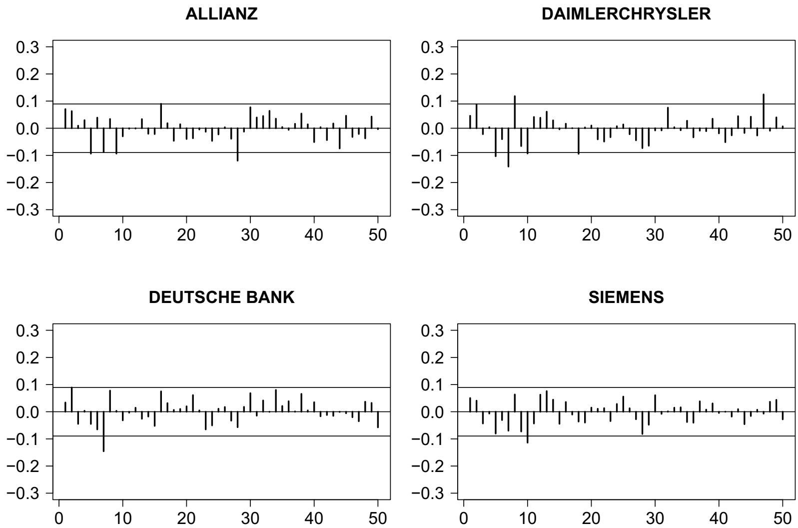

Figure 13.2 *Four share returns: sample ACF of returns.*

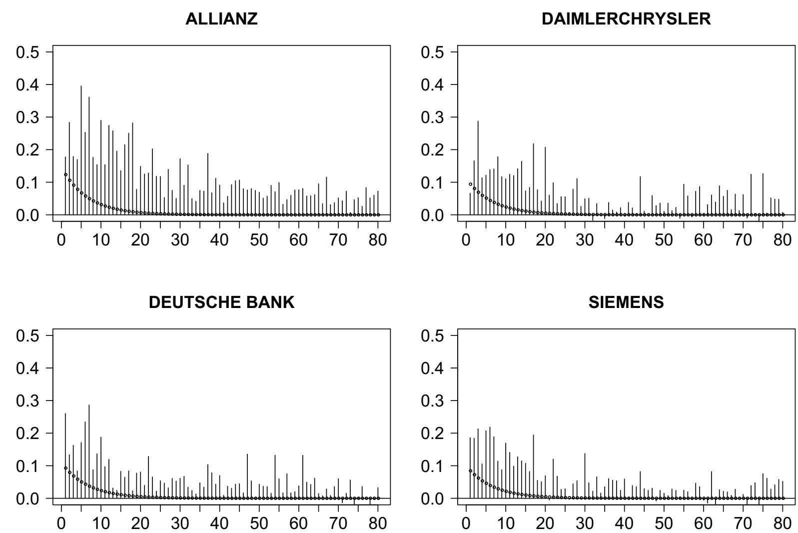

Figure 13.3 *Four share returns: sample ACF of squared returns, plus ACF of squared returns for two-state HMM (smooth curve).*

$\mathbf{\Sigma}$ for states 1 and 2 respectively:

$$\begin{pmatrix} 9.162 & 3.981 & 4.496 & 4.542 \\ 3.981 & 4.602 & 2.947 & 3.153 \\ 4.496 & 2.947 & 4.819 & 3.317 \\ 4.542 & 3.153 & 3.317 & 4.625 \end{pmatrix} \quad \begin{pmatrix} 1.985 & 1.250 & 1.273 & 1.373 \\ 1.250 & 1.425 & 1.046 & 1.085 \\ 1.273 & 1.046 & 1.527 & 1.150 \\ 1.373 & 1.085 & 1.150 & 1.590 \end{pmatrix}$$

standard deviations:

	ALV	DBK	DCX	SIE
state 1	3.027	2.145	2.195	2.151
state 2	1.409	1.194	1.236	1.261

correlation matrices for states 1 and 2 respectively:

$$\begin{pmatrix} 1.000 & 0.613 & 0.677 & 0.698 \\ 0.613 & 1.000 & 0.626 & 0.683 \\ 0.677 & 0.626 & 1.000 & 0.703 \\ 0.698 & 0.683 & 0.703 & 1.000 \end{pmatrix} \quad \begin{pmatrix} 1.000 & 0.743 & 0.731 & 0.773 \\ 0.743 & 1.000 & 0.709 & 0.721 \\ 0.731 & 0.709 & 1.000 & 0.738 \\ 0.773 & 0.721 & 0.738 & 1.000 \end{pmatrix}$$

One notable feature of this model is the clear ordering by volatility that emerges for these shares. The two states are one of high and one of low volatility, for all four shares. There was also a clear ordering by volatility in the three-state model; the states were of high, intermediate and low volatility, i.e. for all four shares. For a different set of shares, however, this ordering might well not emerge.

A feature of the two-state model, but not the three-state model, is that the ranges of correlations within each state are narrow and non-overlapping: 0.61–0.70 in state 1 (the more volatile state), and 0.71–0.77 in state 2.

We now compare properties of the two-state model with the corresponding properties of the observations. For all four shares, means and standard deviations agree extremely well. The same is not true of kurtosis. For the model the kurtoses of the four shares are 4.51 3.97 3.94, and 3.82, and for the observations the figures are 6.19, 4.69, 5.43, and 4.81 respectively. Although the model gives kurtosis well above 3 for each of the shares, the sample kurtosis is in each case much higher. However, much of this discrepancy is due to a few outlying returns in excess of 6% in absolute value, especially at the start of the period. The kurtosis of Allianz, in particular, is much reduced if one caps absolute returns at 6%.

We also compared sample and model autocorrelations of returns, autocorrelations of squared returns, and cross-correlations at lag zero. These cross-correlations match extremely well, and the ACFs of returns well (in that the model autocorrelations are very low and those of the sample negligible): see Figure 13.2. But the ACFs of squared returns do not match well: see Figure 13.3. This is similar to the finding of Rydén, Teräsvirta and Åsbrink (1998) that HMMs could not capture the sample ACF behaviour of absolute returns. Bulla and Bulla (2007) discuss the use of hidden *semi*-Markov models rather than HMMs in order to represent the squared-return behaviour better.

13.3 Financial series III: discrete state-space stochastic volatility models

Stochastic volatility (SV) models have attracted much attention in the finance literature; see e.g. Shephard (1996) and Omori *et al.* (2007). The first form of the model which we consider here, and the best-known, is the (Gaussian) SV model without leverage.

13.3.1 Stochastic volatility models without leverage

In such a model, the returns x_t on an asset $(t = 1, 2, \dots, T)$ satisfy

$$x_t = \varepsilon_t \beta \exp(g_t/2), \qquad g_{t+1} = \phi g_t + \eta_t, \tag{13.1}$$

where $|\phi| < 1$ and $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independent sequences of independent normal random variables, with ε_t standard normal and $\eta_t \sim \mathrm{N}(0, \sigma^2)$. (This is the 'alternative parametrization' in terms of β indicated by Shephard (1996, p. 22), although we use ϕ and σ where Shephard uses γ_1 and σ_η.) We shall later allow ε_t and η_t to be dependent, in which case the model will be said to accommodate leverage; for the moment we exclude that possibility. This model has three parameters, β, ϕ and σ, and for identifiability reasons we constrain β to be positive. The model is both simple, in that it has only three parameters, and plausible in its structure, and some properties of the model are straightforward to establish. For instance, if $\{g_t\}$ is stationary, it follows that $g_t \sim \mathrm{N}(0, \sigma^2/(1-\phi^2))$. We shall assume that $\{g_t\}$ is indeed stationary.

But the principal difficulty in implementing the model in practice has been that it does not seem possible to evaluate the likelihood of the observations directly. Much ingenuity has been applied in the derivation and application of (*inter alia*) MCMC methods of estimating the parameters even in the case in which ε_t and η_t are assumed to be independent. By 1996 there was already a 'vast literature on fitting SV models' (Shephard, 1996, p. 35).

The dependence structure of such a model is precisely that of an HMM, but, unlike the models we have discussed so far, the underlying Markov process $\{g_t\}$ — essentially the log-volatility — is continuous-valued, not discrete. What we describe here, as an alternative to the estimation methods described, e.g., by Shephard (1996) and Kim, Shephard and

Chib (1998), is that the state-space of $\{g_t\}$ should be discretized into a sufficiently large number of states to provide a good approximation to continuity, and the well-established techniques and tractability of HMMs exploited in order to estimate the parameters of the resulting model. In particular, the ease of computation of the HMM likelihood enables one to fit models by numerical maximization of the likelihood, and to compute forecast distributions. The transition probability matrix of the Markov chain is structured here in such a way that an increase in the number of states does not increase the number of parameters; only the three parameters already listed are used.

The likelihood of the observations $\mathbf{x}^{(T)}$ is given by the T-fold multiple integral

$$p(\mathbf{x}^{(T)}) = \int \dots \int p(\mathbf{x}^{(T)}, \mathbf{g}^{(T)})\, \mathrm{d}\mathbf{g}^{(T)},$$

the integrand of which can be decomposed as

$$\begin{aligned} p(g_1) \prod_{t=2}^{T} p(g_t \mid g_{t-1}) \prod_{t=1}^{T} p(x_t \mid g_t) & \\ &= p(g_1)\, p(x_1 \mid g_1) \prod_{t=2}^{T} p(g_t \mid g_{t-1})\, p(x_t \mid g_t), \end{aligned}$$

where p is used as a general symbol for a density. By discretizing the range of g_t sufficiently finely, we can approximate this integral by a multiple sum, and then evaluate that sum recursively by the methods of HMMs.

In detail, we proceed as follows. Let the range of possible g_t-values be split into m intervals (b_{i-1}, b_i), $i = 1, 2, \dots, m$. If $b_{i-1} < g_t \leq b_i$, we shall say that g_t is in state i and denote this, as usual, by $C_t = i$. We denote by g_i^* a representative point in (b_{i-1}, b_i), e.g. the midpoint. The resulting T-fold sum approximating the likelihood is

$$\begin{aligned} &\sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \dots \sum_{i_T=1}^{m} \Pr(C_1 = i_1)\, n(x_1; 0, \beta^2 \exp(g_{i_1}^*)) \\ &\quad \times \prod_{t=2}^{T} \Pr(C_t = i_t \mid C_{t-1} = i_{t-1})\, n(x_t; 0, \beta^2 \exp(g_{i_t}^*)), \quad (13.2) \end{aligned}$$

and the transition probability $\Pr(C_t = j \mid C_{t-1} = i)$ is approximated by

$$\begin{aligned} \gamma_{ij} &= N(b_j; \phi g_i^*, \sigma^2) - N(b_{j-1}; \phi g_i^*, \sigma^2) \\ &= \Phi((b_j - \phi g_i^*)/\sigma) - \Phi((b_{j-1} - \phi g_i^*)/\sigma). \end{aligned}$$

Here $n(\bullet; \mu, \sigma^2)$ is used to denote a normal density with mean μ and variance σ^2, N the corresponding (cumulative) distribution function, and Φ the standard normal distribution function. We are in effect saying that, although g_t has (conditional) mean ϕg_{t-1}, we shall proceed as if that mean were $\phi \times$ the midpoint of the interval in which g_{t-1} falls. With the assumption that the approximating Markov chain $\{C_t\}$ is stationary, (13.2) therefore gives us the usual matrix expression for the likelihood:

$$\boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2)\ldots\boldsymbol{\Gamma}\mathbf{P}(x_T)\mathbf{1}' = \boldsymbol{\delta}\boldsymbol{\Gamma}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2)\ldots\boldsymbol{\Gamma}\mathbf{P}(x_T)\mathbf{1}',$$

$\boldsymbol{\delta} = \mathbf{1}(\mathbf{I} - \boldsymbol{\Gamma} + \mathbf{U})^{-1}$ being the stationary distribution implied by the t.p.m. $\boldsymbol{\Gamma} = (\gamma_{ij})$, and $\mathbf{P}(x_t)$ being the diagonal matrix with i th diagonal element equal to the normal density $n(x_t; 0, \beta^2 \exp(g_i^*))$.

It is then a routine matter to evaluate this approximate likelihood and maximize it with respect to the three parameters of the model, transformed to allow for the constraints $\beta > 0$, $|\phi| < 1$, and $\sigma > 0$. However, in practice one also has to decide what m is, what range of g_t-values to allow for (i.e. b_0 to b_m), whether the intervals (b_{i-1}, b_i) should (e.g.) be of equal length, and which value g_i^* to take as representative of (b_{i-1}, b_i). Of these decisions, the choice of m can be expected to influence the accuracy of the approximation most. In the applications we describe in Sections 13.3.2 and 13.3.4, we have used equally spaced intervals represented by their midpoints.

13.3.2 Application: FTSE 100 returns

Shephard (1996, p. 39, Table 1.5) presents *inter alia* SV models fitted to the daily returns on the FTSE 100 index for the period 2 April 1986 to 6 May 1994; the return on day t is calculated as $100 \log(s_t/s_{t-1})$, where s_t is the index value at the close of day t. Using the technique described above, with a range of values of m and with g_t-values from -2.5 to 2.5, we have fitted models to FTSE 100 returns for that period.

Table 13.6 summarizes our findings and shows that the parameter estimates are reasonably stable by $m = 50$. Although we cannot expect our results to correspond exactly with those of Shephard, it is notable that our value of β (approximately 0.866) is very different from his (-0.452). The other two parameters are of roughly the same magnitude. We conjecture that the label β in column 1 of Shephard's Table 1.5 is inconsistent with his use of that symbol on p. 22, and therefore inconsistent with our usage. (We tested our code using simulated series.) Provisionally, our explanation for the apparent discrepancy is simply that the parameter β in Shephard's Table 1.5 is not the same as our parameter β.

Table 13.6 *SV model without leverage fitted to FTSE 100 returns, 2 April 1986 to 6 May 1994, plus comparable model from Shephard (1996).*

m	β	ϕ	σ
20	0.866	0.964	0.144
50	0.866	0.963	0.160
100	0.866	0.963	0.162
500	0.866	0.963	0.163
Shephard	β	γ_1	σ_η
(simulated EM)	−0.452	0.945	0.212

13.3.3 Stochastic volatility models with leverage

In the SV model without leverage, as described above, there is no feedback from past returns to the (log-) volatility process. As noted by Cappé *et al.* (2005, p. 28), this may be considered unnatural. We therefore discuss now a second, more general, form of the model.

As before, the returns x_t on an asset $(t = 1, 2, \ldots, T)$ satisfy

$$x_t = \varepsilon_t \beta \exp(g_t/2), \qquad g_{t+1} = \phi g_t + \eta_t, \tag{13.3}$$

where $|\phi| < 1$, but now ε_t and η_t are permitted to be dependent. More specifically, for all t

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}) \qquad \text{with} \qquad \mathbf{\Sigma} = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix},$$

and the vectors $\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}$ are assumed independent.

This model has four parameters: β, ϕ, σ and ρ. We constrain β to be positive. If one does so, the model is equivalent to the model (2) of Omori *et al.* (2007), via $\beta = \exp(\mu/2)$ and $g_t = h_t - \mu$. It is also — with notational differences — the discrete-time ASV1 model (2.2) of Yu (2005); note that our η_t corresponds to Yu's v_{t+1}. Yu contrasts the ASV1 specification with that of Jacquier, Polson and Rossi (2004), and concludes that the ASV1 version is preferable. The parameter ρ is said to measure leverage; it is expected to be negative, in order to accommodate an increase in volatility following a drop in returns. The structure of the model is conveniently represented by the directed graph in Figure 13.4.

The likelihood of the observations $\mathbf{x}^{(T)}$ is in this case also given by the multiple integral

$$\int \ldots \int p(\mathbf{x}^{(T)}, \mathbf{g}^{(T)})\, \mathrm{d}\mathbf{g}^{(T)}, \tag{13.4}$$

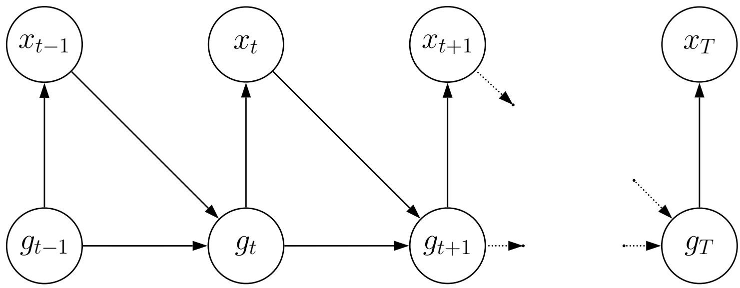

Figure 13.4 *Directed graph of stochastic volatility model with leverage.*

but here the integrand is decomposed as

$$p(g_1)\prod_{t=2}^{T} p(g_t \mid g_{t-1}, x_{t-1}) \prod_{t=1}^{T} p(x_t \mid g_t)$$
$$= \quad p(g_1)\, p(x_1 \mid g_1) \prod_{t=2}^{T} p(g_t \mid g_{t-1}, x_{t-1})\, p(x_t \mid g_t);$$

notice the dependence of $p(g_t \mid g_{t-1}, x_{t-1})$ on x_{t-1}.

We can approximate this integral as well by discretizing the range of g_t and evaluating the sum recursively. But in order to approximate this likelihood thus, we need the conditional distribution of g_{t+1} given g_t and x_t, or equivalently — since $x_t = \varepsilon_t \beta \exp(g_t/2)$ — given g_t and ε_t.

This is the distribution of η_t given ε_t, except that ϕg_t is added to the mean. The distribution of η_t given ε_t is $\mathrm{N}(\rho\sigma\varepsilon_t, \sigma^2(1-\rho^2))$; hence that of g_{t+1}, given g_t and ε_t, is

$$\mathrm{N}(\phi g_t + \rho\sigma\varepsilon_t, \sigma^2(1-\rho^2)).$$

Writing this distribution in terms of the observations x_t rather than the 'observation innovations' ε_t, we conclude that the required conditional distribution of g_{t+1} is

$$g_{t+1} \sim \mathrm{N}\left(\phi g_t + \frac{\rho\sigma x_t}{\beta\exp(g_t/2)}, \sigma^2(1-\rho^2)\right).$$

Hence $\Pr(g_t = j \mid g_{t-1} = i, x_{t-1} = x)$ is approximated by

$$\begin{aligned}\gamma_{ij}(x) &= N(b_j;\, \mu(g_i^*, x), \sigma^2(1-\rho^2)) - N(b_{j-1};\, \mu(g_i^*, x), \sigma^2(1-\rho^2)) \\ &= \Phi\left(\frac{b_j - \mu(g_i^*, x)}{\sigma\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{b_{j-1} - \mu(g_i^*, x)}{\sigma\sqrt{1-\rho^2}}\right),\end{aligned}$$

where we define

$$\mu(g_i^*, x) = \phi g_i^* + \frac{\rho\sigma x}{\beta \exp(g_i^*/2)}.$$

In this case, therefore, the approximate likelihood is

$$\boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}(x_1)\mathbf{P}(x_2)\boldsymbol{\Gamma}(x_2)\ldots\boldsymbol{\Gamma}(x_{T-1})\mathbf{P}(x_T)\mathbf{1}',$$

with $\boldsymbol{\delta}$ here representing the distribution assumed for C_1, and $\boldsymbol{\Gamma}(x_t)$ the matrix with entries $\gamma_{ij}(x_t)$. This raises the question of what (if any) distribution for g_1 will produce stationarity in the process $\{g_t\}$, i.e. in the case of the model with leverage. In the case of the model without leverage, that distribution is normal with zero mean and variance $\sigma^2/(1-\phi^2)$, and it is not unreasonable to conjecture that that may also be the case here.

Here we know that, given g_t and ε_t,

$$g_{t+1} \sim \mathrm{N}(\phi g_t + \rho\sigma\varepsilon_t, \sigma^2(1-\rho^2)).$$

That is,

$$g_{t+1} = \phi g_t + \rho\sigma\varepsilon_t + \sigma\sqrt{1-\rho^2}Z,$$

where Z is independently standard normal. If it is assumed that $g_t \sim \mathrm{N}(0, \sigma^2/(1-\phi^2))$, it follows that g_{t+1} is (unconditionally) normal with mean zero and variance given by $\phi^2(\sigma^2/(1-\phi^2)) + \rho^2\sigma^2 + \sigma^2(1-\rho^2) = \sigma^2/(1-\phi^2)$. In the 'with leverage' case also, therefore, the stationary distribution for $\{g_t\}$ is $\mathrm{N}(0, \sigma^2/(1-\phi^2))$, and that is the distribution we assume for g_1. The distribution we use for C_1 is that of g_1, discretized into the intervals (b_{i-1}, b_i).

13.3.4 Application: TOPIX returns

Table 13.7 *Summary statistics of TOPIX returns, calculated from opening prices 30 Dec. 1997 to 30 Dec. 2002, both inclusive.*

no. of returns	mean	std. dev.	max.	min.	+	−
1232	−0.02547	1.28394	5.37492	−5.68188	602	630

Using the daily opening prices of TOPIX (the Tokyo Stock Price Index) for the 1233 days from 30 December 1997 to 30 December 2002,

Table 13.8 *SV model with leverage fitted to TOPIX returns, opening prices 30 Dec. 1997 to 30 Dec. 2002, both inclusive, plus comparable figures from Table 5 of Omori* et al.

m	β	ϕ	σ	ρ
5	1.199	0.854	0.192	−0.609
10	1.206	0.935	0.129	−0.551
25	1.205	0.949	0.135	−0.399
50	1.205	0.949	0.140	−0.383
100	1.205	0.949	0.142	−0.379
200	1.205	0.949	0.142	−0.378
From Table 5 of Omori et al.*:*				
posterior mean,				
'unweighted'	1.2056	0.9511	0.1343	−0.3617
'weighted'	1.2052	0.9512	0.1341	−0.3578
95% interval	(1.089,1.318)	(0.908,0.980)	(0.091, 0.193)	(−0.593, −0.107)
Parametric bootstrap applied to the model with $m = 50$*:*				
95% CI:	(1.099, 1.293)	(0.826, 0.973)	(0.078, 0.262)	(−0.657, −0.050)
correlations:				
β		0.105	−0.171	0.004
ϕ			−0.752	−0.192
σ				0.324

both inclusive, we get a series of 1232 daily returns x_t with the summary statistics displayed in Table 13.7. This summary agrees completely with the statistics given by Omori *et al.* (2007) in their Table 4, although they state that they used closing prices. (We compute daily returns as $100 \log(s_t/s_{t-1})$, where s_t is the price on day t, and we use the estimator with denominator T as the sample variance of T observations. The data were downloaded from `http://index.onvista.de` on 5 July 2006.)

Again using a range of values of m, and with g_t ranging from -2 to 2, we have fitted an SV model with leverage to these data. The results are summarized in Table 13.8. All parameter estimates are reasonably stable by $m = 50$ and, as expected, the estimate of the leverage parameter ρ is consistently negative. The results for $m = 50$ agree well with the two sets of point estimates ('unweighted' and 'weighted') presented by Omori *et al.* in their Table 5, and for all four parameters our estimate is close to the middle of their 95% interval.

We also applied the parametric bootstrap, with bootstrap sample size

500, to our model with $m = 50$, in order to estimate the (percentile) bootstrap 95% confidence limits, the standard errors and the correlations of our estimators. Some conclusions are as follows. The standard error of $\hat{\rho}$ (0.160) is relatively and absolutely the highest of the s.e.s, and there is a high negative correlation (-0.752) between the estimators of ϕ and σ. Overall, our conclusions are consistent with the corresponding figures in Table 5 of Omori *et al.*

13.3.5 Discussion

Alternative forms of feedback from the returns to the volatility process in an SV model are possible, and indeed have been proposed and implemented in a discrete state-space setting by Rossi and Gallo (2006). They attribute to Calvet and Fisher (2001, 2004) the first attempt to build accessible SV models based on high-dimensional regime switching. Fridman and Harris (1998) also present and implement a model involving such feedback; see their Equation (7). Their route to the evaluation, and hence maximization, of the likelihood is via recursive numerical evaluation of the multiple integral which gives the likelihood; see Equation (13.4) above.

A convenient feature of the model structure and estimation technique we have described is that the normality assumptions are not crucial. Firstly, one can use as the state process some Markov chain other than one based on a discretized Gaussian AR(1). Secondly, one can replace the normal distribution assumed for ε_t by some other distribution.