9 Special cases of nonlinear and non-Gaussian models

9.1 Introduction

In this chapter we shall discuss the classes of non-Gaussian and nonlinear models that we shall consider in Part II of this book; we leave aside the analysis of observations generated by these models until later chapters.

A general form of the *nonlinear non-Gaussian state space model* is given by

$$y_t \sim p(y_t|\alpha_t), \qquad \alpha_{t+1} \sim p(\alpha_{t+1}|\alpha_t), \qquad \alpha_1 \sim p(\alpha_1), \tag{9.1}$$

for $t = 1, \ldots, n$. We shall assume throughout that

$$p(Y_n|\alpha) = \prod_{t=1}^{n} p(y_t|\alpha_t), \qquad p(\alpha) = p(\alpha_1) \prod_{t=1}^{n-1} p(\alpha_{t+1}|\alpha_t), \tag{9.2}$$

where $Y_n = (y_1', \ldots, y_n')'$ and $\alpha = (\alpha_1', \ldots, \alpha_n')'$. The observation density $p(y_t|\alpha_t)$ implies a relationship between the observation vector y_t and state vector α_t. The state update density $p(\alpha_{t+1}|\alpha_t)$ implies a relationship between the state vector of the next period α_{t+1} and the state of the current period α_t. If relationships in both $p(y_t|\alpha_t)$ and $p(\alpha_{t+1}|\alpha_t)$ are linear we say that the model is a *linear non-Gaussian state space model*. If all densities $p(y_t|\alpha_t)$, $p(\alpha_{t+1}|\alpha_t)$ and $p(\alpha_1)$ are Gaussian but at least one relationship in $p(y_t|\alpha_t)$ or $p(\alpha_{t+1}|\alpha_t)$ is nonlinear we say that the model is a *nonlinear Gaussian state space model*. In Section 9.2 we consider an important special form of the general linear non-Gaussian model and in Sections 9.3, 9.4, 9.5 and 9.6 we consider special cases of some subclasses of models of interest, namely exponential family models, heavy-tailed models, stochastic volatility model and other financial models. In Section 9.7 we describe some classes of nonlinear models of interest.

9.2 Models with a linear Gaussian signal

The multivariate model with a linear Gaussian signal that we consider here has a similar state space structure to (3.1) in the sense that observational vectors y_t are determined by a relation of the form

$$p\left(y_t|\alpha_1, \ldots, \alpha_t, y_1, \ldots, y_{t-1}\right) = p(y_t|Z_t\alpha_t), \tag{9.3}$$

Time Series Analysis by State Space Methods, Second Edition, J Durbin, S J Koopman © J Durbin, S J Koopman 2012. Published in 2012 by Oxford University Press

where the state vector α_t is determined independently of previous observations by the relation

$$\alpha_{t+1} = T_t\alpha_t + R_t\eta_t, \qquad \eta_t \sim \mathrm{N}(0, Q_t), \tag{9.4}$$

with the disturbances η_t being serially independent, for $t = 1, \ldots, n$. We define

$$\theta_t = Z_t\alpha_t, \tag{9.5}$$

and refer to θ_t as the *signal*. The density $p\,(y_t|\theta_t)$ can be non-Gaussian, nonlinear or both. In the case $p(y_t|\theta_t)$ is normal and θ_t is linear in y_t, the model reduces to the linear Gaussian model (3.1). While we begin by considering a general form for $p\,(y_t|\theta_t)$, we shall pay particular attention to three special cases:

1. Observations which come from exponential family distributions with densities of the form

$$p\,(y_t|\theta_t) = \exp[y_t'\theta_t - b_t(\theta_t) + c_t(y_t)], \qquad -\infty < \theta_t < \infty, \tag{9.6}$$

   where $b_t(\theta_t)$ is twice differentiable and $c_t(y_t)$ is a function of y_t only;
2. Observations generated by the relation

$$y_t = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim p(\varepsilon_t), \tag{9.7}$$

   where the ε_t's are non-Gaussian and serially independent.
3. Observations generated by a fixed mean but a stochastically evolving variance over time as in

$$y_t = \mu + \exp(\frac{1}{2}\theta_t)\varepsilon_t, \qquad \varepsilon_t \sim p(\varepsilon_t), \tag{9.8}$$

   where μ is the mean and the ε_t's are not necessarily Gaussian.

The model (9.6) together with (9.4) and (9.5) where η_t is assumed to be Gaussian was introduced by West, Harrison and Migon (1985) under the name the *dynamic generalised linear model*. The origin of this name is that in the treatment of non-time series data, the model (9.6), where θ_t does not depend on t, is called a *generalised linear model*. In this context θ_t is called the *link function*; for a treatment of generalised linear models see McCullagh and Nelder (1989). Further development of the West, Harrison and Migon model is described in West and Harrison (1997). Smith (1979, 1981) and Harvey and Fernandes (1989) gave an exact treatment for the special case of a Poisson observation with mean modelled as a local level model; their approach, however, does not lend itself to generalisation. In Section 9.3 we discuss a range of interesting and often used densities that are part of the exponential family.

The model (9.7) is similar to the linear Gaussian state space model of Part I with the difference that at least one element of the observation disturbance vector ε_t is non-Gaussian. The typical example is when the observations are contaminated by outliers. In such cases the Gaussian density is not sufficiently strong in the tails of the distribution. The Student's t distribution and the mixture of normals may be more appropriate for $p(\varepsilon_t)$. The details are provided in Section 9.4.

The model (9.8) is known as the *stochastic volatility* (SV) model and is regarded as the parameter-driven counterpart of the observation-driven *generalised autoregressive conditionally heteroscedasticity* (GARCH) model that is described by Engle (1982) and Bollerslev (1986). For a collection of articles that represents the developments of the SV model we refer to Shephard (2005).

9.3 Exponential family models

For model (9.6), let

$$\dot{b}_t(\theta_t) = \frac{\partial b_t(\theta_t)}{\partial \theta_t} \quad \text{and} \quad \ddot{b}_t(\theta_t) = \frac{\partial^2 b_t(\theta_t)}{\partial \theta_t \partial \theta_t'}. \tag{9.9}$$

For brevity, we will write $\dot{b}_t(\theta_t)$ as $\dot{b}_t$ and $\ddot{b}_t(\theta_t)$ as $\ddot{b}_t$ in situations where it is unnecessary to emphasise the dependence on θ_t. Assuming that the relevant regularity conditions are satisfied, it follows immediately by differentiating the relation $\int p(y_t|\theta_t)\,\mathrm{d}y_t = 1$ once and twice that

$$\mathrm{E}(y_t) = \dot{b}_t \quad \text{and} \quad \mathrm{Var}(y_t) = \ddot{b}_t.$$

Consequently $\ddot{b}_t$ must be positive definite for nondegenerate models. The standard results

$$\begin{aligned} \mathrm{E}\left[\frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t}\right] &= 0, \\ \mathrm{E}\left[\frac{\partial^2 \log p(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'}\right] + \mathrm{E}\left[\frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t}\frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t'}\right] &= 0, \end{aligned} \tag{9.10}$$

are obtained directly from (9.6).

9.3.1 Poisson density

For our first example of an exponential family distribution, suppose that the univariate observation y_t comes from a Poisson distribution with mean μ_t. For example, y_t could be the number of road accidents in a particular area during the month. Observations of this kind are called *count data*.

The logdensity of y_t is

$$\log p(y_t|\mu_t) = y_t \log \mu_t - \mu_t - \log(y_t!). \tag{9.11}$$

Comparing (9.11) with (9.6) we see that we need to take $\theta_t = \log \mu_t$ and $b_t = \exp \theta_t$ with $\theta_t = Z_t \alpha_t$, so the density of y_t given the signal θ_t is

$$p(y_t|\theta_t) = \exp[y_t\theta_t - \exp\theta_t - \log y_t!], \qquad t = 1, \ldots, n. \tag{9.12}$$

It follows that the mean $\dot{b}_t = \exp\theta_t = \mu_t$ equals the variance $\ddot{b}_t = \mu_t$. Mostly we will assume that η_t in (9.4) is generated from a Gaussian distribution but all or some elements of η_t may come from other continuous distributions.

9.3.2 Binary density

An observation y_t has a binary distribution if the probability that $y_t = 1$ has a specified probability, say π_t, and the probability that $y_t = 0$ is $1 - \pi_t$. For example, we could score 1 if Cambridge won the Boat Race in a particular year and 0 if Oxford won.

Thus the density of y_t is

$$p(y_t|\pi_t) = \pi_t^{y_t}(1-\pi_t)^{1-y_t}, \qquad y_t = 1, 0, \tag{9.13}$$

so we have

$$\log p(y_t|\pi_t) = y_t[\log \pi_t - \log(1-\pi_t)] + \log(1-\pi_t). \tag{9.14}$$

To put this in form (9.6) we take $\theta_t = \log[\pi_t/(1-\pi_t)]$ and $b_t(\theta_t) = \log(1 + e^{\theta_t})$, and the density of y_t given the signal θ_t is

$$p(y_t|\theta_t) = \exp[y_t\theta_t - \log(1 + \exp\theta_t)], \tag{9.15}$$

for which $c_t = 0$. It follows that mean and variance are given by

$$\dot{b}_t = \frac{\exp\theta_t}{1+\exp\theta_t} = \pi_t, \qquad \ddot{b}_t = \frac{\exp\theta_t}{(1+\exp\theta_t)^2} = \pi_t(1-\pi_t),$$

as is well-known.

9.3.3 Binomial density

Observation y_t has a binomial distribution if it is equal to the number of successes in k_t independent trials with a given probability of success, say π_t. As in the binary case we have

$$\log p(y_t|\pi_t) = y_t[\log \pi_t - \log(1-\pi_t)] + k_t \log(1-\pi_t) + \log \begin{pmatrix} k_t \\ y_t \end{pmatrix}, \tag{9.16}$$

with $y_t = 0, \ldots, k_t$. We therefore take $\theta_t = \log[\pi_t/(1-\pi_t)]$ and $b_t(\theta_t) = k_t \log(1+\exp\theta_t)$ giving for the density of y_t in form (9.6),

$$p(y_t|\theta_t) = \exp\left[y_t\theta_t - k_t\log(1+\exp\theta_t) + \log\begin{pmatrix} k_t \\ y_t \end{pmatrix}\right]. \qquad (9.17)$$

9.3.4 Negative binomial density

There are various ways of defining the negative binomial density; we consider the case where y_t is the number of independent trials, each with a given probability π_t of success, that are needed to reach a specified number k_t of successes. The density of y_t is

$$p(y_t|\pi_t) = \begin{pmatrix} k_t - 1 \\ y_t - 1 \end{pmatrix} \pi_t^{k_t}(1-\pi_t)^{y_t-k_t}, \qquad y_t = k_t, k_{t+1}, \ldots, \qquad (9.18)$$

and the logdensity is

$$\log p(y_t|\pi_t) = y_t\log(1-\pi_t) + k_t[\log\pi_t - \log(1-\pi_t)] + \log\begin{pmatrix} k_t - 1 \\ y_t - 1 \end{pmatrix}. \quad (9.19)$$

We take $\theta_t = \log(1-\pi_t)$ and $b_t(\theta_t) = k_t[\theta_t - \log(1-\exp\theta_t)]$ so the density in the form (9.6) is

$$p(y_t|\theta_t) = \exp\left[y_t\theta_t - k_t\{\theta_t - \log(1-\exp\theta_t)\} + \log\begin{pmatrix} k_t - 1 \\ y_t - 1 \end{pmatrix}\right]. \qquad (9.20)$$

Since in nontrivial cases $1-\pi_t < 1$ we must have $\theta_t < 0$ which implies that we cannot use the relation $\theta_t = Z_t\alpha_t$ since $Z_t\alpha_t$ can be negative. A way around the difficulty is to take $\theta_t = -\exp\theta_t^*$ where $\theta_t^* = Z_t\alpha_t$. The mean $\mathrm{E}(y_t)$ is given by

$$\dot{b}_t = k_t\left[1 + \frac{\exp\theta_t}{1-\exp\theta_t}\right] = k_t\left[1 + \frac{1-\pi_t}{\pi_t}\right] = \frac{k_t}{\pi_t},$$

as is well-known.

9.3.5 Multinomial density

Suppose that we have $h > 2$ cells for which the probability of falling in the ith cell is π_{it} and suppose also that in k_t independent trials the number observed in the ith cell is y_{it} for $i = 1, \ldots, h$. For example, monthly opinion polls of voting preference: Labour, Conservative, Liberal Democrat, others.

Let $y_t = (y_{1t}, \ldots, y_{h-1,t})'$ and $\pi_t = (\pi_{1t}, \ldots, \pi_{h-1,t})'$ with $\sum_{j=1}^{h-1}\pi_{jt} < 1$.

Then y_t is multinomial with logdensity

$$\log p(y_t|\pi_t) = \sum_{i=1}^{h-1} y_{it}\left[\log \pi_{it} - \log\left(1 - \sum_{j=1}^{h-1}\pi_{jt}\right)\right] + k_t \log\left(1 - \sum_{j=1}^{h-1}\pi_{jt}\right) + \log C_t, \tag{9.21}$$

for $0 \leq \sum_{i=1}^{h-1} y_{it} \leq k_t$ where

$$C_t = k_t!/\left[\prod_{i=1}^{h-1} y_{it}!\left(k_t - \sum_{j=1}^{h-1} y_{jt}\right)!\right].$$

We therefore take $\theta_t = (\theta_{1t}, \ldots, \theta_{h-1,t})'$ where $\theta_{it} = \log[\pi_{it}/(1-\sum_{j=1}^{h-1}\pi_{jt})]$, and

$$b_t(\theta_t) = k_t \log\left(1 + \sum_{i=1}^{h-1}\exp\theta_{it}\right),$$

so the density of y_t in form (9.6) is

$$p(y_t|\theta_t) = \exp\left[y_t'\theta_t - k_t\log\left(1 + \sum_{i=1}^{h-1}\exp\theta_{it}\right)\right] \times C_t. \tag{9.22}$$

9.3.6 Multivariate extensions

Multivariate generalisations of discrete distributions in the exponential family class are usually not straightforward extensions of their univariate counterparts. We therefore do not consider such generalisations here. However, it is relatively straightforward to have a panel of variables which are independent of each other at time t, conditional on signal θ_t. Denote the ith variable in a panel of discrete time series by y_{it} for $i = 1, \ldots, p$. Models with densities of the form

$$p(y_t|\theta_t) = \prod_{i=1}^{p} p_i(y_{it}|\theta_t),$$

where $p_i(y_{it}|\theta_t)$ refers to an univariate density, possibly in the class of the exponential family. Each density p_i can be different and can be mixed with continuous densities. The variables in the panel only share the time series property implied by θ_t. The vector dimension of θ_t can be different from $p \times 1$. In typical cases of interest, the dimension of θ_t can be less than p. In this case, we effectively obtain a nonlinear non-Gaussian version of the dynamic factor model discussed in Section 3.7.

9.4 Heavy-tailed distributions

9.4.1 *t*-distribution

A common way to introduce error terms into a model with heavier tails than those of the normal distribution is to use Student's t. We therefore consider modelling ε_t of (9.7) by the t-distribution with logdensity

$$\log p(\varepsilon_t) = \log a(\nu) + \frac{1}{2}\log\lambda - \frac{\nu+1}{2}\log\left(1+\lambda\varepsilon_t^2\right), \tag{9.23}$$

where ν is the number of degrees of freedom and

$$a(\nu) = \frac{\Gamma\left(\frac{\nu}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}, \quad \lambda^{-1} = (\nu-2)\sigma_\varepsilon^2, \quad \sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_t), \quad \nu > 2, \quad t = 1,\dots,n.$$

The mean of ε_t is zero and the variance is σ_ε^2 for any ν degrees of freedom which need not be an integer. The quantities ν and σ_ε^2 can be permitted to vary over time, in which case λ also varies over time.

9.4.2 Mixture of normals

A second common way to represent error terms with tails that are heavier than those of the normal distribution is to use a mixture of normals with density

$$p(\varepsilon_t) = \frac{\lambda^*}{\left(2\pi\sigma_\varepsilon^2\right)^{\frac{1}{2}}}\exp\left(\frac{-\varepsilon_t^2}{2\sigma_\varepsilon^2}\right) + \frac{1-\lambda^*}{\left(2\pi\chi\sigma_\varepsilon^2\right)^{\frac{1}{2}}}\exp\left(\frac{-\varepsilon_t^2}{2\chi\sigma_\varepsilon^2}\right), \tag{9.24}$$

where λ^* is near to one, say 0.95 or 0.99, and χ is large, say from 10 to 100. This is a realistic model for situations when outliers are present, since we can think of the first normal density of (9.24) as the basic error density which applies $100\lambda^*$ per cent of the time, and the second normal density of (9.24) as representing the density of the outliers. Of course, λ^* and χ can be made to depend on t if appropriate. The investigator can assign values to λ^* and χ but they can also be estimated when the sample is large enough.

9.4.3 General error distribution

A third heavy-tailed distribution that is sometimes used is the general error distribution with density

$$p(\varepsilon_t) = \frac{w(\ell)}{\sigma_\varepsilon}\exp\left[-c(\ell)\left|\frac{\varepsilon_t}{\sigma_\varepsilon}\right|^\ell\right], \qquad 1 < \ell < 2, \tag{9.25}$$

where

$$w(\ell) = \frac{2[\Gamma(3\ell/4)]^{\frac{1}{2}}}{\ell[\Gamma(\ell/4)]^{\frac{3}{2}}}, \qquad c(\ell) = \left[\frac{\Gamma(3\ell/4)}{\Gamma(\ell/4)}\right]^{\frac{\ell}{2}}.$$

Some details about this distribution are given by Box and Tiao (1973, §3.2.1), from which it follows that $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ for all ℓ.

9.5 Stochastic volatility models

In the standard state space model (3.1) the variance of the observational error ε_t is assumed to be constant over time. In the analysis of financial time series, such as daily fluctuations in stock prices and exchange rates, return series will usually be approximately serially uncorrelated. Return series may not be serially independent, however, because of serial dependence in the variance. It is often found that the observational error variance is subject to substantial variability over time. This phenomenon is referred to as *volatility clustering*. An allowance for this variability in models for such series may be achieved via the *stochastic volatility* (SV) *model*. The SV model has a strong theoretical foundation in the financial theory on option pricing based on the work of the economists Black and Scholes; for a discussion see Taylor (1986). Further, the SV model has a strong connection with the state space approach as will become apparent below.

Denote the first (daily) differences of a particular series of asset log prices by y_t. Financial time series are often constructed by first differencing log prices of some portfolio of stocks, bonds, foreign currencies, etc. A basic SV model for y_t is given by

$$y_t = \mu + \sigma \exp\left(\frac{1}{2}\theta_t\right)\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0,1), \qquad t = 1,\ldots,n, \tag{9.26}$$

where the mean μ and the average standard deviation σ are assumed fixed and unknown. The signal θ_t is regarded as the unobserved log-volatility and it can be modelled in the usual way by $\theta_t = Z_t\alpha_t$ where α_t is generated by (9.4). In standard cases θ_t is modelled as an AR(1) process with Gaussian disturbances, that is $\theta_t = \alpha_t$ where α_t is the state process

$$\alpha_{t+1} = \phi\alpha_t + \eta_t, \qquad \eta_t \sim \mathrm{N}\big(0, \sigma_\eta^2\big), \qquad 0 < \phi < 1, \tag{9.27}$$

for $t = 1,\ldots,n$ and with $\alpha_1 \sim \mathrm{N}[0, \sigma_\eta^2 \,/\, (1-\phi^2)]$. In other cases, the generality of the state equation (9.4) for α_t can be fully exploited. The model can be regarded as the discrete time analogue of the continuous time model used in papers on option pricing, such as Hull and White (1987). Since the model is Gaussian, $\mathrm{E}(y_t|\theta_t) = \mu$ and $\mathrm{Var}(y_t|\theta_t) = \sigma^2\exp(\theta_t)$, it follows that the logdensity of the SV model (9.26) is given by

$$\log p(y_t|\theta_t) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\theta_t - \frac{1}{2\sigma^2}(y_t-\mu)^2\exp(-\theta_t).$$

Further statistical properties of y_t are easy to determine. However, the model is not linear so the techniques described in Part I of this book cannot provide an exact solution for statistical analysis. For a review of work and developments of the SV model see Shephard (1996), Ghysels, Harvey and Renault (1996) and Shephard (2005).

Parameter estimation for the SV models based on maximum likelihood has been considered elsewhere as a difficult problem. Linear Gaussian techniques only offer approximate maximum likelihood estimates of the parameters and can only be applied to the basic SV model (9.26). The techniques we develop in the following chapters of this book, however, provide analyses of SV models based on simulation methods which can be made as accurate as is required.

Various extensions of the SV model can be considered. In the remainder of this section we discuss a number of such extensions together with related models.

9.5.1 Multiple volatility factors

In empirical work it has been observed that volatility often exhibits long-range dependence; see for example Andersen, Bollerslev, Diebold and Labys (2003). Ideally, log-volatility θ_t is modelled by a fractionally integrated process, for example; see Granger and Joyeau (1980). Inference for the SV model (9.26) with a long memory process for θ_t is often based on the spectral likelihood function; see, for example, Breidt, Crato and de Lima (1998) and Ray and Tsay (2000). Exact maximum likelihood methods have recently been considered by Brockwell (2007). In our framework, we can approximate the long-range dependence in the log-volatility θ_t by considering it as a sum of independent autoregressive factors, that is

$$\theta_t = \sum_{i=1}^{q} \theta_{it},$$

where each θ_{it} represents an independent process as in (9.27). The most commonly used specification is the two factor model ($q = 2$), where one factor can be associated with the long-run dependence and the other with the short-run dependence; see the discussion in Durham and Gallant (2002).

9.5.2 Regression and fixed effects

The basic SV model (9.26) captures only the salient features of changing volatility in financial series over time. The model becomes more precise when the mean of y_t is modelled by incorporating explanatory variables. For example, the SV model may be formulated as

$$b(L)y_t = \mu + c(L)'x_t + \sigma \exp\left(\frac{1}{2}\theta_t\right)\varepsilon_t,$$

where L is the lag operator defined so that $L^j z_t = z_{t-j}$ for $z_t = y_t, x_t$ and where $b(L) = 1 - b_1 L - \cdots - b_{p^*} L^{p^*}$ is a scalar lag polynomial of order p^*; the column vector polynomial $c(L) = c_0 + c_1 L_1 + \cdots + c_{k^*} L^{k^*}$ contains $k^* + 1$ vectors of coeffients and x_t is a vector of exogenous explanatory variables. Note that the lagged value y_{t-j}, for $j = 1, \ldots, p^*$, can be considered as an explanatory variable to be added to exogenous explanatory variables. An illustration is provided by

Tsiakas (2006) who introduce dummy effects to account for a seasonal pattern in the volatility. Koopman, Jungbacker and Hol (2005) consider a regression variable that contains information on the unobserved log-volatility process. Such regression effects can be incorporated into the SV model by letting the signal θ_t depend on explanatory variables.

9.5.3 Heavy-tailed disturbances

The Gaussian density $p(\varepsilon_t)$ in the SV model can be replaced by a density with heavier tails such as the t-distribution. This extension is often appropriate because many empirical studies find outlying returns (mostly negative but also positive) due to unexpected jumps or downfalls in asset prices caused by changes in economic conditions or turmoil in financial markets. Key examples are the 'black Monday' crash in October 1987 and the world-wide banking crisis in the second half of 2008. The resulting excess kurtosis found in time series of financial returns can be modelled by having a standardised Student's t distribution for ε_t in (9.26) and its density given by (9.23). The dynamic properties of logvolatility and the thickness of tails are modelled separately as a result. Examples of this approach can be found in Fridman and Harris (1998), Liesenfeld and Jung (2000) and Lee and Koopman (2004).

9.5.4 Additive noise

In the empirical finance literature it is widely recognised that financial prices or returns which are observed at very short time intervals are subject to noise due to discrete observed prices, market regulations and market imperfections. The last source is related to strategic trading behaviour and is commonly caused by differences in the amount of information that traders have about the market. This phenomenon is collectively referred to as *market micro-structure effects* which become more and more apparent as prices are observed at smaller and smaller time intervals; see Campbell, Lo and MacKinlay (1997) and the references therein for a further discussion.

The basic SV model assumes that financial returns only have one source of error. In case the returns are observed at a higher frequency, the SV model should be extended with an additive noise component to account for market microstructure. The additive noise can be represented by a Gaussian disturbance term with constant variance. More specifically, we have

$$y_t = \mu + \sigma \exp\left(\frac{1}{2}\theta_t\right)\varepsilon_t + \zeta_t, \qquad \varepsilon_t \sim \mathrm{N}(0,1), \quad \zeta_t \sim \mathrm{N}(0,\sigma_\zeta^2), \tag{9.28}$$

where all disturbances are serially uncorrelated. The disturbance term ζ_t represents market microstructure effects in the returns. This model has been considered by Jungbacker and Koopman (2005).

9.5.5 Leverage effects

Another characteristic of financial time series is the phenomenon of *leverage*. The volatility of financial markets may adapt differently to positive and negative shocks. It is often observed that while markets might remain more or less stable when large positive earnings have been achieved but when huge losses have to be digested, markets become more unpredictable in the periods ahead. The start of the banking crisis in September 2008 is a clear illustration of the leverage effect. In the seminal paper of Black (1976), the phenomenon of leverage is described. In terms of the SV model (9.26) and (9.27), the leverage effect occurs if a negative return ($\varepsilon_t < 0$) increases the volatility ($\eta_t > 0$) more than a positive return ($\varepsilon_t > 0$) of the same magnitude decreases it ($\eta_t < 0$). The leverage effect is incorporated in the SV model by allowing correlation between the disturbances of the state and the observation equation; see Yu (2005) for a detailed discussion. In the case of our basic SV model (9.26) and (9.27), we achieve this by having

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \mathrm{N}\left(0, \begin{bmatrix} 1 & \sigma_\eta \rho \\ \sigma_\eta \rho & \sigma_\eta^2 \end{bmatrix}\right),$$

for $t = 1, \ldots, n$. The correlation coefficient ρ is typically negative, implying that negative shocks in the return are accompanied by positive shocks in the volatility and vice versa.

The nonlinear state space formulation of the SV model with leverage requires both θ_t and η_t in the state vector α_t to account for the nonlinear relationship. For this purpose, a more convenient specification of the SV model with leverage is proposed by Jungbacker and Koopman (2007) where the model is reformulated as

$$y_t = \sigma \exp(\frac{1}{2} h_t^*) \left\{ \varepsilon_t^* + \mathrm{sign}(\rho) \xi_{2t} \right\}, \qquad \varepsilon_t^* \sim \mathrm{N}(0, 1 - |\rho|),$$

where

$$h_{t+1}^* = \phi h_t^* + \sigma_\xi \left(\xi_{1,t} + \xi_{2t} \right), \qquad \xi_{1t} \sim \mathrm{N}(0, 1 - |\rho|), \qquad \xi_{2t} \sim \mathrm{N}(0, |\rho|),$$

for $t = 1, \ldots, n$, with $h_1^* \sim N\{0, \sigma_\xi^2 (1 - \phi^2)^{-1}\}$. The disturbances ε_t^*, ξ_{1t} and ξ_{2t} are mutually and serially independent for $t = 1, \ldots, n$. In terms of the general formulation (9.4), we have $\alpha_t = (h_t^* \ , \ \sigma_\xi \xi_{2,t})'$, $\xi_t = \sigma_\xi (\xi_{1,t} \ , \ \xi_{2,t+1})'$ and

$$\theta_t = \alpha_t, \quad \alpha_{t+1} = \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix} \alpha_t + \xi_t, \qquad \begin{array}{lcl} \xi_t & \sim & \mathrm{N}\left\{0, \sigma_\xi^2 \mathrm{diag}(1 - |\rho|, |\rho|)\right\}, \\ \alpha_1 & \sim & \mathrm{N}\left\{0, \sigma_\xi^2 \mathrm{diag}([1 - \phi^2]^{-1}, |\rho|)\right\}, \end{array}$$

for $t = 1, \ldots, n$. The observations $y_1, \ldots, y_n$ have the conditional density of the form (9.3) and is given by

$$\log p(y_n | \theta) = \sum_{t=1}^{n} \log p(y_t | \theta_t),$$

where

$$\log p(y_t|\theta_t) = c - \frac{1}{2}h_t^* - \frac{1}{2}\sigma^{-2}\exp(-h_t^*)(1-|\rho|)^{-1}\{y_t - \sigma\exp(\frac{1}{2}h_t^*)\text{sign}(\rho)\xi_{2,t}\}^2,$$

for $t = 1, \ldots, n$ where c is some constant.

9.5.6 Stochastic volatility in mean

As investors require a larger expected return if the risk is large, it seems reasonable to expect a positive relationship between volatility and returns. Empirical evidence however points to a negative influence of volatility on returns; see, for example, French, Schwert and Stambaugh (1987). This effect can be explained by assuming a positive relationship between expected return and *ex-ante* volatility. Koopman and Hol-Uspensky (2002) proposed capturing this so-called volatility feedback effect by including volatility as a regression effect in the mean function. Such a model is labelled as the *SV in Mean* (SVM) model and its simplest form is given by

$$y_t = \mu + d\exp(\theta_t) + \sigma\exp\left(\frac{1}{2}\theta_t\right)\varepsilon_t,$$

where d is the risk premium coefficient which is fixed and unknown. Other forms of the SVM model may also be considered but this one is particularly convenient.

9.5.7 Multivariate SV models

Consider a $p \times 1$ vector of differenced series of asset log prices $y_t = (y_{1t}, \ldots, y_{pt})'$ with constant mean $\mu = (\mu_1, \ldots, \mu_p)'$ and stochastic time-varying variance matrix V_t. The basic version of the multivariate stochastic volatility model can be given by

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim \text{N}(0, V_t), \qquad t = 1, \ldots, n, \tag{9.29}$$

where time-varying variance matrix V_t is a function of the scalar or vector signal θ_t as given by (9.5), that is $V_t = V_t(\theta_t)$. The model (9.29) implies that $y_t|\theta_t \sim \text{N}(\mu, V_t)$, for $t = 1, \ldots, n$. We discuss three possible specifications of the variance matrix $V_t(\theta_t)$ below. Other multivariate generalisations of the univariate stochastic volatility model can be considered as well. A more extensive discussion of the multivariate SV model is presented in Asai and McAleer (2005). A treatment of the three multivariate models below is given by Jungbacker and Koopman (2006).

The first multivariate SV model is based on a single time-varying factor. We can take the variance matrix V_t as a constant matrix and scale it by a stochastically time-varying scalar θ_t, that is

$$V_t = \exp(\theta_t)\Sigma_\varepsilon, \qquad t = 1, \ldots, n. \tag{9.30}$$

This multivariate generalisation of the univariate SV model implies observations y_t with time-varying variances and covariances but with correlations that are constant over time. The conditional density $p(y_t|\theta_t)$ is given by

$$p(y_t|\theta_t) = -\frac{p}{2}\log 2\pi - \frac{p}{2}\theta_t - \frac{1}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}\exp(-\theta_t)s_t, \qquad t = 1,\dots,n,$$

with scalar $s_t = (y_t - \mu)'\Sigma_\varepsilon^{-1}(y_t - \mu)$. The formulation (9.30) was originally proposed by Quintana and West (1987) where they used Bayesian methods for inference. Shephard (1994a) proposed a similar model and referred to it as the local scale model. A further extension for the linear Gaussian state space model where all variances are scaled by the common stochastic scalar $\exp(\theta_t)$ is considered by Koopman and Bos (2004).

The second model has stochastically time-varying variances but constant correlations. We consider a $p \times 1$ vector of log-volatilities $\theta_t = (\theta_{1t},\dots,\theta_{pt})'$. The multivariate extension of the basic SV model (9.26) that is considered by Harvey, Ruiz and Shephard (1994) is given by

$$y_t = \mu + D_t\varepsilon_t, \qquad D_t = \exp\{\frac{1}{2}\text{diag}(\theta_{1t},\dots,\theta_{pt})\}, \qquad \varepsilon_t \sim \text{N}(0,\Sigma_\varepsilon), \quad (9.31)$$

for $t = 1,\dots,n$. The conditional variance matrix of y_t is given by $V_t = D_t\Sigma_\varepsilon D_t$ where the constant matrix Σ_ε is effectively a correlation matrix with unity values on its leading diagonal. The variance matrix V_t is a stochastic function of time but the correlations are constant over time. The conditional density is given by

$$p(y_t|\theta_t) = -\frac{p}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{p}\theta_{it} - \frac{1}{2}\log|\Sigma_\varepsilon| - \frac{1}{2}s_t'\Sigma_\varepsilon^{-1}s_t, \qquad t = 1,\dots,n.$$

where $s_t = D_t^{-1}(y_t - \mu)$ is a $p \times 1$ vector with its ith element equal to $s_{it} = \exp(-0.5\theta_{it})(y_{it} - \mu_i)$ for $i = 1,\dots,p$.

The third model has time-varying variances and correlations. It is based on model (9.29) with the variance matrix decomposed as $V_t = CD_t^2C'$ where matrix C is a lower unity triangular matrix and D_t is specified as in (9.31). The variance matrix is effectively subject to a Cholesky decomposition with a time-varying D_t^2. In this specification both the variances and the correlations implied by V_t are time-varying. The resulting model belongs to a class of multivariate SV models that were originally proposed by Shephard (1996) and further extended and analysed by Aguilar and West (2000) and Chib, Nardari and Shephard (2006). The general class of this model allows for a number of $r < p$ 'volatility factors' where θ_t is an $r \times 1$ vector and the $p \times r$ matrix C contains loading factors and includes an additive disturbance vector with constant variances in (9.29).

9.5.8 Generalised autoregressive conditional heteroscedasticity

The generalised autoregressive conditional heteroscedasticity (GARCH) model, a special case of which was introduced by Engle (1982) and is known as the ARCH model, is a widely discussed model in the financial and econometrics literature. A simplified version of the GARCH(1, 1) model is given by

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0,1), \\ \sigma_{t+1}^2 &= \alpha^* y_t^2 + \beta^* \sigma_t^2, \end{aligned} \tag{9.32}$$

where the parameters to be estimated are α^* and β^*. For a review of the GARCH model and its extensions see Bollerslev, Engle and Nelson (1994).

It is shown by Barndorff-Nielsen and Shephard (2001) that recursion (9.32) is equivalent to the steady state Kalman filter for a particular representation of the SV model. Consider the model

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, & \varepsilon_t &\sim \mathrm{N}(0,1), \\ \sigma_{t+1}^2 &= \phi \sigma_t^2 + \eta_t, & \eta_t &> 0, \end{aligned} \tag{9.33}$$

for $t = 1, \ldots, n$, where disturbances ε_t and η_t are serially and mutually independently distributed. Possible distributions for η_t are the gamma, inverse gamma or inverse Gaussian distributions. We can write the model in its squared form as follows

$$y_t^2 = \sigma_t^2 + u_t, \qquad u_t = \sigma_t^2\big(\varepsilon_t^2 - 1\big),$$

which is in a linear state space form with $\mathrm{E}(u_t) = 0$. The Kalman filter provides the minimum mean squared error estimate a_t of σ_t^2. When in steady state, the Kalman update equation for a_{t+1} can be represented as the GARCH$(1,1)$ recursion

$$a_{t+1} = \alpha^* y_t^2 + \beta^* a_t,$$

with

$$\alpha^* = \phi \frac{\bar{P}}{\bar{P}+1}, \qquad \beta^* = \phi \frac{1}{\bar{P}+1},$$

where $\bar{P}$ is the steady state value for P_t of the Kalman filter which we have defined for the local level model in Section 2.11 and for the general linear model in Subsection 4.3.4. We note that $\alpha^* + \beta^* = \phi$.

9.6 Other financial models

9.6.1 Durations: exponential distribution

Consider a series of transactions in a stock market in which the tth transaction x_t is time-stamped by the time τ_t at which it took place. When studying the behaviour of traders in the market, attention may be focused on the duration between successive transactions, that is $y_t = \Delta\tau_t = \tau_t - \tau_{t-1}$. The duration y_t with mean μ_t can be modelled by a simple exponential density given by

$$p(y_t|\mu_t) = \frac{1}{\mu_t}\exp(-y_t/\mu_t), \qquad y_t, \mu_t > 0. \tag{9.34}$$

This density is a special case of the exponential family of densities and to put it in the form (9.6) we define

$$\theta_t = -\frac{1}{\mu_t} \quad \text{and} \quad b_t(\theta_t) = \log\mu_t = -\log(-\theta_t),$$

so we obtain

$$\log p(y_t|\theta_t) = y_t\theta_t + \log(-\theta_t). \tag{9.35}$$

Since $\dot{b}_t = -\theta_t^{-1} = \mu_t$ we confirm that μ_t is the mean of y_t, as is obvious from (9.34). The mean is restricted to be positive and so we model $\theta_t^* = \log(\mu_t)$ rather than μ_t directly. The durations in financial markets are typically short at the opening and closing of the daily market hours due to heavy trading in these periods. The time stamp τ_t is therefore often used as an explanatory variable in the mean function of durations and in order to smooth out the huge variations of this effect, a cubic spline is used. A simple durations model which allows for the daily seasonal is then given by

$$\begin{aligned} \theta_t &= \gamma(\tau_{t-1}) + \psi_t, \\ \psi_t &= \rho\psi_{t-1} + \chi_t, \qquad \chi_t \sim \mathrm{N}\big(0, \sigma_\chi^2\big), \end{aligned}$$

where $\gamma(\cdot)$ is the cubic spline function and χ_t is serially uncorrelated. Such models can be regarded as state space counterparts of the influential *autoregressive conditional duration* (ACD) model of Engle and Russell (1998).

9.6.2 Trade frequencies: Poisson distribution

Another way of analysing market activity is to divide the daily market trading period into intervals of one or five minutes and record the number of transactions in each interval. The counts in each interval can be modelled by a Poisson density for which the details are given in Subsection 9.3.1. Such a model would be a basic discrete version of what Rydberg and Shephard (2003) have labelled as BIN models.

9.6.3 Credit risk models

A firm can obtain its credit rating from commercial agencies such as Moody's and Standard & Poors. A migration from one rating class to another is indicative of the performance of the firm but also of the economic conditions under which the firm and its trading and financial partners operate. Credit risk indicators aim to provide an insight into the overall direction of the rating migrations for an industry or for the economy as a whole. Such indicators are of key interest to financial regulators and economic policy makers. The construction of a credit risk indicator from a database of credit ratings of firms is a challenging task since we need to consider a large database with the rating history for each firm. In a credit risk model, the rating itself can be regarded as a stochastic variable but also the duration at which the firm keeps it rating before it enters into a different rating category. Since the ratings are measured in classes such as AAA, AA, A, BBB, and so on, the rating variable is inherently a non-Gaussian variable. Koopman, Lucas and Monteiro (2008) accommodate the stylised properties of credit rating migrations by an intensity-based duration model for different types of migrations and driven by a common signal θ_t with the possibility of including explanatory variables as well. The common signal represents the overall credit risk indicator. This modelling framework can be cast in the general class of models discussed in Section 9.2.

A simplification of the analysis can be obtained by compressing the available data into numbers of upgrades, downgrades and defaults in each week or month for different categories. Such counts are usually small; especially when we focus on specific groups of firms (manufacturing, banking, transport, etc.). Hence we treat these counts as coming from a binomial distribution. This is the approach taken by Koopman and Lucas (2008) who consider a panel of N time series of counts y_{ijt} where index i refers to a transition type (e.g. number of downgrades of firms in higher credit rating classes, downgrades of firms that have lower ratings, upgrades, defaults), index j is for a group of firms and index t is for the time period. The counts can be regarded a the number of 'successes' in k_{ijt} independent trials (number of firms in the appropriate group at time t) with a given probability of success, say π_{ijt}. A parsimonious model specification for probability π_{ijt} is given by

$$\pi_{ijt} = \frac{\exp \theta^*_{ijt}}{1 + \exp \theta^*_{ijt}}, \qquad \theta^*_{ijt} = \mu_{ij} + \lambda'_{ij}\theta_t,$$

where θ_t is a signal vector that we can specify as (9.5) and where the scalar coefficients μ_{ij} and the vector coefficients λ_{ij} are treated as unknown fixed parameters that need to be estimated. The constants μ_{ij} and the factor loading vectors λ_{ij} can be pooled into a smaller set of unknown coefficients. The binomial density (9.16) discussed in Subsection 9.3.3 may be appropriate for y_{ijt}. When we further assume that the observations conditional on the dynamic factors are independent of each other, we can formulate the conditional density at time t as the product of the individual densities for all i and j. Hence we have shown that this modelling framework for a credit risk analysis fits naturally in a multivariate extension of the exponential family models that we discussed in Section 9.3. Further extensions with economic and financial variables and with constructed business cycle indicators as explanatory variables are considered by Koopman, Lucas and Schwaab (2011).

9.7 Nonlinear models

In this section we introduce a class of nonlinear models which is obtained from the standard linear Gaussian model (3.1) in a natural way by permitting y_t to depend nonlinearly on α_t in the observation equation and α_{t+1} to depend nonlinearly on α_t in the state equation. Thus we obtain the model

$$y_t = Z_t(\alpha_t) + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, H_t), \tag{9.36}$$

$$\alpha_{t+1} = T_t(\alpha_t) + R_t\eta_t, \qquad \eta_t \sim \mathrm{N}(0, Q_t), \tag{9.37}$$

for $t = 1, \ldots, n$, with $\alpha_1 \sim \mathrm{N}(a_1, P_1)$ and where $Z_t(\cdot)$ and $T_t(\cdot)$ are differentiable vector functions of α_t with dimensions p and m respectively. In principle it would be possible to extend this model by permitting ε_t and η_t to be non-Gaussian but we shall not pursue this extension in this book. Models with general forms similar to this were considered by Anderson and Moore (1979).

A simple example of the relation (9.36) is a nonlinear version of the structural time series model in which the trend μ_t and seasonal γ_t combine multiplicatively and the observation error ε_t is additive, giving

$$y_t = \mu_t \gamma_t + \varepsilon_t;$$

a model of this kind has been considered by Shephard (1994b). A related and more general model is proposed by Koopman and Lee (2009) and is based on the specification

$$y_t = \mu_t + \exp(c_0 + c_\mu \mu_t)\gamma_t + \varepsilon_t,$$

where c_0 and c_μ are unknown coefficients. In these nonlinear models, the magnitude of the seasonal fluctuations in the realisations of y_t depend on the trend in the series. This feature of a time series is often encountered and the typical action taken is to transform y_t by taking logs. The models here provide an alternative to this data transformation and become more relevant when observations cannot be transformed by taking logs due to having negative values in the data.