2

The local level model

A basic example of the state space model is the local level model. In this model the *level* component is allowed to vary in time. The level component can be conceived of as the equivalent of the intercept a in the classical regression model (1.1). As the intercept determines the level of the regression line, the level component plays the same role in state space modelling. The important difference is that the intercept in a regression model is fixed whereas the level component in a state space model is allowed to change from time point to time point. In case the level component does not change over time and is fixed for all time points, the level component is equivalent to the intercept. In other words, it is then a global level and applicable for all time points. In case the level component changes over time, the level component applies locally and therefore the corresponding model is referred to as the local level model.

The local level model can be formulated as

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad && \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + \xi_t, \qquad && \xi_t \sim \text{NID}(0, \sigma_\xi^2) \end{aligned} \tag{2.1}$$

for $t = 1, \ldots, n$, where μ_t is the unobserved level at time t, ε_t is the observation disturbance at time t, and ξ_t is what is called the level disturbance at time t. In the literature on state space models, the observation disturbances ε_t are also referred to as the *irregular component*. The observation and level disturbances are all assumed to be serially and mutually independent and normally distributed with zero mean and variances σ_ε^2 and σ_ξ^2, respectively. The first equation in (2.1) is called the *observation* or *measurement* equation, while the second equation is called the *state* equation. Since the level equation in (2.1) defines a *random walk* (see Chapter 10), the local level model is also referred to as the *random walk plus noise* model (where the noise refers to the irregular component).

The second equation in (2.1) is crucial in time series analysis. In the state equation, time dependencies in the observed time series are dealt with by letting the state at time $t+1$ be a function of the state at time t. Therefore, it takes into account that the observed value of the series at time point $t+1$ is usually more similar to the observed value of the time series at time point t than to any other previous value in the series.

When the state disturbances are all fixed on $\xi_t = 0$ for $t = 1, \ldots, n$, model (2.1) reduces to a *deterministic* model: in this case the level does not vary over time. On the other hand, when the level is allowed to vary over time, it is treated as a *stochastic* process. In Section 2.1 we discuss the results of the analysis of the log of the number of UK drivers KSI with a deterministic level. Then in Section 2.2, the latter results are compared with those obtained with a stochastic level component. As the local level model is not appropriate for the UK drivers KSI series, the model is also applied to the annual numbers of road traffic fatalities in Norway in Section 2.3.

2.1. Deterministic level

If the level disturbances in (2.1) are all fixed on $\xi_t = 0$ for $t = 1, \ldots, n$, it is easily verified that:

$$
\begin{aligned}
&\text{for } t = 1: && y_1 = \mu_1 + \varepsilon_1, \\
& && \mu_2 = \mu_1 + \xi_1 = \mu_1 + 0 = \mu_1 \\
&\text{for } t = 2: && y_2 = \mu_2 + \varepsilon_2 = \mu_1 + \varepsilon_2, \\
& && \mu_3 = \mu_2 + \xi_2 = \mu_2 + 0 = \mu_1 \\
&\text{for } t = 3: && y_3 = \mu_3 + \varepsilon_3 = \mu_1 + \varepsilon_3, \\
& && \mu_4 = \mu_3 + \xi_3 = \mu_3 + 0 = \mu_1
\end{aligned}
$$

and so on.

Summarising, in this case the local level model (2.1) simplifies to

$$y_t = \mu_1 + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2) \tag{2.2}$$

for $t = 1, \ldots, n$. Therefore, in this special situation everything relies on the value of μ_1, the value of the level at time $t = 1$. Once this value is established, it remains constant for all other time points $t = 2, \ldots, n$.

Generally, in state space models the value of the unobserved state at the beginning of the time series (i.e. at $t = 1$) is unknown. There are two ways to deal with this problem. Either the researcher provides the first value, based on theoretical considerations, or some previous research, for example. Or this first value is *estimated* by a procedure that falls within the class of state space methods. Since nothing is usually known about the initial value of the state, the second approach is usually followed in practice, and will be used in all further analyses discussed in the present book. In state space modelling, the second approach is called *diffuse initialisation*.

In classical regression analysis the unknown parameters are the intercept and the regression coefficients, for which estimates can be obtained analytically. In state space methods the unknown parameters include the observation and state disturbance variances. These latter parameters are also known as *hyperparameters*. Unlike classical regression analysis, when a state space model contains two or more hyperparameters (i.e. disturbance variances) the (maximum likelihood) estimation of these hyperparameters requires an iterative procedure. The iterations aim to maximise the likelihood value with respect to the hyperparameters (see also Chapter 11). Numerical optimisation methods are employed for this task and they are based on an iterative search process to find the maximum in a numerically efficient way.

Since the variance of the level disturbances σ^2_ξ is fixed at zero, only two parameters need to be estimated in model (2.2). These two parameters are μ_1 and σ^2_ε. Using the diffuse initialisation method, the analysis of the log of the number of UK drivers KSI with the deterministic level model yields the following results:

```
it0    f=     0.3297597 df=9.731e-007 e1=2.690e-006 e2=3.521e-008
Strong convergence
```

This output reflects the numerical search procedure where `it0` refers to the initialisation step, `f` is the logged likelihood value for the hyperparameter value considered at iteration 0 whereas `df` is the first derivative of the likelihood function with respect to the hyperparameter and evaluated at the value of the hyperparameter at iteration 0. The values `e1` and `e2` indicate other measures of convergence of the maximisation procedure. In the numerical maximisation of the likelihood function, no iterations are required for the estimation of the parameters of the deterministic level model. This is in agreement with the fact that the parameter estimates of classical linear regression models can be determined analytically. The value of the log-likelihood function that is maximised in state space methods is 0.3297597. The maximum likelihood estimate of the variance of the observation disturbances is $\widehat{\sigma}_\varepsilon^2 = 0.029353$, and the maximum likelihood estimate of the level for $t = 1$ is $\widehat{\mu}_1 = 7.4061$. The resulting equation for model (2.2) is

$$y_t = 7.4061 + \varepsilon_t.$$

Now, the sum of the log of the monthly number of UK drivers KSI in the period 1969–1984 happens to be 1421.97. Since

$$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t = \frac{1}{192}1421.97 = 7.4061$$

for this time series, and

$$s_y^2 = \frac{1}{n-1}\sum_{t=1}^{n} (y_t - \bar{y})^2 = 0.029353,$$

this extremely simple state space model actually computes the mean and variance of the observed time series.

Thus, the best fitting decomposition based on model (2.2) is

$$y_t = \bar{y} + (y_t - \bar{y}). \tag{2.3}$$

This is not surprising, since it is well known that the best estimate for the parameter μ minimising the least-squares function

$$f(\mu) = \sum_{t=1}^{n} (y_t - \mu)^2$$

equals

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} y_t,$$

the mean of variable y.

The level for model (2.2) is displayed in Figure 2.1, together with the observed time series. As the figure illustrates, the deterministic level is indeed a constant and does not vary over time as a result. Figure 2.2 contains a plot of the observation disturbances ε_t corresponding to the deterministic level model. Just as in the classical regression analysis discussed in Chapter 1, the disturbances ε_t of the deterministic level model are not randomly distributed in this case, and follow a very systematic pattern. In fact, the irregular component in Figure 2.2 simply consists

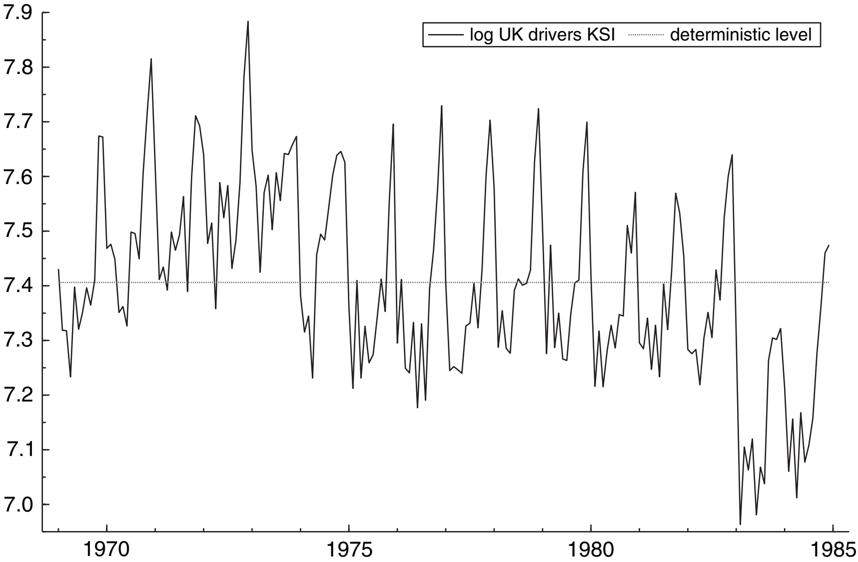

Figure 2.1. Deterministic level.

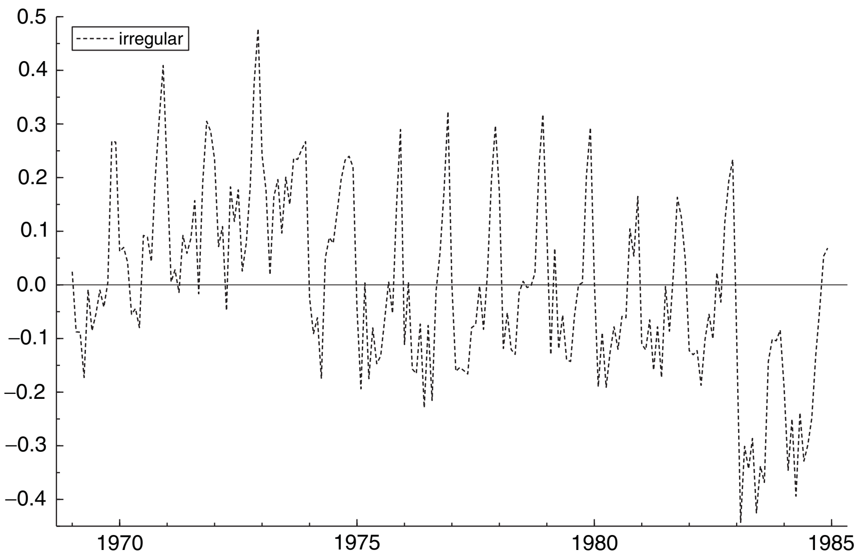

Figure 2.2. Irregular component for deterministic level model.

Table 2.1. Diagnostic tests for deterministic level model and log UK drivers KSI.

	statistic	value	critical value	assumption satisfied
independence	$Q(15)$	415.210	25.00	–
	$r(1)$	0.699	±0.14	–
	$r(12)$	0.677	±0.14	–
homoscedasticity	$H(64)$	2.058	1.67	–
normality	N	0.733	5.99	+

of the deviations of the observed time series from its mean, as already implied by (2.3).

Diagnostic tests for the assumptions of independence, homoscedasticity, and normality of the residuals of the analysis are presented in Table 2.1. A discussion of the exact definition, computation and interpretation of these diagnostic tests is postponed until Section 8.5. Even without this knowledge, however, it is easily seen that the values of the autocorrelations at lags 1 and 12 (see also Chapter 1), which are $r(1) = 0.699$ and $r(12) = 0.677$, respectively, both far exceed the 95% confidence limits of $\pm 2/\sqrt{n} = \pm 0.144$ for this time series with $n = 192$ observations.

The high amount of dependency between the residuals is confirmed by the very large value of the Q-test in Table 2.1. The Q-statistic is a general omnibus test that can be used to check whether the combined first k (in this case 15) autocorrelations in the correlogram deviate from zero. Since $Q(15) = 415.210$ and because this value is much larger than $\chi^2_{(15;0.05)} = 25.00$ (see Table 2.1), evaluated as a whole the first 15 autocorrelations significantly deviate from zero, meaning that the null hypothesis of independence must be rejected.

The H-statistic in Table 2.1 tests whether the variances of two consecutive and equal parts of the residuals are equal to one another. In the present case, the test shows that the variance of the first 64 elements of the residuals is unequal to the variance of the last 64 elements of the residuals, because $H(64) = 2.058$ is larger than the critical value of $F_{(64,64;0.025)} \approx 1.67$. This means that the assumption of homoscedasticity of the residuals is also not satisfied in the present analysis.

Finally, the N-statistic in Table 2.1 tests whether the skewness and kurtosis of the distribution of the residuals comply with a normal or Gaussian distribution. Since $N = 0.733$ is smaller than the critical value of $\chi^2_{(2;0.05)} = 5.99$ (see Table 2.1), the null hypothesis of normally distributed residuals is not rejected.

Summarising, the residuals of the deterministic level model neither satisfy the assumption of independence nor that of homoscedasticity; only the assumption of normality is not violated.

In order to compare the different state space models illustrated in the present book, throughout the Akaike Information Criterion (AIC) will be used:

$$\text{AIC} = \frac{1}{n}\left[-2n \log L_d + 2(q + w)\right],$$

where n is the number of observations in the time series, $\log L_d$ is the value of the diffuse log-likelihood function which is maximised in state space modelling, q is the number of diffuse initial values in the state, and w is the total number of disturbance variances estimated in the analysis. When comparing different models with the AIC the following rule holds: smaller values denote better fitting models than larger ones. A very useful property of this criterion is that it compensates for the number of estimated parameters in a model, thus allowing for a fair comparison between models involving different numbers of parameters. In the deterministic level model only one variance is estimated (σ_ε^2), and one initial value (μ_1). Therefore, the Akaike information criterion for the analysis of the log of the number of drivers KSI with the deterministic level model equals

$$\text{AIC} = \frac{1}{192}\left[-2(192)(0.3297597) + 2(1 + 1)\right] = -0.638686.$$

In the following sections, this value will be used for purposes of comparison with other state space models.

2.2. Stochastic level

When the level μ_t in model (2.1) is allowed to vary over time, on the other hand, the following results are obtained when estimating the hyperparameters by the method of maximum likelihood.

```
it0    f=     0.5673434 df=   0.08018 e1=     0.2550 e2=  0.003223
it1    f=     0.5799665 df=    0.1032 e1=     0.3199 e2=    0.3542
it2    f=     0.6404443 df=   0.08408 e1=     0.2048 e2=   0.02733
it3    f=     0.6424964 df=   0.03334 e1=     0.1025 e2=  0.003279
it4    f=     0.6429869 df=   0.02961 e1=    0.09162 e2= 0.0006207
it5    f=     0.6449777 df=  0.006552 e1=    0.02114 e2=  0.004098
it6    f=     0.6451632 df=  0.002400 e1=   0.007856 e2=  0.001422
it7    f=     0.6451949 df= 0.0004676 e1=   0.001543 e2= 0.0007765
it8    f=     0.6451960 df=3.338e-005 e1= 0.0001103 e2= 0.0001597
it9    f=     0.6451960 df=3.557e-006 e1=8.776e-006 e2=1.508e-005
Strong convergence
```

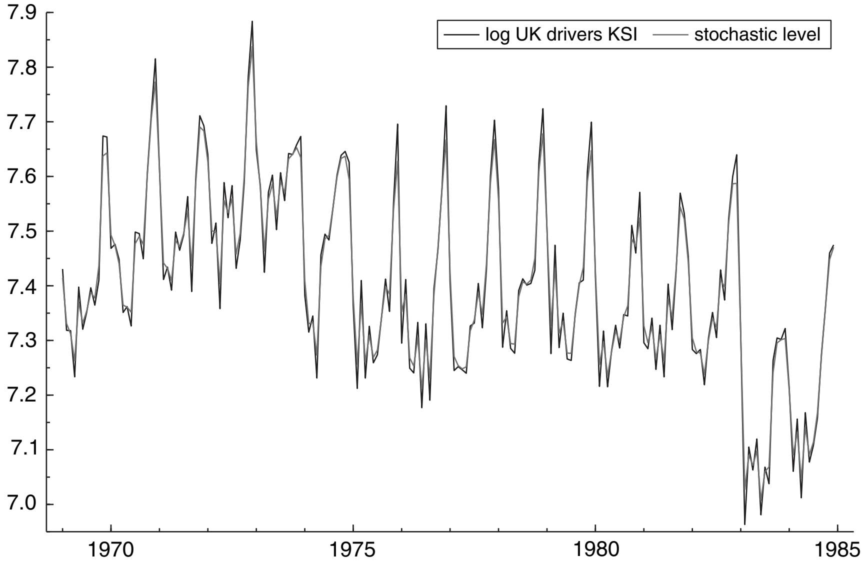

Figure 2.3. Stochastic level.

The algorithm converges in nine iterations. At convergence the value of the log-likelihood function is 0.6451960. The maximum likelihood estimate of the variance of the irregular component is $\widehat{\sigma}_{\varepsilon}^2 = 0.00222157$ and of the level disturbance variance is $\widehat{\sigma}_{\xi}^2 = 0.011866$. The maximum likelihood estimate of the initial value of the level at time point $t = 1$ is $\widehat{\mu}_1 = 7.4150$.

The stochastic level is illustrated in Figure 2.3, together with the observed time series. It shows that the observed time series is recovered quite well when the level is allowed to vary over time. It is nevertheless questionable whether the local level is appropriate for describing all the dynamics in the time series y_t.

Figure 2.4 contains a plot of the irregular component for this analysis. In this figure, the systematic pattern that was found in the residuals of the previous analysis is absent, and the observation disturbances seem to be much closer to independent random values, or – as is also said in control engineering where state space methods originated – to *white noise*.

To some extent, this is confirmed by the diagnostic tests of the residuals given in Table 2.2. The autocorrelation at lag 1 no longer deviates from zero, and the value of the overall Q-test for independence is much smaller than in the previous analysis. The test for heteroscedasticity is also no longer significant. However, both the values of $r(12)$ (the autocorrelation

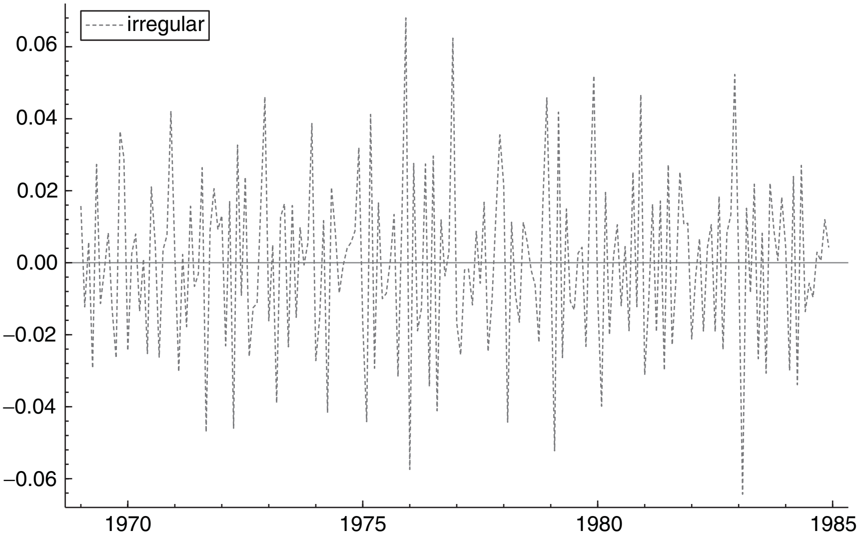

Figure 2.4. Irregular component for local level model.

at lag 12) and of the general Q-test still indicate significant serial correlation in the residuals. Moreover, according to Table 2.2 the residuals of the local level model do not satisfy the assumption of normality.

In the stochastic level model two variances are estimated (σ_ε^2 and σ_ξ^2), and one diffuse element (μ_1). Therefore, the Akaike information criterion for this analysis equals

$$\text{AIC} = \frac{1}{192}\,[-2(192)(0.6451960) + 2(1+2)] = -1.25914.$$

This value is much smaller than for the deterministic level model, meaning that the stochastic level model fits the data better.

In conclusion, the stochastic level model appears to be an improvement upon the deterministic level model. A lot of the dependencies between the observation disturbances in Figure 2.2 have disappeared in Figure 2.4.

Table 2.2. Diagnostic tests for local level model and log UK drivers KSI.

	statistic	value	critical value	assumption satisfied
independence	$Q(15)$	105.390	23.68	–
	$r(1)$	0.009	± 0.14	+
	$r(12)$	0.537	± 0.14	–
homoscedasticity	$H(64)$	1.064	1.67	+
normality	N	13.242	5.99	–

Moreover, the Akaike information criterion indicates that the stochastic level model yields a better representation of the time series than the deterministic level model. However, the diagnostic tests in Table 2.2 also reveal that the stochastic level model is by no means the appropriate model for describing the time series at hand, as will become clearer in Chapter 4. In the next section, therefore, an analysis is presented where the local level model provides a more adequate description of the data.

2.3. The local level model and Norwegian fatalities

Applying the local level model to the log of the annual number of road traffic fatalities in Norway as observed for the 34 years of 1970 through to 2003 (see Appendix B and Figure 2.5), the following results are obtained.

```
it0    f=     0.7755299 df=    0.1692 e1=    0.5779 e2=  0.006216
it1    f=     0.8205220 df=    0.1248 e1=    0.4053 e2=  0.009750
it2    f=     0.8464841 df=   0.02166 e1=   0.06664 e2=   0.01080
it3    f=     0.8468295 df=  0.005806 e1=   0.01800 e2= 0.0007435
it4    f=     0.8468620 df= 0.0003182 e1= 0.0009326 e2= 0.0003626
it5    f=     0.8468622 df=1.945e-005 e1=5.699e-005 e2=2.894e-005
Strong convergence
```

At convergence the value of the log-likelihood function is 0.8468622. The maximum likelihood estimate of the irregular variance is $\hat{\sigma}_{\varepsilon}^2 = 0.00326838$,

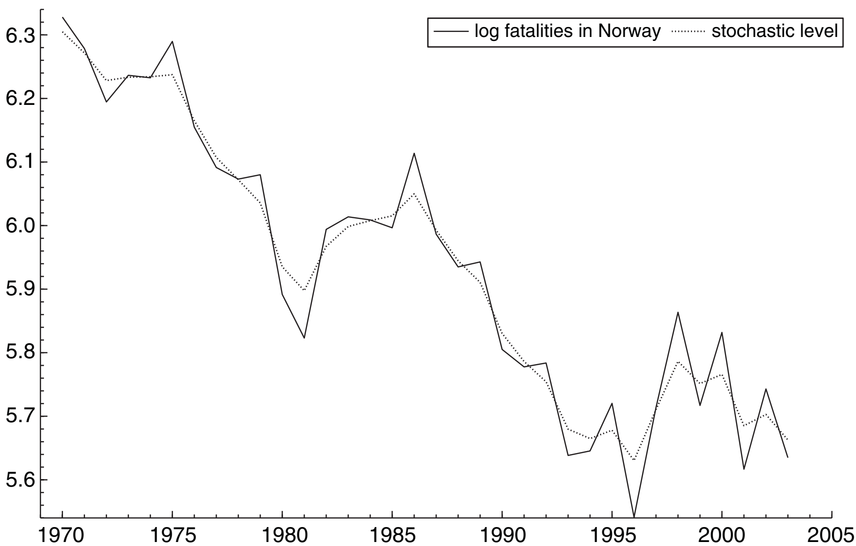

Figure 2.5. Stochastic level for Norwegian fatalities.

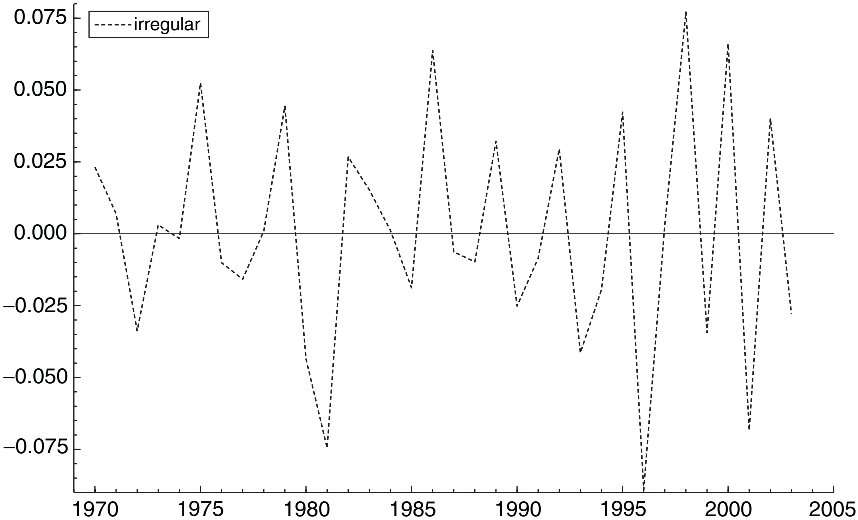

Figure 2.6. Irregular component for Norwegian fatalities.

while the maximum likelihood estimate of the variance of the level disturbances equals $\widehat{\sigma}_{\xi}^2 = 0.0047026$. The maximum likelihood estimate of the initial value of the level at time point $t = 1$ is $\widehat{\mu}_1 = 6.3048$. The stochastic level is illustrated in Figure 2.5, together with the observed time series.

Figure 2.6 contains a plot of the irregular component. The diagnostic tests for independence, homoscedasticity, and normality of the residuals of this analysis are given in Table 2.3. The autocorrelations at lags 1 and 4 are well within the 95% confidence limits of $\pm 2/\sqrt{n} = \pm 0.343$ for this time series. Moreover, since $Q(10) < \chi^2_{(9;0.05)}$, $H(11) < F_{(12,12;0.025)}$, and $N < \chi^2_{(2;0.05)}$ (see also Section 8.5), these tests indicate that the residuals satisfy all of the assumptions of the local level model (2.1).

Table 2.3. Diagnostic tests for local level model and log Norwegian fatalities.

	statistic	value	critical value	assumption satisfied
independence	$Q(10)$	6.228	16.92	+
	$r(1)$	−0.127	±0.34	+
	$r(4)$	−0.105	±0.34	+
homoscedasticity	$H(11)$	1.746	3.28	+
normality	N	1.191	5.99	+

The value of the Akaike information criterion for this analysis equals

$$\text{AIC} = \frac{1}{34}\,[-2(34)(0.8468622) + 2(1+2)] = -1.51725,$$

which is a great improvement upon the deterministic level model applied to these data, since the AIC value for the deterministic model equals 0.040245. Adding a slope component (see Chapter 3) to model (2.1) does not improve the description of this time series, as this results in an AIC value of only −1.28035.