4

The local level model with seasonal

Most readers will probably have understood that an essential aspect of the UK drivers KSI series has been overlooked in the analyses discussed so far. The time series in Figure 1.2 has a yearly recurring pattern. The nature of this pattern becomes even more clear in Figure 4.1 where vertical lines separate each calendar year in the observed time series of Figure 1.2.

Inspecting the monthly development for each year in Figure 4.1, the following regularity emerges: more drivers are killed or seriously injured at the end of a year than in other periods of a year. In time series analysis, this recurring pattern is referred to as a *seasonal* effect. Whenever a time series consists of hourly, daily, monthly, or quarterly observations with respective periodicity of 24 (hours), 7 (days), 12 (months), or 4 (quarters), one should always be on the alert for possible seasonal effects in the series.

In a state space framework, the seasonal effect can be modelled by adding a seasonal component either to the local level model or to the local linear trend model. Since it was found in the previous chapter that the slope component is redundant in describing the time series in Figure 4.1, the investigation of the effect of adding a seasonal component will be restricted to the local level model. In the case of quarterly data, this takes the following form:

$$
\begin{aligned}
y_t &= \mu_t + \gamma_t + \varepsilon_t, & \varepsilon_t &\sim \text{NID}(0, \sigma_\varepsilon^2) \\
\mu_{t+1} &= \mu_t + \xi_t, & \xi_t &\sim \text{NID}(0, \sigma_\xi^2) \\
\gamma_{1,t+1} &= -\gamma_{1,t} - \gamma_{2,t} - \gamma_{3,t} + \omega_t, & \omega_t &\sim \text{NID}(0, \sigma_\omega^2) \\
\gamma_{2,t+1} &= \gamma_{1,t}, \\
\gamma_{3,t+1} &= \gamma_{2,t},
\end{aligned}
\tag{4.1}
$$

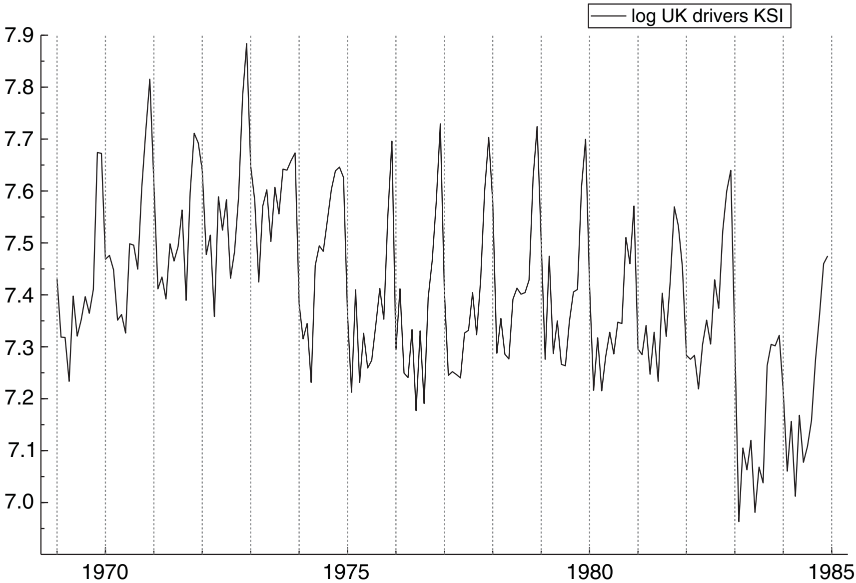

Figure 4.1. Log of number of UK drivers KSI with time lines for years.

for $t = 1, \ldots, n$, where $\gamma_t = \gamma_{1,t}$ denotes the seasonal component. The disturbances ω_t in (4.1) allow the seasonal to change over time. The initial values μ_1, $\gamma_{1,1}$, $\gamma_{2,1}$ and $\gamma_{3,1}$ are treated as fixed and unknown coefficients.

In contrast with the level and slope components, where each component requires one state equation, the seasonal component generally requires $(s - 1)$ state equations where s is given by the periodicity of the seasonal. For quarterly data (where we have $s = 4$), three state equations are needed, as is shown in (4.1). The fourth and fifth equations are identities which can be interpreted as follows. Define $\gamma_{i,t}$ as the ith quarter of time period t. Then the fourth equation tells you that the quarter of the next period $t + 1$ is the next quarter $i + 1$ from the current period t. Since this is a fact of life we cannot add disturbances to such identity equations!

The third equation in (4.1) can also be written as

$$\gamma_{t+1} = -\gamma_t - \gamma_{t-1} - \gamma_{t-2} + \omega_t, \qquad (4.2)$$

for $t = s - 1, \ldots, n$. We notice that the time index for (4.2) starts at $s - 1 = 3$. Since it follows from (4.1) that $\gamma_1 = \gamma_{1,1}$, $\gamma_2 = \gamma_{1,2} = \gamma_{2,1}$ and $\gamma_3 = \gamma_{1,3} = \gamma_{2,2} = \gamma_{3,1}$, we also treat γ_1, γ_2 and γ_3 as fixed and unknown coefficients. Given a set of values for $\{\gamma_1, \gamma_2, \gamma_3\}$, the recursion (4.2) is valid for $t = s - 1, \ldots, n$.

When the seasonal effect γ_t is not allowed to change over time, we require $\omega_t = 0$ for all $t = s-1, \ldots, n$. This is achieved by setting $\sigma^2_\omega = 0$. It follows that

$$\sum_{j=0}^{s-1} \gamma_{t-j} = 0, \tag{4.3}$$

for $t = s, \ldots, n$. When the seasonal is allowed to vary over time, that is $\sigma^2_\omega > 0$, (4.3) is not satisfied due to the random increments of ω_t. However, the expectation of seasonal disturbance ω_t equals zero. As a result, the expectation of the sum $\gamma_t + \gamma_{t-1} + \ldots + \gamma_{t-s+1}$ also equals zero for $t = s, \ldots, n$.

Since the log of the number of UK drivers KSI in Figure 4.1 consists of monthly instead of quarterly data, the periodicity of the seasonal is $s = 12$, implying that the modelling of (4.1) requires a total of 12 state equations (one for the level and 11 for the seasonal). The seasonal specification in (4.1) is called a *dummy seasonal*. It may be noted that other specifications than the dummy seasonal can be used too. For example, the *trigonometric seasonal* can be considered. For details about such alternative specifications of the seasonal we refer to Durbin and Koopman (2001), as these are beyond the scope of the present book.

4.1. Deterministic level and seasonal

Fixing the level and seasonal disturbances ξ_t and ω_t in (4.1) to zero, the analysis of the time series in Figure 4.1 using diffuse initialisation of the values of the 12 state equations at $t = 1$ yields the following results:

```
it0    f=     0.4174873 df=1.613e-006 e1=4.871e-006 e2=5.340e-008
Strong convergence
```

As is the case for all completely deterministic models, the estimation process requires no iterations. At convergence the value of the log-likelihood function is 0.4174873. The maximum likelihood estimate of the variance of the observation disturbances is $\widehat{\sigma}^2_\varepsilon = 0.0175885$. The maximum likelihood estimate of μ_1 is $\widehat{\mu}_1 = 7.4061$. Since the level is deterministic we have $\widehat{\mu}_t = \widehat{\mu}_1 = 7.4061$ for $t = 1, \ldots, n$. Therefore, the estimated deterministic level is again equal to the mean of the observed time series (see also Section 2.1). At this point, we refrain from giving the maximum likelihood estimates of the initial values of the 11 state equations required

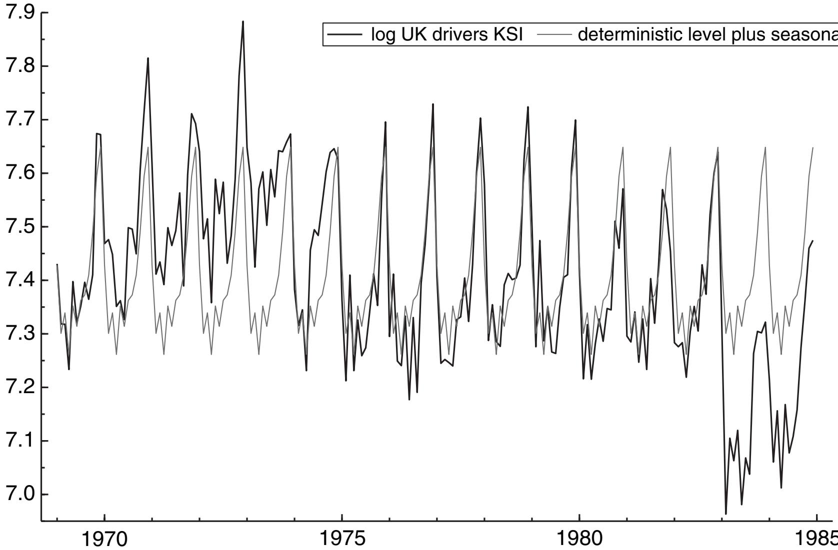

Figure 4.2. Combined deterministic level and seasonal.

for the modelling of the seasonal, because these are not very informative in the present context.

The combined deterministic level and seasonal are displayed in Figure 4.2, while these two components are plotted separately in Figures 4.3 and 4.4, respectively.

By denoting $\bar{y}$ as the overall mean of the log of the numbers of drivers KSI and $\bar{y}_j$ as the mean of the log of the numbers of drivers KSI for month j in the series ($j = 1, \ldots, s$), the deterministic level and seasonal model is given by

$$\widehat{y}_t = \widehat{\mu}_t + \widehat{\gamma}_t = \bar{y} + \left(\bar{y}_j - \bar{y}\right)$$

for $t = 1, \ldots, n$. Note that

$$\sum_{j=0}^{s-1} \widehat{\gamma}_{t-j} = \sum_{j=1}^{s} (\bar{y}_j - \bar{y}) = 0,$$

from which it follows that the seasonal component satisfies (4.3). The deterministic level and seasonal model actually performs a one-way ANOVA with 12 treatment levels (see, e.g., Kirk, 1968). The F-test for the seasonal (with denominator $\widehat{\sigma}^2_{\varepsilon} = 0.0175885$) is $F_{(11,180)} = 12.614$ and this is very significant ($p < 0.01$). The F-test is based on the assumption of random errors. However, as Figure 4.5 clearly indicates, the observation

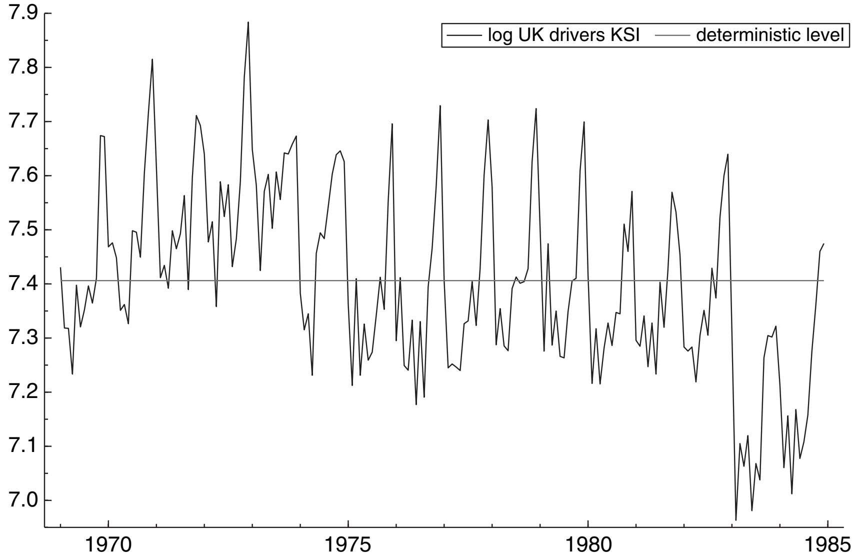

Figure 4.3. Deterministic level.

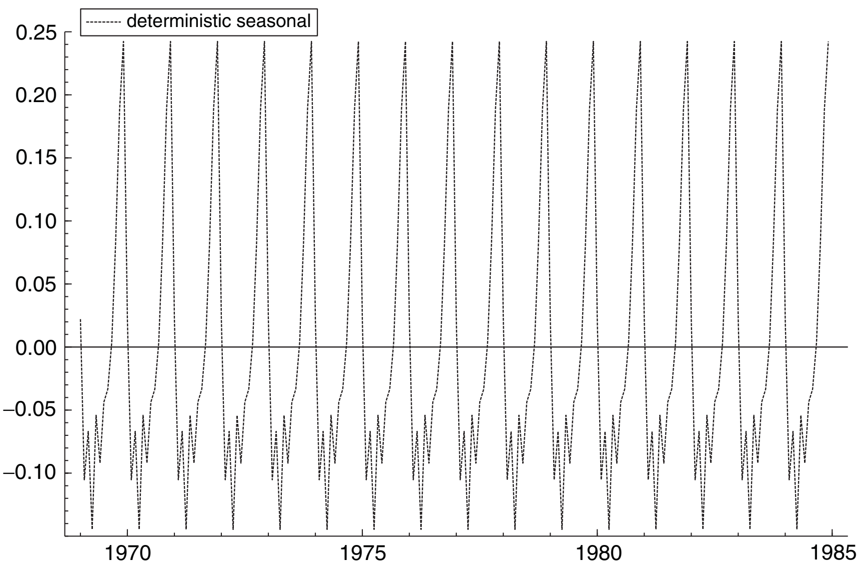

Figure 4.4. Deterministic seasonal.

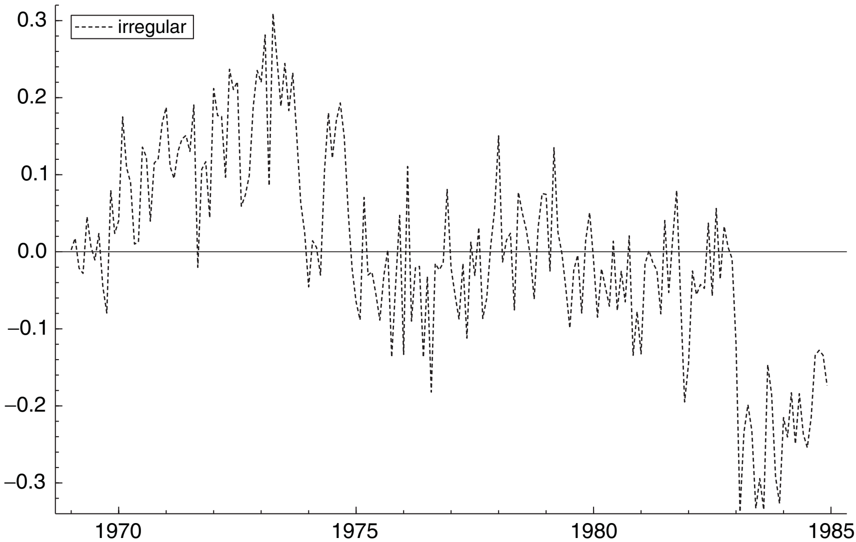

Figure 4.5. Irregular component for deterministic level and seasonal model.

disturbances of the deterministic level and seasonal model are not independently distributed, and the F-test is therefore seriously flawed.

This is confirmed by the results of the diagnostic tests in Table 4.1. They show that the residuals do not satisfy any of the assumptions, except for normality.

Since we are dealing with monthly data, model (4.1) contains 12 state equations for which 12 initial state values need to be estimated. Given the fact that in addition one variance is estimated for the deterministic level and seasonal model, the Akaike information criterion for this model equals

$$\text{AIC} = \frac{1}{192}\,[-2(192)(0.4174873) + 2(12+1)] = -0.699558.$$

Table 4.1. Diagnostic tests for deterministic level and seasonal model and log UK drivers KSI.

	statistic	value	critical value	assumption satisfied
independence	$Q(15)$	751.580	25.00	–
	$r(1)$	0.724	± 0.14	–
	$r(12)$	0.431	± 0.14	–
homoscedasticity	$H(60)$	3.400	1.67	–
normality	N	1.971	5.99	+

Therefore, the AIC of the deterministic level and seasonal model is, somewhat surprisingly, not as good as that of the deterministic linear trend model (−0.796896), although it is slightly better than the deterministic level model (−0.638686).

In the previous chapters it was found that deterministic state space models are identical to some form of classical regression analysis. This suggests that the deterministic level and seasonal model must also have its counterpart in classical regression analysis. The question is: which classical regression model is involved here? Results identical to the deterministic level and seasonal model presented above are obtained by performing the following classical multiple regression analysis.

Eleven variables are constructed according to the following recipe. The first variable is given the value 11 (i.e. $s-1$) whenever an observation in the time series falls in the month of January, and minus one for all the other months of the year. The second variable is set equal to 11 whenever an observation in the time series falls in the month of February and minus one elsewhere. And so on, until the eleventh and last variable. This last variable is given the value 11 for the month of November and minus one elsewhere. A classical multiple regression analysis with an intercept variable together with these 11 'dummy' variables against the log of UK drivers KSI, yields an estimate of the intercept identical to the level shown in Figure 4.3, while the sum of the 11 dummy variables weighted by their respective regression coefficients is identical to the seasonal in Figure 4.4. The overal sum of the seasonal effect in one year is obviously equal to zero.

4.2. Stochastic level and seasonal

The level and the seasonal in (4.1) can be allowed to vary over time. In that case, the following results are obtained:

```
it0    f=     0.6967041 df=    0.1701 e1=     0.7878 e2=  0.003672
it5    f=     0.8803781 df=   0.08417 e1=     0.4735 e2=  0.002996
it10   f=     0.9353563 df=   0.01276 e1=    0.04076 e2=  0.001999
it15   f=     0.9369055 df= 0.0002212 e1= 0.0007954 e2= 0.0001283
it18   f=     0.9369063 df=6.131e-006 e1=1.809e-005 e2=8.189e-006
Strong convergence
```

At convergence the value of the log-likelihood function is 0.9369063. The maximum likelihood estimate of the irregular variance is $\hat{\sigma}_\varepsilon^2 = 0.00341592$ and the maximum likelihood estimates of the state variances are given by $\hat{\sigma}_\xi^2 = 0.000935947$ and $\hat{\sigma}_\omega^2 = 0.00000050$, respectively. Plots of the

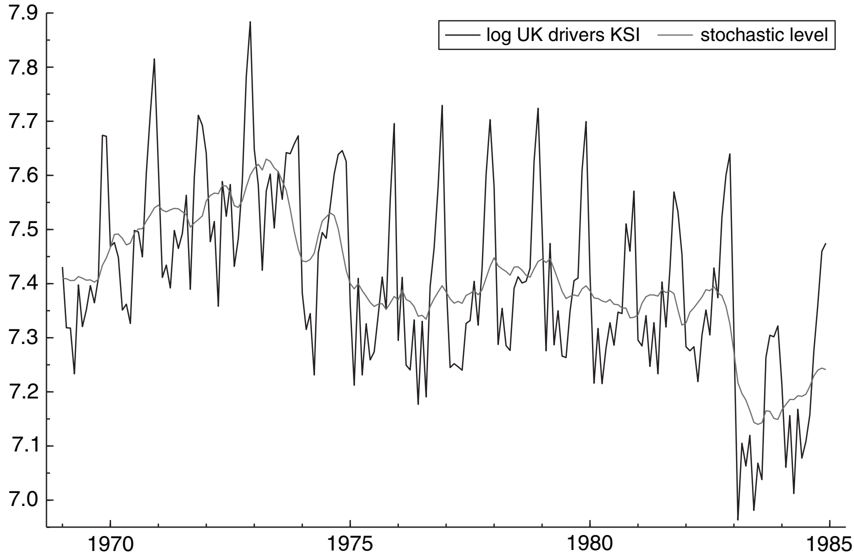

Figure 4.6. Stochastic level.

stochastic level and seasonal obtained from this analysis are displayed in Figures 4.6 and 4.7, respectively. The variance of the seasonal disturbances is very small. This indicates that the seasonal pattern in the observed time series hardly changes over the years, which is confirmed by inspection of Figure 4.7.

For a better understanding of the interpretation of the seasonal component in Figure 4.7, we focus on the first year of the seasonal component (i.e. on 1969), see Figure 4.8. It shows that the largest number of drivers in Great Britain were killed or seriously injured in the months of November and December of 1969, while April 1969 resulted in the smallest number. This pattern is repeated in all the other years of the series.

The irregular component for the stochastic level and seasonal model is displayed in Figure 4.9. The residuals of the stochastic model are much closer to independent random values than those obtained with the deterministic model (see Figure 4.5). Whether 'much closer' is close enough can be determined by the diagnostic tests in Table 4.2.

The first autocorrelation in the correlogram does not deviate from zero but also the autocorrelation at lag 12 is close to zero. This is the first of our analyses where we yield such a satisfactory result for this KSI series. In all previous analyses of the series, the autocorrelation at lag 12 was found to be unacceptably large, see Tables 2.1, 2.2, 3.1, and 3.2.

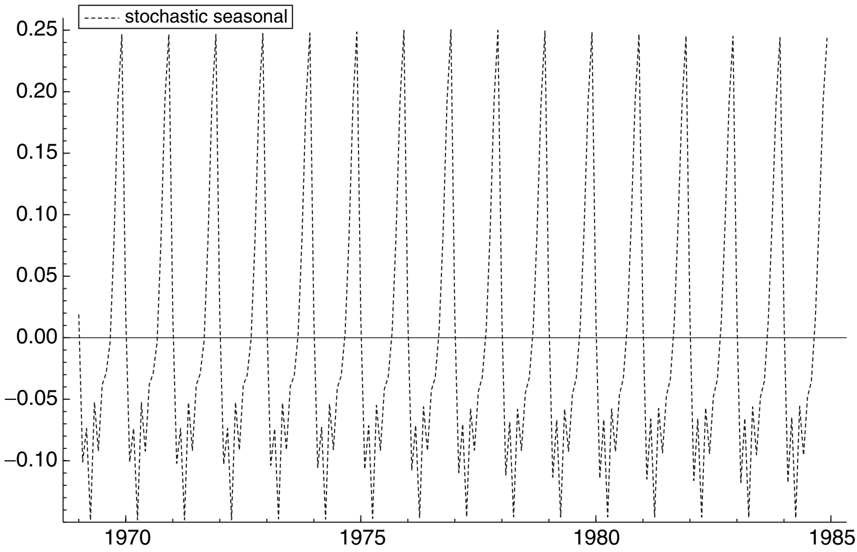

Figure 4.7. Stochastic seasonal.

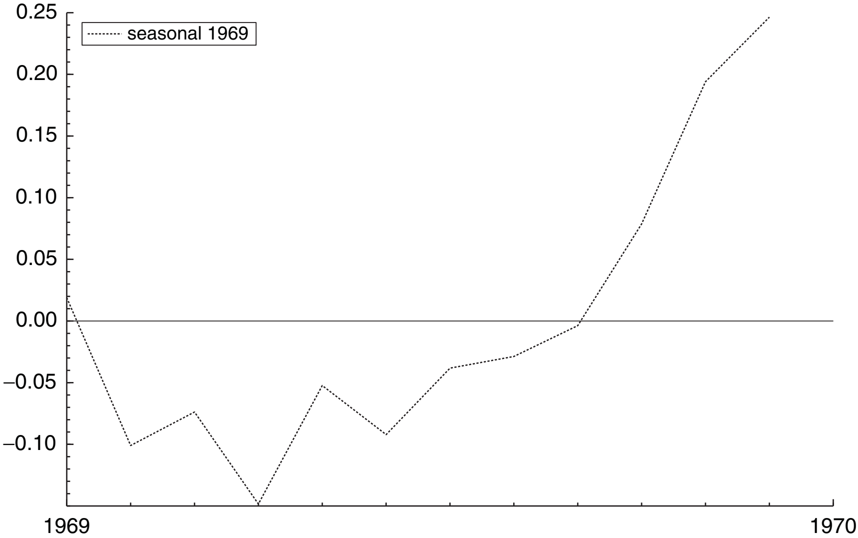

Figure 4.8. Stochastic seasonal for the year 1969.

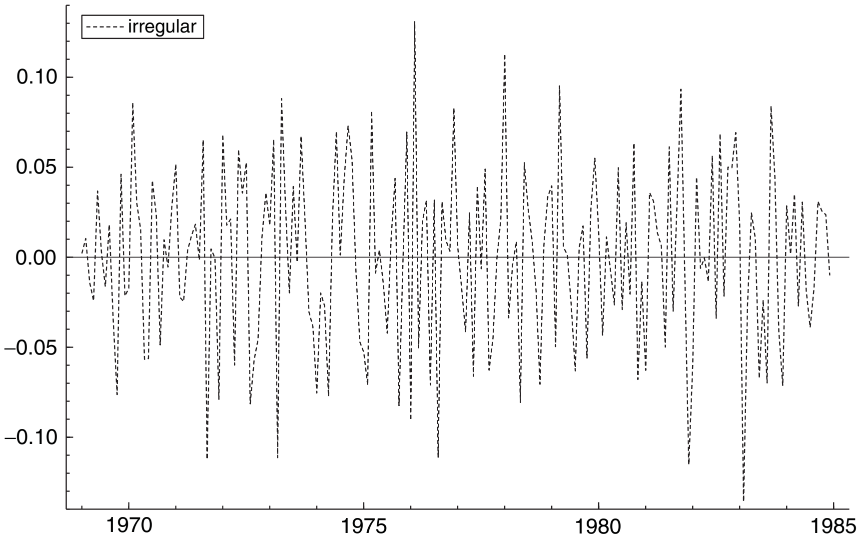

Figure 4.9. Irregular component for stochastic level and seasonal model.

The same applies to the general Q-test for independence based on the first 15 autocorrelations, which is for the first time smaller than the critical value of $\chi^2_{(13;0.05)} = 22.36$. The reason of these satisfactory results is that the the seasonality is explicitly modelled in the present analysis, whereas the residuals of the local level and local linear trend model contained the neglected seasonality in monthly data. Since the assumptions of homoscedasticity and normality are also realistic (see Table 4.2), the residuals of this analysis satisfy all the required criteria.

The Akaike information criterion for the stochastic level and seasonal model equals

$$\text{AIC} = \frac{1}{192}\left[-2(192)(0.9369063) + 2(12 + 3)\right] = -1.71756,$$

Table 4.2. Diagnostic tests for stochastic level and seasonal model and log UK drivers KSI.

	statistic	value	critical value	assumption satisfied
independence	$Q(15)$	14.150	22.36	+
	$r(1)$	0.039	±0.14	+
	$r(12)$	0.014	±0.14	+
homoscedasticity	$H(60)$	1.060	1.67	+
normality	N	5.289	5.99	+

indicating that this is the preferred model for the log of the UK drivers KSI series so far, even though it requires the estimation of a total of 15 parameters: one variance for the irregular component, two variances for the level and seasonal component, and 12 initial values of the state (one for the level, and 11 for the seasonal). Moreover, the present model also fits the data much better than the classical multiple regression analysis obtained with deterministic level and seasonal components.

Since the variance of the seasonal disturbances is found to be almost zero, in the next section we present the results of the analysis of the UK drivers KSI series with a stochastic level and a deterministic seasonal.

4.3. Stochastic level and deterministic seasonal

Fixing the seasonal disturbances ω_t in model (4.1) to zero, but still allowing the level to vary over time yields the following results:

```
it0    f=     0.9362753 df=  0.003305 e1=   0.01239 e2= 0.0001078
it1    f=     0.9362925 df=  0.003487 e1=   0.01310 e2= 0.0003366
it2    f=     0.9363240 df=  0.002234 e1=  0.008362 e2= 0.0003377
it3    f=     0.9363352 df=  0.001322 e1=  0.004066 e2= 0.0002726
it4    f=     0.9363361 df= 0.0002666 e1= 0.0008200 e2=4.323e-005
it5    f=     0.9363361 df=1.145e-005 e1=3.522e-005 e2=8.119e-006
Strong convergence
```

At convergence the value of the log-likelihood function is 0.9363361. The maximum likelihood estimate of the variance of the irregular component is $\hat{\sigma}_\varepsilon^2 = 0.00351385$, and the maximum likelihood estimate of the variance of the level disturbances is $\hat{\sigma}_\xi^2 = 0.000945723$. Plots of the results of this analysis are not shown here, because they are very similar to the ones presented in Section 4.2. The same applies to the results of the diagnostic tests which are very similar to those given in Table 4.2.

The Akaike information criterion for this model equals

$$\text{AIC} = \frac{1}{192}\left[-2(192)(0.9363361) + 2(12 + 2)\right] = -1.72684$$

indicating a slight improvement upon the previous model: the small reduction in the value of the log-likelihood function is compensated by the fact that the present model requires the estimation of only two variances instead of three in the previous model.

The AIC value of −1.72684 for the stochastic level and deterministic seasonal model is a significant improvement upon the local level model, which yields an AIC value of −1.25914. Therefore, and in contrast with the slope component, the addition of a seasonal component is essential in obtaining a good description of the time series at hand.

In this chapter the first *realistic and appropriate description* of the log of the number of UK drivers KSI is presented by combining a stochastic level with a deterministic seasonal component. Furthermore it is shown that a stochastic state space model can be reduced to its equivalent classical regression model by *fixing all state disturbances to zero*. This means that classical linear regression models can be viewed as deterministic state space models.

4.4. The local level and seasonal model and UK inflation

We end this chapter by discussing the results of the analysis of a time series consisting of quarterly UK inflation figures (as given in Appendix D, and displayed at the top of Figure 4.10) with the local level and seasonal model. In economics, the inflation is an important variable that refers to a rise in the general level of prices (deflation usually refers to a fall in prices). Economic policy makers find it important to have a good estimate of inflation. In practice, inflation is taken as the relative price change, usually expressed in a percentage.

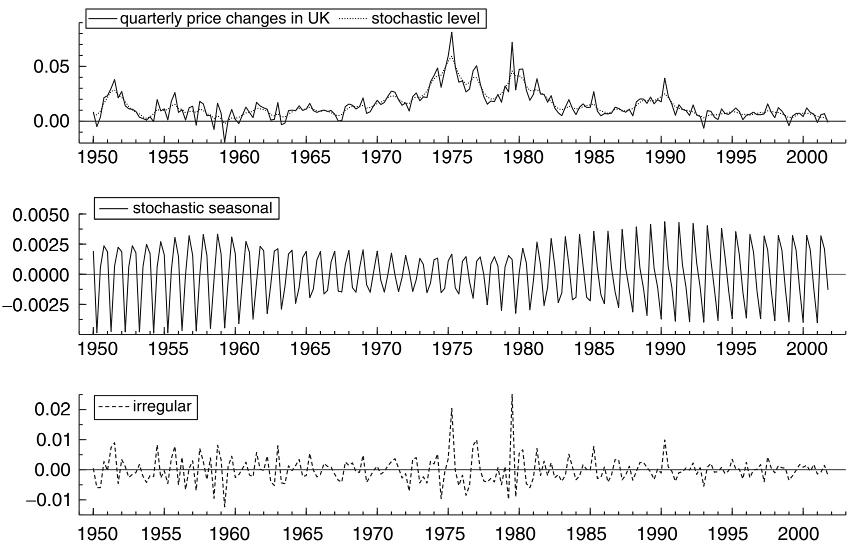

Figure 4.10. Stochastic level, seasonal and irregular in UK inflation series.

The percentage change of the price level over a quarter is not considered to be a reliable estimator of inflation. Instead, quarterly time series of price changes are analysed by time series models to assess inflation. The local level model is an appropriate candidate for this purpose. The final estimate of the level is then an appropriate estimator of the underlying rate of inflation as this represents the underlying inflation for the intermediate and longer term. Inflation relates to average household purchases that can be subject to seasonal variations due to events such as Christmas and summer holiday. As we are dealing with quarterly data, we include a stochastic seasonal component with a periodicity of $s = 4$ in the local level model. This approach of measuring inflation is illustrated by applying it to quarterly price changes in the United Kingdom for the 52 years from 1950 through to 2001 (yielding a total of $n = 52 \times 4 = 208$ observations). The estimation of the parameters in model (4.1) applied to the UK inflation series gives the following results:

```
it0    f=      3.023196 df=    0.1800 e1=     1.119 e2=  0.002894
it1    f=      3.069515 df=    0.1586 e1=     1.015 e2=   0.01299
it2    f=      3.164341 df=    0.1016 e1=    0.5279 e2=   0.01150
it5    f=      3.194490 df=   0.02758 e1=    0.1484 e2=  0.001452
it10   f=      3.198464 df=4.081e-005 e1= 0.0002241 e2=5.183e-005
it11   f=      3.198464 df=3.960e-006 e1=2.175e-005 e2=3.472e-006
Strong convergence
```

At convergence the value of the log-likelihood function is 3.198464. The maximum likelihood estimate of the irregular variance is $\widehat{\sigma}^2_\varepsilon = 3.3717 \times 10^{-5}$ and the maximum likelihood estimates of the variances of the level and seasonal disturbances are equal to $\widehat{\sigma}^2_\xi = 2.1197 \times 10^{-5}$ and $\widehat{\sigma}^2_\omega = 0.0109 \times 10^{-5}$, respectively.

The estimate of the final value of the level at time point $t = 208$ is $\widehat{\mu}_{208} = 0.0020426$. This is our estimate of inflation. As a result, relative prices have increased overall by 0.20% in the final months of 2001. This is rather low. The evolution of inflation over time is reflected by the estimated level component and is presented in the upper graph of Figure 4.10, together with the observed price changes. It is noteworthy that the periods of highest inflation in the UK occurred in the middle of the 1970s and at the end of the 1970s. These periods coincide with the well-known oil and energy crises in the 1970s.

Graphs of the stochastic seasonal and irregular components are also displayed in Figure 4.10. Although the variance of the seasonal disturbances is smaller than that of the other two components, the changes over time in the estimated seasonal component of inflation series are clearly visible. The level component reflects the underlying level of inflation and

Table 4.3. Diagnostic tests for local level and seasonal model and UK inflation series.

	statistic	value	critical value	assumption satisfied
independence	$Q(10)$	7.573	15.507	+
	$r(1)$	0.049	± 0.14	+
	$r(4)$	−0.0622	± 0.14	+
homoscedasticity	$H(68)$	2.738	1.48	−
normality	N	171.550	5.99	−

its evolution over time is quite smooth. The residuals of this level plus seasonal model are close to independent random values (white noise). Some outlier observations appear in the irregular component but apart from these, the residuals seem quite random. Whether the residuals of the local level and seasonal model are close enough to a random process (see Section 10.1.2 for the definition of a random process) can be established by inspection of the diagnostic tests given in Table 4.3.

The last column in Table 4.3 shows that the diagnostics for independence are quite satisfactory. However, the assumptions of homoscedasticity and normality tests are clearly violated. The local level and seasonal model is therefore able to represent the dynamic features in the UK inflation series, but there are also some aspects in the series that still need to be accounted for. Specifically, the neglect in the present model of the large shocks in the estimated irregular component for the UK inflation series at the time points corresponding to the second quarter of 1975 and to the third quarter of 1979 deserve closer attention. It should not come as a surprise that these two time points are related to the world-wide oil and energy crises in the 1970s. An appropriate treatment of these 'outlier observations' will be discussed in Section 7.4.

The AIC for the present model equals

$$\text{AIC} = \frac{1}{208}\left[-2(208)(3.198464) + 2(4+3)\right] = -6.32962,$$

and this value will be used for reference purposes in Chapter 7.

In Chapters 5 and 6, components of the state are introduced that can be used to obtain *explanations* for the observed developments of a time series. The discussion of these components will be illustrated by adding them to the UK drivers KSI series. To keep the exposition as simple as possible, the seasonal component will temporarily be removed from these analyses, even though this component is clearly essential in describing the UK drivers KSI series. In the next two chapters, we are not concerned with the appropriateness of the models when applied to the UK drivers KSI series (and diagnostic residual tests will therefore not be presented). We mainly focus on various issues concerning the inclusion of explanatory variables in the state space models of Chapters 2 and 3. Nevertheless, in Chapter 7 – where a model is presented for the combined *description* and *explanation* of the log of the UK number of drivers KSI – the seasonal component will be added to the state equations.