8 Adding seasonal effects

A seasonal effect is a repeating pattern with a known fixed period, p say. For identifiability we constrain seasonal effects to cancel out over the whole period. Thus if we model seasonal effects additively, then at any point the sum of all p effects must be equal to zero with probability 1. As a general rule, it is always better to model seasonality explicitly than to proceed with someone else's deseasonalised data. Seasonality, like everything else, is partly a matter of beliefs.

Suppose the data y_t are expected to have a seasonal effect of period p, and denote the p seasonal factors by the state vector ψ_t. The seasonal factor model is then

$$\begin{aligned} y_t &= E_p^\intercal \psi_t + \nu_t & \nu_t &\sim \langle 0,\, V\rangle \\ \psi_t &= P\,\psi_{t-1} + \omega_t & \omega_t &\sim \langle \mathbf{0},\, W\rangle \end{aligned}$$

where $E_p = (1, 0, \dots, 0)^\intercal$, as before, and

$$P = \begin{pmatrix} \mathbf{0} & I_{p-1} \\ 1 & \mathbf{0}^\intercal \end{pmatrix}.$$

The matrix P is known as a 'permutation matrix'. Its effect is to switch the first component of ψ_{t-1} to the end of the vector and bump all of the other components up one: you could say it 'rotates' ψ_{t-1} by one time-period. Consequently the p components of ψ_t appear consecutively in the observation equation. This is an important and subtle point: at any time t we have beliefs about all four seasonal effects, but only one is being used in the observation equation.

In order that ψ_t represents seasonal factors, we must impose the identifying constraint that $\mathbf{1}^\intercal\psi_t = 0$ with probability 1, where $\mathbf{1}$ is the unit vector. In other words we want $\mathsf{E}[\mathbf{1}^\intercal\psi_t \mid D_t] = \mathbf{1}^\intercal m_t = 0$ and $\mathsf{Var}\,[\mathbf{1}^\intercal\psi_t \mid D_t] = \mathbf{1}^\intercal C_t\,\mathbf{1} = 0$. What is clever about DLMs is the way in which this constraint can be made to hold for all t by careful choice of m_0, C_0 and W.

Suppose, inductively, that we have $\mathbf{1}^\intercal m_{t-1} = 0$ and $C_{t-1}\mathbf{1} = \mathbf{0}$ (which implies that $\mathbf{1}^\intercal C_{t-1}\,\mathbf{1} = 0$). Our prior mean and variance for $\psi_t \mid D_{t-1}$ are $a_t = P m_{t-1}$ and $R_t = P C_{t-1} P^\intercal + W$. As $\mathbf{1}^\intercal P = \mathbf{1}^\intercal$, we have $\mathbf{1}^\intercal a_t = \mathbf{1}^\intercal m_{t-1} = 0$, and $R_t\,\mathbf{1} = W\mathbf{1}$. If we require that W is chosen so that $W\mathbf{1} = \mathbf{0}$, then we must have $R_t\,\mathbf{1} = \mathbf{0}$, or, equivalently, $\mathbf{1}^\intercal R_t = \mathbf{0}^\intercal$. For our adjusted mean and variance we have

$$\begin{aligned} m_t &= a_t + \big(R_t\,E_p\big)\,Q_t^{-1}\,\big(y_t - E_p^\intercal\,a_t\big) \\ C_t &= C_{t-1} + \big(R_t\,E_p\big)\,Q_t^{-1}\big(R_t\,E_p\big)^\intercal \end{aligned}$$

Thus we must have $\mathbf{1}^\intercal m_t = 0$ and $C_t\,\mathbf{1} = \mathbf{0}$. Now to make this work for all t we must start with m_0 and C_0 such that $\mathbf{1}^\intercal m_0 = 0$ and $C_0\,\mathbf{1} = \mathbf{0}$. In this case, if $W\mathbf{1} = \mathbf{0}$ then $\mathbf{1}^\intercal\psi_t = 0$ with probability 1, as required.

How do we ensure that C_0 and W satisfy the constraint? One common approach when the seasonal factors are considered to be exchangeable is to construct the variance matrices in the form

$$C_0 = c_0\big(p\,I_p - \mathbf{1}\mathbf{1}^\intercal\big)$$

for some scalar $c_0 \geq 0$, for then

$$C_0\,\mathbf{1} = c_0\big(p\,I_p - \mathbf{1}\mathbf{1}^\intercal\big)\mathbf{1} = c_0\big(p\mathbf{1} - \mathbf{1}p\big) = \mathbf{0},$$

as required. A similar scalar $w \geq 0$ will be required for W. The value w determines the amount by which the seasonal factors can change from period to period and should be chosen with this in mind.

Another approach is to specify C_0 as above, but to use *discounting* to specify W. Recollect that we can recover an implicit value for W_t as

$$W_t = \frac{1-\delta}{\delta}\,P C_{t-1} P^\intercal$$

for given discount rate $\delta \in (0,1]$. As $P^{\intercal}\mathbf{1} = \mathbf{1}$ and $C_{t-1}\,\mathbf{1} = \mathbf{0}$, so we see that $W_t\,\mathbf{1} = \mathbf{0}$ as required. So discounting automatically ensures that the seasonal factors sum to zero with probability 1. Discounting is very popular in seasonal models.

9 Example: Property transactions

9.1 Some real data!

Our data comprise the number of property transactions in England and Wales, quarterly from 1959q1 to 1999q2. The source of this data is *Economic Trends*, Annual Supplement 1999, Table 4.5, p. 196 (which may be found on the ground floor of the Main Library). This data is plotted in Figure 9.3. It is clear that this data is locally trending, as you might expect from the notorious boom/bust behaviour of the residential housing market. It is also possible that this data is seasonal, although this is not immediately obvious from the Figure.

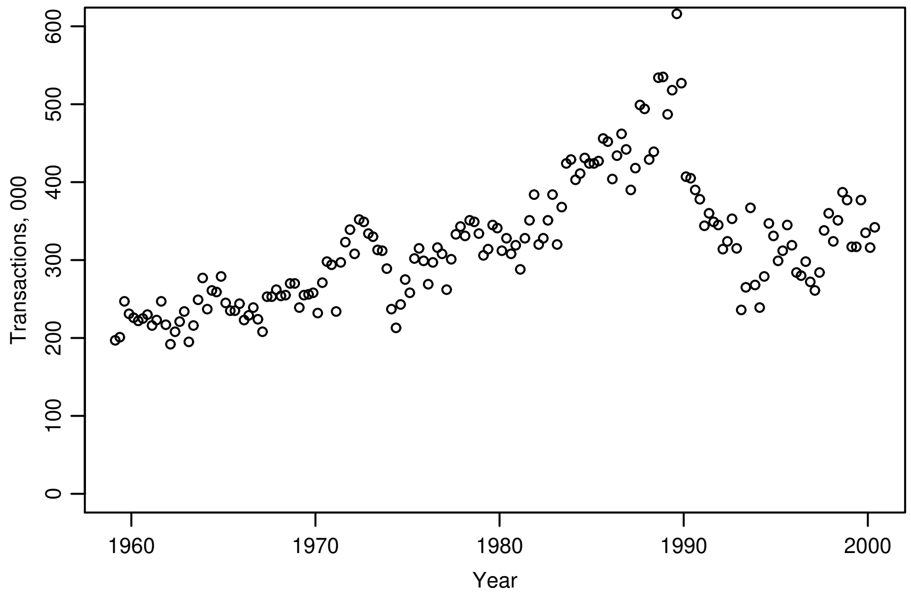

Figure 9.3: Number of property transactions, England and Wales, 000s

9.2 Our Dynamic Linear Model

We transform our data by taking the natural logarithm, and so the error terms should be interpreted as proportionate effects on the original scale. Remember that if $z = \log y$ and z has a gaussian distribution with mean μ and variance σ^2, then y has a lognormal distribution with mean $\mu' = \exp(\mu + 0.5\sigma^2)$ and variance $(\sigma')^2 = (\mu')^2 \left\{\exp(\sigma^2) - 1\right\}$. We are going to suppose that $\log y$ is approximately gaussian, so that we can compute the mean and variance for y from the mean and variance of $\log y$ using these two relations.

In order to capture both the trend and the seasonal we use the direct sum of a local linear trend model and a four-period seasonal model, to give

$$y_t = \begin{pmatrix} E_2 \\ E_4 \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} \theta_t \\ \psi_t \end{pmatrix} + \nu_t \qquad \nu_t \sim \langle 0,\, v \rangle$$

$$\begin{pmatrix} \theta_t \\ \psi_t \end{pmatrix} = \begin{pmatrix} L_2 & \mathbf{0} \\ \mathbf{0} & P \end{pmatrix} \begin{pmatrix} \theta_{t-1} \\ \psi_{t-1} \end{pmatrix} + \omega_t \qquad \omega_t \sim \langle \mathbf{0},\, W \rangle$$

for $t = 1, \ldots, T$, where

$$W = \begin{pmatrix} \begin{pmatrix} w_1 + w_2 & w_2 \\ w_2 & w_2 \end{pmatrix} & \mathbf{0} \\ \mathbf{0} & w_3 \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \end{pmatrix}$$

We need to specify four variance terms:

$$v = 0.05^2 \quad w_1 = 0.02^2 \quad w_2 = 0.005^2 \quad w_3 = 0.005^2/3.$$

Remember that these are proportions. We also need some initial beliefs

(which are less important in this application as we have plenty of data)

$$m_0 = \left(\log(200),\, 0,\, 0,\, 0,\, 0,\, 0\right)^{\top} \quad C_0 = 10 \times W$$

where we are careful to ensure that the initial beliefs satisfy the seasonality conditions.

9.3 Results

In order to incorporate all of the data, we first update the state vector using our data $y_1, y_2, \ldots, y_T$ until we get to $\theta_T \mid D_T$, and then we filter back again. This gives us $\theta_1, \ldots, \theta_T \mid D_T$. With this information we can do the following operations.

Extracting the trend

The trend in our model is the first component of the state vector, θ_{t1}. Figure 9.4 show the mean of the trend with ± 2 standard deviation (s.d.) error bands superimposed upon the original data. Differences between the trend and the data are the sum of the seasonal component plus noise. From this Figure the boom/bust nature of the property is very clear. We have major booms in the early Seventies and late Eighties. The other 'boomlets' show a striking relation to the electoral cycle, which cynical commentators attribute to the effect of the governing party attempting to engineer low interest rates and falling taxes in the year prior to the election. If I was cynical I might be tempted to include the electoral cycle as a second source of seasonality.

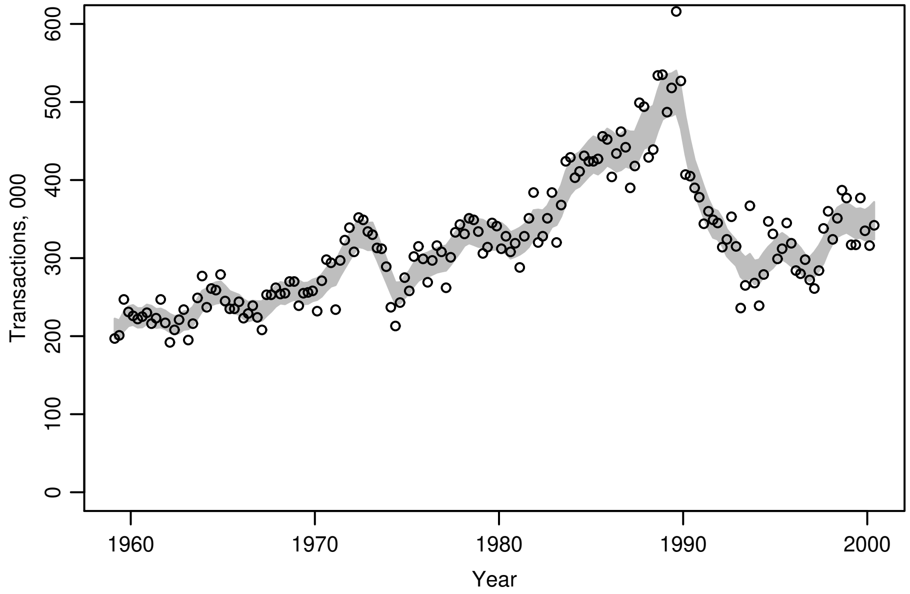

Figure 9.4: Trend property transactions, with ± 2 s.d. error bands shown in gray.

Extract the seasonal components

Second, we may be interested in the seasonal factors themselves. For example, are they constant over time? The seasonal factors are given by ψ_{t1}, the third component of the joint state vector. The series $\mathsf{E}[\psi_{t1} \mid D_T]$

gives the contemporaneous expected seasonal factor for each period, and then we can separate these out into the four quarters. Figure 9.5 shows the means of the four seasonal factors over time: as the data is in logs these are proportions. It is strange to note that the number of property transactions is becoming more seasonal: from $\pm 3\%$ in the early 1960s to $\pm 10\%$ at the end of the 1990s. In particular people seem to be deferring more from the first and second quarters until the third quarter. Take this to be a cautionary tale about the non-constancy of seasonal factors.

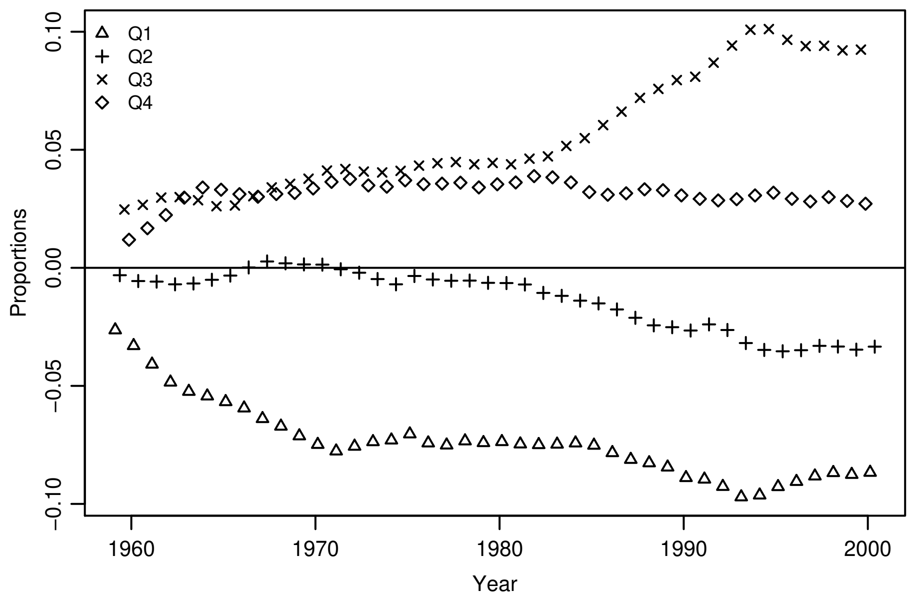

Figure 9.5: The four seasonal factors

Make a forecast

Third, we can make forecasts of future numbers of transactions, including seasonal factors if required. For this we use the mean response function μ_{T+k}, $k = 1, 2, \ldots$. Figure 9.6 shows the three year forecast with ± 2 s.d. error bands. This forecast probably goes a little further into the future than is warranted by the model. The mean forecast is highlighted in the centre of each bar. A slight upward trend is apparent, and the seasonal pattern is easy to spot given that the underlying mean is linear. So ... good news for estate agents then.

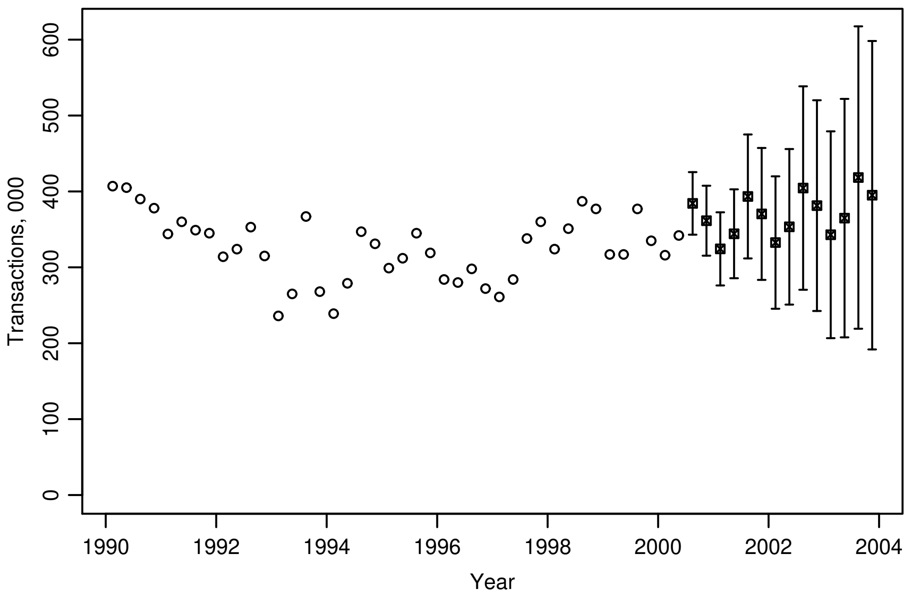

Figure 9.6: Three year forecast, with ± 2 s.d. error bands

9.4 Missing data

We are fortunate in having all of the data for this example, but that won't stop us from pretending that some of it was missing. Suppose that the data for 1980 was missing, and we would like to impute that data, i.e. compute beliefs about $\mu_{1980q1,\ldots,q4} \mid D_T$. On the way through the data on the update pass, we must set $m_t = a_t$ and $C_t = R_t$ for the four 1980 values. On the way back down again, we filter as normal. The result is shown in Figure 9.7. The 1980 values have been replaced with the mean of the mean response function. The 'true' values are not shown, but all four lie within the the ± 2 s.d. error bands.

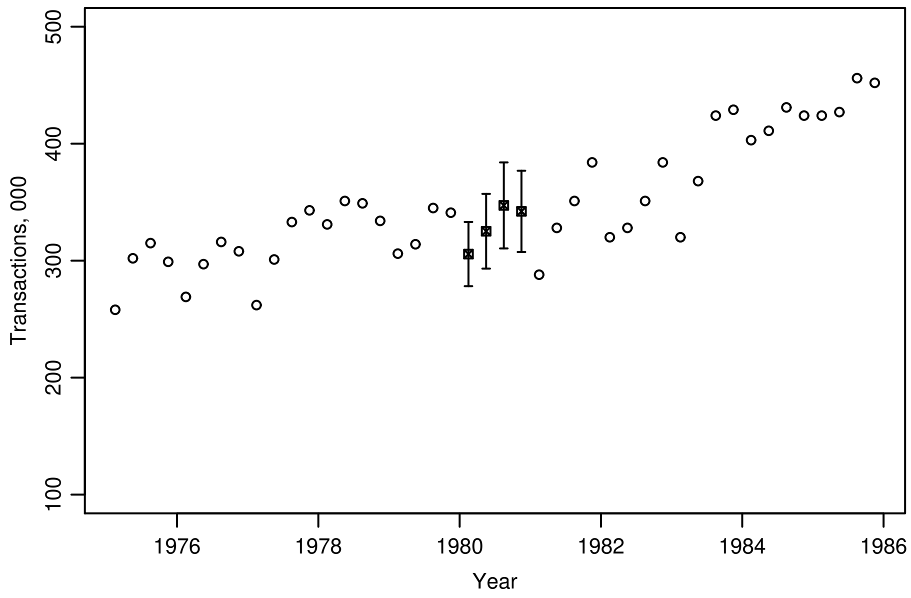

Figure 9.7: Imputing missing data by filtering. The imputed means are shown with ± 2 s.d. error bands.