West & Harrison Chapter 10, Problem 7

The Dans

Thursday, March 16, 2017

The quarterly sales and cost index of a confectionary product are given as

	SALES				COST			
	Quarter				Quarter			
Year	1	2	3	4	1	2	3	4
1975	157	227	240	191	10.6	8.5	6.7	4.1
1976	157	232	254	198	1.9	0.4	1.1	1.9
1977	169	234	241	167	0.2	2.9	4.1	−1.4
1978	163	227	252	185	−4.0	−4.5	−5.3	−8.4
1979	179	261	264	196	−12.8	−13.2	−10.1	−4.6
1980	179	248	256	193	−1.1	−0.1	0.0	−2.5
1981	186	260	270	210	−5.1	−6.4	−8.0	−6.5
1982	171	227	241	170	−3.7	−1.3	6.1	16.5
1983	140	218	208	193	22.9	23.9	18.0	8.3
1984	184	235	245	209	2.9	0.7	−2.4	−7.0
1985	206	260	264	227	−9.8	−10.6	−12.3	−13.2

The goal is to fit a DLM to the sales with first-order polynomial, regression on cost, and full seasonal components. Based on previous information,

- the underlying level of the series when cost is zero is 220 with a nominal standard error of 15
- the regression coefficient of cost is estimated as -1.5 with a standard error of about 0.7
- the seasonal factors for the four quarters of the first year are expected to be -50, 25, 50, -25, with nominal standard errors of 25, 15, 25, and 15, respectively
- the trend, regression, and seasonal components are initially assumed to be uncorrelated
- the observational variance is estimated as 100, with initial degrees of freedom of 12

Using this information, we are to analyze the series along the following lines:

(a) **Using the information provided for the seasonal factors above, apply Theorem 8.2 to derive the appropriate initial prior that satisfies the zero-sum constraint.**

Theorem 8.2: Imposing the constraint $\mathbf{1}'\boldsymbol{\phi} = 0$ *on the prior* $(\boldsymbol{\phi}_0|\mathcal{D}_0^*) \sim N(\mathbf{m}_0^*, \mathbf{C}_0^*)$ *and writing* $U = \mathbf{1}'\mathbf{C}_0^*\mathbf{1}$ *and* $\mathbf{A} = \mathbf{C}_0^*\mathbf{1}/U$ *gives the revised joint prior*

$$(\boldsymbol{\phi}_0|\mathcal{D}_0) \sim \mathrm{N}(\mathbf{m}_0, \mathbf{C}_0),$$
$$\mathbf{m}_0 = \mathbf{m}_0^* - \mathbf{A}\mathbf{1}'\mathbf{m}_0^*,$$
$$\mathbf{C}_0 = \mathbf{C}_0^* - \mathbf{A}\mathbf{A}'U.$$

Using the given information for the seasonal components,

$$\mathbf{m}_0^* = (-25, -50, 25, 50)' \qquad \mathbf{C}_0^* = \begin{pmatrix} 225 & 0 & 0 & 0 \\ 0 & 625 & 0 & 0 \\ 0 & 0 & 225 & 0 \\ 0 & 0 & 0 & 625 \end{pmatrix}.$$

Applying Theorem 8.2 yields

$$
\begin{aligned}
U &= 1700 \\
\mathbf{A} &= (0.1323529, 0.3676471, 0.1323529, 0.3676471)' \\
\mathbf{m}_0^s &= (-25, -50, 25, 50)' \\
\mathbf{C}_0^s &= \begin{pmatrix} 195.22 & -82.72 & -29.78 & -82.72 \\ -82.72 & 395.22 & -82.72 & -229.78 \\ -29.78 & -82.72 & 195.22 & -82.72 \\ -82.72 & -229.78 & -82.72 & 395.22 \end{pmatrix}.
\end{aligned}
$$

This leads to the initial prior satisfying the zero-sum constraint for the seasonal components to be $(\boldsymbol{\phi}_0|\mathcal{D}_0) \sim \mathrm{N}(\mathbf{m}_0^s, \mathbf{C}_0^s)$. Given a state vector $\boldsymbol{\theta}_t = (\theta_{1,t}, \theta_{2,t}, \phi_{1,t}, \dots, \phi_{4,t})'$, the initial prior for the DLM is of the form $(\boldsymbol{\theta}_0|\mathcal{D}_0) \sim \mathrm{N}(\mathbf{m}_0, \mathbf{C}_0)$, where

$$
\begin{aligned}
\mathbf{m}_0 &= (220, -1.5, -25, -50, 25, 50)' \\
\mathbf{C}_0 &= \begin{pmatrix} 225.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.49 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 195.22 & -82.72 & -29.78 & -82.72 \\ 0.00 & 0.00 & -82.72 & 395.22 & -82.72 & -229.78 \\ 0.00 & 0.00 & -29.78 & -82.72 & 195.22 & -82.72 \\ 0.00 & 0.00 & -82.72 & -229.78 & -82.72 & 395.22 \end{pmatrix}.
\end{aligned}
$$

(b) **Identify the full initial prior quantities for the 6-dimensional state vector $\boldsymbol{\theta}_1$ and the observational variance V. Write down the defining quantities $\mathbf{a}_1, \mathbf{R}_1, n_0$, and S_0 based on the above initial information.**

The DLM model in this case is defined as $\{\mathbf{F}_t, \mathbf{G}, V, \mathbf{W}_t\}$, with

$$
\mathbf{F}_t = (1, \mathrm{Cost}_t, 1, 0, 0, 0)' \qquad \mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.
$$

Using this information,

$$
\begin{aligned}
\mathbf{a}_1 &= \mathbf{G}\mathbf{m}_0 = (220, -1.5, 25, 50, -25, -50)' \\
\mathbf{R}_1 &= \mathbf{G}\mathbf{C}_0\mathbf{G}' + \mathbf{W}_t = \begin{pmatrix} 225.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.49 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 395.22 & -82.72 & -229.78 & -82.72 \\ 0.00 & 0.00 & -82.72 & 195.22 & -82.72 & -29.78 \\ 0.00 & 0.00 & -229.78 & -82.72 & 395.22 & -82.72 \\ 0.00 & 0.00 & -82.72 & -29.78 & -82.72 & 195.22 \end{pmatrix} + \mathbf{W}_t \\
n_0 &= 12 \\
S_0 &= 100.
\end{aligned}
$$

Therefore, $(\boldsymbol{\theta}_1|\mathcal{D}_0) \sim \mathrm{T}_{12}(\mathbf{a}_1, \mathbf{R}_1)$ and $(V|\mathcal{D}_0) \sim$ Inverse Gamma$(n_0/2, d_0/2)$, where $d_0 = n_0 S_0 = 1200$.

(c) **Use three discount factors to structure the evolution matrices of the model: δ_T for the constant intercept term, δ_R for the regression coefficient, and δ_S for the seasonal factors. Consider initially the values $\delta_T = \delta_S = 0.9$ and $\delta_R = 0.98$. Fit the model and perform the retrospective, filtering calculations to obtain filtered estimates of the state vector and all model components over time.**

The filtered estimates of each component of the state vector are shown in figure 1. The one-step ahead forecast and 95% confidence interval can be seen in figure 2.

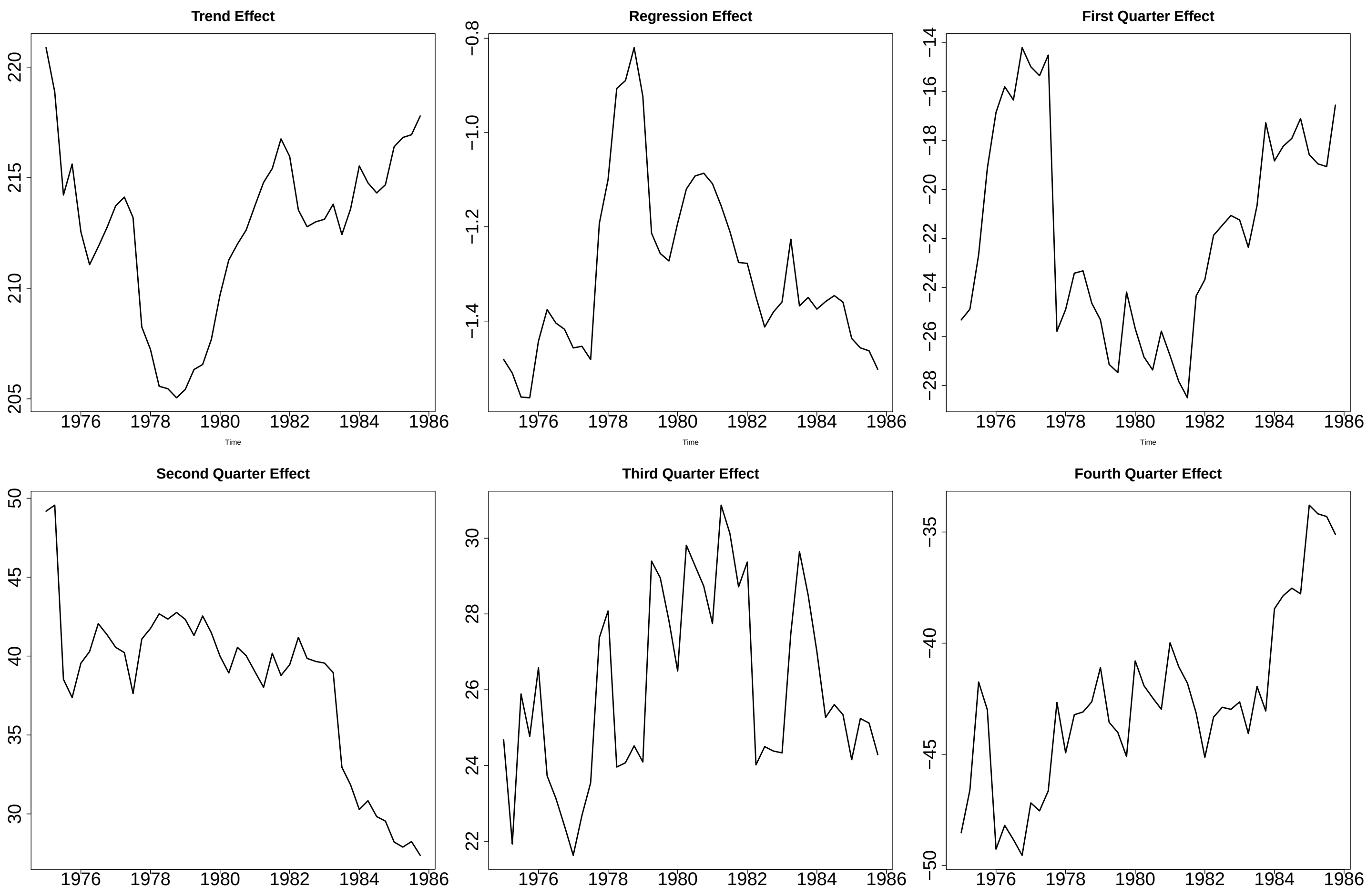

Figure 1: Filtered estimates for all components of the state vector

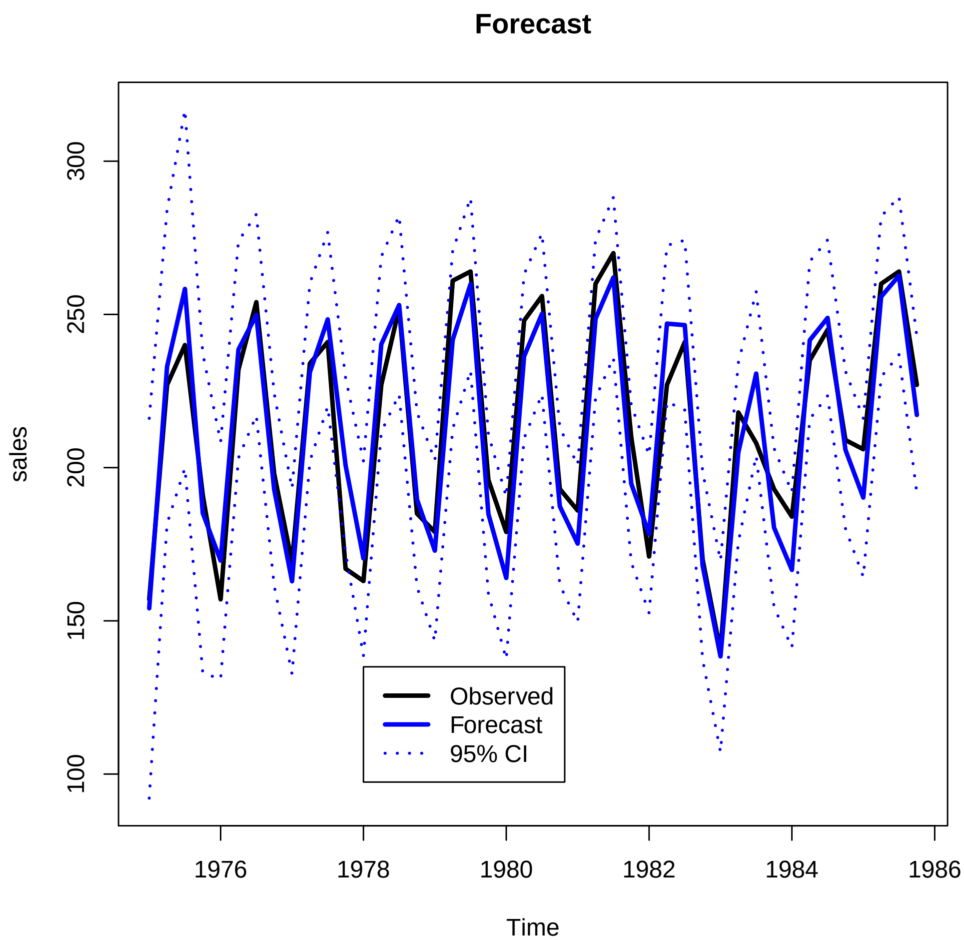

Figure 2: One-step ahead forecast estimate and 95% confidence interval

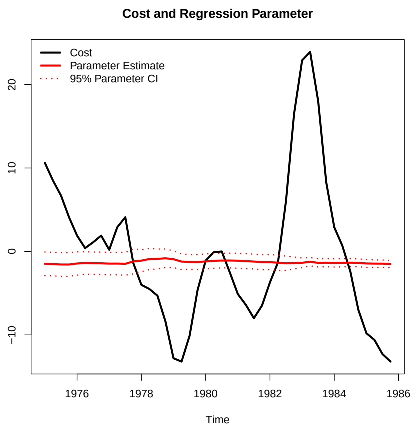

Figure 3: Plot of the observed cost and a summary of the distribution of the regression parameter

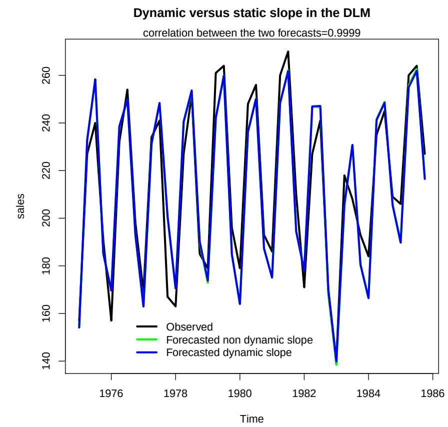

Figure 4: Plot of the forecasts using a dynamic slope (blue) and a nondynamic slope (green)

(d) **Based on this analysis, verify the findings in West et al. [1987] to the effect that the regression parameter on Cost is, in retrospect, rather stable over time.**

Figure 3 shows the values for Cost relative to the estimates of the regression parameter over time. It does not appear that the values for the regression parameter change very much over time, always hovering between -1.5 and -0.8

To confirm this analysis on the scale of the forecast, a similar model to the model in part (c) was fit, but with the slope parameter forced to be non-dynamic (by setting the discount factor to 1). Figure 4 shows the forecast functions for the dynamic and fixed slope models. Almost no difference is seen, and the correlation between the two forecast functions is 0.9999.

Based on these two analyses, a dynamic slope appears to be unnecessary for this dataset.

(e) **Produce step ahead forecasts from the end of the data in the fourth quarter of 1985 for the next three years. The estimated values of Cost to be used in forecasting ahead are given by**

Year	Quarter 1	2	3	4
1986	8.4	10.6	7.2	13.0
1987	−2.9	−0.7	−6.4	−7.0
1988	−14.9	−15.9	−18.0	−22.3

For this problem, the dynamic model for slope was used even though it wasn't really necessary.

Appealing to Harrison and West [1999] page 199, and letting $F(k)$ be the F matrix at time k, the k-step forecast distributions are $T_{38}(Ft(k), Qt(k))$ where $Ft(k) = F(k)^\top * at(k)$, $Qt(k) = F(k)^\top * Rt(k) * F(k)$, $at(k) = G * at(k-1)$, and $Rt(k) = G * Rt(k-1) * G^\top * (diag(\frac{1}{\delta_t^k}, \frac{1}{\delta_r^k}, \frac{1}{\delta_s^k}, \frac{1}{\delta_s^k}, \frac{1}{\delta_s^k}, \frac{1}{\delta_s^k}))$ The step ahead forecasts and 90% posterior predictive intervals are shown figure 5.

References

Jeff Harrison and Mike West. *Bayesian forecasting & dynamic models*. Springer New York, 1999.

Raquel Prado and Mike West. *Time series: modeling, computation, and inference*. CRC Press, 2010.

M West, PJ Harrison, and A Pole. Bats: A user guide. *Warwick Research Report*, page 114, 1987.

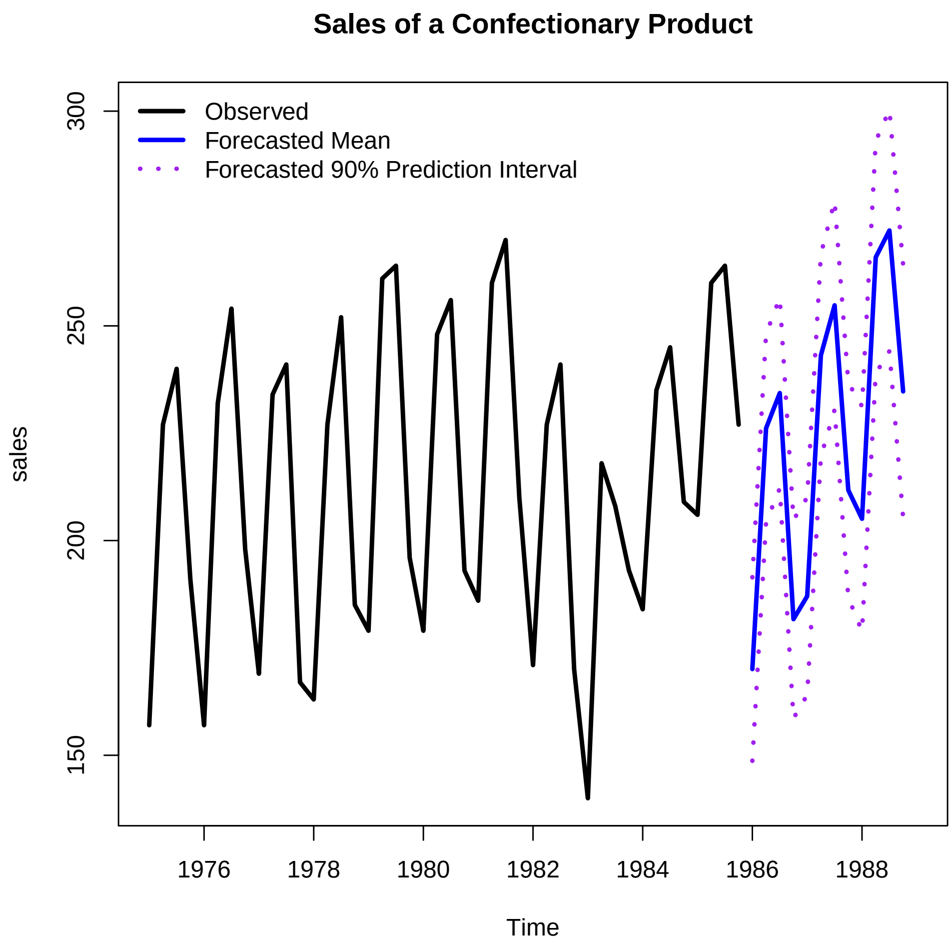

Figure 5: Step ahead forecasts for 1986 - 1988