An automatic report for the dataset : 07-call-centre

The Automatic Statistician

Abstract

This report was produced by the Automatic Bayesian Covariance Discovery (ABCD) algorithm.

1 Executive summary

The raw data and full model posterior with extrapolations are shown in figure 1.

Figure 1: Raw data (left) and model posterior with extrapolation (right)

The structure search algorithm has identified six additive components in the data. The first 2 additive components explain 94.5% of the variation in the data as shown by the coefficient of determination ($R^2$) values in table 1. The first 3 additive components explain 99.1% of the variation in the data. After the first 4 components the cross validated mean absolute error (MAE) does not decrease by more than 0.1%. This suggests that subsequent terms are modelling very short term trends, uncorrelated noise or are artefacts of the model or search procedure. Short summaries of the additive components are as follows:

- A linearly increasing function. This function applies until Feb 1974.
- A very smooth monotonically increasing function. This function applies from Feb 1974 onwards.
- A smooth function with marginal standard deviation increasing linearly away from Feb 1964. This function applies until Feb 1974.
- An exactly periodic function with a period of 1.0 years. This function applies until Feb 1974.
- Uncorrelated noise. This function applies until May 1973 and from Oct 1973 onwards.

Model checking statistics are summarised in table 2 in section 4. These statistics have not revealed any inconsistencies between the model and observed data.

The rest of the document is structured as follows. In section 2 the forms of the additive components are described and their posterior distributions are displayed. In section 3 the modelling assumptions of each component are discussed with reference to how this affects the extrapolations made by the
model. Section 4 discusses model checking statistics, with plots showing the form of any detected discrepancies between the model and observed data.

2 Detailed discussion of additive components

2.1 Component 1: A linearly increasing function. This function applies until Feb 1974

This component is linearly increasing. This component applies until Feb 1974.

This component explains 77.5% of the total variance. The addition of this component reduces the cross validated MAE by 86.5% from 492.5 to 66.3.

![Posterior of component 1](image1)

![Sum of components up to component 1](image2)

Figure 2: Pointwise posterior of component 1 (left) and the posterior of the cumulative sum of components with data (right)

![Residuals after component 1](image3)

Figure 3: Pointwise posterior of residuals after adding component 1
2.2 Component 2: A very smooth monotonically increasing function. This function applies from Feb 1974 onwards

This component is a very smooth and monotonically increasing function. This component applies from Feb 1974 onwards.

This component explains 75.4% of the residual variance; this increases the total variance explained from 77.5% to 94.5%. The addition of this component reduces the cross validated MAE by 51.95% from 66.30 to 31.86.

![Posterior of component 2](image1)

**Figure 4:** Pointwise posterior of component 2 (left) and the posterior of the cumulative sum of components with data (right)

![Residuals after component 2](image2)

**Figure 5:** Pointwise posterior of residuals after adding component 2

2.3 Component 3: A smooth function with marginal standard deviation increasing linearly away from Feb 1964. This function applies until Feb 1974

This component is a smooth function with a typical lengthscale of 3.5 months. The marginal standard deviation of the function increases linearly away from Feb 1964. This component applies until Feb 1974.

This component explains 84.0% of the residual variance; this increases the total variance explained from 94.5% to 99.1%. The addition of this component reduces the cross validated MAE by 24.10% from 31.86 to 24.18.

![Posterior of component 3](image3)

**Figure 6:** Pointwise posterior of component 3 (left) and the posterior of the cumulative sum of components with data (right)
2.4 Component 4: An exactly periodic function with a period of 1.0 years. This function applies until Feb 1974

This component is exactly periodic with a period of 1.0 years. The shape of this function within each period has a typical lengthscale of 3.9 weeks. This component applies until Feb 1974.

This component explains 66.6% of the residual variance; this increases the total variance explained from 99.1% to 99.7%. The addition of this component reduces the cross validated MAE by 13.35% from 24.18 to 20.95.

2.5 Component 5: Uncorrelated noise. This function applies until May 1973 and from Oct 1973 onwards

This component models uncorrelated noise. This component applies until May 1973 and from Oct 1973 onwards.

This component explains 74.0% of the residual variance; this increases the total variance explained from 99.7% to 99.9%. The addition of this component reduces the cross validated MAE by 0.00% from 20.95 to 20.95. This component explains residual variance but does not improve MAE which suggests that this component describes very short term patterns, uncorrelated noise or is an artefact of the model or search procedure.
2.6 Component 6: Uncorrelated noise. This function applies from May 1973 until Oct 1973

This component models uncorrelated noise. This component applies from May 1973 until Oct 1973. This component explains 100.0% of the residual variance; this increases the total variance explained from 99.9% to 100.0%. The addition of this component reduces the cross validated MAE by 0.00% from 20.95 to 20.95. This component explains residual variance but does not improve MAE which suggests that this component describes very short term patterns, uncorrelated noise or is an artefact of the model or search procedure.

3 Extrapolation

Summaries of the posterior distribution of the full model are shown in figure 13. The plot on the left displays the mean of the posterior together with pointwise variance. The plot on the right displays three random samples from the posterior.
Below are descriptions of the modelling assumptions associated with each additive component and how they affect the predictive posterior. Plots of the pointwise posterior and samples from the posterior are also presented, showing extrapolations from each component and the cumulative sum of components.

3.1 Component 1 : A linearly increasing function. This function applies until Feb 1974

This component is assumed to stop before the end of the data and will therefore be extrapolated as zero.

3.2 Component 2 : A very smooth monotonically increasing function. This function applies from Feb 1974 onwards

This component is assumed to continue very smoothly but is also assumed to be stationary so its distribution will eventually return to the prior. The prior distribution places mass on smooth functions with a marginal mean of zero and a typical lengthscale of 8.5 years. [This is a placeholder for a description of how quickly the posterior will start to resemble the prior].

Figure 13: Full model posterior with extrapolation. Mean and pointwise variance (left) and three random samples (right)

Figure 14: Posterior of component 1 (top) and cumulative sum of components (bottom) with extrapolation. Mean and pointwise variance (left) and three random samples from the posterior distribution (right).
3.3 Component 3: A smooth function with marginal standard deviation increasing linearly away from Feb 1964. This function applies until Feb 1974

This component is assumed to stop before the end of the data and will therefore be extrapolated as zero.

3.4 Component 4: An exactly periodic function with a period of 1.0 years. This function applies until Feb 1974

This component is assumed to stop before the end of the data and will therefore be extrapolated as zero.
3.5 Component 5: Uncorrelated noise. This function applies until May 1973 and from Oct 1973 onwards

This component assumes the uncorrelated noise will continue indefinitely.

3.6 Component 6: Uncorrelated noise. This function applies from May 1973 until Oct 1973

This component is assumed to stop before the end of the data and will therefore be extrapolated as zero.
Figure 19: Posterior of component 6 (top) and cumulative sum of components (bottom) with extrapolation. Mean and pointwise variance (left) and three random samples from the posterior distribution (right).

4 Model checking

Several posterior predictive checks have been performed to assess how well the model describes the observed data. These tests take the form of comparing statistics evaluated on samples from the prior and posterior distributions for each additive component. The statistics are derived from autocorrelation function (ACF) estimates, periodograms and quantile-quantile (qq) plots.

Table 2 displays cumulative probability and \( p \)-value estimates for these quantities. Cumulative probabilities near 0/1 indicate that the test statistic was lower/higher under the posterior compared to the prior unexpectedly often i.e. they contain the same information as a \( p \)-value for a two-tailed test and they also express if the test statistic was higher or lower than expected. \( p \)-values near 0 indicate that the test statistic was larger in magnitude under the posterior compared to the prior unexpectedly often.

<table>
<thead>
<tr>
<th>#</th>
<th>ACF</th>
<th>Periodogram</th>
<th>QQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min</td>
<td>min loc</td>
<td>max</td>
</tr>
<tr>
<td>1</td>
<td>0.509</td>
<td>0.490</td>
<td>0.676</td>
</tr>
<tr>
<td>2</td>
<td>0.604</td>
<td>0.425</td>
<td>0.636</td>
</tr>
<tr>
<td>3</td>
<td>0.784</td>
<td>0.512</td>
<td>0.599</td>
</tr>
<tr>
<td>4</td>
<td>0.830</td>
<td>0.489</td>
<td>0.379</td>
</tr>
<tr>
<td>5</td>
<td>0.508</td>
<td>0.498</td>
<td>0.508</td>
</tr>
<tr>
<td>6</td>
<td>0.511</td>
<td>0.507</td>
<td>0.502</td>
</tr>
</tbody>
</table>

Table 2: Model checking statistics for each component. Cumulative probabilities for minimum of autocorrelation function (ACF) and its location. Cumulative probabilities for maximum of periodogram and its location. \( p \)-values for maximum and minimum deviations of QQ-plot from straight line.

No statistically significant discrepancies between the data and model have been detected but model checking plots for each component are presented below.

4.1 Model checking plots for components without statistically significant discrepancies

4.1.1 Component 1: A linearly increasing function. This function applies until Feb 1974

No discrepancies between the prior and posterior of this component have been detected
4.1.2 Component 2: A very smooth monotonically increasing function. This function applies from Feb 1974 onwards

No discrepancies between the prior and posterior of this component have been detected

4.1.3 Component 3: A smooth function with marginal standard deviation increasing linearly away from Feb 1964. This function applies until Feb 1974

No discrepancies between the prior and posterior of this component have been detected
Figure 22: ACF (top left), periodogram (top right) and quantile-quantile (bottom left) uncertainty plots. The blue line and shading are the pointwise mean and 90% confidence interval of the plots under the prior distribution for component 3. The green line and green dashed lines are the corresponding quantities under the posterior.

4.1.4 Component 4: An exactly periodic function with a period of 1.0 years. This function applies until Feb 1974

No discrepancies between the prior and posterior of this component have been detected

Figure 23: ACF (top left), periodogram (top right) and quantile-quantile (bottom left) uncertainty plots. The blue line and shading are the pointwise mean and 90% confidence interval of the plots under the prior distribution for component 4. The green line and green dashed lines are the corresponding quantities under the posterior.

4.1.5 Component 5: Uncorrelated noise. This function applies until May 1973 and from Oct 1973 onwards

No discrepancies between the prior and posterior of this component have been detected

No discrepancies between the prior and posterior of this component have been detected.

Figure 25: ACF (top left), periodogram (top right) and quantile-quantile (bottom left) uncertainty plots. The blue line and shading are the pointwise mean and 90% confidence interval of the plots under the prior distribution for component 6. The green line and green dashed lines are the corresponding quantities under the posterior.