Statistical Models for Projecting the Situation

“Prediction is very difficult, especially about the future.”

Unknown

Statistical models are often used to formally characterize underlying relationships in data. In terms of situational awareness, as the chapter title says, statistical models can useful for projecting (i.e., forecasting) the current situation into the future. The goal is to provide a decision maker with an understanding of what the near-term trends are likely to be for a particular situation.

A benefit of statistical modeling, appropriately done, is that such models also provide estimates of forecast uncertainty. This is an important part of modeling that is sometimes overlooked or ignored. Simply put, a forecast without some quantification of the uncertainty inherent in the forecast is not particularly useful. Without the quantification of the uncertainty, a decision maker cannot know how much credence to give the forecast.

This chapter focuses on statistical models useful for the projection step of situational awareness, including time series and regression-based models. As described in Chapter 3, projection is the third and highest level of situational awareness, involving the ability to project the future status of the elements in the environment. Given perception and comprehension of the situation (situational awareness levels 1 and 2), level 3 SA is achieved by using this information in an appropriate model to project likely future states of the environment that are important or useful for decision making.

In addition, the chapter describes how statistical models can also be used for “preprocessing” (also referred to as “preconditioning”) biosurveillance data. The terms preprocessing and preconditioning refer the use of models to account for (as much as possible) the systematic effects present in biosurveillance data. Doing so is important for the appropriate application of many of the early event detection methods discussed in Chapters 6-8.

Footnote 1: Often attributed to Niels Bohr, though variations of the quote have been traced back to other sources early in the 19th century. Variations of the quote have also been attributed to Yogi Berra, Mark Twain, and Samuel Goldwyn.
CHAPTER OBJECTIVES

Upon completion of this chapter, you should be able to:

- Describe how statistical modeling can be used to enhance situational awareness and why forecasting is necessary for achieving situational awareness level 3.
- Explain why it is important to calculate uncertainty bounds for a forecast and calculate the bounds for some simple models.
- Define what it means to preprocess or precondition data and why preprocessing is important for biosurveillance data, particularly with respect to early event detection. Explain how modeling can be used to preprocess biosurveillance data, particularly in terms of eliminating or mitigating autocorrelation.
- Construct simple smoothing, regression, and time series models of biosurveillance data, including
  - Moving average models, simple exponentially weighted moving average models, and double and triple exponential smoothing models;
  - Cross-sectional regression models and autoregressive and adaptive regression models; and.
  - Autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) time series models.
- Describe the change point analysis methodology and how it can be used in combination with early event detection methods to enhance situational awareness.
### MATHEMATICAL NOTATION

- $\beta_0$: Regression intercept parameter
- $\hat{\beta}_0$: Estimated regression intercept
- $\beta_i$: Regression (partial) slope parameter, $i = 1, \ldots, p$
- $\hat{\beta}_i$: Estimated (partial) slope parameter, $i = 1, \ldots, p$
- $\delta$: Holt-Winters model seasonal component parameter; order of the integrated part of an ARIMA model
- $\epsilon$: Linear regression model error term
- $\hat{\epsilon}_i$: Residual for observation $i$
- $\bar{\epsilon}$: Average of the residuals
- $\gamma$: Holt-Winters model trend component parameter
- $i, j$: Indices for either time or the observations
- $I$: Indicator variable
- $k$: Lag
- $\lambda$: Exponentially weighted moving average parameter; Holt-Winters and Brown’s model mean component parameter
- $\text{MAE}(\hat{y})$: Mean absolute error of the forecast $\hat{y}$
- $\text{MAPE}(\hat{y})$: Mean absolute percent error of the forecast $\hat{y}$
- $\text{MSE}(\hat{y})$: Mean square error of the forecast $\hat{y}$
- $n$: Sample size
- $N(0, \sigma^2)$: Normal distribution with mean 0 and variance $\sigma^2$
- $p$: Number of independent variables in a linear regression model or the order of an autoregressive model
- $\phi_i$: $i$th coefficient in an autoregressive model, $i = 1, \ldots, p$
- $q$: Order of a moving average model model
- $r$: Sample correlation
- $r_k$: Sample autocorrelation for lag $k$
- $\sigma_\epsilon$: Standard deviation of the error term $\epsilon$
- $\hat{\sigma}_\epsilon$: Estimated standard deviation of the error term $\epsilon$
- $\hat{\sigma}_i$: Estimated standard deviation of the $i$th period forecast
- $S_t$: Estimated seasonal component at time $t$ in the Holt-Winters model
- $t$: The current or latest time period
- $T_t$: Estimated trend component at time $t$ in the Holt-Winters model
- $\theta_i$: $i$th coefficient in a moving average model, $i = 1, \ldots, q$
- $x$: An “independent variable” in regression
- $y$: An observation; the “dependent variable” in regression
- $\hat{y}_i$: A model’s estimate for the $i$th observation or forecast for time $i$
5.1 Modeling Time Series Data

As discussed in Chapter 2, time series or longitudinal data consist of a sequence of observations, often measured in successive time periods spaced at equal time intervals. Examples in biosurveillance include daily syndrome counts and lab reports; in sentinel physician systems examples include weekly reports of the number of patients presenting with flu symptoms. Longitudinal data, by definition, have a natural temporal ordering and this makes time series analysis distinct from other common modeling approaches for cross-sectional data analysis. Time series analysis is also distinct from spatial data analysis where the observations are typically related by geographical location.

Time series models are generally based on the idea that observations which are close together in time will be more closely related than observations further apart. That is, time series models are based on and try to exploit autocorrelation. In addition, time series models are often prospective, where they use the natural ordering of time to forecast the value in some future period in terms of past values. This is important in prospective monitoring, where the incorporation of information from future observations is clearly impossible.

Models for time series data can have many forms in order to represent different types of real-world processes. When modeling variations in the level of a process, three broad classes of practical importance are autoregressive (AR) models, integrated (I) models, and moving average (MA) models. Combinations of these ideas produce autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models.

5.1.1 Purposes of Modeling

As described in Chapter 2, biosurveillance data often have systematic effects (i.e., explainable trends and patterns). These can include day-of-the-week effects, where patient health-seeking behavior systematically varies according to the day of the week. It may also include seasonal effects where, for example, influenza-like illness is generally higher in the winter months of the year compared to the summer months.

These trends and patterns can be used to build models and the models can then be used both to better understand and characterize historical trends, to assess how the current state compares to historical trends, and to forecast what is likely to occur in the near future. This latter point is important if biosurveillance systems are to facilitate achieving the third level of situational awareness: projection. In statistical terms, projection is called forecasting.

Furthermore, many of the early event detection (EED) methods discussed in Chapters 6-8 are most effective when the systematic components of biosurveillance data are removed. This is best accomplished by first modeling the data, where the model is used to estimate the systematic effects, and then
using the EED methods on the model residuals. The residuals are what remains after the modeled values (often referred to as the “fitted values”) are subtracted from the raw data. This idea of subtracting the systematic effects from biosurveillance data is often referred to as preprocessing.

5.1.2 Examples of Preprocessing

As originally demonstrated in Chapter 4, the clinic GI and ILI data are relatively modestly autocorrelated (see the correlograms in Figures 4.1 and 4.2 and associated discussion on page 90) because of the presence of systematic effects. To remove these effects, the data is preprocessed by fitting a model to the data and then taking the difference between the model estimates and the actual data. That is, for data $y_1, \ldots, y_n$ estimates are calculated via a model: $\hat{y}_1, \ldots, \hat{y}_n$. The residuals are then $\hat{\epsilon}_i = y_i - \hat{y}_i$, $i = 1, \ldots, n$.

Figures 5.1 and 5.2 are correlograms of clinic GI and ILI residuals, where the residuals are calculated as the difference between the lowess mean estimate (first discussed in Section 2.2.2 and shown in Figures 2.4 and 2.5) and the actual observations. Using Equation 4.17 for the residuals, the autocorrelation for lag $k$ is calculated as

$$r_k = \frac{\sum_{i=k+1}^{n} (\hat{\epsilon}_i - \bar{\epsilon})(\hat{\epsilon}_{i-k} - \bar{\epsilon})}{\sum_{i=1}^{n} (\hat{\epsilon}_i - \bar{\epsilon})^2}, k = 1, 2, \ldots, 100, \quad (5.1)$$

where $\bar{\epsilon} = \frac{1}{n} \sum_{i=1}^{n} \hat{\epsilon}_i$.

Compare these correlograms to those calculated from the original data in Figures 4.1 and 4.2. Note how the autocorrelation has been substantially reduced. Thus, for this data it seems that the autocorrelation is mainly a function of the long-term (i.e., annual) cycle in the data. Once removed, the residuals seem to be roughly independent. Furthermore, as shown in Chapter B, the residuals fit a normal distribution well.

In a similar vein, Table 5.1 is the correlation matrix for the hospital GI data residuals, which were calculated in the same way as the clinic data just discussed. Compared to the original data (see Table 4.2 on page 85), Table 5.1 shows that the correlations between hospitals has been reduced substantially (though they were fairly modest to begin with). However, the data do still retain some small positive correlations suggesting that the lowess model has not accounted for all the systematic effects in the data.

5.1.3 An Example of Forecasting

To illustrate the idea of forecasting, consider the clinic GI syndrome data, first depicted in Figure 2.4 with a lowess line overlaid. The lowess line shows some fairly pronounced trends in GI over the year and, as discussed in Chapter 2,
Fig. 5.1. Correlogram for clinic gastrointestinal syndrome residuals for various lags, $1 \leq k \leq 100$. When comparing this to Figure 4.1 it’s apparent that the autocorrelation present in the first 20 days of so of the raw data has been reduced in the residuals.

Fig. 5.2. Correlogram for clinic ILI syndrome residuals for various lags, $1 \leq k \leq 100$. Compared to Figure 4.2, the autocorrelation of the residuals is much less the autocorrelation of the raw data.
5.1 Modeling Time Series Data

Correlation matrix for the GI residuals ($\hat{\epsilon}_{ij}$) from the seven metropolitan hospitals.

<table>
<thead>
<tr>
<th>Correlation (r)</th>
<th>Hospital #1</th>
<th>Hospital #2</th>
<th>Hospital #3</th>
<th>Hospital #4</th>
<th>Hospital #5</th>
<th>Hospital #6</th>
<th>Hospital #7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital #1</td>
<td>1.00</td>
<td>0.06</td>
<td>0.03</td>
<td>0.07</td>
<td>0.07</td>
<td>0.03</td>
<td>0.10</td>
</tr>
<tr>
<td>Hospital #2</td>
<td>1.00</td>
<td>1.00</td>
<td>-0.04</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Hospital #3</td>
<td>1.00</td>
<td>0.02</td>
<td>0.10</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Hospital #4</td>
<td>1.00</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital #5</td>
<td>1.00</td>
<td>0.00</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital #6</td>
<td>1.00</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital #7</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1. Correlation matrix for the GI residuals ($\hat{\epsilon}_{ij}$) from the seven metropolitan hospitals.

the clinic data also has a weekly cycle in which people are more likely to come into the clinic on Thursdays (see Figure 2.6 and associated discussion).

Figure 5.3 shows the results of fitting a “Holt-Winters” forecasting model (to be discussed in Section 5.2.2) to the clinic GI data. The black line in Figure 5.3 is the forecast for each day based on all of the previous days’ data, where the reader will remember that there were 252 days of clinic data (which is the grey line in the figure).

Fig. 5.3. A “Holt-Winters” forecasting model (to be discussed in Section 5.2.2) to the clinic GI data. The black line in Figure 5.3 is the forecast for each day based on all of the previous days’ data.
Note how the forecasting model, like the lowess line in Figure 2.4, helps show the underlying trends in GI because it damps out some of the volatility in the raw data. However, also note that unlike the lowess, the forecasting model can be used to project what the GI trend is likely to be in the future. In Figure 5.3, a 30-day forecast is shown, where the model clearly carries the upward trend that occurred from day 225 to day 252 forward into the future.

A key point is that uncertainty bounds can (and should) be estimated and displayed around a forecast. In Figure 5.3 the uncertainty bounds are shown as the dotted lines extending from day 253 onward. These uncertainty bounds show the range of what is possible, in terms of the actual daily GI counts that might be observed, where note that they get wider the farther one looks into the future. This should seem reasonable and intuitive, since the further one forecasts into the future the more uncertainty there is inherent in the forecast.

In Figure 5.3 the uncertainty bounds get very wide after only about a week into the future. This means that the inherent variation in the daily counts makes forecasting past a week or so into the future highly uncertain and suggests that, at least for this particular data, precise long-term forecasts are not possible. That said, since biosurveillance systems are designed to collect data in near real time, this is not particularly troubling because the forecast can be recalculated each day as new data is received.

The particular Holt-Winters model depicted in Figure 5.3 allows for trends in the data but not cycles. The rest of this chapter will describe a number of different types of models, each with its pros and cons and which may be more or less suited to modeling any specific type of data.

### 5.2 Smoothing Models

The basic assumption behind moving average and other “smoothing” models is that the time series is “locally stationary,” meaning that the process that generates the observed data is roughly constant over a short period of time, though perhaps with a mean that slowly varies over a longer period of time. Under these conditions, it is appropriate to take a moving (i.e., local) average to estimate the current value of the mean, and this value is then used as a short-term forecast.

Chapter 2 introduced one type of smoothing model, the lowess. Because the values estimated by the lowess depend on observations both before and after the time period in question, this type of smoothing model is useful for retrospective modeling but not for prospective modeling. As such, it will not be discussed further here. Instead, this section will focus on moving average models, both simple and exponentially weighted.
5.2 Simple Moving Average Models

The moving average is often called a smoothed version of the original data because averaging has the effect of smoothing out variation in the original time series. As discussed in Chapter 2 (see page 90) the amount of smoothing is adjusted by the size of the “window” of data used in the moving average calculation.

To illustrate the idea of moving average forecasting, consider the hospital #7 GI syndrome data, first depicted using a lowess line in Figure 2.7. That figure shows a fairly regular and pronounced annual cycle in GI and, as discussed in Chapter 2, the hospital data also has a weekly cycle (similar to that shown on page 39 for respiratory syndrome).

Figure 5.4 shows 2-, 7-, and 14-day moving averages superimposed over the raw data. As forecasts, these moving averages differ slightly from the definition given in Equation 4.13. Here, the \( d \)-period moving average forecast for time \( t + 1 \), \( \hat{y}_{t+1} \), is the moving average calculated from the \( d \) days prior to but not including day \( t + 1 \) (since at time \( t \) the results for time \( t + 1 \) have not been observed):

\[
\hat{y}_{t+1} \triangleq \bar{y}_{t+1} = \frac{1}{d} \sum_{i=t-d+1}^{t} y_i.
\]

(5.2)

Fig. 5.4. Two-, 7- and 14-day moving average forecasts of gastrointestinal syndrome daily counts for hospital #7 superimposed over the raw time series data.

Now, in Figure 5.4 it is actually quite difficult to visually discern which of the moving averages best fit the data. This is partially due to the fact that the graph is rather small, but even when enlarged it is difficult to tell “by eye” whether the 2-, 7-, or 14-day moving average is best. For example, see Figure 5.5, which now plots only the last year of data; here the lines for all three moving averages seem to match the data fairly well.
To address this, three metrics are commonly used to judge time series model fit. The first is the mean square error (MSE) of the forecasts, defined as

$$\text{MSE}(\hat{y}) = \frac{1}{n-d} \sum_{t=d+1}^{n} (y_t - \hat{y}_t)^2. \tag{5.3}$$

The MSE of the forecasts is the average squared difference between the observations and their forecasted values where in Equation 5.3 $\hat{y}_t \equiv \bar{y}_t$. Note that the summation in Equation 5.3 is over the entire set of data for which both an observation and its forecast exist. Here it starts at time $t = d + 1$ because the first $d$ periods of data are required to calculate the first moving average forecast (and thus there can be no forecasts for time periods $t = 1, \ldots, d$).

For the hospital #7 data, the MSE for the 2-day moving average is 42.2, for the 7-day moving average it is 32.8, and for the 14-day moving average it is 33.4. In fact, looking over all possible $d$-day moving average forecasts, the MSE is minimized for $d = 7$, where the fact that it is a multiple of the shortest periodicity in the data should not be a surprise (c.f. the discussion on page 90). Also, given that Figures 2.7 and 5.4 show these data have a fairly pronounced seasonal cycle, it it is reasonable that the preferred averaging period is relatively short. By way of comparison, the MSE for hospital #1 is minimized with a 35-day moving average, which is consistent with Figure 2.7 that shows the hospital #1 data maintains a fairly constant level of GI over time with no pronounced cycles.
The other two metrics used to assess time series model fit are the *mean absolute error* (MAE) and the *mean absolute percent error* (MAPE). They are defined as

\[
\text{MAE}(\hat{y}) = \frac{1}{n-d} \sum_{t=d+1}^{n} |y_t - \hat{y}_t|, \tag{5.4}
\]

and

\[
\text{MAPE}(\hat{y}) = \frac{100}{n-d} \sum_{t=d+1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|, \tag{5.5}
\]

where as in Equation 5.3 the summation starts at time \( t = d + 1 \) because the forecast is based on the \( d \)-period moving average. In general, as with the MSE, the MAE and MAPE are calculated over the set of data for which both an observation and its forecast exist. In terms of hospital #7, the MAE and MAPE are consistent with the MSE results and indicate that a 7-day moving average is the preferred model for this data.

Figure 5.6 shows the last year of data and associated 7-day moving average, where it shows the moving average visually follows the data quite well. Figure 5.6 also shows a 30-day forecast into the future. With a moving average model, the forecast is rather simple; all it consists of is the forecast for the next period extended indefinitely into the future (as depicted by the line horizontal in the figure after day 980). More sophisticated models yet to be discussed can incorporate trends, cycles, and other systematic effects into the forecast, but not the moving average, and that is is a significant limitation of this method for forecasting.

Now, as described in the opening to this chapter, it is important to quantify the uncertainty in any forecast. In Figure 5.6 the uncertainty is depicted as bounds around the forecast (denoted by the dotted lines). If the last observed value is at time \( t \), the bounds are calculated as the forecasted value at time \( t + j \) as

\[
\hat{y}_{t+j} \pm 2\hat{\sigma}_{t+j}, j = 1, 2, 3, \ldots, \tag{5.6}
\]

where \( \hat{\sigma}_{t+j} \) is the estimated standard deviation of the observed data around the \( j \)th period ahead forecast. Unfortunately, there is no underlying statistical theory for how to calculate \( \hat{\sigma}_{t+j} \) (Nau, 2012), so it is empirically estimated from the data by calculating

\[
\hat{\sigma}_{t+j} = \sqrt{\frac{1}{n-d-j} \sum_{t=d+j+1}^{n} (y_t - \hat{y}_{t+j})^2}, \tag{5.7}
\]

where \( d = 7 \) in Figure 5.6.

Readers who have taken a statistics course will recognize the bounds as approximate 95 percent confidence intervals at each time period. They are calculated as the \( j \)th period ahead forecast value (which as previously discussed is a constant at value \( \hat{y}_{t+j} = \hat{y}_{t+1}, j = 1, 2, \ldots \) plus or minus twice the estimated standard deviation for the \( j \)th period ahead forecast.
5.2.2 Exponentially Weighted Moving Average Models

An issue with the simple moving average is that the forecast uses the $d$ previous observations with equal emphasis in the calculation and then completely ignores all the observations prior to them. That is, Equation 5.2 can be written as

$$\bar{y}_{t+1} = \frac{1}{d} \sum_{i=1}^{d} y_{t-d+i},$$  \hspace{1cm} (5.8)$$

where the weight $w_i$ is defined as $w_i = 0$ for $i = 1, \ldots, t-d$ and $w_i = 1/d$ for $i = t-d+1, \ldots, t$.

While the choice of $d$ is based on a best fit according to some metric such as the MSE, it seems rather arbitrary to use observation $y_{t-d+1}$ in the calculation with as much weight as the most recent observation $y_t$, particularly if $d$ is large, and it seems just as arbitrary to completely ignore observation $y_{t-d}$ which immediately precedes observation $y_{t-d+1}$.

**Simple Exponentially Weighted Moving Average Models**

An alternative is the *simple exponentially weighted moving average* model which gives successively less weight to observations further back in time according to a user-defined parameter $\lambda$, $0 < \lambda \leq 1$. The exponentially weighted moving average forecast is calculated as
\[ \hat{y}_{t+1} \triangleq \bar{y}_{t+1} = \lambda y_t + \lambda (1 - \lambda) y_{t-1} + \lambda (1 - \lambda)^2 y_{t-2} + \cdots \] (5.9)

\[ = \sum_{i=1}^{t} \lambda (1 - \lambda)^{i-1} y_{t-i+1}, \]

where now the weight for observation \( t - i + 1 \) is \( w_i = \lambda (1 - \lambda)^{i-1} \). The forecast \( \hat{y}_{t+1} \) is still a weighted average of past observations; however, the more recent the observation the greater its weight. Note that smaller values of \( \lambda \) put relatively more emphasis on past observations, while larger values put more emphasis on recent observations. For \( \lambda = 1 \) the exponentially weighted moving average forecast reduces to just the last observation: \( \hat{y}_{t+1} = y_t \).

Equation 5.9 can be reduced to a simpler recursive equation in which the next forecast value is the weighted average of the current observation and the current forecast:

\[ \bar{y}_{t+1} = \lambda y_t + (1 - \lambda) \hat{y}_t. \] (5.10)

Thus, the forecast for the next time period can be calculated with just the value of the current observation \( y_t \) and its forecast \( \hat{y}_t \). To use Equation 5.10, an initial forecast value \( \hat{y}_1 \) must be defined. A reasonable choice is \( \hat{y}_1 = y_1 \) where, because of the exponential smoothing, the effect of this initial choice diminishes rapidly.

The obvious question is how to choose the smoothing parameter \( \lambda \). Generally speaking, for processes with shifts or short-run trends, larger values of \( \lambda \) are to be preferred as they put more weight on recent observations and thus make the forecasts more sensitive to shifts. On the other hand, smaller values of \( \lambda \) are better for stationary or nearly-stationary data – meaning data that come from a process that has few and small trends, cycles or other changes, if any – because the more stationary the process the more useful is older data.

The point is that the appropriate choice for \( \lambda \) depends on the specific data being modeled. Thus, a preferred method for choosing \( \lambda \) is by finding the value which minimizes the MSE (or MAE or MAPE, as preferred) on a set of historical data equivalent to what is to be expected in the future. That’s essentially an empirical exercise, much like in the previous section with the selection of \( d \).

As in the simple moving average, if the last observed value is at time \( t \), the uncertainty bounds for time \( t + j, j = 1, 2, 3, \ldots \), are calculated as

\[ \bar{y}_{t+j} \pm 2\hat{\sigma}_{y_{t+j}} \] (5.11)

where

\[ \hat{\sigma}_{y_{t+j}} = \hat{\sigma}_{y_t} \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2j}]}, \] (5.12)

and where, as with the simple moving average, \( \bar{y}_{t+j} = \bar{y}_{t+1} \) for \( j = 2, 3, 4, \ldots \). What is required, then, is an estimate of the standard deviation of the observations at time \( t \), \( \hat{\sigma}_{y_t} \), which is calculated in the usual way (i.e., Equation 4.6 on page 81) from a recent window of the data, \( y_{t-d}, y_{t-d+1}, \ldots, y_t \).
Example 5.1. Calculate the exponentially weighted moving average forecasts for times $t = 1, \ldots, 11$, using $\lambda = 0.3$, for the following data:

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t$</td>
<td>7</td>
<td>13</td>
<td>7</td>
<td>12</td>
<td>11</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

Solution: Table 5.2 gives the forecasts, calculated as $\hat{y}_{t+1} = 0.3 \times y_t + 0.7 \times \bar{y}_t$. The forecast for time $t = 11$ is $\hat{y}_{11} = 9.7$. Given that the data was randomly generated from a Poisson distribution with mean $\mu = 10$, the forecast at time 11 is actually quite close to the true value.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$y_t$</th>
<th>$\hat{y}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>7.00</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>8.80</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>8.26</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>9.38</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>9.87</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9.61</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>9.13</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>9.09</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>7.86</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>9.70</td>
</tr>
</tbody>
</table>

Table 5.2. Forecasts for Example 5.1, where $\hat{y}_{t+1} = 0.3 \times y_t + 0.7 \times \bar{y}_t$.

Example 5.2. Returning to the Hospital #7 GI data, calculate the exponentially weighted moving average using $\lambda = 0.1, 0.3, \text{and} 0.9$. Plot the moving averages over the data to visually show how they fit to the data. Then determine the value of $\lambda$ that minimizes the MSE. Compare and contrast the results to the 7-day moving average forecast shown in Figure 5.6.

Solution: Figure 5.7 is a plot of the exponentially weighted moving averages for $\lambda = 0.1, 0.3, \text{and} 0.9$ versus the data. As with the moving averages in Figure 5.4, it is difficult to visually discern which of the forecasts best fit the data.
However, looking carefully at Figure 5.7 it does seem that while the forecasts for $\lambda = 0.9$ match quite closely to the prior period, they are often quite a bit off from the data of the forecasted period. For data such as this, where there is quite a bit of volatility in the GI count from day to day, this suggests that smaller values of $\lambda$ are likely to perform better for forecasting. That said, for $\lambda = 0.1$ it also seems like the forecasts during the spike in GI around day 900 seem to lag behind the actual data. This suggests that values for $\lambda$ that are excessively small may result in degraded forecasts as well.

![Fig. 5.7. A plot of the exponentially weighted moving averages for $\lambda = 0.1$, 0.3, and 0.9 versus the hospital #7 GI data.](image)

Figure 5.8 is a plot of MSE versus $\lambda$, for $0.01 \leq \lambda \leq 0.99$. It shows that the MSE is minimized at $\lambda = 0.16$ with a value of 32.17. This is slightly less than the MSE for the 7-day moving average (32.8). This suggests the exponentially weighted moving average is a slightly better model for this data than a simple moving average. However, it is important to remember that neither of these models can incorporate trends or cycles into their forecasts, which is a significant impediment to their application to biosurveillance.

Figure 5.9 is the final exponentially weighted moving average model with $\lambda = 0.16$ showing a 90-day forecast and uncertainty bounds.
**Fig. 5.8.** A plot of MSE versus $\lambda$ for an exponentially weighted moving average model of hospital #7 GI data. MSE is minimized at $\lambda = 0.16$, suggesting this is the preferred choice of the smoothing parameter for this data.

**Fig. 5.9.** A plot of the exponentially weighted moving averages for $\lambda = 0.16$ for the hospital #7 GI data with a 90-day forecast and uncertainty bounds. The MSE for this model is 32.206.
More Complex Exponentially Weighted Moving Average Models

For biosurveillance data with trends, a better alternative is double exponential smoothing, also referred to as Brown linear exponential smoothing (Brown, 1963, 1959), which not only takes a weighted average of the most recent observation and the most recent forecast, but it also takes a weighted average of the most recent change and the most recent forecast change. As such, double exponential smoothing can incorporate trends in the data into its forecast.

The model is as follows. First, using exponential smoothing for a chosen $\lambda$ value, at time $t$ calculate the smoothed mean $\bar{y}_t$ and trend $T_t$ for the data:

$$\bar{y}_t = \lambda y_t + (1 - \lambda)\bar{y}_{t-1},$$
$$T_t = \lambda(\bar{y}_t - \bar{y}_{t-1}) + (1 - \lambda)T_{t-1}.$$ (5.13)

Then the forecast at time $t + 1$ is the sum of the smoothed mean and the estimated trend at time $t$: $\hat{y}_{t+1} = \bar{y}_t + T_t$. A generalization to Brown’s method is due to Holt (1957) and Winters (1960), which allows separate smoothing parameters for the mean and the trend:

$$\bar{y}_t = \lambda y_t + (1 - \lambda)\bar{y}_{t-1},$$
$$T_t = \gamma(\bar{y}_t - \bar{y}_{t-1}) + (1 - \gamma)T_{t-1},$$ (5.14)

where as before $0 < \lambda \leq 1$ and similarly $0 < \gamma \leq 1$.

Of course, some biosurveillance data also have periodic (for example, weekly or seasonal) cycles, for which the Holt-Winters triple exponential smoothing model may be appropriate. The Holt-Winters model, as the name suggests, uses three weighted averages: the most recent observation and forecast; the most recent change and forecast change; and, the most recent cyclical deviation and forecast cycle value.

The triple exponential smoothing model is:

$$\bar{y}_t = \lambda y_t + (1 - \lambda)\bar{y}_{t-1},$$
$$T_t = \gamma(\bar{y}_t - \bar{y}_{t-1}) + (1 - \gamma)T_{t-1},$$
$$S_t = \delta(y_t - \bar{y}_t) + (1 - \delta)S_{t-s},$$ (5.15)

where, as with $\lambda$ and $\gamma$, $0 < \delta \leq 1$. The user chooses $s$, which is the periodicity over which the cyclical term is calculated, and the calculations must keep track of $s S_{t-s}$ terms, one for each time period. The forecast at time $t + 1$ is then the sum of the smoothed mean, the estimated trend, and the cyclical component at time $t$: $\hat{y}_{t+1} = \bar{y}_t + T_t + S_t$.

In addition to the foregoing additive models, there are also multiplicative models and a host of variants. These models get increasingly more challenging to fit where, for example, in the triple exponential smoothing model one must choose values for three parameters according to some criteria (such as minimize the MSE) as well as the starting values for $\bar{y}_1$, $T_1$, and $S$. Similarly,
the calculations for the uncertainty bounds of the forecasts for some number of periods into the future become increasingly complicated. These details are important, but also beyond the level of this text.

Suffice it to say that many of these methods have been incorporated into various statistical software packages, which often mitigates the need to understand the specific computational details. For those readers interested in additional detail, see the Additional Reading section at the end of this chapter. What is most important is that the appropriate model is fit to the data that gives both appropriate forecasts, meaning forecasts that incorporate both trends and cycles as necessary, and that provide correct uncertainty bounds for the forecasts.

\[\text{Example 5.3. Compare and contrast the exponentially weighted moving average model of the previous example with Holt-Winters double and triple exponential smoothing applied to the hospital #7 GI data. Use the models to generate 90-day forecasts and associated uncertainty bounds. Which one of the models is most appropriate for this data?}\]

\[\text{Solution: Figures 5.10, 5.11, and 5.12 plot the results from fitting a Holt double exponential model, Holt-Winters triple exponential model with } s = 7, \text{ and Holt-Winters triple exponential model with } s = 365, \text{ respectively. For the triple exponential models, } s = 7 \text{ and } s = 365 \text{ were chosen since there is some evidence of a weekly cycle in the data and it’s also reasonable to assume that an annual cycle may be present. For comparison, Figure 5.9 plots the results from fitting a exponentially weighted moving average model.}\]

\[\text{Visually, all four models have pros and cons. The simple exponential smoothing model in Figure 5.9 has the desirable feature of being the most simple and visually looks like the forecast is reasonable when compared to the past. On the other hand, during the outbreak around day 900 this model would not be able to incorporate the upward and downward trends. In comparison, the double exponential model in Figure 5.10 picks up the trend in the data, though clearly extending the trend out 90 days is unrealistic.}\]

\[\text{Now, the triple exponential model in Figure 5.11 with } s = 7 \text{ can model both trends in the data and a weekly cycle, and the latter is evident in the prediction (and incorporated into the uncertainty bounds). However, the forecast weekly cycle seems to strongly reflect the cycle in the last few weeks of the data, which does not seem to be as visually present in the earlier part of the data. Finally, the triple exponential model in Figure 5.12 with } s = 365 \text{ can model both trends in the data and a annual cycles in the data. Here the forecast is much noisier than any of the other models, though in some ways it}\]

\[\text{\footnote{These models were fit using the R statistical software package using the } \textbf{HoltWinters} \text{ function to fit the models and the } \textbf{forecast} \text{ library to calculate the forecasts and associated uncertainty bounds.}}\]
Fig. 5.10. A plot of the Holt double exponential model with $\lambda = 0.2$ and $\gamma = 0.06$ for the hospital #7 GI data with a 90-day forecast and uncertainty bounds.

Fig. 5.11. A plot of Holt-Winters triple exponential model with $\lambda = 0.16$, $\gamma = 0.003$, $\delta = 0.05$, and $s = 7$ for the hospital #7 GI data with a 90-day forecast and uncertainty bounds.
Fig. 5.12. A plot of Holt-Winters triple exponential model with $\lambda = 0.14$, $\gamma = 0.0$, $\delta = 0.63$, and $s = 365$ for the hospital #7 GI data with a 90-day forecast and uncertainty bounds.

looks much more realistic. However, it’s not clear visually whether this model is or is not modeling the data particularly well.

To quantitatively compare the model fits, Table 5.3 gives the MSE, MAE, and MAPE for each. The table shows that the double exponential model fits least well, followed by the triple exponential model with $s = 7$. In terms of the simple exponential model and the triple exponential model with $s = 365$, the simple exponential model has the smallest MSE, while the triple exponential model has the smallest MAE and MAPE. The former suggests that the triple exponential model can result in large deviations from the actual counts, which when squared results in an inflated MSE, but that the deviations on average are smaller than those with the simple exponential model.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\lambda$</th>
<th>$\gamma$</th>
<th>$\delta$</th>
<th>$s$</th>
<th>MSE</th>
<th>MAE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Exponential</td>
<td>0.16</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>32.206</td>
<td>4.163</td>
<td>0.314</td>
</tr>
<tr>
<td>Double Exponential</td>
<td>0.2</td>
<td>0.06</td>
<td>–</td>
<td>–</td>
<td>34.267</td>
<td>4.364</td>
<td>0.322</td>
</tr>
<tr>
<td>Triple Exponential</td>
<td>0.16</td>
<td>0.003</td>
<td>0.05</td>
<td>7</td>
<td>33.035</td>
<td>4.219</td>
<td>0.311</td>
</tr>
<tr>
<td>Triple Exponential</td>
<td>0.14</td>
<td>0.0</td>
<td>0.63</td>
<td>365</td>
<td>32.995</td>
<td>3.820</td>
<td>0.249</td>
</tr>
</tbody>
</table>

Table 5.3. MSE, MAE, and MAPE for exponential models fit to hospital #7 data.
It is interesting to note that the two triple exponential models achieve essentially the same MSE, which suggests that there is benefit to incorporating both a weekly and an annual cycle into a forecast model. Unfortunately, the standard Holt-Winters triple exponential model, as defined in Equation 5.15, can only explicitly incorporate the period of one cycle.

5.3 Regression-based Models

Linear regression is a statistical methodology typically used for modeling the relationship between a “dependent variable” (usually denoted as \(y\)) and one or more “independent” or explanatory variables (usually denoted as \(x_1, \ldots, x_p, p \geq 1\)). With only one independent variable, the model is often referred to as simple linear regression; when there is more than one independent variable it is referred to as multiple regression.

Linear regression is typically applied to cross-sectional data, where the dependent variable is modeled as a linear function of the independent variables. The unknown model parameters (intercept and slope or partial slopes) are estimated from the data. This section begins with an explanation of linear regression in the standard cross-sectional data context and then proceeds to describe approaches to applying linear regression to time series data and particularly biosurveillance data.

5.3.1 Linear Regression

A simple linear regression model models a dependent variable as a linear function of a single independent variable. Mathematically, the model is

\[
y = \beta_0 + \beta_1 x + \epsilon, \quad (5.16)
\]

where \(\beta_0\) and \(\beta_1\) are the true (and unobserved) intercept and slope, and \(\epsilon\) is the error term which can be interpreted as random noise. The standard assumptions for a linear regression model are:

- There is a linear relationship between \(y\) and \(x\).
- The error term is a random variable from a symmetric distribution with mean zero and constant variance. Often it is assumed that the error term is normally distributed, and then the assumption can be expressed as \(\epsilon \sim N(0, \sigma^2_{\epsilon})\).

Note that from these assumptions it follows that for a given value of \(x\) the dependent variable has a normal distribution with mean \(\beta_0 + \beta_1 x\) and variance
\( \sigma^2 \). Often this is written as \( Y \sim N(\beta_0 + \beta_1 x, \sigma^2) \). The key take-away of this result is that the dependent variable must be continuous for linear regression modeling to be appropriate.

When fitting a linear regression model, the slope and intercept are estimated from data, \( \{(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)\} \), so that they minimize the mean square error. This is often referred to as least squares or ordinary least squares. Denoting the estimated intercept and slope as \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \), the fitted model is

\[
\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i = 1, \ldots, n,
\]

and the chosen values of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) minimize the mean square error of the residuals

\[
\text{MSE}(\hat{\varepsilon}) = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left( y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right)^2,
\]

(5.18)

where the residuals are defined as \( \varepsilon_i = y_i - \hat{y}_i, \quad i = 1, \ldots, n \).

The MSE of the residuals is fundamentally different from the MSE of the forecasts first presented in Equation 5.3. With the MSE of the residuals, the fitted values \( \hat{y}_1, \ldots, \hat{y}_n \) are calculated based on a single regression model fit to a single set of data, and each of the fitted values correspond to an observed value within the sample of data. For the MSE of the forecast, the \( \hat{y}_{t+1} \) values are forecasts based on prior data and each of the forecasted values may be based on fitting separate models to different sets of prior data. Hence, the MSE of the residuals is a within-sample measure of how well a single model fits a single set of data while the MSE of the forecasts is an out-of-sample measure of how well a forecasting methodology predicts future values.

Given that the two MSEs are different, it should not be a surprise that the 2 in the denominator in Equation 5.18 is not equivalent to the \( d \) in Equation 5.3. The reason for subtracting 2 from \( n \) in Equation 5.18 is so the MSE is an unbiased estimate of \( \sigma^2 \).

Figure 5.13 illustrates a regression line fit to a simulated set of data, where in the simulation \( \beta_0 = 10 \) and \( \beta_1 = 2 \). For the \( n = 100 \) simulated paired observations of dependent and independent variables shown, the fitted regression line has an estimated intercept \( \hat{\beta}_0 \) = 12.4 and estimated slope \( \hat{\beta}_1 \) = 1.9. Also shown are two other lines for visual comparison.

The fitted regression line differs from the true line because of randomness in the data, which occurs because of the error term in Equation 5.16. For the simulated data, \( \epsilon \sim N(0, \sqrt{5}) \). The randomness arising from the error term is evident in Figure 5.13 because the way data is distributed around the line.

As previously discussed, the fitted regression line achieves the smallest MSE for this set of data, even when compared to the MSE for the true line. It is literally the line that ‘best fits’ the data according to the criteria of minimizing the mean square error.
Multiple linear regression is similar to simple linear regression but with more independent variables. For \( p \) independent variables, the model is

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon
\]

\[
= \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \epsilon,
\]

where the \( \beta_i, i = 1, \ldots, p \), are referred to now as partial slopes. The same assumptions from simple linear regression apply, with the additional assumption that the independent variables (and linear functions of subsets of the independent variables) are independent, or nearly so. As with simple linear
regression, the values for the intercept and partial slopes are estimated from the data so that they minimize the MSE, where the MSE is defined as

$$\text{MSE}(\hat{y}) = \frac{1}{n-p-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$  \hspace{1cm} (5.20)

and where

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \hat{\beta}_2 x_{i,2} + \cdots + \hat{\beta}_p x_{i,p}.$$  \hspace{1cm} (5.21)

This definition for MSE is actually the general expression for regression, while that given in Equation 5.18 is the specific form for simple linear regression. That is, in simple linear regression \( p = 1 \) and thus \( n - p - 1 \) in Equation 5.20 becomes \( n - 2 \) for simple linear regression. As in Equation 5.18, the reason for subtracting \( p + 1 \) from \( n \) in the denominator is to that the MSE is an unbiased estimator of \( \sigma^2_\epsilon \).

Multiple regression allows for fitting more complicated models, both in the sense of adding more independent variables, but also in terms of adding quadratic and higher order terms, which then allows for a nonlinear relationship between the dependent variable and the independent variable or variables. For example, consider the case of just one independent variable \( x \), but for which there is a quadratic relationship between \( y \) and \( x \). This can be modeled as

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$  \hspace{1cm} (5.22)

Now, while the dependent variable must be continuous, independent variables can be either continuous or discrete. For example, consider a multiple regression model where \( y \) is patient weight and the goal is to model weight as a function of age and gender. Denote age, a continuous variable, by \( x \), and denote gender, a binary variable, by \( I \). The use of the \( I \) notation is to emphasize that this is an indicator variable, where \( I = 0 \) means the patient is female and \( I = 1 \) means the patient is male. The model, then, is

$$y = \beta_0 + \beta_1 x + \beta_2 I + \epsilon.$$  \hspace{1cm} (5.23)

The resulting model assumes the lines fitted for men and women have the same slopes but different intercepts. That is, with a little algebra Equation 5.23 reduces to

$$y = \begin{cases} 
\beta_0 + \beta_1 x + \epsilon, & \text{for female patients (i.e., } I = 0) \\
(\beta_0 + \beta_2) + \beta_1 x + \epsilon, & \text{for male patients (i.e., } I = 1),
\end{cases}$$  \hspace{1cm} (5.24)

where the slopes for both males and females is \( \beta_1 \), while the intercept is \( \beta_0 \) for females and \( \beta_0 + \beta_2 \) for males.

For a categorical variable with \( m \) levels, \( m - 1 \) indicator variables are required in the linear regression model. Further, if \( m \) indicator variables are used, a regression model cannot be fit as the model violates the assumption
of independence (because membership in any category is completely defined by knowledge of the other \(m - 1\) categories). Thus, for example, 6 indicator variables can be used in a regression model to account for day-of-the-week effects in biosurveillance data.

Now, the assumption of equal slope for males and females in Equations 5.23 and 5.24 may not be realistic. Different slopes can be accommodated with interaction terms. For example, consider the model

\[
y = \beta_0 + \beta_1 x + \beta_2 I + \beta_3 xI + \epsilon, \tag{5.25}
\]

where \(x\) and \(I\) are defined as before. The model can be restated as

\[
y = \begin{cases} 
\beta_0 + \beta_1 x_1 + \epsilon, & \text{for female patients} \\
(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_2 + \epsilon, & \text{for male patients},
\end{cases} \tag{5.26}
\]

where the intercepts by gender are the same as before, and so is the slope of the line for females, but the slope for males is now \(\beta_1 + \beta_3\).

Depending on the complexity of the data and the phenomenon being modeled, even more complicated models can be fit. For example, quadratic terms could be added to Equation 5.25 to allow for nonlinearities. However, all of the regression models described thus far are cross-sectional in nature and thus of limited use for biosurveillance forecasting. That said, as the next section shows, regression can be adapted for use with time series data.

### 5.3.2 Autoregression

An autoregressive (AR) model is similar to the regression models of the previous section, but they are applied to time series data and use previous values of the dependent variable for predicting the current value of the dependent variable. An autoregression model of order \(p\), denoted AR(\(p\)), is

\[
y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \epsilon, \tag{5.27}
\]

where the term order refers to the maximum time lag used in the model. The simplest model is an AR(1), where the current value of \(y\) only depends on the last value:

\[
y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon. \tag{5.28}
\]

As with the exponential models, autoregressive models can be challenging to fit where, for example, the choice of \(p\) requires fitting multiple models, and the appropriate calculations for the uncertainty bounds of the forecasts can be quite complicated. These details are important but, as with the exponential models, the use of a good statistical software package will take care of the computational intricacies.
Example 5.4. Fit an autoregressive model to the hospital #7 GI data, determining the order and coefficients that result in a ‘best fit.’ Compare and contrast this model to the exponential models in Example 5.3. Is the autoregressive model more or less appropriate for this data?

Solution: Figure 5.14 shows the best fit autoregressive model, for order $p = 8$, which is

$$y_t = \beta_0 + 0.178y_{t-1} + 0.099y_{t-2} + 0.067y_{t-3} + 0.108y_{t-4} +$$
$$0.090y_{t-5} + 0.153y_{t-6} + 0.085y_{t-7} + 0.081y_{t-8}.$$

![Figure 5.14](image-url)  

**Fig. 5.14.** A plot of the autoregression model results with $p = 8$ for the hospital #7 GI data with a 90-day forecast and uncertainty bounds.

Note how the forecast quickly smooths out once it has extended just a couple of multiples of the order into the future. Also note that the figure shows an interesting pattern in the uncertainty bounds that occurs about every 30 days or so. Why this is occurring is not clear, but it is also not particularly important since, for biosurveillance, forecasts of a week or two into the future are of primary interest, and the forecasts will be refit each day anyway.

---

3 The model was fit using the `ar` function in R.
For the autoregressive model, MSE = 31.753, MAE = 4.179, and MAPE = 0.314. Comparing these to those from the exponential models in Table 5.3, the autoregressive model metrics are equal to or less than the minimum values across all of the exponential models. This suggests the autoregressive model is a better model to use with the hospital #7 data.

5.3.3 Adaptive Regression

Burkom et al. (2006) first proposed the adaptive regression model with sliding baseline to model the systematic component of the syndromic surveillance data. The idea is as follows. At time $t$, fit a linear regression of the observations from the past $n$ days, $\{y_t, y_{t-1}, \ldots, y_{t-n+1}\}$, on time relative to the current period as represented by the set of integers $\{n, n-1, \ldots, 1\}$. That is, the data for the regression are $\{(y_t, n), (y_{t-1}, n-1), \ldots, (y_{t-n+1}, 1)\}$, where the integers are the “independent variables” and where it is assumed that $t \geq n$.

The model is for day $i$, $t \geq i \geq t - n + 1$

$$y_i = \beta_0 + \beta_1 \times (i - t + n) + \epsilon,$$

(5.29)

where the model if fit using ordinary least squares and, as with other linear regression models, $\epsilon$ is the error term which is assumed to follow a symmetric distribution with mean 0 and standard deviation $\sigma_\epsilon$.

The “with sliding baseline” in the name means to repeat this process each day, always using the most recent $n$ observations as the sliding baseline in the regression to calculate the forecast for time $t + 1$. That is, having fit the regression at time $t$, use the slope and intercept from the data, $\hat{\beta}_0(t)$ and $\hat{\beta}_1(t)$ – where the $t$ in parentheses is intended to stress that the slope and intercept will be re-estimated at each time period – to forecast the observation for the next time period as

$$\hat{y}_{t+1} = \hat{\beta}_0(t) + (n + 1) \times \hat{\beta}_1(t).$$

(5.30)

When using regression to predict future observations, the question naturally arises as to how much historical data should be used for the regression’s sliding baseline. Burkom et al. (2006) recommended an 8-week sliding baseline ($n = 56$ for a 7-day week). In a simulation study Fricker et al. (2008a) found that the optimal $n$ varied depending on the data, ranging from $n = 15$ to $n = 56$ time periods, where the optimal choice was based on minimizing the MSE of the forecast.

All other factors being equal, regressions based on a shorter sliding baseline will less accurately model the underlying systematic effects in the data than those based on longer sliding baselines. However, although a longer sliding baseline theoretically should allow for fitting a more detailed regression
model, the longer sliding baseline will also make the resulting forecasts less responsive to quick changes in disease incidence. Further, in syndromic surveillance the amount of available data may be limited and older data may be of less relevance due to changing trends or phenomena. Hence, there is a trade-off to be made between the amount of historical data used in a particular model and the predictive accuracy of that model.

As shown in Fricker et al. (2008a), the standard deviation of the forecast error at time $t + 1$ for a simple linear adaptive regression is

$$
\sigma_{t+1} = \sigma_\epsilon \sqrt{\frac{(n+2)(n+1)}{n(n-1)}}.
$$

(5.31)

where $\sigma_\epsilon$ is the standard deviation of $\epsilon$ in Equation 5.30 and $\sigma_{t+1}$ is the standard deviation of the prediction error at time $t+1$. Since $\sigma_\epsilon$ will not be known, it is estimated from the data as the square root of the MSE (often referred to as the root mean square error or RMSE) in Equation 5.20, and then used in Equation 5.33 to estimate $\sigma_{t+1}$.

Figures 5.15 and 5.16 illustrate the application of the adaptive regression method to the Hospital #7 GI data. Specifically, Figure 5.15 shows the adaptive regression line with $n = 56$ fit to the daily counts for days 924 to 979. The dark dot is the forecast for day 980 ($\hat{y}_{980} = 10.78$), the white dot is the actual value for day 980 ($y_{980} = 19$), and the dotted line denotes the uncertainty bounds around $\hat{y}_{980}$.

To calculate the uncertainty bounds, $\sigma_\epsilon$ is estimated from the data (using Equation 5.20) as $\hat{\sigma}_\epsilon = 5.14$. Then the uncertainty bounds for $\hat{y}_{980}$, calculated as a 95% prediction interval, is

$$
\hat{y}_{980} \pm 1.96\sigma_{980} = \hat{y}_{980} \pm 1.96 \times 5.14 \sqrt{\frac{(58)(57)}{(56)(55)}} = 10.78 \pm 10.44.
$$

As Figure 5.15 shows, the actual observed count for day 980 falls within the uncertainty bounds.

Figure 5.16 illustrates how the adaptive regression adjusts to the sliding baseline of the most recent $n$ observations. In the figure, the lines are the adaptive regressions for the last month of Hospital #7 daily GI syndrome counts and the dots are their associated predicted values for the next day, where each regression line is is based on the previous $n = 56$ observations. What the figure shows is how regression lines change over time, conforming to ‘best fit’ the sliding baseline of data. As a result, the predictions follow the trends in the data.

As with other regression models, Equation 5.30 can also be adapted to allow for nonlinearities and other effects by adding additional terms into the equation. For example, a model that also includes day-of-the-week effects (for a 7-day week) is
5.3 Regression-based Models

Fig. 5.15. An adaptive regression line fit to Hospital #7 daily GI syndrome counts for days 924 to 979. The dark dot is the predicted value, the dotted line shows the uncertainty bounds for the predicted value, and the white dot is the actual value for day 980.

Fig. 5.16. Adaptive regression lines used to predict the last month of Hospital #7 daily GI syndrome counts.
150 Statistical Models for Projecting the Situation

\[ y_t = \beta_0 + \beta_1 \times (i - t + n) + \beta_2 I_{\text{Mon}} + \beta_3 I_{\text{Tues}} \]
\[ + \beta_4 I_{\text{Wed}} + \beta_5 I_{\text{Thurs}} + \beta_6 I_{\text{Fri}} + \beta_7 I_{\text{Sat}} + \epsilon, \]  \hspace{1cm} (5.32)

where the Is are indicator variables – \( I = 1 \) on the relevant day of the week and \( I = 0 \) otherwise – and \( \epsilon \) is the error term which is assumed to follow a symmetric distribution with mean 0 and standard deviation \( \sigma_\epsilon \).

It is important to note that the appropriate calculations for the standard deviation of the forecast error at time \( t + 1 \) depend on the form of the regression, where for the model in Equation 5.32 Fricker et al. (2008a) showed

\[ \sigma_{t+1} = \sigma_\epsilon \sqrt{\frac{(n + 7)(n - 4)}{n(n - 7)}}. \]  \hspace{1cm} (5.33)

Also, while the data may contain nonlinearities, a linear model may still provide good predictions. See Fricker et al. (2008a) for additional details.

**Example 5.5.** For the hospital #7 GI syndromic surveillance data, determine the optimal value of \( n \), where “optimal” means minimizing the MSE of the forecast.

**Solution:** Figure 5.17 shows the forecasts from an simple linear adaptive regression model with a sliding baseline of size \( n = 56 \). The grey line shows the actual data and the black line shows the forecasts, and visually it looks like the forecasts are a noticeable delay in reflecting the trends in the data. In fact, the forecast MSE is 36.903, which is substantially worse than the exponentially weighted moving average and autoregressive models of Examples 5.3 and 5.4.

Figure 5.18 shows the forecast values for the optimal simple linear adaptive regression model which, for the hospital #7 GI data, is \( n = 30 \). Using a sliding baseline of size \( n = 30 \) results in a forecast MSE of 34.919, which is a substantial improvement and a visual comparison of the two figures shows the fit is better. However, for this data it seems adaptive regression does not perform nearly as well as either autoregressive or exponential smoothing models. This result is consistent with Burkom et al. (2006).

---

### 5.4 ARMA and ARIMA Models

ARMA and ARIMA models are a general class of models useful for forecasting many different types of time series data. The acronyms stand for autoregressive
**Fig. 5.17.** Forecasts from an simple linear adaptive regression model with a sliding baseline of size $n = 56$. The grey line shows the actual data and the black line line shows the forecasts. Visually the forecasts have a noticeable delay in reflecting the trends in the data.

**Fig. 5.18.** Forecasts from an simple linear adaptive regression model with a sliding baseline of size $n = 30$. The grey line shows the actual data and the black line line shows the forecasts. Compared to Figure 5.17, the forecasts more closely track with the actual data.
moving average and autoregressive integrated moving average models. In addition to the smoothing and regression-based models of the previous sections, ARMA and ARIMA models are additional ways to model autocorrelated time series data.

5.4.1 Autoregressive Moving Average (ARMA) Models

Given time series data \( \{y_1, y_2, \dotsc, y_t\} \), an ARMA\((p,q)\) model consists of two parts, an autoregressive (AR) part and a moving average (MA) part, where \( p \) is the order of the autoregressive part of the model and \( q \) is the order of the moving average part.

The autoregressive part of the model is just as described in Section 5.3.2. However, in this section, so that the notation is consistent with conventions in the literature, the coefficients for an AR\((p)\) model will be denoted by \( \phi_0, \phi_1, \dotsc, \phi_p \).

On the other hand, the moving average part of the model is not the same as the moving average models presented in Section 5.2.1. Rather, in the context of ARMA models (and ARIMA models in the next section), a first order moving average model, MA\((1)\), is

\[
\hat{y}_t = \theta_0 + \epsilon_t - \theta_1 \epsilon_{t-1},
\]

where the \( \epsilon \)'s are assumed to have a normal distribution with mean zero and constant variance. For a \( q \)th order moving average model, MA\((q)\), the model is

\[
\hat{y}_t = \theta_0 + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}.
\]

A moving average model is similar to a linear regression of the current value of the series against the \( q \) previous error terms. The distinction in the MA\((q)\) model is that the error terms are propagated forward into future values of the time series. Fitting an MA model is more complicated than the AR models because the error terms are unobserved and the MA models are also harder to interpret compared AR models. As with some of the previous modeling approaches, the details of model fitting will be left to a good statistical software package and interpretation of the model is of less interest than accurate forecasts.

For the ARMA\((p,q)\) model, the autoregressive and moving average components are combined into one model:

\[
\hat{y}_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}.
\]

Quite complex models can result if the \( p \) and \( q \) are large. A standard strategy is to try to identify parsimonious models, meaning models with small \( p \) and \( q \), that fits the data well.
Example 5.6. Fit an ARMA model to the hospital #7 GI data, determining the $p$ and $q$ that minimize the forecast MSE. Compare and contrast this model to the previous models. Is the ARMA model to be preferred for this data?

Solution: The optimal model is ARMA(8,6)\(^4\) which achieves a forecast MSE 30.516. Thus the addition of the moving average portion of the ARMA model improves the fit over the pure AR model of Example 5.4 and is better than the exponentially weighted moving average models fit in Example 5.3. Hence it the ARMA(8,6) model is the preferred model for the hospital #7 GI syndromic surveillance data.

5.4.2 Autoregressive Integrated Moving Average (ARIMA) Models

An ARIMA model is a generalization of the ARMA model that are used for longitudinal data with long-term trends. An initial differencing step (which is the “integrated” part of the model) is used to first remove the trend prior to fitting the autoregressive and/or moving average coefficients.

The model is generally denoted as ARIMA($p,d,q$) where $p$, $d$, and $q$ are the order values of the autoregressive, integrated, and moving average parts of the model. When one or more of the terms is zero, the ARIMA model reduces to a simpler form, many of which have already been discussed. For example, an ARIMA(1,0,0) is an pure first order autoregressive model and an ARIMA(8,0,6) model is the ARMA(8,6) model that gave the optimal fit in Example 5.6.

To fit an ARIMA model, the first step is to identify the order of differencing (i.e., $d$) needed to remove the long-term trend or trends in the data. No trend in the data is modeled by setting $d = 0$, a linear trend is modeled by setting $d = 1$, and a quadratic trend is modeled by $d = 2$. In terms of biosurveillance, these types of trends might arise if the population being monitored is either growing or shrinking over time and that change is reflected in the rate of patients presenting.

Example 5.7. Fit an ARIMA model to the hospital #7 GI data, determining the $p$, $d$, and $q$ that minimize the forecast MSE. Compare and contrast this model to the optimal ARMA model of Example 5.6. Is the optimal ARIMA model to be preferred over the optimal ARMA for this data?\(^4\)

\(^4\) The model was fit in R using the arima function.
Solution: The optimal model is of order $d = 0$, so the ARMA(8,6) of Example 5.6 is the best model. This should actually be intuitive from the plot of the hospital #7 data where there are short term fluctuations in the data, corresponding to seasonal outbreaks and day-to-day noise, but the plots do not show any sort of sustained long-term trends. That is, after each outbreak, the time series returns to a non-outbreak state that is characterized by a very constant average of about 14 patients per day presenting with GI syndrome symptoms.

5.5 Change Point Analysis

Change point analysis (Taylor, 2012a) is a retrospective analytical approach useful for identifying discrete points in time at which the mean of a time series has changed. While not a forecasting method, when it is applied to non-stationary biosurveillance data it is useful for helping to determine when changes, such as outbreaks, start and end.

The idea of change point analysis (CPA) is to model time series data with a set of sample means calculated from contiguous sequences of the data. The time periods immediately after the end of each contiguous sequence are the “change points,” where they represent those times at which the mean of the time series has changed. CPA is a computationally intensive method that uses a number of statistical techniques beyond the scope of this book (particularly the bootstrap) and some that are discussed in later chapters (namely a variant of the CUSUM described in Chapter 7).

The idea of CPA, however, is straightforward to describe. It works as follows. CPA begins by modeling time series simply using the sample mean calculated from all the data, which is the same as assuming there are no change points. That is, for a set of data $y_1, \ldots, y_t$, the sample mean is

$$\bar{y} = \frac{1}{t} \sum_{i=1}^{t} y_i \quad (5.37)$$

and, using MSE as a measure of fit,

$$\text{MSE}_0 = \frac{1}{t} \sum_{i=1}^{t} (y_i - \bar{y})^2. \quad (5.38)$$

That is, in the first stage, the model assumes the fitted values for the data are all equal to the sample mean: $\hat{y}_i = \bar{y}$, $i = 1, \ldots, t$.

Next, a change point $k$ is identified, $1 < k \leq t$, and the model is redefined so that each part of the data is now modeled by its associated sample mean. Thus, given a change point is identified, there are now two sample means,
\[ \bar{y}_1 = \frac{1}{k-1} \sum_{i=1}^{k-1} y_i \] and \[ \bar{y}_2 = \frac{1}{t-k+1} \sum_{i=k}^{t} y_i, \] (5.39)

and the MSE is calculated as

\[ \text{MSE}_1 = \frac{1}{k-1} \sum_{i=1}^{k-1} (y_i - \bar{y}_1)^2 + \frac{1}{t-k+1} \sum_{i=k}^{t} (y_i - \bar{y}_2)^2. \] (5.40)

Now the model is \( \hat{y}_i = \bar{y}_1 \) for \( i = 1, \ldots, k-1 \) and \( \hat{y}_i = \bar{y}_2 \) for \( i = k, \ldots, t \) and, if this model fits better then the previous one, \( \text{MSE}_1 < \text{MSE}_0 \).

In the next step, each of the two segments of the data are examined for change points. For each segment, if a change point is identified then the method proceeds as in the previous step.

For example, imagine that change points \( j \) and \( l \) are identified, one in each segment of the data, so that there are now three change points: \( 1 < j < k < l \leq t \). Then there are four sample means,

\[ \bar{y}_1 = \frac{1}{j-1} \sum_{i=1}^{j-1} y_i, \quad \bar{y}_2 = \frac{1}{k-j} \sum_{i=j}^{k-1} y_i, \]

\[ \bar{y}_3 = \frac{1}{l-k} \sum_{i=k}^{l-1} y_i, \quad \text{and} \quad \bar{y}_4 = \frac{1}{t-l+1} \sum_{i=l}^{t} y_i, \] (5.41)

and the MSE is

\[ \text{MSE}_2 = \frac{1}{j-1} \sum_{i=1}^{j-1} (y_i - \bar{y}_1)^2 + \frac{1}{k-j} \sum_{i=j}^{k-1} (y_i - \bar{y}_2)^2 + \]

\[ \frac{1}{l-k} \sum_{i=k}^{l-1} (y_i - \bar{y}_3)^2 + \frac{1}{t-l+1} \sum_{i=l}^{t} (y_i - \bar{y}_4)^2. \] (5.42)

The model at this stage is \( \hat{y}_i = \bar{y}_1 \) for \( i = 1, \ldots, j-1 \), \( \hat{y}_i = \bar{y}_2 \) for \( i = j, \ldots, k-1 \), \( \hat{y}_i = \bar{y}_3 \) for \( i = k, \ldots, l-1 \), and \( \hat{y}_i = \bar{y}_4 \) for \( i = l, \ldots, t \), and the method continues to iterate until no more change points are found.

Figure 5.19 shows the results of applying CPA to the hospital #7 GI syndrome data, where CPA identified 19 change points. The sample means seem to visually well-describe the data, though in the last outbreak period around day 900 the discrete steps of the sample means looks to be a bit artificial.

Table 5.4 shows the results output from Taylor’s software (Taylor, 2012b), where the specific change points are identified along with confidence intervals.
and associate confidence levels for each change point, as well as the sample means before and after each change point.\(^5\)

Now, the previous description of CPA is not quite accurate as reduction in MSE is not the criteria either for determining whether to include a change point or when to stop the procedure. Indeed, if that was the stopping criterion, the procedure would not stop until the time series was divided up into \(t\) segments each comprised only of one point (since then the MSE is zero). Rather, the CPA uses a cumulative sum (CUSUM) method to identify candidate change points and the bootstrap to determine if candidate change points are statistically significant (and thus should be kept). In addition, Taylor (2012a) recommends using a backward elimination procedure to ensure that as candidate change points are added that the previously identified change points remain statistically significant (and, if not, are removed). Readers interested in the details should consult Taylor (2012a) and Kass-Hout et al. (2012).

\(^5\) In addition to Taylor’s software, a collaborative site has been established with open source software developed specifically for biosurveillance: https://sites.google.com/site/changepointanalysis/.

---

**Fig. 5.19.** Results of applying CPA to the hospital #7 GI syndrome data, where CPA identified 19 change points. The sample means seem to visually well-describe the data, though in the last outbreak period around day 900 the discrete steps of the sample means looks to be a bit artificial.
It’s important to mention that the bootstrap methodology as implemented in the software used to find the change points in the hospital #7 GI data and as posted on the collaborative site assume the data is not autocorrelated. That is clearly not true for this data, though the results as shown in Figure 5.19 look reasonable. Nonetheless, additional research is required to modify the software so that it appropriately handles autocorrelated data. One possibility is to first preprocess the data by modeling the systematic components in the data using one of the methods in this chapter, and use CPA on the residuals that remain after the systematic effects are subtracted.

Returning to the application of CPA to biosurveillance, CPA is most relevant for establishing situational awareness level 2 (comprehension) and technically it belongs in Chapter 4. However, it is included here as it is also closely related to the problem of early event detection and thus is something of a bridge to Chapters 6-8. CPA is included here with the other methods for situational awareness because it is most useful for retrospective analyses of time series data. In comparison, early event detection is focused on prospectively identifying changes.

<table>
<thead>
<tr>
<th>Change Point</th>
<th>Confidence Interval</th>
<th>Confidence Level</th>
<th>Mean Before</th>
<th>Mean After</th>
</tr>
</thead>
<tbody>
<tr>
<td>57 (57, 58)</td>
<td>100%</td>
<td>12.6</td>
<td>48.0</td>
<td></td>
</tr>
<tr>
<td>59 (58, 59)</td>
<td>94%</td>
<td>48.0</td>
<td>16.4</td>
<td></td>
</tr>
<tr>
<td>119 (94, 141)</td>
<td>100%</td>
<td>16.4</td>
<td>20.3</td>
<td></td>
</tr>
<tr>
<td>160 (138, 190)</td>
<td>99%</td>
<td>20.3</td>
<td>25.6</td>
<td></td>
</tr>
<tr>
<td>192 (187, 195)</td>
<td>100%</td>
<td>25.6</td>
<td>15.6</td>
<td></td>
</tr>
<tr>
<td>216 (193, 305)</td>
<td>92%</td>
<td>15.6</td>
<td>13.5</td>
<td></td>
</tr>
<tr>
<td>405 (399, 414)</td>
<td>100%</td>
<td>13.5</td>
<td>19.4</td>
<td></td>
</tr>
<tr>
<td>496 (485, 515)</td>
<td>100%</td>
<td>19.4</td>
<td>24.3</td>
<td></td>
</tr>
<tr>
<td>565 (560, 567)</td>
<td>100%</td>
<td>24.3</td>
<td>14.7</td>
<td></td>
</tr>
<tr>
<td>727 (718, 730)</td>
<td>100%</td>
<td>14.7</td>
<td>20.3</td>
<td></td>
</tr>
<tr>
<td>738 (735, 752)</td>
<td>93%</td>
<td>20.3</td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>820 (815, 822)</td>
<td>91%</td>
<td>14.9</td>
<td>23.7</td>
<td></td>
</tr>
<tr>
<td>827 (822, 828)</td>
<td>96%</td>
<td>23.7</td>
<td>16.4</td>
<td></td>
</tr>
<tr>
<td>836 (835, 836)</td>
<td>96%</td>
<td>16.4</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>838 (838, 845)</td>
<td>93%</td>
<td>6.5</td>
<td>19.3</td>
<td></td>
</tr>
<tr>
<td>881 (870, 883)</td>
<td>93%</td>
<td>19.3</td>
<td>27.0</td>
<td></td>
</tr>
<tr>
<td>889 (887, 896)</td>
<td>100%</td>
<td>27.0</td>
<td>35.4</td>
<td></td>
</tr>
<tr>
<td>915 (911, 921)</td>
<td>100%</td>
<td>35.4</td>
<td>27.6</td>
<td></td>
</tr>
<tr>
<td>937 (934, 938)</td>
<td>99%</td>
<td>27.6</td>
<td>15.8</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: CPA output from Taylor’s software (Taylor, 2012b), where the specific change points are identified along with confidence intervals and associate confidence levels for each change point, as well as the sample means before and after each change point.
Indeed, CPA has been implemented in BioSense 2.0 expressly for the purpose of improving situational awareness and augmenting the early event detection methods. As the current director of the CDC’s Division of Informatics Solutions and Operations said,

“When we combined forecasting with CPA, we were actually able to get timely picture of the situation and to answer questions; such as, is the incidence going up, down or is stable and when combined with EARS we were able to make sure not to miss sudden or subtle changes” (Kass-Hout, 2012a).

5.6 Discussion & Summary

This chapter began with the use of statistical models for preprocessing biosurveillance data and then came full-circle back to the application of preprocessing when CPA is used for autocorrelated biosurveillance data. Preprocessing is also important for the proper application of many of the early event detection (EED) methods discussed in Chapters 6-8. As with CPA, most of these methods are designed for data that are not autocorrelated. Given that autocorrelation is inherent in biosurveillance data, preprocessing can be used to eliminate, or at least mitigate, autocorrelation by first modeling and then removing the systematic effects in the data and then using CPA and the EED methods on the residuals.

The main focus of this chapter is on statistical models useful for the projection step of situational awareness, including smoothing, regression-based, and time series models. The examples in this chapter were almost all centered around one set of data, for which it turned out that an ARMA model gave the best forecasts. However, this is not a general result and it’s almost surely true that other models will be better for other types and sets of data. What is generally true is that no one method will be universally applicable or optimal. Biosurveillance system designers and operators should allow for and apply a suite of modeling tools in order to find and then use the best model.

In spite of its length, this chapter has really only scratched the surface of smoothing, regression, and time series modeling. Each of these topics is worthy of, and indeed is the subject of, numerous textbooks. This chapter has mainly introduced the general ideas of each of the types of models and illustrated their application to biosurveillance data via the examples. Not mentioned in this chapter are the details and intricacies of model fitting, including model diagnostics for checking whether assumptions are violated, and numerous model variants and extensions. Readers interested in learning more about a particular methodology should consult one or more of the references in the Additional Reading section.
Additional Reading

For those who would like to delve more deeply into the material, consider:

- For additional reading on preprocessing, including other methods and applications, see Hagen et al. (2011), Shmueli & Burkom (2010), Lotze et al. (2008), Fricker et al. (2008a) Fricker et al. (2008b), and Burkom et al. (2006).

- While almost any statistics textbook will provide at least a brief introduction to linear regression, an excellent text focused just on regression is *Introduction to Linear Regression Analysis* (Montgomery et al., 2012). In addition, *Time Series Analysis: Regression Techniques* by (Ostrom, 1990) is a good introduction to the application of regression to modeling time series data.

- In terms of time series modeling, a good introduction to smoothing and ARIMA models is *Introduction to Time Series Analysis and Forecasting* (Montgomery et al., 2008). *Introduction to Time Series and Forecasting* (Brockwell & Davis, 2002) is a more advanced treatment with a focus on ARMA and ARIMA models. A good advanced text is *Time Series Analysis: Forecasting and Control* (Box et al., 2008).

- For other examples of the application of models to biosurveillance data, see Hagen et al. (2011) who applied adaptive regression to syndromic surveillance data; Burkom et al. (2006) who compared a loglinear regression model, an adaptive regression model with sliding baseline, and a Holt-Winters method for generalized exponential smoothing; the CDC’s cyclical regression models discussed in Hutwagner et al. (2003b); log-linear regression models in Farrington et al. (1996); and time series models in Reis & Mandl (2003). Also see Fricker et al. (2008a) and Fricker et al. (2008b) for the use of adaptive regression on simulated biosurveillance data, and Shmueli & Burkom (2010) for additional discussion of the use of regression and time series methods for syndromic surveillance.

- For more on CPA applied to biosurveillance data, see Kass-Hout et al. (2012), where they compare Taylor’s method to two alternative methods. Readers interested in collaborating on developing CPA software or downloading open source programs should see Kass-Hout (2012b).