An empirical Bayes procedure for adaptive forecasting of shrimp yield

David G. Whiting *, H. Dennis Tolley, Gilbert W. Fellingham

Department of Statistics, Brigham Young University, Provo, UT, 84602, USA

Accepted 9 July 1999

Abstract

In order to make decisions regarding when to harvest cultured shrimp, short-term forecasts of future growth must be compared with expected future costs and changes in the value of the shrimp harvested. As with most agricultural products, researchers and business professionals collect and maintain data on the various factors that influence growth. Though many general fixed relationships between, say salinity, temperature, time and turbidity are known, exactly how these affect a particular crop changes from year to year. Vitality of the stock, unmeasured characteristics of the pond, and other factors which vary over time have a significant affect on the yield of a harvest. Consequently, to make reasonable short-term forecasts, forecasting models must include both the general factors gathered from past experience and crop specific variations from these overall effects. This paper presents a method of obtaining such a forecasting equation using a general mixed model setup currently available in many computer packages. These general formulas are illustrated with a real life example. SAS code necessary to implement the analysis is also included. The paper concludes by presenting a general equation for short-term forecasts using the estimates obtained from a SAS analysis. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Bayes procedure; Shrimp; Forecasting

1. Introduction

For a corporation involved in the commercial farming of freshwater or marine organisms, there are a number of variables affecting profitability which must regularly be taken into account. One of the more important factors in determining net income is the decision of when to harvest the crop of organisms. If the company harvests too early,
In this paper, we illustrate a statistical approach for the adaptive forecasting of shrimp yield in the context of the following scenario. A shrimp farming corporation operates a series of 20 ostensibly identical 30-acre ponds in which the Pacific white shrimp (*Penaeus vannamei*) is the sole species being grown. These ponds are traditionally harvested after a 22–27 week growth period, each period roughly corresponding to the dry and wet (summer and winter, respectively) seasons. The operators collect extensive data on each pond for each season and year including, but not limited to, daily measurements of morning and evening air and water temperature, daily measurements of salinity and water turbidity, and weekly measurements of average shrimp weight. Inasmuch as the variables influencing the growth of *P. vannamei* have been addressed elsewhere (see e.g., Bray et al., 1994; Wyban et al., 1995), we will not specifically deal with them here. Likewise, Leung and Hochman (1990); Hochman et al. (1990); Leung et al. (1993), and Tian et al. (1993) have considered shrimp farming in a broad economic framework. For instance, Hochman et al. (1990) and Leung and Hochman (1990) present a complete management model for determining the optimal stocking and harvesting schedules of farmed shrimp based on a series of decision rules. Rather than addressing all of the issues pertaining to such a management model, we focus exclusively on the methodology needed to compute an adaptive forecast given data from previous years and year-to-date data for the current year. This component could then easily be included in the growth estimation portion of the economic model provided by either Hochman et al. (1990) or Leung and Hochman (1990).

Prediction of future growth entails two components. First, a regression equation is typically used to model the average growth rate of the organism over time. Subsequently, the individual characteristics of the current crop of organisms which cause their growth rate to differ from previous crops are also accounted for in the model. These include such things as weather and environmental conditions peculiar to the current year, the viability of the original stock, and so forth. As the crop matures, the importance of these characteristics becomes more evident and therefore should be used in conjunction with modeled averages to predict growth. Thus, the problem is how to use past years’ information adjusted for current conditions to determine the optimal harvest time during the current year.

The approach requires the following steps: (1) the assessment of which measured variables are significant predictors of growth, (2) the estimation of variability across ponds, years, and seasons, (3) the calculation of a random coefficients growth curve model for the historical data, and (4) the use of conditional probability to give adaptive forecasts of shrimp yield. In statistical nomenclature, this is referred to as either an empirical Bayes procedure (Carlin and Louis, 1996) or, in traditional statistical terms, as an estimated best linear unbiased prediction (EBLUP) procedure. In short, we use historical data plus current measurements to yield updated predictions.
2. Growth curve models

Wishart (1938) was the first to consider the problem of analyzing growth rates. These models, known as random coefficient models in the statistical literature, have since received considerable attention (see, for example, Rao, 1959; Potthoff and Roy, 1964; Grizzle and Allen, 1969; Laird and Ware, 1982; or Littell et al., 1996, chapter ).

Suppose we measure the total mass of shrimp in each pond on a weekly basis for a given year and season, say summer 1995. Conceptually, “pond” is the experimental unit whose growth we measure over time. Let us assume that shrimp growth rates for summer 1995 are similar across ponds due to identical weather conditions, similar feed, turbidity, salinity, and so on. A simple linear model to describe this average growth is

\[ y_{ij} = \beta_0 + \beta_1 w_{ij} + \beta_2 w_{ij}^2 + e_{ij} \]  

where \( y_{ij} \) is the yield, \( w_{ij} \) is the week number, \( w_{ij}^2 \) is the week squared, and \( e_{ij} \) is the error associated with the \( i \)th pond and \( j \)th week. The reason for including the square of the weeks is to account for a nonlinear growth rate over time. Explicitly, when the initial rate of growth is nearly linear but tapers off as the shrimp get closer to full size, the growth curve flattens out. The square term in the model provides for this.

The 20 ponds can be thought of as a random sample from a hypothetical population of summer 1995 ponds represented by Eq. 1. Define the growth curve equation for a specific pond \( i \) as

\[ y_{ij} = b_{0i} + b_{1i} w_{ij} + b_{2i} w_{ij}^2 + e_{ij} \]  

where \( e_{ij} \) are assumed to be Gaussian errors with mean zero and variance \( \sigma^2 \). The coefficients \( b_{0i} \), \( b_{1i} \) and \( b_{2i} \) are realizations of \( \beta_0 \), \( \beta_1 \) and \( \beta_2 \) for the \( i \)th pond, respectively. In other words, the random growth curves for each pond given in Eq. 2 are regression lines which deviate about the overall population growth curve given by (1). To illustrate, in Fig. 1 the dark line represents a hypothetical population growth curve and the two lighter lines represent individual growth curves from two hypothetical ponds.

Random Coefficient Growth Curve

![Random Coefficient Growth Curve](image)

Fig. 1. Random coefficient growth curves for two hypothetical ponds.
From Eq. (2), we see that the error term \( e_{ij} \) for the overall linear growth model (1) consists of deviations from the population parameters plus random error, i.e.,

\[
e_{ij} = (b_{0i} - \beta_0) + (b_{1i} - \beta_1)w_{ij} + (b_{2i} - \beta_2)w_{ij}^2 + e_{ij}.
\]

We can rewrite the growth curve Eq. (1) as

\[
y_{ij} = \beta_0 + \beta_1w_{ij} + \beta_2w_{ij}^2 + (b_{0i} - \beta_0) + (b_{1i} - \beta_1)w_{ij} + (b_{2i} - \beta_2)w_{ij}^2 + e_{ij}.
\]

For purposes of prediction, the \( \beta \) coefficients and \( \sigma^2 \) are estimated from previous years’ data. The \( \beta \) coefficients, however, are peculiar to the specific pond, season and year.

To further refine the model in Eq. (1) for the \( i \)th pond and \( j \)th week, we add terms for the covariates temperature, \( t_{ij} \), salinity, \( s_{ij} \), and turbidity, \( u_{ij} \). The new model is

\[
y_{ij} = \beta_0 + \beta_1w_{ij} + \beta_2w_{ij}^2 + \beta_3t_{ij} + \beta_4s_{ij} + \beta_5u_{ij} + \beta_6t_{ij}w_{ij}
\]

\[
+ \beta_7s_{ij}w_{ij} + \beta_8u_{ij}w_{ij} + \beta_9t_{ij}w_{ij}^2 + \beta_{10}s_{ij}w_{ij}^2 + \beta_{11}u_{ij}w_{ij}^2 + e_{ij}.
\]  

(3)

Here \( \beta_0 \) is the coefficient for the \( y \)-intercept, \( \beta_1 \) is the coefficient for linear growth in weeks and \( \beta_2 \) is the coefficient for growth in squared weeks. The coefficients for temperature, salinity and turbidity are given by \( \beta_3 \), \( \beta_4 \) and \( \beta_5 \), respectively. The remaining coefficients, \( \beta_6 \) through \( \beta_{11} \), are for various interactions of interest. Note that further interactions (e.g., \( \beta_{12}t_{ij}s_{ij}w_{ij} \)) could have been included in this model.

It is illustrative to write Eq. (3) as

\[
y_{ij} = (\beta_0 + \beta_3t_{ij} + \beta_4s_{ij} + \beta_5u_{ij}) + (\beta_1 + \beta_6t_{ij} + \beta_7s_{ij} + \beta_8u_{ij})w_{ij}
\]

\[
+ (\beta_2 + \beta_9t_{ij} + \beta_{10}s_{ij} + \beta_{11}u_{ij})w_{ij}^2 + e_{ij}.
\]  

(4)

Here \( \beta_0 \), \( \beta_3 \), \( \beta_4 \) and \( \beta_5 \) are the coefficients for the intercept of the growth curve, \( \beta_1 \), \( \beta_6 \), \( \beta_7 \) and \( \beta_8 \) are the coefficients for the linear slope of the curve, and \( \beta_2 \), \( \beta_9 \), \( \beta_{10} \) and \( \beta_{11} \) are the coefficients for the quadratic portion of the model. More specifically, for example, \( \beta_7 \) and \( \beta_{10} \) measure the effect of salinity on linear and quadratic growth, respectively.

In certain situations it may be reasonable to assume a no-intercept model, in which case \( \beta_0 \), \( \beta_3 \), \( \beta_4 \) and \( \beta_5 \) from Eq. (4) would be set equal to zero rather than estimated, yielding

\[
y_{ij} = (\beta_1 + \beta_6t_{ij} + \beta_7s_{ij} + \beta_8u_{ij})w_{ij} + (\beta_2 + \beta_9t_{ij} + \beta_{10}s_{ij} + \beta_{11}u_{ij})w_{ij}^2 + e_{ij}.
\]  

(5)

In summary, the yield in a pond can be modeled as a growth curve which increases weekly. The method for forecasting growth to determine when to harvest the shrimp consists of two components. First, we estimate the growth coefficients and the amount of pond-to-pond variability from historical data. Second, we use the year-to-date data to
estimate adjustments to the historical growth curve coefficients. These corrected coefficients can then be used to forecast growth in the current year.

3. Random coefficient growth curve analysis

The analyses presented here were carried out using the Mixed procedure from the statistical software package SAS. (Instructions for the SAS package can be found in SAS *Language Reference Manual*, Version 6, SAS Institute) Note that SAS’s Proc Mixed is a very powerful procedure for analyzing many types of mixed linear models, random coefficient growth curves being only one. For more details on mixed models, see e.g., Scheffe (1959); Searle et al. (1992), or Christensen (1987). Additionally, see Littell et al. (1996) for a thorough discussion and some excellent examples of mixed model analysis in SAS.

3.1. Preliminary analysis

An initial investigation of the data entailed creating a growth curve model for all 4 years and both seasons. The purpose of such a preliminary analysis is to decide on an appropriate model upon which later prediction will be based. Our model was refined as follows.

After comparing results from using the general Eq. 4 and the no-intercept Eq. 5, the no-intercept model was deemed most appropriate for modeling shrimp growth. The reasons for this are two-fold. First, the operators did not measure yield until 5 or 6 weeks after shrimp were introduced into the ponds (i.e., when the shrimp were large enough to be captured by the nets being used for measurement). Thus, information on these first few weeks of growth was unavailable and, in practical terms, yield could be considered to be zero at week 0. Secondly, since our goal is prediction at the upper end of the growth curve, the growth behavior at this low end of the curve had very little influence on the forecasts of interest. In other words, using the no-intercept constraint did not significantly alter our prediction results.

This preliminary analysis also showed that growth rates for individual ponds were not necessarily similar from year to year. Furthermore, the variability found within growth periods was not uniform. For instance, the growth curves for a set of ponds in summer 1994 were markedly more variable than the growth curves for the same set of ponds in summer 1996.

Most importantly, growth rates were shown to be consistent within seasons across different years but not between seasons within the same or different years. In short, summer growth rates differed significantly from winter growth rates. As shown in Fig. 2, although winter growth has a faster linear increase than summer, the growth rate also flattens out earlier. Fig. 3 shows the estimated growth curves for each year for the summer and winter seasons. Note that the variability of growth rates within a season between years is small compared with the variability between seasons.

Furthermore, the variables influential in predicting growth differed for the two seasons. While temperature and salinity were found to be statistically significant
predictors in summer, salinity and turbidity were significant winter growth predictors. Since the predictive factors were not the same for summer and winter, a separate growth rate model was constructed for each season.

Stating that a factor is not statistically significant does not imply that it is unimportant in growth. For instance, the operators of the shrimp farm felt that turbidity was influential primarily in the first few weeks of growth. However, at least for the summer season, growth for that period may have been adequately described by other factors already in the model. In other words, if turbidity is highly correlated with one or more factors currently in the model, its inclusion is not necessary for purposes of prediction.

The exclusion of temperature in winter illustrates another reason a factor which has an obvious impact on growth may be found insignificant. Specifically, if the factor shows little variation over the growth period, its role in prediction may be minimal. In the winter seasons, temperature levels fluctuated very little over time with an observed standard deviation of only 1.4°C, hence its exclusion is not surprising. Thus, a factor may be statistically insignificant in the growth curve model although its impact on growth is unquestioned.

3.2. Summer growth curve analysis

A detailed description of the growth curve analysis for the summer season follows. The SAS commands and output for this analysis may be found in Appendices A and B, respectively. Recall that for summer, turbidity was found to be an insignificant predictor.
Fig. 3. Estimated growth rates for 1994–1997.

of growth. Since a no-intercept model was chosen for both summer and winter seasons, the model equation for summer is identical to Eq. (5) except that $\beta_s$ and $\beta_{1i}$, the coefficients for the effect of turbidity on linear and quadratic growth, respectively, are omitted. The summer model is thus

$$y_{ij} = (\beta_1 + \beta_6 t_{ij} + \beta_7 s_{ij})w_{ij} + (\beta_2 + \beta_9 t_{ij} + \beta_{10} s_{ij})w_{ij}^2 + \epsilon_{ij}.$$

(6)
Assuming yield data starts at week one and lasts until week \( n \), the matrix form of the growth curve model for the \( i \)th pond can be expressed as \( y_i = X_i \beta + \varepsilon \) or, more explicitly,

\[
\begin{bmatrix}
y_{i1} \\
y_{i2} \\
y_{i3} \\
y_{in}
\end{bmatrix} = 
\begin{bmatrix}
1 & t_{i1} & s_{i1} & 1 & t_{i1} & s_{i1} \\
2 & 2t_{i2} & 2s_{i2} & 4 & 4t_{i2} & 4s_{i2} \\
3 & 3t_{i3} & 3s_{i3} & 9 & 9t_{i3} & 9s_{i3} \\
. & . & . & . & . & . \\
n & nt_{in} & ns_{in} & n^2 & n^2t_{in} & n^2s_{in}
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_6 \\
\beta_7 \\
\beta_2 \\
\beta_9 \\
\beta_{10}
\end{bmatrix} + 
\begin{bmatrix}
\varepsilon_{i1} \\
\varepsilon_{i2} \\
\varepsilon_{i3} \\
\varepsilon_{i4} \\
\varepsilon_{in}
\end{bmatrix}
\]

(7)

The three components of the historical data necessary for adaptive forecasting are (1) the vector of coefficients \( \beta \), (2) the covariance matrix of \( \beta \) denoted \( \Psi \) and (3) the overall variance \( \sigma^2 \). We label the estimates of these three quantities \( \hat{\beta}, \hat{\Psi} \) and \( \hat{\sigma}^2 \), respectively.

Note that the \( \beta \) vector obtained from SAS includes coefficients for each year, i.e.,

\[
\beta^* = (\beta_{i(1994)}, \beta_{i(1995)}, \beta_{i(1996)}, \beta_{i(1997)}, \ldots, \beta_{i(10(1994)}, \beta_{i(10(1995)}, \beta_{i(10(1996)}, \beta_{i(10(1997)})
\]

The six SAS ESTIMATE statements yield the estimates of \( \beta \) given in Eq. (7) averaged over all 4 years. Likewise, the covariance matrix of \( \beta^* \) output by the SAS command COVB is a \( 24 \times 24 \) matrix. The estimates of \( \Psi \) may be calculated by using the matrix equation

\[
\Psi = A\Psi^gA'
\]

where \( \Psi^g \) is the covariance matrix of \( \beta^* \) output by SAS and

\[
A = 
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \otimes \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix},
\]

with \( \otimes \) denoting the Kronecker product operator.

The estimates of \( \sigma^2, \beta \) and \( \Psi \) for summer are \( \hat{\sigma}^2 = 1.613 \),

\[
\hat{\beta} = 
\begin{bmatrix}
-2.81 \times 10^0 \\
9.79 \times 10^{-2} \\
1.70 \times 10^{-2} \\
1.72 \times 10^{-1} \\
-4.76 \times 10^{-3} \\
8.26 \times 10^{-4}
\end{bmatrix}
\]

and

\[
\hat{\Psi} = 
\begin{bmatrix}
4.6 \times 10^{-2} & -2.2 \times 10^{-3} & -1.3 \times 10^{-3} & 6.6 \times 10^{-5} & -1.3 \times 10^{-4} & 1.2 \times 10^{-6} \\
1.2 \times 10^{-4} & 6.7 \times 10^{-5} & -3.7 \times 10^{-6} & 5.9 \times 10^{-6} & -4.0 \times 10^{-7} \\
4.8 \times 10^{-5} & -2.4 \times 10^{-6} & -5.3 \times 10^{-6} & 2.1 \times 10^{-7} & -9.6 \times 10^{-9} \\
1.3 \times 10^{-7} & 2.6 \times 10^{-7} & -5.7 \times 10^{-7} & 2.9 \times 10^{-8} \\
\end{bmatrix}
\]
respectively. (Note also that the lower diagonal portion of the matrix has been omitted as it is symmetric.)

4. Empirical Bayes analysis

The philosophy underlying empirical Bayes analysis is to combine data from previous years with current, year-to-date data in producing forecasts which reflect the overall form of historical patterns but are adapted for current conditions.

Recall that are the historical components used in forecasting. Suppose we have collected 8 weeks of data for a particular pond in the current year. Organizing this data as in Eq. (7), let

\[ y_0 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} \quad \text{and} \quad X_0 = \begin{bmatrix} 1 & t_1 & s_1 & 1 & t_1 & s_1 \\ 2 & 2t_2 & 2s_2 & 4 & 4t_2 & 4s_2 \\ 3 & 3t_3 & 3s_3 & 9 & 9t_3 & 9s_3 \\ 4 & 4t_4 & 4s_4 & 16 & 16t_4 & 16s_4 \\ 5 & 5t_5 & 5s_5 & 25 & 25t_5 & 25s_5 \\ 6 & 6t_6 & 6s_6 & 36 & 36t_6 & 36s_6 \\ 7 & 7t_7 & 7s_7 & 49 & 49t_7 & 49s_7 \\ 8 & 8t_8 & 8s_8 & 64 & 64t_8 & 64s_8 \end{bmatrix} \]

be a column vector containing weight measurements and a matrix of independent variables, respectively. To forecast weight for a future week, say week 10, we construct a row vector \( x' \) by entering values for that week in the same order as the values in the rows of \( X_0 \). Thus,

\[ x' = (10, 10t, 10s, 100, 100t, 100s). \]

Since temperature and salinity are unknown for the week we are predicting, we can either enter average values based on the past few weeks or enter an outside forecasted value. For example, if a cold front is expected, one could place the forecasted temperature value into \( t \). Additionally, if our interest is to predict for a specific pond, we may use the pond-adjusted values of \( \hat{\beta} \). Note that the covariance matrix \( \hat{\Psi} \) would remain the same.

The matrix equation used to perform an empirical Bayes forecast of shrimp yield is given by

\[ \hat{y}_0 = x'_0 \hat{\beta} + x'_0 \hat{\Psi} X_0 (X_0 \hat{\Psi} X_0^{-1} + \hat{\sigma}^2 I)^{-1} (y_0 - X_0 \hat{\beta}). \]

where \( \hat{y}_0 \) is the forecasted yield and \( I \) is an identity matrix of appropriate dimension. In the context of this problem, the empirical Bayes estimator is identical to the estimated best linear unbiased predictor (BLUP) from linear model theory. (See Christensen, 1987 for the derivation of Eq. (8) and Maritz and Lwin, 1989 or Carlin and Louis, 1996 for further information on empirical Bayes estimation.)
Table 1
Values for 1-, 2- and 3-week ahead forecasts for a pond from summer 1997

<table>
<thead>
<tr>
<th>Week</th>
<th>Actual</th>
<th>Lag 1</th>
<th>Lag 2</th>
<th>Lag 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>3.8</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>8</td>
<td>4.3</td>
<td>4.21</td>
<td>3.90</td>
<td>~</td>
</tr>
<tr>
<td>9</td>
<td>4.7</td>
<td>4.74</td>
<td>4.87</td>
<td>4.56</td>
</tr>
<tr>
<td>10</td>
<td>5.6</td>
<td>5.35</td>
<td>5.42</td>
<td>5.56</td>
</tr>
<tr>
<td>11</td>
<td>5.2</td>
<td>6.09</td>
<td>6.07</td>
<td>6.12</td>
</tr>
<tr>
<td>12</td>
<td>6.6</td>
<td>6.83</td>
<td>6.84</td>
<td>6.82</td>
</tr>
<tr>
<td>13</td>
<td>7.7</td>
<td>7.79</td>
<td>7.59</td>
<td>7.62</td>
</tr>
<tr>
<td>14</td>
<td>8.5</td>
<td>8.73</td>
<td>8.55</td>
<td>8.38</td>
</tr>
<tr>
<td>15</td>
<td>9.4</td>
<td>9.62</td>
<td>9.48</td>
<td>9.33</td>
</tr>
<tr>
<td>16</td>
<td>10.9</td>
<td>10.54</td>
<td>10.36</td>
<td>10.24</td>
</tr>
<tr>
<td>17</td>
<td>11.0</td>
<td>11.46</td>
<td>11.23</td>
<td>11.10</td>
</tr>
<tr>
<td>18</td>
<td>12.5</td>
<td>12.13</td>
<td>12.08</td>
<td>11.92</td>
</tr>
<tr>
<td>19</td>
<td>13.0</td>
<td>12.70</td>
<td>12.69</td>
<td>12.70</td>
</tr>
<tr>
<td>20</td>
<td>13.6</td>
<td>13.53</td>
<td>13.33</td>
<td>13.25</td>
</tr>
<tr>
<td>21</td>
<td>14.3</td>
<td>14.66</td>
<td>14.30</td>
<td>13.97</td>
</tr>
<tr>
<td>22</td>
<td>14.6</td>
<td>15.68</td>
<td>15.52</td>
<td>15.08</td>
</tr>
<tr>
<td>23</td>
<td>15.4</td>
<td>16.25</td>
<td>16.61</td>
<td>16.39</td>
</tr>
<tr>
<td>24</td>
<td>16.2</td>
<td>16.30</td>
<td>17.16</td>
<td>17.55</td>
</tr>
<tr>
<td>25</td>
<td>17.0</td>
<td>16.59</td>
<td>17.03</td>
<td>18.09</td>
</tr>
<tr>
<td>26</td>
<td>16.9</td>
<td>16.74</td>
<td>17.20</td>
<td>17.76</td>
</tr>
<tr>
<td>27</td>
<td>16.8</td>
<td>17.32</td>
<td>17.25</td>
<td>17.80</td>
</tr>
</tbody>
</table>

In summary, the empirical Bayes Eq. (8) has the following components: (1) historical data gives us values for $\beta$, $\Psi$ and $\sigma^2$, (2) year-to-date data gives us values for $y_0$ and $X_0$, and (3) $x_0$ is created by plugging in the week number as well as average or predicted values for the covariates temperature and salinity.

![Fig. 4. Forecasted values for a pond from summer 1997.](image)
5. Forecasting example

We present as an example the forecasted growth curves for an individual pond from the summer of 1997. For lag-one forecasting, data from weeks 1–6 were used to predict week 7, then weeks 1–7 were used to predict week 8, and so on. Forecasts were computed at lags of 1, 2, and 3 weeks. Temperature and salinity values for the weeks being predicted were obtained by averaging the respective values from the two previous weeks. Table 1 shows the actual values plus the 1-, 2-, and 3-week ahead forecasts for this pond. Fig. 4 displays these data graphically.

From Fig. 4, note that the lag 1, lag 2, and lag 3 forecasts all follow the shape of the growth curve quite well. Although the lag 2 and lag 3 forecasts give slightly higher predictions of shrimp yield than were actually observed in the last few weeks of growth, they both do a remarkable job of predicting the point on the curve at which growth ceased and the curve flattened out. All other economic factors being equal, the optimal harvest time for this particular pond is week 25, a fact predicted at least three weeks earlier.

6. Conclusions

We have illustrated how the empirical Bayes method can be used to incorporate both historical data and data obtained on the current crop to make short-term predictions. This entailed estimating the overall growth curve as a function of weeks, salinity, temperature and turbidity. Analyses were done using the Proc Mixed procedure in SAS. Data on the current crop is also included in the empirical Bayes setup so that deviations of the current year from previous years can be included in the predictions. The example provided illustrated the ability of the prediction method to forecast 2 or 3 weeks into the future a downturn in the overall rate of growth. Using this in conjunction with other economic factors allows decision makers to estimate profit from harvest before the decision to harvest must be made.

The concept of adjusting estimates based on recent data epitomizes some of the advantages of Bayesian and empirical Bayesian methods. In the past, however, computational demands made implementation of methods such as the shrimp example impractical. Now that the computational difficulties have been reduced, the major reason for not using these methods appears to be a lack of understanding regarding their existence and ease of application. As illustrated here, the analysis to implement these procedures is not far removed from the common analysis of covariance approach. The results of the analysis can then be used to construct a prediction equation using matrix methods. This prediction equation constantly updates itself according to the most recent data available.

Acknowledgements

The authors wish to thank the associate editor and referees for their many helpful comments.
Appendix A. SAS code

DATA shrimp;
  INFILE 'shrimp.dat';
  INPUT pond year weeks weight temp turb sal;
  week2 = weeks * weeks;
RUN;
/* * Summer Random Coefficients Growth Curve Model */
PROC MIXED;
  CLASS year pond;
  MODEL weight =
    weeks * year temp * weeks * year sal * weeks * year
    week2 * year temp * week2 * year sal * week2 * year
  / SOLUTION NOINT COVB;
  RANDOM weeks week2 / TYPE = UN SUBJECT = pond SOLUTION;
  ESTIMATE 'weeks' weeks * year 0.25 0.25 0.25 0.25;
  ESTIMATE 'temp * weeks' temp * weeks * year 0.25 0.25 0.25 0.25;
  ESTIMATE 'sal * weeks' sal * weeks * year 0.25 0.25 0.25 0.25;
  ESTIMATE 'week2' week2 * year 0.25 0.25 0.25 0.25;
  ESTIMATE 'temp * week2' temp * week2 * year 0.25 0.25 0.25 0.25;
  ESTIMATE 'sal * week2' sal * week2 * year 0.25 0.25 0.25 0.25;
  MAKE 'COVB' OUT = summer;
RUN;
/* * Create Covariance Matrix for Summer */
PROC IML;
  USE summer;
  READ all var_num into V;
  V = V [1:24,2:25 ];
  name1 = {'weeks', 'temp', 'sal', 'week2', 'temp2', 'sal2'};
  names = {'weeks', 'temp * weeks', 'sal * weeks',
            'week2', 'temp * week2', 'sal * week2'};
  x = J(4,1)/4;
  a = I(6);
  y = a @ x;
  summer = y' * V * y;
  print summer [label = "'Summer Covariance Matrix'"
                rowname = names colname = names];
  se = sqrt (diag (summer));
  print se [label = "'Standard Errors'"
            rowname = names colname = names];
RUN;
ENDSAS;
Appendix B

B.1. SAS output

Covariance Parameter Estimates (REML)

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>POND</td>
<td>0.03450224</td>
</tr>
<tr>
<td>UN(2,1)</td>
<td>POND</td>
<td>-0.00108508</td>
</tr>
<tr>
<td>UN(2,2)</td>
<td>POND</td>
<td>0.00003881</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>1.61322927</td>
</tr>
</tbody>
</table>

ESTIMATE statement results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>DF</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>weeks</td>
<td>-2.80645053</td>
<td>0.21371564</td>
<td>2073</td>
<td>-13.13</td>
</tr>
<tr>
<td>temp*weeks</td>
<td>0.09787990</td>
<td>0.00693497</td>
<td>2073</td>
<td>14.11</td>
</tr>
<tr>
<td>sal*weeks</td>
<td>0.01696653</td>
<td>0.00349486</td>
<td>2073</td>
<td>4.85</td>
</tr>
<tr>
<td>week2</td>
<td>0.17237805</td>
<td>0.01150535</td>
<td>2073</td>
<td>15.46</td>
</tr>
<tr>
<td>temp*week2</td>
<td>-0.000476407</td>
<td>0.00036008</td>
<td>2073</td>
<td>13.23</td>
</tr>
<tr>
<td>sal*week2</td>
<td>-0.00082625</td>
<td>0.00017068</td>
<td>2073</td>
<td>-4.84</td>
</tr>
<tr>
<td>Summer weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>temp*week2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sal*week2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weeks</td>
<td>0.0456744 (0.0000656)</td>
<td>-0.001339 (7.2432e-6)</td>
<td>-0.000132</td>
<td>-0.002227</td>
</tr>
<tr>
<td>sal*weeks</td>
<td>-0.000132 (2.5812e-7)</td>
<td>-5.335e-6 (-5.733e-7)</td>
<td>-5.335e-6</td>
<td>0.0000668</td>
</tr>
<tr>
<td>week2</td>
<td>-0.002227 (-3.714e-4)</td>
<td>0.0000668 (-3.99e-7)</td>
<td>5.8817e-6</td>
<td>0.0001243</td>
</tr>
<tr>
<td>temp*week2</td>
<td>0.0000656 (1.2966e-7)</td>
<td>-2.3e-6 (-9.545e-9)</td>
<td>2.5812e-7</td>
<td>-3.714e-6</td>
</tr>
<tr>
<td>sal*week2</td>
<td>7.2432e-6 (-9.545e-9)</td>
<td>2.1237e-7 (2.9133e-8)</td>
<td>-5.733e-7</td>
<td>-3.99e-7</td>
</tr>
</tbody>
</table>

References


