ABSTRACT

Substantial increases in milk price volatility have resulted from changes in federal dairy policies. For a dairy farm, however, monthly gross milk receipts are a function of unit price and quantity produced. Both can vary substantially over time. Therefore, to be effective, risk management strategies must address milk and input price volatility (price risk management) and fluctuations in milk production per cow and cow numbers (production risk management). Herd milk production through time can be modeled as a discrete stochastic process using finite Markov chains. Cows at time \( t = 0 \) are assigned to homogeneous production cells in four-dimensional arrays with coordinates determined by parity (1,2,3), week in milk (1, \ldots ,104), pregnancy status (0,1), and week pregnant (1, \ldots ,40). The processes of aging, pregnancy, involuntary cull, voluntary cull, abortion, dry-off, and freshening from week \( i-1 \) to week \( i \) are accounted for, using nonstationary transition probabilities. Bayesian estimates of transition probabilities are derived from historical herd data, assuming that individual outcomes are from Bernoulli distributions. The values of parameters \( \theta_i \) for the Bernoulli distributions are unknown but have prior distributions that follow beta distributions with parameters \( \alpha_i \) and \( \beta_i \) estimated from historical data. Herd observations are then used to generate posterior distributions of \( \theta_i \), also from beta distributions. Projecting from one week to the next is accomplished by moving virtual animals from one production cell to the next based on the transition probability assigned to that path. Summing production estimates and variances of all independent cells provides for an expected herd production with an associated variance. As expected, the forecast variance increases with time, reflecting increased uncertainty of distant projections. Model validation presents an interesting problem because future observations used for validation are under human control and are not independent of the forecast.

Key words: forecasting, discrete Markov process, risk management

INTRODUCTION

Radical changes in federal dairy policies have resulted in significant increases of milk price volatility in the United States (USDA, 1999). Dairy operations are now facing important income risk management questions. Issues of liquidity and cash flows are often as important as profitability (St-Pierre et al., 2000). New futures and options markets are now available as instruments to manage the variance associated with output (milk) price risk. However, gross income is the product of price per unit multiplied by the number of units produced (yield). Therefore, it may be as important to have the ability to forecast the number of production units (cows) and their productivity (pounds of milk per cow) as it is to forecast the price of milk. Additionally, a forecast must include an estimate of expectation (mean) and dispersion (variance) if it is to be useful in risk management. A forecasting model of milk and dairy products prices has been developed using a Bayesian vector autoregression approach (Petrov, 1999). The model appears to generate forecast errors of less than 5% of milk prices over a forecast horizon of up to 6 mo. However, there is no model available to forecast short- and medium-term herd structure and productivity (expectations and variances). The objectives of this research were: 1) to develop a discrete, stochastic forecasting model using finite Markov chains, 2) to determine an algorithm for the estimation...
of the model’s parameters, and 3) to derive procedures for calculating the variance of the forecast.

A DISCRETE MARKOV MODEL

The model segments the herd in multiple disjointed sets, each representing a group of animals of uniform physiological status and with identical production ability (in a probabilistic sense). Animals of milking age are assigned to a four-dimensional matrix with the following four attributes: parity (PAR = 1,2,3), pregnancy status (PREG = 0,1), week in milk (WIM = 1,2,...,104), and week pregnant (WP = 1,...,40). Growing animals are assigned to a three-dimensional matrix with indices: week of growth (WG = 1,...,156), pregnancy status (PREG = 0,1), and week pregnant (1,...,40). The conceptual structure of the model is presented in Figure 1. At time t, a cell for the milking herd contains the expected number of animals (of PAR = i₁, PREG = i₂, WIM = i₃, and WP = i₄), the variance of this expected number, the expected milk production per animal in the cell, and its variance. Forecasting is done by advancing time (weekly steps) and transferring virtual animals using finite Markov chains (Hernandez-Lerma, 1989; Van Dijk, 1984).

Based on biology, a number of cells are never populated (e.g., cows at WIM = 3 and WP = 5) and are ignored in the computer implementation of the model. Likewise, of the more than 10,000 mathematically possible chains linking a cell (i₁, i₂, i₃, i₄) at time t to all cells at time t + 1, only a few can physically or biologically take place. Those are represented symbolically in Table 1. A maximum of four chains leave a cell at time t to time t + 1. The proportion of animals moving from a cell at time t to another cell at time t + 1 is determined from a transition probability kᵢ with 0 ≤ kᵢ ≤ 1. The sum of all kᵢ from a given cell at time t must be equal to 1 (i.e., all animals must be accounted for from one period to the next). Figure 2 is a diagram of the processes of aging and culling from time t to time t + 1 for a specific cell, in this example, for first-parity animals, open in their first WIM at time t. There is a probability that some animals will be culled, which is set to 0.75% in this example. This rate reflects a 1.5% death loss in the first 2 wk of lactation. The other transition probability affecting the cell is that some cows will increase WIM by 1 wk without becoming pregnant. In the example, this probability accounts for the remaining animals (100 – 0.75%) or 99.25%.

Although there is a maximum of four chains leaving a given cell at time t, the number of chains arriving at a given cell at time t + 1 is variable, but often equal to one (Table 2). All cells with a PREG index equal to 1 at time t (pregnant) receive only one chain from time t – 1. There are 6240 cow cells and 3120 replacement heifer cells with a PREG index equal to 1. This has an important repercussion on the size of the problem. The model has a grand total of 6554 cow cells and 3276 replacement heifer cells. Each of these could, from an algebraic standpoint, be connected to 9830 cells from time = t to time = t + 1, which would require the estimation of 96,628,900 transition probabilities. Fortunately, biology reduces considerably the number of actual transition probabilities.

EXPECTATION AND VARIANCE OF MILK PRODUCED

The expected herd milk production (M) at time t (Mᵢ) is simply the sum of the expected milk production of all cells:
Table 1. Markov chains required to represent aging, pregnancy, culling, abortion, dry-off, and parturition.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Pregnancy</th>
<th>Actions1</th>
<th>Symbolic2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milking</td>
<td>0</td>
<td>a</td>
<td>WIM + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p</td>
<td>PREG = 1, WIM + 1, WP + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>Cull cell</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>a</td>
<td>WIM + 1, WP + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>PREG = 0, WIM + 1, WP = 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>d</td>
<td>WIM = 0, WP + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>Cull cell</td>
</tr>
<tr>
<td>Dry</td>
<td>1</td>
<td>a</td>
<td>WP + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>f</td>
<td>PREG = 0, WIM = 1, WP = 0, PAR + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>Cull cell</td>
</tr>
<tr>
<td>Replacement heifer</td>
<td>0</td>
<td>a</td>
<td>WG + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p</td>
<td>PREG = 1, WP = 1, WG + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>Cull cell</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>a</td>
<td>WP + 1, WG + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>PREG = 0, WP = 0, WG + 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>f</td>
<td>PREG = 0, WP = 0, PAR = 1, WIM = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>Cull cell</td>
</tr>
</tbody>
</table>

1 = Aging, b = abort, c = culling, d = dry-off, f = freshening, and p = initiates pregnancy.
2PREG = Parity (1,2,3), PREG = Pregnancy status (0,1), WG = week of growth (1, ..., 156), WIM = week in milk (1, ..., 104), WP = week pregnant (1, ..., 40), and cull cell = stack counting the number of animals culled.

\[
E(M^t) = \sum_j E(M^t)_j \quad [1]
\]

and because cells are disjointed, the variance of \( M^t \) is simply

\[
VAR(M^t) = \sum_j VAR(M^t)_j \quad [2]
\]

This equation states that if we can calculate the estimated milk production and variance of each cell, then we can easily calculate the expected milk production and variance of the whole herd. We are left with the complex problem of estimating expectation and variance of milk production for each cell.

At time \( t \), the expected milk production of a given cell \( j \) with indices \((PAR = i_1, PREG = i_2, WIM = i_3, \text{ and WP} = i_4)\) is

\[
E(M^t_j) = (nt_{i_1,i_2,i_3,i_4} - 1) \times k^a_{i_1,i_2,i_3,i_4} \times MPC_{i_1,0,i_3,0} + \ldots + (nt_{i_1,i_2,i_3,i_4} - 1) \times k^b_{i_1,i_2,i_3,i_4} \times MPC_{i_1,0,i_3,0} \quad [3]
\]

if \PREG = 0 at time \( t \);

\[
E(M^t_j) = (nt_{i_1,i_2,i_3,i_4} - 1) \times k^a_{i_1,i_2,i_3,i_4} \times MPC_{i_1,0,i_3,1} \quad [4]
\]

if \PREG = 1 and WP > 1 at time \( t \);

\[
E(M^t_j) = (nt_{i_1,i_2,i_3,i_4} - 1) \times k^P_{i_1,i_2,i_3,i_4} \times MPC_{i_1,1,i_3,1} \quad [5]
\]

if \PREG = 1 and WP = 1 at time \( t \) where

- \( n_{i_1,i_2,i_3,i_4} \) = number of cows in cell \((i_1, i_2, i_3, i_4)\) at time \( t-1 \);
- \( k^a_{i_1,i_2,i_3,i_4} \) = transition probability of simply aging by a week for cows in cell \((i_1, i_2, i_3, i_4)\) at time \( t-1 \);
- \( k^b_{i_1,i_2,i_3,i_4} \) = transition probability of aborting for cows in cell \((i_1, i_2, i_3, i_4)\) at time \( t-1 \);
- \( k^P_{i_1,i_2,i_3,i_4} \) = transition probability of becoming pregnant for cows in cell \((i_1, 0, i_3, 0)\) at time \( t-1 \), and
- \( MPC_{i_1,i_2,i_3,i_4} \) = milk per cow for cows in cell \((i_1, i_2, i_3, i_4)\) at time \( t \).

These equations are really not as complex as they look. For example, using equation [4] to calculate the expected production from first-parity cows, pregnant, 20 WIM, and 10 wk pregnant at time \( t \), we have

---

**Figure 2.** Diagram representing the process of aging or culling from a cell at time \( t \) to corresponding cells at time \( t + 1 \).
where \( Y_{ijklm} \) is the TD milk measurement for lactation \( i \) on TD \( j \) for a cow that is in age class \( k \) and season \( \times \) WIM \( \times \) parity class \( (P_i) \) on TD, and \( e_{ijklm} \) is the residual error. Similar to Stanton et al. (1992), the random effect \( L_i \) assumes an average repeatability between TD of 0.46. The fixed effect of age is represented by 60 monthly classes, whereas seasons of calving are defined as winter (September through March) and summer (April through August). Test day is considered a random effect to reflect its random nature in the forecast of future milk production.

The derivation of the variance of \( M^t \) is more complex because it involves the product of three random variables (n, k, and MPC). To simplify the notation, we relabeled the three random variables n, k, and MPC as \( X_1, X_2, \) and \( X_3 \). Without loss of generality, we can suppose that \( X_i \) has mean \( \theta_i \) and \( \text{VAR}(X_i) = \sigma_i^2 \). Consider again the function \( g(X_1, X_2, X_3) = X_1 \times X_2 \times X_3 = g(X) \) for brevity. If \( g_1(\theta) \) is \( \delta g(X)/\delta X_i \) evaluated at \( \theta_1, \theta_2, \) and \( \theta_3 \), then we have a Taylor expansion:

\[
g(X) = g(\theta) + \sum_{i=1}^{3} g_1(\theta) \cdot (X_i - \theta_i) + 0(n^{-1}) \quad [8]
\]

where \( 0(n^{-1}) \) is some unknown residual function, assumed to be negligible in the vicinity of \( \theta_i \). By definition

\[
\text{VAR}(g(X)) = \text{E}\{\sum_{i=1}^{3} (g_1(\theta) \cdot (X_i - \theta_i))^2\} + 0(n^{-1}) \quad [9]
\]

\[
= \sum_{i=1}^{3} (g_1(\theta))^2 \cdot \text{VAR}(X_i) \quad [10]
\]

Thus,

\[
\text{VAR}(X_1 \times X_2 \times X_3) = \theta_1^2 \theta_2^2 \sigma_1^2 + \theta_1^2 \theta_3^2 \sigma_2^2 + \theta_2^2 \theta_3^2 \sigma_3^2 + \theta_1^2 \theta_2^2 \sigma_1^2 + \theta_1^2 \theta_3^2 \sigma_2^2 + \theta_2^2 \theta_3^2 \sigma_3^2 \quad [11]
\]

It should be apparent that the distribution of \( M^t \) is not normal because it is the product of a binomial variate \((n)\), a beta variate \((k)\), and a normal variate \((\text{MPC})\). However, because \( M^t = \sum J_j \) for which \( j \) is a relatively large number (>1000), then \( M^t \) will tend to normality because of a Central Limit effect. Variables \( n_{ijklm} \) are calculated recursively. In the initial period, they are set as constants based on the current status of cows in the herd. The \( \text{E}(\text{MPC}_{ijklm}) \) and \( \text{VAR}(\text{MPC}_{ijklm}) \) are estimated from herd historical production data using a mixed model as described previously. The estimation of
transition probabilities, however, proved to be difficult from historical data. Because of the large number of cells, the matrix is generally sparse and, in many instances, frequencies are too low to generate useful and accurate estimates. In general, precision of parameter estimates increases with the number of observations. The amount of data required to obtain a satisfactory parameter estimate is dependent on the parameter, the assumed distribution, and, of course, the degree of precision required for a satisfactory status. Our model requires the estimation of numerous transitional probabilities for the implementation of Markov chains. Often, transition probabilities must be estimated from just a few observations; there are instances where such probability parameters must be estimated in the absence of prior data. Lastly, estimations of transition probabilities are not independent from one another. For example, the probability that open cows become pregnant between WIM $i_2$ and $i_2 + 1$ is not independent of the probability that open cows become pregnant between WIM $i_2$ and $i_2 - 1$. That is, the conception rate should not vary widely from week to week but should smoothly vary through WIM. In the next section, we describe a procedure for estimating transition probabilities using a Bayesian approach with prior distributions to correct for sparsity of data.

**BAYESIAN ESTIMATES OF TRANSITION PROBABILITIES**

Bayesian estimation theory and methods are discussed in numerous textbooks (Berger, 1985; Carlin and Louis, 1996). Because few animal scientists are exposed to Bayesian statistics, it is important to describe our methods in greater details than required for those already familiar with Bayesian statistics.

**Specifying a Prior Distribution**

Consider the problem of statistical inference in which observations are to be taken from a distribution for which the probability distribution function (PDF) is $f(x | \Theta)$, where $\Theta$ is a parameter having an unknown value. It is assumed that the unknown value of $\Theta$ must lie in a specified parameter space $\Omega$. In many instances, before any observations from $f(x | \Theta)$ are available, we can summarize our previous information and knowledge about where in $\Omega$ the value of $\Theta$ is likely to lie by constructing a probability distribution for $\Theta$ on the set $\Omega$. In other words, outside prior sources of knowledge indicate that $\Theta$ is more likely to lie in certain regions of $\Omega$ than others. The relative likelihood of the different regions can be expressed in terms of probability distribution of $\Theta$, because it represents the relative likelihood that the true value of $\Theta$ lies in each of various regions of $\Omega$ prior to observing any values from $f(x | \Theta)$.

**Controversial Nature of Prior Distributions**

The statistical concept of prior distribution is very controversial. This controversy is closely related to the meaning of probability. Some statisticians believe that a prior distribution can always be chosen for the parameter $\Theta$. They believe that this distribution is a subjective probability distribution in the sense that it represents subjective beliefs about where the true value of $\Theta$ is likely to lie. They also think that a prior distribution is no different from any other statistical probability distribution. These people adhere to the Bayesian philosophy of statistics.

Other statisticians believe that in many problems, it is not appropriate to speak of a probability distribution of $\Theta$ because the true value of $\Theta$ is not a random variable at all but a fixed number whose value happens to be unknown. These people believe that a prior distribution can be assigned to a parameter $\Theta$ only when there is extensive previous information about the relative frequencies of $\Theta$ estimates.

The method that we propose for the estimation of transition probabilities is Bayesian because it assumes that $\Theta$ follows a prior distribution. The method, however, uses objective distribution estimates as opposed to subjective estimates. This approach is generally referred to as Empirical Bayes estimation (Berger, 1985; Carlin and Louis, 1996).

**The Posterior Distribution**

Suppose that $n$ random variables, $x_1, \ldots, x_n$, form a random sample from a distribution for which the PDF is $f(x | \Theta)$. The value of the parameter $\Theta$ is unknown and the prior PDF of $\Theta$ is $\xi(\Theta)$. Because the random variables $x_1, \ldots, x_n$ form a random sample from the distribution for which the PDF is $f(x | \Theta)$, it follows that their joint PDF is:

$$f_n(x_1, \ldots, x_n | \Theta) = f(x_1 | \Theta) \cdots f(x_n | \Theta) \quad [12]$$

If we use the vector notation $x = (x_1, \ldots, x_n)$, then the joint PDF can be written as $f_n(x | \Theta)$.

In a Bayesian context, $\Theta$ itself has a distribution for which the PDF is $\xi(\Theta)$. The n-dimensional marginal joint PDF $g_n(x)$ of $x_1, \ldots, x_n$ can be written in the form:

$$g_n(x) = \int f_n(x | \Theta) \xi(\Theta) \, d\Theta \quad [13]$$

Also, the conditional PDF of $\Theta$ given that $X_1 = x_1, \ldots, X_n = x_n$, which is denoted by $\xi(\Theta | x)$, must be equal
to the joint PDF of \( X_1, \ldots, X_n \) and \( \Theta \) divided by the marginal joint PDF of \( X_1, \ldots, X_n \). Therefore, we have:

\[
\xi(\Theta | x) = \frac{f_n(x | \Theta) \xi(\Theta)}{g_n(x)} [14]
\]

This is called the posterior distribution function of \( \Theta \) because it is the distribution of \( \Theta \) after the values of \( X_1, \ldots, X_n \) have been observed.

**The Likelihood Function**

The denominator on the right side of equation [14] is simply the integral of the numerator over all possible values of \( \Theta \). Although the value of this integral depends on the observed values \( X_1, \ldots, X_n \), it does not depend on \( \Theta \), and it may be treated as a constant when the right side of the equation is regarded as a PDF of \( \Theta \). Therefore, we may write the following relation:

\[
\xi(\Theta | x) \propto f_n(x | \Theta) \xi(\Theta) [15]
\]

When the joint PDF of the observations in a random sample is regarded as a function of \( \Theta \) for given values of \( X_1, \ldots, X_n \), it is called the likelihood function. In this terminology, the relation in [15] states that the posterior PDF of \( \Theta \) is proportional to the product of the likelihood function and the prior PDF of \( \Theta \).

**Conjugate Prior Distributions**

One last aspect needs to be explained before we can describe the Bayesian estimation of transitional probabilities. In working with the likelihood function [15], certain prior distributions are particularly convenient for use with samples from certain other distributions. For example, if a random sample is taken from a Bernoulli distribution for which the value of the parameter \( \Theta \) is unknown, and if the prior distribution of \( \Theta \) is a beta distribution, the posterior distribution of \( \Theta \) will again be a beta distribution.

**EXPECTATIONS AND VARIANCES OF TRANSITION PROBABILITIES**

**Prior Distribution**

In a Markov chain,

\[
n_{i+1} = n_i \times k_i, [16]
\]

where \( i \) indicates the time period and \( n_i \) is the number of objects connecting time \( i \) to time \( i + 1 \). With this formulation, the \( n_{i+1} \) are assumed to follow a Bernoulli distribution with parameter \( \Theta_i \). In a Bayesian sense, \( \Theta_i \) is a random variable that is assumed to follow a beta distribution. The first step in our procedure is to calculate the values of parameter \( \alpha_i \) and \( \beta_i \) of a beta distribution based on the knowledge from all periods \( 1, 2, \ldots, i \), scaled to the average knowledge at any given period.

In our model, events are modeled as a series of Markov chains. For example, Figure 3 illustrates the series of Markov chains that describes the establishment of pregnancy for animals of parity \( i \). Let \( n_i \) be the number of animals in cell \( i \) at time \( t-1 \) and \( m_i \) be the number of animals in cell \( i \) that became pregnant at time \( t \). Let the initial values of \( \alpha \) and \( \beta \) be:

\[
\alpha_0 = \sum_{i=1}^{n} m_i [17]
\]

\[
\beta_0 = \sum_{i=1}^{n} n_i - \sum_{i=1}^{n} m_i = \sum_{i=1}^{n} n_i - \alpha_0. [18]
\]

An initial estimate for \( k_i \) could be calculated as follows:

\[
k_i^0 = \frac{\alpha^0}{\alpha^0 + \beta^0} [19]
\]

where \( k_i^0 = k_{i-1}^0 = \ldots = k_0^0 \).

These estimates, however, carry too much weight because many of the prior observations at different time periods came from the same animals. Prior estimates are therefore weighted by the average number of observations at each time period. The prior estimates of \( \alpha \) (\( \alpha^{Pr} \)) and \( \beta \) (\( \beta^{Pr} \)) are then:

\[
\alpha^{Pr} = \frac{(\alpha^0)}{\alpha^0 + \beta^0} \times \frac{(\alpha^0 + \beta^0)}{i-1} = \frac{\alpha^0}{i-1} [20]
\]

Likewise;
Table 3. Example of the empirical Bayes estimation procedure applied through five week-in-milk periods.

<table>
<thead>
<tr>
<th>Periods</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_i</td>
<td>100</td>
<td>92</td>
<td>80</td>
<td>50</td>
<td>50</td>
<td>372</td>
</tr>
<tr>
<td>m_i</td>
<td>8</td>
<td>5</td>
<td>12</td>
<td>0</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>Uncorrected k_i</td>
<td>0.080</td>
<td>0.054</td>
<td>0.150</td>
<td>0.000</td>
<td>0.080</td>
<td>0.080</td>
</tr>
</tbody>
</table>

$\alpha^0 = 29$  
$\beta^0 = 343$  
$\alpha^P = 29 + 5 = 5.8$  
$\beta^P = 343 + 5 = 68.6$

$\alpha^P = 13.8$  
$\beta^P = 160.6$  
$E(k_i^P) = 4.15 \times 10^{-4}$  
$\text{Var}(k_i^P) = 0.078$

$\beta^P = \frac{\beta^0}{i - 1}$  
$E(k_i^P) = \frac{\alpha^P}{\alpha^P + \beta^P} = 14\%$

that is, the prior estimates of all $k_i$ are all the same and each follows a beta distribution with parameters $\alpha^P$ and $\beta^P$.

Posterior Distribution

The posterior distributions of each $k_i$ follows a beta distribution with the following parameters:

$\alpha_i^P = \alpha^P + m_i$  
$\beta_i^P = \beta^P + (n_i - m_i)$

that is, the posterior is based on the prior (same for all $k_i$), plus the observations specific to each period. For those periods without observation, the posterior is equal to the prior (the average transition probability). For those periods with substantially more observations than the average period, the posterior will be largely dependent on the observations with little weight put on the prior. Under this procedure, transition probabilities across time periods are relatively smooth unless solid evidence from observations points to the contrary.

Therefore the posterior distribution of $k_i$, noted as $k_i^P$, is a beta distribution with an expectation (mean) and variance:

$E(k_i^P) = \frac{\alpha^P}{\alpha^P + \beta^P}$  
$\text{Var}(k_i^P) = \frac{\alpha^P \beta^P}{(\alpha^P + \beta^P)^2}$

Table 3 provides an example of the application of the Bayes procedure for estimating transition probabilities. In this example, the number of periods is limited to five for the reason of simplicity.

EXAMPLE

The model was used to produce a production forecast for a herd operating in the Eastern United States. At time $t = 0$, the herd consisted of 494 total cows of which 426 were milking. Milk sales totaled 92,990 kg for the week. Figure 4a shows the forecasted number of milking animals over the following 50 wk. The 95% confidence region is also shown. The herd had expanded in the prior year by purchasing 100 milking animals. The model projects a 14% reduction in the number of milking cows over the next 20 wk. This number is expected to plateau after 30 wk at approximately 380 cows with a lower 95% confidence band at 360 cows.

The total milk production forecast is shown in Figure 4b. According to the forecast, weekly production is expected to decline from the current 93,000 kg/wk to less than 70,000 kg/wk over the course of the next 30 wk. The decline in milk production around wk 30 is counter-intuitive considering the large number of cows projected to calve between wk 27 and 31 (Figure 5). The majority of these calvings, however, involve first-lactation animals that produce less per week in their first month of lactation than the herd average. This effect is evident in Figure 4c where the forecast for milk production per cow is shown.

VALIDATION

Model validation raises an interesting question: how do you validate a model where dynamic changes in the
target values are based on information related to the forecast? For example, assume that the model is forecasting a reduction in production 6 mo from now due to a smaller number of cows in production. Sometime during the next 6 mo (and likely before 6 mo lapse), the manager will realize that he is facing a reduction in production and he will take action to prevent or correct this drop. A combination of culling and purchasing of animals will take place, with the result that the forecasted drop in production will not occur, although the forecast may have been very accurate. This problem is similar in nature to that of forecasting aircraft-landing sites. Most airplanes eventually land where they were supposed to even though a model could forecast that they were about to miss the runway from 500 miles out. The problem is not with the forecast but with the pilot who, fortunately for the passengers, takes corrective action so that the plane lands on the runway. The value of an advanced forecast is that the plane will follow a more direct and economical route to the landing site. Likewise, an advance forecast would minimize costs associated with corrective actions in the physical production of milk. Wells (1996) derived a cost function that incorporates both system costs and dynamic costs of a dynamic model of financial markets. Using this information, the author showed that his model was in fact the least cost trajectory. The same approach could be used in our application. Alternatively, engineering techniques used for missile guidance systems could be expanded to our model validation (Hernandez-Lerma, 1989, 1996; Yin, 1998).

**APPLICATION**

Most commercial farms already have in place an electronic production database either through their Dairy Herd Improvement Association or as a stand-alone system operating on a personal computer or as a specialized commercial hardware/software product. Thus, the acquisition of the data used for estimating the parame-

Figure 4. Fifty-week forecast of the number of milking cows (a), total weekly milk production (b), and milk production per cow for the example herd. Shown are the expected values (•–•) with their 95% lower and upper confidence range (○–○).
ter of model [7] could be done automatically and electronically. Although difficult just a few years ago, the estimation of the fixed and random parameters of model [7] is now commonly done and various commercial software programs are available to accomplish this. In the introduction, we referred to a forecasting model of milk and dairy products prices that we have developed using a Bayesian vector autoregression approach (Petrov, 1999). This model requires price series that are electronically available from the USDA. Solutions to the autoregressive model can again be found using various commercial software programs. The whole process needs to be integrated, possibly as stand-alone software operating on a personal computer or as a web-based system accessible through the Internet. Dairy producers and their advisors could then examine the impact of various culling, animal purchasing, and milk marketing strategies and tactics on the expected future income and its variance.

CONCLUSION

Dynamics of a herd through time can be represented successfully as finite Markov processes. Problems associated with the estimation of a large number of transition probabilities from sparse data are completely resolved using empirical Bayes methods. We were also successful at deriving an equation to approximate the variance of the product of three independent random variables. The model's forecasting behavior is in line with prior expectations for such models in that forecasting variance increases monotonically with time. Further work is needed on validation procedures.

REFERENCES